SUPPORTING INFORMATION

Assessing sequence plasticity of a virus-like nanoparticle by evolution toward a versatile scaffold for vaccines and drug delivery

Yuan Lu[1], Wei Chan[1], Benjamin Y. Ko[1],

Christopher C. VanLang[1], James R. Swartz[1, 2, *]

1. Department of Chemical Engineering, Stanford University, Stanford, CA 94305

2. Department of Bioengineering, Stanford University, Stanford, CA 94305

[*] Correspondence to James R. Swartz, Department of Chemical Engineering, Stanford University, Stanford, CA 94305-5025, USA; email: jswartz@stanford.edu; telephone: 650-723-5398; fax: 650-725-0555

SI Materials and Methods

Plasmid Construction. The gene sequence encoding the human Hepatitis B core (HBc) antigen of subtype adyw (1) (UniProt accession number: P03147) with the C-terminus truncated at amino acid 149 was optimized for E. coli tRNA concentrations and was synthesized from oligonucleotides designed with DNAworks v3.0. The vector pET24a-HBc149 was generated by ligation (T4 DNA ligase, New England Biolabs, Ipswich, MA) of the optimized HBc protein gene into the pET-24a(+) vector (Novagen, San Diego, CA) at the Nde I and Xho I restriction sites. To incorporate methionine analogues, two mutations (M66S and L76M) were introduced. pET24a-HBc149-M66S-L76M was transformed into DH5a cells and the plasmid was purified with Qiagen Plasmid Maxi Kit (Qiagen, Valencia, CA) for use in cell-free protein synthesis (CFPS). All mutants were constructed using QuikChange PCR (Stratagene, La Jolla, CA). The sequences of 10 different variants with different cysteine mutations and 17 different variants with different AHA sites or different negative charge mutations are shown in *Table S1-3*.

Cell-free Protein Synthesis (CFPS). CFPS was conducted using the PANOx-SP (PEP, amino acids, nicotinamide adenine dinucleotide (NAD), oxalic acid, spermidine, and putrescine) cell-free system as described previously (2) with several modifications. The standard PANOx-SP CFPS reaction mixture includes: 1.2 mM ATP, 0.85 mM each of GTP, UTP, and CTP, 33 mM phosphoenol pyruvate (Roche Molecular Biochemicals, Indianapolis, IN), 170 mM potassium glutamate, 10 mM ammonium glutamate, 16 mM magnesium glutamate, 1.5 mM spermidine, 1.0 mM putrescine, 0.17 mg/mL folinic acid, 45 μg/mL plasmid, approximately 100−300 μg/mL T7 RNA polymerase, 2 mM of each of the 20 unlabeled amino acids, 0.33 mM NAD, 0.26 mM Coenzyme A (CoA), 2.7 mM potassium oxalate, and 0.28 volumes of *E. coli* KC6 S30 extract (3). For global replacement of methionines in HBc proteins, methionine was left out of cell-free reaction mixtures, and substituted with 1 mM azidohomoalanine (AHA) (Medchem Source LLP, Federal Way, WA) with an azide moiety. All reagents were obtained from Sigma−Aldrich (St. Louis, MO) unless otherwise noted.

CFPS reactions to produce the HBc protein were conducted at 30 °C for 6 h. Small-scale CFPS reactions were carried out in 20 μL volumes in 1.5 mL microcentrifuge tubes. Preparative-scale reactions used 6 mL volumes with 1 mL per well in 6-well tissue culture plates (BD Falcon #3046, BD, Franklin Lakes, NJ). 8.4 μM L-[U-^{14}C]-Leucine (PerkinElmer, Waltham, MA) was

added to small-scale reactions and to 20 μL aliquots of preparative-scale reactions for measuring protein yields using a previously described trichloroacetic acid protein precipitation protocol (4) and a Beckman LS3801 liquid scintillation counter (Beckman Coulter, Fullerton, CA).

The production of GM-CSF, IM9-scFv and CpG DNA with an alkyne moiety were described by Patel *et al* (5). The production of flagellin with an alkyne moiety was described by Lu *et al* (6).

We use "percent solubility" when we describe how much protein is soluble and "assembly efficiency" when we refer to how much soluble protein assembles into VLPs.

The term "percent solubility" is defined as percent of amount of proteins in the supernatant after centrifugation to total amount of proteins in the sample before centrifugation.

$$Percent\ solubility = \frac{Amount\ of\ soluble\ proteins}{Total\ amount\ of\ proteins} \times 100\%$$

The term "assembly efficiency" is defined as percent of amount of soluble proteins assembled into VLP to total amount of soluble proteins in the sample.

$$VLP\ assembly\ efficiency = \frac{\begin{matrix}Amount\ of\ soluble\ proteins \\ that\ assemble\ into\ VLPs\end{matrix}}{Amount\ of\ soluble\ proteins} \times 100\%$$

Size-exclusion Chromatography (SEC). To remove unincorporated L-[U-14C] leucine, the cell-free product was immediately dialyzed in 6-8000 MWCO Spectra/Por molecular-porous Membrane Tubing (Spectrum Labs, Rancho Dominguez, CA) against Dialysis Buffer (10 mM Tris-HCl, pH 7.4, 0.5 M NaCl) with 1 mM DTT, with two buffer exchanges in series. The dialyzed cell-free reaction product was loaded on an SEC column packed with Sepharose 6 FastFlow resin (GE Healthcare). The running buffer was the same as the dialysis buffer but with 1 mM DTT. The protein concentrations of eluate fractions were determined based on radioactivity. For small-scale SEC, 2.2 ml Sepharose 6 FastFlow resin was placed in a gravity-flow disposable polystyrene column (Thermo Scientific, #29920) and equilibrated with 5 column volumes of Dialysis Buffer. Following sample loading, the column was washed with Dialysis Buffer, and VLP fractions (9-11) were collected for protein concentration measurement and SDS-PAGE analysis.

The stability of purified VLPs were analyzed by SEC in different running buffers. The purified VLPs were initially formulated in the assembly buffer. The buffers were then exchanged into

different buffers (the assembly buffer, PBS buffer, and 10 mM Tris-HCl (pH 7.4) buffer without NaCl salt) by dialysis. The SEC running buffer was the same as the dialysis buffer.

Sucrose Gradient Sedimentation. The isolated VLPs from SEC were first dialyzed against the Dialysis Buffer to remove DTT and then oxidized to form disulfide bonds by adding 20 mM diamide at room temperature for 1 h. The oxidants were removed by dialysis against the Dialysis Buffer with two rounds of buffer exchange. The oxidized VLPs were then subjected to velocity sedimentation. Ten to forty percentage weight per volume continuous sucrose density gradients were prepared in Dialysis Buffer as indicated in Polyallomer 16 × 102 mm Centrifuge Tubes (Beckman Coulter) with the Gradient Master Ver3.05L Gradient Maker (Biocomp Instruments, Inc., Fredericton, Canada). The VLP products (200 μL) were layered on top of the sucrose and centrifuged at 31,000 rpm in a Beckman Coulter SW-32.1 swinging bucket rotor (Fullerton, CA) in a Beckman L8-M ultracentrifuge at 4 ^{o}C for 7 h with profile 7 slow acceleration and deceleration. One-half milliliter fractions were collected, and the concentration in each fraction was determined by radioactivity measurement.

The stability of HBc VLPs were analyzed by sucrose gradient centrifugation after one freeze-thaw cycle. The VLPs in the assembly buffer were flash frozen in the liquid N_2 and stored at -80 ^{o}C. After one week, the VLPs were thawed on ice and then analyzed by sucrose gradient centrifugation.

SDS-PAGE and Autoradiography. NuPAGE Novex precast gels and reagents were purchased from Invitrogen (Carlsbad, CA). For reducing SDS-PAGE, samples were denatured for 5 min at 95 ^{o}C in loading buffer (1X LDS running buffer and 50 mM dithiothreitol). For non-reducing SDS-PAGE, samples were only mixed with LDS running buffer, without addition of dithiothreitol and heat treatment. The samples were loaded onto a 10% (w/v) Bis-Tris precast gel with SeeBlue Plus2 molecular weight protein standard, and electrophoresed in MES/SDS running buffer. SimplyBlue SafeStain was used to stain and fix the gels according to the manufacturer's recommendations. The gels were dried using a gel dryer model 583 (Bio-Rad, Richmond, CA), before exposure to storage phosphor screen (Molecular Dynamics), which was subsequently scanned using a Typhoon Scanner (GE Healthcare).

