รายงานวิจัยฉบับสมบูรณ์

การพัฒนาวิธีการใหม่เพื่อรบกวนกระบวนการประกอบอนุภาค และกระบวนการสร้างไวรัสเอชไอวีที่สมบูรณ์ด้วยโปรตีนแองไครินรีพีท

ชัชชัย ตะยาภิวัฒนา

สิงหาคม 2557

สัญญาเลขที่ **BRG5480004**

รายงานวิจัยฉบับสมบูรณ์

การพัฒนาวิธีการใหม่เพื่อรบกวนกระบวนการประกอบอนุภาค และกระบวนการสร้างไวรัสเอชไอวีที่สมบูรณ์ด้วยโปรตีนแองไครินรีพีท

ชัชชัย ตะยาภิวัฒนา
แขนงวิชาภูมิคุ้มกันวิทยาคลินิก
ภาควิชาเทคนิคการแพทย์
คณะเทคนิคการแพทย์
มหาวิทยาลัยเชียงใหม่

สนับสนุนโดยสำนักงานกองทุนสนับสนุนการวิจัย

กิตติกรรมประกาศ

งานวิจัยนี้ได้รับทุนสนับสนุนประเภท ทุนพัฒนานักวิจัย ประจำปี 2553 จากสำนักงานกองทุนสนับสนุนการวิจัย ผู้วิจัยขอขอบพระคุณมา ณ ที่นี้ด้วย

ผู้วิจัยขอขอบคุณ คณาจารย์, บุคลากร และนักศึกษา ภาควิชาภูมิคุ้มกันวิทยาคลินิก ที่มีส่วนร่วมอำนวยความสะดวก และสนับสนุนในเชิงวิชาการ โดยเฉพาะ รศ.ดร.วัชระ กสิณฤกษ์ ที่ให้แนวคิดและสนับสนุนงบประมานส่วนหนึ่งสำหรับการวิจัยในครั้งนี้ นอกจากนี้ผู้วิจัยขอขอบคุณเจ้าหน้าที่ของสำนักงานกองทุนสนับสนุนการวิจัยที่ช่วยประสานงานทุกด้าน

ท้ายนี้ขอขอบพระคุณ ศ.ดร.สถิตย์ สิริสิงห และ รศ.ดร.กำจัด มงคลกุล ซึ่งเป็นทั้งผู้ให้โอกาสและเป็นต้นแบบในการทำงานทางวิทยาศาสตร์ของผู้วิจัย

INDEX

Abstract	1
Introduction	2
Objective	8
Materials and Methods	9
Results and Discussion	14
Conclusion	48
References	49
List of Publications	52

ABSTRACT

Project Code : BRG5480004
Project Title : Implement of novel strategy to interfere the assembly and maturation processes of human immunodeficiency virus (HIV-1) by artificial ankyrin repeat proteins
Investigator : Mr. Chatchai Tayapiwatana
Div. of Clinical Immunology, Fac. of Associated Medical Sciences
Chiang Mai University
E-mail Address : asimi002@hotmail.com
Project Period : 3 year

Acquired Immune Deficiency Syndrome (AIDS) is a viral infectious disease that has threatened human lives for decades. Despite the attempts to fight this deadly disease, more than one-in-hundred adults in Thailand is infected with HIV. The attempts to cure the disease usually involve interruption of HIV life cycle which is a sequence of steps regulated both by viral and cellular proteins. The highly active antiretroviral therapy (HAART) in controlling HIV-1 is success. However, HAART is not totally effective in all cases especially in the multi-drug resistant, and has problematic side effects. Hence, several strategies have been emerged to apply as the novel therapeutic approaches including amalgamations of gene- and immune-therap. The most popular therapeutic molecules applied for this purpose are immunoglobulin fragments i.e. single chain variable fragment (ScFv) as intrabody format. However, due to the complex structures their intracellular functions are limited in certain compartments i.e. cytoplasm.

According to the increasing knowledge of protein-protein interactions and the development of advance selection technologies, novel binding molecules based on protein 'scaffolds' have been explored to mimic the binding principle of immunoglobulins with virtually target proteins. One of proteins in this architecture, the ankyrin repeat proteins, is very attractive scaffold to generate the specific molecular binders. These proteins mediate certain important protein-protein interactions in virtually all species and are found in all compartments of the cells indicating that these proteins can adapt for many different environments. In this study, the ankyrin repeat proteins will be created the artificial protein library and specific binding molecules to HIV-1 matrix and capsid domain (MA-CA) will be isolated. In contrast to ScFv intrabody, the isolated ankyrin binders will naturally fold and be active in cytoplasm. These binders will be applied to evaluate their intracellular function in the interference of viral life cycle. The successful innovation will be an alternative strategy for HIV gene therapy in the future.

Keywords: HIV, Ankyrin, intracellular, gene therapy

INTRODUCTION

Acquired Immune Deficiency Syndrome (AIDS) is a viral infectious disease that has threatened human lives for decades. It is caused by the Human Immunodeficiency Virus (HIV). Despite the attempts to fight this deadly disease, more than one-in-hundred adults in Thailand is infected with HIV. The attempts to cure the disease usually involve interruption of HIV life cycle which is a sequence of steps regulated both by viral and cellular proteins. The highly active antiretroviral therapy (HAART) in controlling HIV-1 is success. However, HAART is not totally effective in all cases especially in the multi-drug resistant, and has problematic side effects. Hence, several strategies have been emerged to apply as the novel therapeutic approaches including amalgamations of gene- and immune-therapy [1-5]. The most popular therapeutic molecules applied for this purpose are immunoglobulin fragments i.e. single chain variable fragment (ScFv) as intrabody format. However, due to the complex structures their intracellular functions are limited in certain compartments i.e. cytoplasm.

Since the increasing knowledge of protein-protein interactions and superior understanding of protein engineering and the further development of advance selection technologies, novel binding molecules based on protein frameworks 'scaffolds' have been explored to mimic the binding principle of immunoglobulins with virtually target proteins. Several protein-binding scaffolds have been generated to overcome the limitation of antibody and its derivative molecules in terms of stability, ease of modification, robustness and cost-efficient production. Nowadays, different types of scaffold proteins have been successfully generated the artificial library and isolated the specific binding molecules with various targets [6-10]. However, the binding pocket of different protein scaffolds is unique. It might be more or less suited with some targets depending on the nature of the target molecules or their epitope to which the binding protein should be directed [9].

Focusing on repeat proteins, the repeating structural units (repeats) stack together to form elongated protein domain (repeat domains). The binding surface of this scaffold is variable in size as the number of repeats can be varied [11-13]. One of proteins in this architecture, the ankyrin repeat proteins, is very attractive scaffold to generate the specific molecular binders. These proteins mediate many important protein-protein interactions in virtually all species and are found in all compartments of the cells indicating that these proteins can adapt to many different environments. Moreover, this scaffold have been generated the large library and isolated several specific binding molecules. Therefore, it is of interest applying this scaffold to gain the desired intracellular functions.

In this study, the ankyrin repeat proteins will be created the artificial protein library and specific binding molecules to HIV-1 matrix and capsid domain (MA-CA) will be isolated. In contrast to ScFv intrabody, the isolated ankyrin binders will naturally fold and be active in cytoplasm. These binders will be applied to evaluate their intracellular function in the interference of viral life cycle. The successful innovation will be an alternative strategy for HIV gene therapy in the future.

The gene therapy for the HIV/AID treatment

The success of highly active antiretroviral therapy (HAART) in controlling HIV-1 offers hope, but the difficulties of this therapy and the ability of HIV-1 to mutate into drug-resistant variants necessitate continuous development of new therapies. In addition, HAART is not totally effective and has problematic side effects. Drug-resistant HIV-1 is increasingly frequent [14-16]. Investigation into additional therapeutic approaches should therefore

continue. The gene therapy will probably not replace pharmacotherapy, but may rather play an important supporting role. The goals of anti-HIV-1 gene therapy are to deliver transgenes: (a) to hematopoietic progenitor cells (HSC) to protect their differentiated progeny from HIV-1; (b) directly to HIV-1-susceptible cells, to render them resistant to HIV-1 infection or inhibit HIV-1 replication in them; (c) to immunize against HIV-1 antigens; and (d) to inhibit HIV-1 in discrete organ target sites (e.g., central nervous system) [17]. Many different transgenes have been reported. Protein-based anti-HIV-1 strategies have been the single largest area of anti-HIV-1 gene transfer trials in humans [18, 19]. Briefly, protein structures developed to inhibit HIV-1 include transdominant negative mutants, intrakines, toxins, single-chain antibodies, and DNA-based vaccines. Indeed, very recent works suggest that intracellular HIV-1-specific single-chain variable fragment antibodies (ScFv) can target and redirect essential HIV-1 proteins away from required subcellular compartments and block the function or processing of such essential proteins as HIV-1 gp120 [20], Rev[21], Gag [22], reverse transcriptase (RT) [23], and integrase (IN) [24].

An alternative non-antibody scaffolds for novel binding functions

Over the past three decades, monoclonal antibodies have developed into the favorable binding proteins and currently become various applications in research, diagnostics and therapy. However, several applications require high specificity and a defined molecular composition. It has become obvious that the whole molecule of antibodies suffers from some fundamental disadvantages for example in the therapeutic; their Fc regions are not really required. The constant Fc region mediates immunological effector functions and often leads to undesired interactions and rather their large molecules limited tissue penetration [25, 26]. The engineered antibodies in the form of ScFv, Fab and multivalent fragments have been obtained from synthetic libraries or B cells [27-29]. They are made of two different polypeptides, the light (V_L) and heavy (V_H) chains. The complex architecture of their antigen-binding site, which is formed by six hypervariable loops, is difficult to manipulate and also requires complicated cloning steps for recombinant expression and the generation of synthetic libraries. Moreover, the stability of these molecules relies on disulfide bonds which can occur in the appropriate condition in the cells [30].

These limitations and the increased understanding of other natural binding proteins inspired scientists to transfer the concept of a universal binding site from the antibody structure to alternative protein frameworks (‘scaffolds’). The term ‘scaffold’ in the meaning of protein engineering has been described as a polypeptidic framework with a high tolerance of its fold for modifications such as multiple insertions, deletions or substitutions. This conformational stability enables the directed randomization and drastic change within a defined region of the protein to gain the certain novel properties whereas its structure and original physicochemical behavior remains conserved [31]. The generation of novel binding molecules based on protein frameworks is a concept that has been strongly promoted during the past decade years. At present, the different types of protein scaffolds that have been successfully exploited for the construction of artificial binding proteins can be classified into the following groups as showed in Figure 1.

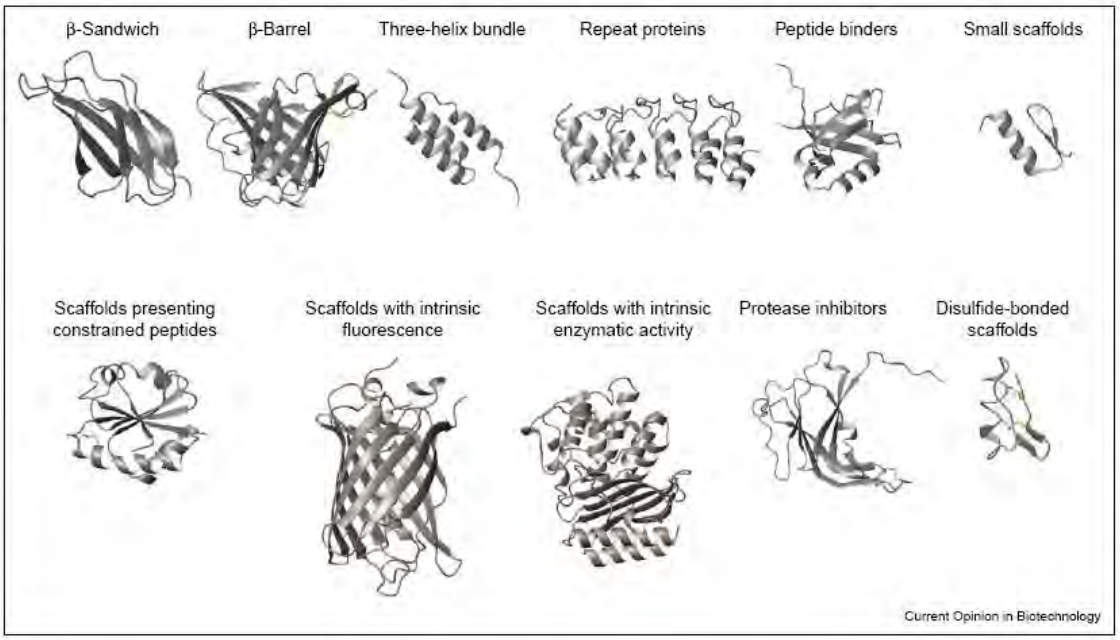

Figure 1. The different protein backbones used as scaffolds for the generation of protein-binding agents, classified in groups [6].

In addition to the specific binding, these artificial binding proteins provide favorable characteristics such as robustness, ease of modification and cost-efficient production.

Evaluation of repeat proteins

Repeat proteins, the most abundant natural protein classes specialized in binding property. They are found in all phyla and involved in diverse biological processes, such as cell cycle control, transcriptional regulation, cell differentiation, cellular scaffolding or bacterial invasion [13, 32-34]. Their unique architecture features repeating structural units which stack together to form elongated repeat domains displaying variable and modular target-binding surfaces [35]. The tertiary structures of several proteins with structural repeats are shown in Figure 2.

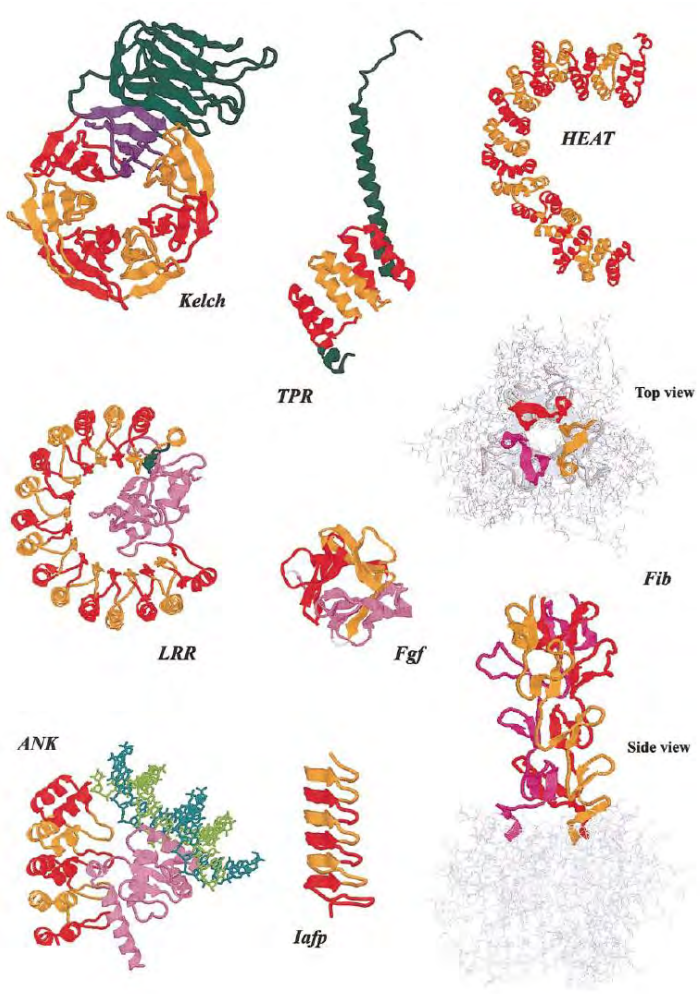

Figure 2. The architecture of several repeat proteins [35].

The other scaffold motifs such as affibodies [36], lipocalins [37], green fluorescent protein (GFP) [38] and fibronectin type III domains (FNIII) [39] are the binding molecule which have the functionally partitioned protein architectures consisting of a structural framework and variable target-binding surface loops. These scaffolds binding surface is limited by the size. Therefore, the repeat proteins have evolved another successful binding strategy. They feature repeating structural units stack together to form elongated protein domains with a continuous target-binding surface which is variable in size as the number of repeats can be varied. For such open structures, there is no theoretical limit on their repeat number since incremental addition of repeats is not sterically hindered. These rod-like or superhelical structures present an extensive solvent-accessible surface that is well suited to binding large substrates such as proteins and nucleic acids [35].

The ankyrin repeat protein

In 1987, Breeden and Nasmyth reported a ~33 residue repeating motif in the sequence of two yeast cell-cycle regulators, Swi6p and Cdc10p, and in the Notch and LIN-12 developmental regulators from *Dorsophila melanogaster* and *Caenorhabditis elegans* [40]. Subsequently, the discovery of this sequence containing 24 copies in cytoskeletal protein ankyrin led to the naming of this motif as the ankyrin (ANK) repeat. The ankyrin repeat proteins carry out a wide variety of biological activities and have been detected in organisms ranging from viruses to humans. These molecules are present in the nucleus, cytoplasm and

the extracellular milieu. Although, some molecules appear to contain only ANK repeat, others contain insertions between the repeats. Recently, more sophisticated homology search algorithms have now identified ~19,276 sequences in 3608 proteins identified from the non-redundant protein database (http://smart.embl-heidelberg.de/help /smart_glo ssary.shtml) [41]. The ANK repeat is a motif of 33 amino acid residues and its defined structure exhibits a canonical helix-turn-helix conformation, in which two antiparallel α-helices stacked side by side and connected by a series of interventing β-hairpin motifs. The extended β-sheet projects outward at an approximately 90º angle from the helical pairs resulting in a characteristic L-shaped cross-section (Figure 3).

Figure 3. The characteristic of ankyrin repeat (ANK) protein architecture. The arrangement of α-helices and β-hairpins was sown as cylinders and arrows respectively [13].

From the statistical analyses of ANK sequences by Mosavi et al [34] and Kohl et al [42], the consensus sequences of this motif have been identified. There is a consistent pattern of key residues which are well conserved to keep the structural integrity of this motif and some residues which are conserved. In 2003, the first combinatorial libraries of ANK repeat named Designed Ankyrin Repeat Protein library (DARPins) was generated from this consensus [43]. Several specific binders have been isolated from this library using ribosome display and phage display [42, 44-49]. However, the construction of DARPins library contained the fixed number of internal repeat which was 2 (N2C) and 3 (N3C) repeats. The amino acid in the random positions was designed to use the mixture of degenerated codon which codes the same proportion of all amino acid except cysteine, glycine and proline. According to natural ANK sequences, the number of repeat found in a single protein varies greatly. Therefore, in this study, we are interested in the construction of artificial ankyrin repeat library which mimic the natural sequence in term of varying number of repeat and also the amino acid which present in the natural protein.

Directed evolution

Directed evolution is a method used in protein engineering to exploit the power of natural selection to evolve proteins or RNA with desirable properties not found in nature. In order to modify or optimize an existing protein, a library of variant gene is designed and constructed with goals in mild. First, library members need to be sufficiently similar on sequence to the starting protein to share a similar structure and function. Second, library members need to be sufficiently different in sequence from the starting protein to be slightly different in structure and thus in the functional property of interest. Other factors that contribute to library design and construction are the difference between accessible library size and the theoretical sequence space of interest, the limited number of approaches, the use and control of randomization, the natural versus synthetic origin of the diverse population and library quality.

A typical directed evolution experiment involves mainly two steps which are diversification and selection. The diversification is the process to create large libraries of variant genes by mutation and/or recombination. Nowadays, several established methods have been proposed. The three major classes of library-construction methods applied to protein engineering are based on random mutagenesis, recombination and site-directed diversification. The random mutagenesis mimics the incidence of such errors over the millions of years of natural evolution, but vastly increases the rate of mutagenesis by artificial increasing the error rate of DNA replication such as UV light, x-ray radiation, chemical mutagen, and error-prone PCR. Recombination method such as DNA shuffling mimics a second mechanism of natural evolution in the exchange of pieces between related genes using homologous recombination. Site-directed diversification is the method that directed to a specific position or set of positions, and the remaining protein sequence is fixed as wild-type. The most wildly used method is to synthesize a set of oligonuclotides where the wild-type bases in each codon of interest are replaced by mixtures of degenerated nucleotides such as NNN, NNS or NNK.

The second step, selection is the process to isolate molecules or protein possessing the desired properties. This step needs the display technologies which physically linked the phenotype (polypeptides or proteins) displayed on a certain platform to their corresponding genotype (gene). These technologies include virus/phage display, cell display, ribosome display, mRNA display and covalent DNA display (CDT) with phage display being by far the most utilized as shown in Figure 4.

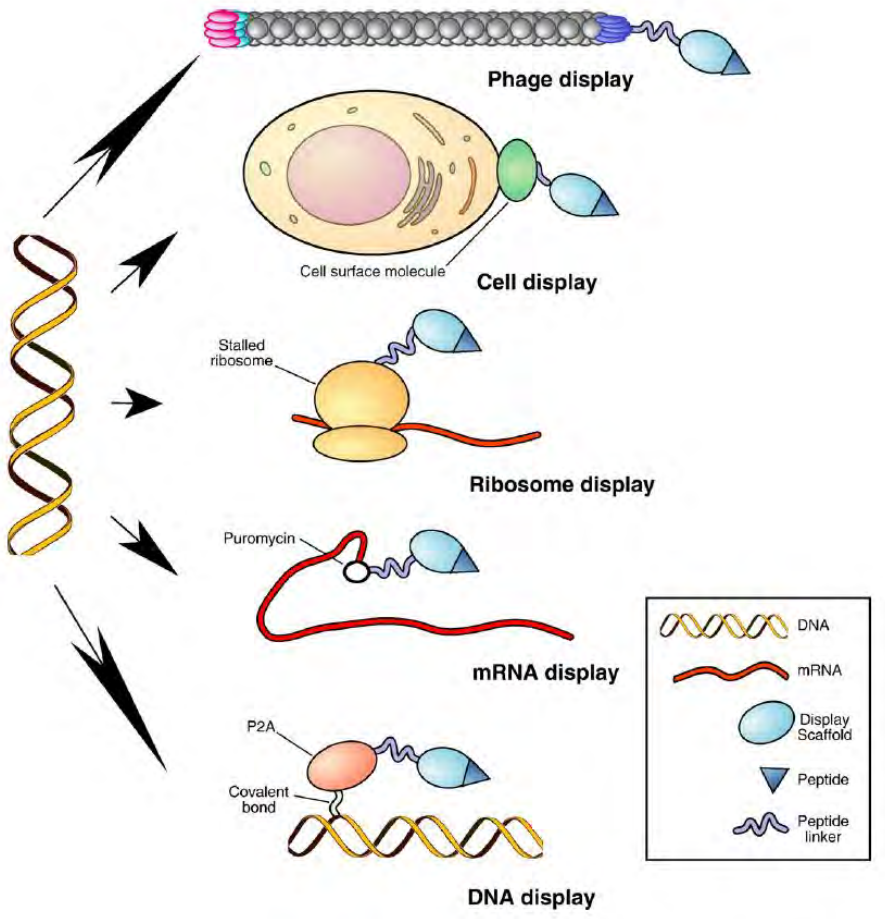

Figure 4. Schema of available display technologies [50].

Focusing on phage display, this technique was first described in 1985 by George Smith et al. They demonstrated that the linkage between phenotype and genotype could be established in filamentous bacteriophage and established the new technology of phage display [51]. The foreign DNA fragment is inserted into the genome of the filamentous phage and the encoded foreign peptide is displayed as a fusion to one of the coat proteins on the surface of phage. Since then, phage display has been used as a powerful method to isolate target-specific polypeptides with high affinity[52]. In this study, this technique will be applied to isolate the artificial ankyrin repeat proteins which bind specifically to matrix (MA) and capsid (CA) domains of HIV-1 for the intracellular application in the interference of viral production.

OBJECTIVE

1 To construct the artificial ankyrin repeat proteins libraries.
2 To produce the fusion protein of HIV-1 matrix and capsid (H_6MA-CA) target protein using baculovirus expression system.
3 To select the specific H_6MA-CA binders using phage display strategy.
4 To investigate the binding activity of H_6MA-CA binders.
5 To evaluate the intracellular function of selected binders in the interference of viral assembly and viral maturation.
6 To define the mechanisms which involve in HIV replication blockage

MATERIALS AND METHODS

Construction of artificial ankyrin repeat protein library

Sequence analysis and design of oligonucleotides fragments for library construction

The sequence of artificial ankyrin repeat proteins library is based on the sequence design designated as designed ankyrin repeat proteins (DARPins) library previously described by Plücktun A with two minor modifications:

- Limited changes in the Ankyrin repeat concensus sequence has been introduced in order to create a non symetric restriction site (*Bsm* BI). The presence of this site between consecutive ankyrin repeat coding fragments simplify library construction and module shuffling processes. The 3D structure of modified consensus was subjected to Discovery 2.5 program using DARPin E3_5 (1JM0) as the template to analyze the optimized structure comparing with the template structure.
- The randomization scheme used in the variable position of the ankyrin repeats was targeted to mimic the natural residue frequency at each variable position. The frequency of amino acids in each variable position was computed from Ankyrin domain families as defined in Pfam and Prosite. The conservation at each position in multi alignment of ankyrin modules can be visualized as a sequence logo generated using the WebLogo3 server (http://weblogo.berkeley.edu/).

The amino acid distribution at each variable position was approximated by encoding each these positions with a set of partially degenerated codons. The repeat sequences were made by a set of oligonucleotides. The sequence coding a single repeat was divided into four fragments: Va (variable fragment a), Vb (variable fragment b), Vc (variable fragment c) and C1 (constant fragment) as shown in **Figure 5**. The C1 fragment was designed to have the site for *Bsm* BI restriction enzyme. The sequences of each fragment were detailed in **Table 1**.

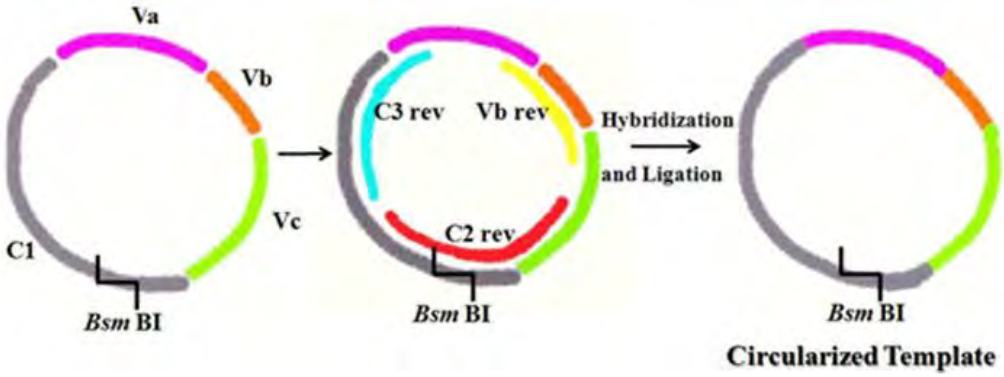

Figure 5 The generation of circularized template. All fragments of Va, Vb, Vc, C1 and short bridging fragments (Vb rev, C2 rev and C3 rev) are mixed together with an equal molar ratio and performed the hybridization and ligation resulting in the circularized template.

Table 1 Oligonucleotides for the artificial ankyrin repeat protein library construction.

Oligonucleotides	Sequence (5'->3')
Variable fragment a (Va)	amino acid position: 33, 1, X, X, 4, X, 6, 7
Va1	cgt gac vdk vdk ggt vdk acc ccg
Va2	cgt gac vdk vdk ggt dmy acc ccg
Va3	cgt gac vdk dmy ggt vdk acc ccg
Va4	cgt gac vdk dmy ggt dmy acc ccg
Va5	cgt gac dmy vdk ggt vdk acc ccg
Va6	cgt gac dmy vdk ggt dmy acc ccg
Va7	cgt gac dmy dmy ggt vdk acc ccg
Va8	cgt gac dmy dmy ggt dmy acc ccg
Va9	cgt gac raa vdk ggt dmy acc ccg
Va10	cgt gac raa dmy ggt vdk acc ccg
Va11	cgt gac raa vdk ggt vdk acc ccg
Va12	cgt gac raa dmy ggt dmy acc ccg
Va13	cgt gac vdk van ggt dmy acc ccg
Va14	cgt gac dmy van ggt vdk acc ccg
Va15	cgt gac dmy van ggt dmy acc ccg
Va16	cgt gac vdk van ggt vdk acc ccg
Va17	cgt gac vdk dmy ggt van acc ccg
Va18	cgt gac dmy vdk ggt van acc ccg
Va19	cgt gac dmy dmy ggt van acc ccg
Va20	cgt gac vdk vdk ggt van acc ccg
Va21	cgt gac vdk dmy ggt tgg acc ccg
Va22	cgt gac dmy vdk ggt tgg acc ccg
Va23	cgt gac vdk vdk ggt tgg acc ccg
Va24	cgt gac dmy dmy ggt tgg acc ccg
Variable fragment b (Vb)	amino acid position: 8, 9, X
Vb1	ctg cac ctg
Vb2	ctg cac tgg
Vb3	ctg cac tac
Vb4	ctg cac rtc

Table 1 Oligonucleotides for the artificial ankyrin repeat protein library construction (con't).

Oligonucleotides	Sequence (5'->3')
Variable fragment c (Vc)	amino acid position: 11, 12, X, X, 15, 16, 17
Vc1	gct gcg kck kck ggt cat ctg
Vc2	gct gcg kck var ggt cat ctg
Vc3	gct gcg kck aac ggt cat ctg
Vc4	gct gcg kck sgy ggt cat ctg
Vc5	gct gcg kck yay ggt cat ctg
Vc6	gct gcg kck ntg ggt cat ctg
Vc7	gct gcg tac var ggt cat ctg
Vc8	gct gcg tac kck ggt cat ctg

Vc9	gct gcg tac yay ggt cat ctg
Vc10	gct gcg cgy kck ggt cat ctg
Vc11	gct gcg cgy var ggt cat ctg
Vc12	gct gcg cgy aac ggt cat ctg
Vc13	gct gcg cgy sgy ggt cat ctg
Vc14	gct gcg cgy yay ggt cat ctg
Vc15	gct gcg cgy ntg ggt cat ctg
Vc16	gct gcg var kck ggt cat ctg
Vc17	gct gcg var var ggt cat ctg
Vc18	gct gcg var aac ggt cat ctg
Vc19	gct gcg var sgy ggt cat ctg
Vc20	gct gcg var yay ggt cat ctg
Vc21	gct gcg ktk kck ggt cat ctg
Vc22	gct gcg ktk var ggt cat ctg
Vc23	gct gcg ktk aac ggt cat ctg
Vc24	gct gcg ktk sgy ggt cat ctg
Vc25	gct gcg ktk yay ggt cat ctg
Constant fragment	amino acid position: 18-32*
C1	gaa atc gtt cgt ctc ctg ctg gaa cac ggc gca gac gta aac gcg

Table 1 Oligonucleotides for the artificial ankyrin repeat protein library construction (con't).

Oligonucleotides	**Sequence (5'->3')**
Bridging fragments	
Vb1rev	cgc agc cag gtg cag cgg ggt
Vb2Rev	cgc agc cca gtg cag cgg ggt
Vb3rev	cgc agc gta gtg cag cgg ggt
Vb4rev	cgc agc gay gtg cag cgg ggt
C2rev	cag gag acg aac gat ttc cag atg acc
C3rev	gtc acg cgc gtt tac gtc tgc gcc gtg ttc cag

Footnote: * Recognition site for *Bsm* BI is represented as underlined letters.
Neucleotides at randomized positions are higlighted in red color.
Code for mixed bases nomenclature:
d = g or a or t
k = g or t
m = a or c
v = g or a or c
r = a or g
s = g or c
y = c or t

Preparation of DNA cassettes encoding a single internal repeat for library construction.

All synthetic fragments (Va, Vb, Vc, and C1) were hybridized with reverse oligonucleotides linkers (bridging fragments; Vb rev, C2 rev, and C3 rev) at equal molar by heating at 95ºC for 5 minutes followed by cooling down to 25ºC at the rate 0.1 second/minute. To generate the circularized template, the hybridized product was ligated by T4 DNA ligase enzyme (NEB, Pickering, Ontario) resulting in the circular templates. The ligation product was purified using NucleoSpin® Extract II (Macherey- Nagel, Düren, Germany). The circularized templates were hybridized with random hexamer primers provided in the illustra TempliPhi 100 amplification kit and amplified by Phi29 polymerase (GE healthcare Bio-Sciences Co. Piscataway, NJ) at 30ºC for 15 hr. The amplification process using Phi29 polymerase is based on the rolling circle amplification (RCA) processes as shown in **Figure 6.**

The polymerized product was incubated at 65ºC for 15 minutes and subsequently treated with *Bsm* BI (NEB, Pickering, Ontario) at 55ºC for 4 hr. The *Bsm* BI-treated product represented mixture of a monomeric fragment of DNA cassette was purified using NucleoSpin® Extract II. The complete digestion was analyzed by agarose gel electrophoresis.

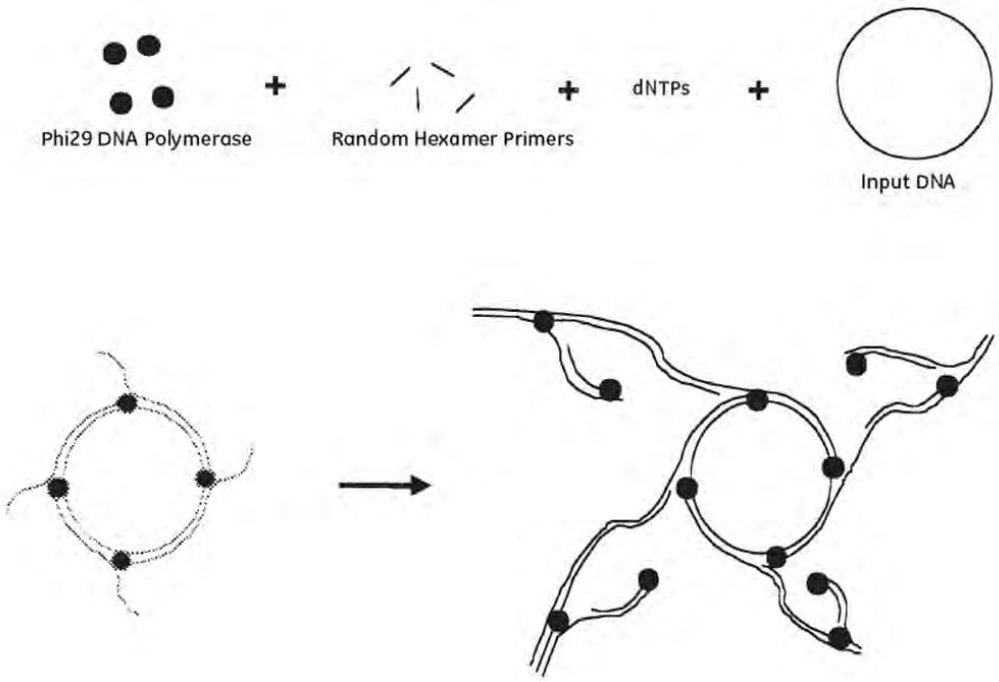

Figure 6 Schematic of the TempliPhi amplification process. Random hexamer primers anneal to the circular template DNA at multiple sites. Phi29 DNA polymerase extends each of these primers. When the DNA polymerase reaches a downstream extended primer, strand displacement synthesis occurs. The displaced strand is rendered single-stranded and available to be primed by more hexamer primers. The process continues isothermal amplification in 4-6 hr at 30ºC without the need for thermal cyclin.

Generation of intermediate vector

The phagemid pHDiExDsbA-Ank15 vector was used for the construction of intermediate vector termed acceptor vector (pHDiExDsbA-AccV). The synthetic adaptor cassette containing restriction sites required for the library construction process;

Ank ModAcc Fw (5'-TGCTGGAGGAGGGTACCGAGGAGACCTGCTA ATTGCTGAAGCACGGTGCTGACGTTAACGCTAAT-3') and Ank ModAcc Rev (5'-GATCATTAGCGTTAACGTCAGCACCGTGCTTCAGCAATTAGCAGGTCTC CTCGGTACCCC CTCC-3') were mixed at the equal molar ratio (5 μM) and hybridized by heating at 95ºC for 5 minutes followed by cooling down to 25ºC with the constant rate at 0.1ºC/second. The hybridized cassettes were cloned into *Not* I and *Bcl* I site of pHDiExDsbA-Ank15 and transformed into competent XL-1 Blue resulting in pHDiExDsbA-AccV vector.

Library construction

The pHDiExDsbA-AccV vector (30 μg) was treated with *Kpn* I at 37ºC for 4 hr and then purified using NucleoSpin® Extract II. The purified product was further digested with *Bsm* BI at 55ºC for 16-18 hr and purified using NucleoSpin® Extract II. The mixture of single-stranded oligonucleotides, Biotin link *Kpn* I (biotin-5'-ACGACAGGGTAC-3') and Link *Kpn* I Rev (5'-CTGTCTGT-3') was pre-hybridized prior to the ligation step with treated acceptor vector to seal the *Kpn* I site. The ligation product was further purified by NucleoSpin® Extract II. The DNA cassette which was prepared in the previous step and the intermediate vector (pHDiExDsbA-AccV vector linked with biotinylated-*Kpn* I linker) were mixed at the molar ratio 1:5. DNA fragments corresponding to single ankyrin modules were hetero polymerized in the acceptor vector with T4 DNA ligase at 20 ºC for 4 hr. Due to the non palyndromic *Bsm* B1 cohesive ends, the modules can be ligated in only one orientation. The excess cassettes were eliminated by capturing the biotinylated acceptor vector on pre-washed streptavidin coated magnetic beads (Sigma, St. Louis, MO). Beads were agitated at RT for 1 hr and washed three times with NEB buffer 3 (NEB, Pickering, Ontario). The vectors which are captured on beads were released out by incubating with *Bsp* MI enzyme at 50ºC for 3 hr resulting in the cohesive ends designed to compatible with *Bsm* BI site at 3' end of the polymerized DNA cassette. The mixture of released vector was glued to generate the circularized vector by T4 DNA ligase at 20ºC for 4 hr, as shown in **Figure 7**. The ligation product was purified and then electroporated into XL-1 Blue electrocompetent cells. The transformed cells were spread on the 20 cm x 20 cm plates of 2YT agar containing ampicillin (100 μg/ml) and 1% (w/v) glucose. Some of the transformed cells were taken to perform the serial dilution and plated on 2YT agar containing 100 μg/ml ampicillin and 1% (w/v) glucose to evaluate the library size. The number of colonies of each dilution was use for library size calculation. Colonies on the solid media were collected after overnight growth, pooled and stored at -80ºC in the 2YT medium containing 20% (v/v) glycerol. The DNA was extracted from a sample of pooled cells for restriction analysis and collective transformation in an expression host.

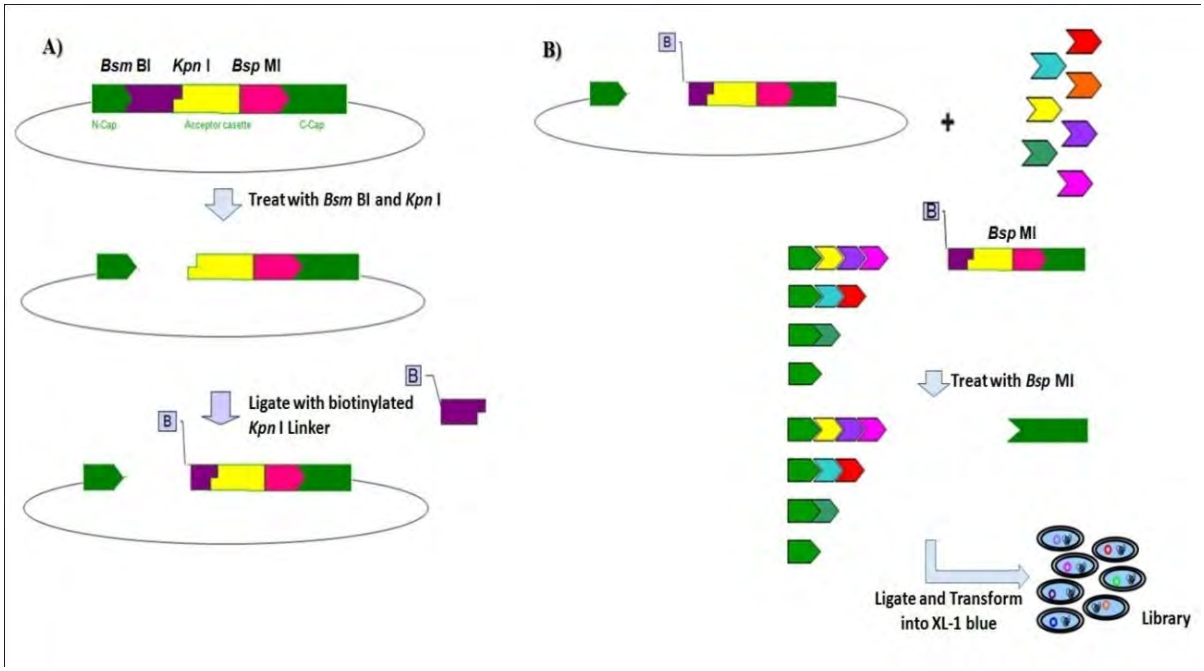

Figure 7 Process for library generation. **(A)** The intermediate vector was treated with *Bsm* BI and *Kpn* I restriction enzymes to generate the cohesive ends for polymerization with DNA cassettes at C-terminus of N-cap and for ligation with biotinylated linker. **(B)** The polymerization process was undergone randomly during the ligation process with the mixture of DNA cassettes. Free-DNA cassettes were removed by trapping the vector on streptavidin beads and washed. Then vectors were released from the beads by digestion with *Bsp* MI to generate the compatible end with *Bsm* BI and closed by the ligation process. Finally, released vectors were closed by the ligation process and then transformed into *E. coli*.

Characterization of the library

Thirty randomized colonies from the library were picked and extracted plasmids to analyze the sequence by the standard sequencing method. To validate the distribution of the repeat number, plasmid form pooled clones was treated with *Nde* I and *Hind* III at 37ºC for 3 hr and separated using gel electrophoresis. Moreover, the soluble protein expression was evaluated by colony filtration (CoFi) blot analysis as described elsewher. Briefly, the pooled plasmid from the library was transformed into *E. coli* BL21 (DE3). Seventy two of randomized single colonies were picked and cultured in 200 µl of LB broth containing ampicillin (100 µg/ml), and 1% (w/v) glucose in 96-well plate with shaking at 550 rpm, 37ºC for 5 hr. Pre-cultured cells were transferred to culture on Durapore® membrane filter (Millipore, Billerica, MA) which pre-covered on LB agar containing ampicillin (100 µg/ml), and 1% (w/v) glucose using 96-well replicator at 37ºC for 16-18 hr. The membrane containing colonies was transferred to another LB agar induction plate containing ampicillin (100 µg/ml), and 1mM IPTG for protein induction and incubated at 37ºC for 4 hr. The membrane was taken out from agar and placed on the pre-wet filter sandwich containing layers of nitrocellulose membrane on the top of three pieces of Whatman 3MM paper (GE healthcare Bio-Sciences Co. Piscataway, NJ). The filter was then subjected to freeze-thawing process at -80ºC for 10 minutes following by incubation at 37ºC for 10 minutes. Three consecutive repeats were performed. Consequently, the Durapore membrane was discarded. Soluble proteins were recovered on the nitrocellulose membrane and probed with a mouse anti-His tag antibody (GenScript, Piscataway, NJ) at dilution 1:5,000 in blocking solution (5% BSA in TBST) after the blocking step. The membrane was washed three times with

TBST (0.05% Tween 20 in TBS) and then revealed with goat anti-mouse immunoglobulins conjugated Alexa Fluor® 680 at dilution 1:10,000 in blocking solution. After washing step, the membrane fluorescent signal was observed using Odyssey® infrared scanner (LI-COR Biosciences, Lincoln, NE) with excitation at 680 nm and emission at 700 nm. Positive clones were detected by quantitative fluorescence signal analysis comparing to signal from background and negative control.

Preparation of phage-displayed artificial ankyrin repeat protein library

Five hundred microliters of stock library were cultured in 500 ml 2X YT broth (1.6% (w/v) tryptone, 1% (w/v) yeast extract, and 0.5% (w/v) sodium chloride) containing ampicillin (100 µg/ml), 1% (w/v) glucose and tetracycline (10 µg/ml) at 37ºC with shaking at 200 rpm until the $OD_{600\ nm}$ was as reached 0.5. The bacterial culture was infected further with 20 MOI of K07 helper phages and then incubated at 37ºC for 30 minutes without shaking and followed by 37ºC for 30 minutes with shaking at 150 rpm. Phage-infected cells were centrifuged at 1,200 g for 10 minutes at 25 °C. Pellets were resuspended and cultured in 500 ml of 2X YT broth containing ampicillin (100 µg/ml), and kanamycin (70 µg/ml) with shaking at 200 rpm at 30 °C for 16-18 hr. Bacterial culture was clarified by centrifugation 1,200 g for 10 minutes at 4°C. The culture supernatant was collected and phage particles were harvested by PEG/NaCl (20% (w/v) PEG 8000, 2.5 M NaCl) precipitation. Finally, the pellets were resuspended with 10 mM Tris-buffer saline (TBS) pH 7.4.

Phage titration

The concentration of phage was titrated using two methods, which are measuring by UV spectrophotometer as described elsewhere (Barbas, 2001) and re-infecting into *E. coli* strain XL-1 Blue resulting in the concentration in particles per volume (particles/ml) and colony forming unit per volumn (CFU/ml), respectively. The first method, precipitated phages were diluted at dilution 1: 50 in TBS and scanned from 240 to 320 nm with UV-2450/2550 spectrophotometer (Shimadzu, Columbia, MD). The absorbance at 269 nm (A_{269}) which is a maximum peak was used for calculation as follow:

$$\text{Phage particles per ml} = \frac{(\text{Adjusted } A_{269}) \times (6 \times 10^6) \times (\text{dilution factor})}{(\text{The number of nucleotides in the phage genome})}$$

The second method, fifty microliter of serial dilutions of phage were used to infect 450 µl of cultured *E. coli* (OD_{600} at 0.5) and incubated at 37ºC for 30 minutes. Fifty microliters of viral-infected bacteria were spread on LB agar containing ampicillin (100 µg/ml). Plates were incubated at 37ºC overnight and the ampicillin resistant colonies were counted and calculated for phage concentration using the formula below.

$$\mathbf{A = B \times C \times (1000/v)}$$

A = The concentration of phage (CFU/ml)
B = The number of ampicillin resistant colonies.
C = Dilution factor of viral infected bacteria.
V = Volume (µl) of viral infected bacteria.

