

Fluorescent protein tagged hepatitis B virus capsid protein with long glycine-serine linker that supports nucleocapsid formation

Jiang-Yan Chen^{a, b, c, #}, Chun-yang Gan^{a, #}, Xue-fei Cai^{a, #}, Wen-lu Zhang^a,
Quan-xin Long^a, Xia-fei Wei^a, Yuan Hu^a, Ni Tang^a, Juan Chen^a, Haitao Guo^d,
Ai-long Huang^{a,c, *}, Jie-li Hu^{a, c,*}

Key Laboratory of Molecular Biology on Infectious Diseases, Ministry of Education, Department of Infectious Diseases, the Second Affiliated Hospital of Chongqing Medical University, Chongqing, China ^a; Department of Clinical Laboratory, The First Affiliated Hospital of Shantou University Medical College, Shantou, Guangdong, China ^b; Collaborative Innovation Center for diagnosis and treatment of infectious diseases (CCID), Hangzhou, China ^c; Department of Microbiology and Immunology, Indiana University School of Medicine ^d.

Running Head: Hepatitis B virus core protein engineering

* Address correspondence to Jie-li Hu, hujieli1977@163.com and Ai-long Huang, ahuang1964@163.com.

JY. C, CY. G and XF. C contributed equally to this work.

Highlights

- A solution was developed to construct fusion proteins of Hepatitis B virus core protein in which the functions of the fused domains were kept.
- The solution contains two principles: (1) fuse heterogeneous proteins at the N terminus of HBc; (2) use long Glycine-serine linkers between the two domains.
- EGFP-HBc and RFP-HBc constructed following the principles supported the formation of nucleocapsid in HepG2 cells.

Abstract

Fusion core proteins of Hepatitis B virus can be used to study core protein functions or capsid trafficking. A problem in constructing fusion core proteins is functional impairment of the individual domains in these fusion proteins, might due to structural interference. We reported a method to construct fusion proteins of Hepatitis B virus core protein (HBc) in which the functions of fused domains were partially kept. This method follows two principles: (1) fuse heterogeneous proteins at the N terminus of HBc; (2) use long Glycine-serine linkers between the two domains. Using EGFP and RFP as examples, we showed that long flexible G₄S linkers can effectively separate the two domains in function. Among these fusion proteins constructed, GFP-G₄S186-HBc and RFP-G₄S47-HBc showed the best efficiency in rescuing the replication of an HBV replicon deficient in the core protein expression, though both of the two

fusion proteins failed to support the formation of the relaxed circular DNA. These fluorescent protein-tagged HBcs might help study related to HBc or capsids tracking in cells.

Keywords: Hepatitis B virus; core protein; engineering; fusion proteins.

Introduction

Hepatitis B virus (HBV) is an enveloped virus which packages its DNA genome in an inner icosahedra capsid. The capsids are built with core proteins (HBc), which are 183-185 amino acid residues long. According to the function, HBc can be divided into two domains: the assembly domain (amino acid residues 1-140, Fig. 1A) which is necessary and sufficient to form dimers, and the carboxy-terminal domain (CTD) (residues 150–183, Fig. 1A) which has a high affinity for RNA (Porterfield et al., 2010) and is needed for pgRNA encapsidation (Nassal, 1992). By providing a functional room for genome DNA replication, the capsids apparently are indispensable for the fulfillment of HBV life cycle (for reviews, see (Seeger and Mason, 2000, Nassal and Schaller, 1993, Seeger and Mason, 2015)).

Structure of HBV capsid has been studied extensively by both cryo-EM and X-ray crystallography (Bottcher et al., 1997, Crowther et al., 1994, Wynne et al., 1999, Dryden et al., 2006, Conway et al., 1997). Bacterially expressed core proteins spontaneously assemble into two types of capsids, resembling the

native capsids morphologically (Dryden et al., 2006), composed of 180 or 240 core proteins arranged with $T = 3$ or $T = 4$ icosahedral symmetry, respectively (Bottcher et al., 1997, Conway et al., 1997, Crowther et al., 1994, Wynne et al., 1999). The dimeric spikes formed by 4-helix bundles protrude from the surface of capsids. At the tips of the spikes are the major immunodominant region (MIR) of the HBc molecule presented by amino acids 78–82 (Bottcher et al., 1997, Wynne et al., 1999; Fig. 1A). The two N termini per dimer were located on either side of the spike-stem, at the level at which it enters the shell (Conway et al., 1998), and the C termini line the capsid interior (Zlotnick et al., 1997, Wang et al., 2012; Fig. 1A).

Fusion core proteins have been exploited to study core protein functions. Green fluorescent protein (GFP) is usually fused with HBc to facilitate observation. Schaller's lab used an N-terminally fused GFP-HBc to investigate the factors influencing the intracellular distribution of HBc (Weigand et al., 2010). The same lab used a line of N-terminally fused GFP-DHBcs to identify the nuclear localization signal that mediates nuclear pore association of the duck hepatitis B virus nucleocapsid (Mabit et al., 2001). A COOH-terminally truncated core protein with GFP inserted into the immunodominant loop (amino acids 78–83) (Kratz et al., 1999) has been used to investigate the interaction of hepatitis B virus capsids with nuclear pore complexes in permeabilized HeLa cells (Lill et al., 2006).

The above studies using GFP-HBc fusion proteins provided much helpful

information about the biological features of HBc. However, whether these fusion proteins keep the authentic functions of HBc is not so clear. One fact is that the N-terminally fused GFP-HBc is unable to form capsids (Weigand et al., 2010). Although the core protein with GFP inserted into the immunodominant loop did allow capsids formation (Kratz et al., 1999, Lill et al., 2006), our results showed that this fusion protein did not support HBV DNA replication in HepG2 cells (Chen et al., 2013). These data indicated that the functions of HBc were not fully kept in these GFP-HBcs, might attributable to the structural interference between the two proteins fused together. A core fusion protein fails to keep the functions of the core domain means that the core domain has been different from the wild type. Thus, studying core protein using this kind of core fusion proteins might be unsuitable in cases requiring a highly functional core domain. Here, we reported a solution for the problem of functional interference on HBc imposed by the specific foreign proteins in fusion proteins. This solution contains two key points: (i) fuse the foreign proteins at the N terminus of HBc; (ii) use long flexible linkers between the two proteins to be fused.

Materials and Methods

Plasmids.