Purification of HBc VLPs. The CFPS products were centrifuged for 15 min at 15,000 g to remove aggregates. After collecting the supernatant, saturated ammonium sulfate was added dropwise while mixing to a final concentration of 1.2 M (7). The supernatant was mixed for an additional 10 min at room temperature and centrifuged at 10,000 x g to pellet the precipitate. The precipitate was resuspended in Dialysis Buffer (10 mM Tris-HCl, pH 7.4, 0.5 M NaCl) with 10 mM DTT by vortexing. The mixture was then dialyzed against Dialysis Buffer with 1 mM DTT to remove residual ammonium sulfate and centrifuged for 15 min at 15,000 g to remove aggregates. The supernatant was loaded onto a Sepharose 6 Fast Flow column pre-equilibrated with the same buffer, and fractions were collected for analysis by SDS-PAGE. The VLP fractions were pooled and concentrated by ultrafiltration using an Amicon stirred cell with a 3 kDa molecular weight cutoff regenerated cellulose membrane. VLP fractions were then oxidized to form disulfide bonds by adding 20 mM diamide. The oxidants were removed by dialysis against the Dialysis Buffer with five rounds of buffer exchange. The endotoxins in VLP solutions were removed by phase separation using 0.5% Triton X-114 (8, 9). The endotoxin level was measured using Pierce LAL Chromogenic Endotoxin Quantitation Kit.

Azide-Alkyne Conjugation and Purification. The [3 + 2] cycloaddition click reactions were conducted in an anaerobic glovebox (Coy Laboratories, Grass Lake, MI) to preserve the reduced state of the tetrakis(acetonitrile)copper(I) hexafluorophosphate catalyst ($[(CH_3CN)_4Cu]PF_6$) (Sigma Aldrich, St. Louis, MO). Cu (I) catalyst was added to reactions at 1 mM in addition to 0.5 mM of the enhancer ligand, tris(triazolylmethyl) amine (TTMA), to improve the rate of the click reactions. HBc VLPs and functional molecules (flagellin, GM-CSF or CpG DNA) were mixed with the Cu (I) catalyst and TTMA enhancer with 0.01% Tween 20. Before addition of the Cu (I) catalyst, click reaction components were deoxygenated in 1.5 mL microcentrifuge tubes for 1 h in the anaerobic glovebox. The click reactions for attaching functional molecules to HBc VLPs were conducted overnight.

Conjugation efficiency of functional molecules (flagellin, GM-CSF or CpG DNA) were estimated by autoradiography after SDS-PAGE. The HBc proteins we used for click reaction were radioactive. The functional molecules were non-radioactive. The click-reaction products were analyzed by reducing SDS-PAGE. The catalyst Cu(I) was not added to the reaction as the control. The autoradiography of the gel was performed after SDS-PAGE. The HBc band and

conjugate band could be seen on the autoradiograph. Based on the density of control band and conjugate band, we could estimate the conjugation efficiency by densitometry using ImageJ software.

Transmission Electron Microscopy. A 5 μL sample of a purified 5 nM VLP solution was applied to a carbon coated copper/Formvar grid and negatively stained with 1% w/v uranyl acetate, pH 4. Photographs were taken with a Gatan Orius CCD camera in a JEOL JEM1400 electron microscope at 120 kV acceleration voltage.

Immunization of Mice. Six- to eight-week-old BALB/c mice were obtained and housed at the Laboratory Animal Facility at Stanford University Medical Center (Stanford, CA). All animal experiments were conducted following the Laboratory Animal Facility and National Institute of Health guidelines. The study protocol was approved by the Stanford University Institutional Animal Care and Use Committee. Ten mice per group were vaccinated intradermally with 3 μg KLH or HBc VLP formulated in PBS buffer. The endotoxin levels in the injected solutions are < 0.04 EU/dose. Mice were vaccinated on days 0, 10 and 20, and retro-orbital bled on days 9, 19 and 29 to assess immune responses. On day 36, the spleens were removed from three mice per group for T-cell proliferation assay.

ELISA Assay. In an enzyme-linked immunosorbent assay (ELISA), 50 μL of antigen proteins at 1 μg/mL concentrations were coated on 96-well ELISA plates (NUNC MaxiSorp) and allowed to bind overnight at 4 °C. We used the common Carbonate-Bicarbonate buffer (pH 9.6) as the coating buffer. Plates were then washed four times with PBS buffer and blocked with Blocking Buffer (PBS buffer with 1% (w/v) BSA) at room temperature for 1 h. After washing four times with Washing Buffer (PBS buffer with 0.05% (w/v) Tween 20), 50 μL of dilutions of mouse sera in Blocking Buffer were then added to the plates and incubated at room temperature for 1 h. Antibody 13A9 (IgG2b) binding to the region AA 135-140 of HBc antigen protein was used as the standard for quantitating the antibody levels. Plates were washed four times again with Washing Buffer before adding 0.1 μg/mL of peroxidase-conjugated monoclonal anti-mouse IgG(H+L) antibody (KPL) in Blocking Buffer and incubating at room temperature for 1 h. Plates were washed six times again with Washing Buffer before developing with 50 μL of Ultra-TMB

substrate (Pierce) for 10 min and quenching with 30 μL of 2 M H_2SO_4. Each well was measured at OD_{450} using a VersaMax microplate reader. Each data point indicates the mean of triplicate assay results and error bars represent standard deviation.

T-cell Proliferation Assay. Spleen cells were collected by squeezing the spleen through a 70 μm cell strainer, erythrocytes were removed by ACK lysis buffer (0.15 M NH_4Cl, 10 mM $NaHCO_3$, 0.1 mM EDTA, pH 7.3). Splenocytes were cultured at 2×10^5 cells/well in round-bottom 96-well plates in RPMI 1640 (10% heat-inactivated FBS, 2 mM L-glutamine, 50 μM 2-ME, 100 U/ml penicillin, 100 μg/ml streptomycin sulfate) in the presence of antigens for 4 days. Each well was pulsed with 0.5 μCi [^{3}H]-Thymidine (Amersham, Buckinghamshire, UK) 16 hours before harvesting. All proliferation assays were performed in triplicate. Results are expressed as the stimulation index (SI), which represents the ratio between the mean c.p.m. obtained in the presence and absence of antigen (10, 11).