Phage filtration using *Strep*-tactin® coated magnetic beads

To enrich the population of coding phage in the library, 1 x 10^{13} particles of phage library were diluted in filter-sterile 2% BSA in TBS and incubated with pre-washed *Strep*-

tactin® coated magnetic beads (IBA GmbH, GÖttingen, Germany) with agitation at 25ºC for 1 hr. Unbound phage particles were eliminated by washing nine times with washing buffer (sterile-filtered 0.1% Tween 20 in TBS). Bound phage particles were specifically eluted with 500 µl of 2.5 mM desthiobiotin and collected by centrifugation. The eluted phage particles in the supernatant were rescued by infecting *E. coli*. The infected cells were spread on LB agar containing ampicillin (100 µg/ml) and incubated at 37ºC overnight. The antibiotic resistant colonies were collected, pooled and stored at -80ºC in the 2YT medium containing 20% (v/v) glycerol. The phage particles were prepared from the glycerol stock for the next round filtration. This step was performed for three rounds. DNA was extracted from pooled clones of each round and transformed into *E. coli* BL21 (DE3) to evaluate the soluble expression by CoFi blot analysis as described in the previous step.

Production of H_6MA-CA and H_6-CA recombinant protein by baculovirus expression system

Vector construction

A DNA fragment coding for H_6MA-CA was generated by standard PCR method using the following pairs of primers: the first pair consisted of primers 5'-CTAGCATGGGTGCGAGAG-3' and 5'-CATGGGTGCGAGAGCG-3' and the second pair consisted of primers 5'-CTTACTACAAAACTCTTGCTTTATG-3' and 5'-GTACCTTACTACAAAACTCTTGC-3'. Two PCR reactions were performed using the HIV-1 plasmid pNL4-3 as the template. The PCR products from both reactions were then mixed, denatured, and hybridized to obtain DNA fragments with *Nhe* I and *Kpn* I cohesive ends, resulting in the H_6MA-CA-encoding fragment. This fragment was competent for ligation to pBlueBac4.5-His intermediate vector linearized with *Nhe* I and *Kpn* I. Plasmid pBlueBac4.5-His was derived from pBlueBac4.5 (Invitrogen, San Diego, CA) by insertion of a sequence coding for the 6Histidine (H_6) tag and a GSGSAS linker upstream to the *Nhe* I site. The H_6MA-CA-encoding fragment was cloned into the *Nhe* I and *Kpn* I sites of pBlueBac4.5 intermediate vector using the standard ligation method as described in previous step to generate pBlueBac-H_6MA-CA vector.

Production of recombinant H_6MA-CA in Sf9 cells

Sf9 cells were cotransfected with pBlueBac4.5-H_6MA-CA (10 µg) and Bac-N-Blue™ DNA using Cellfectin® II reagent following the Bac-N-Blue™ transfection and expression manual (Invitrogen, San Diego, CA). These two vectors recombined inside the cells resulting in the recombinant AcMNPV DNA (**Figure 8**) containing genes encoding for proteins for viral replication, H_6MA-CA and LacZ for blue plaque selection as shown in **Figure 9**. The recombinant virus (BV-H_6MA-CA) in the supernatant was isolated using the blue plaque selection method as described in the instruction manual. Briefly, the culture supernatant of transfected cells was harvested after 48 hr of incubation. Serial dilutions of the supernatant containing the recombinant virus were used for infection with Sf9 cells seeded in 6-well plates. Plate was rotated gently and incubated at 27ºC for 1 hr. Medium was discarded after incubation prior to carefully adding the TNM-FH media containing 2% of low melting agar and 150 µg/ml of 5-bromo-4-chloro-3-indoly-β-D-galactoside (X-gal). The blue plaques were picked up and resuspended in Grace's Insect medium and used for infection in Sf9 cells. After 48 hr of infection, infected cells were harvested and lysed cells by the freeze-thawing method. The clarified lysate was yielded by centrifugation at 15,000 g, 4ºC for 30 minutes. The presence of H_6MA-CA protein was detected by western immunoblotting. The cell lysate was separated in 12% SDS-PAGE and then transferred to nitrocellulose membrane. The

blotted membrane was blocked with 5% skim milk in TBS at RT for 1 hr and then detected using monoclonal anti-His tag antibody diluted 1:5,000 in the blocking solution at RT for 1 hr with slow agitation. After an incubation time, the membrane was washed with TBST (0.05% Tween 20 in TBS) and revealed using goat anti-mouse immunoglobulins conjugated HRP at dilution 1:8,000 in 5% skim milk in TBS. Unbound conjugate was washed five times. The specific bands were visualized using TMB membrane peroxidase substrate (KPL, Gaithersburg, MD). H_6MA-CA protein was recovered from a clarified Sf9 cell lysate by affinity chromatography on HisTrap column using ÄKTA prime™ plus (GE healthcare Bio-Sciences Co. Piscataway, NJ). Protein concentration of purified H_6MA-CA was determined by Bradford protein assay (Thermo Fisher Scientific Inc., Rockford, IL). Purified H_6MA-CA protein was separated in 15% SDS-PAGE under reducing condition and then the gel was stained using Coomassie's blue to determine the purity.

Recombinant H_6-CA was produced by baculovirus expression system. A gene fragment coding CA was amplified from parental vector (pNL4-3) by standard PCR protocol using pair of primers: FWD_p24 *Nhe* I, 5'-GAGGAGGAGGTGCTATA GTGCAGAACCTCCAG-3' and REV_p24 *Kpn* I, 5'-GAGGAGGAGCTGGT ACCTTACAAAACTCTTGCTTTATGGCC-3'. PCR fragment was treated with *Nhe* I and *Kpn* I and subsequently cloned into pBlueBac-H_6MA-CA plasmid resulting pBlueBac-H_6-CA. This vector was co-transfected with Bac-N-Blue^{TM} DNA using the standard protocol as described. The subsequent steps for recombinant viral production, protein expression and detection were followed the same protocol as described in H_6MA-CA production.

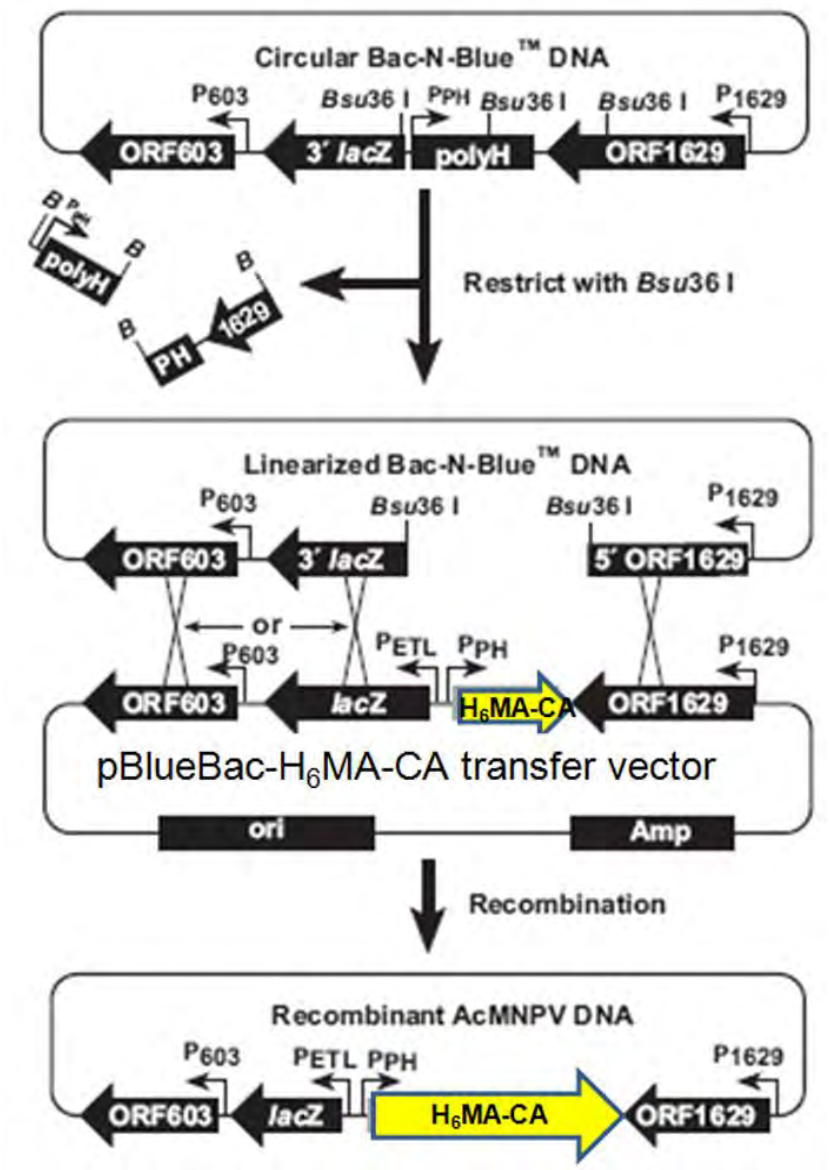

Figure 8 Recombination between Bac-N-Blue™ DNA and pBlueBac-H_6MA-CA transfer vector (modified from the instruction of Bac-N-Blue™ transfection kit, Invitrogen).

.

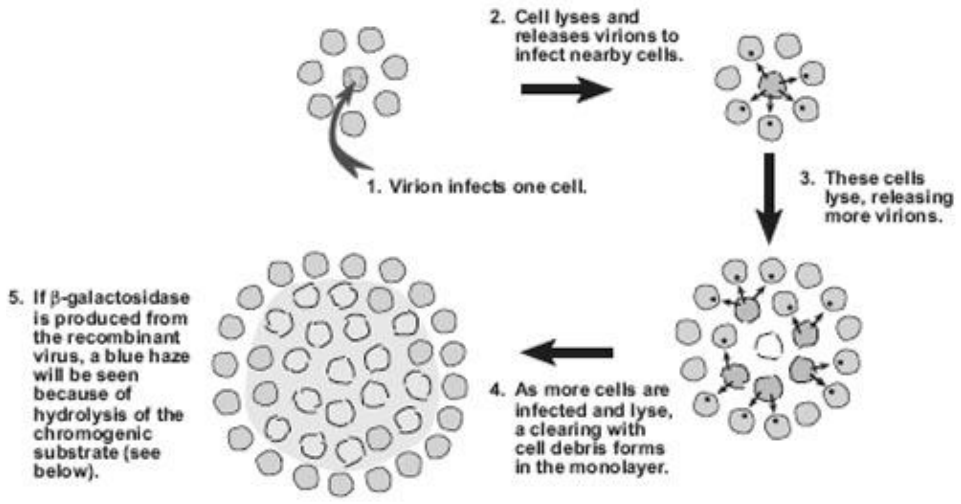

Figure 9 Formation of blue plaques. The infected cells are lyzed and released virions to infect nearby cells. The plaque with cell debris as observed a clearing area indicated more cells are infected and lyzed. In the presence of X-gal, a blue haze will be seen because of hydrolysis of the substrate (modified from the instruction of Bac-N-Blue™ transfection kit, Invitrogen).

Selection of specific binders from the constructed artificial ankyrin repeat protein library

Selection procedure

Microtiter plate (NUNC, Roskilde, Denmark) was coated with 100 µl of purified H_6MA-CA protein (20 µg/ml) or A3 protein diluted in sterile PBS overnight at 4ºC. A3 protein was used as the folded target protein for evaluating the quality of constructed artificial ankyrin library. Purified A3 protein was produced as described. Coated wells were washed four times with sterile-filtered TBST. The wells were blocked the non-specific binding with 200 µl of sterile-filtered blocking buffer (2% (w/v) BSA in TBST) at RT for 1 hr with shaking at 150 rpm in Eppendorf Thermomixer® (Eppendorf, Hauppauge, NY). After the washing step, 100 µl (1 x 10^{11} particles) of diluted phage were added and incubated at RT for 1 hr with shaking. After an incubation time, wells were washed 20 times with TBST and 10 times with TBS to eliminate unbound phage. 100 µl of 0.1M Glycine pH2.5 was added and incubated at RT for 10 minutes with shaking. Then 12.5 µl of 1M TrisHCl pH8 was added to neutralize the acid condition. The eluted phage was collected and subsequently infected with 5 ml of XL-1 Blue ($O._{D600nm}$ = 0.6-0.8) at 37ºC for 30 minutes. The infected cells were centrifuged at 1,200 g 25ºC for 10 minutes and re-suspended with 1 ml of 2X YT broth. The cells were plate on LB agar containing ampicillin (100 µg/ml). Some of the cells were taken to evaluate the number of phage output by performing serial dilution in 2X YT broth and 50 µl of each dilution was spread on LB agar containing ampicillin (100 µg/ml). The colonies on agar were collected and used for phage preparation as described in previous step to perform the next round of selection. In addition for the third round of selection, free antigen was used as specific elutor in the elution step instead of 0.1M Glycine pH2.5 to obtain the specific binders.

Screening of specific binding clones

Phage particles were prepared in small scale using the 96-wells format. Briefly, random colonies from each round of selection were picked and culture in 150 µl of 2X YT broth containing ampicillin (100 µg/ml) and 1% (w/v) glucose in 96-well plate (master plate) with shaking at 600 rpm in Eppendorf Thermomixer®, 37ºC overnight. 10 µl of overnight culture was transferred to new plate containing 150 µl of 2X YT broth containing ampicillin (100 µg/ml), tetracycline (10 µg/ml), and 1% (w/v) glucose and cultured at 37ºC with shaking for 4 hr. Bacterial cells were infected with 20 MOI of KO7 helper phage and incubated at 37ºC without shaking for 30 minutes and at 37ºC with shaking for 30 minutes. 200 µl of infected cells was transferred to the 96-well storage plate containing 1.5 ml of 2X YT broth adding ampicillin (100 µg/ml) and kanamycin (70 µg/ml). The plate was covered with gas permeable adhesive seals (Thermo scientific, Surrey, UK) and cultured at 30ºC with shaking for 16-18 hr. Phages in the culture supernatant were collected after the centrifugation and used for phage ELISA to determine the binding activity.

Hundred microliters of 20 µg/ml of purified H_6MA-CA or A3 diluted in PBS was added into a microtiter plate and left for overnight at 4ºC in the moisture chamber. The coated wells were washed four times with TBST. The wells were then blocked the non-specific binding with 200 µl of blocking solution at RT for 1 hr with shaking. After the washing step, 100 µl of the culture supernatant of each clone containing phage particles was added and incubated at RT for 1 hr. 100 µl of mouse anti-M13 conjugated HRP (GE healthcare Bio-Sciences Co. Piscataway, NJ) at dilution 1:5,000 in blocking buffer was added after washing step. The wells were then washed again prior to adding 100 µl SureBlue™ TMB Microwell Substrate (KPL, Gaithersburg, MD) and measured the optical density at 450 nm (OD_{450}) after adding 1 N HCl by a MTP-120 ELISA plate reader (Corona Electric, Japan).

Production of soluble ankyrin binders

Construction of pQE-30 expressing soluble ankyrin binders

pQE-30 ankyrin acceptor vector was constructed by inserted the hybridization product of two synthetic oligonucleotides, pQE-Ank-Adapt-Fw (5'-GATCCGCGGC CGCAAACGCGTAAA-3') and pQE-Ank-Adapt-Re (5'-AGCTTTTACGCGTTTGC GGCCGCG-3'), into *Bam* HI and *Hind* III sites on pQE-30 vector resulting in additional *Not* I restriction site in pQE-30. The DNA fragment encoding ankyrin binders was isolated from pHDiExDsbA phagemid vector by treated with *Not* I and *Hind* III and then subcloned into the same sites on acceptor vector resulting in pQE-30 containing DNA fragment encoding ankyrin binders. These vectors were transformed into chemical competent M15 [pREP4] (Qiagen, Hilden, Germany).

Expression and purification of soluble ankyrin binders

M15 [pREP4] containing certain vectors were cultured in 500 ml of LB broth containing ampicillin (100 µg/ml), kanamycin (25 µg/ml), and 1% (w/v) glucose at 37ºC with shaking until OD_{600nm} reached 0.8. Bacterial cells were induced for protein production by adding 1mM IPTG and cultured at 30ºC for 4 hr with shaking. The induced cells were collected by centrifuged at 3000 rpm 4ºC for 30 minutes. The cell pellets were resuspended with lysis buffer (TBS pH7.4, complete EDTA free protease inhibitor, 1 µg/ml lysozyme) and then subjected to three cycles of freeze-thawing. The lyzed cells were centrifuged at 15,000 g 4ºC for 30 minutes. The soluble form of ankyrin binders was purified by injected

the clarified lysate into HisTrap column (GE healthcare Bio-Sciences Co. Piscataway, NJ) followed by Sephadex G-75. The purified proteins were analyzed the purity by separated in 15% SDS-PAGE and performed Coomassie's blue staining.

Biotinylation of soluble ankyrin binders and monoclonal antibodies

The purified ankyrin binders, and monoclonal antibodies were chemically linked to biotin molecule following the instruction of the EZ-Link Sulfo-NHS-LC-Biotin kit (ThermoScientific, Rockford, IL). Briefly, 100 µM of purified protein was mixed with 500 µM of Sulfo-NHS-Biotin solution at the final volume 2 ml and then incubated at 25ºC for 1 hrs. Free biotin molecules were removed by applying the mixture into pre-equilibrated Zeba™ Desalt Spin Column (ThermoScientific, Rockford, IL) and allowed sample to absorb into resin. The column was centrifuged at 1,000 g for 2 minutes. The biotinylated protein solution in flow-through was collected and measured the concentration using NanoDrop 2000 (ThermoScientific, Rockford, IL). The purified proteins were evaluated the biotinylation efficiency by dot blot analysis. 10 µmole of biotinylated proteins was spotted on nitrocellulose membrane. The membrane was occupied with blocking solution and revealed by Extravidin-HRP (Sigma, St Louis, MO) at dilution 1:5,000 in blocking solution for 1 hr with shaking. The BM Blue POD Substrate, precipitating (Roche, Mannheim, Germany) was added to visualize the reaction.

Generation of stable cell line expressing selected ankyrin binders

Construction of pCEP4 vector harboring ankyrin binders

Two versions of vectors which are pCEP4Myr-Ank-GFP and pCEP4Cyt-Ank-GFP vector were constructed. These two vectors were used for expressing and directing ankyrin protein to inner membrane of cells (Myr) and cytoplasm (Cyt), respectively. The DNA encoding ankyrin repeat proteins, which bind to H_6MA-CA were fused with or without myristoylation signal and green fluorescent protein (GFP) at N- and C- terminus respectively, by overlapping PCR using two sets of primers, as shown in **Table 1**. Genes encoding ankyrin binders were amplified from pHDiExDsbA-encoding ankyrin binders vector and DNA encoding GFP fragment was amplified from pTriEx-GFP using Accuprime™ *Pfx* DNA polymerase (Invitrogen, Carlsbad, CA). PCR products encoding ankyrin binders and GFP were used as a template for the second round of PCR resulting in the recombined fragment of these two genes. The PCR products of the second round PCR were treated with *Kpn* I and *Xho* I restriction enzymes and cloned into corresponding sites of pCEP4 based vector. The sequences of these vectors were analyzed by standard sequencing method.

Table 1 The sequences of primers for generation of pCEP4 expressing ankyrin binders

Primers	Sequences (5'-->3')
Primers for ankyrin amplification	
Fw KpnIAnkMyr-pCep4	gag gag gag ctg gta cca tgg gga gta gca aga gca agg cgg ccg cag acc t
Fw KpnIAnk-pCep4	gag gag ctg gta cca tgg cgg ccg cag acc tgg gta ag
Rev AnkHindIII-pCep4	acc aga gaa gct ttg cag gat ttc agc cag gtc ctc gt
Primers for GFP amplification	
Fw HindIIIGFP-pCep4	ctg caa agc ttc tct ggt acg att gat gac gac ga
Rev GFPHisXhoI-pCep4	gac ctc ctc gag cta tta gtg atg gtg gtg atg gtg act agt ttt gt

Establishment of stable cell line expressing MA-CA binders

1 x 10^6 cells of SupT1 cell line were electroporated with pCEP4Myr-Ank-GFP and pCEP4Cyt-Ank-GFP vector using the Nucleofector™ (Amaxa, Koeln, Germany) with the Nucleofector™ transfection reagent V following protocol T-16. pCEP4 based vector contains the hygromycin resistance gene for selection of stable lines. The transfected cells were selected by limiting dilution and maintained in complete RPMI containing hygromycin B (Hyclone) ranging from 10-400 μg/ml. The transfected cells were observed the expression and the localization of protein under the fluorescence microscopy and flow cytometry.

RESULTS and DISCUSSION

Design of artificial ankyrin repeat motif for library construction

The ultimate goal in library construction is to design the consensus repeat modules which contain the fixed conserve-framework residues and the randomized potential target interaction residues. The consensus proposed in this study was based on the previous constructed library called 'Designed Ankyrin Repeat Proteins (DARPins) library' with an advance in the strategy for library generation to comprise the randomization not only in amino acids but also the number of repeats. Moreover, degenerated codons in interacting residues were designed to gain the similar proportion with natural sequences. The single internal repeat of ankyrin template was subjected to align with the databases. The 3485 and 1180 sequences of natural ankyrin repeats from Prosite (PS50088) and Pfm (PF00023) database were retrieved. Only repeats without extra insertions or deletions were recruited for alignment and subsequently submitted to generate the sequence logo, the graphical representation of an amino acid multiple sequence alignment on website (WebLogo, http://weblogo.berkeley.edu/). Each logo consists of stacks of symbols, one stack for each position in the sequence. The overall height of the stack indicates the sequence conservation at that position, while the height of symbols within the stack indicates the relative frequency of each amino or nucleic acid at that position.

The sequence logo of obtained natural ankyrin repeats was generated as shown in **Figure 10A**. The order of amino acid was re-numbered with the respect to DARPin library. The resulting alignment of all sequences yielded the consensus containing highly randomized residues at position 2, 3, 5, 10, 13, 14, 17, and 33. Other positions were defined as framework residues. Second step, the consensus sequence was defined based on DARPins and resulting alignments with some modifications as shown in **Figure 10B**. Amino acids at positions 26 and 33 were preserved as histidine (H) and arginine (R), respectively. Glutamic acid (E) and valine (V) at position 21 and 22 were changed to arginine (R) and leucine (L) for inserting *Bsm* BI restriction site. Additionally, lysine (K) at position 25 was substituted with E to prevent the electrostatic problem with R at position 21.

The 3D structure of our consensus with substitutions (1D4_binder) was generated using Accelrys Discovery Studio version 2.5 software. The minimized structure of 1D4_binder was superimposed with template (E3_5 template). As shown in **Figure 11**, the structure of generated consensus was not critically changed from the template, so that these substitutions might not alter the structure of protein. Finally, the library was created with the variation of amino acids at positions 2, 3, 5, 10, 13, and 14 as shown the consensus comparing with DARPins in **Figure 10B**. All of these variable positions are localized at the first α-helices and oriented towards the binding site ankyrin domain as demonstrated in **Figure 12**. From 3D structures of complexes of ankyrin domains with target proteins, the β-turns and the first α-helices of protein involved in the interactions**.** Focusing on the distribution of amino acid at these variable positions, residues in each position from all obtained sequences were analyzed and presented as the graph in **Figure 10A**. Proline (P) expressed very low frequency at positions 2 and 3, or absent at 5, 10, 13 and 14. Cystein (C) also presented with low proportion and did not include in these design to prevent the disulfide bonding problem. The design of degenerated codons was based on these observed natural sequences with no code for these two residues and stop codon. The proportion of coded amino acid of this library was shown in **Table 2**.

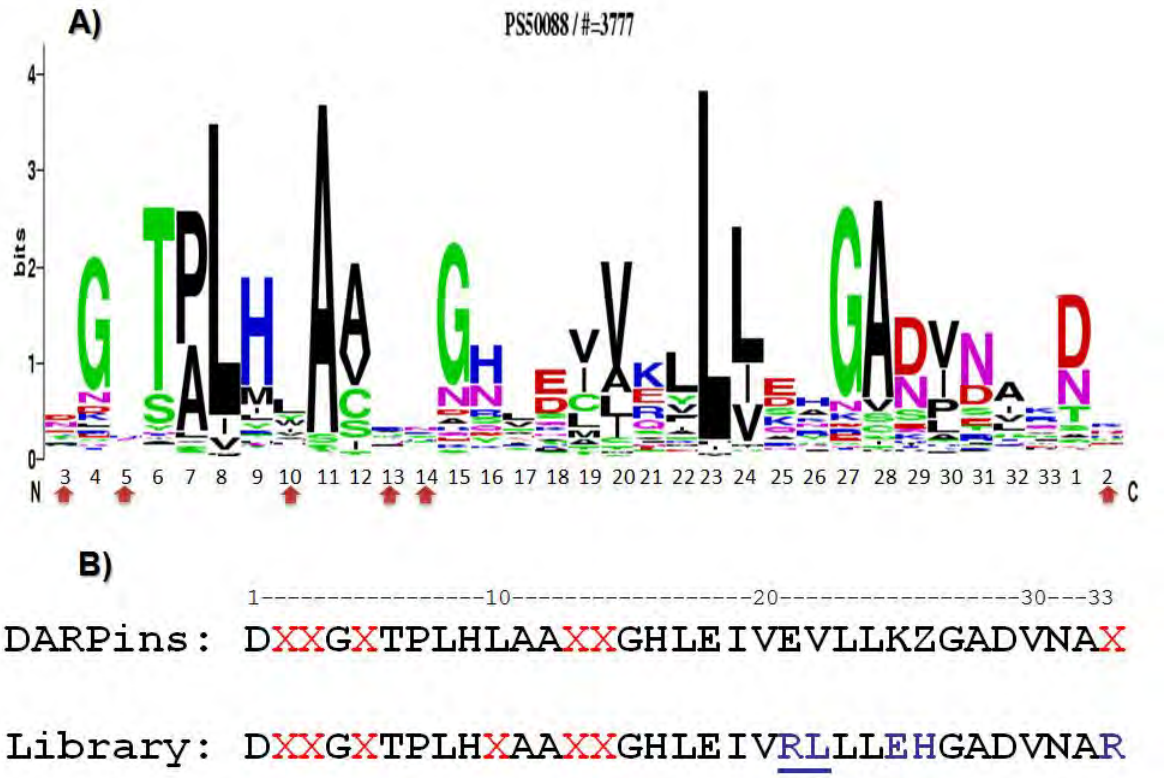

Figure 10 Sequence analysis of natural ankyrin repeats. (A) The sequence logo of natural ankyrin repeat generated from WebLogo. Amino acids were presented in the single letter code. The red arrows indicated the positions for generating the artificial ankyrin library. **(B)** The consensus sequence of constructed library comparing with DARPins. The red letters refer to the randomized position and blue letters represent the altered residues. Recognition sites for *Bsm* BI in consensus were underlined.

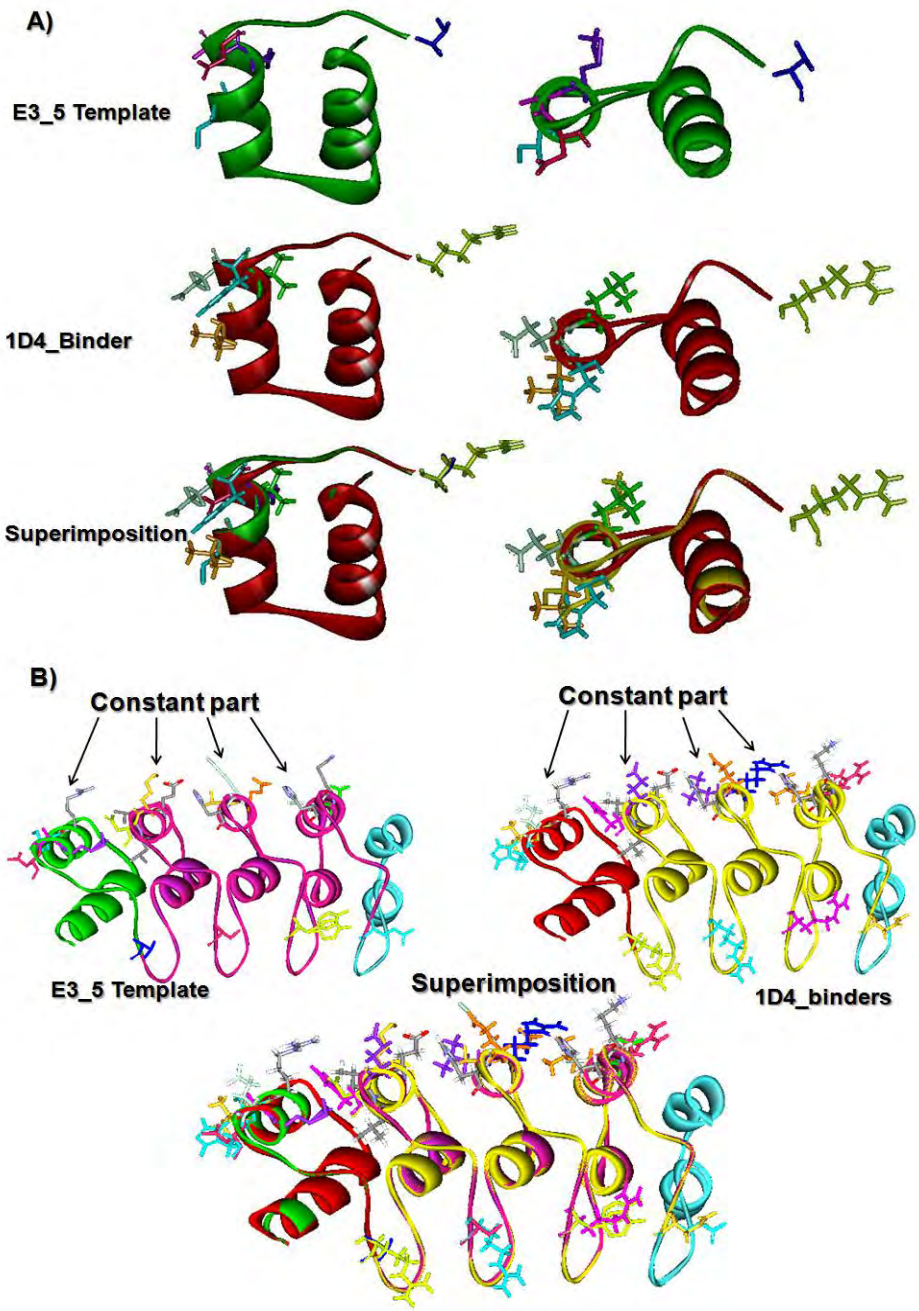

Figure 11 Optimized structures of designed consensus. A single domain **(A)** and whole molecule of ankyrin repeat **(B)** were illustrated as ribbon. Substituted residues located at the constant part were depicted as stick.

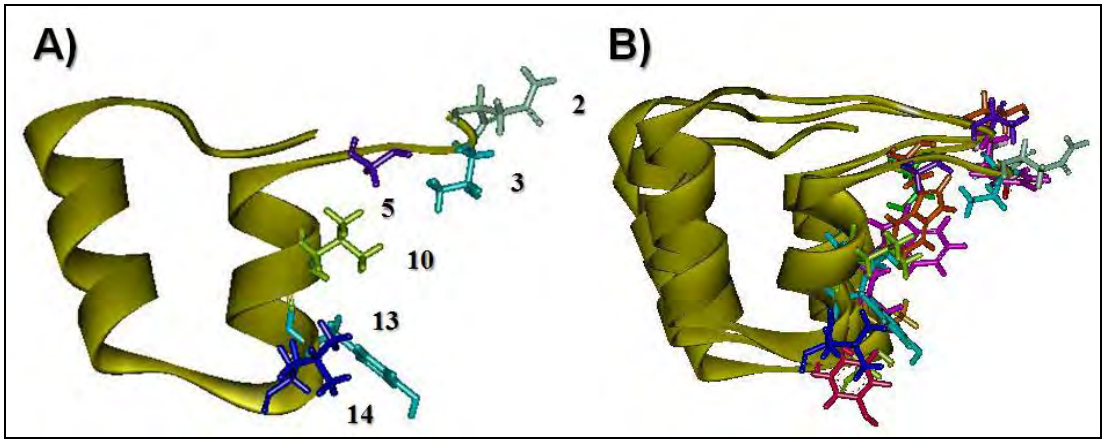

Figure 12 Model structure of internal repeat. A single repeat **(A)** or triple repeats **(B)** were presented in yellow ribbon. Amino acids at variable positions were indicated as the stick pattern.

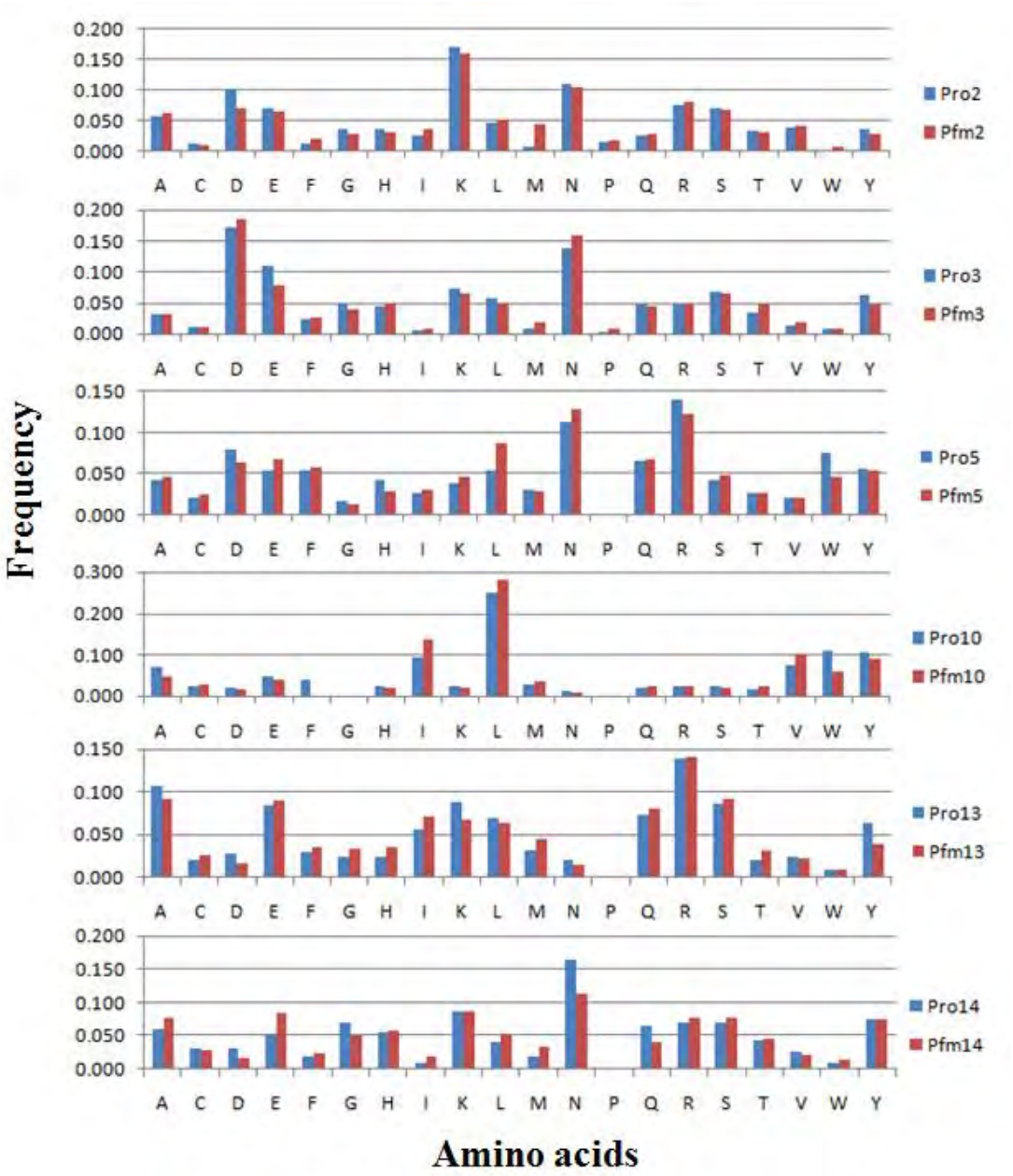

Figure 13 Amino acid distribution of natural ankyrin sequence. Amino acids were presented as a single letter code. Frequency obtained from Prosite and Pfam were depicted in blue and red bar, respectively. Each graph represented the frequency at each position as indicated the number.

Table 2 Proportion of amino acid coding form mixed bases at each randomization.

Randomized Positions	Mixed bases*	Amino acids (single letter code)
2, 3, 5	VDK	D, E, G(2), H, I, K, L(2), M, N, R(3), S, V
	DMY	A(2), D(2), N(2), S(2), T(2), Y(2)
	RAA	E, K
	VAN	D(2), E(2), H(2), K(2), N(2), Q(2)
	TGG	W
10	CTG	L
	TGG	W
	TAC	Y
	RTC	I, V
13, 14	KCK	A(2), S(2)
	VAR	E(2), K(2), Q(2)
	SGY	G(2), R(2)
	YAY	H(2), Y(2)
	NTG	L(2), M, V

*Mix bases refer to: N= A/T/G/C Y= C/T
V= A/C/G R= A/G
D= A/G/T S= C/G
K= G/T
M= A/C

Library construction

The artificial ankyrin library constructed in this study was made by a variable number of ankyrin modules with the internal variable residues flanked between N- and C-capping using the directional polymerization of a microgene or fragment corresponding exactly only one repeat into phagemid vector. To generate this library, the defined consensus as indicated above was fragmented into four parts which were Va, Vb, Vc, and C1. Va, Vb, and Vc encoded for amino acid positions 31-7, 8-10, and 11-17, respectively. All three fragments contained the random positions 2, 3, and 5 in Va, 10 in Vb, and 13 and 14 in Vc. Constant fragment, C1, encoded for residues 18-32 with the recognition site for *Bsm* BI. All fragments were annealed with complementary bridging oligonucleotides to generate the circular DNA template coding exactly for one repeat and closed by ligation. The mixture of circularized templates was amplified using Phi 29 polymerase (Dean et al., 2001) resulting in long homopolymers of repeats that were then separated into monomers by treated with *Bsm* BI and verified by gel electrophoresis as shown in **Figure 14**. The monomeric fragments were undergone the polymerization directly onto treated acceptor vector which was captured on the magnetic beads at the 5' compatible site of N-terminal capping by ligation process. In this step, the number of modules in the library is variable and controlled by the molar ratio of acceptor vector and monomers. Consequently, the vector containing heteropolymerized repeats were treated by *Bsp* MI to release vectors from beads and to seal vectors with the

designed-compatible cohesive extremities to *Bsm* BI. The released vectors were closed by ligation and electroporated into XL-1 Blue electrocompetent cells.

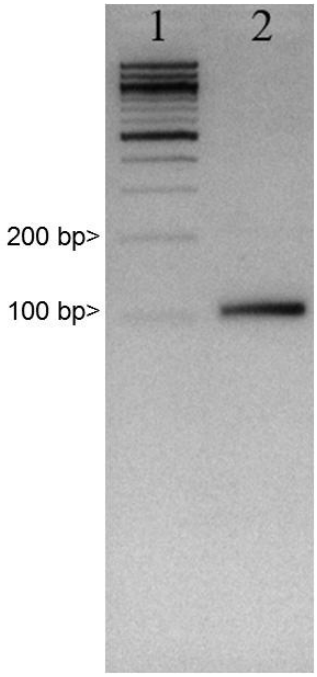

Figure 14 Agarose gel electrophoresis indicated the complete digestion of homopolymer into monomer.

Lane 1: 100 base pair DNA marker

Lane 2: Monomer of DNA coding one repeat.

Characterization of constructed library

An artificial ankyrin library was constructed with 1.9 x 10^8 independent clones so call 'naïve library'. The distribution of repeat number was evaluated by treated plasmid pool from the library with restriction enzymes located on each side of coding sequence and separated on gel electrophoresis. As shown in **Figure 15A**, this result demonstrated an apparent pattern of discrete fragments certified a variation in the number of repeat between 0-15, with maximum length of 2 to 6. Sequences of thirty randomized clones were analyzed. The selected clones had the internal repeat between 1 to 5 modules, as observed the expected length by restriction analysis. Ten of these selected clones contained the coding sequence and the remaining clones had coding modules and frame-shift modules resulting in internal stop codon. These unexpected sequences could result either from oligonucleotides synthesis, or from miss-assembly and amplification.

For evaluating the proportion of soluble protein expression in constructed library, a plasmid pools from library was transformed into *E. coli* BL21 (DE3) strain and 72 randomly selected clones were evaluated soluble protein expression by colony filtration blot (CoFi blot) analysis. The positive clones were analyzed by comparing the fluorescence intensity of tested clones with that of negative control. The result revealed that 34% (24/72) of selected clones were expressed as soluble proteins (**Figure 15B**). These results suggested that the proportion

of coding sequence in library is closed to the expressing clones detected by CoFi blot. Focusing on the non-coding clones, this fraction may result the problem in panning step. An enrichment of coding fraction was done by phage pre-selection or phage filtration using *Strep*-tactin® coated magnetic beads to eliminate protein-free viral particles. *Strep*-tag domain encoded upstream of ankyrin repeat fused to pIII domain of phage binds to *Strep*-tactin which is coated on magnetic beads. After phage binding step, free viral particles were washed out. Captured particles were specifically eluted using desthiobiotin and rescued by infection into *E. coli* XL-1 blue. After two round of filtration, the soluble protein expression was evaluated by CoFi blot as described. Interestingly, fraction of coding clones was enriched to 42% (30/72) and 60% (42/72) after the first and second round, respectively (**Figure 16**). Although, the elution step was performed using specific molecule, non coding clones still present in the filtrated library. This phenomenon could be described by reasons that protein-free viral particles may stick to displayed particles during phage precipitation and washing step is not enough to completely remove unbound particles. Therefore, naïve and filtrated libraries were taken to prepare phage library for the selection process.

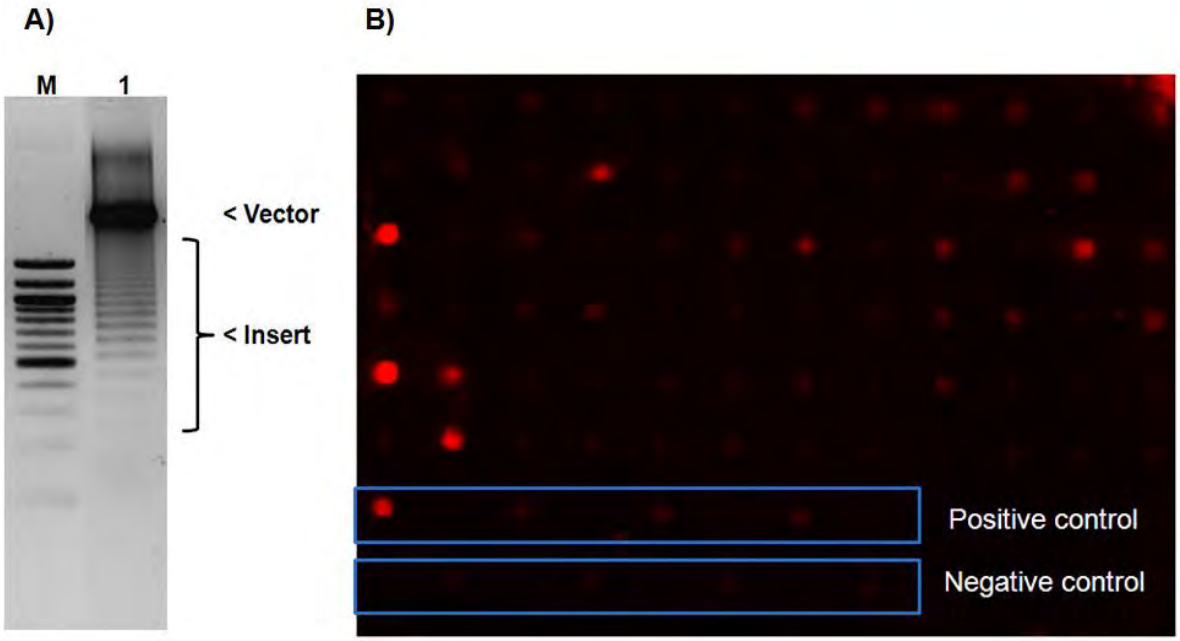

Figure 15 Characterization of constructed ankyrin library. (A) The distribution of number of repeat was evaluated by restriction analysis. Vector was extracted from plasmid pooled of library and treated with *Nde* I and *Hind* III. **(B)** The evaluation of soluble protein expression by CoFi blot analysis. Seventy two of randomly clones were picked up for testing. pHDiExDsbA-Ank15 and pHDiExDsbA-AccV-transformed *E. coli* were used as positive and negative control, respectively.

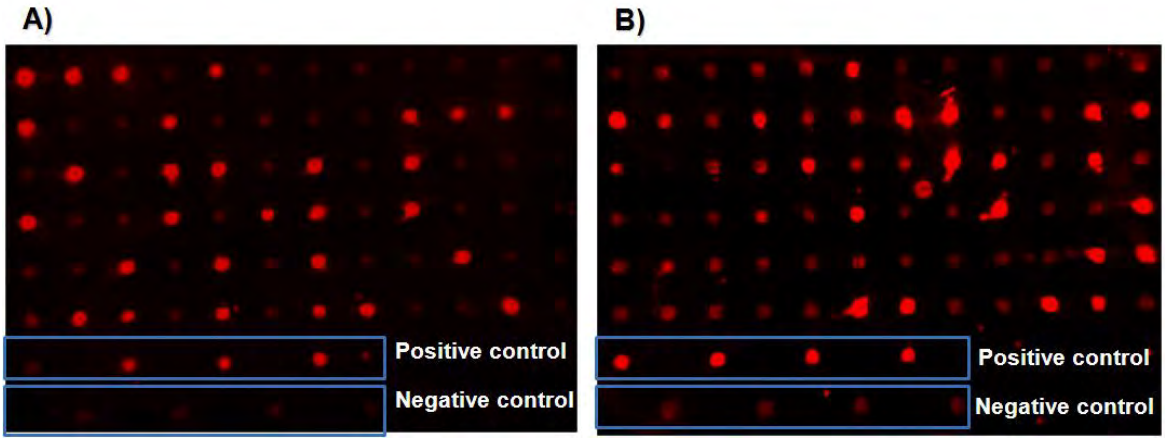

Figure 16 Enrichment of coding clones by phage filtration using *Strep*-tactin® coated magnetic beads. Seventy-two clones from first round **(A)** and second round **(B)** were randomly selected for testing the expression by CoFi blot. Plasmid pHDiExDsbA-Ank15 and pHDiExDsbA-AccV were used as positive and negative control, respectively.