HBV1.1c⁻ was constructed previously, in which the 40th amino acid of core protein was mutated to a stop codon (GAA to TAA) (Chen et al., 2013). WtHBc expresses wild type HBc of genotype D (Chen et al., 2013) derived

from plasmid PCH9/3091 (Nassal, 1992). G-C12-HBc expresses the HBc with the extra 12 amino acids (Fig 1, B) from HBc of genotype G inserted between its 2nd and 3rd amino acid. To construct G-C12-HBc, a fragment was amplified from WtHBc with primers F G12 and Rc456. The 5' end of F G12 carried a sequence that can anneal to the sequence immediately upstream of the 2nd codon of the HBc gene on WtHBc. Primer Rc456 can anneal to the sequence in the HBc gene (nt434-456). This fragment was used to replace the corresponding HBc gene region by Fragment Substitution Reaction (FSR) method developed previously (Hu et al., 2012). Briefly, a FSR system of 20 μ l was constructed which contained 200ng fragment F G12+ Rc456, 10ng plasmid WtHBc, 10 μ l PrimeStar HS 2 \times premix (Takara) and H₂O. The reaction was performed on a PCR instrument under following condition: Initial denaturation at 94 $^{\circ}$ C for 3 min, followed by 18 cycles of denaturation at 94 $^{\circ}$ C for 15 sec, annealing at 55 $^{\circ}$ C for 15 sec, and extension at 72 $^{\circ}$ C for 3 min. The resultant FSR product was purified with a PCR Purification Kit (Roche Diagnostics), digested with *Dpn* I (Promega, Madison, WI), and then transformed into competent JM109 cells. The clone contained the replaced fragment was confirmed by sequencing. 4 \times G-C12-HBc is the same as G-C12-HBc except that it contains 3 copies more G-C12 at the N terminus of HBc. To construct G-C12-HBc, a fragment was first amplified from G-C12-HBc with primers F4G3 and R1798. This fragment was amplified with primer F4G2 and Rc456, and the resulting fragment was then amplified with

primers F4G1 and Rc456. The last fragment was used to replace the corresponding region on G-C12-HBc by FSR method. G₄S12-HBc is a plasmid that expresses the HBc with a 12 amino acid residues-long Glycine-Serine linker (GSG₄SG₄S) at its N terminus. To construct G₄S12-HBc, the fragment amplified from WtHBc with primers F G₄S12 and R1798 was replaced to WtHBc by FSR method. G₄S47-HBc, G₄S92-HBc, and G₄S186-HBc are the same as G₄S12-HBc except that longer Glycine-Serine linkers were added to the N terminus of HBc respectively. To construct G₄S47-HBc (G₄S47 means a linker with a length of 47 amino acids, and the sequence of which was presented in Fig. 1B), a fragment was amplified first from G₄S12-HBc with primers F47gly 2 and R1798, and this fragment was used as template to amplify a fragment with primers F47gly 1 and Rc456. The last fragment was then replaced to WtHBc by FSR method. To construct G₄S92-HBc, primers F BsmBI1 and R BsmBI1 were used to amplify the fragment 1 (frag1) from G₄S47-HBc. Primers F BsmBI2 and R BsmBI2 were used to amplify the fragment 2 (frag2) from G₄S47-HBc. Primers F BsmBI vect and R BsmBI vect were used to amplify the fragment vector (vector) from WtHBc. The above three fragments were ligated together by Golden gate method (Engler et al., 2009, Engler et al., 2008, Weber et al., 2011). Specifically, a 10µl Golden gate system was constructed which contained 20ng of the fragment each, 0.75µl BsmB I (NEB), 1µl DTT (10mM), 0.25µl T7 ligase (NEB), 1µl ATP (10mM), 1µl Tango buffer (Fermentas) and H₂O. Reaction parameters: 37°C for 5 min,

22 °C for 5 min, 25 cycles. The resultant product was transformed into competent JM109 cells. The right clones were confirmed by sequencing. To construct G₄S186-HBc, the 4 fragments were amplified respectively from G₄S47-HBc using primer pairs F BsmBI1 + R BsmBI1, F BsmBI2 + R BsmBIa, F BsmBI1b + R BsmBI1b, and F BsmBIc + R BsmBI2. These 4 fragments were ligated with the fragment vector (the same as the fragment vector above) by Golden gate method. To construct GFP-G₄S47-HBc, GFP-G₄S92-HBc and GFP-G₄S186-HBc, the EGFP fragment amplified from pEGFP-N1 with primers FGFP-HBC and RGFP-HBC was replaced to G₄S186-HBc by FSR. There are 4 possible annealing sites between the fragments and the glycine-serine clustered repeat sequence on G₄S186-HBc. FSR product after *Dpn* I digestion was transformed, and three GFP-HBc constructs with G₄S47, G₄S92 and G₄S186 linkers were obtained by clone screening and sequencing. Similarly, to construct RFP-G₄S47-HBc and RFP-G₄S186-HBc, the RFP fragment amplified from pAD-RFP (kindly provided by Dr. TongChuan He) with primers FRFP-HBC and RRFP-HBC was replaced to G₄S186-HBc. Two RFP-HBC constructs with G₄S47 and G₄S186 linkers were obtained by clone screening. All the sequences of the primers used are listed in table 1.

Cell culture and transfection

HepG2 cells were grown at 37 °C under 5% CO₂ in modified Eagle's medium supplemented with 10% (v/v) fetal bovine serum. For transfection, the cells were seeded into six-well plates and allowed to adhere overnight. On the

following day, when the cells were 60%–70% confluent, the culture medium was replaced with fresh medium and 4 μ g of HBV construct was transfected into the cells in each well using 8 μ L Lipofectamine 2000 (Invitrogen), according to the instructions provided by the supplier. The culture medium was changed every 2 days, and the cells were harvested on day 5 after transfection.

Western blot analysis.

Cell lysis samples resolved on 10% SDS-PAGE gels were transferred onto PVDF membranes (Millipore). The membranes were blocked with 5% skim milk powder and probed with specific primary antibodies against GFP (Origen), RFP (Abcam) and GAPDH (Beyotime), respectively. Bound antibody was revealed with secondary antibodies conjugated with HRP and the chemiluminescent substrate ECLplus (Amersham Pharmacia), and then visualized by Fusion Fx5 system (Vilber Lourmat).

Viral DNA analysis

Intracellular viral core DNA was extracted as following. Cells from one well of a 6-well plate were lysed with 400 μ l of lysis buffer containing 10 mM Tris-HCl (pH 8.0), 1 mM EDTA, 1% NP-40 at room temperature for 30 min. Cell debris and nuclei were removed by centrifugation, and the supernatant was mixed with 3 μ l of 1 M MgCl₂ and DNase I (Promega; added to 40 U/ml) to digest the contaminating plasmid. After 3 hr at 37 °C, the reaction was terminated with 10 mM EDTA, and then 140 ml of 35% polyethylene glycol (PEG) 8000 containing 1.5 M NaCl was added. After incubation for 1 hr on ice,

the viral nucleocapsids were pelleted by centrifugation at 11,000g for 5 min at 4 °C, and then digested (3 hr at 45 °C) in 400 µl of buffer containing 1 mg/ml proteinase K (Promega). The digestion mixture was extracted twice with phenol. The DNA was precipitated with ethanol and dissolved in TE buffer (10 mM Tris-HCl [pH 8.0], 1 mM EDTA). For southern blot, DNA samples were resolved in 1% agarose gels and transferred to positively charged nylon membranes (Roche). Hybridization and detection were performed with DIG DNA Labeling and Detection Kit (Roche), according to the manufacturer's instructions.

Viral RNA analysis

Trizol reagent (Life technologies, Carlsbad, USA) was used to extract total RNA according to the manufacturer's instructions. Encapsidated RNA was extracted as previously described (Guo et al., 2003, Mao et al., 2013) with modifications. Briefly, cells in a well of a 12 well-plate were lysed with 250µl of lysis buffer containing 10 mM Tris-HCl (pH 8.0), 1 mM EDTA and 1% NP-40 at room temperature for 30 min. Cell debris and nuclei were removed by centrifugation, and the supernatant was mixed with 1.5µl of 1 M CaCl₂ and 10U micrococcal nuclease (NEB) to digest the unprotected RNA for 15min at 37 °C. Add 6µl of 0.5M EDTA to stop the reaction. Then 750µl of Trizol LS (Life technologies) was used to extract the encapsidated pgRNA according to the manufacturer's instructions. For Northern blot, 5µg of total RNA of each sample or encapsidated pgRNA from each well was separated on a 1.2%

agarose gel containing 2% formaldehyde and then transferred to a positively charged nylon membrane (Roche Diagnostics). HBV RNA was detected by a digoxigenin labeled RNA probe prepared by DIG Northern Starter Kit (Roche Diagnostics) according to the manufacturer's instructions. The template for RNA probe preparation was the 0.7kb fragment amplified from PCH9/3091 using primers F1085 and R1798 T₇.