SI Figures

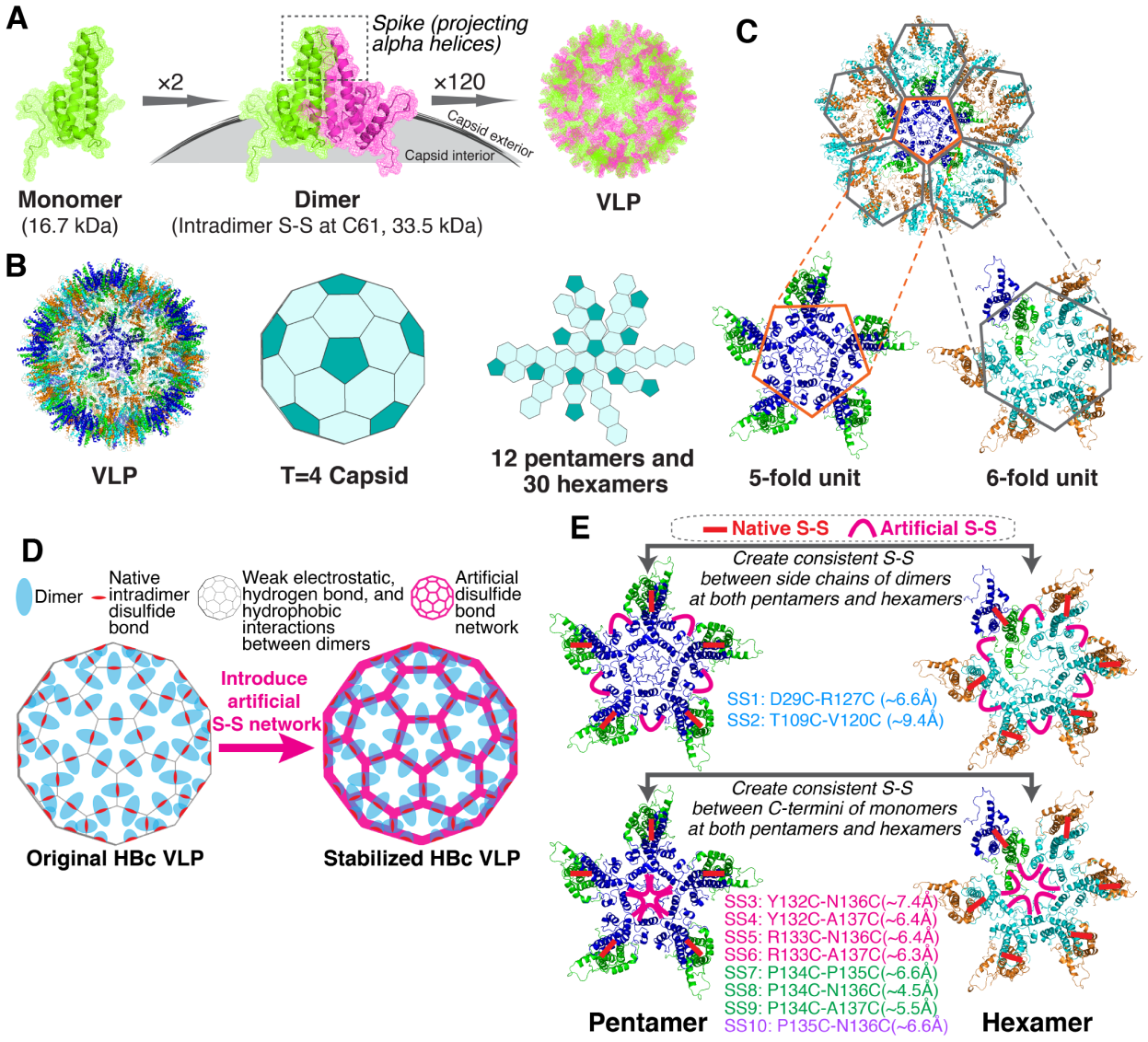


Fig. S1. Illustration of Hepatitis B core protein (HBc) VLP assembly and construction of artificial disulfide network. (A) HBc VLP is formed by self-assembly of 120 dimers. (B) The T=4 capsid structure of Hepatitis B core protein (HBc_{1-149}) VLP constructed from 12 pentamers and 30 hexamers. (C) The structures of the pentamer and the hexamer subunits. (D) Schematic for the introduction of artificial disulfide (S-S) bond network. (E) Selection of possible S-S bonds between side chains of dimers or between C-terminus of monomers for both the pentamers and the hexamers.

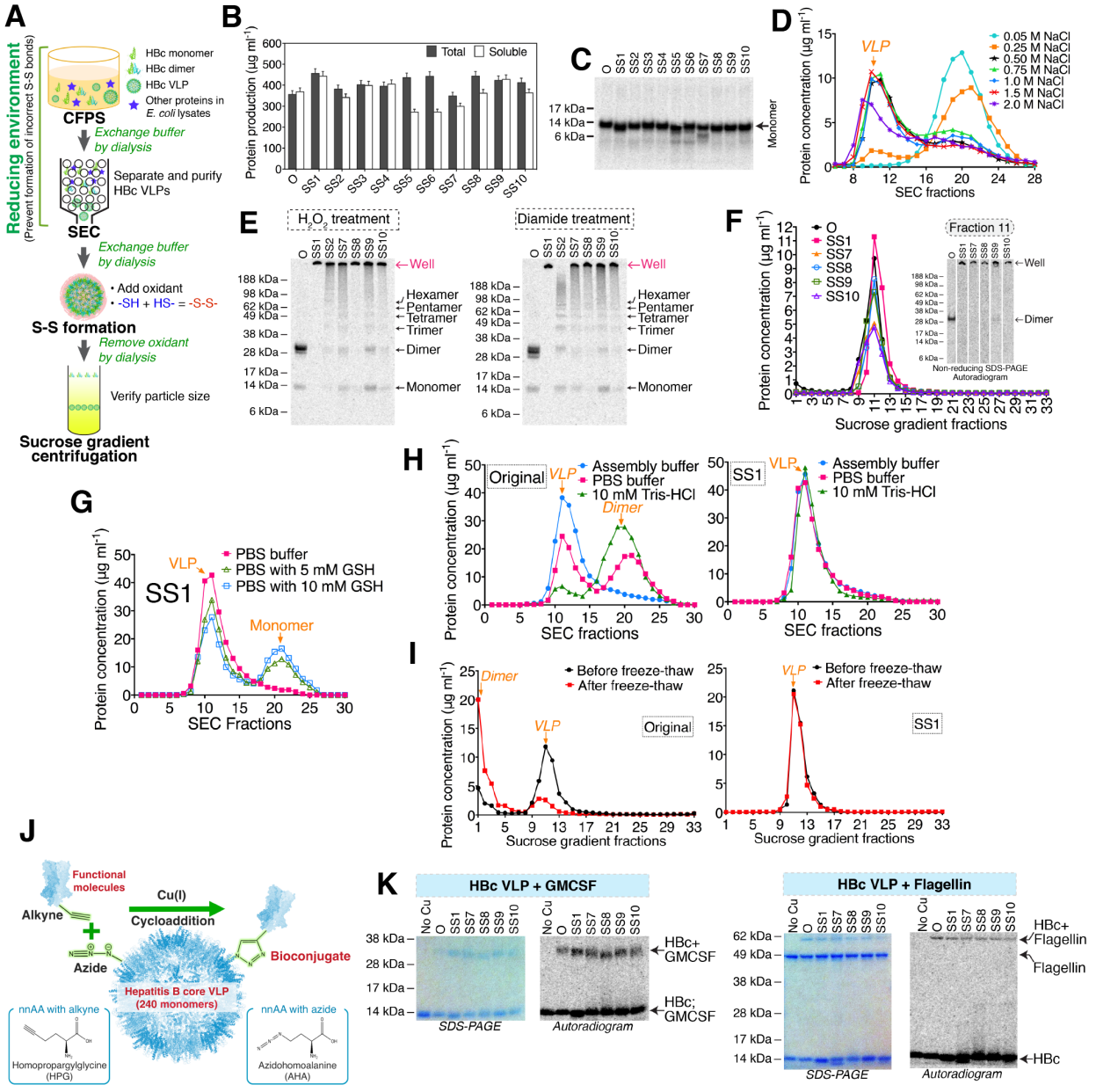


Fig. S2. Stabilization of HBc VLP with artificial disulfide bonds. (A) The procedure for the HBc protein synthesis, VLP self-assembly, VLP purification, S-S bond formation and the verification of correct S-S bonding. The dialysis buffer: 10 mM Tris-HCl(pH 7.4), 0.5 M NaCl. (B) CFPS yields of orginal HBc protein and mutants. (C) Autoradiogram analysis of non-reducing SDS-PAGE. (D) Self-assembly analysis of HBc VLP in the 10 mM Tris-HCl (pH 7.4) buffer with different NaCl concentrations using size-exclusion chromatography (SEC). (E) The non-reducing SDS-PAGE and the autoradiogram analysis after the oxidization treatment of purified VLPs. The