Production and purification of recombinant H_6MA-CA and H_6-CA

The DNA fragment coding for MA-CA, a non-N-myristoylated, carboxyterminal-truncated version of HIV-1 Gag polyprotein containing the matrix (MA) and capsid (CA) domain, was cloned into *Nhe* I and *Kpn* I site of pBlueBac4.5-His resulting in pBlueBac4.5-H_6MACA. This constructed transfer plasmid was co-transfected with linearized BV DNA into Sf 9 cells. The recombinant BV, abbreviated BV-H_6MA-CA, from transfected cells was isolated using blue plaque selection. The infected cells harboring recombinant BVs showed the blue plaque appearance after beta-galactosidase staining (**Figure 17A)** were picked and used for protein production. The cells infected with wild-type AcMNPV were observed the precipitated polyhedrin inside the cells as shown in **Figure 17B**. The presence of recombinant H_6MA-CA in the infected-cells was analyzed by Western immunoblotting using anti-His and anti-Gag antibodies. The reactive bands referring to the molecular size of H_6MA-CA protein were detected in all isolated plaques as shown in **Figure 17C and D**. The large amount of recombinant H_6MA-CA protein was produced in Sf9 cells infected with BV-H_6MA-CA. Infected cells were harvested at 48 hr post infection (pi), and recombinant H_6MA-CA were recovered from clarified cell lysate by Ni^{2+}-NTA-agarose column. The recombinant protein could be isolated with high purity as shown in **Figure 18**. This purified protein was dialyzed against PBS and ready for isolating the binders and detecting the binding activity in the next experiments. In addition, the production of recombinant H_6-CA protein was accomplished as demonstrated by detection with monoclonal anti-CA, clone M88 and G18 in ELISA and Western immunoblotting (data not shown). Cell lysate of BV-H_6-CA infected Sf9 cells was subjected as antigen for proving the target compartment of selected ankyrin binders.

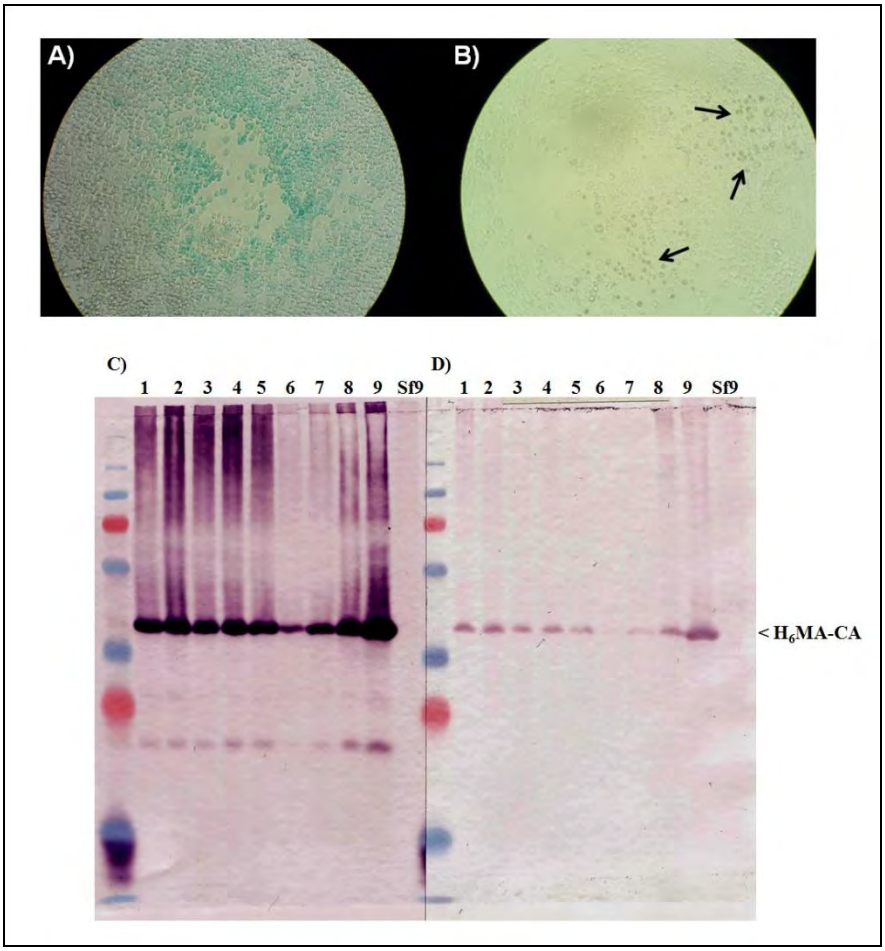

Figure 17 Isolation of recombinant BV-H_6MA-CA. (A) The blue plaque appearance of Sf9 cells infected with recombinant BV-H_6MA-CA. **(B)** The precipitated polyhedrin protein as indicated with arrow heads of Sf9 infected with wild-type AcMNPV. Western immunoblotting result demonstrated the presence of recombinant protein in cells infected with virus from each plaques using anti-His tag **(C)** and anti-Gag **(D)** antibodies, respectively. The number indicated the given number of virus from isolated plaques. The lysate of non-infected Sf9 cell was used as negative control.

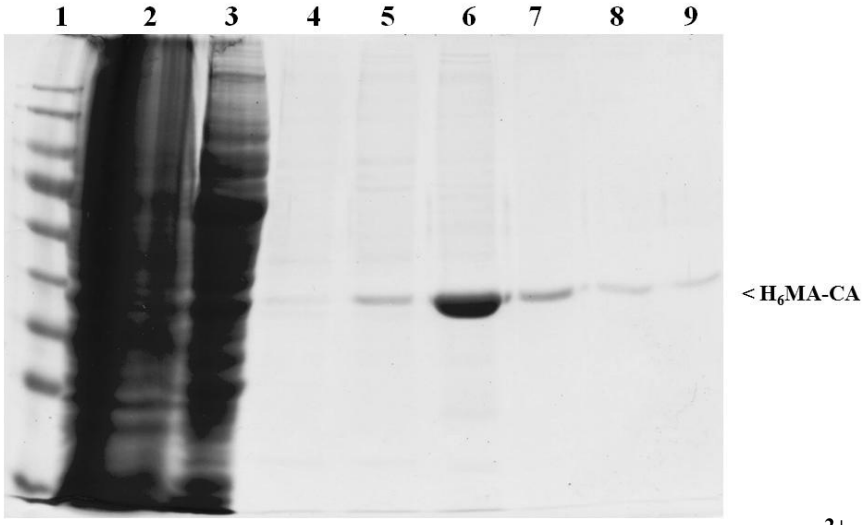

Figure 18 Purification of recombinant H_6MA-CA protein using Ni^{2+}-NTA-agarose column. Lane 1, pre-stained protein marker; Lane 2, cell lysate; Lane 3, flow-through fraction; Lane 4-9, eluted fractions. All samples were separated under denaturing conditions in 12% SDS-PAGE. The separated protein in the gel was visualized by Coomassie blue R250 staining.

Isolation of H_6MA-CA and A3 binders by phage selection

The phage display library used for the selection process was amplified from naïve and filtrate libraries and collected as two forms which are in precipitated form and in culture supernatant without precipitation process. The objective to use these two forms was to evaluate that the precipitation process causes the phages stick together and results in the non-specific binding. The phage titration from the first round of selection showed that precipitated phages from both naïve and filtrate library had background 10 times over the phage from supernatant while phage input and output were found at the same number. This finding guided us to use the phage without precipitation for lowering the background in next round of selection. Phage screening was performed after the second round of selection. Focusing on H_6MA-CA binders, interestingly, one reactive clone of five tested clone was detected from filtrate library after the first round of selection. Moreover, the more number of positive clones were observed from filtrate library (10 of 33) comparing with that from naïve library (4 of 32) as shown in **Figure 19A**. To further performing the third round of selection, phages input were amplified form a pool clones of eluted phage form filtrate library. In addition, to recover the specific binders, soluble antigen was used for elution instead of using acid solution. Released phages were collected at time points. Randomized clones of each eluted fractions were selected for phage screening as shown in **Figure 19B**. Almost 70% of clones in all elute fractions represented the binding activity over the background. All positive clones from first and second round, and top five clones which showed the high binding activity of each elute fractions in third round were examined the number of repeat by restriction analysis. Most of selected binders (32 of 39) contained 2 internal repeats as shown the molecular size at 370 base pairs and the other binders appeared at 470 base pairs referring to 3 repeats (**Figure 20**). The sequences of positive clones were analyzed by standard sequencing method. From sequencing results, selected binders were divided into two major groups relying on the number of internal repeats. In the group containing two repeats (**Figure 21A**), first group, four identical sequences were found from the first and second round of selection (1A9, 1A12, 1C12, 1, and 1D10). Another clone, 1D3, had nine different amino acids located in variable positions. Reactive candidates from third round were classified in three groups. 6A5, 7F7,

and 6D12 have the same sequences as the first group at variable positions but different amino acids in consensus. Sequences of three clones, 6E9, 7D11, and 7G7, hold some different positions in consensus and the most of random positions. Another group, 7A8 and 7C6, also differ in some positions in consensus and fives residues in variable positions. Focusing on group containing three repeats (**Figure 22B**), four groups could be classified. Clone 1B4, 1B8, and 1D8 were the most abundant sequences where as clone 1D4, 6B4, and 6G5 had its individual sequence. Interestingly, the most observed amino acid substitutions which are E17K, V18F, D28E in N-capping repeat occurred in the consensus of binders selected from the third round. Finally, clones containing three internal repeats without substitutions in consensus, 1B8, and 1D4, and also 6B4 were selected to investigate the binding activity *in vitro*.

In this study, A3 protein was also used as the target molecule for evaluating this constructed artificial ankyrin repeat library in term of possibility to be selected specific binders. This protein is an artificial alpha-helicoidal repeat protein (αRep) based on thermostable HEAT-like repeats as described (Urvoas et al., 2010). It's also folded cooperatively and very stable. Reactive binders could be isolated. In the first round of selection, five of 33 tested clones from naïve library gave the positive signals while none of positive clone was found from filtrate phage (data not shown). Moreover, many binders could be selected from the third round of selection. These positive clones were characterized the repeat number and sequence as described (data not shown). The binding activities of these reactive clones (2D3, and 2E3) with the most abundant sequence and without substitutions in consensus were further evaluated by ELISA and Isothermal Titration Calotimetry (ITC)

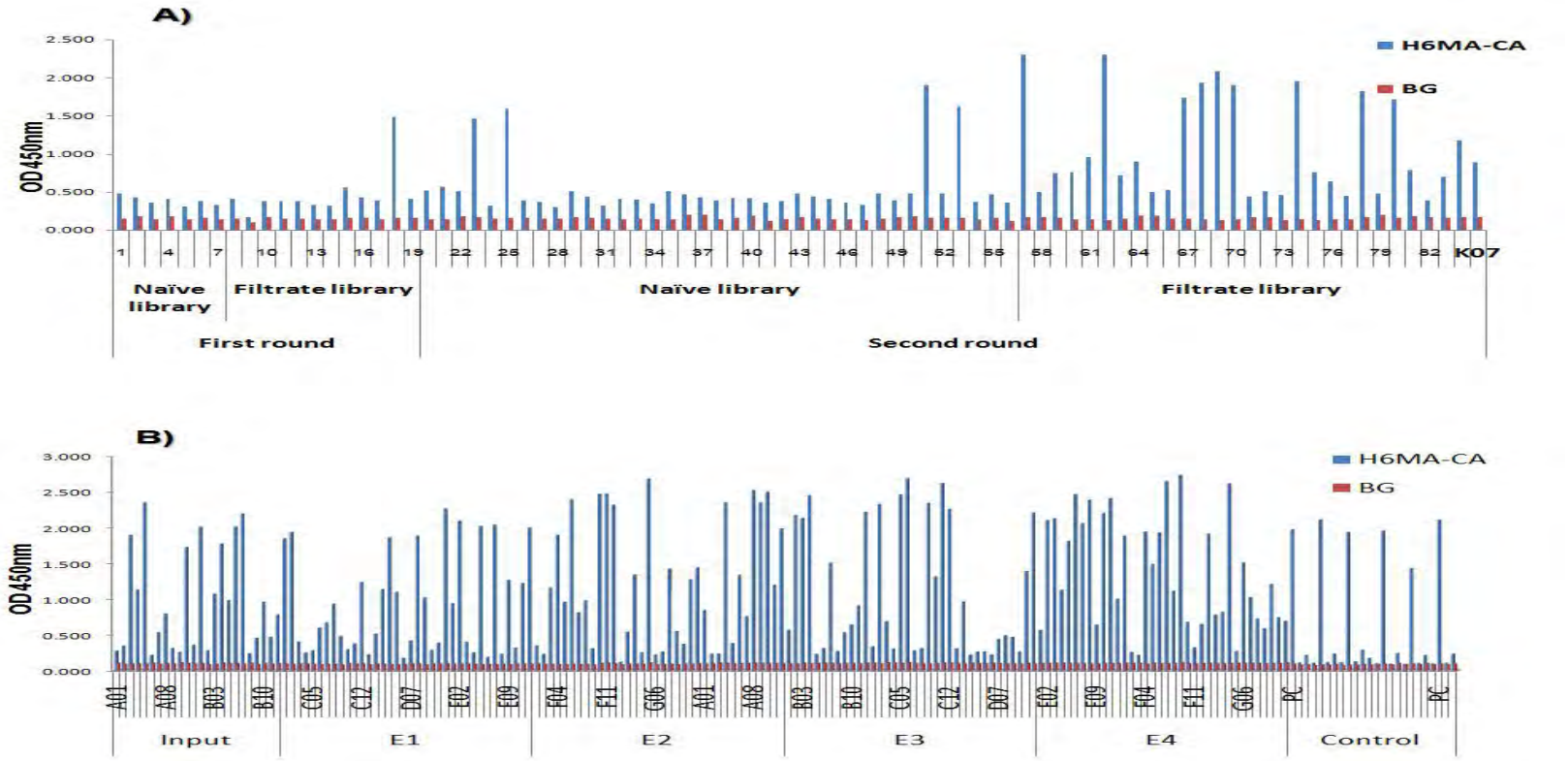

Figure 19 Phage ELISA for screening of the H_6MA-CA binders. Culture supernatants containing phage particles from first and second round of selection (**upper panel**) and third round selection (**lower panel**) were added into H_6MA-CA coated wells and un-coated wells (BG). The bound phages were traced with HRP-conjugated anti-M13 antibody. The signal was measured at 450 nm after developing the TMB substrate. The positive and negative clones of the first round were used as control in the screening of third round. The sequential elution of third round selection using free H_6MA-CA was indicated as E1, E2, E3, and E4.

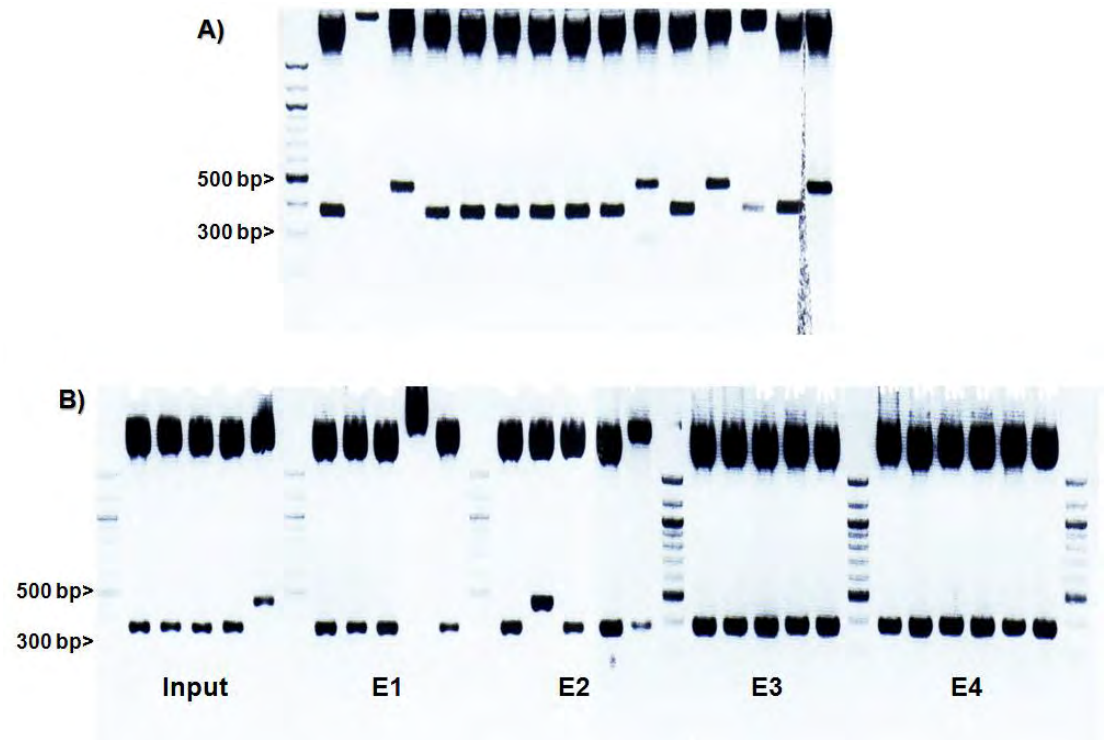

Figure 20 Distribution of repeat number of selected H_6MA-CA binders. Plasmids were extracted from positive clones of first and second round **(A)**, and third round **(B)** of selection and double treated with *Not* I and *Hind* III. Treated vectors were analyzed by gel electrophoresis. Clones from sequential elution were indicated as input, E1, E2, E3, and E4, respectively.

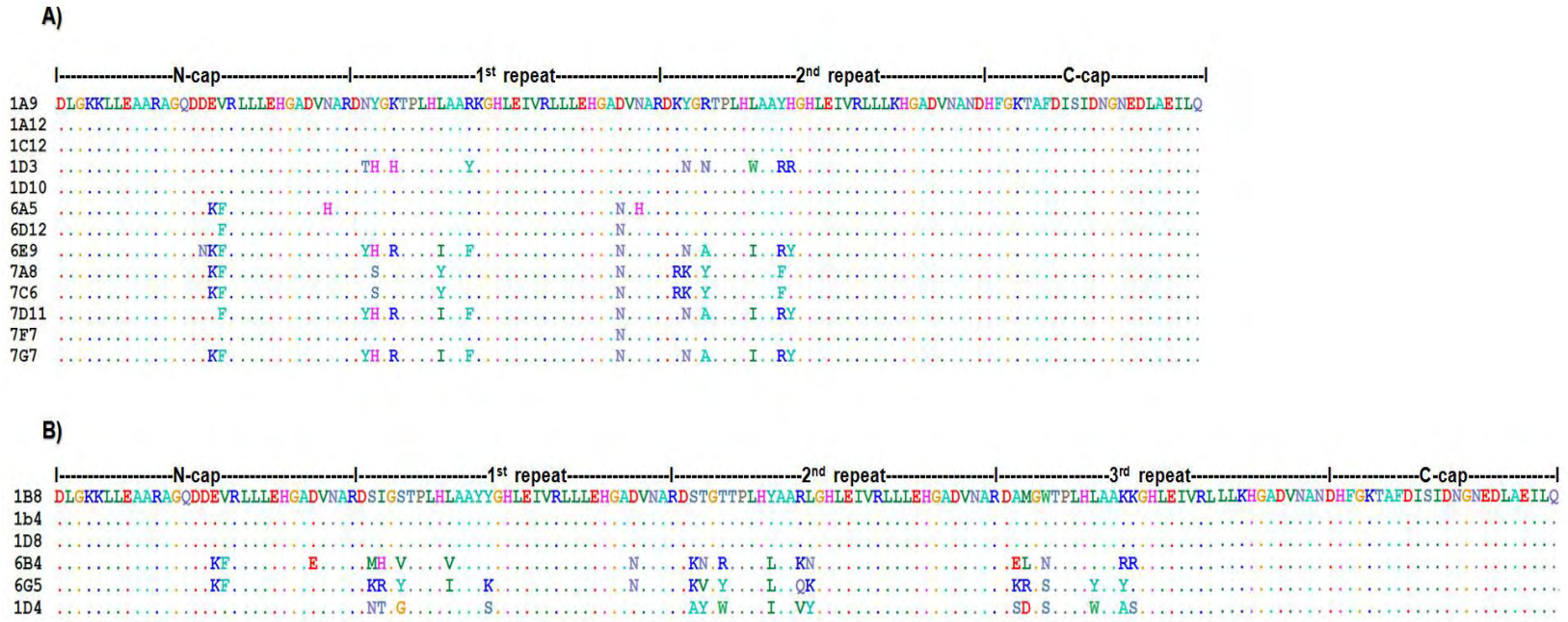

Figure 21 Sequence analyses of H_6MA-CA binders. Selected binders were analyzed the sequence and divided into two groups relying on their internal repeat number which are two repeats **(A)** and three repeats **(B)**.

Generation of monoclonal antibodies against H_6MA-CA

The titers of polyclonal antibodies gainst H_6MA-CA after the third immunization from mice were determined by indirect ELISA. Polyclonal antibodies of mouse A had the higher signal than that of mouse B at the same dilution without any background (**Figure 22**). Spleenocytes of mouse A were collected to generate the hybridoma cells. The culture supernatants from wells containing hybridoma cells were collected and screened for antibody reactivity by ELISA and Western immunoblotting.

Culture supernatants of 278 clones were taken for screening the reactivity by ELISA. As shown in **Figure 23A**, thirteen of all tested clones showed the high signal over the background demonstrating the reactivity of secreated antibody from hybridoma clones but another clones expressed not significant signal over background (data not shown). All of positive clones were further characterized the epitope by Western immunoblotting (**Figure 23B**). Cell lysate of BV-H_6MA-CA infected cells were seperated and transferred to the PVDF membrane. Anti-His tag and anti-Gag were used for localized the molecular size of H_6MA-CA fusion protein and partially cleaved form of H_6MA at ~40 kDa and ~17kDa, respectively. No reactive band was observed on membranes reacted with clones: G106, M7, M8, M25, G57, M68, and M102 concluding that these clones bind to the conformational epitope. Another clones which are G46, G52, G53, and M48 exhibited 2 positive bands (~40 kDa and ~17kDa). This result indicated that their epitope presented on MA domain. Furthermore, clones G18, and M88 bind to CA domain as revealed two reactive bands at ~40 kDa and ~24 kDa.

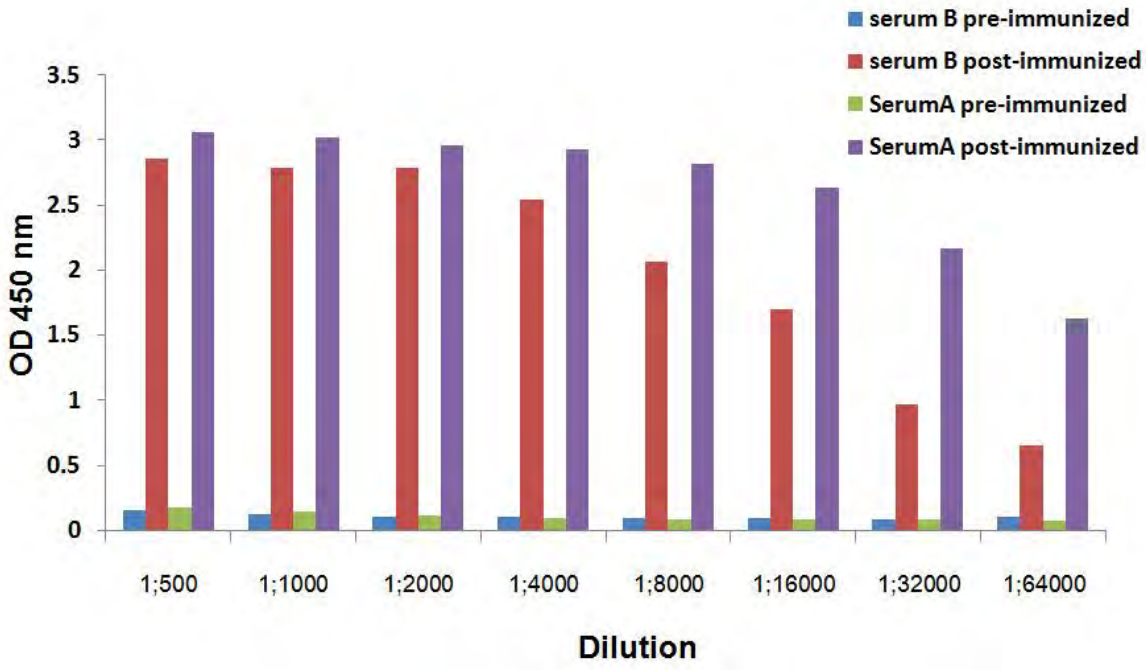

Figure 22 The reactivity of polyclonal antibodies in immunized mice serum. The serial diluted serum were collected after third immunization and detemined the antibody reactivity by indirect ELISA. The pre-immunized sera were tested at the same time.

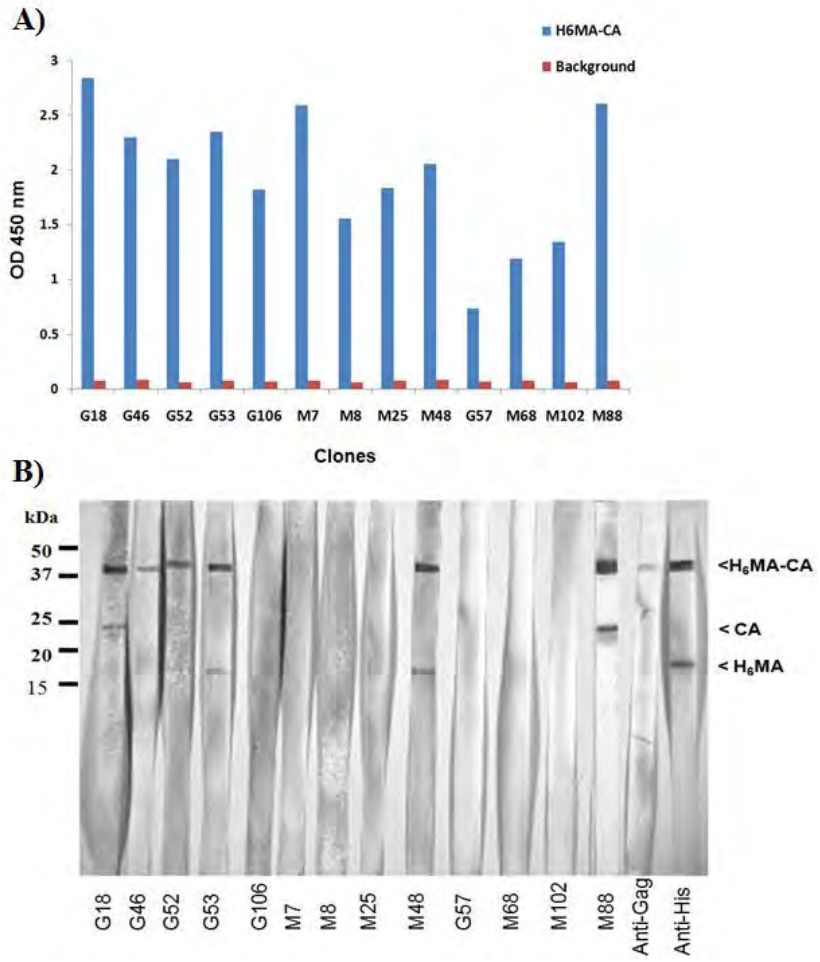

Figure 23 Characterization of antibody reactivity. (A) Culture supernatants from hybridoma clones were added into H_6MA-CA coated wells. The captured antibodies were detected by HRP-conjugated goat anti-mouse immunoglobulins and signals were measured at 450 nm after adding substrate and HCl. **(B)** Cell lysate of infected cells was separated and probed with culture supernatants. The bound antibodies on membrane were revealed by HRP-conjugated goat anti-mouse immunoglobulins and the picture were captured after adding TMB membrane substrate. The molecular sizes of recombinant proteins were indicated in the picture.

Binding activity of selected ankyrin binders

Three clones of reactive binders were first chosen to evaluate the binding activity *in vitro*. Gene coding for these binders were cut from phagemid vector and subsequently cloned into pQE-30 expression vector for high level of protein expression in *E. coli* M15 [pREP4] strain. Ankyrin binders were expressed and purified using HisTrap column follow by size exclusion using Sephadex G-75. The purity of purified ankyrin binders were evaluated by running in 15% SDS-PAGE under reducing condition with overload amount of protein and visualizing by Coomassie's blue staining (**Figure 24A**). Hence ankyrin binders and H_6MA-CA protein hold the same tag; this might limit detection system in ELISA-based method. Purified ankyrin binders were chemically biotinylated using EZ-Link Sulfo-NHS-LC-Biotin kit and verified by dot blot (**Figure 24B**). All of three selected candidates could be conjugated with biotin as shown the blue

color developed from reaction in the second membrane comparing with no signal developed from non-biotinylated protein.

The reactivity of biotinylated binders was assessed by dot blot analysis (**Figure 25A**). H_6MA-CA or A3 coated on membrane were reacted with biotinylated proteins and subsequently traced by HRP-conjugated extravidin. Reactive dots were visualized by adding TMB membrane. As shown in **Figure 25A**, blue dots were observed only at H_6MA-CA. This result indicated that ankyrin binders remained their binding activity after biotinylation and also expressed the specificity to their antigen. To confirm the specific interaction, biotinylated binders were mixed with its non-biotinylated form and irrelevant binders (2D3, the A3 binder) and then added into H_6MA-CA coated wells. The bound-biotinylated binders were detected by reacting with HRP-conjugated extravidin. Interestingly, as presented in **Figure 25B**, the inhibition could be observed in the mixture of biotinylated-protein with its free form (blue bar) while irrelevant ankyrin showed not significant effect (red bar). This result confirmed that, all three selected binders bound to their target molecule. Moreover, 1D4 exhibited the highest signal follow by 1B8, and 6B4 respectively.

Furthermore, competitive ELISA was done for evaluating the specificity of selected binders with various inhibitors. As resulted in **Figure 26**, the inhibition effect was observed while ankyrin clones 1B8, 1D4, 6B4 and soluble H_6MA-CA were used as inhibitors for biotinylated-1B8, and -1D4 but not ankyrin clone 2E3 which is A3 binders or A3 protein. These results indicated that all binders might react at the same or nearby epitopes on target protein. Remarkably, clone 1D4 represented the highest binding signal with target protein (blue bar, no inhibitor) and exhibited the best inhibiting effect. This clone was, hence, chosen for evaluating its binding activity by ITC method. Additionally, the binding activity of these binders was evaluated by performing sandwich ELISA using monoclonal antibodies against CA domain as demonstrated the method in **Figure 27C, upper panel**. Biotinylated-ankyrin 1D4 and 1B8 could bind to the captured-target protein on the well (**Figure 27C, lower panel**). These observed results confirmed the binding reaction of previous experiments. Besides, Biotinylated-antibodies also exhibited the binding activity on both antibodies-coated wells indicating that these two antibodies binds to different epitopes or the multiple recognition sites presented on the wells.

The epitope localization of ankyrin binders was analyzed by western immunoblotting as described. Clones 1B8, 1D4, and 6B4 reacted to the separated proteins on membrane and resulted positive band at molecular size of H_6MA-CA but not at H_6MA (**Figure 28A**). The indirect ELISA using H_6-CA recombinant protein captured on Ni-coated wells was performed to confirm the recognition epitope of these three ankyrin binders. Positive signals were detected with all H_6MA-CA binders (1B8, 1D4, and 6B4) but not A3 binder (2D3) (**Figure 28B**). This result indicated that, the epitope of these three H_6MA-CA binders was situated at CA domain.

Additional confirmation of affinity and specificity was obtained from isothermal titration calorimerty (ITC) experiments. Titration of increasing amounts of 1D4 into sample cell containing purified H_6MA-CA provided a dissociation constant (K_d) of 0.45 μM (**Figure 28A**). ITC also confirmed specificity, as 1D4 binding to A3 was not detected in this assay (**Figure 28B**) whereas 2D3 reacting to A3 exhibited a K_d of 18 nM (**Figure 28C**). These data also validated the competition assays as an accurate measurement of affinity. Moreover, the useful obtained data form a single ITC experiment was the stoichiometry (N) which deals with the quantitative relationships between protein and its ligand or the number of binding sites. From the fitting curve, the calculated stoichiometry value was 0.656 indicating the molar ratio of H_6MA-

CA and 1D4 was around two. This finding supported the result of sandwich ELISA. The reason would be explained by the dimerization of this protein (**Figure 28D**) as described in previous study (Gamble et al., 1997) resulting in the multiple recognition site that presented on the wells.

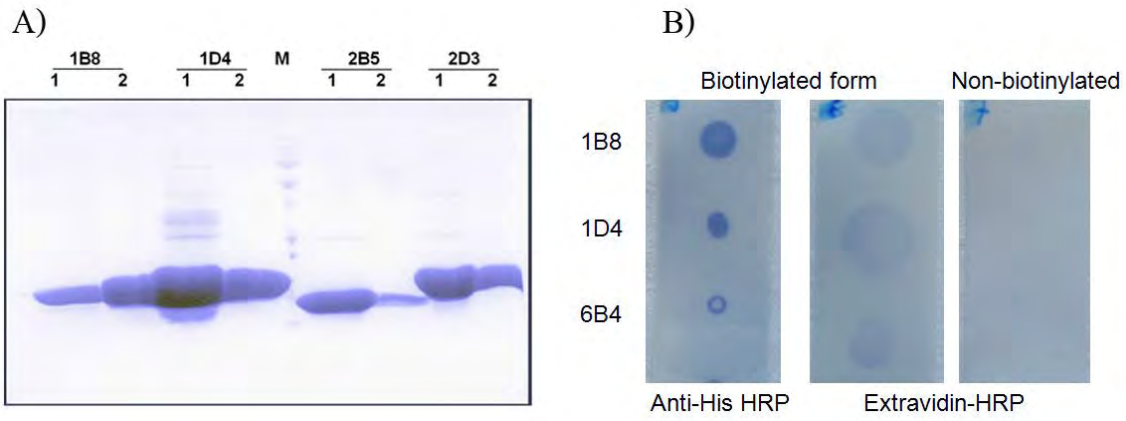

Figure 24 Production and biotinylation of purified ankyrin binders. Ankyrin binders were expressed in *E. coli* and purified using two step; HisTrap column (**A, Lane 1 of each clone**) followed by size exclusion column (**A, Lane 2 of each clone**). Purified proteins were chemically linked with biotin and verified by dot blot analysis **(B)** using HRP conjugated anti-His and extravidin. The same amount of non-biotinylated protein was employed as control.

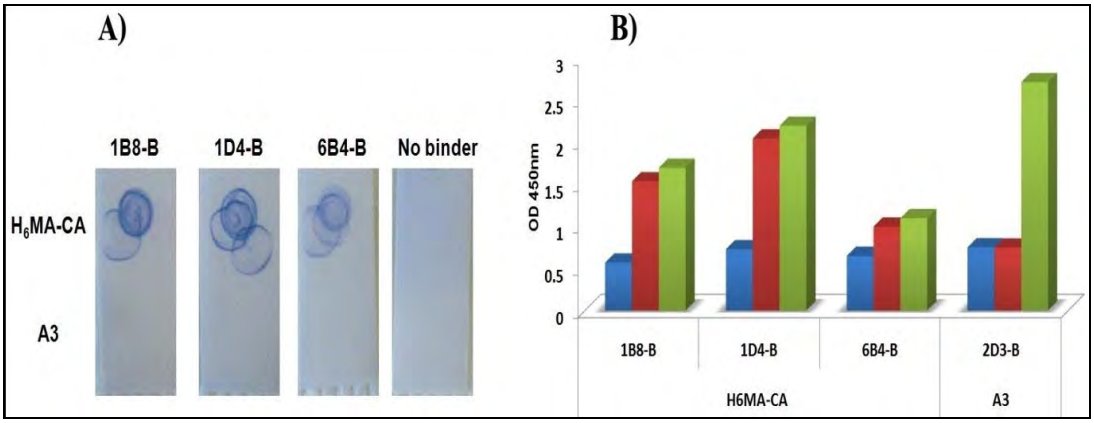

Figure 25 Binding activities of biotinylated ankyrin binders. (A) Dot blot analysis. Biotinylated-ankyrin binders were reacted with H_6MA-CA and A3 coated on membrane. The reactive dots were developed precipitating substrate after revealing membrane with HRP-conjugated extravidin. **(B)** Competitive ELISA. The mixture of biotinylated-binders with its free form (,Blue bars), with irrelevant ankyrin binders (,Red bars), and with no inhibitor (, Green bars) were added into H_6MA-CA- or A3-coated wells as indicated. Bound-ankyrin binders were detected by adding HRP-conjugated extravidin. The signal was measured after TMB substrate developing.

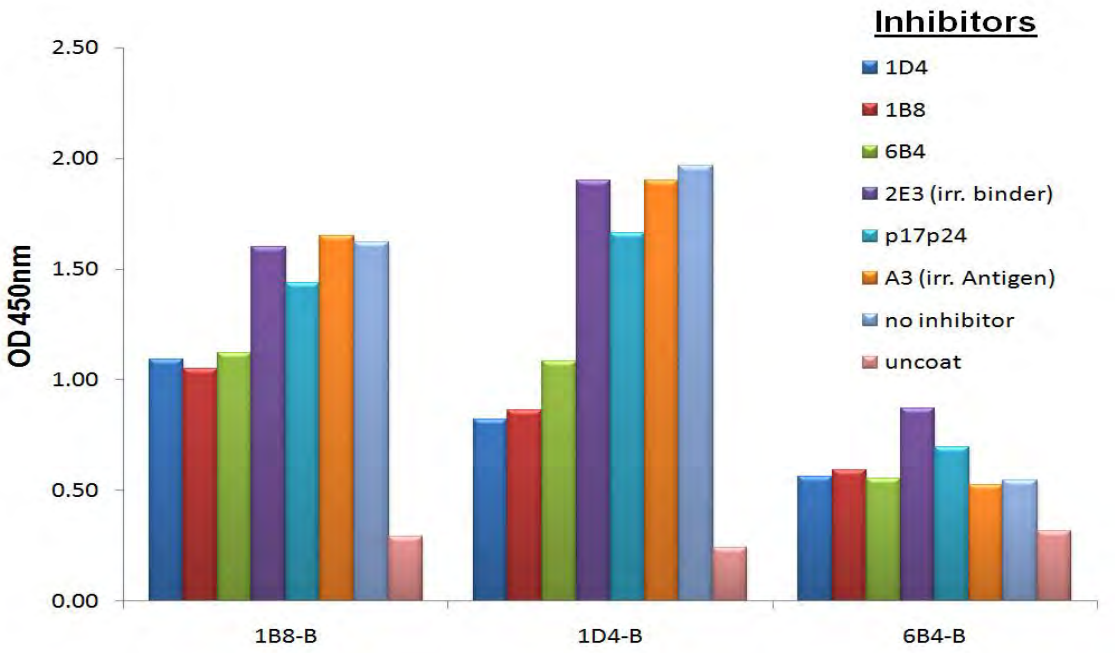

Figure 26 Specific reactivity of ankyrin binders. The binding of Biotinylated-ankyrin binders were inhibited with various inhibitors including free H_6MA-CA binders, H_6MA-CA protein. A3 binders and A3 protein were used as negative control.

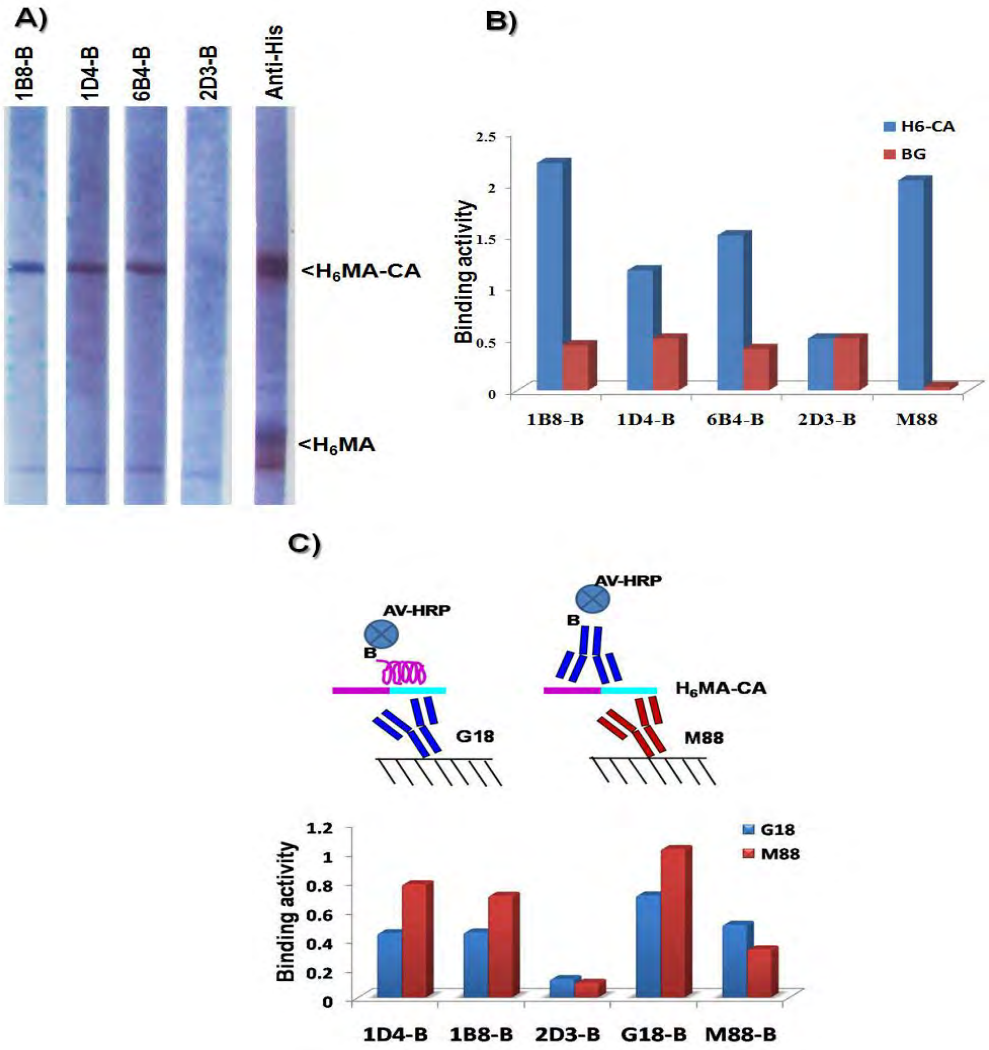

Figure 27 Epitope localization of selected binders. (A) Binding reactivity of ankyrin with cell lysate of BV-H_6MA-CA blotted membrane. Anti-histidine tag was used for indicating the molecular size of target protein. **(B)** Indirect ELISA of H_6-CA captured on Ni-treated plate. Monoclonal anti-CA (clone M88) and ankyrin 2D3 were used as positive control and irrelevant binder control. **(C)** Schematic demonstrating the sandwich ELISA, upper panel. The binding reactivity of biotinylated-binders reacting with the antibody-captured H_6MA-CA on the wells, lower panel and with the cell lysate of BV-H_6MA-CA blotted membrane

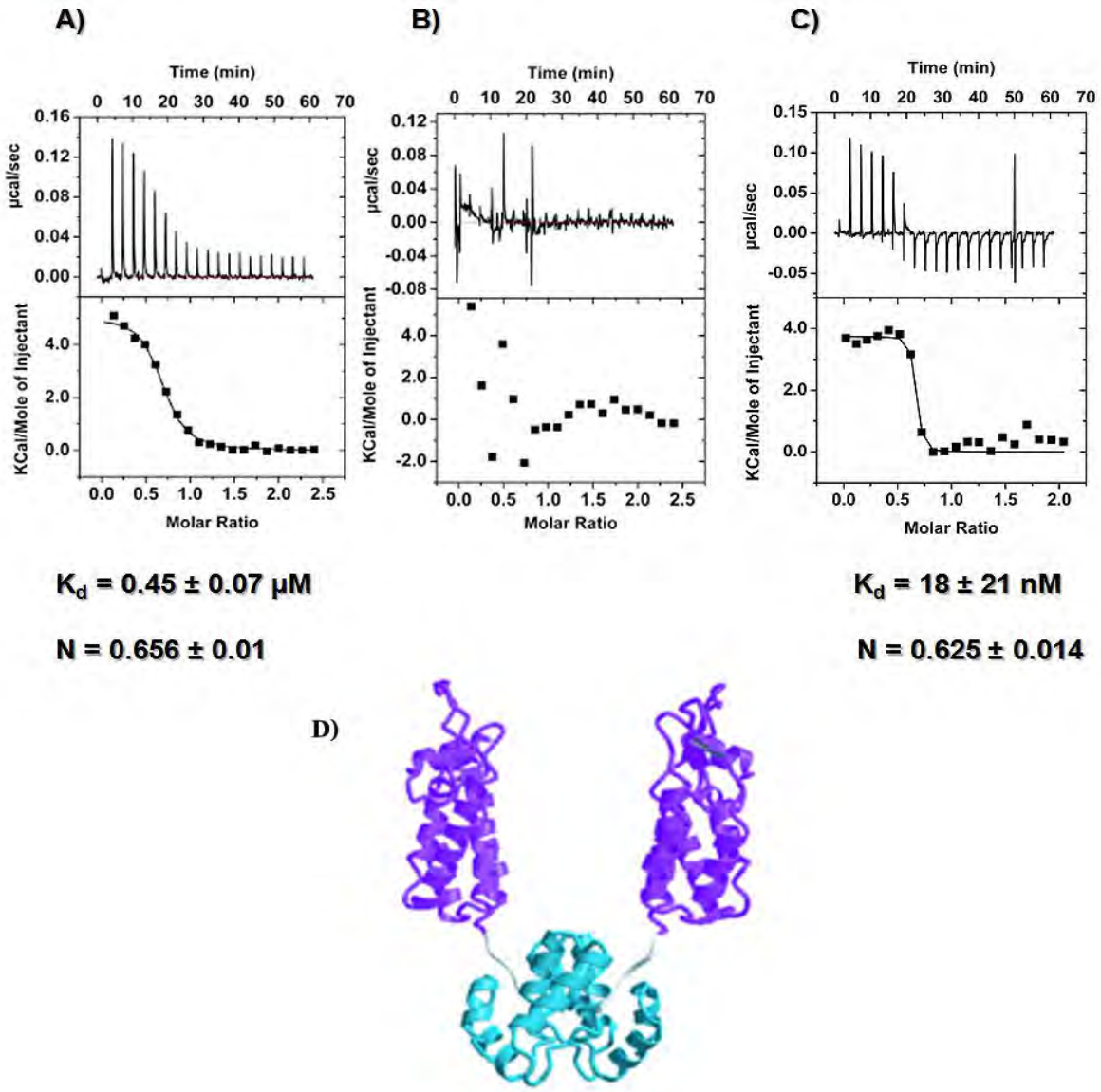

Figure 28 Affinity and specificity of selected ankyrin binder. ITC experiment was performed to measure the accurate binding activity of 1D4 with H_6MA-CA **(A)**, 2D3 with A3 **(C)**, and to evaluate the specific binding of 1D4 **(B)**. The picture represented model for the intact HIV-1 capsid dimer **(D)**. The CA(146–231) dimer (cyan) is shown covalently linked to the CA(1–151) domain (Gamble et al., 1997)

Hindrance of HIV-1 maturation by interfering protease activity

HIV-1 protease plays a key role in viral maturation by cleavage of the nascent polypeptide into mature proteins. In our lab, we have developed method based on ELISA to detect the activity of HIV-1 protease and its variants by taking an advantage of antibody against C-terminus of MA domain, anti-p17, (manuscript is in preparation). In this study, this method was applied to assess the interfering effect of selected ankyrin binders to block the cleavage process. H_6MA-CA protein captured on nickel-treated plate was reacted with ankyrin binders

before treated with HIV-protease. The cleaved form of substrate was evaluated by detecting the presence of C-terminus of MA. As shown in **Figure 29**, positive signal was observed as same as no inhibitor control. This result demonstrated that all three selected clones exhibited no interfering effect on HIV-1 protease activity

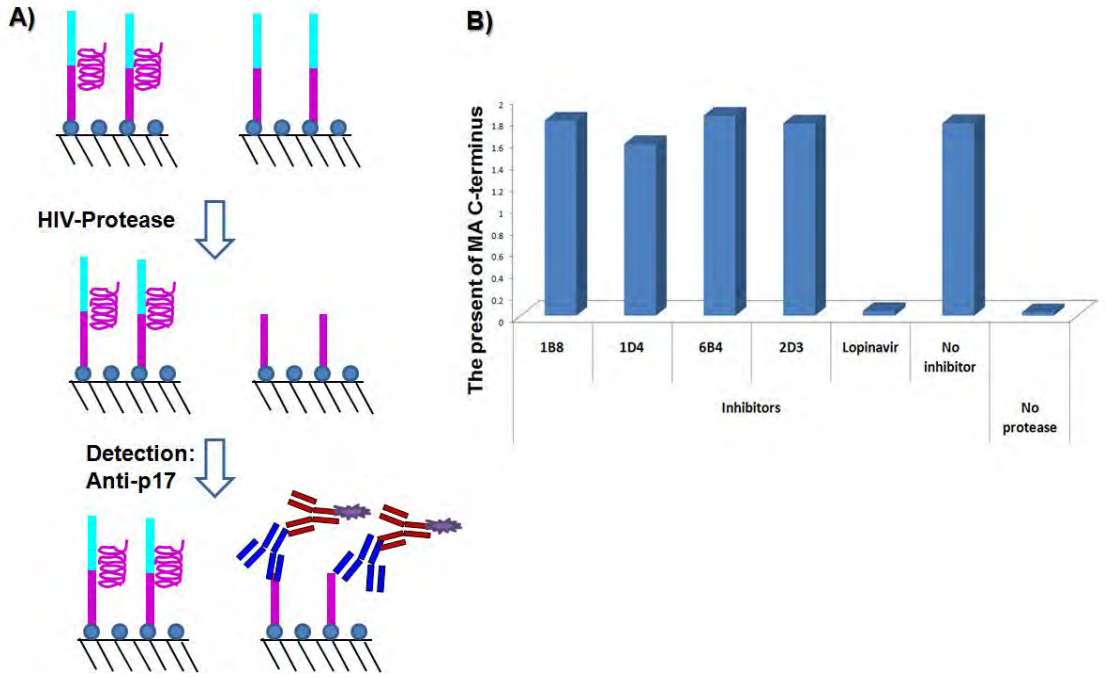

Figure 29 Interference of HIV-1 protease activity by selected ankyrin binders. (A) Schematic represents ELISA system for testing the protease activity. H_6MA-CA –captured on microwells were reacted with ankyrin binders before treating with protease. The cleaved form of substrate was detected using monoclonal anti-p17 and subsequently traced by goat anti-mouse IgS conjugated HRP. **(B)** Interfering effect of selected ankyrin binders on HIV-1 protease activity. Lopinavir was used as a protease inhibitor control.