Core particle analysis.

Intracellular HBV capsids and associated viral DNA were analyzed as described previously (Guo et al., 2006, Guo et al., 2007, Xu et al., 2010). Briefly, cells in each well of a 6-well plate were lysed by 400 μ l buffer containing 10 mM Tris-HCl (pH 7.6), 100 mM NaCl, 1 mM EDTA, and 0.1% NP-40. Cell debris was removed by centrifugation at 12,000 \times g for 5 min. Twenty microliters of the clarified cell lysates were resolved on nondenaturing 1% agarose gels and transferred to nitrocellulose filters by blotting with TNE buffer (10 mM Tris-HCl, pH 7.6, 150 mM NaCl, and 1 mM EDTA). HBV capsids were detected by probing the membrane with the antibody against the HBV core protein or against GFP mentioned above. Bound antibody was revealed by HRP-conjugated secondary antibodies and the chemiluminescent substrate ECLplus, and then visualized by Fusion Fx5 system. To detect capsid-associated HBV DNA, the membranes were treated with buffer containing 0.5 N NaOH and 1.5 M NaCl for 5 min, followed by neutralization with buffer containing 1 M Tris-HCl and 1.5 M NaCl for 5 min. The viral DNA

was detected with DIG DNA Labeling and Detection Kit, according to the manufacturer's instructions. For extracellular capsids assay, 600 μ l of 35% PEG8000 was added to 2ml of culture medium. Stand the mixture overnight at 4 °C after rotating for 30min. Capsids were then precipitated by centrifugation with 10,000g for 5min. The pellet was dissolved with 20 μ l TNE buffer and applied to native agarose gel assay as described above.

Immunofluorescence Assay

Cells on glass coverslips were fixed with 2% paraformaldehyde in PBS, and 0.1% Triton-X100 was used to permeabilize cell membrane for 1hr. Samples were then blocked for 1hr with blocking buffer (10%FBS and 2%BSA in PBS). For HBc detection, the cells were incubated with the DAKO rabbit polyclonal antibody (1/500 in the blocking buffer) for 1.5hr. After rinsing with PBS 3 times, the cells were incubated with the secondary antibody (Alex Fluro 594 goat anti rabbit antibody, Life technologies) and DAPI (Thermo Fisher Scientific) for 1hr. For fluorescent protein detection, only DAPI was used. The confocal images were collected with the Leica DMI8 confocal microscope through a 40 \times objective.

Results

The N terminus of core protein tolerates modification

To keep the function of HBc in a fusion protein, we first considered the fusion sites. As mentioned above, heterogenous peptides or proteins have been

linked to 3 sites, i.e. N terminus, C terminus and MIR, of HBc, resulting in fusion proteins that can form capsids (Roose et al., 2013). We chose the N terminus of HBc as our modification target based on the following reasons: (1) We speculated that a protein fused to the C terminus of authentic HBc might interfere the function of CTD, and in turn the package of pgRNA, since the C terminus lines the capsid interior (Fig. 1A). (2) Our previous results showed that the HBc with GFP inserted in MIR did not support HBV DNA replication in HepG2 cells (Chen et al., 2013). (3) HBc of genotype G HBV has an extra 12 amino acid residues at its N terminus, in contrast to other genotypes (Fig.1B). This extra 36 nucleotide sequence had been shown to increase core protein at the level of translation, and do not influence genome maturity when introduced into genotype A and D clones (Li et al., 2007). This implies that the N terminus of HBc might tolerate modification. (4) The polymerase in the nucleocapsids formed by the HBc with an extra 23 amino acid residues at the N terminus did presented endogenous polymerase activity, though this modified HBc supported HBV DNA replication less efficiently than wild type in hepatoma cells (Hui et al., 1999).

The next thing to be considered was the type and length of linker. A linker serves as a separator to decrease the interference between the two proteins fused. We tested 2 types of linkers. One is the extra 12-residues presented in the N terminus of genotype G HBc. The reason we chose this peptide is that it's a naturally evolved sequence, and do not influence HBV replication when

introduced into the HBc of genotype A and D (Li et al., 2007). The other one tested is the usually used Glycine-Serine flexible linker.

We tested the function of the HBcs fused different linkers alone at their N terminus. Replication rescue experiment was conducted to test the function of modified HBcs. The plasmids expressing different HBcs were co-transfected respectively with HBV1.1c, a plasmid expressing all other components required for HBV DNA replication except for HBc. The replication of HBV DNA will be rescued if a functional HBc is provided. As shown in Fig.1 D, G-C12-HBc rescued HBV DNA replication while 4×G-C12-HBc not, suggesting that extended G-C12 peptide (Fig. 1C) do not obstruct the function of core protein significantly but 4×G-C12 (Fig. 1C) do. Specifically, no pgRNA was encapsidated in the 4×G-C12-HBc co-transfection experiment (Fig.1D, lower panel). Co-transfection of G₄S12-c and G₄S47-c with HBV1.1c confirmed that both rescued HBV DNA replication well (Fig. 1D), and the G₄S47-c even supported replication better than WtHBc (Fig.1D). More encapsidated pgRNA was detected in the G₄S47-c transfection samples (Fig.1D). This result encouraged us to add longer G₄S linkers to the N terminus of core protein, with the speculation that longer linkers might separate fused proteins better. Interestingly, core protein with G₄S linkers even longer than HBc itself (G₄S186-HBc) in length supported HBV DNA replication (Fig.1D), indicating these linkers themselves do not disrupt HBC function.

GFP-HBc with long GS linkers keeps the functions of individual domains.

We further tested whether those long GS linkers described above can serve as effective “separators” to keep functions of the 2 individual domains in an HBc fusion protein. We constructed plasmids expressing the fusion protein EGFP-HBcs. The EGFP fragment amplified from pEGFP-N1 was cloned into the plasmid G₄S186-HBc by Fragment Substitution Reaction (FSR) method developed previously (Fig. 2 A and reference (Hu et al., 2012)). One end of the EGFP fragment can anneal to the immediate upstream of the G₄S186, and the other end can anneal to one of the 4 possible sites (because G₄S186 is roughly 4×G₄S47. Fig. 2A). Among the picked clones after transformation of FSR product, we identified 3 constructs of GFP-G₄S47-HBc, GFP-G₄S92-HBc and GFP-G₄S186-HBc by sequencing. These three clones correspond respectively to the molecules produced by annealing of the EGFP fragment to the site 1, 2 and 4 indicated in Fig. 2A.