SEC fractions 9-11 were pooled as the purified VLPs. Hydrogen peroxide and diamide were used as oxidants. (F) Sucrose gradient centrifugation analysis of original HBc VLP and oxidized SS1, SS7, SS8, SS9 and SS10 VLPs. (G) Size-exclusion chromatography (SEC) analysis of HBc Original SS1 VLPs after treatment with GSH. GSH is commonly the most abundant low molecular mass thiol in animal and plant cells. Intracellular GSH concentrations usually range from 0.5 to 10 mM.To demonstrate the selective decomposition of the disulfide-stabilized capsid in the reducing intracellular environment, we dialyzed the disulfide-stabilized HBc Original SS1 VLP against PBS buffer with 5 mM or 10 mM GSH for 12 hours. The VLPs were then analyzed by size-exclusion chromatography (SEC). The SEC running buffer was the same as the dialysis buffer. The SEC analysis showed that around 35% to 45% of VLP proteins disassembled after treatment with 5 mM or 10 mM GSH in physiological PBS buffer. Most likely, a higher percentage became leaky. This demonstration experiment proved that the S-S bonds in HBc VLP could confer conditional stability, which potentially could make the VLP open under the reducing environment inside the cell for delivery applications. (H) SEC analysis of purified VLPs in different running buffers. (I) Sucrose gradient centrifugation analysis of HBc VLPs after one freeze-thaw cycle. (J) Diagram of Cu(I)-catalyzed [3 + 2] cycloaddition click chemistry reaction for the direct coupling of functional molecules to HBc VLPs. (K) Reducing SDS-PAGE analysis of click-reaction products. Cu(I) was not added to the reaction as the control. HBc VLPs (monomer: 16.7 kDa) were radioactive. Flagellin (52.7 kDa) and GMCSF (16.1 kDa) were not radioactive.

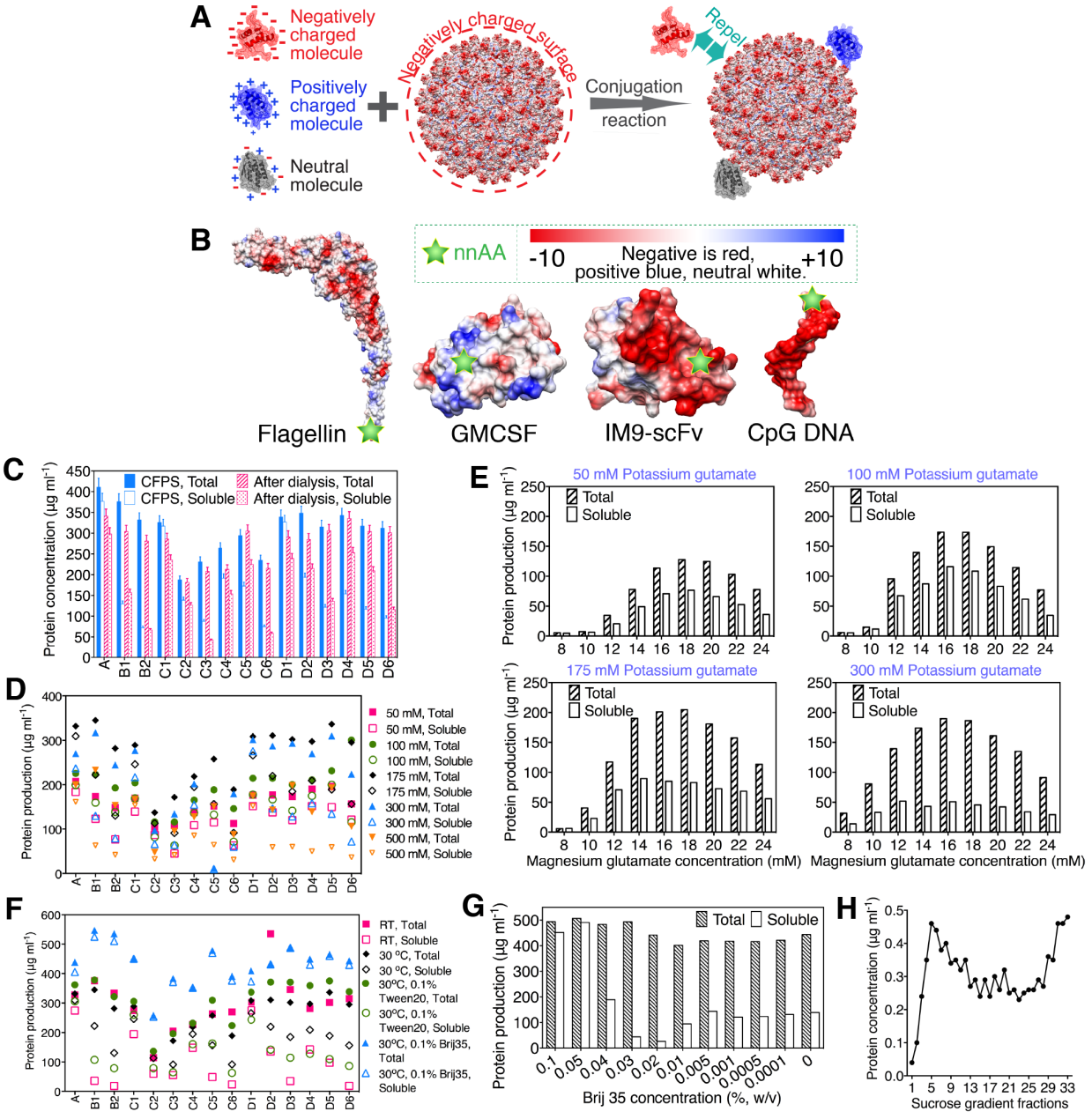


Fig. S3. Optimization of CFPS conditions for improving the soluble CFPS yields and the VLP assembly of HBc mutants. (A) Illustration of HBc VLP conjugation with desired surface additions. Negatively charged molecules are less suitable for conjugation to this VLP. (B) The surface charge distribution of the four attachment molecules at physiological pH and the position of nnAAs. (C) The CFPS yields and soluble yields after dialysis against buffer with 0.5 M NaCl. (D) Effects of potassium glutamate concentration on the CFPS yields of different mutants. The potassium glutamate concentration in standard CFPS system is 175 mM. (E) Effects of Mg^{2+} concentration on the CFPS yield of the mutant B1 (D78AHA) at different potassium glutamate

concentrations. The Mg^{2+} concentration in standard CFPS system is 16 mM. (F) Effects of temperature and the addition of detergent on the CFPS yields of different mutants. Addition of the detergent Brij 35 improved the soluble yield greatly. (G) Effects of Brij 35 concentration on the CFPS yield of the mutant B1 (D78AHA). The HBc protein was completely soluble when the Brij 35 concentration in the CFPS system was above 0.05% (w/v). (H) Sucrose gradient centrifugation (SGC) analysis of HBc protein (mutant B1 (D78AHA)) from the CFPS system with 0.05% (w/v) Brij 35. The results showed that Brij 35 improved the soluble yield but disrupted the VLP assembly.