Intracellular function of ankyrin protein on viral assembly

DNA fragment encoding ankyrin binder clone 1D4 was separated amplified using two sets of primers to generate two versions of expression vectors. Each product was subsequently performed overlapping PCR with gene coding for green fluorescence protein (GFP) which is used as a reporter and then subcloned into pCEP4-based vector. The resulting vectors which are pCEP4Myr1D4-GFP containing myristoylation signal at N-terminus of ankyrin for targeting protein to membrane and pCEP4Cyt1D4-GFP for cytoplasmic expression as designated in **Figure 30**. For the establishment of stable lines, these two vectors were separately transfected into SupT1 cell line. pCEP4-based plasmid is an episomal mammalian expression that uses the cytomegalovirus (CMV) immediate early enhancer/ promoter for high level transcription of

recombinant genes. Epstein-Barr Virus replication origin (oriP) and nuclear antigen which is encoded by this plasmid to permit extrachromosomal replication in mammalian cells. Moreover, this vector also carries the hygromycin B resistance gene for stable selection in transfected cells. Thus, in this experiment, transfected cells were maintained in medium containing hygromycin B.

The stable cells were observed the protein expression under the fluorescence microscopy (**Figure 31A**) and flow cytometer (**Figure 31B**). Interestingly, myristoylation signal efficiently directed the GFP-fused ankyrin protein to the membrane as shown the green fluorescence located at the plasma membrane (**Figure 31A upper panel**). In constrast, non-myristoylated ankyrin diffused throughout the cells (**Figure 31B, lower panel**). Additional data from flow cytometry demonstrated that 95% GFP positive cells were observed with both stable lines. These resulted indicated that almost of the cells expressed ankyrin protein and ready to be evaluated its intracellular function. Culture supernatants containing viral particles were collected after eleven days of challenging Sup-T1 stable lines with 5 MOI of HIV_{NL4-3} to determine p24 level. The result showed that the level of p24 observed in both stable lines was significantly decreased comparing with Sup-T1 control. Surprisingly, stable line expressing membrane-bound 1D4 demonstrated the superior effect on disturbance of viral replication than cytoplasmic 1D4. This result revealed that directing therapeutic molecules to functional area of target molecules is the convincing concept.

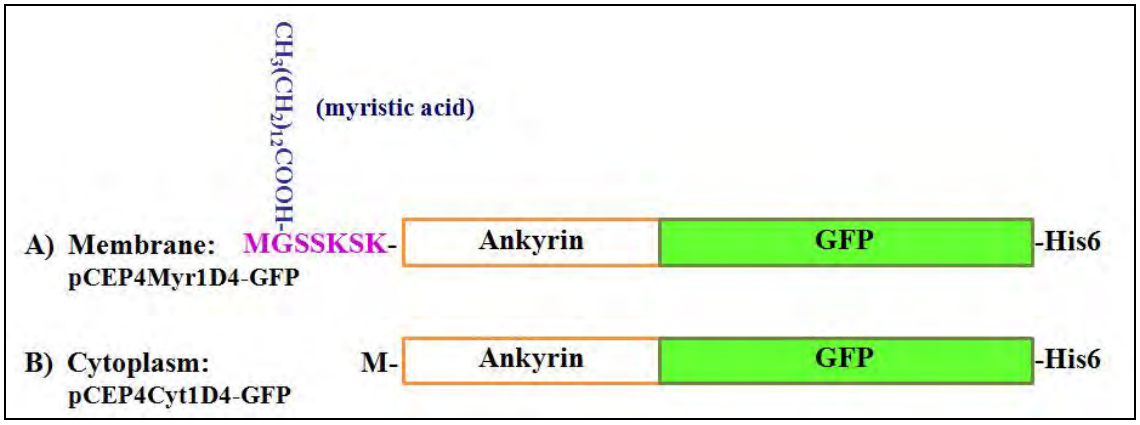

Figure 30 Schematic representation of two constructed vector for stable lines generation. pCEP4-based vector was inserted with two different DNA fragment coding ankyrin binder (1D4), green fluorescence protein (GFP) and Histidine tag with myristoylation signal **(Membrane Myr^{+}, A)** and without signal **(Cytoplasm, Myr^{-}, B)** at N-terminus. Myristic acid covalently linked with glycine (G) was highlighted in blue letters.

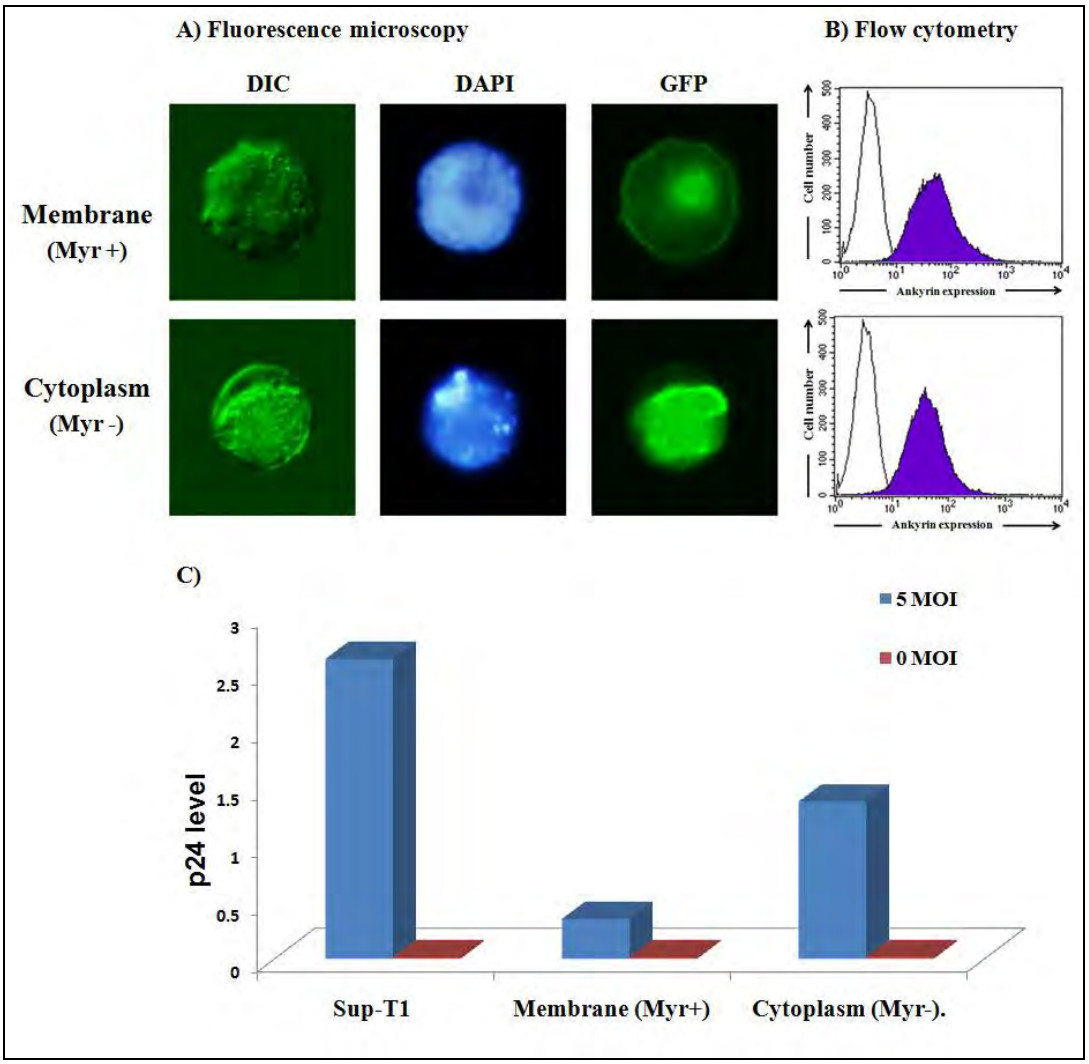

Figure 31 Interference of HIV assembly by ankyrin binder 1D4 in different cellular compartments. (A) Intracellular expression and localization of ankyrin binder 1D4 were observed under fluorescence microscopy. Cell shape, nucleus, and GFP-fused ankyrin protein were visualized by bright view (DIC), DAPI staining (DAPI), and green fluorescence channel (GFP), respectively. **(B)** Flow cytometry demonstrated the population of GFP positive cells. **(C)** The level of p24 measured from stable cell lines challenged with HIV_{NL4-3}. Sup-T1 was used as control.

CONCLUSION

This study aimed to discover novel protein-based agents as intracellular inhibitors for intervening HIV replication by gene-targeting strategy. Considering the superior of ankyrin repeat proteins to antibodies, this molecular scaffold was selected to generate the artificial protein library and select for the specific binding molecules to HIV matrix (MA) and capsid proteins (CA). High amount of target molecule which expressed as fusion protein to histidine tag (H_6MA-CA) was produced in insect cells by using baculovirus expression system. The artificial library was constructed with respect to mimic the natural ankyrin sequence. The amino acids at random positions were designed to serve the distribution occurred in nature and consensus residues were modified at some positions to create the recognition site for restriction enzyme without disturbing the conformation of ankyrin.

Several specific binders to H_6MA-CA were isolated from this library by phage display technology. Besides, a number of A3 specific binding molecules were isolated in parallel to evaluate the efficiency of this constructed library. The DNA fragments encoding specific binder were subsequently transferred to cytoplasmic expression vector to produced high level of soluble proteins. The binding activity of isolated candidates was analyzed by various techniques. In addition, monoclonal antibodies against H_6MA-CA were generated in parallel. All taken ankyrin binders, 1D4, 1B8, and 6B4, bind specifically to target molecule with distinct degree of activity as shown by ELISA. The specific site of these ankyrin binders was located at CA domain as demonstrated by western immunoblotting and indirect ELISA. Moreover, this area is not overlap to the epitope of monoclonal antibodies against CA domain. The best ankyrin candidate (1D4) was selected to further evaluate its binding constant and functions. Measuring by ITC method, 1D4 had its binding constant about 0.45 μM with the molar ratio to H_6MA-CA at 0.6 indicating one to two mole of 1D4 to H_6MA-CA.

In order to examine the activity of 1D4 *in vivo*, two individual vectors were constructed to generate the stable Sup-T1 cell lines expressing membrane-bound and cytoplasmic 1D4 for HIV challenge. The preliminary data of viral assembly interference showed the remarkable decreasing of p24 (CA) level from cells harboring the membrane-bound 1D4 in contrast to the cytoplasmic 1D4 expressing cells. This result provides a novel concept of targeting HIV inhibiting molecules intracellulary to interfere the normal life-cycle in viral assembly. The forthcoming aspect of this discovery will be significant for stem cell gene-therapy.

REFERENCES

1. Berkhout, B., *Toward a durable anti-HIV gene therapy based on RNA interference*. Ann N Y Acad Sci, 2009. 1175: p. 3-14.
2. Check, E., *Pioneering HIV treatment would use interference and gene therapy*. Nature, 2005. 437(7059): p. 601.
3. Luque, F., et al., *Gene therapy for HIV-1 infection: are lethal genes a valuable tool?* Cell Mol Biol, 2005. 51(1): p. 93-101.
4. Von Laer, D., S. Hasselmann, K. Hasselmann, *Gene therapy for HIV infection: what does it need to make it work?* J Gene Med, 2006. 8(6): p. 658-67.
5. Trkola, A., et al., *Delay of HIV-1 rebound after cessation of antiretroviral therapy through passive transfer of human neutralizing antibodies*. Nat Med, 2005. 11(6): p. 615-22.
6. Binz, H.K. and A. Plückthun, *Engineered proteins as specific binding reagents*. Curr Opi Biotechnol, 2005. 16(4): p. 459-469.
7. Forrer, P., et al., *A novel strategy to design binding molecules harnessing the modular nature of repeat proteins*. FEBS Letters, 2003. 539(1-3): p. 2-6.
8. Hey, T., et al., *Artificial, non-antibody binding proteins for pharmaceutical and industrial applications*. Trends in Biotechnology, 2005. 23(10): p. 514-522.
9. Nygren, P.A. and A. *Skerra, Binding proteins from alternative scaffolds*. J Immunol Methods, 2004. 290(1-2): p. 3-28.
10. Hosse, R.J., A. Rothe, and B.E. Power, *A new generation of protein display scaffolds for molecular recognition*. Protein Sci, 2006. 15(1): p. 14-27.
11. Montgomery, J., et al., *Identification and predicted structure of a leucine-rich repeat motif shared by Leishmania major proteophosphoglycan and Parasite Surface Antigen 2*. Mol Biochem Parasitol, 2000. 107(2): p. 289-95.
12. Groves, M.R. and D. Barford, *Topological characteristics of helical repeat proteins*. Curr Opin Struct Biol, 1999. 9(3): p. 383-9.
13. Sedgwick, S.G. and S.J. Smerdon*, The ankyrin repeat: a diversity of interactions on a common structural framework*. Trends Biochem Sci, 1999. 24(8): p. 311-6.
14. Wensing, A.M. and C.A. Boucher, *Worldwide transmission of drug-resistant HIV*. AIDS Rev, 2003. 5(3): p. 140-55.
15. Lee, C.C., et al., *Primary drug resistance and transmission analysis of HIV-1 in acute and recent drug-naive seroconverters in Singapore*. HIV Med, 2009. 10(6): p. 370-7.
16. Tang, J.W. and D. Pillay, *Transmission of HIV-1 drug resistance.* J Clin Virol, 2004. 30(1): p. 1-10.
17. Strayer, D.S., et al., *Current status of gene therapy strategies to treat HIV/AIDS*. Mol Ther, 2005. 11(6): p. 823-842.
18. Dornburg, R. and R.J. Pomerantz, *HIV-1 gene therapy: promise for the future*. Adv Pharmacol, 2000. 49: p. 229-61.
19. Rondon, I.J. and W.A. Marasco, *Intracellular antibodies (intrabodies) for gene therapy of infectious diseases*. Annu Rev Microbiol, 1997. 51: p. 257-83.
20. Marasco, W.A., W.A. Haseltine, and S.Y. Chen, *Design, intracellular expression, and activity of a human anti-human immunodeficiency virus type 1 gp120 single-chain antibody*. Proc Natl Acad Sci U S A, 1993. 90(16): p. 7889-93.

21. Vercruysse, T., et al., *An intrabody based on a llama single-domain antibody targeting the N-terminal alpha-helical multimerization domain of HIV-1 rev prevents viral production*. J Biol Chem, 2010. 285(28): p. 21768-80.
22. Levin, R., et al., *Inhibition of early and late events of the HIV-1 replication cycle by cytoplasmic Fab intrabodies against the matrix protein, p17*. Mol Med, 1997. 3(2): p. 96-110.
23. Bagasra, O., *Potent inhibition of human immunodeficiency virus type 1 replication by an intracellular anti-Rev single-chain antibody.* Proc Natl Acad Sci U S A, 1998. 95(17): p. 10344.
24. Levy-Mintz, P., et al., *Intracellular expression of single-chain variable fragments to inhibit early stages of the viral life cycle by targeting human immunodeficiency virus type 1 integrase*. J Virol, 1996. 70(12): p. 8821-32.
25. Dearden, C.E., *Role of antibody therapy in lymphoid malignancies*. Br Med Bull, 2007. 83(1): p. 275-290.
26. Blakey, D.C., *Drug targeting with monoclonal antibodies*. Acta Oncol, 1992. 31(1): p. 91-7.
27. Jia, L., et al., *Screening of human antibody Fab fragment against HBsAg and the construction of its dsFv form*. Int J Biol Sci, 2008. 4(2): p. 103-10.
28. Dantas-Barbosa, C., M.M. Brigido, and A.Q. Maranhao, *Construction of a human Fab phage display library from antibody repertoires of osteosarcoma patients*. Genet Mol Res, 2005. 4(2): p. 126-40.
29. Bostrom, J. and G. Fuh, *Design and construction of synthetic phage-displayed Fab libraries*. Methods Mol Biol, 2009. 562: p. 17-35.
30. McAuley, A., et al., *Contributions of a disulfide bond to the structure, stability, and dimerization of human IgG1 antibody CH3 domain*. Protein Sci, 2008. 17(1): p. 95-106.
31. Skerra, A., *Engineered protein scaffolds for molecular recognition*. J Mol Recognit, 2000. 13(4): p. 167-87.
32. Tewari, R., et al., *Armadillo-repeat protein functions: questions for little creatures*. Trends Cell Biol. 20(8): p. 470-481.
33. Koprivova, A., et al., *Identification of a pentatricopeptide repeat protein implicated in splicing of intron 1 of mitochondrial nad7 transcripts*. J Biol Chem.
34. Mosavi, L.K., et al., *The ankyrin repeat as molecular architecture for protein recognition*. Protein Sci, 2004. 13(6): p. 1435-48.
35. Andrade, M.A., C. Perez-Iratxeta, and C.P. *Pointing, Protein repeats: structures, functions, and evolution*. J Struct Biol, 2001. 134(2-3): p. 117-31.
36. Nord, K., et al., *Binding proteins selected from combinatorial libraries of an alpha-helical bacterial receptor domain*. Nat Biotechnol, 1997. 15(8): p. 772-7.
37. Skerra, A., *'Anticalins': a new class of engineered ligand-binding proteins with antibody-like properties*. J Biotechnol, 2001. 74(4): p. 257-75.
38. Abedi, M.R., G. Caponigro, and A. Kamb, *Green fluorescent protein as a scaffold for intracellular presentation of peptides*. Nucleic Acids Res, 1998. 26(2): p. 623-30.
39. Koide, A., et al., *The fibronectin type III domain as a scaffold for novel binding proteins*. J Mol Biol, 1998. 284(4): p. 1141-51.

40. Breeden, L. and K. Nasmyth, *Similarity between cell-cycle genes of budding yeast and fission yeast and the Notch gene of Droso*phila. Nature, 1987. 329(6140): p. 651-4.
41. Li, J., A. Mahajan, and M.D. Tsai, *Ankyrin repeat: a unique motif mediating protein-protein interactions*. Biochemistry, 2006. 45(51): p. 15168-78.
42. Kohl, A., et al., *Allosteric Inhibition of Aminoglycoside Phosphotransferase by a Designed Ankyrin Repeat Protein*. Structure, 2005. 13(8): p. 1131-1141.
43. Binz, H.K., et al., *Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins*. J Mol Biol, 2003. 332(2): p. 489-503.
44. A, K., et al., *Allosteric inhibition of a kinase by a designed ankyrin repeat protein inhibitor*. Structure, 2005.
45. Christian Zahnd, et al., *A Designed Ankyrin Repeat Protein Evolved to Picomolar Affinity to Her2*. J. Mol. Biol., 2007. 369: p. 1015-1028.
46. Huber, T., et al., *In vitro selection and characterization of DARPins and Fab fragments for the co-crystallization of membrane proteins: The Na+-citrate symporter CitS as an example*. J. Struct. Biol, 2007. 159(2): p. 206-221.
47. Interlandi, G., et al., *Characterization and Further Stabilization of Designed Ankyrin Repeat Proteins by Combining Molecular Dynamics Simulations and Experiments*. J. Mol. Biol, 2008. 375(3): p. 837-854.
48. Vogel, M., al., *Designed Ankyrin Repeat Proteins as Anti-Idiotypic-Binding Molecules*. Ann. N.Y. Acad. Sci. , 2007. 1109: p. 9-18.
49. Plückthun, A., et al., *Intracellular kinase inhibitors selected from combinatorial libraries of designed ankyrin repeat proteins*. J Biol Chem 2005.
50. Sergeeva, A., et al., *Display technologies: application for the discovery of drug and gene delivery agents*. Adv Drug Deliv Rev, 2006. 58(15): p. 1622-54.
51. Smith, G.P., *Filamentous fusion phage: Novel expression vectors that display cloned antigens on the virion surface*. Science, 1985. 228(4705): p. 1315-1317.
52. Paschke, M., *Phage display systems and their applications*. Applied Micro. Biotechnol, 2006. 70(1): p. 2-11.

LIST of PUBLICATIONS

1. Praditwongwan W, Chuankhayan P, Saoin S, Wisitponchai T, Lee VS, Nangola S, Hong SS, Minard P, Boulanger P, Chen CJ, Tayapiwatana C*. Crystal structure of an antiviral ankyrin targeting the HIV-1 capsid and molecular modeling of the ankyrin-capsid complex. J Comput Aided Mol Des. 2014 Jul 5.
2. Nangola S, Urvoas A, Valerio-Lepiniec M, Khamaikawin W, Sakkhachornphop S, Hong SS, Boulanger P, Minard P, Tayapiwatana C*. Antiviral activity of recombinant ankyrin targeted to the capsid domain of HIV-1 Gag polyprotein. Retrovirology 2012, 9;17.
3. Nangola S, Minard P, Tayapiwatana C*. Appraisal of translocation pathways for displaying ankyrin repeat protein on phage particles. Protein Expr Purif. 2010, 74; 156-61.

Protein Expression and Purification 74 (2010) 156–161

Contents lists available at ScienceDirect

Protein Expression and Purification

journal homepage: www.elsevier.com/locate/yprep

Appraisal of translocation pathways for displaying ankyrin repeat protein on phage particles

Sawitree Nangola [a,b], Philippe Minard [b], Chatchai Tayapiwatana [a,c,*]

[a] Division of Clinical Immunology, Department of Medical Technology, Faculty of Associated Medical Sciences, Chiang Mai University, Chiang Mai 50200, Thailand
[b] Laboratoire de Modélisation et d'Ingénierie des Protéines, UMR8619, Université de Paris-Sud, Orsay Cedex 91405, France
[c] 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

ARTICLE INFO

Article history:
Received 2 April 2010
and in revised form 20 August 2010
Available online 25 August 2010

Keywords:
Phage display
Signal sequences
Protein translocation pathways
Ankyrin repeat protein
pIII domain

ABSTRACT

Depending on the molecular properties of the proteins of interest (POI), the rate of success in displaying proteins on phage particles is unpredictable. Formation of polypeptide tertiary structure in the cytoplasm occasionally results in low level display on viral particles. Here we assessed the influence of different leader peptides on the display of a premature cytoplasmic folding protein, ankyrin repeat protein (ARP), *via* the minor coat protein pIII. These peptides include the Sec, SRP and Tat pathways. The results demonstrated that the Sec and SRP pathways were capable of displaying the protein on the viral particle, whereas the Tat pathway failed to do so. Interestingly, the Tat pathway efficiently directed ARP through its translocon without fusing with pIII. Furthermore, the soluble form of ARP was detected in *Escherichia coli* periplasm.

Introduction

Phage display, first described by George Smith in 1985, is a technique with a high potential for expressing functional recombinant proteins (phenotypes) by fusing their responsible gene (genotype) to a gene encoding the phage minor (pIII)[1] or major (pVIII) coat protein [1]. In order to display peptides or proteins on the phage particles, the gene encoded protein of interest (POI) is cloned into a phagemid vector between the signal sequence and the phage coat proteins. The signal sequences encoded by the phagemid vector are recognized by several translocation systems. The major mechanisms of proteins translocation across the cytoplsmic membrane in *Escherichia coli* are the Sec pathway, the signal recognition particle (SRP) pathway, and the twin-arginine translocation (Tat) pathway. The Sec machinery is the major route of targeting processes which direct post-translational proteins in an unfolded state, while the SRP pathway is the most important for cotranslational proteins. These two systems drive proteins through the SecYEG channel. The Tat pathway is independent of the Sec pathway, and is capable of secreting folded proteins through the Tat translocon [2].

Series of phagemid vectors containing various signal sequences corresponding to the different translocation pathways have been established to improve the efficiency for displaying certain proteins on phage particles. Certain favorable signal sequences, such as the OmpA signal sequence (*OmpA*ss), the PhoA signal sequence (*PhoA*ss), or the PelB signal sequence (*PelB*ss) involving the Sec pathway have been used in phagemid vectors and prokaryotic expression vectors [3–5]. However, the limiting factor of using these vectors is only an unfolded conformation of the protein could be transported across the cytoplasmic membrane of *E. coli*. Several molecules would not be facilitated through this machinery because of the premature folding in cytoplasm such as an intracellular carboxylesterase of *Bacillus subtilis* or green fluorescent protein (GFP) [6,7]. This problem has been circumvented by directing these two proteins *via* the Tat pathway [7]. In addition, this system has been used to display CD147 molecule on phage particles *via* pVIII [8]. Since the Tat system has a folding quality-control, some proteins such as an antibody derivative, scFv, or other disulfide bond containing protein require the appropriate condition in periplas-

* Corresponding author at: Division of Clinical Immunology, Department of Medical Technology, Faculty of Associated Medical Sciences, Chiang Mai University, Chiang Mai 50200, Thailand. Fax: +66 53 946042.
E-mail addresses: sawitree2727@hotmail.com (S. Nangola), philippe.minard@u-psud.fr (P. Minard), asimi002@hotmail.com, asimi002@chiangmai.ac.th (C. Tayapiwatana).

[1] *Abbreviations used:* pIII, gene III coat protein; pVIII, gene VIII coat protein; POI, protein of interest; SRP, signal recognition particle; Tat, twin-arginine translocation pathway; *OmpA*ss, OmpA signal sequence; *PhoA*ss, PhoA signal sequence; *PelB*ss, PelB signal sequence; scFv, single variable fragment; ARP(s), ankyrin repeat protein(s); *DsbA*ss, DsbA signal sequence; TorAss, TorA signal sequence; MOI, multiplicity of infection; PEG, polyethylene glycol; BSA, bovine serum albumin; ELISA, enzyme-linked immunosorbent assay; IPTG, Isopropyl β-D-1 thiogalactopyranoside; TSD, Tris-sucrose-dithiothreitol method; SDS–PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; PVDF, polyvinylidene fluoride (PVDF) membrane; PBS, phosphate buffer saline; ECL, enhanced chemiluminescent substrate; kDa, kilo Dalton.

1046-5928/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.pep.2010.08.010

mic space for folding that result in the efficiently transported by this system [9]. Recently, the SRP pathway has been applied for improving the translocation efficiency of various molecules including scFv binding to gpD and thioredoxin of *E. coli* [10].

Focusing on the premature cytoplasmic folding protein, especially ankyrin repeat proteins (ARPs), the successful secretion of this protein family is complicated by its exceptional stability. Successful display of very stable proteins may have important practical applications as this is a prerequisite to evaluate their potential as novel classes of alternative scaffolds (specific binding molecules, with physical properties more favorable than natural antibodies). Natural ARPs are presented in the cytoplasm of cells and function in the protein–protein interaction of numerous proteins [11]. The careful analysis of the sequence determinant of the ARPs has been able to generate the design of a new family of artificial ankyrin repeat proteins named Designed Ankyrin Repeat Proteins (DARPins) [12]. A combinatorial library based on this design was created and used for isolating high-affinity binding molecules specific to various target molecules by ribosome display [13–16]. DARPins are very stable molecules and are naturally fast folding in the cytoplasm, which may give rise to inefficient translocation to the periplasm and consequently impact on the level of display using phage particles. This phenomenon has been evaluated by comparing the effect of two secretory pathways, the Sec and SRP pathways, on the displaying efficiency of DARPins. The Tat pathway, however, was not included in the report [10].

Since the Tat system allows secretion of folded proteins from the *E. coli* cytosol, it might take advantage over the Sec system. In particular, for those proteins which possess accelerated folding or are highly stable, to be reversing to unfolded conformation before translocation machinery could take place. Thus, it is desirable to assess the possibilities of the Tat pathway on the translocation of ARP not only for the purpose of phage display, but also in regard to the intracellular isolation technique [17]. In this study, we compared the influence of three leader peptides involved in the Sec pathway (*Ompa*ss), the SRP pathway (*Dsba*ss), and the Tat pathway (*Tora*ss) on the display of ARP on phage particles. In addition, the affect of pIII domain was also evaluated in regards to the displaying of protein on phage particle using the Tat pathway.

Materials and methods

Construction of vectors

Three series of phagemid vectors containing *Ompa*ss, *Tora*ss and *Dsba*ss were constructed. These constructs based on the E3_5 backbone as the synthetic ankyrin model are 96% identical to the generated Ank15 [12]. There are seven amino acid substitutions (N58H, S65Y, L67N, T68D, T70H, A78S, and T79N) located in the variable positions of the ankyrin repeats that differentiate Ank15 from the previously described E3_5. To generate phagemid vector containing *Ompa*ss, the gene encoding Ank15 was amplified from vector pHDiEx-Ank15 (P Minard, unpublished data) using primers FwAnk_*Sfi* I and RevAnk_*Sfi* I (Table. 1). The PCR product was treated with *Sfi* I, and then cloned into the *Sfi* I site of the pComb3H phagemid vector (kindly provided by C. F. Barbas, Scripps Institute, USA). The phagemid vector containing *Tor*Ass was created from two oligonucleotides, FAnk_*Xho* I and RAnk_*Spe* I (Table. 1), which were used as primers to amplify the ankyrin gene from pHDiEx-Ank15 vector. The PCR product was treated with *Xho* I and *Spe* I and then cloned into *Xho* I and *Spe* I sites of pTat8-CD147 phagemid [8] and transformed into XL-1 Blue competent cells (Stratagene, La Jolla, CA). The resulted vector was further digested with *Sap* I and *Spe* I and cloned into restricted pComb3H phagemid vector. To generate the vector containing *Dsba*ss, two single-stranded oligonucleotides, DsbA F_*Nde* I and DsbA R_*Not* I (Table. 1) were mixed and hybridized by heating at 95 °C for 5 min followed by cooling from 95 °C to 25 °C at the rate of 0.1 °C/s. This resulted in the double-stranded oligonucleotide which generated the *Nde* I site at the 5 end and the *Not* I site at the 3 end which was then cloned into pHDiEx-Ank15 vector in place of the *Ompa*ss and *Strep*-tag II genes. All of the ligation products were separately transformed into the *E. coli* XL-1 Blue competent cells. The transformed cells were selected on Luria–Bertani (LB) agar containing 100 μg/ml of ampicillin. Finally, the DNA sequence from each clone was analyzed by standard DNA sequencing methods using *Bgl* II sense primer. The constructed vectors were designated pCom3H-Ank15, pTat3-Ank15 and pHDiExDsbA-Ank15 respectively. The pTat3-Ank15 was further transformed into HB2151 non-suppressor *E. coli* host strain (kindly provided by Dr. A.D. Griffiths, MRC, Cambridge, UK) for expression of Ank15 without pIII domain.

Table 1
The sequences of oligonucleotides using for vector constructions.

Primer name	Sequence*
FwAnk_*Sfi* I	5'-GAGGAGGAGCTGGCCAGGCGGCCGCGGCATGGTCTCATCCTCAG-3' (*Sfi* I)
RevAnk_*Sfi* I	5'-GAGGAGGAGGTGGCCGGCCTGGCCGTGATGGTGATGGTGGTGGTT-3' (*Sfi* I)
DsbA F_*Nde* I	5'-TATGAAGAAAATCTGGCTGGCCCTGGCAGGCCTGGTTCTGGCTTT (*Nde* I) CAGCGCTTCCGCTGCGTGGTCCCATCCGCAATTCGAAAAAGC-3' (*Not* I)
DsbA R_ *Not* I	5'-GGCCGCTTTTTCGAATTGCGGATGGGACCACGCAGCGGAAGCGCT (*Not* I) GAAAGCCAGAACCAGGCCTGCCAGGGCCAGCCAGATTTTCTTCA-3' (*Nde* I)
Bgl II sens	5'-GCAGAGATCTGCGAAATTAATAC-3'
FAnk_*Xho*I	5'-GAGGAGGAGCTCTCGAGGCGGCATGGTCTCATCCT-3' (*Xho* I)
RAnk_*Spe*I	5'-GAGGAGGAGGTACTAGTGTGATGGTGATGGTGGTG-3' (*Spe* I)

* Nucleotide sequences recognized by restriction endonucleases used for cloning of DNA fragments were underlined with the enzyme names.

Phage production

To amplify phage displaying Ank15 protein on their particles, 10 ml of 2YT broth containing 1% (w/v) glucose, ampicillin (100 μg/ml), and tetracycline (10 μg/ml) were inoculated with the glycerol stocks of *E. coli* XL-1 Blue harboring the constructed vectors and then cultivated at 37 °C with shaking at 200 rpm for 18 h. Fifty microliters of pre-cultured cells were transferred to 50 ml of 2YT broth and cultivated at 37 °C with shaking at 200 rpm until the OD at 600 nm reached 0.6. The bacterial culture was further infected with 20 MOI of KO7 helper phages, and then incubated at 37 °C for 30 min without shaking followed by 37 °C for 30 min with shaking. Phage-infected cells were centrifuged at 3000 rpm for 10 min at 25 °C. Pellets were resuspended and cultured in 50 ml of 2YT broth containing ampicillin (100 μg/ml), and kanamycin (70 μg/ml) with shaking at 200 rpm at 30 °C for 16–18 h. Bacteria were centrifuged at 3000 rpm for 10 min at 4 °C, and the phages were harvested by PEG/NaCl (20% (w/v) PEG 8000, 2.5 M NaCl) precipitation. Finally, the pellets were resuspended with 10 mM Tris-buffer saline (TBS) pH 7.4.

Detection of fusion protein on phage particles by ELISA

The serial dilutions of phage particles in binding buffer (25 mM Tris–HCl, 2 mM EDTA, 140 mM NaCl, pH 7.6) containing 2% (w/v) BSA were added into *Strep*-Tactin coated microtiter plates (IBA GmbH, Göttingen, Germany) and incubated at room temperature for 1 h. After washing, a horseradish peroxidase conjugated anti-M13 monoclonal antibody (GE Healthcare, Piscataway, NJ) was

added and incubated at room temperature for 1 h. The TMB/H_2O_2 substrate (KPL, Gaithersburg, MD) was added after the washing step. Finally, 1 N HCl was added prior absorbance measuring at 450 nm.

Protein expression and extraction

The bacterial cells harboring the vector were grown in 10 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), at 37 °C for 18 h with shaking. One hundred microliters of pre-cultured bacteria were inoculated in 100 ml of the same medium containing 1% (w/v) glucose and ampicillin (100 μg/ml) and continued shaking at 37 °C until the optimal density (OD) at 600 nm reached 0.8. To induce protein expression, IPTG was added to the culture at a final concentration of 1 mM. The bacteria were cultivated at 30 °C for 4 h. After the induction time, the bacterial cells were then collected by centrifugation at 3000 rpm for 30 min at 4 °C.

The extraction of protein from the whole cell lysate was obtained following the instructions of B-Per II bacterial protein extraction reagent (PIERCE, Rockford, IL). The periplasmic protein was extracted using the Tris–sucrose–dithiothreitol (TSD) method. This method has previously been described to selectively extract periplasmic proteins without cytoplasmic proteins [18]. Briefly, the cells were washed three times with 20 mM Tris–HCl pH 7.5 and then treated with 200 μl of TSD buffer (100 mM Tris–HCl, pH 8.0, 15% sucrose and 1 mM dithiothreitol) and incubated on ice for 10 min. The cell pellet was collected by centrifugation at 12,000 rpm for 10 min at 4 °C and resuspended gently with 200 μl of deionized water then incubated on ice for 10 min. The periplasmic fraction was collected by centrifugation at 12,000 rpm for 10 min at 4 °C. The whole cell lysates and periplasmic fractions were analyzed by 12% SDS–PAGE and Western blot analysis.

SDS–PAGE and western blot analysis

The whole cell lysates, periplasmic fractions, and 1×10^{12} particles of each phage were separated in a 12% SDS–PAGE gel under reducing condition. The separated proteins in SDS–PAGE were analyzed by Coomassie blue staining and Western blot analysis. Proteins were transferred onto a polyvinylidene fluoride (PVDF) membrane (Amersham Biosciences, Buckinghamshire, UK). To block non-specific binding, 5% (w/v) skim milk in PBS was added to the blotted membrane to block non-specific binding. After washing with 0.05% (v/v) Tween 20 in PBS, the membrane was incubated with mouse anti-pIII mAb (Exalpha Biologicals, Inc., Watertown, MA) and monoclonal anti-His tag antibody (GenScript, Piscataway, NJ). Following the washing step, peroxidase-conjugated goat anti-mouse immunoglobulins (KPL, Gaithersburg, MD) were added to the membrane. The peroxidase reaction was visualized by using an enhanced chemiluminescent (ECL) substrate detection system (GE Healthcare, Buckinghamshire, UK). The reactive bands were visualized by exposing the membrane to X-ray film (Kodak, Rochester, NY).

Results

Phagemid vectors with three difference leader peptides

To appraise the effect of difference translocation pathways on phage displaying ARP on viral particles, three phagemid vectors were constructed and designated as pComb3H-Ank15, pTat3-Ank15, and pHDiExDsbA-Ank15. Each vector contains *OmpA*ss, *TorA*ss, and *DsbA*ss, respectively followed by the gene encoding *Strep*-tag II, Ank15, Histidine tag and C-terminal domain of phage protein III (residues 275–424) under the IPTG inducible lac promoter. Amino acid sequences of each signal sequence are shown in the single letter code. The amino acid sequence of Ank15 was presented in Fig. 1. There are seven amino acid substitutions that differentiate Ank15 from the previously described E3-5 (N58H, S65Y, L67N, T68D, T70H, A78S and T79N). A production scheme of all vectors is shown in Fig. 2. All vectors were transformed into bacterial cell strain XL-1 Blue for phage production and HB2151 for soluble protein production.

The expression of ankyrin repeat protein on phage particles

Phage particles, namely OmpA, TorA and DsbA phage were amplified from bacteria containing pComb3H-Ank15, pTat3-Ank15 and pHDiExDsbA-Ank15 vectors, respectively. For the detection of fusion proteins on phage particles, phage displaying *Strep*-tag II fused-Ank15 protein was analyzed by phage ELISA using *Strep*-Tactin coated microtiter plates. The detection of bound phage with an anti-M13 antibody conjugated with HRP was shown as observed absorbance values (Fig. 3). The signal of DsbA phage was 1.4-fold higher than the OmpA phage at the same dilution. No such positive results were observed for the TorA phage and

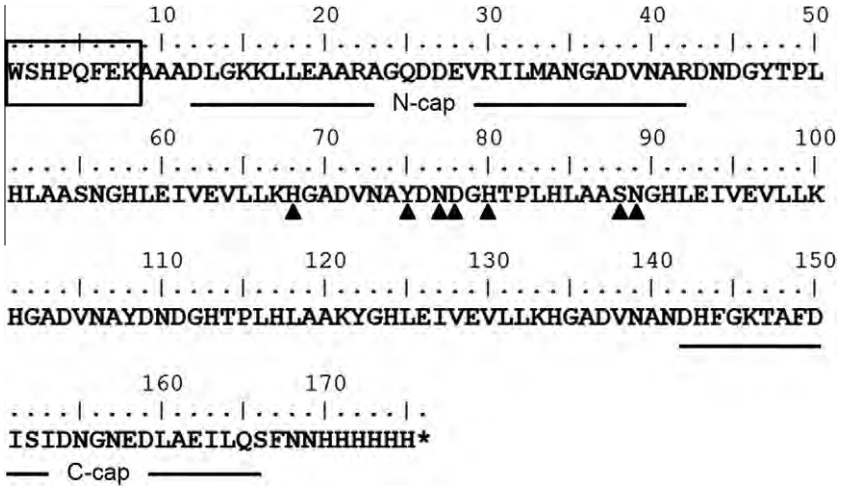

Fig. 1. The amino acid sequence of Ank15. The *Strep*-tag II sequence is highlighted in the box. N-cap and C-cap refer to the *n*-terminal capping repeat and the c-terminal capping repeat of Ank15 protein, respectively. The closed triangles represented the positions of amino acid substitution in the internal repeats of E3_5 [12].

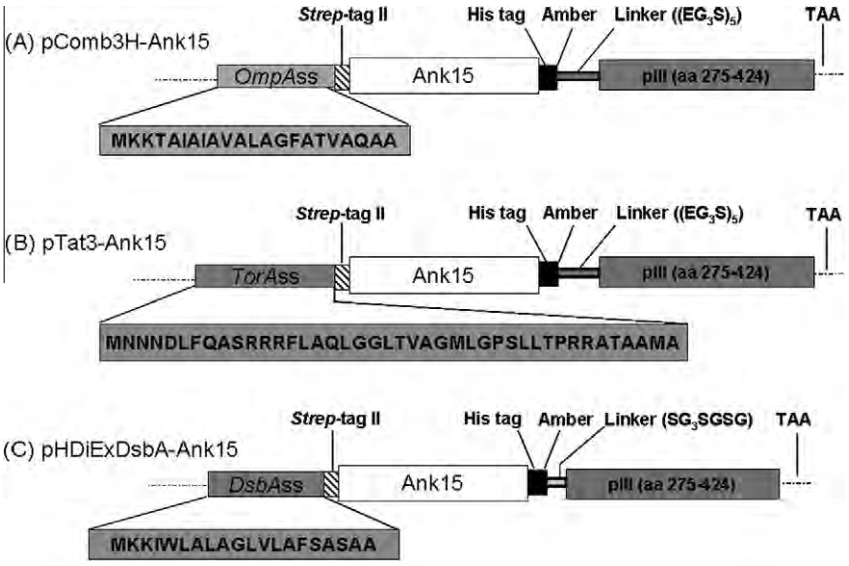

Fig. 2. Construction of phagemid vectors. Three types of phagemid vectors are shown (A) pComb3H-Ank15, (B) pTat3-Ank15 and (C) pHDiExDsbA-Ank15. The following abbreviations are used: *OmpA*ss, OmpA signal sequence; *TorA*ss, TorA signal sequence; and *DsbA*ss, DsbA signal sequence. The amino acid sequences of the signal peptide are shown in single letter code.