Western blotting confirmed the expression of these HBc fusion proteins (Fig. 2B). To assess the function of these fusion proteins, we transfected these constructs into HepG2 cells. Green fluorescence can be easily observed 24h after transfection (Fig. 2C), indicating that the function of EGFP was retained in all the 3 fusion proteins. Almost all of the green fluorescence (from GFP) exactly colocalized with the red fluorescence (from HBc) in the 3 GFP-HBc transfection groups, indicating that these fusion proteins were expressed. Notably, the fluorescence forms distinct foci. This pattern is different sharply from that of the EGFP expressed from plasmid pEGFP-N1 (Fig. 2C), or the

GFP-HBc fusion protein reported previously (Weigand et al., 2010), which diffused evenly throughout the whole cells. The foci of fluorescence implied that these fusion proteins might aggregate themselves together through the interaction between core proteins. The WtHBc presented less foci than those GFP-HBcs, and distributed more evenly both in the cytoplasm and the nuclei.

To confirm the particle formation of the GFP-HBc proteins, cytoplasm lysis were resolved on native agarose gels after co-transfection of these constructs with HBV1.1c⁻ respectively. A GFP antibody and an HBc antibody were used to detect capsid-like particle. As shown in Fig 2D, capsid-like particles formed by GFP-G₄S47-HBc, GFP-G₄S92-HBc and GFP-G₄S186-HBc can be detected with both of the antibodies, while WtHBc core particle can only be detected by the HBc antibody. The signals appeared at the same location when using both GFP and HBC antibodies suggested that the particles contain the corresponding fusion proteins. As expected, the migration velocity was WtHBc > GFP-G₄S47-HBc > GFP-G₄S92-HBc > GFP-G₄S186-HBc. To characterize these capsid-like particles further, we assayed the particle-associated HBV DNA. HBV DNA probe was used to detect HBV DNA in these capsid-like particles. As shown in Fig. 2D, only the capsid-like particles formed by GFP-G₄S186-HBc contained detectable HBV DNA. The same location of HBV DNA and the core particles indicated that the DNA was synthesized in the capsid-like particles formed by G₄S186-HBc. And this kind of capsid-like particles must be functional particles resembling wild type capsids. No HBV

DNA in GFP-G₄S92-HBc particles was detected, probably attributed to its low quantity, since the HBV DNA appeared in the extracted cytoplasm DNA, though very weak (Fig. 2 F), must be synthesized in functional capsids. In conclusion, EGFP-HBc fusion protein linked with G₄S186 linker keeps the function of both EGFP and HBc well.

RFP-HBc with long GS linkers keeps the functions of individual domains.

As another example of HBc fusion protein, RFP gene was inserted at the N terminus of HBc linked with G₄S47 and G₄S186 linkers, using FSR method described above (Fig. 3 A). Expression of these two proteins in HepG2 cells was confirmed by Western blotting (Fig. 3 B). A line of evidence supported that these fusion proteins keep, at least partly, the function of both of the two fused domains. (1) The two fusion proteins both emitted red fluorescence when transfected into HepG2 cells, while the distribution patterns were somewhat different from that of WtHBc (Fig. 3C); (2) Intracellular capsid-like particles can be detected both by an RFP and an HBc antibody at the same location (Fig. 3 D); (3) Intracellular core DNA can be detected when these two constructs were co-expressed with HBV1.1c⁻ respectively. RFP-G₄S47-HBc supported HBV DNA replication better than RFP-G₄S186-HBc, implying less interference RFP imposed on HBc structure. Noteworthy is that two bands were detected of the capsids formed by RFP-G₄S186-HBc, implying two kinds of capsids formed. Further study is needed to identify whether the lower band

represents capsids composed of fewer copies of RFP-G₄S186-HBc proteins. Capsid DNA was not clearly detected both for RFP-G₄S47-HBc and RFP-G₄S186-HBc, might attributed to their less efficiency to support HBV DNA replication compared to the WtHBc (Fig. 3 F).

We next co-transfected RFP-G₄S47-HBc with GFP-G₄S186-HBc into HepG2 cells, to observe if these 2 proteins colocalize. As shown in Fig 2C, these 2 proteins apparently colocalized together, especially in the foci which distributed both in the cytoplasm and nuclei. By contrast, in the cells co-transfected with RFP and GFP expressing plasmids, both of the 2 proteins evenly distributed throughout the cells and no apparent colocalization can be observed (Fig. 3C).

Discussion

We provided a way to keep the individual function of the proteins being fused in an HBc fusion protein. One purpose of constructing fusion proteins is to endow a protein with new function, while retaining its original function. Unfortunately, the proteins being linked together may interfere with each other in many cases. As for HBc, it functions as a building block of capsids, a packager of HBV pgRNA and polymerase, and a helper during the synthesis of HBV DNA in capsids (for a review, see (Zlotnick et al., 2015)). Keeping one of these functions in a fusion protein, like the capacity of capsids formation, may be relatively easy to achieve, as demonstrated by the numerous examples for

vaccine purpose (Pumpens and Grens, 2001, Roose et al., 2013, Ulrich et al., 1998). To our knowledge, however, no HBc fusion protein has been reported in which more functions of HBc have been successfully kept, except the one with an extra 23aa stretch at the N terminus of HBc (Hui et al., 1999). Some of the HBc fusion proteins expressed here can form functional capsids. HBV DNA replicated in capsids indicated that pgRNA and polymerase have been packaged, and they must interact with the capsids appropriately. The retention of these functions after the addition of GFP and RFP tags provided us a new tool to study the dynamics of HBc or the capsids and the interaction between HBcs and themselves and other virus or host factors in cells.

All of the HBc fusion proteins constructed here formed apparent foci in cells, largely in the cytoplasm and less in the nuclei (Fig. 2C and Fig. 3C). WtHBc presented foci only in a minor part of the cells in our experiment, and the EGFP and RFP did not form this kind of foci (Fig. 2C and Fig. 3C). Possible reason for this difference include: (1) the fusion made the original proteins (the fluorescent proteins or the HBc) tend to aggregate, forming larger particles than a normal capsid which usually contains 240 HBcs. (2) more antibodies bound to the capsid like particles formed by the HBc fusion proteins, making these particles look like larger than a normal one. We have no evidence to distinguish these hypotheses till now, and further studies are warranted to make this clear.

In addition, co-transfection experiment clearly demonstrated that GFP-G₄S186-HBc and RFP-G₄S47-HBc had an apparent colocalization (Fig.

3C). This colocalization cannot be explained by the interaction between GFP and RFP, since the co-transfection experiment with GFP and RFP expressing plasmids did not show a colocalization of these 2 proteins (Fig. 3C). There probably is an interaction between the HBcs in the 2 fusion proteins, presumably an interaction between 2 heterogeneous monomers (a GFP-G₄S186-HBc monomer and a RFP-G₄S47-HBc monomer), or that between 2 heterogeneous dimers (a GFP-G₄S186-HBc dimer and a RFP-G₄S47-HBc dimer). If this is the case, the colocalized fluorescence would be used to reflect the interaction.