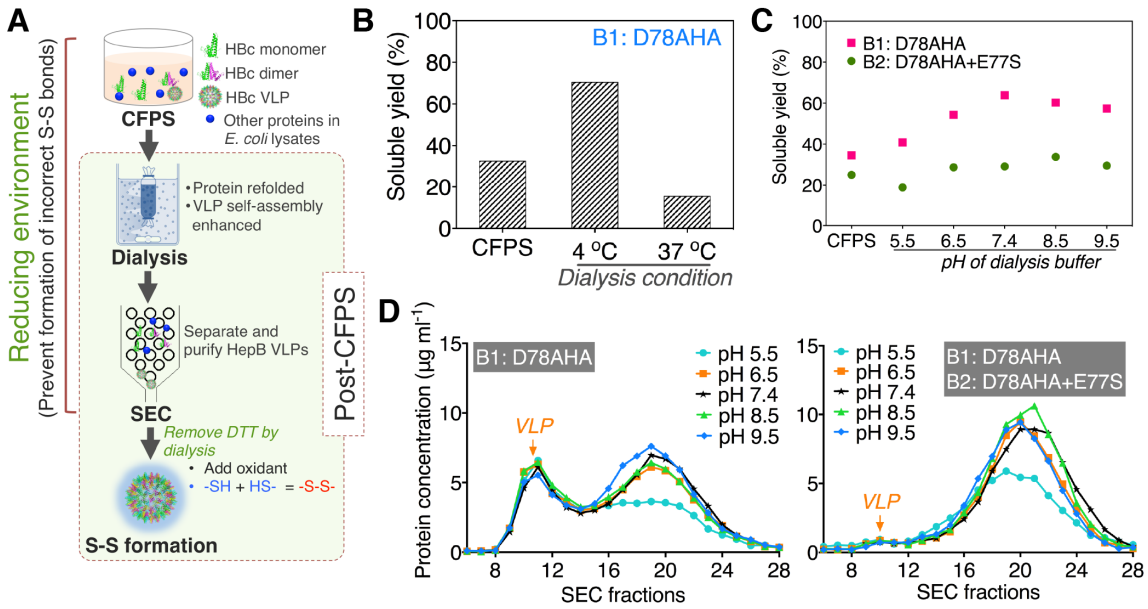


Fig. S4. Optimization of dialysis conditions for improving the soluble yield and the VLP assembly of HBc mutants. (A) The procedure for the CFPS, dialysis, SEC and S-S bond formation. The dialysis after CFPS is a key step for the HBc protein folding and VLP assembly. The dialysis buffer is 10 mM Tris-HCl, pH 7.4, 0.5 M NaCl. (B) Effects of dialysis temperature on the soluble yield of the mutant B1 (D78AHA). The standard temperature in the dialysis step is 4 ^{o}C. Higher dialysis temperature (37 ^{o}C) decreased the solubility of HBc protein. (C) Effects of pH of dialysis buffer on the soluble yield of the mutant B1 (D78AHA) and B2 (D78AHA+E77S). The standard pH of the dialysis buffer is 7.4. The pH 7.4 was still the best. (D) SEC analysis of the CFPS product after the dialysis at different pH values. The SEC running buffer was the same as the dialysis buffer. pH did not affect the VLP assembly.

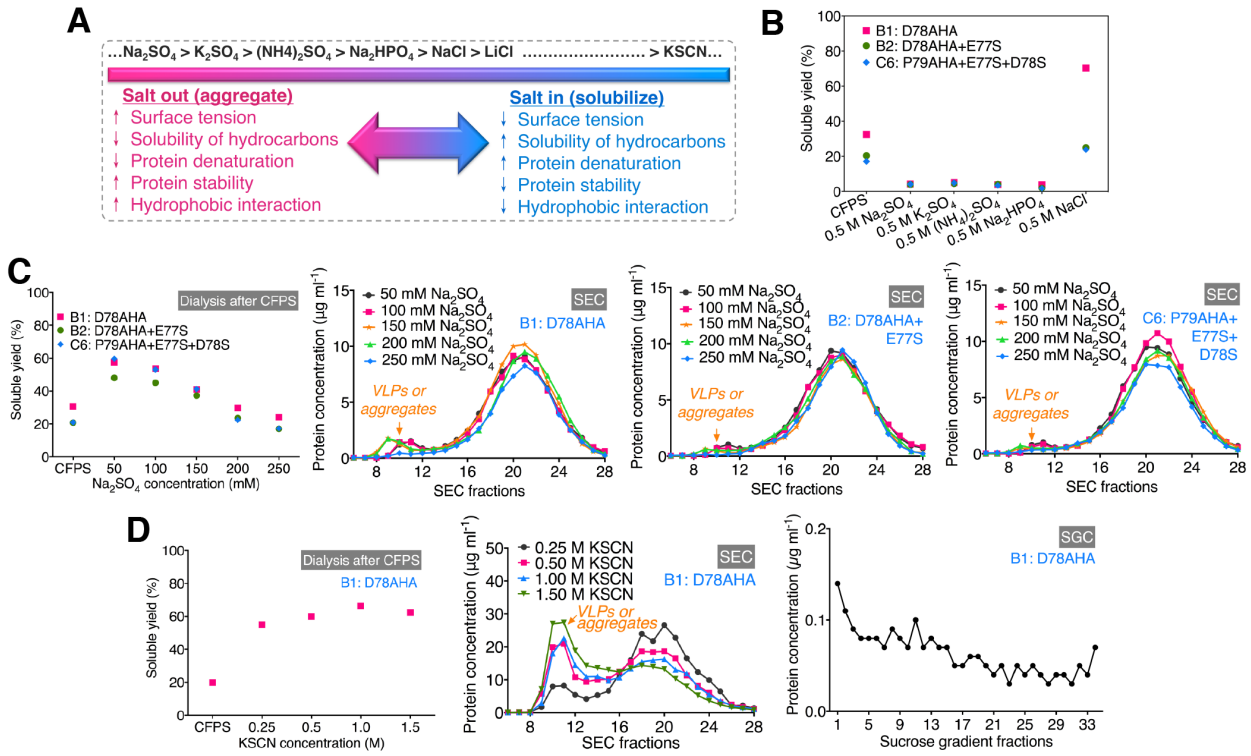


Fig. S5. Effects of salts on the solubility and the VLP assembly of HBc mutants (B1 (D78AHA), B2 (D78AHA+E77S), and C6 (P79AHA+E77S+D78S)). We used 10 mM Tris-HCl (pH 7.4) buffer with different salts in the dialysis step after CFPS. (A) Salts in Hofmeister series. (B) Effects of different salts (Na_2SO_4, K_2SO_4, $(NH_4)_2SO_4$, Na_2HPO_4 and NaCl) on the solubility of HBc mutants. (C) Effects of different concentrations of Na_2SO_4. The analyses of solubility after dialysis and size-exclusion chromatography (SEC) showed that Na_2SO_4 could improve the solubility a little but disrupt the VLP assembly. (D) Effects of different concentrations of KSCN. The analyses of solubility after dialysis, size-exclusion chromatography (SEC), and sucrose gradient centrifugation (SGC) showed that KSCN denatured the HBc antigen protein.