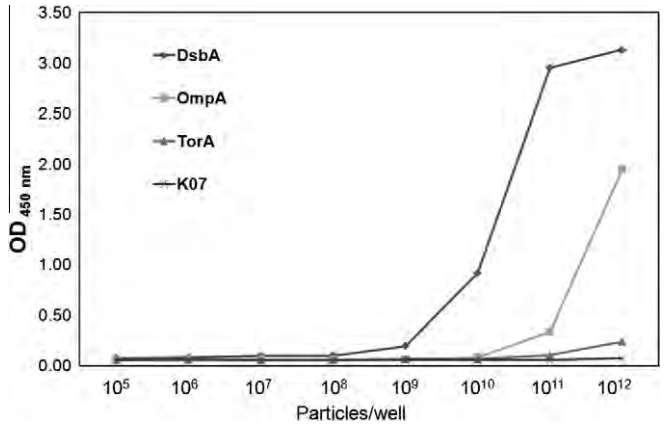

Fig. 3. The display of Ank15 on phage particles. Serial dilutions of each phage were reacted with *Strep*-Tactin coated microtiter plates. The bound phages were detected by adding an anti-M13 antibody conjugated with HRP. The signals of each reaction were measured at 450 nm after adding TMB substrate and stopping the reaction with 1 N HCl.

K07 helper phage. The result demonstrated that Ank15 fusion protein could be expressed and incorporated on phage particles using the *DsbA* and *OmpA* signal sequence.

The presence of fusion protein on phage particles was evaluated by western blot analysis. As shown in Fig. 4, the reactive bands were detected by using mouse anti-pIII antibody and mouse anti-His tag antibody respectively. The pIII wild type migrates in SDS–PAGE at an apparent molecular mass of 65 kDA, corresponding to the result from previous study [19]. The proper molecular size of the fusion protein was observed at 40 kDa for OmpA phage and at 38 kDa for DsbA phage while detecting with two specific antibodies. No reactive bands corresponding to fusion protein appeared from TorA phage and K07 phage, suggesting that the fusion protein could be incorporated on phage particles using *OmpA*ss and *DsbA*ss. However, the intensity of the reactive band from DsbA phage was stronger than that from OmpA phage, suggesting an unequal level of fusion protein incorporated on phage particles. *DsbA*ss showed an enhancement of Ank15 displayed on phage particles, in agreement with result from previous study [10].

The expression and translocation of ankyrin repeat protein

For the detection of protein expression and evaluation of the effect of three chosen signal sequences on directing the ankyrin repeat protein to the periplasmic space, the XL-1 Blue cells harboring suitable vectors were cultured for protein expression without infected with K07 helper phage. The result from the whole cell lysates (Fig. 5A) showed the reactive bands at the proper molecular size (38–40 kDa) indicating that all bacterial cells could express the fusion protein regardless of vector types examined. However, the positive bands of fusion protein in the periplasmic fraction could be observed when directing Ank15 with *OmpA*ss and *DsbA*ss but not with *TorA*ss. This phenomenon suggested that

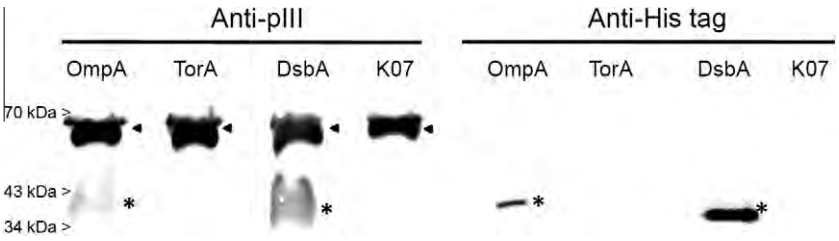

Fig. 4. The comparative displaying efficiency of fusion protein on phage particles. Phage produced from bacteria containing appropriate phagemid: OmpA phage (lanes 1 and 5), TorA phage (lanes 2 and 6), DsbA phage (lanes 3 and 7) and K07 phage (lanes 4 and 8) were separated by SDS–PAGE under reducing conditions and transferred to a PVDF membrane. The reactive bands were analyzed with mouse anti-pIII antibody and mouse anti-His tag antibody. The wild type pIII and Ank15-pIII are shown with closed triangles and asterisks, respectively.

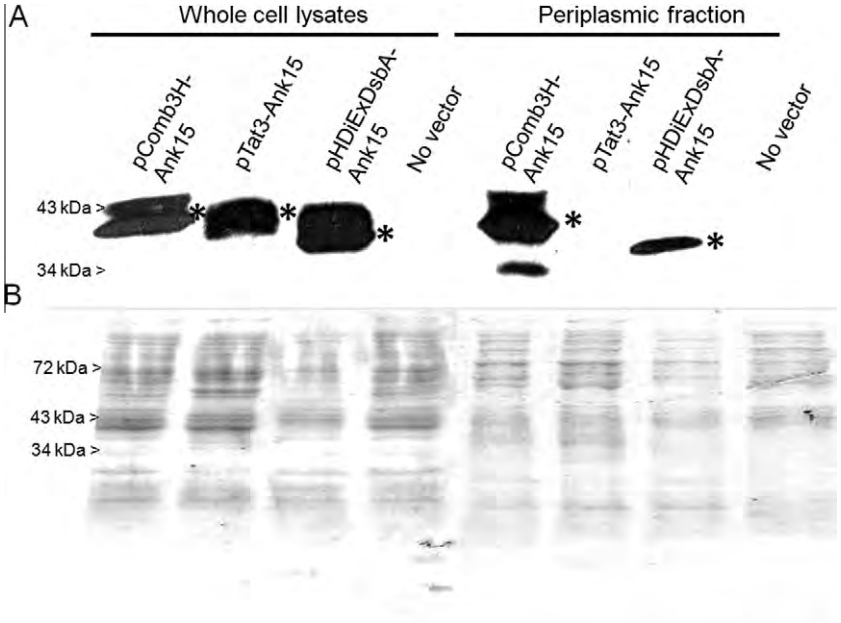

Fig. 5. The expression and translocation of fusion protein in bacterial cells. The whole cell lysates and periplasmic fractions of XL-1 Blue cells harboring individual vector were separated under reducing condition by SDS–PAGE and transferred to PVDF membrane. The reactive band (asterisk) was analyzed with mouse anti-His tag antibody (A), and the loading control was shown by the Coomassie blue staining of SDS–PAGE (B).

the absence of fusion protein on TorA phage was due to the incompatibility of the translocation system. The loading amount of each bacterial protein extract was comparable, demonstrated by Coomassie blue staining on the SDS–PAGE (Fig. 5B).

The effect of Tat machinery on the translocation of ankyrin repeat protein

To determine the influence of pIII domain in Ank15 translocation by Tat machinery, the whole cell lysate and periplasmic fraction of IPTG induced pTat3-Ank15 transformed HB2151 strain were analyzed by immunoblotting. Interestingly, the reactive bands in both fractions could be detected (Fig. 6). This result indicated that the Ank15 protein was able to be expressed and could be directed into the periplasm via *Tor*Ass with the correct conformation.

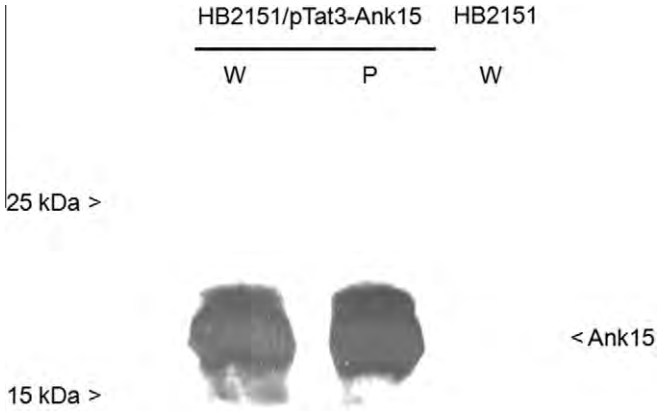

Fig. 6. The effect of pIII on the translocation of Ank15. The whole cell lysate (W) and periplasmic fraction (P) of HB2151 harboring pTat3-Ank15 vector were separated under reducing condition by SDS–PAGE and transferred to PVDF membrane. The reactive band was analyzed using mouse anti-His tag antibody.

Discussion

The phage display technique is an efficient method for displaying a wide range of peptides and proteins on the surface of phage particles. However, not every protein or peptide can be incorporated into the viral particles. The protein must be synthesised and recognized by several translocation systems for directing to the periplasm where the assembly process occurs. The translocation systems that have been applied for phage display are the Sec, Tat, and SRP pathways. Each individual system is suitable for targeting particular protein species. In this study, different leader peptides using the Sec, Tat and SRP pathways were compared in regard to phage display and the directing efficiency of ARP.

Three series of phagemid containing different signal sequences were constructed for displaying the Ank15 on phage particles. The results from ELISA (Fig. 3) and western immunoblotting (Fig. 4) demonstrated that Ank15 protein could be expressed and displayed on phage particles *via DsbA*ss and *OmpA*ss. The fusion protein in the periplasmic fraction could also be observed when directing the protein with *DsbA*ss and *OmpA*ss, but not with *TorA*ss (Fig. 5). This observation correlated with the display level of Ank15 on the phage particles. Naturally, ARP obtains its conformation rapidly in the cytoplasm. The achievement of directing Ank15 protein can be explained by the effective function of cytoplasmic chaperone SecB in slowing down folding and storing this protein in an unfolded state until it is transported through the periplasmic membrane [20,21]. This finding was described previously in displaying DARPins E3_19 [10].

The Ank15 fusion protein failed to be translocated into the periplasmic space and did not display on phage using *TorA*ss even though the Tat machinery could direct free Ank15 through its translocon. This result indicated that pIII domain disrupted the display of ARP on Tat pathway in contrast to Sec and SRP pathways. At present, the explanations why the pIII fusion protein could not be incorporated on the phage particles *via* the Tat pathway remains unclear. Some possible reasons for this phenomenon have been described [9]. In general, the Tat machinery demonstrates the presence of a folding quality-control [9,22]. The inability of *TorA*ss to direct functional phage display may be related to both folding and targeting problems. Therefore, (i) the folding of pIII

protein requires the appropriate environment in the periplasmic space and pIII typically do not fold in the reducing environment of the cytoplasm, resulting in the incorrect folding problem of pIII [19], (ii) while the pIII is kept in an unfolded conformation by cytoplasmic chaperones, it is not suitable to export via this pathway [23], (iii) the appropriate mechanism (the Sec machinery) is required to transport pIII across the inner membrane [24].

To validate whether translocation of Ank15 by Tat machinery was interfered by pIII domain, we applied the advantage of amber stop codon (UAG) to switch from Ank15-pIII to free Ank15 in HB2151 strain (Fig. 2). Remarkably, the cell fractionation of HB2151 strain revealed the successful targeting of the ankyrin protein through the periplasm (Fig. 6). This result indicated that free ARP could be targeted and translocated through the periplasmic space of *E. coli via* the Tat pathway. Therefore, the pIII molecule is involved in the refractory transport of ARP into the periplasm and in the display on the phage particles.

Recently, intracellular selection techniques in *E. coli* have been developed based on the ability of the Tat system to efficiently co-translocate non-covalent complexes of two folded polypeptides. This assay was able to practically detect interactions between intracellular scFv antibodies and their cognate antigen [25]. Furthermore, our findings suggest that although ARP is not effectively transclocated by the Tat pathway when fused to pIII domain, the translocation of folded ankyrin by Tat pathway is possible. Tat dependent translocation of complex between ARP in complex with cognate partners seem to be applicable for this technique in the future. In conclusion, the display of ARP on phage particles could be accomplished when the Sec- and SRP-dependent signal peptides were used, whereas the same direction of the fusion protein using the Tat-dependent pathway was not successful. However, the Tat pathway efficiently directed ARP through its translocon, without fusing with pIII, and appeared as a soluble protein in the *E. coli* periplasm.

Acknowledgments

This work was supported by CHE-PhD-THA scholarship from the Commission on Higher Education, Ministry of Education, Thailand. The authors also express grateful acknowledgement to the Thailand Research Fund, the Research Chair Grant of National Sciences and Technology Development Agency (Thailand) and the Franco-Thai collaboration program. We would like to thank Dr. Dale Taneyhill and Roma Yumul for proofreading the manuscript.

References

[1] G.P. Smith, Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface, Science 228 (1985) 1315–1317.
[2] P. Fekkes, A.J. Driessen, Protein targeting to the bacterial cytoplasmic membrane, Microbiol. Mol. Biol. Rev. 63 (1999) 161–173.
[3] V.P. Hytonen, O.H. Laitinen, T.T. Airenne, H. Kidron, N.J. Meltola, E.J. Porkka, J. Horha, T. Paldanius, J.A. Maatta, H.R. Nordlund, M.S. Johnson, T.A. Salminen, K.J. Airenne, S. Yla-Herttuala, M.S. Kulomaa, Efficient production of active chicken avidin using a bacterial signal peptide in *Escherichia coli*, Biochem. J. 384 (2004) 385–390.
[4] A. Maggioni, M. Vonitzstein, R. Gerardy-Schahn, J. Tiralongo, Targeting the expression of functional murine CMP-sialic acid transporter to the *E. coli* inner membrane, Biochem. Biophys. Res. Commun. 362 (2007) 779–784.
[5] J. Manosroi, C. Tayapiwatana, F. Götz, R.G. Werner, A. Manosroi, Secretion of active recombinant human tissue plasminogen activator derivatives in *Escherichia coli*, Appl. Environ. Microbiol. 67 (2001) 2657–2664.
[6] M.J. Droge, Y.L. Boersma, P.G. Braun, R.J. Buining, M.K. Julsing, K.G. Selles, J.M. van Dijl, W.J. Quax, Phage display of an intracellular carboxylesterase of *Bacillus subtilis*: comparison of Sec and Tat pathway export capabilities, Appl. Environ. Microbiol. 72 (2006) 4589–4595.
[7] J.D. Thomas, R.A. Daniel, J. Errington, C. Robinson, Export of active green fluorescent protein to the periplasm by the twin-arginine translocase (Tat) pathway in *Escherichia coli*, Mol. Microbiol. 39 (2001) 47–53.
[8] P. Thammawong, W. Kasinrerk, R.J. Turner, C. Tayapiwatana, Twin-arginine signal peptide attributes effective display of CD147 to filamentous phage, Appl. Microbiol. Biotechnol. 69 (2006) 697–703.
[9] M.P. DeLisa, D. Tullman, G. Georgiou, Folding quality control in the export of proteins *via* the bacterial twin-arginine translocation pathway, Proc. Natl. Acad. Sci. USA 100 (2003) 6115–6120.
[10] D. Steiner, P. Forrer, M.T. Stumpp, A. Plückthun, Signal sequences directing cotranslational translocation expand the range of proteins amenable to phage display, Nat. Biotechnol. 24 (2006) 823–831.
[11] L.K. Mosavi, T.J. Cammett, D.C. Desrosiers, Z.Y. Peng, The ankyrin repeat as molecular architecture for protein recognition, Protein Sci. 13 (2004) 1435–1448.
[12] H.K. Binz, M.T. Stumpp, P. Forrer, P. Amstutz, A. Plückthun, Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins, J. Mol. Biol. 332 (2003) 489–503.
[13] M. Vogel, E. Keller-Gautschi, M.J. Baumann, P. Amstutz, C. Ruf, F. Kricek, B.M. Stadler, Designed ankyrin repeat proteins as anti-idiotypic-binding molecules, Annu. NY Acad. Sci. 1109 (2007) 9–18.
[14] H.K. Binz, P. Amstutz, A. Kohl, M.T. Stumpp, C. Briand, P. Forrer, M.G. Grutter, A. Plückthun, High-affinity binders selected from designed ankyrin repeat protein libraries, Nat. Biotechnol. 22 (2004) 575–582.
[15] C. Zahnd, E. Wyler, J.M. Schwenk, D. Steiner, M.C. Lawrence, N.M. McKern, F. Pecorari, C.W. Ward, T.O. Joos, A. Plückthun, A designed ankyrin repeat protein evolved to picomolar affinity to Her2, J. Mol. Biol. 369 (2007) 1015–1028.
[16] P. Amstutz, H.K. Binz, P. Parizek, M.T. Stumpp, A. Kohl, M.G. Grutter, P. Forrer, A. Plückthun, Intracellular kinase inhibitors selected from combinatorial libraries of designed ankyrin repeat proteins, J. Biol. Chem. 280 (2005) 24715–24722.
[17] A.C. Fisher, M.P. DeLisa, Efficient isolation of soluble intracellular single-chain antibodies using the twin-arginine translocation machinery, J. Mol. Biol. 385 (2009) 299–311.
[18] V.E.B. Hernandez, L.M.P. Maldonado, E.M. Rivero, A.P. Barba de la Rosa, J.F. Jimenez-Bremont, L.G.O. Acevedo, A.D.L. Rodriguez, Periplasmic expression and recovery of human interferon gamma in *Escherichia coli*, Protein Expr. Purif. 59 (2008) 169–174.
[19] H. Thie, T. Schirrmann, M. Paschke, S. Dubel, M. Hust, SRP and Sec pathway leader peptides for antibody phage display and antibody fragment production in *E. Coli*, Nat. Biotechnol. 25 (2008) 49–54.
[20] N. Wolff, G. Sapriel, C. Bodenreider, A. Chaffotte, P. Delepelaire, Antifolding activity of the SecB chaperone is essential for secretion of HasA: a quickly folding ABC pathway substrate, J. Biol. Chem. 278 (2003) 38247–38253.
[21] A.J.M. Driessen, SecB, a molecular chaperone with two faces, Trends Microbiol. 9 (2001) 193–196.
[22] S. Panahandeh, C. Maurer, M. Moser, M.P. DeLisa, M. Muller, Following the path of a twin-arginine precursor along the Tat ABC translocase of *Escherichia coli*, J. Biol. Chem. 283 (2008) 33267–33275.
[23] M. Paschke, W. Hohne, A twin-arginine translocation (Tat)-mediated phage display system, Gene 350 (2005) 79–88.
[24] D.A. Marvin, Filamentous phage structure Infection and assembly, Curr. Opin. Struct. Biol. 8 (1998) 150–158.
[25] D. Waraho, M.P. DeLisa, Versatile selection technology for intracellular protein-protein interactions mediated by a unique bacterial hitchhiker transport mechanism, Proc. Natl. Acad. Sci. USA. 106 (2009) 3692–3697.

RETROVIROLOGY

RESEARCH Open Access

Antiviral activity of recombinant ankyrin targeted to the capsid domain of HIV-1 Gag polyprotein

Sawitree Nangola[1,2,3], Agathe Urvoas[3], Marie Valerio-Lepiniec[3], Wannisa Khamaikawin[1,2], Supachai Sakkhachornphop[1,2], Saw-See Hong[4,5], Pierre Boulanger[4,5*], Philippe Minard[3*] and Chatchai Tayapiwatana[1,2*]

Abstract

Background: Ankyrins are cellular mediators of a number of essential protein-protein interactions. Unlike intrabodies, ankyrins are composed of highly structured repeat modules characterized by disulfide bridge-independent folding. Artificial ankyrin molecules, designed to target viral components, might act as intracellular antiviral agents and contribute to the cellular immunity against viral pathogens such as HIV-1.

Results: A phage-displayed library of artificial ankyrins was constructed, and screened on a polyprotein made of the fused matrix and capsid domains (MA-CA) of the HIV-1 Gag precursor. An ankyrin with three modules named Ank^{GAG}1D4 (16.5 kDa) was isolated. Ank^{GAG}1D4 and MA-CA formed a protein complex with a stoichiometry of 1:1 and a dissociation constant of K_d ~ 1 μM, and the Ank^{GAG}1D4 binding site was mapped to the N-terminal domain of the CA, within residues 1-110. HIV-1 production in SupT1 cells stably expressing Ank^{GAG}1D4 in both N-myristoylated and non-N-myristoylated versions was significantly reduced compared to control cells. Ank^{GAG}1D4 expression also reduced the production of MLV, a phylogenetically distant retrovirus. The Ank^{GAG}1D4-mediated antiviral effect on HIV-1 was found to occur at post-integration steps, but did not involve the Gag precursor processing or cellular trafficking. Our data suggested that the lower HIV-1 progeny yields resulted from the negative interference of Ank^{GAG}1D4-CA with the Gag assembly and budding pathway.

Conclusions: The resistance of Ank^{GAG}1D4-expressing cells to HIV-1 suggested that the CA-targeted ankyrin Ank^{GAG}1D4 could serve as a protein platform for the design of a novel class of intracellular inhibitors of HIV-1 assembly based on ankyrin-repeat modules.

Keywords: HIV-1, HIV-1 assembly, Gag polyprotein, CA domain, virus assembly inhibitor, ankyrins, artificial ankyrin library, intracellular antiviral agent

Background

In recent years, significant progress has been made in the control of HIV-1 infections using highly active antiretroviral therapy (HAART). Nevertheless, the occurrence of multi-drug resistant mutants and the side effects of HAART justify the exploration of alternative therapeutic approaches, such as gene therapy [1-5]. Several strategies for anti-HIV gene therapy are currently under development, and certain ones have been tested in hematopoietic cells [6-8]. They can be classified into two major categories: (i) RNA-based agents including antisense, ribozymes, aptamers and RNA interference [9]; (ii) protein-based agents including dominant-negative mutant proteins, intrabodies, intrakines, fusion inhibitors and zinc-finger nucleases [10,11]. The most commonly transduced genes with antiviral potential consist of those encoding derivatives of immunoglobulins. However, the complex structure of these molecules limits their antiviral function within cells, since their stability relies on disulfide bond(s) which

* Correspondence: pierre.boulanger@univ-lyon1.fr; philippe.minard@u-psud.fr; asimi002@hotmail.com
[1]Division of Clinical Immunology, Department of Medical Technology, Faculty of Associated Medical Sciences, Chiang Mai University, Chiang Mai, Thailand 50200
[3]Institut de Biochimie et de Biophysique Moléculaire et Cellulaire (IBBMC) UMR-8619, Université de Paris-Sud et CNRS, Orsay Cedex 91405, France
Full list of author information is available at the end of the article

rarely occur(s) in the reducing conditions of the intracellular milieu [12-16].

Several methods and novel molecules have been developed to overcome the limitations of antibodies and their derivatives (e.g. scFv), in terms of stability, facility of modifications, robustness, and cost-efficient production [13,17-19]. This is the case for molecules based on protein frameworks or scaffolds which interact with potential therapeutic targets by mimicking the binding process of immunoglobulins to their specific antigens. The ankyrin-repeat proteins represent an attractive scaffold to generate this type of specific binders [20,21]. Analysis of the protein sequence-structure relationship in natural ankyrins has defined consensus ankyrin motifs (or modules), and the results have been used to generate large libraries of artificial proteins, called 'Designed Ankyrin-Repeat Proteins' or DARPins. Several DARPins with desired binding specificity to various target molecules have successfully been isolated from such libraries [12,21-27], including competitors of HIV-1 binding to the viral receptor CD4 [28].

Ankyrins mediate many important protein-protein interactions in virtually all species and are found in all cellular compartments, indicating that these proteins can be adapted to function in a variety of environments, intracellular as well as extracellular [12,20,21,23,25,29,30]. For example, lentiviral vectors pseudotyped with HER2/neu-specific DARPins have been found to efficiently transduce their specific targets, HER2/neu-positive cells [31]. The major advantages of ankyrin-repeat proteins reside in (i) their binding specificity and affinity, as observed in DARPins selected from large libraries; (ii) their solubility and stability, even in the reducing conditions of the intracellular milieu; (iii) their sequence features present in DARPins, which are naturally expressed in human cells: as a consequence, ankyrin-repeat proteins are expected not to be as immunogenic as foreign proteins. Artificial ankyrins are therefore promising candidates as protein interfering reagents capable of acting both extra- and intra- cellularly [24].

In the present study, we designed artificial ankyrin molecules targeted to the HIV-1 Gag polyprotein and evaluated their potential as intracellular therapeutic agents which would negatively interfere with HIV-1 replication, and more specifically with the virus particle assembly machinery. To this aim, we constructed a library of ankyrin-repeat protein library expressed at the surface of recombinant filamentous bacteriophages. This phage-displayed library was screened on immobilized matrix (MA)-capsid (CA) domain (MA-CA) of the HIV-1 Gag precursor (Pr55Gag, or more simply Gag), and a panel of Gag-specific artificial ankyrins were isolated. One particular Gag binder, $Ank^{GAG}1D4$, was selected for further characterization. $Ank^{GAG}1D4$ binding site was mapped to the N-terminal domain of the CA, and SupT1 cells that stably expressed $Ank^{GAG}1D4$ showed a reduced permissiveness to HIV-1 infection. The $Ank^{GAG}1D4$-mediated antiviral effect was found to occur at post-integration steps of the HIV-1 life cycle involving the Gag protein assembly and budding machinery. This study demonstrated the potential of ankyrin-repeat proteins as a novel class of intracellular antivirals and suggested that $Ank^{GAG}1D4$ could serve as a protein platform for the design of efficient intracellular inhibitors of HIV-1 assembly.

Results

Design of artificial ankyrin-repeat motifs and construction of an ankyrin library

The construction of an artificial ankyrin library implies the randomization of amino acid residues localized on the accessible surface of ankyrin which has a potential interaction with the desired target, while maintaining intact the conserved residues of the consensus repeat modules which form the rigid framework of ankyrin. The consensus sequence of the ankyrin-repeat modules generated in this study was based on previous DARPins libraries [12-14,23,29,30,32], with minor modifications, as described in the Materials and Methods section. For example, the lysine residue (K) at position 25 was substituted for glutamate (E) to prevent a possible repulsion with the positively charged arginine (R) at position 21 (Tables 1 and Figure 1). Our ankyrin library was generated by randomization of amino acids at positions 2, 3, 5, 10, 13, and 14 (Figure 2 and 3). The amino acid side chains at these positions were all oriented outwards and belonged to the same surface-exposed surface of the ankyrin-repeat module (Figure 3B).

The artificial ankyrin-repeat proteins obtained were made of a variable number of ankyrin modules flanked by N- and C-capping sequences (N-cap and C-cap; Figure 2B). The library was generated using the directional polymerization of one ankyrin microgene, each microgene corresponding to one single repeat motif. Polymerization was realized directly into a phagemid vector [33]. This resulted in proteins with variable numbers of repeats and sequence lengths. The length distribution in the library was determined by digesting the phagemid pool with restriction enzymes whose sites were located on both sides of the ankyrin coding sequence, followed by analysis of the DNA fragments by gel electrophoresis. A maximum number of 15 ankyrin repeats was obtained, and the most frequent numbers ranged from 1 to 6.

Our final phage-displayed ankyrin library represented as many as 1.9×10^8 independent clones. The quality of our library was first evaluated by sequencing proteins from randomly picked clones. Nine out of fifteen clones (60%) presented discontinuous sequences, with sequence

Table 1 Randomization schemes used to introduce variability at specific positions of ankyrin repeats [(a)]

Repeat position	Degenerated codons	Subset of encoded amino acid
2	VDK, DMY, RAA	A, D, E, G, H, I, K, L, M, N, Q, R, S, T, V, Y
3	VDK, DMY, VAN	A, D, E, G, H, I, K, L, M, N, Q, R, S, T, V, Y
5	VDK, DMY, VAN, TGG	A, D, E, G, H, I, K, L, M, N, Q, R, S, T, V, W, Y
10	CTG, TGG, TAC, RTC	I, L, V, W, Y
13	KCK, TAC, CGY, VAR	A, E, F, I, K, L, M, Q, R, S, Y
14	KCK, VAR, AAC, SGY, YAY, NTG	A, E, G, H, K, L, M, N, Q, R, S, V, Y

[(a)] For each position indicated on the leftmost column, the degenerated codons used for the microgene synthesis and the corresponding encoded amino-acids are indicated in the middle and rightmost columns, respectively. Standard letter for mixed bases refer to: N = A/T/G/C, V = A/C/G, D = A/G/T, K = G/T, M = A/C, Y = C/T, R = A/G.

frameshifts and stop codons, which likely resulted from errors in oligonucleotide synthesis and/or assembly. Discontinuous sequences occurred with a higher frequency in ankyrins with numerous modules, while most ankyrin molecules with fewer repeats showed correct, open reading frames. The proportion of clones in our library with readthrough ankyrin sequences was also evaluated from the proportion of colonies which expressed C-terminally His-tagged soluble ankyrin protein: 34% (24 out of 72 clones) were found to be positive for the C-terminal His-tag, as monitored by colony filtration blotting (COFI blot; data not shown). We therefore estimated the real diversity of our library to one third of the total number of independent clones, i.e. 6×10^7 independent ankyrin coding sequences.

Production and purification of the viral protein target: HIV-1 Gag precursor H_6MA-CA

The viral target used for screening our phage-displayed ankyrins consisted of the His-tagged recombinant

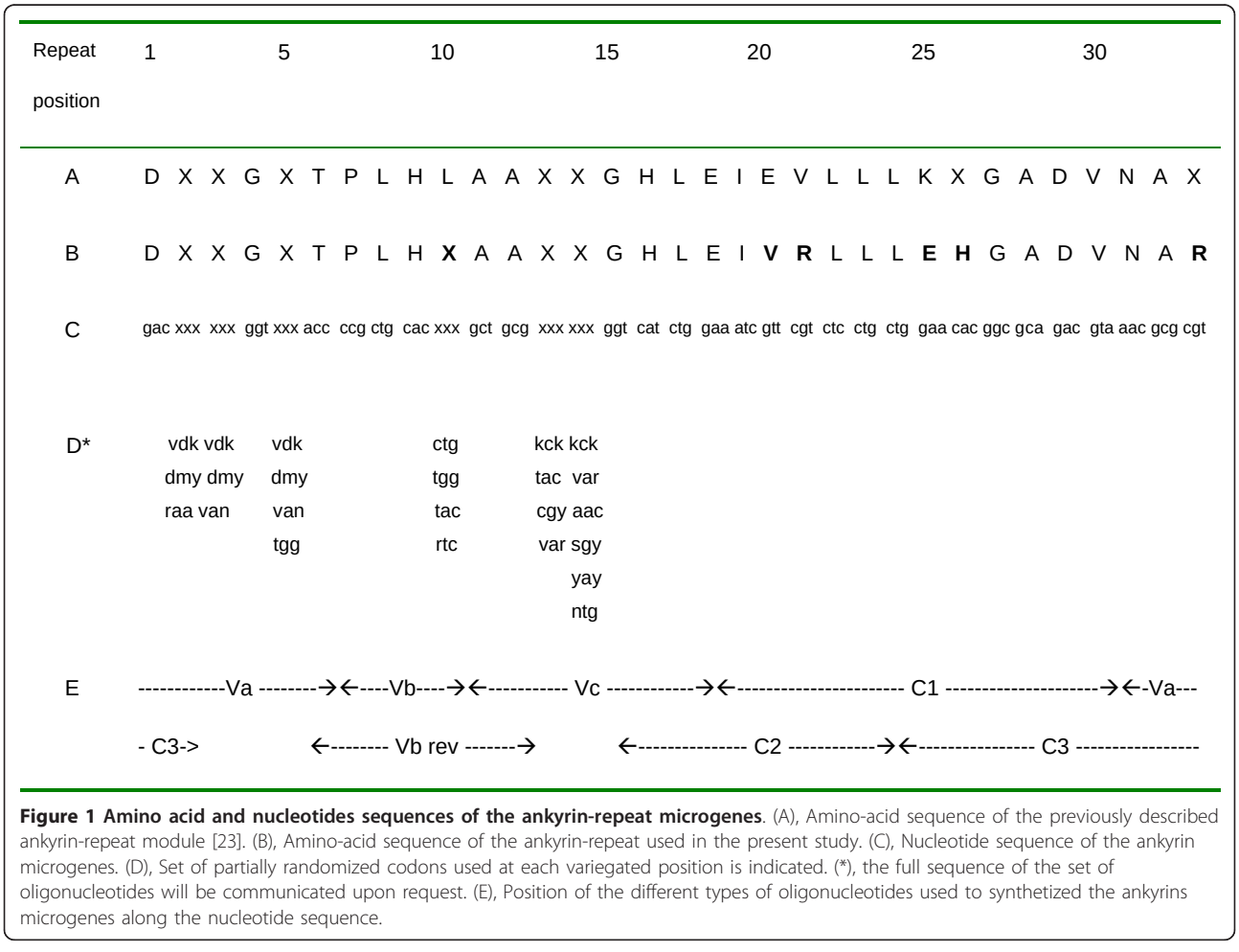

Figure 1 Amino acid and nucleotides sequences of the ankyrin-repeat microgenes. (A), Amino-acid sequence of the previously described ankyrin-repeat module [23]. (B), Amino-acid sequence of the ankyrin-repeat used in the present study. (C), Nucleotide sequence of the ankyrin microgenes. (D), Set of partially randomized codons used at each variegated position is indicated. (*), the full sequence of the set of oligonucleotides will be communicated upon request. (E), Position of the different types of oligonucleotides used to synthetized the ankyrins microgenes along the nucleotide sequence.

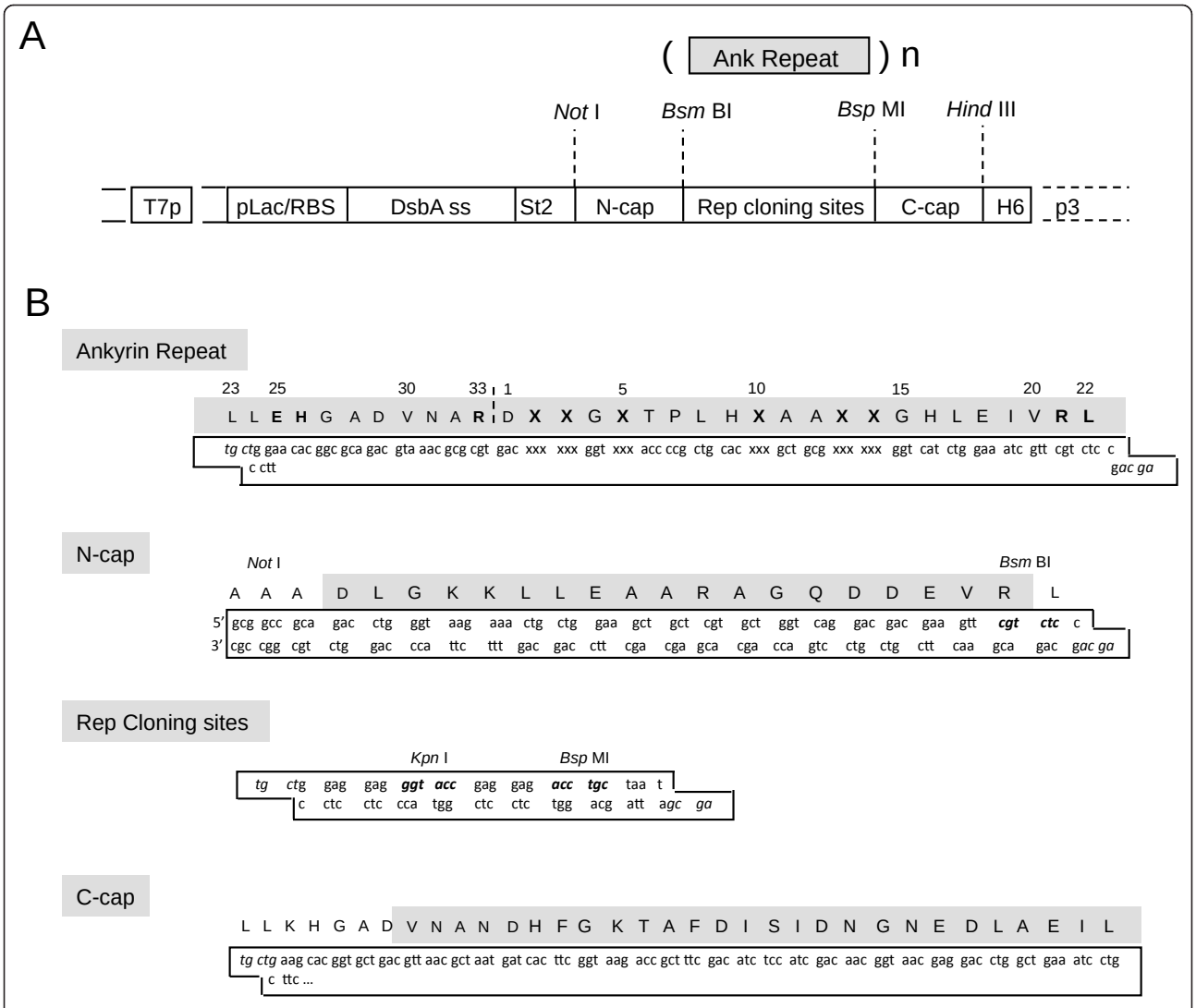

Figure 2 Schematic construction of the ankyrin repeat library using the pHDiEx acceptor vector. **(A)**, The mono-ankyrin microgenes were polymerized by insertion/ligation to pHDiEx double digested by *Bsm* BI and *Kpn* I. The construction was subsequently digested by *Bsp* MI and recircularised by intramolecular ligation. This resulted in the substitution of the region containing the Rep cloning sites by the ankyrin repeats. The number of repeats differed from one clone to another, and usually ranged from 2 to 6 repeats. **(B)**, Detailed sequences of the different DNA regions. Abbreviations: T7p, T7 promoter; pLac/RBS, lactose promoter and Ribosome binding site; DsbA ss, DsbA signal sequence; St2, Strep-Tag2 tag; H6, hexa-histidine tag.

polyprotein H_6MA-CA, corresponding to the MAp17 and CAp24 domains of the HIV-1 Gag precursor. The rationale for screening our ankyrin library on the MA-CA target was not only to search for MA- or/and CA-binders, but also for ankyrin(s) which recognize(s) the MAp17-CAp24 hinge region, which contains the cleavage site of the viral protease (PR). His-tagged recombinant protein H_6CA, which corresponded to the mature capsid protein CAp24, was used to identify the structural domains of the Gag precursor which contained the ankyrin-binding site. Large amounts of recombinant H_6MA-CA and H_6CA proteins were produced in Sf9 cells infected with recombinant baculoviruses BV-H_6MA-CA or BV-H_6CA, and the recombinant Gag proteins purified by affinity chromatography on nickel-sepharose column.

Screening of Gag-binding ankyrins using the phage-display method

The phage-displayed library of ankyrins was amplified using a conventional protocol [34,35], and Gag-binders were isolated by three rounds of selection/elution from surface-immobilized H_6MA-CA protein. Elution of H_6MA-CA-bound phages was performed using acidic buffer for the first two rounds, followed by specific ligand elution using excess of soluble H_6MA-CA protein as the

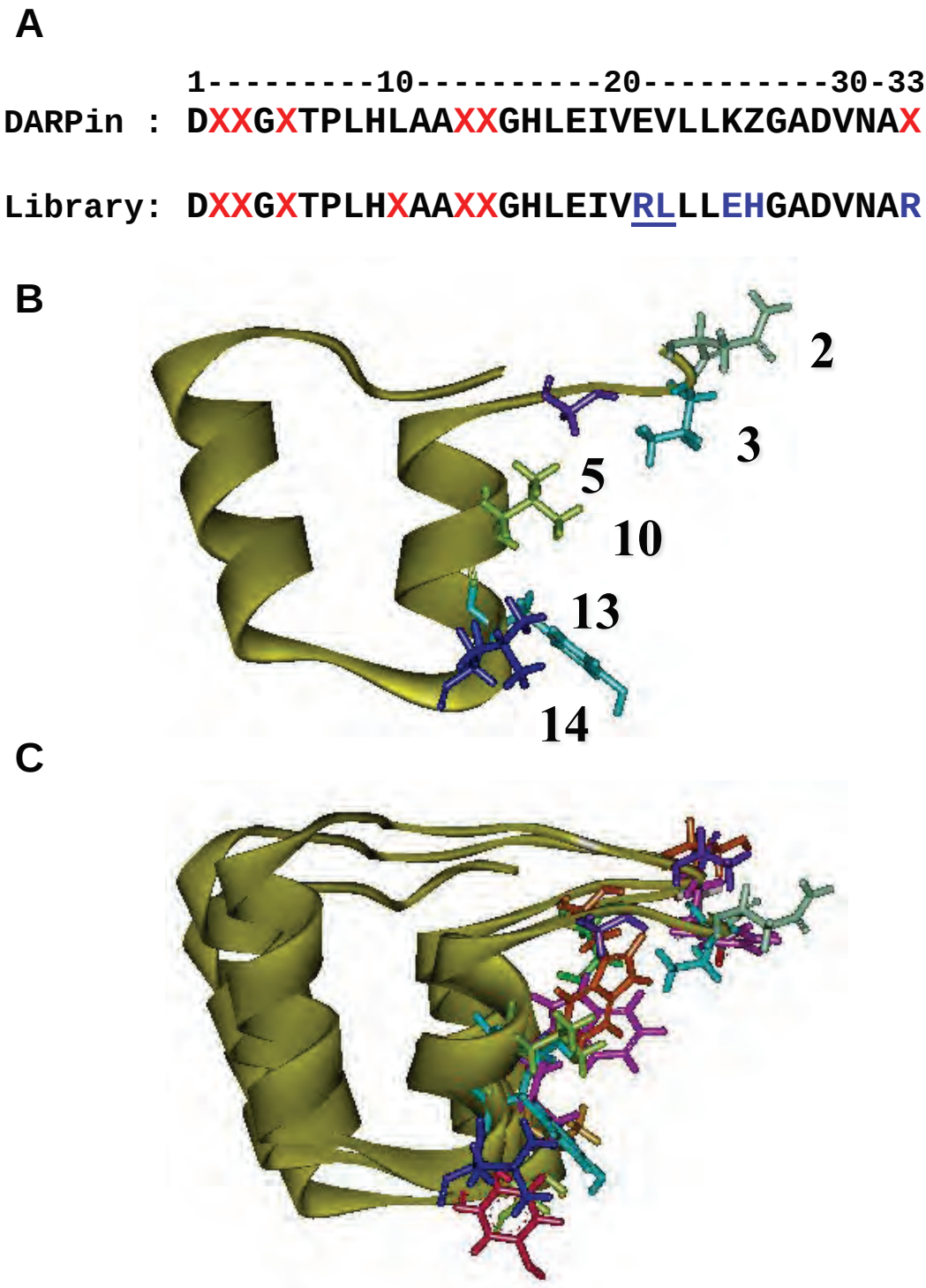

Figure 3 Consensus sequence and three-dimensional model of ankyrin module. **(A)**, Sequence comparison between the consensus DARPin repeat motif and the repeat motif of our ankyrin library. The red letters refer to the positions of random amino acids, and blue letters represent the residues which differ from the consensus DARPin sequence. The position of the recognition site for the restriction nuclease *Bsm* BI is underlined. **(B)**, Structural model of one single ankyrin repeat motif (or module). (**C**), Spatial arrangement of three modules belonging to the same ankyrin linear sequence (triple-repeat motif ankyrin molecule). The fixed structure of the repeat motif is presented as a yellow ribbon. The variable amino acids on the solvent-exposed surface are shown as stick pattern; their respective number in the linear sequence is indicated in panel B.

competititor in the third round [34,35]. Phage clones were picked at random in each eluate, and tested by ELISA for binding to H_6MA-CA. Only 20% of Gag-binders were found in the first eluate, whereas a significant enrichment was observed in the second and third eluates, with 70% of Gag-binders in both. Clones which gave a signal 5-fold over the background signal were picked in all eluates and sequenced. All positive clones showed two or three ankyrin repeats flanked by N-cap and C-cap. Three different clones, referred to as Ank^{GAG}1B8, Ank^{GAG}1D4 and Ank^{GAG}6B4 and containing three ankyrin modules each, were identified several times; they were therefore selected for further studies.

To evaluate the specificity of our Gag-binders, an irrelevant target protein, αRep-A3, was used in lieu of H_6MA-CA. Protein αRep-A3, previously described under the acronym αRep-n4-a in our previous study [33], is an artificial alpha-helicoidal repeat protein (αRep) based on thermostable HEAT-like repeats, which folds cooperatively and shows a high stability [33]. Our phage-displayed ankyrin library was screened on immobilized αRep-A3 protein, and αRep-A3-bound clones were checked for binding specificity and sequenced. One ankyrin clone with a high affinity and specificity for the αRep-A3 target, referred to as Ank^{A3}2D3, was used as the irrelevant control of H_6MA-CA binders in the rest of the present study.

Gag-ankyrin interaction

Gag- and αRep-A3-binding ankyrins were purified, chemically biotinylated, and assayed for their capacity of binding to their specific target *in vitro*. Importantly, no change was detected in the interaction of the three Gag-binders Ank^{GAG}1B8, Ank^{GAG}1D4 and Ank^{GAG}6B4, and of control αRep-A3-binder Ank^{A3}2D3, with their respective substrates, as determined by ELISA (data not shown). This indicated that biotinylation did not alter their Gag- or αRep-A3-specific binding activity.

The degree of Gag-specificity of biotinylated Ank^{GAG}1B8, Ank^{GAG}1D4 and Ank^{GAG}6B4 was evaluated in the presence of specific or nonspecific competitors, and tested in ELISA using H_6MA-CA-coated wells. Controls consisted of Ank^{A3}2D3 and αRep-A3-coated wells. Competitors were (i) the same ankyrin protein in its non-biotinylated form and (ii) non-biotinylated αRep-A3 protein. Ank^{GAG}1B8, Ank^{GAG}1D4, Ank^{GAG}6B4 and Ank^{A3}2D3 were all competed with their respective non-biotinylated versions, while no significant competition was observed between ankyrins Ank^{GAG}1B8, Ank^{GAG}1D4, Ank^{GAG}6B4 on one hand, and αRep-A3 protein on the other hand (Figure 4A). Interestingly, Ank^{GAG}1D4 showed the highest signal of binding to the H_6MA-CA target, and the highest competition effect was observed with non-biotinylated Ank^{GAG}1D4 (Figure 4A).