Different linkers have different influence on the function of fusion proteins. Although the G-C12 peptide did not disturb the function of HBc, the 4×G-C12 peptide significantly impaired it. Apparently, 4×G-C12-HBc did not assemble a capsid that encapsidate pgRNA (Fig.1C). We did not determine whether this kind of fusion protein can form capsids, and how this repeated peptide influence the function of HBc is not clear. All these flexible linkers with different lengths we tested did not influence the function of HBc significantly. A surprising observation is that the HBc added the G₄S47 linker supported HBV DNA replication better than the wild type. Possible explanations for this phenomenon include but not limit to that the nucleocapsids formed by G₄S47-c may not be enveloped and thus accumulated in cells, or that G₄S47-c has a higher efficiency to package pgRNA. While the EGFP was fused to HBc, it was linker G₄S186 but not others that kept the function of HBc best. However,

no clear relaxed circular DNA was detected in GFP-G₄S186-HBc rescued replication (Fig. 2F), indicating a defect in either second-strand DNA synthesis or genome circularization still exists. In the case of RFP-HBcs, the one with G₄S47 linker performed better than that with G₄S186 in supporting HBV DNA replication. Due to the complicated factors influencing the structure and function of a protein, which kinds of linkers to be adopted for the construction of other fusion proteins need to be determined empirically.

Acknowledgments

This work was supported in part by grant 81671997, 81471946, 81661148057 and 81501751 from the National Natural Science Foundation of China, the Major National S&T program grant 2017ZX10202203 and 2017ZX10302201 from Science & Technology Commission of China, grant cstc2015jcyjA10023 and cstc2015jcyjA100006 from Chongqing Science & Technology Commission, the Program for Innovation Team of Higher Education in Chongqing (grant no. CXTDX201601015), and the Leading Talent Program of CQ CSTC (CSTCCXLJRC201719).

Conflict of interest

The authors declared that they have no conflicts of interest to this work.

Reference

- BOTTCHER, B., WYNNE, S. A. & CROWTHER, R. A. 1997. Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy. *Nature*, 386, 88-91.
- CHEN, J., HUANG, R., TAO, Y., HUANG, Y., LUO, Y., HUANG, A. & HU, J. 2013. [Impacts on hepatitis B virus replication by gene engineering at apical loop region of capsid protein]. *Sheng Wu Gong Cheng Xue Bao*, 29, 1663-71.
- CONWAY, J. F., CHENG, N., ZLOTNICK, A., STAHL, S. J., WINGFIELD, P. T. & STEVEN, A. C. 1998. Localization of the N terminus of hepatitis B virus capsid protein by peptide-based difference mapping from cryoelectron microscopy. *Proc Natl Acad Sci U S A*, 95, 14622-7.
- CONWAY, J. F., CHENG, N., ZLOTNICK, A., WINGFIELD, P. T., STAHL, S. J. & STEVEN, A. C. 1997. Visualization of a 4-helix bundle in the hepatitis B virus capsid by cryo-electron microscopy. *Nature*, 386, 91-4.
- CROWTHER, R. A., KISELEV, N. A., BOTTCHER, B., BERRIMAN, J. A., BORISOVA, G. P., OSE, V. & PUMPENS, P. 1994. Three-dimensional structure of hepatitis B virus core particles determined by electron cryomicroscopy. *Cell*, 77, 943-50.
- DRYDEN, K. A., WIELAND, S. F., WHITTEN-BAUER, C., GERIN, J. L., CHISARI, F. V. & YEAGER, M. 2006. Native hepatitis B virions and capsids visualized by electron cryomicroscopy. *Mol Cell*, 22, 843-50.
- ENGLER, C., GRUETZNER, R., KANDZIA, R. & MARILLONNET, S. 2009. Golden gate shuffling: a one-pot DNA shuffling method based on type II restriction enzymes. *PLoS One*, 4, e5553.
- ENGLER, C., KANDZIA, R. & MARILLONNET, S. 2008. A one pot, one step, precision cloning method with high throughput capability. *PLoS One*, 3, e3647.
- GUO, H., ALDRICH, C. E., SAPUTELLI, J., XU, C. & MASON, W. S. 2006. The insertion domain of the duck hepatitis B virus core protein plays a role in nucleocapsid assembly. *Virology*, 353, 443-50.
- GUO, H., JIANG, D., ZHOU, T., CUCONATI, A., BLOCK, T. M. & GUO, J. T. 2007. Characterization of the intracellular deproteinized relaxed circular DNA of hepatitis B virus: an intermediate of covalently closed circular DNA formation. *J Virol*, 81, 12472-84.
- GUO, J. T., PRYCE, M., WANG, X., BARRASA, M. I., HU, J. & SEEGER, C. 2003. Conditional replication of duck hepatitis B virus in hepatoma cells. *J Virol*, 77, 1885-93.
- HU, J. L., CUI, J., DENG, X. Y., ZHANG, W. L., LI, Q. L., GUO, J. J., ZENG, A. Z. & HUANG, A. L. 2009. A new strategy for constructing in vitro replication-competent 1.3 copies of hepatitis B virus genome. *J Virol Methods*, 161, 63-9.
- HU, J. L., CUI, J., GUO, J. J., ZHANG, W. L., CAI, X. F., YUAN, Z. W., LI, Q. L., DENG, X. Y., ZENG, A. Z., HU, Y., TANG, N. & HUANG, A. L. 2012. Phenotypic assay of a hepatitis B virus strain carrying an rtS246T variant using a new strategy. *J Med Virol*, 84, 34-43.
- HUI, E. K., YI, Y. S. & LO, S. J. 1999. Hepatitis B viral core proteins with an N-terminal extension can assemble into core-like particles but cannot be enveloped. *J Gen Virol*, 80 (Pt 10), 2647-59.
- KRATZ, P. A., BOTTCHER, B. & NASSAL, M. 1999. Native display of complete foreign protein domains on the surface of hepatitis B virus capsids. *Proc Natl Acad Sci U S A*, 96, 1915-20.
- LI, K., ZOULIM, F., PICHOU, C., KWEI, K., VILLET, S., WANDS, J., LI, J. & TONG, S. 2007.