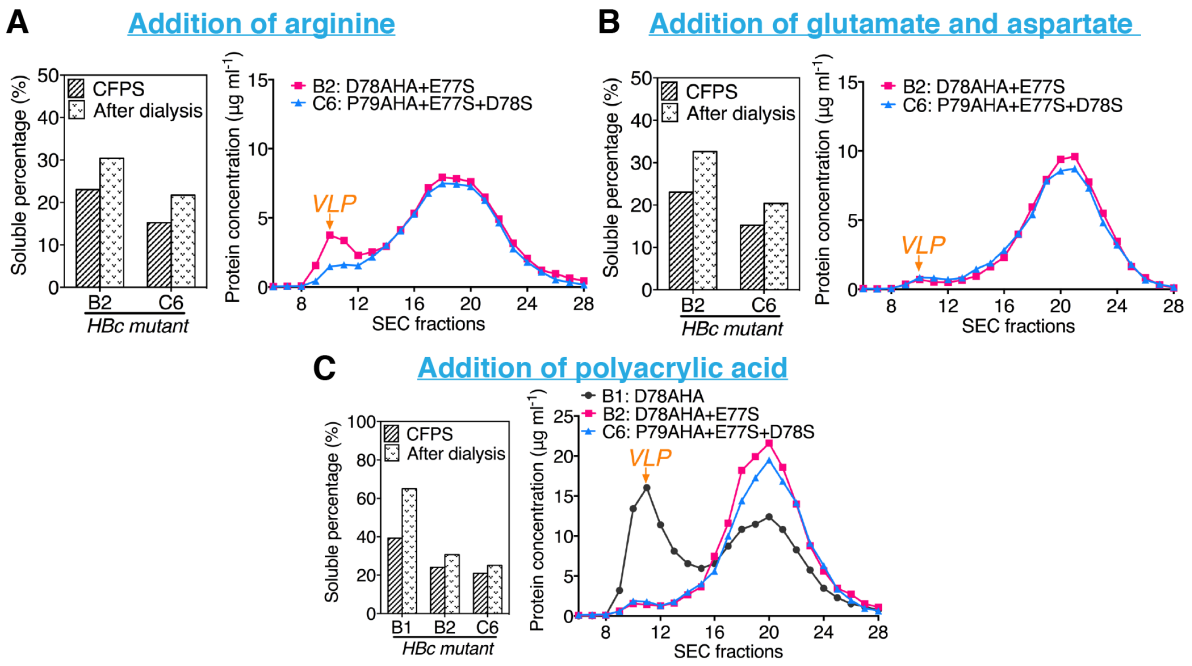


Fig. S6. Effects of amino acid salts and polyacrylic acid (PAA) on the solubility and the VLP assembly of HBc mutants (B1 (D78AHA), B2 (D78AHA+E77S), and C6 (P79AHA+E77S+D78S)). We used 10 mM Tris-HCl (pH 7.4) buffer with different salts in the dialysis step after CFPS. (A) Effects of argnine. The arginine (0.5 M) was added in the dialysis buffer and the pH was adjusted to 7.4. When pH is 7.4, arginine is positively charged, which might affect the assembly of HBc VLP. (B) Effects of salts of glutamic acid and aspartic acid. The glutamic acid (5 mM) and aspartic acid (5 mM) were added in the dialysis buffer and the pH was adjusted to 7.4. When pH is 7.4, glutamic acid and aspartic acid negatively charged, which might affect the assembly of HBc VLP. (C) Effects of PAA. PAA is thought to be able to induce the HBc VLP assembly (Newman *et al.*, 2009). The PAA (1 g/L) was added in the dialysis buffer and the pH was adjusted to 7.4. The analyses of solubility after dialysis and size-exclusion chromatography (SEC) showed that addition of amino acid salts and PAA almost had no effect on the solubility and the VLP assembly of HBc mutants.

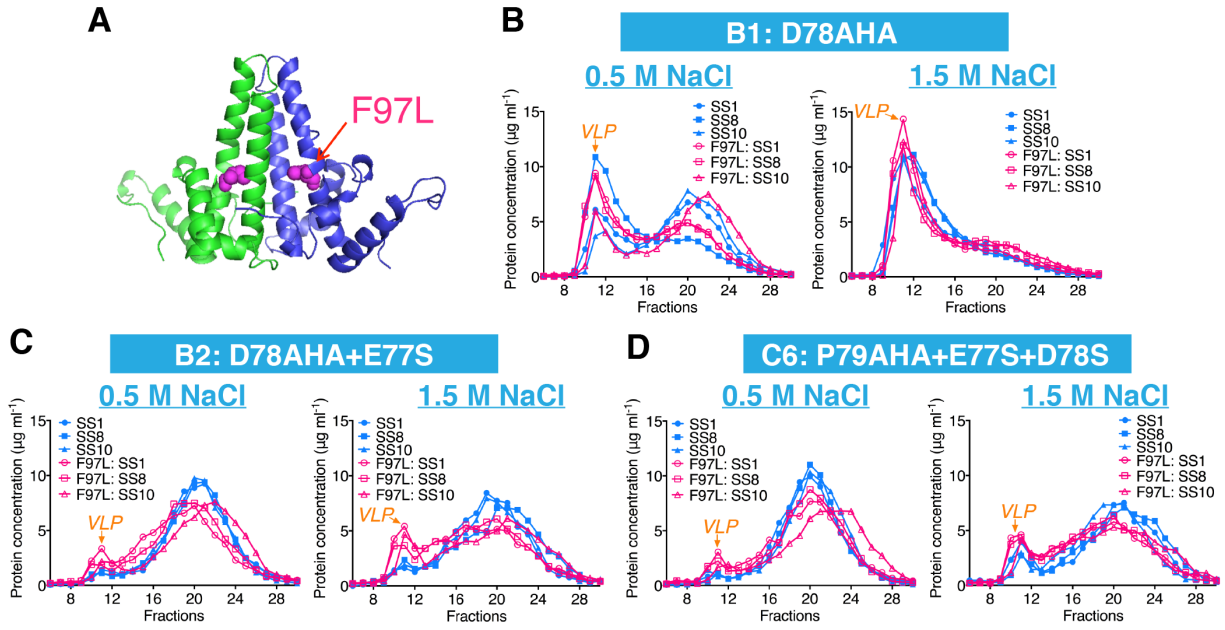
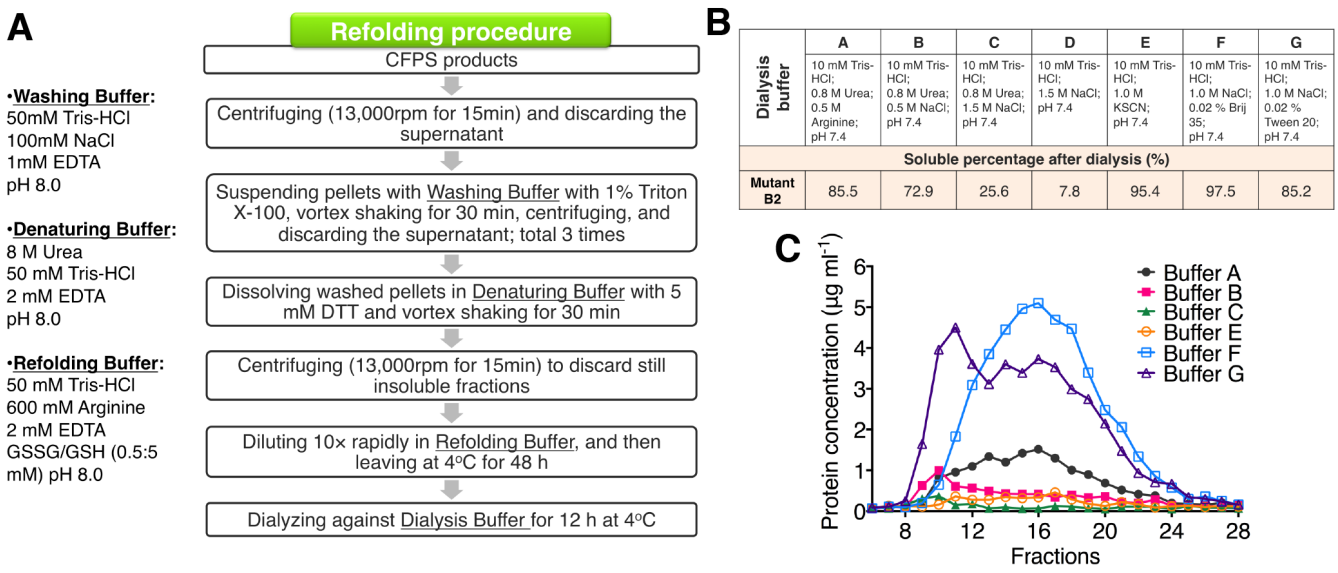


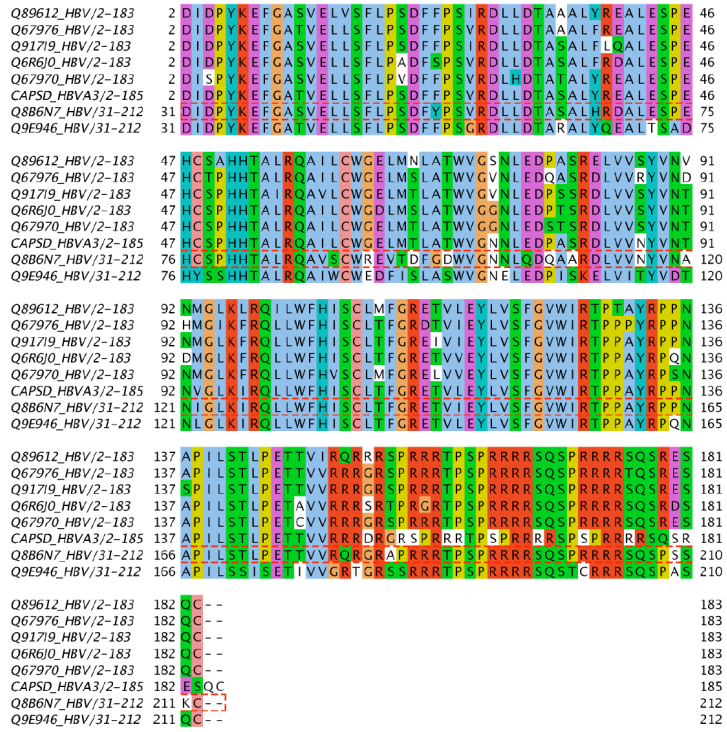
Fig. S7. Effects of mutation F97L and different disulfide bond networks on the solubility and the VLP assembly of HBc mutants (B1 (D78AHA), B2 (D78AHA+E77S), and C6 (P79AHA+E77S+D78S)). We used 10 mM Tris-HCl (pH 7.4) buffer with two different NaCl concentrations (0.5 M and 1.5 M) in the dialysis step after CFPS. Three different disulfide bond networks were tried, including SS1 (D29C-R127C), SS8 (P134C-N136C) and SS10 (P135C-N136C). (A) Position of mutation F97L in the HBc dimer. In chronic hepatitis B virus (HBV) infections, one of the most common mutations to the virus occurs at amino acid 97, where leucine (L) replaces phenylalanine (F). Residue 97 is located in a hydrophobic pocket in the middle of the four-helix. Its mutation could affect virus assembly thermodynamics and kinetics (Ceres *et al.*, 2004). (B) SEC analysis for the mutant b1 with newly introduced mutations. (C) SEC analysis for the mutant b2 with newly introduced mutations. (D) SEC analysis for the mutant c6 with newly introduced mutations. The SEC analysis results demonstrated that introduction of F97L and different S-S bonds did not improve the VLP assembly of mutants B2 and C6. However, the mutation F97L could improve the VLP assembly of mutant B1 a little.