Identification of the ankyrin binding domain on HIV-1 Gag precursor

The structural domain of Pr55Gag recognized by each of the three Gag-binders Ank^{GAG}1B8, Ank^{GAG}1D4 and Ank^{GAG}6B4 was determined by Far Western blot analysis and ELISA. Lysates of H_6MA-CA-expressing Sf9 were analyzed by SDS-PAGE, and proteins transferred to PVDF membranes. Spontaneous cleavage of H_6MA-CA by insect cell proteases resulted in the occurrence of His-tagged N-terminal domain, H_6MA, migrating as the mature matrix protein of the virion, MAp17 (Figure 4B; control, rightmost lane). All three Gag-binders, Ank^{GAG}1B8, Ank^{GAG}1D4 and Ank^{GAG}6B4, reacted with H_6MA-CA on blot, but not with H_6MA (Figure 4B). This indicated that the ankyrin binding site was not located in the MA domain, but in the CA domain. As expected, no reaction was obtained with the control αRep-A3-binder Ank^{A3}2D3 (Figure 4B). The reactivity towards the CA domain was confirmed by indirect ELISA, using recombinant H_6CA protein immobilized on nickel-coated wells. Positive signals with the CA protein were detected with all three Gag-binders, but not with Ank^{A3}2D3 (Figure 4C). This indicated that the binding sites of Ank^{GAG}1B8, Ank^{GAG}1D4 and Ank^{GAG}6B4 on H_6MA-CA protein were all situated in the CA domain.

Biochemical characterization of Gag-binding ankyrins

As shown in Figure 4C, Ank^{GAG}1B8 reacted with H_6CA with the highest apparent affinity. However, DNA sequencing showed several nonconservative amino acid substitutions within the highly structured scaffold domain of the Ank^{GAG}1B8 modules, as well as in Ank^{GAG}6B4. Since these mutations could adversely affect the ankyrin-repeat motifs, Ank^{GAG}1B8 and Ank^{GAG}6B4 were excluded from our next analyses, and only Ank^{GAG}1D4 was selected for further characterization. DNA sequencing revealed that Ank^{GAG}1D4 protein comprised of three ankyrin modules, each containing different types of amino acids at the six assigned positions for variable residues (Figure 5A). The oligo-histidine tag allowed us to purify Ank^{GAG}1D4 protein to homogeneity by using a two-step chromatographic procedure, (i) affinity chromatography (Figure 5B, lane 3), and (ii) gel filtration (Figure 5B, lane 2). SDS-PAGE analysis showed that Ank^{GAG}1D4 migrated with an apparent molecular mass of 16.5 kDa (Figure 5B), consistent with the theoretical mass 17.9 kDa for a protein of 163 amino acid residues.

Mapping of the Ank^{GAG}1D4 binding site on the CA domain

A more refined mapping of the ankyrin binding site on the CA domain was performed using carboxyterminal deletion mutants of Gag expressed as recombinant proteins in baculovirus-infected cells. Gag*amb*276 and

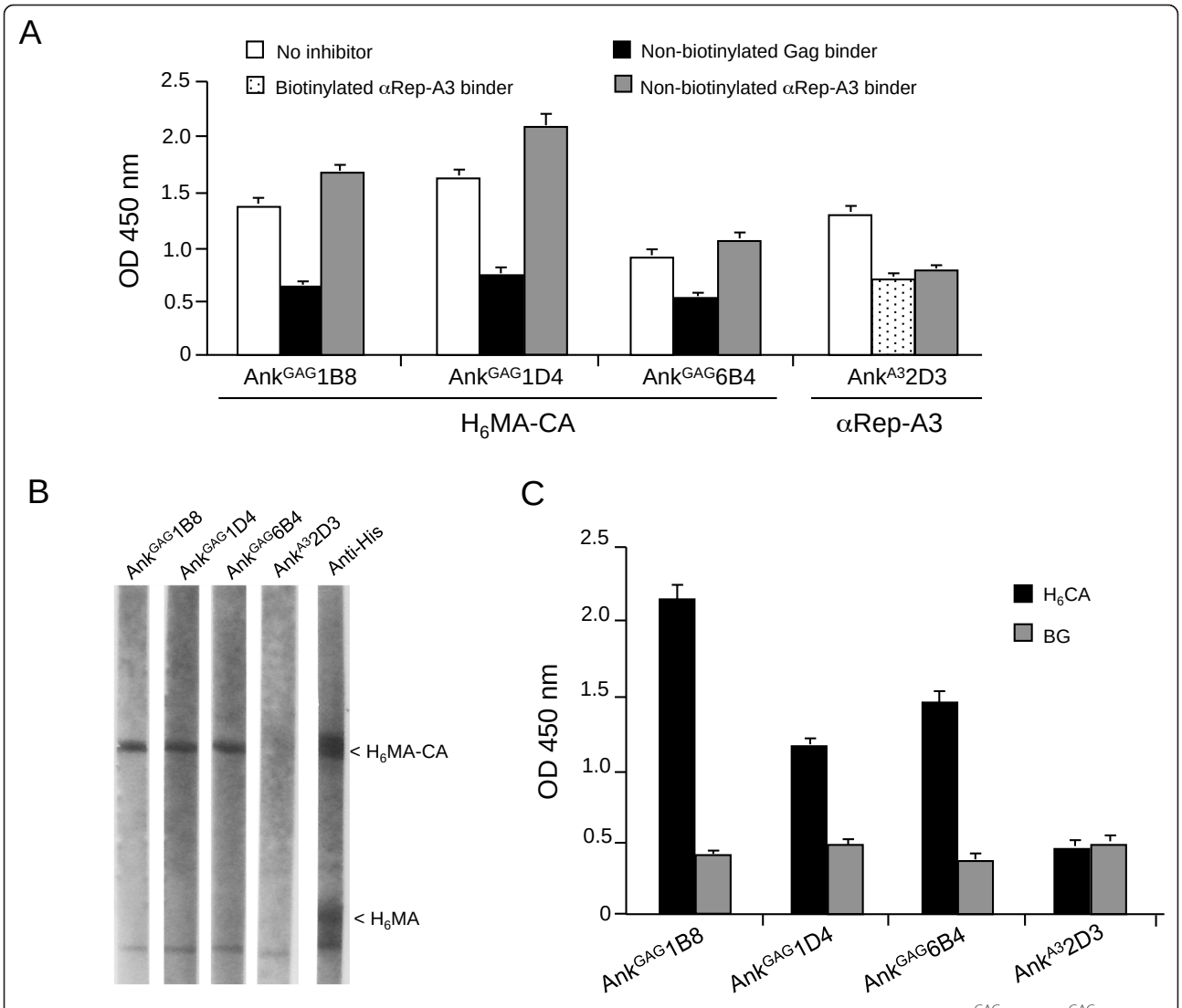

Figure 4 Gag-binding activity of artificial ankyrins. **(A)**, ***Competition ELISA***. Samples of biotinylated Gag-binders Ank^{GAG}1B8, Ank^{GAG}1D4 and Ank^{GAG}6B4, and of control biotinylated αRep-A3-binder Ank^{A3}2D3 were mixed with their corresponding non-biotinylated form (black bars), or mixed with irrelevant soluble target (grey bars), or mixed with buffer containing no inhibitor (white bars). Mixtures were added to H_6MA-CA- or αRep-A3-coated wells, as indicated at the bottom of the panel. Bound-ankyrins were detected by addition of HRP-conjugated extravidin, followed by the TMB substrate. **(B)**, ***Far Western blotting***. Lysates of BV-H_6MA-CA-infected Sf9 cells were electrophoresed in SDS-gel, proteins transferred to a PVDF membrane, and membrane cut into strips. Gag-binding activity was determined by incubation of the strips with the different biotinylated ankyrins Ank^{GAG}1B8, Ank^{GAG}1D4, Ank^{GAG}6B4, and Ank^{A3}2D3, as indicated on top of the strips. On the rightmost strip, the respective positions of the Gag proteins H_6MA-CA and H_6MA were determined using anti-histidine tag antibody (arrowheads). **(C)**, ***Indirect ELISA***. H_6CA was captured on nickel-coated plate, and used as substrate for binding assay of biotinylated ankyrins Ank^{GAG}1B8, Ank^{GAG}1D4, Ank^{GAG}6B4, and Ank^{A3}2D3. Bound-ankyrins were quantitated as in **(A)**. BG, background signal.

Gag*amb*241 mutants carried an amber stop codon in the Pr55Gag sequence at positions 276 and 241, respectively [36]. Both recombinant Gag proteins had in common the MA domain, plus 110 residues of the CA domain for Gag*amb*241, and 145 residues of the CA domain for Gag*amb*276 [36]. Ank^{GAG}1D4 was found to bind to both C-truncated Gag proteins (Figure 6). This restricted the Ank^{GAG}1D4 binding site to the N-terminal region of the CA domain spanning residues 1 to 110, corresponding to positions 132-241 in the Pr55Gag sequence of 500 amino acids.

Gag-binding parameters of Ank^{GAG}1D4

The specificity and binding parameters of Ank^{GAG}1D4 to its H_6MA-CA substrate were determined by microcalorimetry (ITC). Titration of increasing amounts of Ank^{GAG}1D4 protein into sample cell containing purified H_6MA-CA protein gave the approximate value of 1 μM

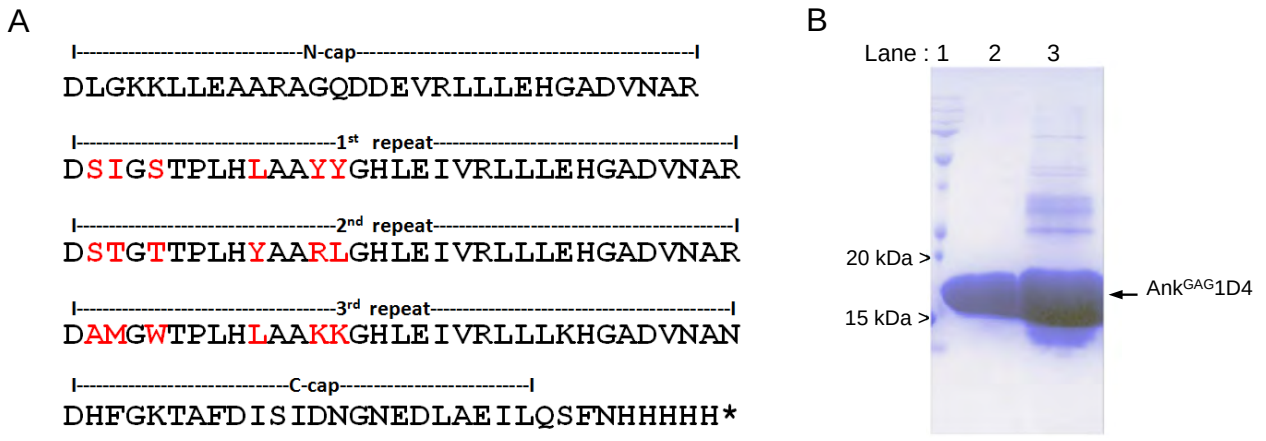

Figure 5 Characterization of Ank^{GAG}1D4. **(A)**, Amino acid sequence of the Ank^{GAG}1D4 protein, represented by using the single letter code. The variable residues at the six predefined positions on the ankyrin solvent-exposed surface were highlighted in red. **(B)**, SDS-PAGE analysis of Ank^{GAG}1D4. The Ank^{GAG}1D4 protein was first isolated by affinity chromatography on nickel-column (HisTrap column; lane 3), and further purified by gel filtration chromatography (Sephadex G-75 column; lane 2). Lane 1, molecular mass markers.

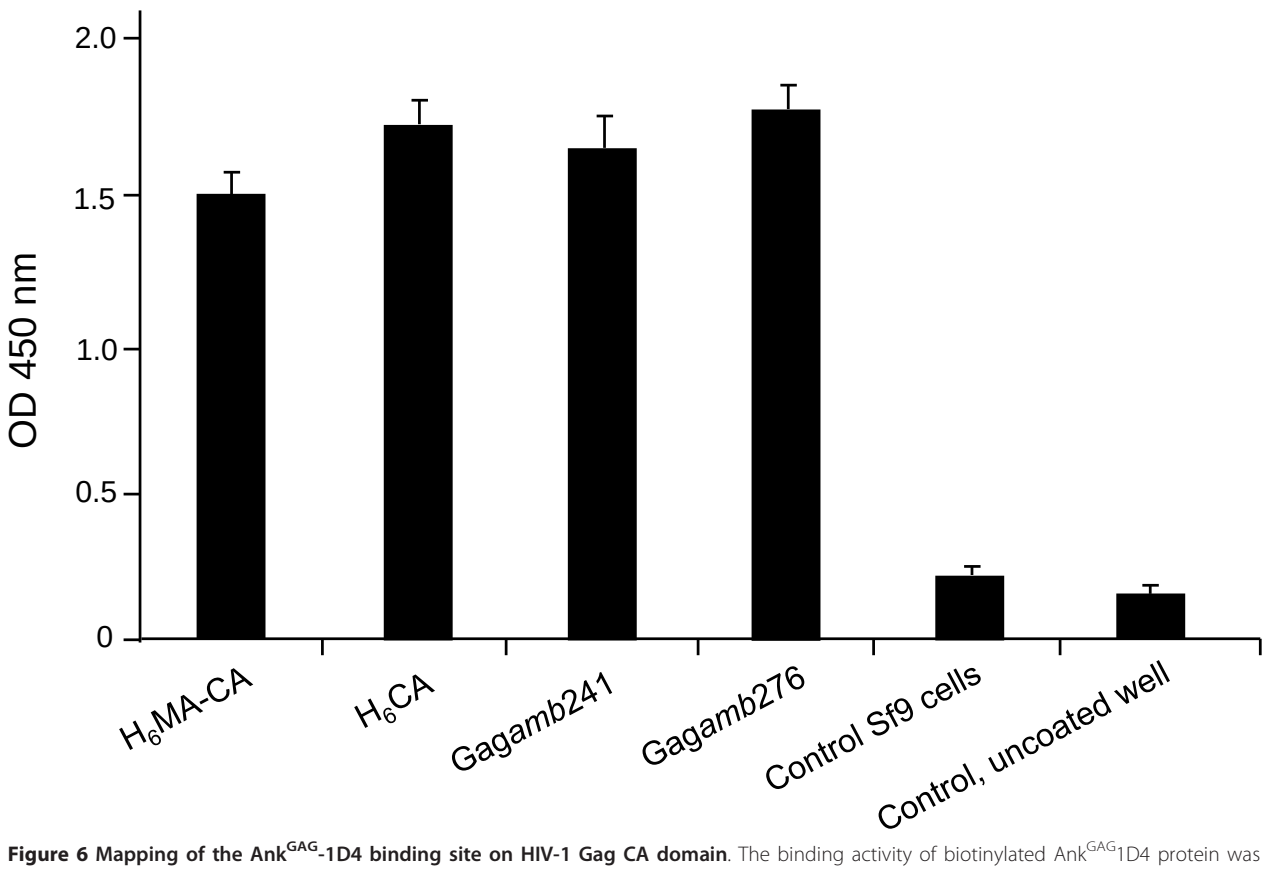

Figure 6 Mapping of the Ank^{GAG}-1D4 binding site on HIV-1 Gag CA domain. The binding activity of biotinylated Ank^{GAG}1D4 protein was tested in ELISA towards surface-immobilized lysates of mock-infected Sf9 cells, or baculovirus-infected Sf9 cells expressing recombinant H_6MA-CA, H_6CA, Gag*amb*241, or Gag*amb*276 protein. The quantity of bound Ank^{GAG}1D4 was assayed by reaction with streptavidin-HRP. Data presented are from triplicate experiments (m ± SEM).

for the dissociation constant (K_d) of the specific reaction of the binder with its target protein (Figure 7; leftmost top panel). In control experiments, no interaction was detected between Ank^{GAG}1D4 and αRep-A3 (Figure 7; middle top panel). By comparison, Ank^{A3}2D3 interacted with its substrate αRep-A3 with a K_d = 18 nM (Figure 7; rightmost top panel).

The stoichiometry (N) of the interacting molecules in the protein complexes and the number of binding sites were calculated from the fitting curves of ITC data. The stoichiometry of protein monomers was found to be N = 0.91 for the pair Ank^{GAG}1D4/H_6MA-CA, and N = 0.62 for the control pair Ank^{A3}2D3/αRep-A3 (Figure 7). The αRep-A3 protein is known to occur as a homodimer [33], and the experimental value of 0.62 was close to the theoretical ratio of 0.5. The data therefore suggested that one molecule of Ank^{A3}2D3 bound to an αRep-A3 homodimer to form a ternary Ank^{A3}2D3/(αRep-A3)$_2$ complex. By contrast, Ank^{GAG}1D4 bound to the H_6MA-CA monomer in a 1-to-1 binary complex.

Construction of SupT1 cell lines stably expressing Gag-binding ankyrin proteins

Two pCEP4-based episomal plasmids encoding the Ank^{GAG}1D4 protein in its Myr+ and Myr0 versions, respectively, and fused to His-tagged GFP at the N-terminus (Figure 8A) were transfected into the SupT1 cell line. Clones that stably expressed Ank^{GAG}1D4-GFP protein (SupT1/Myr+Ank^{GAG}1D4-GFP and SupT1/Myr0Ank^{GAG}1D4-GFP) were identified by fluorescent microscopy, isolated and expanded under the hygromycin-B selection. Two control SupT1 cell lines harboring the pCEP4 plasmids encoding the Myr+ and Myr0 versions of Gag-irrelevant Ank^{A3}2D3-GFP (SupT1/Myr+Ank^{A3}2D3-GFP and SupT1/Myr0Ank^{A3}2D3-GFP) were generated in parallel. Confocal microscopy showed that Myr+Ank^{GAG}1D4-GFP and Myr+Ank^{A3}2D3-GFP localized in both the cytoplasm and the plasma membrane, as expected for N-myristoylated proteins, whereas Myr0Ank^{GAG}1D4-GFP and Myr0Ank^{A3}2D3-GFP showed a diffuse cytoplasmic fluorescence. Flow cytometry analysis showed that almost 80% of ankyrin-expressing cells were GFP-positive, and that Myr+Ank^{GAG}1D4-GFP or Myr0Ank^{GAG}1D4-GFP did not negatively interfere with the surface expression of CD4 (Figure 8B). The status of CD4 molecules, the primary receptors of HIV-1, was important to assess in ankyrin-expressing cells prior to HIV-1 infection, in order to ensure that SupT1/Myr+Ank^{GAG}1D4-GFP and SupT1/Myr0Ank^{GAG}1D4-GFP cells could serve as host cells for testing the functionality of Ank^{GAG}1D4 as antiviral agent.

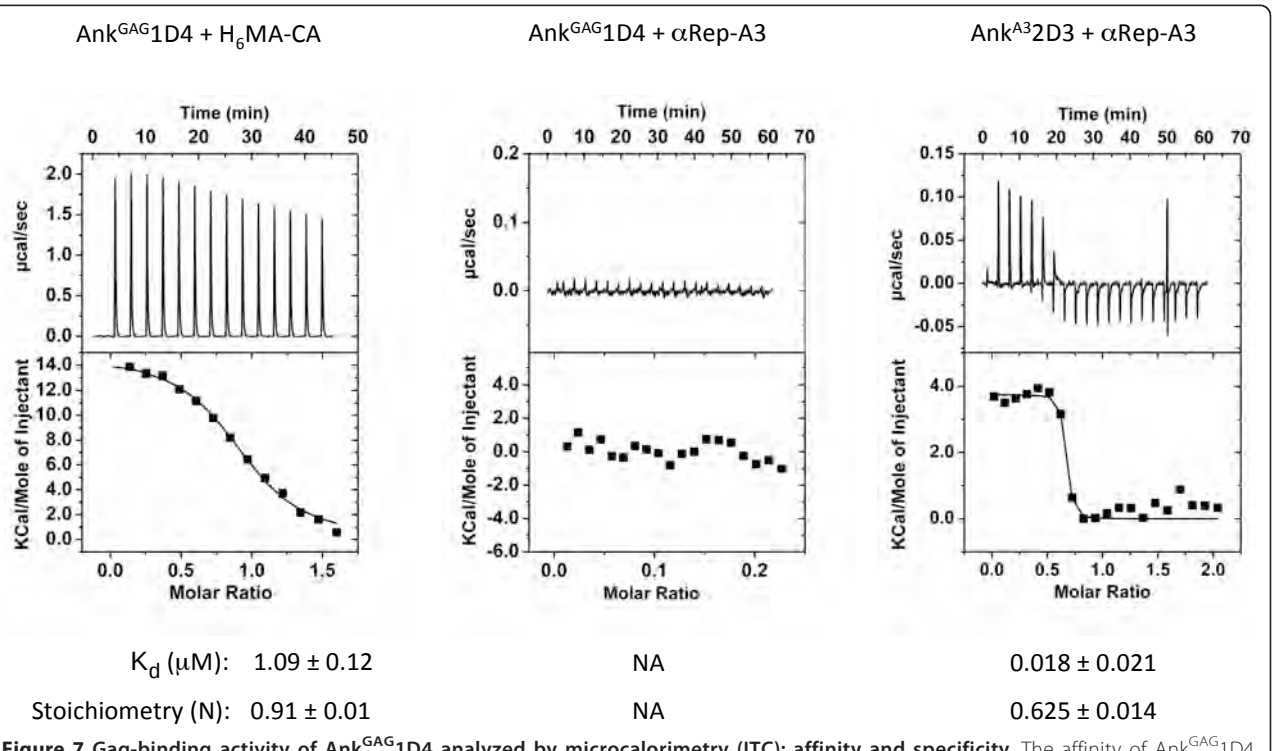

Figure 7 Gag-binding activity of Ank^{GAG}1D4 analyzed by microcalorimetry (ITC): affinity and specificity. The affinity of Ank^{GAG}1D4 towards its Gag substrate H6MA-CA (leftmost panel) and the affinity of Ank^{A3}2D3 towards its αRep-A3 substrate (rightmost panel) were determined by ITC measurements of ankyrin/substrate mixtures at various protein ratios. The Gag-binding specificity was evaluated by ITC analysis of the mixture of Ank^{GAG}1D4 and irrelevant substrate αRep-A3 at various protein ratios (middle panel). The parameters of each binding reaction, K_d and stoichiometry (N), are shown under the corresponding panel. NA, calculation not applicable.

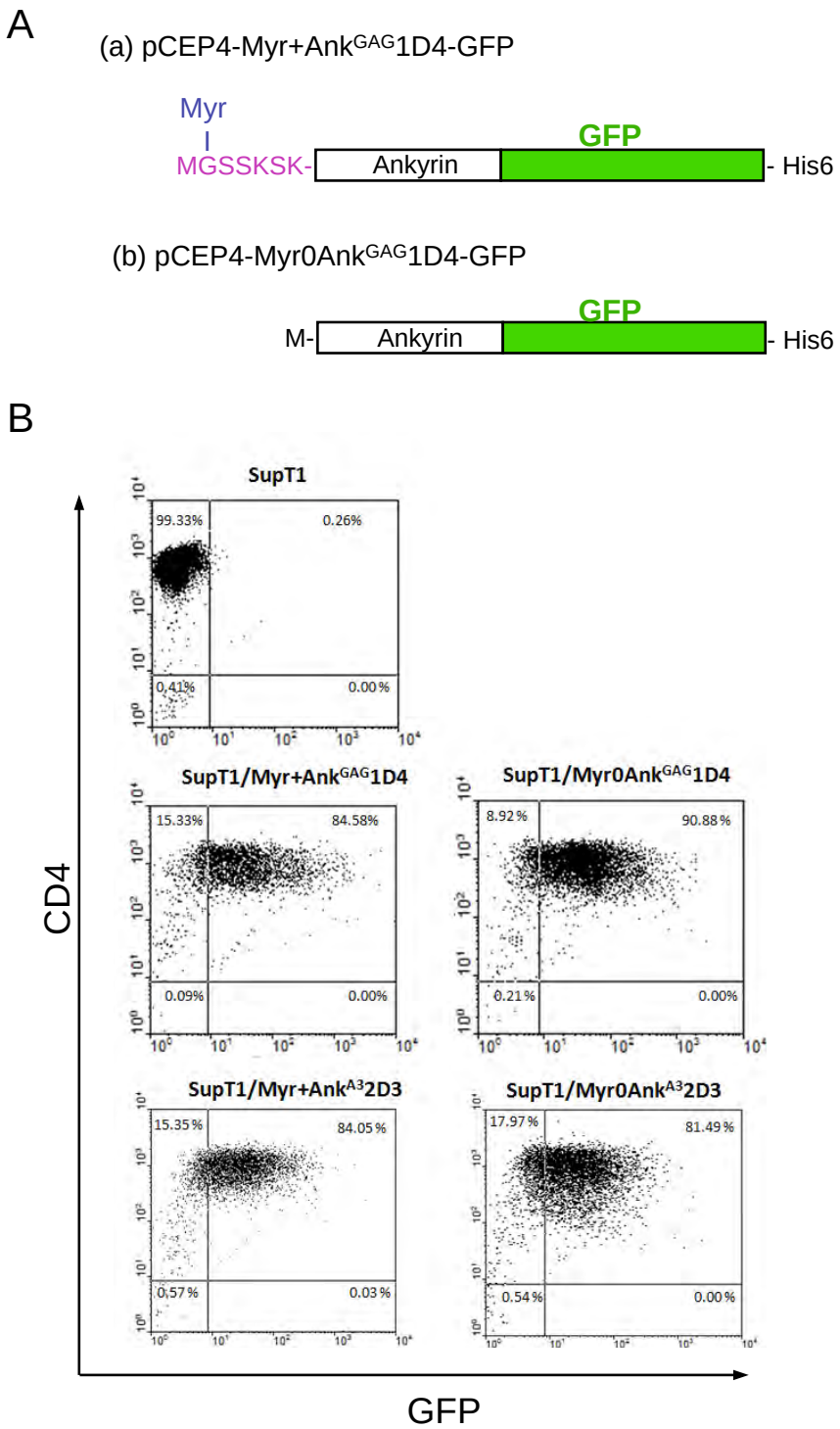

Figure 8 Characterization of SupT1 cells stably expressing artificial ankyrins. **(A)**, ***Ankyrin constructs***. Schematic representation of the artificial ankyrin constructs designed for stable expression in SupT1 cells, using pCEP4-based vector. Histidine-tag and green fluorescence protein (GFP; green box) were inserted at the C-terminus and in-phase with the ankyrin sequence (white box). Addition of a N-myristoylation signal (in purple red letters) to the Myr+Ank^{GAG}1D4-GFP clone resulted in the removal of the N-terminal methionine (M) and covalent linkage of myristic acid chain (Myr; in blue letters) to glycine-2 (G). **(B)**, ***Flow cytometry***. Expression of CD4 molecules at the surface of control SupT1 cells, SupT1/Myr+Ank^{GAG}1D4-GFP and SupT1/Myr0Ank^{GAG}1D4-GFP cells. Flow cytometry analysis was performed on nonpermeabilized cells, using monoclonal antibody against CD4, followed by PE-conjugated goat anti-mouse IgG.

HIV-1 infection of ankyrin-expressing SupT1 cell lines

SupT1/Myr+Ank^{GAG}1D4-GFP, SupT1/Myr0Ank^{GAG}1D4-GFP, SupT1/Myr+Ank^{A3}2D3-GFP and SupT1/Myr0Ank^{A3}2D3-GFP cells were infected with HIV-1_{NL4-3} virus at MOI 10 for 16 h at 37°C. Cells were harvested at day 11 post-infection (pi), and examined in confocal microscopy after permeabilization and reaction with anti-CAp24 mAb and PE-conjugated anti-mouse IgG-F(ab')$_2$ antibody. In infected cells expressing Myr+Ank^{GAG}1D4-GFP, the Gag and GFP signals superimposed in both cytoplasmic compartment and at the plasma membrane, as expected for two N-myristoylated, membrane-targeted partner proteins (Figure 9a). Gag and Gag-irrelevant N-myristoylated ankyrin Myr+Ank^{A3}2D3-GFP were also both addressed to the plasma membrane, but showed a lesser degree of colocalization (Figure 9c). No significant colocalization was detected for Gag and the non-N-myristoylated, Gag-irrelevant ankyrin Myr0Ank^{A3}2D3-GFP (Figure 9d). Interestingly, a significant degree of colocalization was observed for Gag and the non-N-myristoylated Gag-binder Myr0Ank^{GAG}1D4-GFP (Figure 9b). This implied that Gag and Myr0Ank^{GAG}1D4-GFP proteins interacted within the cytoplasm and were addressed as a complex to the plasma membrane, via the N-myristoylated signal carried by the Gag protein partner.

Negative effect of Ank^{GAG}1D4 on HIV-1 production

The possible antiviral activity of Ank^{GAG}1D4 on HIV-1 assembly and budding was first evaluated by syncytium formation. SupT1 cells expressing Myr+Ank^{GAG}1D4-GFP, Myr0Ank^{GAG}1D4-GFP, Myr+Ank^{A3}2D3-GFP and Myr0Ank^{A3}2D3-GFP, respectively, were infected with HIV-1_{NL4-3} virus at MOI 10 for 16 h at 37°C, and examined by phase contrast microscopy at day 11 pi. Numerous syncytia were observed in control samples of HIV-1-infected SupT1 cells (Figure 10, upper row) as well as in HIV-1-infected SupT1/Myr+Ank^{A3}2D3 and SupT1/Myr0Ank^{A3}2D3 cells (Figure 10, two bottom rows). However, very few syncytia were observed in HIV-1-infected SupT1/Myr0Ank^{GAG}1D4 cells (Figure 10, third row from the top), and very rare, if any, in HIV-1-infected SupT1/Myr+Ank^{GAG}1D4 cells (Figure 10, second row from the top).

Virus yields were quantitated in the extracellular medium of ankyrin-expressing SupT1 cells infected with HIV-1_{NL4-3} virus, in the same conditions as above. Culture supernatants were collected at different times pi, and virus progeny titer indirectly determined using ELISA/CAp24. A significant reduction of extracellular levels of CAp24 was observed at days 11 and 13 pi in the supernatants of SupT1/Myr+Ank^{GAG}1D4-GFP and SupT1/Myr0Ank^{GAG}1D4-GFP cells, compared to control cells, nontransduced HIV-1-infected SupT1 cells and SupT1 cells expressing the Gag-irrelevant ankyrin Myr+Ank^{A3}2D3-GFP (Figure 11A). There was a slight decrease of CAp24 levels in the culture medium of SupT1/Myr+Ank^{A3}2D3-GFP, compared to control SupT1 cells, and this effect was less pronounced in SupT1/Myr0Ank^{A3}2D3-GFP cells, which expressed a Gag-irrelevant, non-N-myristoylated ankyrin, (Figure 11A). The possibility that the increase in CAp24 yields at day 13 pi might be due to Ank^{GAG}1D4-escape HIV-1 mutant(s) was investigated: no mutation in the *gag* gene was found in RT-PCR amplicons derived from the HIV-1 progeny of SupT1/Myr+Ank^{GAG}1D4 cells harvested at day 13. However, this did not exclude that Ank^{GAG}1D4-resistant *gag* mutants could be found after a higher number of passages. Long-term cultures of HIV-1-infected SupT1 cells stably expressing Ank^{GAG}1D4 will be necessary to evaluate the viral genetic barrier to Ank^{GAG}1D4.

Viral loads were also determined at day-11 in the extracellular media, and the data confirmed the ELISA/CAp24. A significant inhibitory effect of Myr+Ank^{GAG}1D4 on HIV-1 replication was observed, with an average 600-fold lower virus progeny production, compared to control, nontransduced HIV-1-infected SupT1 cells (2, 500 × 10^7 genome copies/mL; Figure 11B). A significant decrease in HIV-1 production was also observed with the non-N-myristoylated ankyrin Myr0Ank^{GAG}1D4, although to a lesser degree compared to its N-myristoylated version (160-fold less; Figure 11B). These results indicated that the antiviral function of Ank^{GAG}1D4 occurred in both compartments, plasma membrane and cytoplasm, but with a higher efficiency when the ankyrin molecules were addressed to the plasma membrane (Figure 11B). This suggested that the antiviral function carried by Ank^{GAG}1D4 occurred at late step(s) of the virus life cycle, e.g. the assembly and budding of virus particles. As observed in ELISA/CAp24, SupT1 cells expressing Gag-irrelevant, N-myristoylated ankyrin molecule Myr+Ank^{A3}2D3-GFP showed some decrease (17-fold) in virus production (140 × 10^7 genome copies/mL; Figure 11B).

Ank^{GAG}1D4-mediated antiviral activity in HIV-1 life cycle: early versus late step(s)

As a partner of the HIV-1 structural protein CAp24, Ank^{GAG}1D4 might interfere with various step(s) of the HIV-1 life cycle. This included (i) virus uncoating, (ii) intracellular trafficking of incoming viruses, (iii) nuclear import of the viral preintegration complex, at early times, (iv) Gag oligomerisation, (v) virus particle assembly, and (v) extracellular budding, at late times post infection. To address this issue, integration events were evaluated by PCR amplification of *Alu-gag* junctions in HIV-1-infected, Ank^{GAG}1D4-expressing SupT1 cells harvested at day 11 pi. Control samples consisted of HIV-1-infected, nontransfected SupT1 cells and SupT1 cells expressing Gag-irrelevant, Ank^{A3}2D3 ankyrin. There was no significant difference between Ank^{GAG}1D4-expressing SupT1 cells

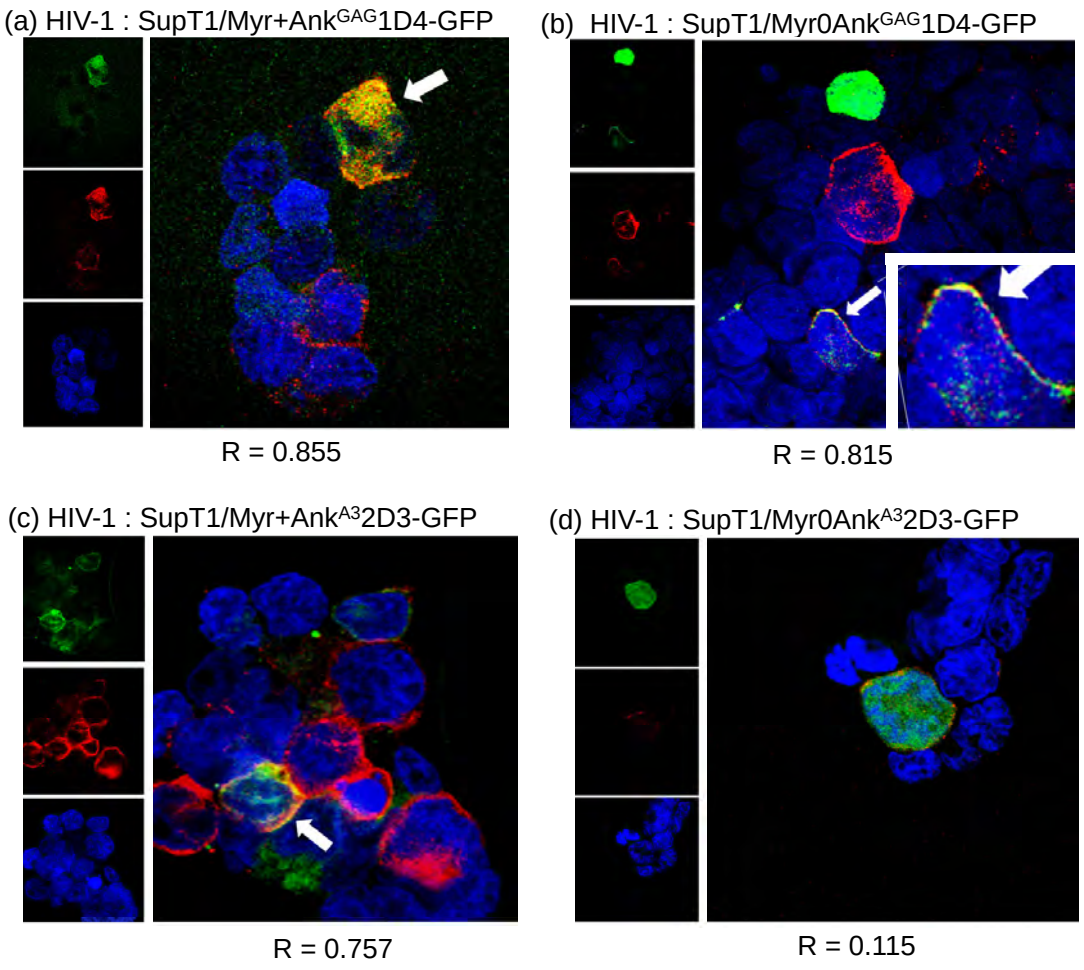

Figure 9 Confocal microscopy of HIV-1-infected, ankyrin-expressing SupT1 cells. HIV-1-infected SupT1/Myr+Ank^{GAG}1D4-GFP (panel **a**), SupT1/Myr0Ank^{GAG}1D4-GFP (**b**), SupT1/Myr+Ank^{A3}2D3-GFP (**c**), and SupT1/Myr0Ank^{A3}2D3-GFP cells (**d**) were collected at day 11 pi, permeabilized, immunolabeled with anti-CAp24 mAb and PE-conjugated anti-mouse IgG F(ab')$_2$ antibody (red signal), and nuclei counter-stained in blue with DAPI. Ankyrins were detected by their GFP-tag. Merged images are enlarged and shown on the right side of each panel. White arrows point to cells showing colocalization of ankyrin and Gag proteins. R, the Pearson correlation coefficient for signal colocalization, was determined using the Olympus FluoView software.

and control cells (Table 2), suggesting that the Ank^{GAG}1D4-mediated antiviral effect took place at the post-integration phase of the virus life cycle. To validate this negative result, control experiments of integration blockage were carried out using the HIV-1 integrase inhibitor Raltegravir™ (RAL). RAL was added at increasing molarities (1, 10 and 100 nM) to the SupT1 cell culture medium 24 h prior to HIV-1 infection, and maintained for 7 days [37]. No significant alteration of the cell viability was observed within this molarity range (Additional File 1). No viral integration was detectable at RAL molarities over 10 nM (Table 3 and Additional File 1), a result which was consistent with the IC_{50} value of 10 nM for RAL [37]. To further dissect the nature of the post-integration blockage of HIV-1 provoked by Ank^{GAG}1D4, the fate of the viral target of Ank^{GAG}1D4, the Gag protein, was analyzed in HIV-1-infected SupT1 cells harvested at late times pi and subjected to cell fractionation. Whole cell lysates and cell fractions were assayed for Gag content by ELISA/CAp24, and the Gag protein pattern analysed by SDS-PAGE and Western blotting. The CAp24 levels were significantly lower in Myr+Ank^{GAG}1D4- and Myr0Ank^{GAG}1D4-expressing cells, compared to control cells expressing no exogenous ankyrin or the Gag-irrelevant ankyrin Ank^{A3}2D3 (Figure 12). A similar decrease was observed in the whole cell lysate and membrane fraction (Figure 12, compare panels A and B), implying that the antiviral effect of Ank^{GAG}1D4 did not involve the trafficking of Gag to the plasma membrane.

Western blot analysis showed a drastic reduction of all Gag protein species in Myr+Ank^{GAG}1D4- and Myr0-Ank^{GAG}1D4-expressing cells compared to control cells

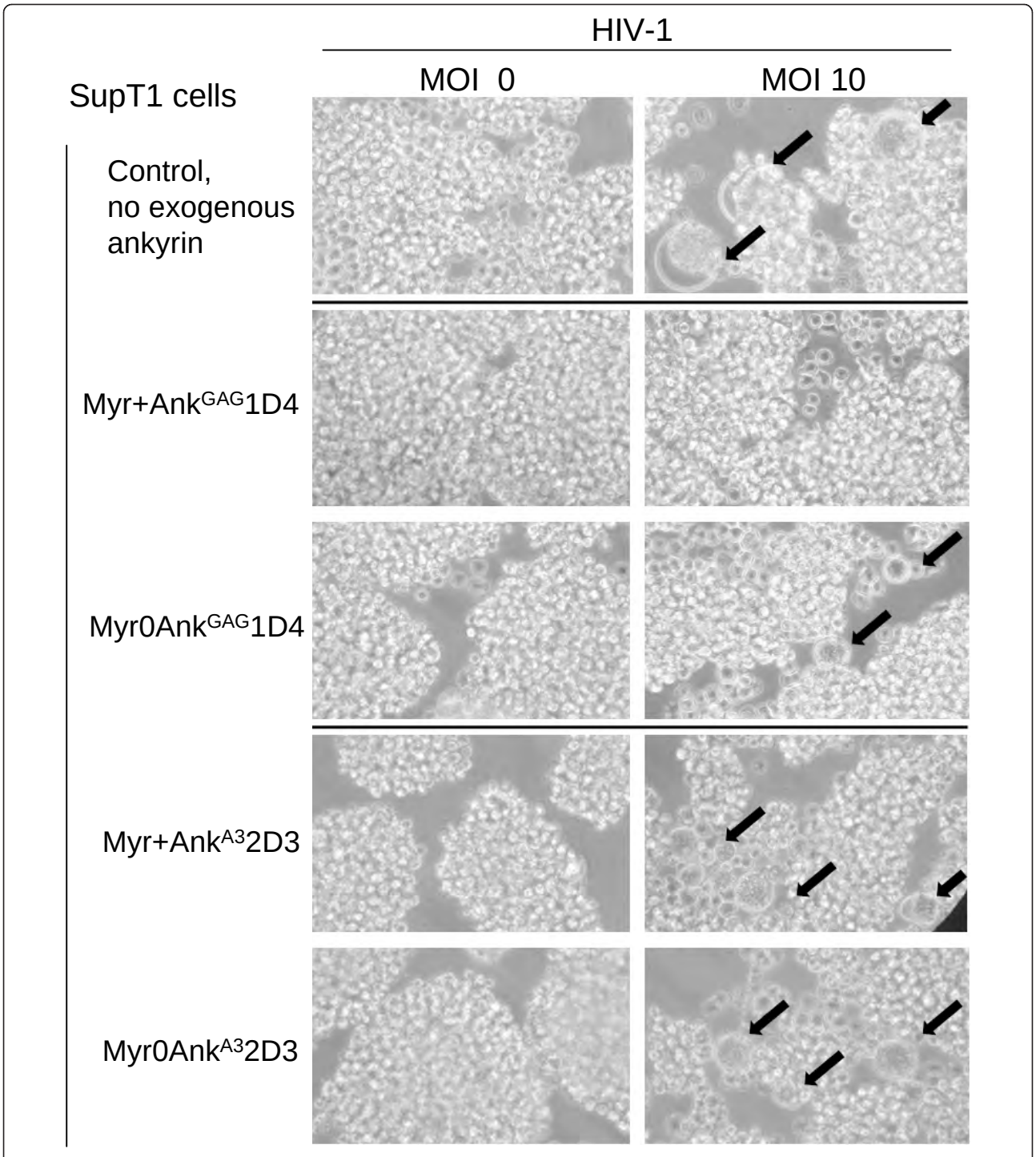

Figure 10 HIV-1-induced syncytium formation. SupT1/Myr+Ank^{GAG}1D4, SupT1/Myr0Ank^{GAG}1D4, SupT1/Myr+Ank^{A3}2D3 and SupT1/Myr0Ank^{A3}2D3 were mock-infected (MOI 0; left column) or infected with HIV-1 (MOI 10, right column). Cells were observed at 400X magnification using an inverted microscope. Black arrows point to syncytia.

(Figure 12C, D). This pattern excluded a possible interference of Ank^{GAG}1D4 with the proteolytic processing of Gag, which might provoke a premature cleavage of the Pr55Gag precursor.

Viral specificity of Ank^{GAG}1D4

The viral specificity of Ank^{GAG}1D4 was evaluated on HIV-Luc and Moloney murine leukemia virus (MLV)-Luc vectors, which express the luciferase-encoding reporter gene.