- Critical role of the 36-nucleotide insertion in hepatitis B virus genotype G in core protein expression, genome replication, and virion secretion. *J Virol*, 81, 9202-15.
- LILL, Y., LILL, M. A., FAHRENKROG, B., SCHWARZ-HERION, K., PAULILLO, S., AEBI, U. & HECHT, B. 2006. Single hepatitis-B virus core capsid binding to individual nuclear pore complexes in HeLa cells. *Biophys J*, 91, 3123-30.
- MABIT, H., BREINER, K. M., KNAUST, A., ZACHMANN-BRAND, B. & SCHALLER, H. 2001. Signals for bidirectional nucleocytoplasmic transport in the duck hepatitis B virus capsid protein. *J Virol*, 75, 1968-77.
- MAO, R., NIE, H., CAI, D., ZHANG, J., LIU, H., YAN, R., CUCONATI, A., BLOCK, T. M., GUO, J. T. & GUO, H. 2013. Inhibition of hepatitis B virus replication by the host zinc finger antiviral protein. *PLoS Pathog*, 9, e1003494.
- NASSAL, M. 1992. The arginine-rich domain of the hepatitis B virus core protein is required for pregenome encapsidation and productive viral positive-strand DNA synthesis but not for virus assembly. *J Virol*, 66, 4107-16.
- NASSAL, M. & SCHALLER, H. 1993. Hepatitis B virus replication. *Trends Microbiol*, 1, 221-8.
- PORTERFIELD, J. Z., DHASON, M. S., LOEB, D. D., NASSAL, M., STRAY, S. J. & ZLOTNICK, A. 2010. Full-length hepatitis B virus core protein packages viral and heterologous RNA with similarly high levels of cooperativity. *J Virol*, 84, 7174-84.
- PUMPENS, P. & GRENS, E. 2001. HBV core particles as a carrier for B cell/T cell epitopes. *Intervirology*, 44, 98-114.
- ROOSE, K., DE BAETS, S., SCHEPENS, B. & SAELENS, X. 2013. Hepatitis B core-based virus-like particles to present heterologous epitopes. *Expert Rev Vaccines*, 12, 183-98.
- SEEGER, C. & MASON, W. S. 2000. Hepatitis B virus biology. *Microbiol Mol Biol Rev*, 64, 51-68.
- SEEGER, C. & MASON, W. S. 2015. Molecular biology of hepatitis B virus infection. *Virology*, 479-480, 672-86.
- ULRICH, R., NASSAL, M., MEISEL, H. & KRUGER, D. H. 1998. Core particles of hepatitis B virus as carrier for foreign epitopes. *Adv Virus Res*, 50, 141-82.
- WANG, J. C., DHASON, M. S. & ZLOTNICK, A. 2012. Structural organization of pregenomic RNA and the carboxy-terminal domain of the capsid protein of hepatitis B virus. *PLoS Pathog*, 8, e1002919.
- WEBER, E., ENGLER, C., GRUETZNER, R., WERNER, S. & MARILLONNET, S. 2011. A modular cloning system for standardized assembly of multigene constructs. *PLoS One*, 6, e16765.
- WEIGAND, K., KNAUST, A. & SCHALLER, H. 2010. Assembly and export determine the intracellular distribution of hepatitis B virus core protein subunits. *J Gen Virol*, 91, 59-67.
- WYNNE, S. A., CROWTHER, R. A. & LESLIE, A. G. 1999. The crystal structure of the human hepatitis B virus capsid. *Mol Cell*, 3, 771-80.
- XU, C., GUO, H., PAN, X. B., MAO, R., YU, W., XU, X., WEI, L., CHANG, J., BLOCK, T. M. & GUO, J. T. 2010. Interferons accelerate decay of replication-competent nucleocapsids of hepatitis B virus. *J Virol*, 84, 9332-40.
- ZLOTNICK, A., CHENG, N., STAHL, S. J., CONWAY, J. F., STEVEN, A. C. & WINGFIELD, P. T. 1997. Localization of the C terminus of the assembly domain of hepatitis B virus capsid protein: implications for morphogenesis and organization of encapsidated RNA. *Proc Natl Acad Sci U S A*, 94, 9556-61.
- ZLOTNICK, A., VENKATAKRISHNAN, B., TAN, Z., LEWELLYN, E., TURNER, W. & FRANCIS,

S. 2015. Core protein: A pleiotropic keystone in the HBV lifecycle. *Antiviral Res*, 121, 82-93.

Figure legends.

Fig. 1, The N terminus of core protein tolerates modification. (A) Structure models of HBc monomer and dimer. (B) Alignment of the start part of the ORFs of HBc gene of A to H genotype. The sequences are from the whole genome of 831 HBV strains retrieved from Genbank(Hu et al., 2009). There is an extra 12aa insertion at the N terminus of HBc of genotype G. (C) Amino acid sequences added at the N terminus of genotype D HBc. (D) Functional assay of core fusion proteins. Plasmids expressing wild type or engineered core proteins were co-transfected with HBV1.1c⁻ into HepG2 cells respectively. Intracellular core DNA was extracted and detected by Southern blot. Total RNA or encapsidated RNA was extracted and detected by Northern blot. RC, relaxed circular DNA; DSL, double-strand linear DNA; SS, single-strand DNA. pgRNA, pregenomic RNA.

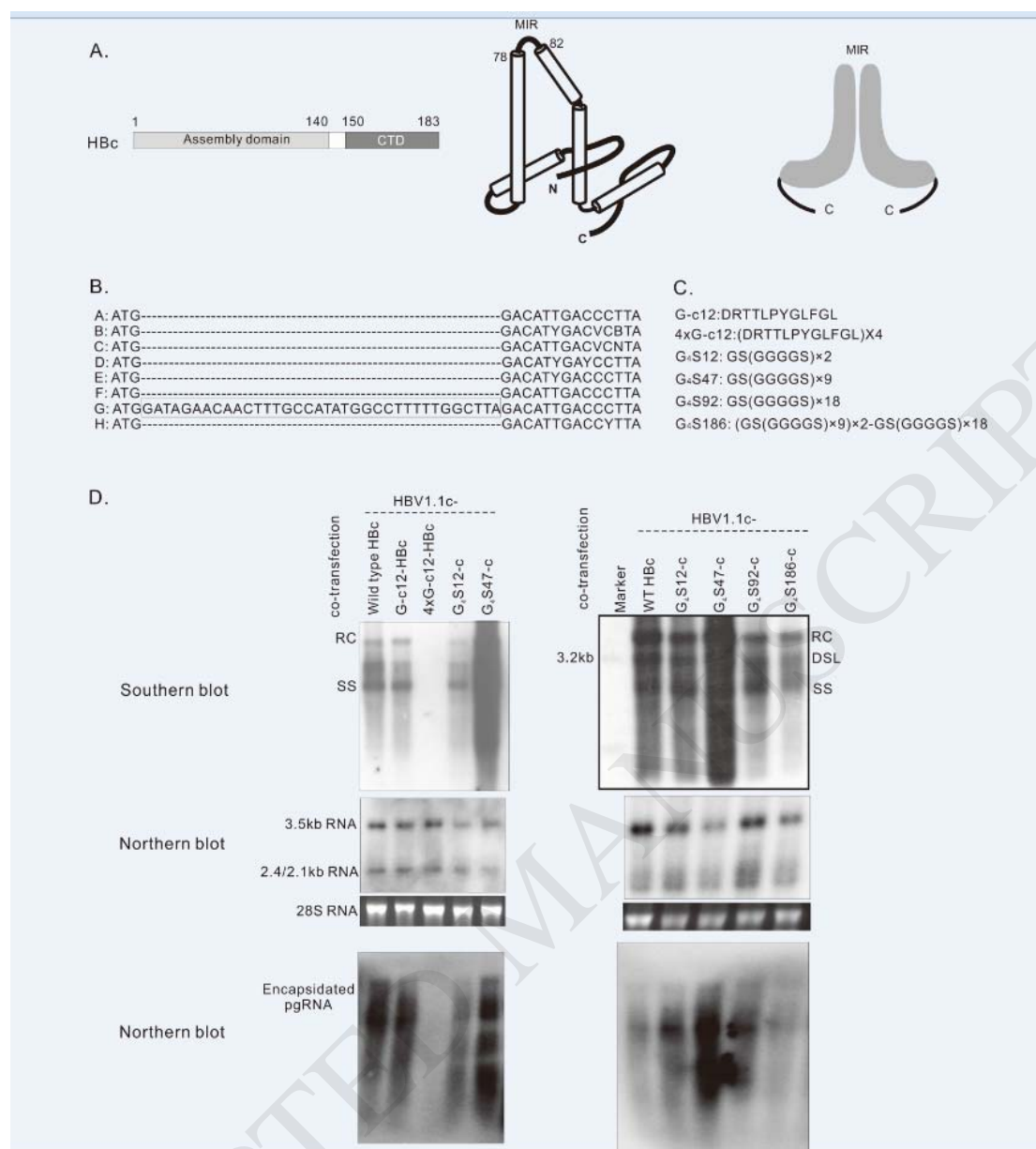


Fig. 2, Functional characterization of EGFP-HBcs. (A) Construction of EGFP-HBc expressing plasmids. The EGFP fragment was amplified from pEGFP-N1 with a pair of primers. At the 5' end (black bar on the mega primers) of the forward primer is a sequence that can anneal to the sequence (black bar on the plasmid) just upstream of the G₄S186 linker (blue bars), and the 5' end (orange bars on the mega primers) of the reverse primer was designed to anneal to the start parts of each of the four G₄S47 blocks. The fragment was then used