Dialysis buffer	A	B	C	D	E	F	G
	10 mM Tris-HCl; 0.8 M Urea; 0.5 M Arginine; pH 7.4	10 mM Tris-HCl; 0.8 M Urea; 0.5 M NaCl; pH 7.4	10 mM Tris-HCl; 0.8 M Urea; 1.5 M NaCl; pH 7.4	10 mM Tris-HCl; 1.5 M NaCl; pH 7.4	10 mM Tris-HCl; 1.0 M KSCN; pH 7.4	10 mM Tris-HCl; 1.0 M NaCl; 0.02 % Brij 35; pH 7.4	10 mM Tris-HCl; 1.0 M NaCl; 0.02 % Tween 20; pH 7.4
Soluble percentage after dialysis (%)							
Mutant B2	85.5	72.9	25.6	7.8	95.4	97.5	85.2

Fig. S8. Refolding of insoluble HBc protein after CFPS reaction. (A) The refolding procedure of mutant B2 (D78AHA+E77S). (B) The soluble percentage in the final 7 different dialysis buffers after protein refolding. (C) The SEC analysis of soluble refolded proteins. The running buffers were the same as the dialysis buffers. The results showed that the protein refolding did not work.

A



B

Number	Name in the seed alignment	UniProt accession number	Host source
1	Q89612_HBV/2-183	Q89612	Homo sapiens (Human) or Pan troglodytes (Chimpanzee)
2	Q67976_HBV/2-183	Q67976	
3	Q9171I9_HBV/2-183	Q9171I9	
4	Q6R6J0_HBV/2-183	Q6R6J0	
5	Q67970_HBV/2-183	Q67970	
6	CAPSD_HBVA3/2-185	P03148	
7	Q8B6N7_HBV/31-212	Q8B6N7	
8	Q9E946_HBV/31-212	Q9E946	
9	HBEAG_GSHV/32-217	P0C6J1	Spermophilus (old world ground squirrels)
10	HBEAG_HHBV/45-305	P0C6K0	Ardeidae (herons)
11	Q6RSG9_9HEPA/45-305	Q6RSG9	Ross's goose
12	Q67849_9HEPA/2-262	Q67849	Anas (ducks)

C

Natural HBc mutants	Mutation at position 77 in natural mutants			
	E→D	E→Q	E→A	E→K
UniProt accession number	Q4R1S8; P0C697	P0C692; O91532; Q4R1S0; O92920; P03149; P0C696; P0C698; Q9QBF2; P0C677; Q8AZ62; Q9WMW8; D2U608; D2U612; Q8B6N7	Q9WMB7; P0C6H4; P0C6I3	Q507F9

Fig. S9. Sequence alignment of the Hepatitis core proteins. (A) Sequence alignment of the Hepatitis core using the Hepatitis core protein family PF00906 from Pfam protein domain database. To identify the natural mutants, we examined a protein sequence alignment of the Hepatitis core using the protein family PF00906 from Pfam protein domain database (Punta *et al.*, 2012). We started with the seed alignment which contained 12 members and removed 4 members not from human host. Of the remaining 8 sequences, none of the seed sequences had both negative charges removed (E77, D78) but one sequence, Q8B6N7 had a mutation at E77. The Q8B6N7 mutant was fairly novel. Comparing it to the rest of around 7000 members of PF00906, it was the only one with its series of mutations around the spike domain. (B) Twelve sequences for the seed alignment of the Hepatitis core protein sequences using the Hepatitis core family PF00906 from Pfam protein domain database. (C) Natural HBc protein mutants with mutations at position 77. The amino acid at the position 77 in original HBc protein is Glutamic acid (E). The net negative charges on the dimer spike are from E77 and D78. D78 is conserved in all natural HBc mutants.