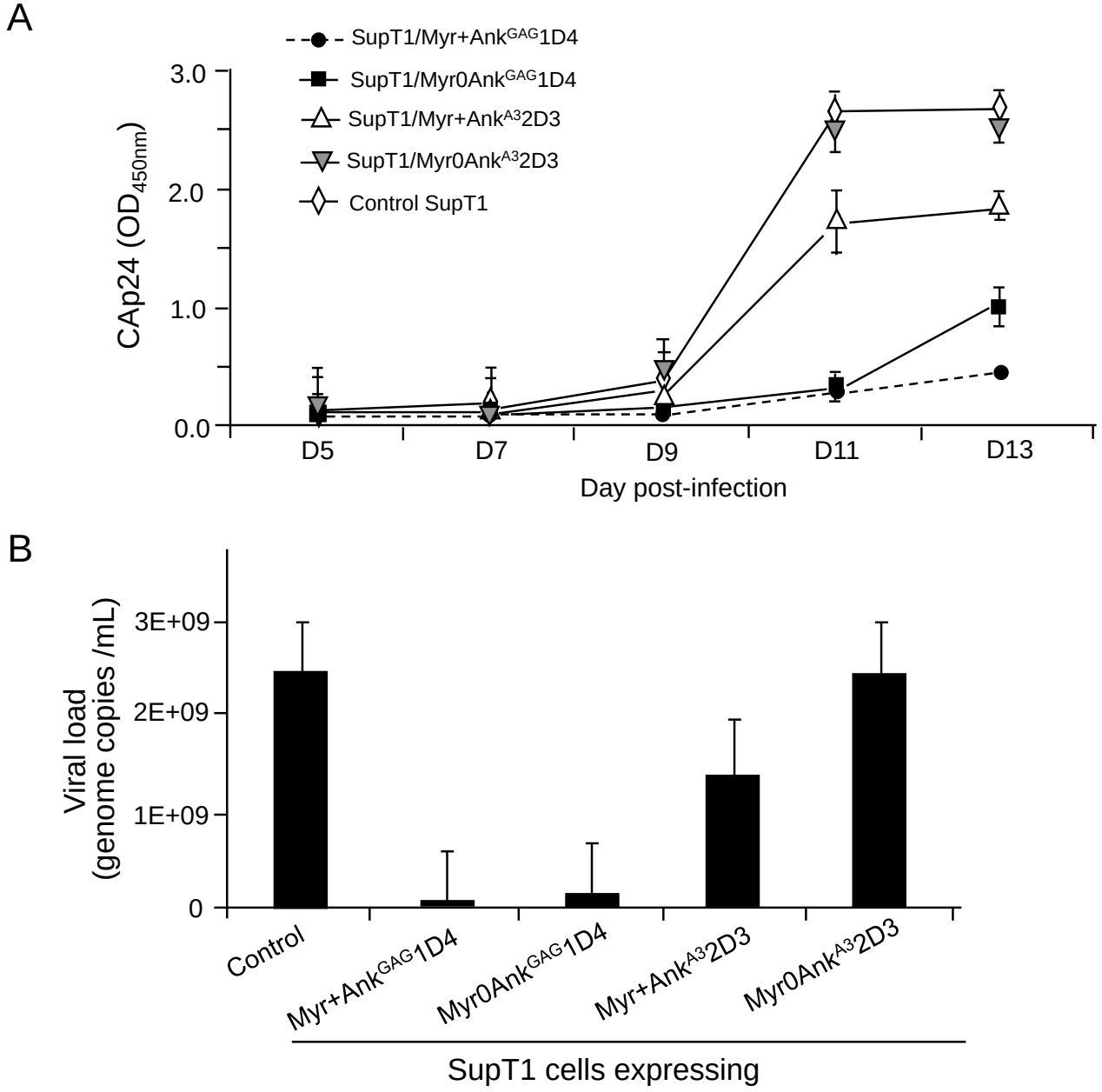

Figure 11 Ank^{GAG}1D4-mediated inhibitory effect on HIV-1 replication. **(A)**, ***CAp24 titration***. SupT1/Myr+Ank^{GAG}1D4 and SupT1/Myr0Ank^{GAG}1D4 stably expressed the N-myristoylation and non-N-myristoylation versions of the H6MA-CA-binding ankiryn Ank^{GAG}1D4, respectively. SupT1/Myr +Ank^{A3}2D3 and SupT1/Myr0Ank^{A3}2D3 expressed the N-myristoylation and non-N-myristoylation versions of the αRep-A3-binding ankyrin Ank^{A3}2D3, respectively. Cells were infected with HIV-1 $_{NL4-3}$ at MOI 10, and cell culture supernatants collected at time intervals (5, 7, 9, and 11 days pi) and virus progeny titers determined by CAp24 assays, using ELISA. Results shown are mean (m) from triplicate experiments ± SEM. **(B)**, ***Viral load***. The copy numbers of viral genome were determined in the culture supernatants collected on day-11, using Cobas Amplicor tests. Data presented are from triplicate experiments (m ± SEM). The average values were the following: Control SupT1 cells = 2, 470 × 10^7 copies/mL; SupT1/Myr+Ank^{GAG}1D4 = 4 × 10^7; SupT1/Myr0Ank^{GAG}1D4 = 15 × 10^7 ; SupT1/Myr+Ank^{A3}2D3 = 140 × 10^7 ; SupT1/Myr0Ank^{A3}2D3 = 245 × 10^7.

HIV-Luc was produced in 293T cells, and MLV-Luc was produced in the GP2-293-Luc packaging cells, which stably express the MLV *gag-pol* gene products and a packageable, luciferase-encoding MLV RNA transcript [38]. HIV-Luc and MLV-Luc producer cells were transfected by the different pCEP4-ankyrin plasmids, and the extracellular media were collected at 72h posttransfection and assayed for vector yields. Vector titers were determined by measuring the luciferase activity in HIV-Luc- and MLV-Luc-infected 293T cells [39]. The expression of

Table 2 HIV-1 integration events in control and ankyrin-expressing SupT1 cells [(a)]

Sequence amplified	No ankyrin	Myr+Ank^{GAG}1D4	Myr0Ank^{GAG}1D4	Myr+Ank^{A3}2D3
Alu-*gag* junctions	28.3 ± 0.2	28.6 ± 0.3	30.5 ± 0.1	27.8 ± 0.2
GAPDH	25.7 ± 0.4	26.1 ± 0.5	25.5 ± 0.5	24.5 ± 0.1

[(a)] The level of HIV-1 integration in SupT1 cells harvested at day 11 pi was evaluated by quantitative PCR amplification of host cell DNA extracts, using primers specific to *Alu-gag* junctions, and to cellular *GAPDH* gene used as the internal control. Figures shown in the Table are *C*ts values, mean ± SD (n = 3).

Myr+Ank^{GAG}1D4 and Myr0Ank^{GAG}1D4 in vector producer cells resulted in a 20-fold lower production of HIV-Luc (Figure 13A), while MLV-Luc production was decreased by 6- to 8-fold (Figure 13B). These results suggested a significant degree of Ank^{GAG}1D4 cross-reactivity between HIV-1 and MLV, two evolutionarily divergent retroviruses. Of note, the expression of non-relevant ankyrins Myr+Ank^{A3}2D3 and Myr0Ank^{A3}2D3 decreased both HIV-Luc and MLV-Luc yields by a factor of 2, indicative of a basal interference level for ankyrins (Figure 13A, B).

Discussion

Although HAART can significantly reduce HIV-1 replication and prolong the life of HIV-infected individuals [1,10], the treatment is lifelong, causes a variety of side-effects with cumulative toxicities, and is responsible for the development of resistant viral strains [40]. Moreover, host genome-integrated proviral DNA persists in latent tissue reservoirs in HIV-1-infected individuals [1,10]. All these disadvantages have led many laboratories to consider the use of anti-HIV gene therapy, either as a stand-alone approach, or as an adjuvant to pharmacological regimens [10,41]. Gene therapy offers the potential of preventing progressive HIV infection by a sustained interference with the intracellular cascade that leads to virus replication. The aim of the present study was to investigate a novel class of intracellular protein-based antiviral agents, which would interact with viral proteins at critical steps of the virus life cycle and act as intracellular inhibitors of viral replication. One particular HIV-1 protein target which still lacks specific inhibitor(s) at the stage of clinical development is the Gag polyprotein precursor Pr55Gag and the late step of viral assembly [41-49]. The viral protein that we used as the target in this study was a truncated form of Gag precursor, consisting of the MAp17 and CAp24 domains (MA-CA), with ankyrin-derived repeat proteins acting as intracellular interactors and potential blockers.

Table 3 HIV-1 integration events in control and Raltegravir-treated SupT1 cells [(a)]

Sequence amplified	Raltegravir (nM)			
	0	1	10	100
Alu-*gag* junctions	26.9 ± 0.2	36.1 ± 0.3	ND	ND
GAPDH	23.3 ± 0.5	23.7 ± 0.4	24.1 ± 0.3	23.9 ± 0.2

[(a)] Aliquots of HIV-1-infected SupT1 cells were pretreated with Raltegravir at 0, 1, 10, and 100 nM, respectively, for 24 hr prior to HIV-1 infection (MOI 10), and the drug maintained at the indicated concentrations for 7 days. Host cell genomic DNA was extracted on day 7 pi, and the level of HIV-1 integration in SupT1 cell lines was evaluated by quantitative PCR amplification of DNA extracts, using primers specific to *Alu-gag* junctions and to cellular *GAPDH* gene as the internal control. Figures shown in the Table are *C*ts values (mean ± SD, n = 3). ND, not detectable (below the threshold of detection).

Ankyrin-repeat proteins or DARPins represent potential candidates of anti-HIV-1 molecules, since they are considered as the best alternative to intracellular antibodies, single chain antibodies or intrabodies, in terms of binding affinity, specificity and stability [12-14,23,26,27,30]. The molecular architecture of ankyrin-derived repeat proteins consists of a common structural framework made of a theoretically unlimited number of helix-turn-helix repeat motifs or modules, arranged in a parallel orientation. The same positions in consecutive repeats can accommodate variable amino acid residues, selected randomly, with the exclusion of cysteine, glycine and proline [23]. All variable amino acid side chains are oriented on the same side of the helices (refer to Figure 3B), and are presented on a large solvent-accessible, hyper-variable surface which is well adapted to bind large surfaces on desired targets [22]. More importantly, the correct folding and stability of these molecules do not depend on disulfide bridge formation and can therefore occur in all environments, including extracellular and intracellular milieu, which is a major advantage over antibodies [26,27,30]. The number of modules is theoretically unlimited, and highly variable, in natural ankyrins, but the number of ankyrin modules in selected binders is generally restricted to two or three, although longer proteins up to fifteen modules are also present in libraries [23]. This is not due to a better functional adaptation of shorter ankyrins, but likely to the proportion of coding sequences in short versus longer proteins. Long ankyrin constructs are more likely to incorporate at least one frameshift mutation; thus the useful diversity of longer sequences is limited. Nevertheless, arrangements of two to three modules have enough surface and variability to generate binders with high affinity and specificity.

The construction of a DARPin library with random amino acids at predetermined, surface exposed positions in the conserved alpha-helical module of ankyrin required the use of oligonucleotides synthesized from

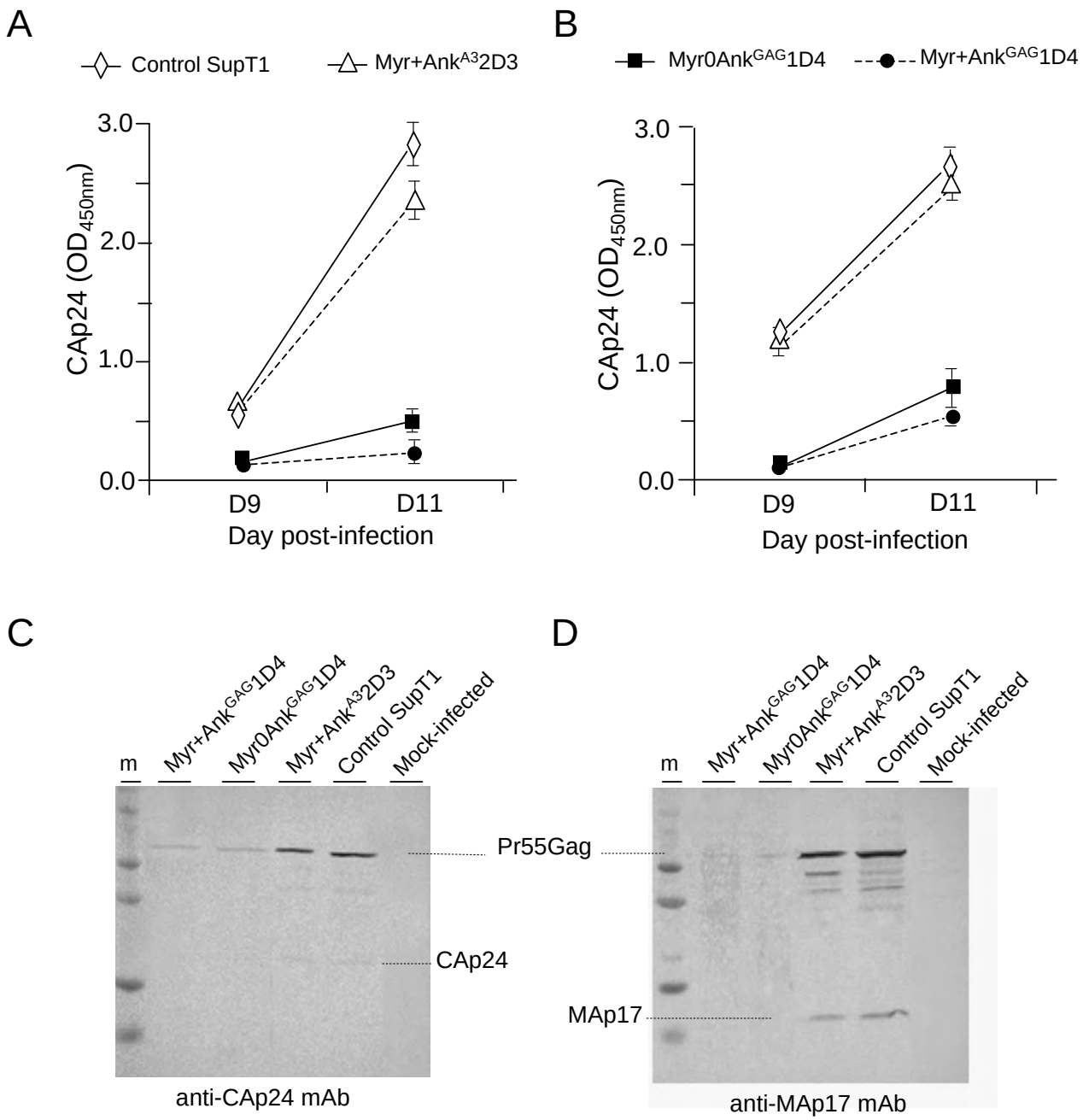

Figure 12 Effect of ankyrin expression on Gag protein levels in HIV-1-infected SupT1 cells. Control and ankyrin-expressing SupT1 cell lines were infected with HIV-1_{NL4-3} inoculum at 10 MOI for 16 h at 37°C. After washing with serum-free medium, cells were resuspended in prewarmed medium containing 500 μg/mL hygromycin B and 10% FCS, and seeded into 6-well plates. Cells were harvested and lysed at day 9 and 11 pi. Cell lysates were fractionated into nuclear pellet, membrane compartment and cytosol, and each fraction assayed for CAp24 content by ELISA, and Gag proteins by SDS-PAGE and Western blot (WB) analysis. Protein concentration in all samples was normalized to 10 μg/mL. **(A)**, ELISA of whole cell lysates (WCL); **(B)**, ELISA of membrane/particulate fractions. Data presented are mean from triplicate experiments, with error bars indicating the standard error to mean. **(C, D)**, SDS-PAGE and WB analysis of WCL. Samples were taken at day 11 pi, and analyzed using anti-CAp24 (C) and anti-MAp17 mAb (D).

trinucleotide synthons. This constituted the major obstacle in the construction of such libraries, since this technology is not commonly accessible. To overcome this inconvenience, we developed an alternative strategy based on a mixture of partially degenerated oligonucleotides and generated a phage-displayed ankyrin-repeat library with a reasonable degree of diversity. What is referred to as 'useful diversity' of the ankyrin library

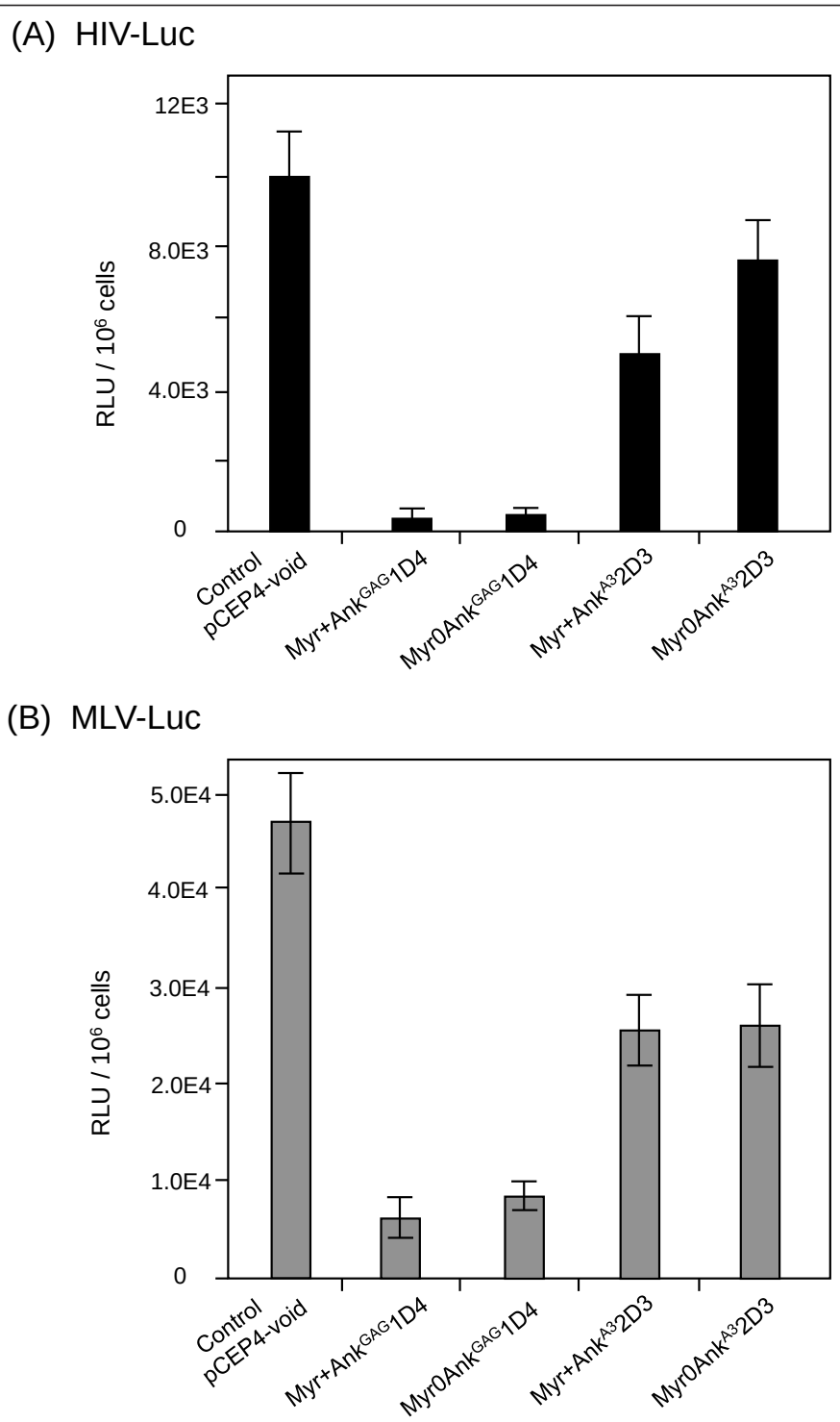

Figure 13 Influence of Ank^{GAG}1D4 on the HIV-Luc (A) and MLV-Luc vector (B) yields. VSV-G-pseudotyped HIV-Luc released by 293T cells cotransfected with pNL4-3Luc(R-E-), phCMV-G and pCEP4-ankyrin **(A)**, and VSV-G-pseudotyped MLV-Luc vector released by GP2-293-Luc cells cotransfected with phCMV-G and pCEP4-ankyrin (**B**), were titrated on 293T cells, using the luciferase assay. RLU, relative light units. Data shown are m ± SEM, *n* = 3.

generated in this study was estimated to be 6×10^7 independent coding sequences.

By screening our phage-displayed library on the MA-CA domains of HIV-1 Gag used as the bait, we isolated Ank^{GAG}1D4, a Gag-specific, trimodular ankyrin with an apparent molecular mass of 16.5 kDa. The potential capacity of Ank^{GAG}1D4 to interfere with HIV-1 replication was evaluated in SupT1 cells expressing the N-myristoylated (SupT1/Myr+Ank^{GAG}1D4) and the non-N-myristoylated (SupT1/Myr0Ank^{GAG}1D4) versions of Ank^{GAG}1D4, respectively. We observed a lower permissiveness to HIV-1 for both cell lines, with a significant reduction of the HIV-1 progeny released in the culture medium, compared to control cells expressing Gag-irrelevant ankyrins or no exogenous ankyrin. The Ank^{GAG}1D4 anti-HIV-1 effect was found to occur at the post-integration phase of the virus life cycle, a result consistent with Gag, the viral structural protein being the target of Ank^{GAG}1D4. The results obtained with the MLV-Luc vector indicated that Ank^{GAG}1D4 could have antiviral effect on phylogenetically distant retroviruses. Interestingly, the interferon-induced cellular protein HERC5, which acts as a host restriction factor of HIV-1 infection, has been found to block both HIV-1 and MLV Gag particle assembly with a similar efficiency [50]. The cross-reactivity observed between HIV-1 and MLV implied that Ank^{GAG}1D4, which was selected on HIV-1 Gag, recognized a conformational structure or/and motif conserved among retroviral Gag prolyproteins.

The mechanism of Ank^{GAG}1D4 activity was further dissected in HIV-1-infected SupT1 cells. No premature cleavage or sequestration of the Pr55Gag precursor in a cellular compartment was observed, but a lower Gag content was found in both Myr+Ank^{GAG}1D4- and Myr0-Ank^{GAG}1D4-expressing cells, compared to control cells. Image analysis of HIV-1-infected, ankyrin-expressing SupT1 cells suggested that the non-N-myristoylated Myr0Ank^{GAG}1D4 bound to nascent or newly synthesized Gag polyprotein within the cytoplasm, and that the Myr0Ank^{GAG}1D4-Gag complex was addressed to the plasma membrane via the N-myristoylation signal carried by Gag. N-myristoylated Myr+Ank^{GAG}1D4, however, was genetically designed for plasma membrane targeting. In both cases, the formation of Ank^{GAG}1D4-Gag protein complexes likely resulted in the depletion of Gag from the pool of molecules available for virus assembly. Although these results suggested that the specific interaction of Ank^{GAG}1D4 with the CA NTD negatively interfered with the Gag assembly and budding pathway at the plasma membrane, some interference with the interaction of Gag with the viral genomic RNA could not be excluded. This early interaction has been shown to occur at perinuclear/centromal sites [51] and could be the target of the cytoplasmic Myr0Ank^{GAG}1D4.

The molecular and cellular mechanism of the Ank^{GAG}1D4 antiviral effect might also involve the plasma membrane anchoring of Gag via its N-myristoylated signal. Several reports have shown a link between membrane anchoring of the Pr55Gag precursor and its translation. N-myristoylated Pr55Gag protein regulates its own translation *in vitro* in the presence of plasma membrane-containing fraction [52,53]. In the present study, SupT1/Myr+Ank^{GAG}1D4 cells showed a lower permissiveness to HIV-1 infection, compared to SupT1/Myr0Ank^{GAG}1D4, suggesting that the addressing of Ank^{GAG}1D4 to the plasma membrane compartment via a N-myristoylated signal significantly increased its efficiency as antiviral agent. Furthermore, the plasma membrane-targeted, Gag-irrelevant ankyrin Myr+Ank^{A3}2D3 showed some negative interference with HIV-1 replication. It might be therefore hypothesized that Myr+Ank^{GAG}1D4 and Myr+Ank^{A3}2D3 occupied anchoring sites in critical domains of the plasma membrane inner leaflet which were required for the insertion of Pr55Gag/genomic RNA complex and the initiation of virus assembly [54-56]. Alternatively, but not exclusively, Myr+Ank^{GAG}1D4 and Myr+Ank^{A3}2D3 might compete with Pr55Gag for N-myristoyl-transferases, resulting in decreased levels of N-myristoylated Pr55Gag molecules competent for plasma membrane anchoring, and viral particle formation and egress. The latter hypothesis is consistent with a previous report describing the inhibitory effect of competing unsaturated fatty acids on viral budding [57].

The binding site of Ank^{GAG}1D4 was mapped to the N-terminal moiety of the CA domain, within the first 110 residues. HIV-1 CA domain is composed of two highly structured subdomains, the N-terminal subdomain (NTD, 1-145) and the C-terminal subdomain (CTD, residues 149-219), separated by a flexible hinge [58-60]. The Ank^{GAG}1D4 binding site encompassed two highly accessible and functionally important regions in the CA NTD: (i) the aminoterminal ß-hairpin, and (ii) the cyclophilin-A (CypA) loop [58], which contains Proline-90 and Isoleucine-91. Pro-90 is the substrate of the CypA *cis-trans* peptidyl-prolyl isomerase, and Pro-90 and Ile-91 are two residues essential for virion incorporation of CypA, an HIV-1 infectivity factor [61-63]. Therefore, besides its effect on virus assembly, Ank^{GAG}1D4 could also decrease the infectivity of HIV-1 virions, via a blockage of the CypA encapsidation.

Retroviral Gag and GagPol polyproteins are incorporated into immature virus particles through Gag-Gag and Gag-GagPol interactions. In this assembly pathway, Gag dimerization, mediated by Gag-RNA interaction, represents a critical step [54-56,64-68]. Crystal analysis of HIV-1 CA has shown that the CTD is involved in the

formation of CA dimers [58-60,69,70]. If the direct interaction of Ank^{GAG}1D4 with the CA NTD negatively interfered with the Gag multimerization process, this would occur via the NTD-NTD hexamerization interface, or the NTD-CTD interface, and not the CTD-CTD dimerization interface [58]. This differed from other peptide inhibitors of HIV-1 Gag assembly which have been shown to target the CTD-CTD interface and block the CA dimerization [44,46,60]. Analysis of the H_6MA-CA/Ank^{GAG}1D4 complex suggested a stoichiometry of 1:1 for the pair of reagents, and a moderate affinity of Ank^{GAG}1D4 for its H_6MA-CA target *in vitro*. As a comparison, the dodecapeptide ITFEDLLDYYGP (abbreviated CAI, for capsid assembly inhibitor), isolated by screening of a phage-displayed peptide library on the HIV-1 CA domain, has been found to bind to CA with a K_d of ~ 15 μM, and to inhibit the Gag multimerization and formation of immature virus particles at an average 50% inhibitory concentration of about 10 μM [44,46].

Conclusions

The present study demonstrated the potential of ankyrin-repeat proteins as a novel class of intracellular antivirals. The data obtained with ankyrin Ank^{GAG}1D4 showed that a significant antiviral effect could be obtained with an ankyrin molecule targeted to a structural protein of the HIV-1 virion, which was the CA domain of the Gag precursor. The Ank^{GAG}1D4 molecule therefore represents an attractive platform for the design of more efficient ankyrin-based intracellular inhibitors of HIV-1 which would negatively interfere with the virus assembly and egress pathway. More generally, the antiviral activity shown by Ank^{GAG}1D4 should contribute to promote the use of ankyrin-repeat proteins as intracellular therapeutic agents against a variety of pathogens.

Methods

Cells

Human embryonic kidney cells HEK-293T cells were obtained from the American Type Culture Collection (ATCC, Manassas, VA), and maintained as monolayers in Dulbecco's modified Eagle's medium (DMEM; Invitrogen) supplemented with 10% fetal bovine serum (FBS; Invitrogen), penicillin (100 U/mL), and streptomycin (100 mg/mL) at 37°C, 5% CO2. GP2-293 and GP2-293-Luc packaging cells stably expressed the Moloney murine leukemia virus (MLV) *gag-pol* gene products, and GP2-293-Luc contained an additional packageable, luciferase-encoding viral RNA transcript [38] expressed from the luciferase reporter vector pLLRN (BD Biosciences Clontech). *Spodoptera frugiperda* (Sf9) cells were maintained as monolayers at 28°C in Grace's insect medium supplemented with 10% fetal bovine serum (FBS) and antibiotics (Invitrogen). They were infected with recombinant baculovirus at a multiplicity of infection (MOI) ranging from 2 to 10 PFU/cell, as previously described [71-74]. SupT1 cell lines stably expressing the ankyrin-repeat proteins were generated using the pCEP4-based vector (Invitrogen). Transfected SupT1 cells were maintained in complete RPMI containing hygromycin B (400 μg/mL).

Plasmids and vectors

Plasmid pQE-30 (Qiagen) was used for production of 6xHis-tagged recombinant proteins in bacterial cells. Plasmid pNL4-3, 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 the wild-type HIV-1 MA-CA and CA domains, and insertion into the pBlueBac4.5 plasmid (Invitrogen). The pBlueBac4.5 transfer vector was recombined with the genome of *Autographa californica* multiple nucleopolyhedrosis virus (AcMNPV), to generate recombinant baculoviruses AcMNPV-H_6MA-CA and AcMNPV-H_6CA were used to produce the H_6MA-CA and H_6CA recombinant proteins. The pCEP4 vector (Invitrogen) was used for constitutive, episomal expression of designed ankyrins from the CMV promoter in SupT1 cells. VSV-G-pseudotyped HIV-1-luciferase vector was recovered from the culture supernatant of 293T cells cotransfected with equal doses of pCEP4-ankyrin, phCMV-G and pNL4-3Luc(R-E-) plasmids (3 μg of each plasmid per 10^6 cell aliquots), as previously described [39]. VSV-G-pseudotyped MLV-luciferase vector was recovered from the culture supernatant of GP2-293-Luc cells cotransfected with equal doses (3 μg/10^6 cells) of pCEP4-ankyrin and phCMV-G. Cell culture supernatants containing the VSV-G-pseudotyped HIV-1 or MLV vectors (abbreviated HIV-Luc and MLV-Luc, respectively) were harvested at 72 h posttransfection, aliquoted and used for infection of 293T cells. HIV-Luc or MLV-Luc vector titers were determined at 24 h pi by luciferase assay of 293T cell lysates [39].

Construction of ankyrin-repeat protein library

The artificial ankyrin library was constructed using a combined phage display/expression vector based on pHDiExDsbA-Ank15 [75]. This vector was used for low level expression of DARPins fused to the M13 g3p truncated protein in phage display experiments. Since this construction had also a T7 promoter and a suppressible stop codon between the DARPins and g3p coding sequences, it could also be used for periplasmic expression of nonfused DARPins in non supE strains of *E. coli* expressing T7 polymerase. DARPins are extremely stable proteins and are efficiently translocated to periplasmic space only provided that they are fused to the SRP export sequence [76]. For cytoplasmic expression, the sequence

encoding soluble ankyrin proteins were inserted into the pQE-30 expression vector (Qiagen), using M15 (pREP4) strain (Qiagen) for expression. Bacterial cells XL-1 Blue MRF' (Stratagene) were used as host cells for the generation of the library and propagation of phages displaying the artificial ankyrin-repeat protein library.

The generation of our artificial ankyrin-repeat proteins library was based on that of DARPins library previously described [23], with the following modifications. A limited number of changes were introduced in the ankyrin-repeat consensus sequence in order to create a type II non-palyndromic restriction site *Bsm* BI within the ankyrin module. This site was used to produce ankyrin module-encoding microgenes from circularly amplified products and for their subsequent directional polymerization, in order to create the library. The *Bsm* BI recognition site was introduced by replacing glutamate-21 by arginine (E21R substitution) and valine-22 by leucine (V22L), using the appropriate nucleotide changes. In order to minimize possible charge repulsion involving the newly introduced R21, a compensatory change was made, consisting of a K-to-E mutation at position 25, which introduced a negative charge in the consecutive turn of the same alpha helix. These modifications were not expected to interfere with the folding or stability of the ankyrin module, since these types of amino acid residues are commonly found at equivalent positions of natural protein with ankyrin repeats. Furthermore, the changes that we created were located on the face of the protein opposite to the binding surface, and therefore should not interfere with the potential binding activity of the artificial ankyrins.

Further changes with respect to previously described libraries were introduced in the design of this library. The side chain of residue located at position 10 (helix-1) was oriented toward the binding surface and was therefore partially randomized, while position 26 and 33, not directly located within the binding surface, were kept constant. In DARPins libraries previously described, the modification of each variable amino acids of the ankyrin repeats were essentially performed randomly, with the exclusion of cysteine (C), glycine (G) and proline (P). This was made possible by using oligonucleotides synthesized from trinucleotide synthons. As this technology is not commonly accessible, we devised an alternative strategy based on a mixture of partially degenerated oligonucleotides, and comprising of the following steps (Figure 2).

(i) The oligonucleotides pools were designed to exclude undesired cysteine residues and to mimic the natural residue frequency of residues at each defined position where amino acid residues could vary, i.e. position numbers 2, 3, 5, 10, 13 and 14 (Table 1). The position-specific, natural distribution of amino-acids frequencies were computed from the natural ankyrin modules collections defined in the Prosite database (PS50088). The choice for the set of partially degenerated codons was in fact a compromise, in order to minimize the numbers of codons (and therefore of oligonucleotides), while maintaining the side chains diversity close to the chemical diversity encountered in natural ankyrin-repeat proteins.

(ii) The repeat sequences were generated by using a set of oligonucleotides containing a set of partially degenerated codons. The sequence coding a single repeat was divided into four fragments (Figure 1): Va (variable fragment a), Vb (variable fragment b), Vc (variable fragment c) and C1 (constant fragment). Each variable fragment was generated by mixing a pool of oligonucleotides with randomized positions encoded with different combination of partially degenerated codons (Table 1).

(iii) All synthetic fragments (Va, Vb, Vc, and C1) were hybridized with reverse oligonucleotides linkers ("bridging" fragments; Vb-rev, C2, and C3) at equal molarity by heating at 95°C for 5 min, followed by progressive refrigeration to 25°C at the rate of 0.1°C/min.

(iv) To generate the circularized template, the hybridized product was ligated by T4 DNA ligase (New England Biolabs, NEB), purified using the NucleoSpin® Extract II kit (Macherey-Nagel), and used as the template for Rolling Circle Amplification (RCA) process, using the Illustra TempliPhi 100 amplification kit (GE Healthcare, Bio-Sciences).

(v) The polymerized product was incubated at 65°C for 15 min and subsequently treated with *Bsm*B I (NEB) at 55°C for 4 h, resulting in a mixture of mono-repeat ankyrin microgenes.

(vi) The mixture of generated fragments was subjected to a hetero-polymerization process for the generation of repeat protein library, using a procedures adapted from a previous work on a different type of repeat protein [33]. In brief, the pool of mono-repeat ankyrin microgenes were inserted into and ligated to a specially designed "acceptor" vector containing the N- and C-cap of DARPins (Figure 2). This vector was first cleaved with *Bsm*B I and *Kpn* I (Fermentas) to generate the cohesive ends compatible with ankyrin repeats microgenes. The *Kpn* I cleavage, although not strictly necessary, was used to minimize the vector recircularization which would compete with ankyrin-repeat polymerisation. Once ankyrin repeats were ligated with N-Cap, vector was cleaved with *Bsp*M I (NEB) and recircularised by intramolecular ligation. This resulted in the elimination of the Rep cloning sites regions and its replacement by a variable number of ankyrin repeats between the N- and C-caps. The ligation product was transfected into electrocompetent XL-1 Blue cells. Transformed cells were selected on LB agar containing ampicillin (100 μg/mL). The number of ankyrin repeats was determined by gel electrophoresis, after digestion of

the vector pool with *Not* I (NEB) and *Hind* III (NEB). The quality of the ankyrin library, based on the proportion of readthrough clones, was evaluated by CoFi blot analysis as previously described [33].

Construction of expression vectors

(i) Baculoviral vectors (AcMNPV). The baculovirus transfer vector encoding His-tagged MA-CA domains of Gag (H_6MA-CA) was generated as described elsewhere [77]. For production of recombinant His-tagged CAp24 domain of Gag (H_6CA), the gene encoding H_6CA was amplified from the parental vector pNL4-3 by standard PCR protocol using pair of primers: FWD-p24 *Nhe* I, 5'-GAGGAGGAGGTGCTATAGTGCAGAACCTCCAG-3' and REV-p24 *Kpn* I, 5'-GAGGAGGAGCTGGTACCTTACAAAACTCTTGCTTTATGGCC-3'. The PCR fragment was treated with *Nhe* I and *Kpn* I and subsequently cloned into the pBlueBac4.5 transfer vector, resulting in plasmid pBlueBac-H_6CA.

(ii) Bacterial cell vectors (pQE-30). Ankyrin genes encoding H_6MA-CA or αRep-A3-binder (αRep- previously described as αRep-n4-a (pdb-code 3LTJ; [33]) were inserted into the pQE-30 ankyrin acceptor vector, designed for soluble protein production. The acceptor vector was constructed by inserting the hybridization product of two synthetic oligonucleotides, pQE-Ank-Adapt-Fw (5'-GATCCGCGGCCGCAAACGCGTAAA-3') and pQE-Ank-Adapt-Re (5'-AGCTTTTACGCGTTTGCGGCCGCG-3'), into the *Bam* HI and *Hind* III sites of the pQE-30 vector, resulting in the insertion of a *Not* I restriction site into pQE-30. Phagemid pHDiExDsbA was treated with *Not* I and *Hind* III, and the resulting *Not* I-*Hind* III fragment was cloned into the same sites of the pQE-30 acceptor vector. The resulting pQE-30 vector contained the gene coding for Gag-binding or αRep-A3-binding ankyrin. All vector constructs were transfected into *E.coli* M15 [pREP4] (Qiagen).

(iii) Mammalian cell vectors (pCEP4). Two versions of ankyrin-coding vectors, pCEP4-Myr$^+$Ank-GFP and pCEP4-Myr0Ank-GFP, were constructed. The N-myristoylated ankyrin-GFP fusion protein expressed by pCEP4-Myr$^+$Ank^{GAG}1D4-GFP was designed to be directed to the plasma membrane, whereas the non-N-myristoylated ankyrin-GFP fusion protein expressed by pCEP4-Myr0-Ank^{GAG}1D4-GFP was designed to localize in the cytoplasm. The DNA encoding the Gag-binders Ank^{GAG}1D4 and control Ank^{A3}2D3 were amplified from their respective pHDiExDsbA-encoding plasmids using two sets of primer with or without the N-myristoylation signal at the 5'end. The gene encoding the green fluorescent protein (GFP) was amplified from pTriEx-GFP [78], using primers of which sequence will be communicated upon request. PCR products encoding Ank^{GAG}1D4 or Ank^{A3}2D3 fused to GFP were recombined by overlapping PCR. The PCR products of the second round were treated with *Kpn* I and *Xho* I (Fermentas) and cloned into corresponding sites of the pCEP4 vector. The sequence of Ank^{GAG}1D4-GFP and Ank^{A3}2D3-GFP, as well as all our other constructs, was verified by standard DNA sequencing.

Production of recombinant H_6MA-CA and H_6CA proteins in baculovirus-infected cells

Sf9 cells were cotransfected with 10 μg each of pBlueBac4.5-H_6MA-CA (or pBlueBac4.5-H_6CA) and Bac-N-Blue™ DNA, using Cellfectin® II reagent, using the conditions recommended by the manufacturer (Invitrogen). The recombinant viruses obtained, BV-H_6MA-CA and BV-H_6CA, were isolated using the blue plaque selection method, and amplified. BV-H_6MA-CA- and BV-H_6CA-infected Sf9 cells were harvested at 48 h postinfection (pi), lysed by freezing and thawing. The cell lysates were clarified by centrifugation at 15, 000 × g for 30 min at 4°C. The presence of recombinant Gag proteins was detected by SDS-PAGE and Western blotting. The nitrocellulose membranes (GE Healthcare Bio-Sciences) were incubated with blocking solution (5% skimmed milk in TBS) for 1 h at RT, and Gag proteins detected using monoclonal anti-His-tag antibody (1:5, 000 dilution in the blocking solution) for 1 h at RT with slow rocking. After washing with TBST (TBS containing 0.05% Tween 20), membranes were incubated with HRP-conjugated goat anti-mouse Ig (1:8, 000 dilution in blocking solution) for 1 h at RT. After two extra washing steps, the Gag proteins were visualized using TMB membrane peroxidase substrate (KPL). His-tagged Gag proteins were purified from clarified Sf9 cell lysates by affinity chromatography on HisTrap column, using ÄKTA prime™ plus (GE Healthcare Bio-Sciences). Protein concentration was determined using the Bradford protein assay (Thermo Fisher Scientific Inc.). Purity of His-tagged Gag proteins was assessed by SDS-PAGE analysis in 15% acrylamide gel and Coomassie blue staining [77].

Phage selection

Microtiter plate (NUNC) was coated with 100 μl H_6MA-CA protein (or αRep-A3 protein) solution at 20 μg/mL in sterile PBS, overnight at 4°C. Purified αRep-A3 protein, produced as described [33], was used as a control for evaluating the quality of our artificial ankyrin library against a properly folded protein target. Plates were washed four times with sterile-filtered TBST. Non-specific binding was prevented by blocking with sterile-filtered blocking buffer (2% BSA in TBST; 200 μl per well) for 1 h at RT with shaking at 150 rpm on an Eppendorf Thermomixer® (Eppendorf). After a washing step with TBST, 100 μl of phage suspension, corresponding to 10^{11} particles, was added per well. After 1 h incubation

at RT with shaking, plates were washed 20 times with TBST and 10 times with TBS. Substrate-bound phages were eluted by postincubation with 100 μl of 0.1 M glycine solution at pH 2.5, for 10 min at RT with shaking, followed with pH neutralization using 12.5 μl 1 M Tris-HCl buffer, pH 8. The eluted phages were mixed with 5 ml of XL-1 Blue cell suspension (OD_{600} 0.6-0.8), and the mixture incubated for 30 min at 37°C. Bacterial cells were centrifuged at 1, 200 × *g* for 10 min at 25°C, pellet resuspended in 1 ml 2X YT broth and plated on LB agar containing ampicillin (100 μg/mL). Bacterial colonies were pooled, and used for phage preparation to perform the next round of phage selection. Individually picked, single colonies of the second and third rounds of selection were screened by phage ELISA.

Expression and purification of soluble ankyrins with Gag-binding activity

M15[pREP4] bacterial cells harboring the pQE30-ankyrin plasmid were grown in 500 ml LB broth supplemented with ampicillin (100 μg/mL), kanamycin (25 μg/mL), and 1% (w/v) glucose, at 37°C with shaking until OD_{600} reached 0.8. Protein expression was induced by addition of 1 mM IPTG, and maintained in culture for 4 hr at 30°C with shaking. Bacteria were pelleted by centrifugation at 1, 200 × *g* for 30 min at 4°C. Pellets were resuspended in lysis buffer and subjected to three cycles of freezing and thawing. Lysis buffer consisted of TBS buffer, pH 7.4, containing 1 μg/mL lysozyme and a cocktail of protease inhibitors (Roche Diagnostics GmbH). Bacterial cell lysates were clarified by centrifugation at 15, 000 × *g* for 30 min at 4°C. The soluble form of Gag-interacting ankyrins was purified from the clarified bacterial lysates by a two-step procedure comprising of affinity chromatography on HisTrap column followed by gel filtration on Sephadex G-75 (GE Healthcare Bio-Sciences). Proteins were analyzed by SDS-PAGE and Coomassie blue staining, or SDS-PAGE and Western blotting, as detailed below.

Biotinylation of soluble ankyrins

Purified Gag-binding ankyrins were chemically biotinylated using the EZ-Link Sulfo-NHS-LC-Biotin kit (ThermoScientific, Rockford, IL). In brief, a solution of purified protein at 100 μM was mixed with a 5-fold molar excess of Sulfo-NHS-Biotin solution in a final volume of 2 ml, and incubated at 25°C for 1 h. Excess reagents and by-products were removed by applying the mixture to a pre-equilibrated Zeba™ Desalt Spin column (ThermoScientific). The column was centrifuged at 1, 000 × g for 2 min, and the biotinylated proteins were recovered in the flow-through fraction. The concentration of biotinylated proteins was determined using the NanoDrop 2000 system (ThermoScientific). The biotinylation efficiency of proteins was qualitatively evaluated using dot-blot analysis. 10 μmol biotinylated proteins was spotted on nitrocellulose membrane, membrane blocked with blocking buffer (5% BSA in TBS), and biotin groups revealed by extravidin-HRP (Sigma) used at dilution 1:5, 000 in blocking buffer (1 h at RT with shaking) and BM Blue POD Substrate (Roche Diagnostics GmbH).

Assessment of ankyrin reactivity towards HIV-1 MA-CA polyprotein

(i) Competitive ELISA. Microtiter plates were coated with 100 μl of purified H_6MA-CA or αRep-A3 (1 μg/mL) diluted in PBS and left overnight at 4°C in a moisture chamber. The coated wells were washed four times with TBST and incubated with 200 μl blocking solution (2% BSA in TBS) for 1 h at RT. After washing, 100 μl biotinylated Gag-binding ankyrin at 10 μM, alone or mixed with an equal molar amount of competitor (non-biotinylated ankyrin or irrelevant ankyrin), was added and incubated for 1 h at RT. Plates were then washed and incubated with extravidin-HRP diluted to 1:5, 000 in blocking solution for 1 h at RT. After washing, 100 μl of TMB substrate was added, and the reaction was blocked by addition of 1 N HCl. OD was measured at 450 nm using a MTP-120 ELISA plate reader (Corona Electric, Ibaraki, Japan).

(ii) Far Western blotting. Lysates of Sf9 cells infected by BV-H_6MA-CA were analyzed by SDS-PAGE and proteins transferred to polyvinylidene fluoride (PVDF) membrane (GE Healthcare Bio-Sciences). Membranes were incubated in blocking buffer (5% BSA in TBS) overnight at 4°C, then postincubated with biotinylated Gag-binding ankyrins at 1 μM for 1 h at RT with gentle rocking. Substrate-bound biotinylated-ankyrins were detected by reaction with extravidin-HRP (diluted to 1:10, 000 in blocking buffer) and TMB membrane peroxidase substrate (KPL).

Mapping of ankyrin binding site on HIV-1 Gag precursor

(i) Specificity assay. The specificity of the Gag-binding ankyrins towards the CA domain was performed by indirect ELISA. Lysates of BV-H_6CA-infected cells were added to nickel pre-coated wells, as described elsewhere [79]. Biotinylated Gag-binding ankyrins were individually reacted with immobilized H_6CA domain for 1 h at 37°C. The binding reaction was monitored by adding extravidin-HRP (dilution 1:5, 000) and TMB substrate. After stopping reaction with 1N HCl, the signals were measured at OD_{450}, as above described.

(ii) Mapping. The ankyrin binding site on the CA domain was determined using Gag amber mutants (Gag*amb*) expressed as recombinant proteins in baculovirus-infected cells [36]. Lysates of Sf9 cells expressing Gag*amb*276 or Gag*amb*241 polyprotein were coated on ELISA plates and reacted with biotinylated Gag-binding ankyrins, as above.

Microcalorimetry analysis of Gag-ankyrin binding parameters

Interaction between proteins was analyzed by isothermal titration calorimetry (ITC), using the MicroCal iTC_{200} isothermal titration microcalorimeter (Microcal), under the conditions described in a previous study [80]. All proteins were diluted in 20 mM phosphate buffer pH7.5, 150 mM NaCl. For each injection, 2 μl of ankyrin solution was added from a computer-controlled 40-μl microsyringe at intervals of 180 s into the protein substrate solution, H_6MA-CA or αRep-A3. A theoretical titration curve was fitted to the experimental data, as previously described [80].