as “mega-primers” to replace the corresponding region on the plasmid G₄S186-HBc. Theoretically, there would be four kinds of products by annealing of the mega-primers to the different positions at the tandem repeat G₄S region on the template. Actually, three kinds of constructs, i.e. GFP-G₄S47-HBc, GFP-G₄S92-HBc and GFP-G₄S186-HBc, were obtained by sequencing. (B) Expression of EGFP-HBc fusion proteins in cultured cells. The 3 plasmids were transfected into HepG2 cells respectively. Western blotting was used to detect the expression of the fusion proteins 48h after transfection. (C) Intracellular localization by immunofluorescence confocal microscopy. The cells seeded on glass coverslips were transfected with the different plasmids. Twenty-four hours after transfection, the cells were assayed by immunofluorescence with the anti-HBc antibody. Nuclei were visualized by staining with DAPI. The green fluorescence of these samples was also observed. (D) Capsids and capsid-associated DNA assay. The HepG2 cells transfected with the 3 plasmids were harvested and lysed 5 days post-transfection, and part of the lysis of each sample was resolved by native agarose gels electrophoresis. For capsids assay, samples in the gels were transferred to an NC membrane, and then detected by using an HBc antibody and a GFP antibody. For capsid-associated DNA assay, samples were transferred to a nylon membrane, and then hybridized by using an HBV DNA probe. (E) HBV RNA assay. Total RNA was extracted from the HepG2 cells co-transfected with HBV1.1c⁻ and the plasmids expressing different core

proteins. Northern blotting was used to detect HBV RNA. (F) Core DNA assay. The 4 plasmids were co-transfected with HBV1.1c⁻ respectively and intracellular core DNA was extracted 5 days after transfection. HBV DNA was detected by Southern blotting. RC, relaxed circular DNA. SS, single-strand DNA.

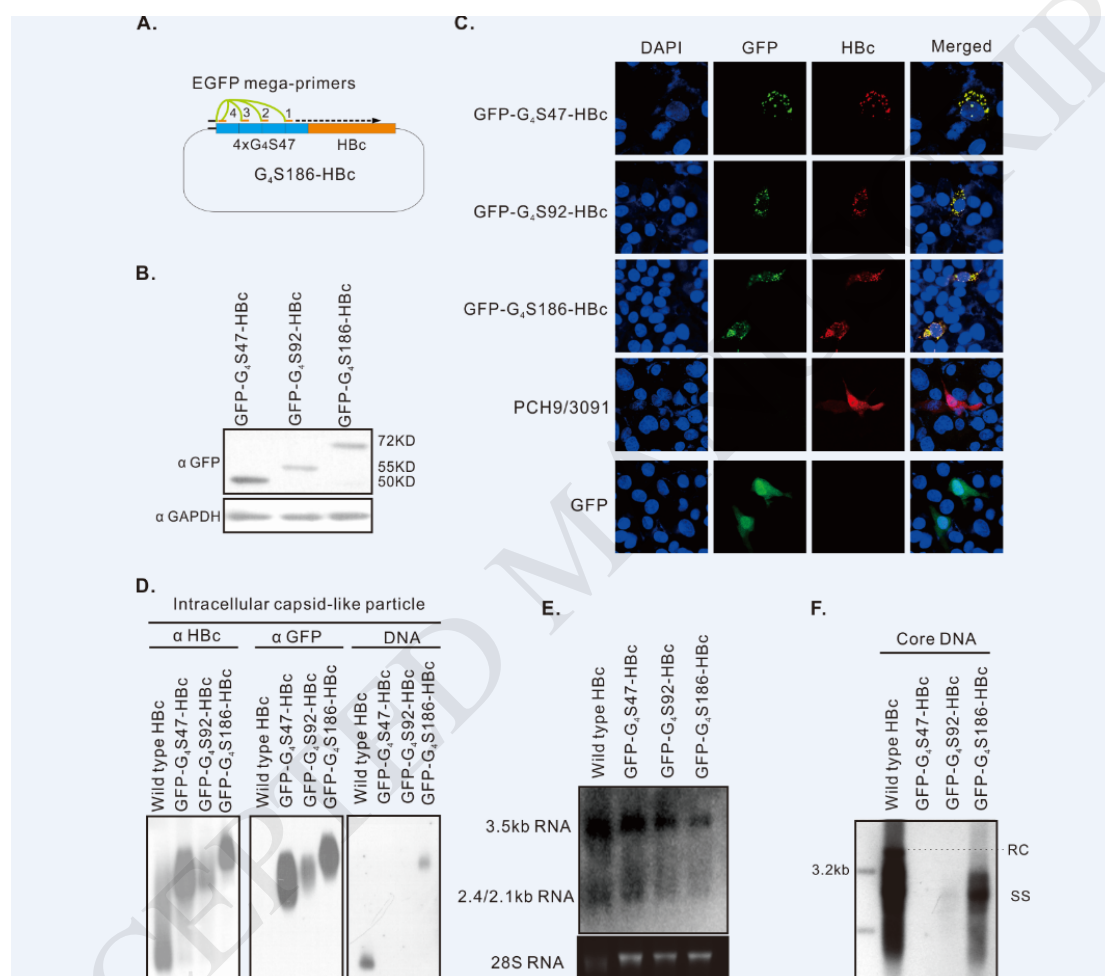
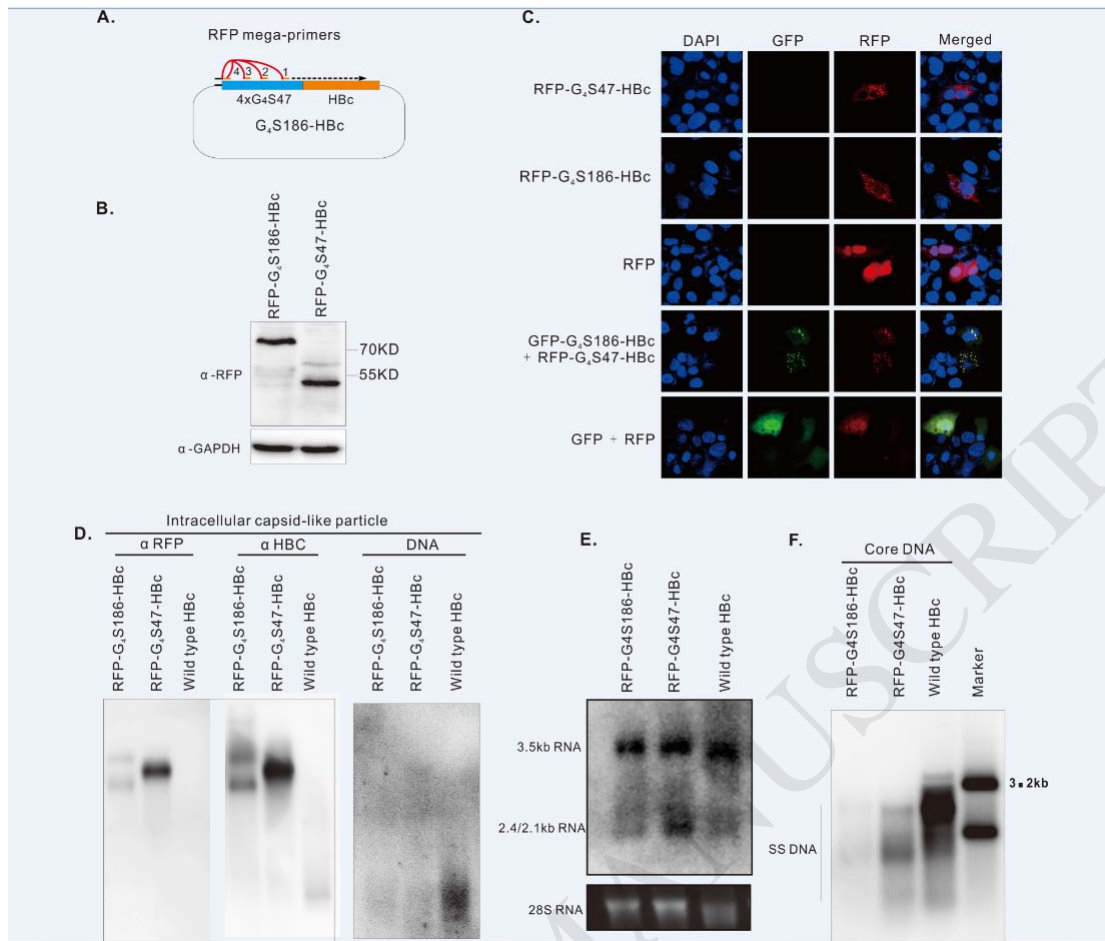


Fig. 3 Functional characterization of RFP-HBcs. (A) Construction of the RFP-HBc vectors. RFP gene was amplified from pAD-RFP and was used as “mega-primers” to replace the corresponding region on the plasmid G₄S186-HBc by FSR. There should be four annealing positions between the “mega-primer” and the tandem repeat region on the template. Two kinds of

constructs, i.e. RFP-G₄S47-HBc and RFP-G₄S186-HBc, were identified by sequencing. (B) Expression of RFP-HBcs in cells. HepG2 cells transfected with the 2 plasmids were assayed by Western blotting to confirm the expression of the fusion proteins. (C) Intracellular localization by immunofluorescence confocal microscopy. The cells transfected with different plasmids were assayed by immunofluorescence with the anti-HBc antibody (no antibody was used for the detection of fluorescent proteins). Nuclei were visualized by staining with DAPI. (D) Capsids and capsid-associated DNA assay. The HepG2 cells in a 6-well plate were transfected with the 2 plasmids plus HBV1.1c⁻ respectively and lysed 5 days post-transfection. Part of the lysis of each sample was assayed as described above except that an RFP antibody was used for capsids assay. (E) HBV RNA assay. Northern blotting was used to detect the HBV RNA extracted from the HepG2 cells co-transfected with HBV1.1c⁻ and the plasmids expressing different core proteins. (F) Core DNA assay. RFP-G₄S47-HBc and RFP-G₄S186-HBc were co-transfected with HBV1.1c⁻ and intracellular core DNA was extracted 5 days after transfection respectively. Southern blotting was used to detect HBV DNA. SS DNA, single-strand DNA.



Table

Primer	Sequence (5'-3')
F1.1c-	GATACCGCCTCAGCTCTGTATCGGTAAGCCTTAGAGTCTCCTG AGCATTG
R1.1c-	CTCGTCGTCTAACAACAGTAGTCT
FHBC	CAATCTCGGGAATCTCAATGTTAGGAGATTAGGTTAAAGGTCT TTGTAC
RHBC	GTACAAAGACCTTTAACCTAATCTCCTAACATTGAGATTCCCG AGATTG
F G12	GGTGGCTTTGGGGCATGGACAGAACAACCTTTGCCATATGGCCT TTTTGGCTTAGACATCGACCCTTATAAAGAATT
F4G 1	GTGGCTTTGGGGCATGGACAGAACAACCTTTACCCTATGGCCTA TTCGGCCTTGACCGTA
F4G2	ATTCGGCCTTGACCGTACTACCTTACCGTACGGATTGTTCCGGTC TAGATCGCACGACTC
F4G3	CTAGATCGCACGACTCTTCCATACGGACTCTTCGGACTTGACA GAACAACCTTTGCCATA
F G12	GTGGCTTTGGGGCATGGACAGAACAACCTTTACCCTATGGCCTA TTCGGCCTT GACATCGACC CTTATAAAGA
R2300	TCTATAAGCTGGAGGAGTGCGA
FG4S12	GCTTAGCCTTGGGTGGCTTTGGGGCATGGGTTTCAGGAGGTGG TGGATCTGGAGGAGGTGGATCTGACATCGACCCTTATAAAGA
R1798	CCAATTTATGCCTACAGCCT
F47gly 2	GAGGTTCCGGTGGCGGCGGCAGTGGTGGTGGAGGCTCTGGTG GTGGAGGCTCTGGAGGCGGAGGTTTCAGGAGGTGGTGGATC
F47gly 1	TTGGGTGGCTTTGGGGCATGGGATCAGGCGGTGGCGGTTTCAG GAGGTGGTGGCTCAGGCGGAGGAGGTTCCGGTGGCGG
Rc456	CTCGTCGTCTAACAACAGTAGTCT
R BsmBI vect	GCTGACCGTCTCTCATGCCCAAAGCCACCCAA
F BsmBI vect	TGCGTCCGTCTCCGACATCGACCCTTATAAAG
F BsmBI1	TGCGTCCGTCTCCATGGGATCAGGCGGTGGCGGTTTC
R BsmBI1	GCTGACCGTCTCACTGATCCACCTCCTCCAGATCCAC
F BsmBI2	TGCGTCCGTCTCATCAGGCGGTGGCGGTTTC
R BsmBI2	GCTGACCGTCTCCTGTGATCCACCTCCTCCAGATC
F BsmBIb	TGCGTCCGTCTCAGGCGGCGGCGGTGGCGGTTTC
R BsmBIa	GCTGACCGTCTCATGCCAGATCCACCTCCTCCAGA
F BsmBIc	TGCGTCCGTCTCCATCTGGATCAGGCGGTGGCGGTTTC

R BsmBIb	GCTGACCGTCTCTAGATCCACCTCCTCCAGA
FGFP-HBC	CCTCCAAGCTTAGCCTTGGGTGGCTTTGGGGCATGGTGAGCA AGGGCGAGGAGCT
RGFP-HBC	TGAGCCACCACCTCCTGAACCGCCACCGCCTGATCCCTTGTA AGCTCGTCCATGC
FRFP-HBC	CCTCCAAGCTTAGCCTTGGGTGGCTTTGGGGCATGGCCTCCTC CGAGGACGT
RRFP-HBC	TGAGCCACCACCTCCTGAACCGCCACCGCCTGATCCGTTATCT AGATCCGGTGGA
F1085	GCAGGCTTTCACCTTCTCGCCA
R1798 T ₇	TAATACGACTCACTATAGGGGACCAATTATGCCTACAGCCT

Table 1. Sequence of primers used to construct the plasmids or for RNA probe preparation.