Construction of cell lines stably expressing ankyrins

Aliquots of SupT1 cells (10^6 cells) were electroporated with pCEP4-based vectors encoding GFP-fused ankyrins with (Myr+) or without (Myr0) the myristoylation signal, using the Nucleofector™ (Lonza, Basel, Switzerland) and the Nucleofector™ transfection reagent V (Lonza), according to protocol T-16. Transfected cells were maintained in complete RPMI containing hygromycin B (400 μg/mL). Four cell lines were generated, SupT1/Myr+Ank^{GAG}1D4-GFP, SupT1/Myr0Ank^{GAG}1D4-GFP, SupT1/Myr+Ank^{A3}2D3-GFP and SupT1/Myr0Ank^{A3}2D3-GFP, respectively. The level of expression of Ank^{GAG}1D4-GFP and control Ank^{A3}2D3-GFP proteins was monitored by flow cytometry of the GFP signal, and cellular localization by confocal fluorescence microscopy. For flow cytometry, cells were blocked by incubation with human AB serum on ice for 30 min. They were reacted with 50 μl of purified anti-CD4 mAb MT4-3 [81] at 20 μg/mL in 1% BSA-PBS-NaN_3 on ice for 30 min. At the end of the incubation time, the cells were washed three times with PBS, and incubated with 25 μl PE-conjugated rabbit anti-mouse F(ab')2 (DAKO) on ice for 30 min. Cells were washed, fixed in 1% paraformaldehyde in PBS, and analyzed by flow cytometry.

HIV-1 challenge

To evaluate the effect of Gag-binding ankyrins on the HIV-1 life cycle, SupT1 cells stably expressing the Myr^+ or Myr0 version of the best Gag binder Ank^{GAG}1D4 and irrelevant control Ank^{A3}2D3, were challenged with HIV-1 virions (NL4-3 strain). Triplicate samples of SupT1/Myr+Ank^{GAG}1D4, SupT1/Myr0Ank^{GAG}1D4, SupT1/Myr+Ank^{A3}2D3 and SupT1/Myr0Ank^{A3}2D3 were infected at MOI 10 for 16 h at 37°C. The virus infectivity titer was determined from the genome copy number measured by quantitative RT-PCR (Roche Diagnostics). Cells were then washed three times with serum-free medium, resuspended in 3 mL of fresh medium containing 400 μg/mL hygromycin B and 10% FCS, and seeded into 6-well plates. Cells were harvested at days 5, 7, 9, 11 and 13, and culture supernatants and cell pellets were separately processed for determination of virus progeny yields and viral integration.

HIV-1 production assay

The yields of extracellular virus were evaluated in triplicate samples of culture supernatants of day 11 pi, using a CAp24 ELISA kit (Genscreen ULTRA HIV Ag-Ab, BioRad). Day-11 samples were also assayed for viral genome copy numbers, using COBAS® AmpliPrep/COBAS TaqMan HIV-1 Test (Roche Diagnostics GmbH). Extracellular budding of virions was also monitored by syncytium formation observed in day-11 samples under an inverted microscope (Olympus).

Gag protein assays

Membrane-bound and particulate form of Gag proteins were determined in HIV-1-infected SupT1 cells subjected to cell fractionation. Triplicate cell samples of days 9 and 11 pi were lysed and extracted using the FractionPREP™ Cell Fractionation System (BioVision, Mountain View, CA), following the manufacturer's instructions. The membrane fraction thus isolated was assayed for HIV-1 Gag protein content, using the CAp24 ELISA kit mentioned above, or SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and Western blotting. Proteins were denatured by heating to 100°C for 2 min in SDS-ß-mercaptoethanol-containing sample buffer, electrophoresed in SDS-containing 15%-polyacrylamide gel [77], and then electrically transferred to a polyvinylidene-fluoride (PVDF) membrane. PVDF membranes were blocked with 5% skimmed milk in PBS containing 0.5% Triton X-100, then probed with anti-CAp24 monoclonal antibody (mAb) G18, or anti-MAp17 mAb M48. Both G18 and M48 mAbs were laboratory-made (W. Kasinrerk; unpublished). Blots were developed using HRP-conjugated goat anti-mouse IgG antibody and TMB membrane peroxidase substrate. Extracellular virus-like particles (VLP) released by MLV Gag-Pol-expressing GP2-293 cells were recovered by ultracentrifugation of the cell culture medium [71,82], and VLP production estimated by SDS-PAGE of VLP pellets and Western blot analysis using rabbit polyclonal antibody to MLV-GagCAp30 protein (antibodies-online Inc., Atlanta, GA). Intracellular content of MLV Gag proteins was analyzed in the same manner, using the whole cell lysate.

HIV-1 integration assay

The number of viral genome copies integrated into the host DNA of control SupT1 or SupT1 expressing Gag-specific (Ank^{GAG}1D4) or irrelevant ankyrin (Ank^{A3}2D3) was determined using a conventional *Alu-gag* qPCR assay [83,84] with some modifications. The first-round of PCR was performed on cellular DNA, extracted using the High Pure PCR Template Preparation Kit (Roche, Mannheim, Germany). The sequences of the first round amplification

primers were: *Alu* forward, 5'-GCC TCC CAA AGT GCT GGG ATT ACA G-3' [84], and HIV-1 *gag* reverse, 5'-GTT CCT GCT ATG TCA CTT CC -3' [83]. The first round reactions were carried out in a volume of 25 μl containing 2.5× master mix (5 PRIME, Gaithersburg, MD), using a standard protocol. The second-round of real-time quantitative PCR of RU5 was performed using 10 μl of diluted (1:8) first-round amplicons. The sequences of primers were: R_FWD, 5'-TTA AGC CTC AAT AAA GCT TGC C-3'; and U5 _REV, 5'-GTT CGG GCG CCA CTG CTA GA-3', and the sequence of RU5 molecular beacon probe was 5'-FAM-CCA GAG TCA CAC AAC AGA CGG GCA CA-BBQ-3' [85]. The reactions were carried out in a final volume of 25 μl containing 2× DyNAmo probe qPCR master mix (Finnzymes, Espoo, Finland), 400 nM RU5 (R_FWD) primer, 400 nM RU5 (U5 _REV) primer, and 140 nM RU5 molecular beacon probe. The reactions were performed in a MJ Mini Thermal Cycler and MiniOpticon Real-Time PCR System (BioRad) with the following thermal program: 20-sec hot start at 95°C followed by 50 cycles of denaturation at 95°C for 3 sec and annealing and extension at 63°C for 30 sec. A primer-probe set, designed to quantify the copy number of the cellular gene glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*), was used to quantify the amount of DNA in each qPCR assay. The *GAPDH* primer sequences were as follows: GAPDH_FWD, 5'-GAA GGT GAA GGT CGG AGT C-3'; and *GAPDH*_REV, 5'-GAA GAT GGT GAT GGG ATT TC-3'. The GAPDHTM molecular beacon probe was designed to contain the following sequence: 5'-FAM-CAA GCT TCC CGT TCT CAG CCT-BBQ-3'. The reactions were carried out as described above. Results were expressed as *C*ts, i.e. the number of cycles (*C*ts) required for the fluorescence signal to cross the threshold value (cycle threshold). Control experiments for the inhibition of provirus integration in HIV-1-infected cells were carried out as follows. The HIV-1 integrase inhibitor Raltegravir™ (RAL; Merck Sharp & Dohme) was added at 0, 1, 10, and 100 nM, respectively, to SupT1 cell culture medium 24 h prior to HIV-1 infection [37]. SupT1 cells were then infected with HIV-1_{NL4-3} at MOI 10, and virus inoculum removed after 16 h. Cells were washed three times with prewarmed, serum-free medium and resuspended in growth medium containing RAL at the above-mentioned concentrations. Cells were divided (1:2) every second day to maintain a cell density of approximately 10^6 cells per mL, and harvested at day-7 pi for *Alu-gag* and *GAPDH* qPCR assays. Cell viability in each sample was assayed using PrestoBlue Cell Viability Reagent (Invitrogen).

HIV-1 gag gene sequencing

Mock-infected or HIV-1-infected SupT1 cells (MOI 10) expressing Myr+Ank^{GAG}1D4 or Myr+Ank^{A3}2D3, were harvested at day-13 pi. A four-step protocol was then applied. (i) Total viral RNA was isolated, using the High Pure Viral RNA kit (Roche Applied Science, Roche, Mannheim, Germany). (ii) The viral RNA thus obtained was reverse transcribed into cDNA, using the Transcriptor High Fidelity cDNA Synthesis Kit with anchored-oligo $(dT)_{18}$ primer (Roche). (iii) The single-stranded cDNA was then amplified, using a proof-reading PCR protocol (Phusion™ High-Fidelity DNA Polymerase; Finnzymes, Espoo, Finland) and the following pair of p24-specific gene primers: FWD_p24 *Nhe* I, 5'-GAGGAGGAGGT GCTAGCCCTATAGTGCAGAACCTCCAG-3' and REV_p24 *Kpn* I, 5'-GAGGAGGAGCTGGTACCTTA-CAAAACTCTTGCTTTATGGCC-3'. (iv) The PCR products were purified using the GeneJET™ PCR purification kit (Fermentas International), and sequenced using standard DNA sequencing method (1st BASE Pte Ltd, Singapore).

Confocal microscopy

Aliquots of HIV-1-infected SupT1 cells (1×10^6 cells, MOI 10) were harvested on day 11 pi, washed with PBS, fixed in 4% formaldehyde in PBS, and permeabilized with 0.2% Triton X-100. After blocking with 10% human AB serum for 30 min at room temperature, cells were incubated with G18 anti-CAp24 mAb at 37°C for 1 hr. After washing twice with PBS containing 1% BSA and NaN_3, cells were incubated with PE-conjugated polyclonal rabbit anti-mouse IgG $F(ab')_2$ (Dako, Denmark), and nuclei counterstained with DAPI. Images were acquired using FluoView laser scanning confocal microscope (Olympus, FV1000; Olympus Optical, Japan).

Conflict of interests

The authors declare that they have no competing interests.

Additional material

Additional file 1: Control, Raltegravir-mediated inhibition of HIV-1 integration in SupT-1 cells. Aliquots of HIV-1-infected SupT1 cells were pretreated with Raltegravir at 0, 1, 10, and 100 nM, respectively, for 24 h prior to HIV-1 infection (MOI 10). The drug was maintained at the indicated concentrations for 7 days, and the cells harvested at day 7 pi. The level of HIV-1 integration in SupT1 cell lines was evaluated by quantitative PCR amplification of host cell DNA extracts, using primers specific to *Alu-gag* junctions and to cellular *GAPDH* gene as the internal control. **(A)**, *Alu-gag* qPCR obtained with the different cell samples. The qPCR assays were performed in triplicate. The colours of the curves correspond to the different Raltegravir molarities, as indicated in (B). **(B)**, Comparison of the mean *C*ts values (m ± SD) for *Alu-gag* and *GAPDH* qPCR. ND, not detectable (below the detection threshold). **(C)**, Cell viability, determined by the PrestoBlue Cell Viability Reagent, and expressed as the percentage of control, untreated cells.

Acknowledgements

The work in Thailand was supported by the National Research University project under the Thailand's Office of the Commission on Higher Education.

The authors also express their gratitude to the Thailand Research Fund, the Research Chair Grant of National Sciences and the Technology Development Agency and the Franco-Thai collaboration program. SN was supported by CHE-PhD-THAI scholarship from the Commission on Higher Education, Ministry of Education, Thailand, and the AMS-IRD URI-174 PHPT Franco-Thai Cooperation Program for High Education and Research. WK and SS were supported by The Royal Golden Jubilee PhD Program.
The work in France was supported by the Centre National de la Recherche Scientifique (CNRS), the French National Agency for Research on AIDS and Viral Hepatitis (Inserm-ANRS contract N° 11344/2011-2013), the University of Lyon and the University of Paris-Sud. SSH is a scientist of the French Institute of Health and Medical Research (Institut National de la Santé et de la Recherche Médicale, Inserm) and the recipient of a Contrat d'Interface Inserm-Hospices Civils de Lyon (CIF-2008-2013). We deeply thank Supattara Suwanpairoj, Monique Martingay and Sylvie Farget for their precious technical help during the completion of this study.

Author details
[1]Division of Clinical Immunology, Department of Medical Technology, Faculty of Associated Medical Sciences, Chiang Mai University, Chiang Mai, Thailand 50200. [2]Biomedical Technology Research Unit, National Center for GeneticEngineering and Biotechnology, National Science and Technology Development Agency at the Faculty of Associated Medical Sciences, Chiang Mai University, Chiang Mai 50200, Thailand. [3]Institut de Biochimie et de Biophysique Moléculaire et Cellulaire (IBBMC) UMR-8619, Université de Paris-Sud et CNRS, Orsay Cedex 91405, France. [4]University Lyon 1, 50, avenue Tony Garnier, 69366 Lyon Cedex 07, France. [5]INRA UMR-754, Retrovirus and Comparative Pathology, 50, avenue Tony Garnier, 69366 Lyon Cedex 07, France.

Authors' contributions
Conceived and designed the experiments: SN AU SSH PB PM CT. Performed the experiments: SN AU MVL WK SS SSH PM. Analyzed the data: SN MVL WK SSH PM CT. Wrote the paper: SN SSH PB PM CT. All authors approved the submitted manuscript.

Received: 15 August 2011 Accepted: 20 February 2012
Published: 20 February 2012

References

1. Arora DR, Gautam V, Gill PS, Mishra N: **Recent advances in antiretroviral therapy in HIV infection.** *J Indian Med Assoc* 2010, **108**:29-34.
2. Check E: **Pioneering HIV treatment would use interference and gene therapy.** *Nature* 2005, **437**:601.
3. Luque F, Oya R, Macias D, Saniger L: **Gene therapy for HIV-1 infection: are lethal genes a valuable tool?** *Cell Mol Biol* 2005, **51**:93-101.
4. Trkola A, Kuster H, Rusert P, Joos B, Fischer M, Leemann C, Manrique A, Huber M, Rehr M, Oxenius A, Weber R, Stiegler G, Vcelar B, Katinger H, Aceto L, Günthard HF: **Delay of HIV-1 rebound after cessation of antiretroviral therapy through passive transfer of human neutralizing antibodies.** *Nat Med* 2005, **11**:615-622.
5. von Laer D, Hasselmann S, Hasselmann K: **Gene therapy for HIV infection: what does it need to make it work?** *J Gene Med* 2006, **8**:658-667.
6. Gonzalo T, Clemente MI, Chonco L, Weber ND, Diaz L, Serramia MJ, Gras R, Ortega P, de la Mata FJ, Gomez R, Lopez-Fernández LA, Muñoz-Fernández MA, Jiménez JL: **Gene therapy in HIV-infected cells to decrease viral impact by using an alternative delivery method.** *Chem Med Chem* 2010, **5**:921-929.
7. Joseph A, Zheng JH, Chen K, Dutta M, Chen C, Stiegler G, Kunert R, Follenzi A, Goldstein H: **Inhibition of in vivo HIV infection in humanized mice by gene therapy of human hematopoietic stem cells with a lentiviral vector encoding a broadly neutralizing anti-HIV antibody.** *J Virol* 2010, **84**:6645-6653.
8. Kitchen SG, Shimizu S, An DS: **Stem cell-based anti-HIV gene therapy.** *Virology* 2011, **411**:260-272.
9. Berkhout B: **Towards a durable anti-HIV gene therapy based on RNA interference.** *Ann NY Acad Sci* 2009, **1175**:3-14.
10. Rossi JJ, June CH, Kohn DB: **Genetic therapies against HIV.** *Nat Biotech* 2007, **25**:1444-1454.
11. Wheeler YY, Chen SY, Sane DC: **Intrabody and intrakine strategies for molecular therapy.** *Mol Ther 8:355-66* 2003, **8**:355-366.
12. Binz HK, Amstutz P, Kohl A, Stumpp MT, Briand C, Forrer P, Grutter MG, Pluckthun A: **High-affinity binders selected from designed ankyrin repeat protein libraries.** *Nat Biotechnol* 2004, **22**:575-582.
13. Binz HK, Amstutz P, Pluckthun A: **Engineering novel binding proteins from nonimmunoglobulin domains.** *Nat Biotech* 2005, **23**:1257-1268.
14. Binz HK, Pluckthun A: **Engineered proteins as specific binding reagents.** *Curr Opin Biotechnol* 2005, **16**:459-469.
15. Guglielmi L, Martineau P: **Expression of single-chain Fv fragments in E. coli cytoplasm.** *Methods Mol Biol* 2009, **562**:215-224.
16. Zhao JX, Yang L, Gu ZN, Chen HQ, Tian FW, Chen YQ, Zhang H, Chen W: **Stabilization of the single-chain fragment variable by an interdomain disulfide bond and its effect on antibody affinity.** *Int J Mol Sci* 2010, **12**:1-11.
17. Hey T, Fiedler E, Rudolph R, Fiedler M: **Artificial, non-antibody binding proteins for pharmaceutical and industrial applications.** *Trends Biotechnol* 2005, **23**:514-522.
18. Lofblom J, Frejd FY, Stahl S: **Non-immunoglobulin based protein scaffolds.** *Curr Opin Biotechnol* 2011, **22**:1-6.
19. Nygren PA, Uhlen M: **Scaffolds for engineering novel binding sites in proteins.** *Curr Opin Struct Biol* 1997, **7**:463-469.
20. Li J, Mahajan A, Tsai MD: **Ankyrin repeat: a unique motif mediating protein-protein interactions.** *Biochemistry* 2006, **45**:15168-15178.
21. Mosavi LK, Cammett TJ, Desrosiers DC, Peng ZY: **The ankyrin repeat as molecular architecture for protein recognition.** *Protein Sci* 2004, **13**:1435-14448.
22. Andrade MA, Perez-Iratxeta C, Ponti CP: **Protein repeats: structures, functions, and evolution.** *J Struct Biol* 2001, **134**:117-131.
23. Binz HK, Stumpp MT, Forrer P, Amstutz P, Pluckthun A: **Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins.** *J Mol Biol* 2003, **332**:489-503.
24. Boersma YL, Plückthun A: **DARPins and other repeat protein scaffolds: advances in engineering and applications.** *Curr Opin Biotechnol* 2011, **22**:849-857.
25. Sedgwick SG, Smerdon SJ: **The ankyrin repeat: a diversity of interactions on a common structural framework.** *Trends Biochem Sci* 1999, **24**:311-316.
26. Stumpp MT, Amstutz P: **DARPins: a true alternative to antibodies.** *Curr Opin Drug Discov Devel* 2007, **10**:153-159.
27. Stumpp MT, Binz HK, Amstutz P: **DARPins: a new generation of protein therapeutics.** *Drug Discov Today* 2008, **13**:695-701.
28. Schweizer A, Rusert P, Berlinger L, Ruprecht CR, Mann A, Corthésy S, Turville SG, Aravantinou M, Fischer M, Robbiani M, Amstutz P, Trkola A: **CD4-specific designed ankyrin repeat proteins are novel potent HIV entry inhibitors with unique characteristics.** *PLoS Pathog* 2008, **4**: e1000109.
29. Schweizer A, Roschitzki-Voser H, Amstutz P, Briand C, Gulotti-Georgieva M, Prenosil E, Binz HK, Capitani G, Baici A, Pluckthun A, Grutter MG: **Inhibition of caspase-2 by a designed ankyrin repeat protein: specificity, structure, and inhibition mechanism.** *Structure* 2007, **15**:625-636.
30. Zahnd C, Wyler E, Schwenk JM, Steiner D, Lawrence MC, McKern NM, Pecorari F, Ward CW, Joos TO, Plückthun A: **A designed ankyrin repeat protein evolved to picomolar affinity to Her2.** *J Mol Biol* 2007, **369**:1015-1028.
31. Münch RC, Mühlebach MD, Schaser T, Kneissl S, Jost C, Plückthun A, Cichutek K, Buchholz CJ: **DARPins: an efficient targeting domain for lentiviral vectors.** *Mol Ther* 2011, **19**:686-693.
32. Huber T, Steiner D, Rothlisberger D, Pluckthun A: **In vitro selection and characterization of DARPins and Fab fragments for the co-crystallization of membrane proteins: the Na+-citrate symporter CitS as an example.** *J Struct Biol* 2007, **159**:206-221.
33. Urvoas A, Guellouz A, Valerio-Lepiniec M, Graille M, Durand D, Desravines DC, van Tilbeurgh H, Desmadril M, Minard P: **Design, production and molecular structure of a new family of artificial alpha-helicoidal repeat proteins (alphaRep) based on thermostable HEAT-like repeats.** *J Mol Biol* 2010, **404**:307-327.
34. Hong SS, Boulanger P: **Protein ligands of human Adenovirus type 2 outer capsid identified by biopanning of a phage-displayed peptide library on separate domains of WT and mutant penton capsomers.** *EMBO J* 1995, **14**:4714-4727.
35. Hong SS, Karayan L, Tournier J, Curiel DT, Boulanger PA: **Adenovirus type 5 fiber knob binds to MHC class I alpha2 domain at the surface of human epithelial and B lymphoblastoid cells.** *EMBO J* 1997, **16**:2294-2306.

36. Carrière C, Gay B, Chazal N, Morin N, Boulanger P: **Sequence requirement for encapsidation of deletion mutants and chimeras of human immunodeficiency virus type 1 Gag precursor into retrovirus-like particles.** *J Virol* 1995, **69**:2366-2377.
37. Delelis O, Malet I, Na L, Tchertanov L, Calvez V, Marcelin AG, Subra F, Deprez E, Mouscadet JF: **The G140S mutation in HIV integrases from raltegravir-resistant patients rescues catalytic defect due to the resistance Q148H mutation.** *Nucleic Acids Res* 2009, **37**:1193-1201.
38. Vu HN, Ramsey JD, Pack DW: **Engineering of a stable retroviral gene delivery vector by directed evolution.** *Mol Ther* 2008, **16**:308-314.
39. Rakotobe D, Tardy J-C, Andre P, Hong SS, Darlix J-L, Boulanger P: **Human Polycomb group EED protein negatively affects HIV-1 assembly and release.** *Retrovirology* 2007, **4**:37.
40. Wensing AM, Boucher CA: **Worldwide transmission of drug-resistant HIV.** *AIDS Rev* 2003, **5**:140-155.
41. Adamson CS, Freed EO: **Novel approaches to inhibiting HIV-1 replication.** *Antiviral Res* 2010, **85**:119-141.
42. Greene WC, Debyser Z, Ikeda Y, Freed EO, Stephens E, Yonemoto W, Buckheit RW, Esté JA, Cihlar T: **Novel targets for HIV therapy.** *Antiviral Res* 2008, **80**:251-265.
43. Muriaux D, Darlix J-L, Cimarelli A: **Targeting the assembly of the human immunodeficiency virus type 1.** *Current Pharmaceutical Design* 2004, **10**:3725-3739.
44. Sticht J, Humbert M, Findlow S, Bodem J, Muller B, Dietrich U, Werner J, Kräusslich H-G: **A peptide inhibitor of HIV-1 assembly in vitro.** *Nat Struct Mol Biol* 2005, **12**:671-677.
45. Tang C, Loeliger E, Kinde I, Kyere S, Mayo K, Barklis E, Sun Y, Huang M, Summers MF: **Antiviral inhibition of the HIV-1 capsid protein.** *J Mol Biol* 2003, **327**:1013-1020.
46. Ternois F, Sticht J, Duquerroy S, Kräusslich H-G: **The HIV-1 capsid potein C-terminal domain in complex with a virus assembly inhibitor.** *Nat Struct Mol Biol* 2005, **12**:678-682.
47. Waheed AA, Freed EO: **Peptide inhibitors of HIV-1 egress.** *ACS Chem Biol* 2008, **3**:745-747.
48. Waheed AA, Ono A, Freed EO: **Methods for the study of HIV-1 assembly.** *Methods Mol Biol* 2009, **485**:163-184.
49. Wainberg MA: **Perspectives on antiviral drug development.** *Antiviral Res* 2009, **81**:1-5.
50. Woods MW, Kelly JN, Hattlmann CJ, Tong JGK, Xu LS, Coleman MD, Quest GR, Smiley JR, Bar SD: **Human HERC5 restricts an early stage of HIV-1 assembly by a mechanism correlating with the ISGylation of Gag.** *Retrovirology* 2011, **8**:95.
51. Poole E, Strappe P, Mok HP, Hicks R, Lever AM: **HIV-1 Gag-RNA interaction occurs at a perinuclear/centrosomal site; analysis by confocal microscopy and FRET.** *Traffic* 2005, **6**:741-755.
52. Anderson EC, Lever AM: **Human immunodeficiency virus type 1 Gag polyprotein modulates its own translation.** *J Virol* 2006, **80**:10478-10486.
53. Ricci EP, Soto Rifo R, Herbreteau CH, Decimo D, Ohlmann T: **Lentiviral RNAs can use different mechanisms for translation initiation.** *Biochem Soc Trans* 2008, **36**:690-693.
54. Alfadhli A, Dhenub TC, Still A, Barklis E: **Analysis of human immunodeficiency virus type 1 Gag dimerization-induced assembly.** *J Virol* 2005, **79**:14498-14506.
55. Jouvenet N, Simon SM, Bieniasz PD: **Imaging the interaction of HIV-1 genomes and Gag during assembly of individual viral particles.** *Proc Natl Acad Sci USA* **106**:19114-19119.
56. Kutluay SB, Bieniasz PD: **Analysis of the initiating events in HIV-1 particle assembly and genome packaging.** *PLoS Pathogens* 2010, **6**: e1001200.
57. Lindwasser OW, Resh MD: **Myristoylation as a target for inhibiting HIV assembly: unsaturated fatty acids block viral budding.** *Proc Natl Acad Sci USA* 2002, **99**:13037-13042.
58. Ganser-Pornillos BK, Cheng A, Yeager M: **Structure of full-length HIV-1 CA: a model for the mature capsid lattice.** *Cell* 2007, **131**:70-79.
59. Ganser-Pornillos BK, von Schwedler UK, Stray KM, Aiken C, Sundquist WI: **Assembly properties of the human immunodeficiency virus type 1 CA protein.** *J Virol* 2004, **78**:2545-2552.
60. Ganser-Pornillos BK, Yeager M, Sundquist WI: **The structural biology of HIV assembly.** *Curr Opin Struct Biol* 2008, **18**:203-217.
61. Colgan J, Yuan HE, Franke EK, Luban J: **Binding of the human immunodeficiency virus type 1 Gag polyprotein to cyclophilin A is mediated by the central region of capsid and requires Gag dimerization.** *J Virol* 1996, **70**:4299-4310.
62. Dorfman T, Weimann A, Borsetti A, Walsh CT, Göttlinger HG: **Active-site residues of cyclophilin A are crucial for its incorporation into human immunodeficiency virus type 1 virions.** *J Virol* 1997, **71**:7110-7113.
63. Zhao Y, Chen Y, Schutkowski M, Fischer G, Ke H: **Cyclophilin A complexed with a fragment of HIV-1 gag protein: insights into HIV-1 infectious activity.** *Structure* 1997, **5**:139-146.
64. Briant L, Gay B, Devaux C, Chazal N: **HIV-1 assembly, release and maturation.** *WJA* 2011, **1**:111-130.
65. Cimarelli A, Darlix J-L: **Assembling the human immunodeficiency virus type 1.** *Cell Mol Life Sci* 2002, **59**:1166-1184.
66. Darlix JL, Cristofari G, Rau M, Péchoux C, Berthoux L, Roques B: **Nucleocapsid protein of human immunodeficiency virus as a model protein with chaperoning functions and as a target for antiviral drugs.** *Adv Pharmacol* 2000, **48**:345-372.
67. Ma YM, Vogt VM: **Rous sarcoma virus Gag protein-oligonucleotide interaction suggests a critical role for protein dimer formation in assembly.** *J Virol* 2002, , **76:** 5452-5462.
68. Ma YM, Vogt VM: **Nucleic acid binding-induced Gag dimerization in the assembly of Rous Sarcoma Virus particles in vitro.** *J Virol* 2004, **78**:52-60.
69. Gamble TR, Yoo S, Vajdos FF, von Schwedler UK, Worthylake DK, Wang H, McCutcheon JP, Sundquist WI, Hill CP: **Structure of the carboxyl-terminal dimerization domain of the HIV-1 capsid protein.** *Science* 1997, **278**:849-853.
70. Ganser BK, Li S, Klishko VY, Finch JT, Sundquist WI: **Assembly and analysis of conical models for the HIV-1 core.** *Science* 1999, **283**:80-83.
71. DaFonseca S, Blommaert A, Coric P, Hong SS, Bouaziz S, Boulanger P: **The 3-O-(3', 3'-dimethylsuccinyl) derivative of betulinic acid (DSB) inhibits the assembly of virus-like particles in HIV-1 Gag precursor-expressing cells.** *Antiviral Ther* 2007, **12**:1185-1203.
72. Hong SS, Boulanger P: **Self-assembly-defective dominant mutants of HIV-1 Gag phenotypically expressed in baculovirus-infected cells.** *J Virol* 1993, **67**:2787-2798.
73. Huvent I, Hong SS, Fournier C, Gay B, Tournier J, Carriere C, Courcoul M, Vigne R, Spire B, Boulanger P: **Interaction and co-encapsidation of HIV-1 Vif and Gag recombinant proteins.** *J Gen Virol* 1998, **79**:1069-1081.
74. Royer M, Hong SS, Gay B, Cerutti M, Boulanger P: **Expression and extracellular release of human immunodeficiency virus type 1 Gag precursors by recombinant baculovirus-infected cells.** *J Virol* 1992, **66**:3230-3235.
75. Nangola S, Minard P, Tayapiwatana C: **Appraisal of translocation pathways for displaying ankyrin repeat protein on phage particles.** *Protein Expr Purif* 2010, **74**:156-161.
76. Steiner D, Forrer P, Pluckthun A: **Efficient selection of DARPins with sub-nanomolar affinities using SRP phage display.** *J Mol Biol* 2008, **382**:1211-1227.
77. Kitidee K, Nangola S, Gonzalez G, Boulanger P, Tayapiwatana C, Hong SS: **Baculovirus display of single chain antibody (scFv) using a novel signal peptide.** *BMC Biotechnol* 2010, **10**:80.
78. Sakkhachornphop S, Jiranusornkul S, Kodchakorn K, Nangola S, Sirisanthana T, Tayapiwatana C: **Designed zinc finger protein interacting with the HIV-1 integrase recognition sequence at 2-LTR-circle junctions.** *Protein Sci* 2009, **18**:2219-2230.
79. Cressey R, Pimpa S, Chewaskulyong B, Lertprasertsuke N, Saeteng S, Tayapiwatana C, Kasinrerk W: **Simplified approaches for the development of an ELISA to detect circulating autoantibodies to p53 in cancer patients.** *BMC Biotechnology* 2008, **8**:16.
80. Nicaise M, Valerio-Lepiniec M, Minard P, Desmadril M: **Affinity transfer by CDR grafting on a nonimmunoglobulin scaffold.** *Protein Sci* 2004, **13**:1882-1891.
81. Pata S, Tayapiwatana C, Kasinrerk W: **Three different immunogen preparation strategies for production of CD4 monoclonal antibodies.** *Hybridoma* 2009, **28**:159-165.
82. Gonzalez G, DaFonseca S, Errazuriz E, Coric P, Souquet F, Turcaud S, Boulanger P, Bouaziz S, Hong SS: **Characterization of a novel type of HIV-1 particle assembly inhibitor using a quantitative luciferase-Vpr packaging-based assay.** *PLoS One* 2011, **6**:e27234.
83. Agosto LM, Yu JJ, Dai J, Kaletsky R, Monie D, O'Doherty U: **HIV-1 integrates into resting CD4+ T cells even at low inoculums as demonstrated with an improved assay for HIV-1 integration.** *Virology* 2007, **368**:60-72.

84. O'Doherty U, Swiggard WJ, Jeyakumar D, McGain D, Malim MH: **A sensitive, quantitative assay for human immunodeficiency virus type 1 integration.** *J Virol* 2002, **76**:10942-10950.
85. Liszewski MK, Yu JJ, O'Doherty U: **Detecting HIV-1 integration by repetitive-sampling Alu-gag PCR.** *Methods* 2009, **47**:254-260.

doi:10.1186/1742-4690-9-17
Cite this article as: Nangola *et al.*: Antiviral activity of recombinant ankyrin targeted to the capsid domain of HIV-1 Gag polyprotein. *Retrovirology* 2012 **9**:17.

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

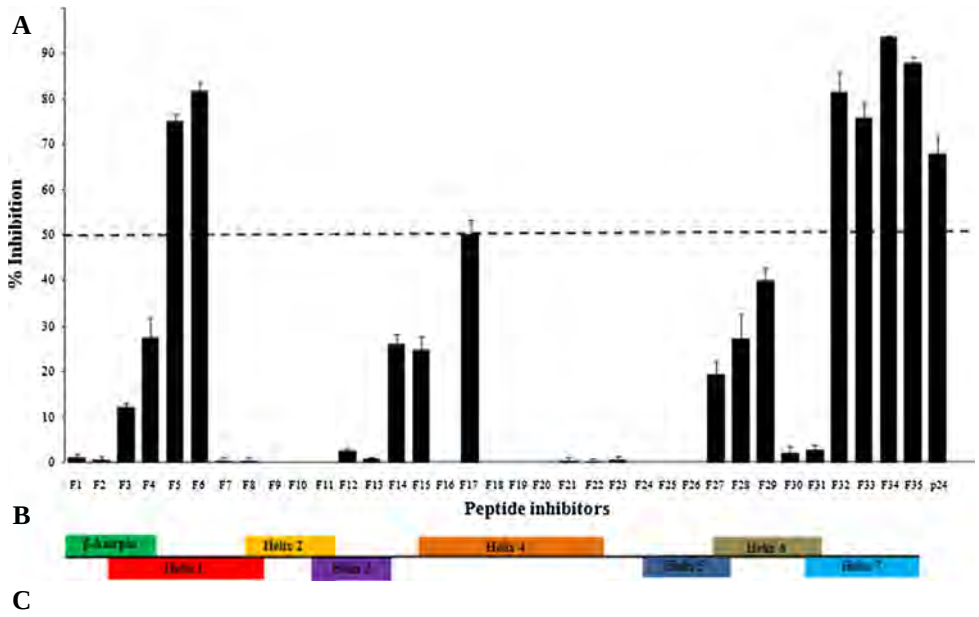

Fig. 1 Ankyrinotope mapping in HIV-1 NTD^{CA}. **A** Data from competition ELISA. Binding of the viral capsid protein CA to Ank^{GAG}1D4 was analyzed in the presence of competing overlapping pentadecapeptides covering the whole N-terminal domain of the CA (NTD^{CA}). Positive control sample for homologous 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

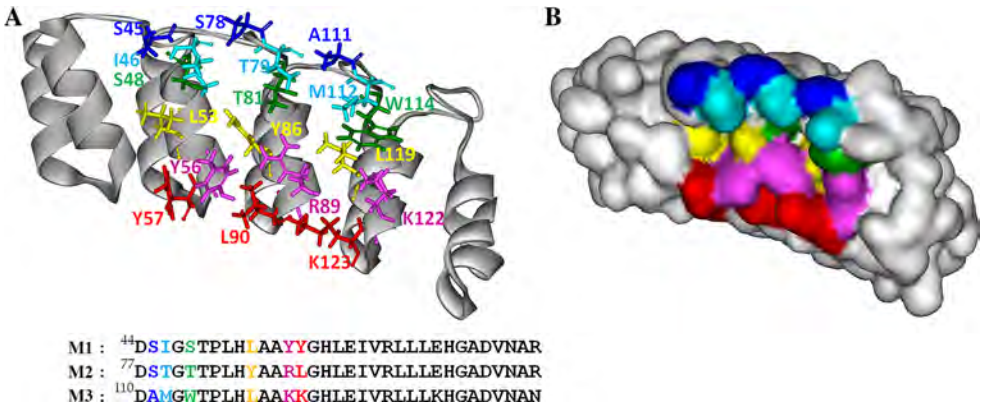

Fig. 2 Crystal structure of the ankyrin Ank^{GAG}1D4. **A** Ribbon model of Ank^{GAG}1D4, and amino acids involved in its CA-binding surface. *Top* the backbone of constant residues is shown by the *grey ribbon trace*, whereas the side chains of variable residues are presented as *coloured sticks*. *Bottom* 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**)

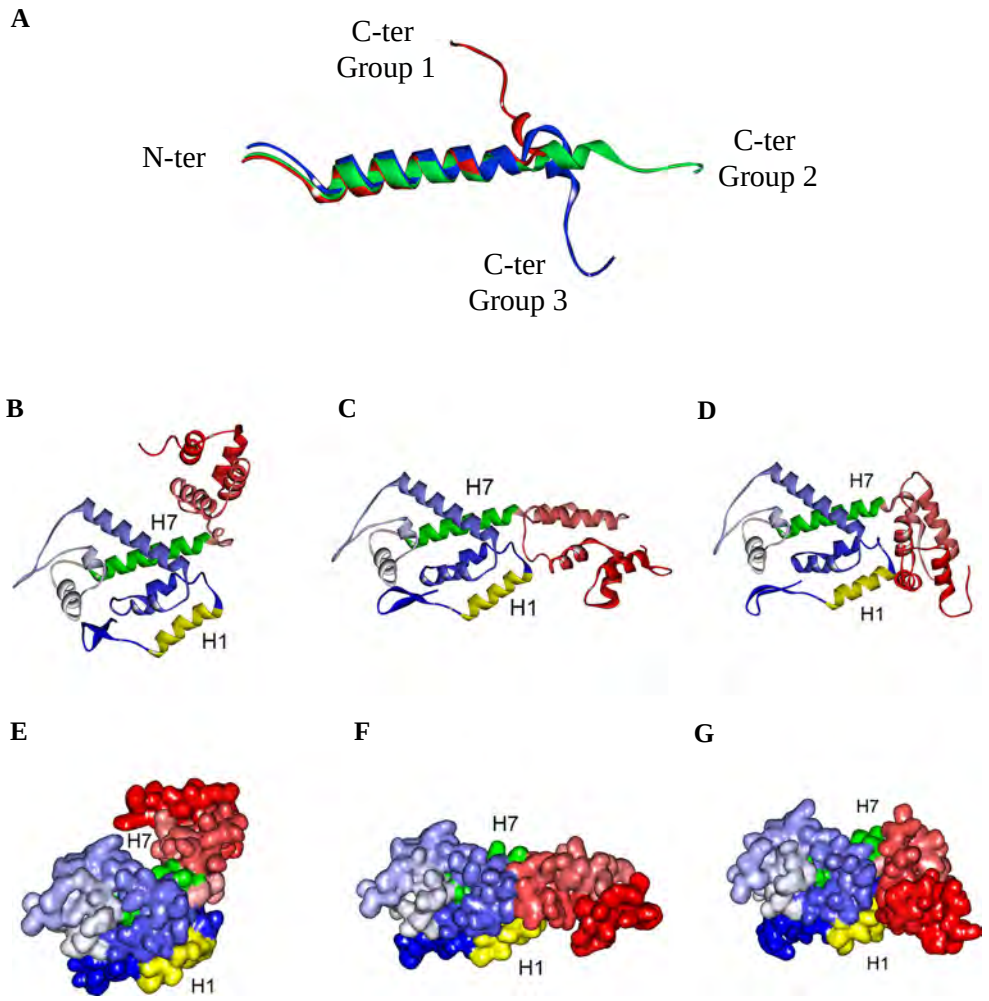

Fig. 3 Structural models of the HIV-1 CA protein. **A** Possible spatial orientations of the YSPTS linker forming the hinge between the N- and C-terminal domains of the CA protein. Each colour represents the candidate selected for each group of models: *red* group 1, *green* group 2, *blue* group 3. 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.

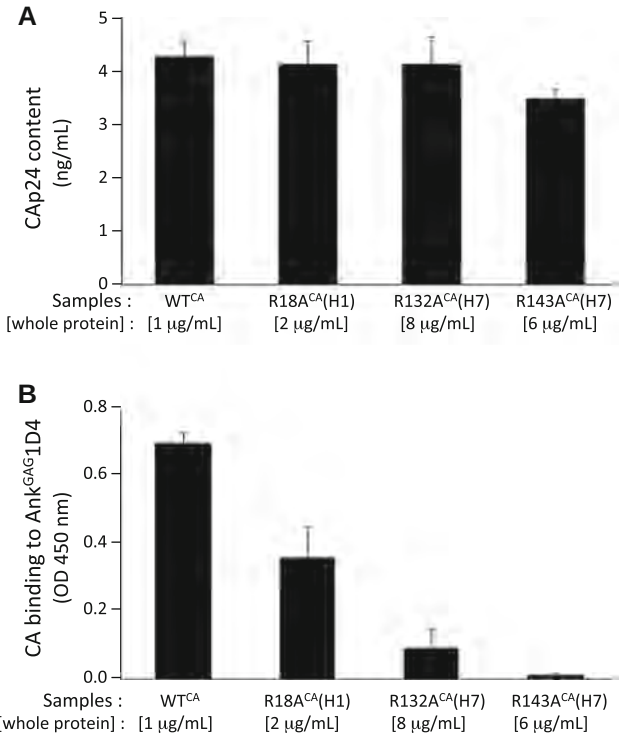

Fig. 4 Influence of R-to-A mutations in the helices 1 and 7 of NTD^{CA} on the interaction of the CA protein with Ank^{GAG}1D4. **A** *Normalization of the CA protein*. The normalization of the CA protein amount in each sample immobilized on ELISA plate was performed using a sandwich CAp24 detection 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

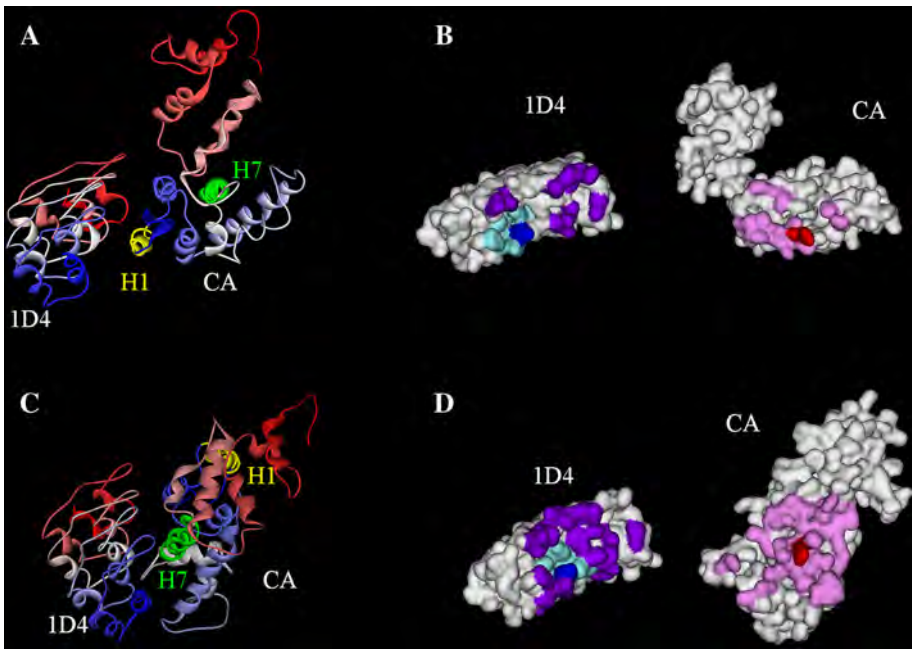

Fig. 5 Structural models of Ank^{GAG}1D4-CA complexes. **A**, **B** Structure of the p269 complex (m1H1). **C**, **D** Structure of the p29 complex (m20H7). The proteins are depicted as ribbon style (**A**, **C**) and surface style (**B**, **D**). In **A**, **C**, the *blue and red gradient colours* 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

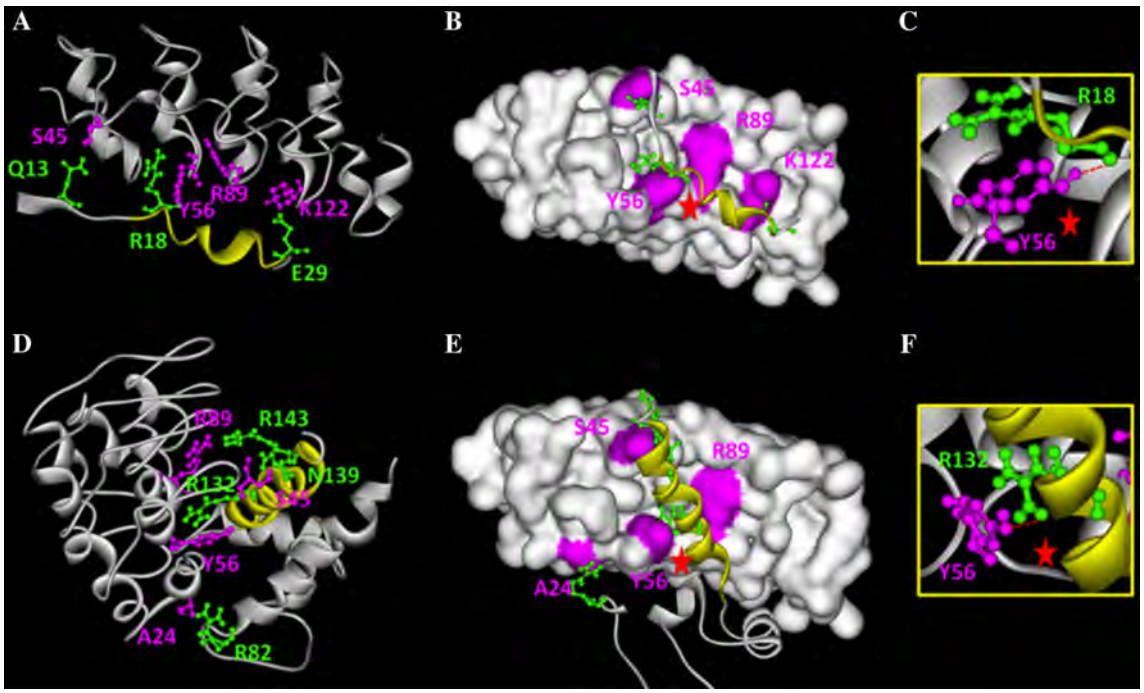

Fig. 6 Amino-acid interactions in the p269 and p29 models of Ank^{GAG}1D4-CA complexes. The Ank^{GAG}1D4 molecules complexed with the helix 1 (m1H1; panels **A**, **B**, and **C**) or helix 7 of NTD^{CA} (m20h7; panels **D**, **E**, and **F**) are shown in different presentations. In the ribbon style 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