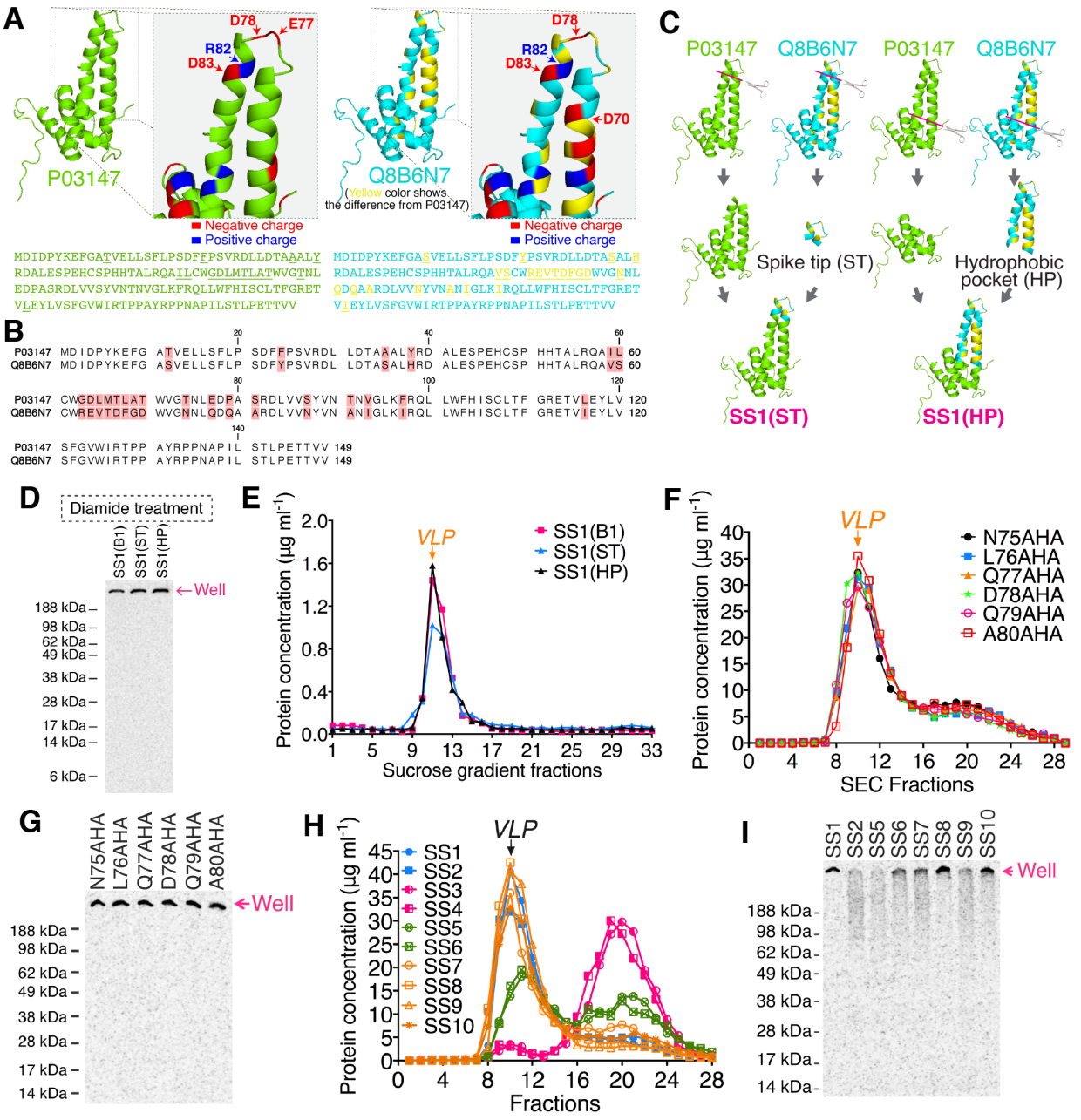


Fig. S10. Transplant of the spike region of natural mutant Q8B6N7 to P03147. (A) The difference between the HBc protein used in this study (P03147) and a natural mutant (Q8B6N7). The sequence differences are underlined and marked in yellow. (B) The alignment of protein sequences (P03147 and Q8B6N7). The differences are highlighted in pink. HBc protein (UniProt accession number: P03147) was originally used in this study. (C) Illustration for the creation of two new mutants (SS1(ST) and SS1(HP)). (D) The non-reducing SDS-PAGE and the autoradiogram analysis after the oxidization treatment of purified VLPs. Diamide was used as the oxidant. (E) Sucrose gradient centrifugation analysis of oxidized VLPs. (F) SEC analysis

after dialysis against buffer with 1.5 M NaCl. Six different nnAA sites (N75AHA, L76AHA, Q77AHA, D78AHA, Q79AHA, and A80AHA) were selected on the spike tip of mutant HBc(HP). (G) The non-reducing SDS-PAGE and the autoradiogram analysis after the oxidization treatment of purified VLPs. Diamide was used as the oxidant. (H) Synthesis of HBc HP 78AHA VLP with disulfide bridges. Ten positions were selected and compared, including SS1 (D29C-R127C), SS2 (T109C-V120C), SS3 (Y132C-N136C), SS4 (Y132C-A137C), SS5 (R133C-N136C), SS6 (R133C-A137C), SS7 (P134C-P135C), SS8 (P134C-N136C), SS9 (P134C-A137C), SS10 (P135C-N136C). Separation of VLPs by size-exclusion chromatography (SEC) using 10 mM Tris-HCl buffer (pH 7.4) with 1.5 M NaCl. The mutants SS3, SS4, SS5, and SS6 might affect the interactions between HBc monomer or dimers, so that they could not self-assemble into VLPs very well. (I) The non-reducing SDS-PAGE and the autoradiogram analysis after the oxidization treatment of purified VLPs. The SEC fractions 8-13 were pooled as the purified VLPs. Diamide was used as the oxidant. Only SS1 particles all stayed in the well of SDS-PAGE gel, which demonstrated that all S-S bonds were formed in SS1 VLPs. Based on these results, mutant SS1 appeared to be the best. Mutant SS8 was also good.

SI Tables

Table S1. The sequences of 10 different HBc variants with different cysteine mutations.

Variants	Protein sequence	DNA encoding sequence
Wild-type (HBc149)	MDIDPYKEFGATVELLSFLP SDFFPSVRDLLDTAAALYRD ALESPEHCSPHHTALRQAIL CWGDLMTLATWVGTNLEDPA SRDLVVSYVNTNVGLKFRQL LWFHISCLTFGRETVLEYLV SFGVWIRTPPAYRPPNAPIL STLPETTVV	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAACTGCTGTCTTTCC TGCCGTCTGATTTCTTCCCGTCTGTTCGTGACCTGCTGGACACCGCGGCAGCACTG TACCGTGACGCGCTGGAATCTCCGGAACATTGTTCTCCGCATCACACTGCGCTGC GTCAGGCGATTCTGTGCTGGGGCGACCTGATGACCCTGGCGACTTGGGTTGGCAC CAACCTGGAAGATCCGGCGTCTCGTGATCTGGTTGTTTCTTACGTTAACACTAACG TTGGTCTGAAATTCCGTCAGCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGGT CGTGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGATTCGTACTCCGCC GGCTTACCGTCCGCCGAACGCACCGATCCTGAGCACCCTGCCGGAAACCACTGTT GTGTAATAA
Original (M66S-L76M)	MDIDPYKEFGATVELLSFLP SDFFPSVRDLLDTAAALYRD ALESPEHCSPHHTALRQAIL CWGDLSTLATWVGTNMEDPA SRDLVVSYVNTNVGLKFRQL LWFHISCLTFGRETVLEYLV SFGVWIRTPPAYRPPNAPIL STLPETTVV	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAACTGCTGTCTTTCC TGCCGTCTGATTTCTTCCCGTCTGTTCGTGACCTGCTGGACACCGCGGCAGCACTG TACCGTGACGCGCTGGAATCTCCGGAACATTGTTCTCCGCATCACACTGCGCTGC GTCAGGCGATTCTGTGCTGGGGCGACCTGagcACCCTGGCGACTTGGGTTGGCACC AACatgGAAGATCCGGCGTCTCGTGATCTGGTTGTTTCTTACGTTAACACTAACGTT GGTCTGAAATTCCGTCAGCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGGTCG TGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGATTCGTACTCCGCCGG CTTACCGTCCGCCGAACGCACCGATCCTGAGCACCCTGCCGGAAACCACTGTTGT GTAATAA
SS1: D29C-R127C	MDIDPYKEFGATVELLSFLP SDFFPSVRCLLDTAAALYRD ALESPEHCSPHHTALRQAIL CWGDLSTLATWVGTNMEDPA SRDLVVSYVNTNVGLKFRQL LWFHISCLTFGRETVLEYLV SFGVWICTPPAYRPPNAPIL STLPETTVV	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAACTGCTGTCTTTCC TGCCGTCTGATTTCTTCCCGTCTGTTCGTTGCCTGCTGGACACCGCGGCAGCACTG TACCGTGACGCGCTGGAATCTCCGGAACATTGTTCTCCGCATCACACTGCGCTGC GTCAGGCGATTCTGTGCTGGGGCGACCTGagcACCCTGGCGACTTGGGTTGGCACC AACatgGAAGATCCGGCGTCTCGTGATCTGGTTGTTTCTTACGTTAACACTAACGTT GGTCTGAAATTCCGTCAGCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGGTCG TGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGATTTGTACTCCGCCGG CTTACCGTCCGCCGAACGCACCGATCCTGAGCACCCTGCCGGAAACCACTGTTGT GTAATAA
SS2: T109C-V120C	MDIDPYKEFGATVELLSFLP SDFFPSVRDLLDTAAALYRD ALESPEHCSPHHTALRQAIL CWGDLSTLATWVGTNMEDPA SRDLVVSYVNTNVGLKFRQL LWFHISCLCFGRETVLEYLC SFGVWIRTPPAYRPPNAPIL STLPETTVV	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAACTGCTGTCTTTCC TGCCGTCTGATTTCTTCCCGTCTGTTCGTGACCTGCTGGACACCGCGGCAGCACTG TACCGTGACGCGCTGGAATCTCCGGAACATTGTTCTCCGCATCACACTGCGCTGC GTCAGGCGATTCTGTGCTGGGGCGACCTGagcACCCTGGCGACTTGGGTTGGCACC AACatgGAAGATCCGGCGTCTCGTGATCTGGTTGTTTCTTACGTTAACACTAACGTT GGTCTGAAATTCCGTCAGCTGCTGTGGTTCCACATCTCTTGCCTGTGCTTCGGTCG TGAAACCGTTCTGGAATACCTGTGTTCTTTTGGTGTTTGGATTCGTACTCCGCCGG CTTACCGTCCGCCGAACGCACCGATCCTGAGCACCCTGCCGGAAACCACTGTTGT GTAATAA
SS3: Y132C-N136C	MDIDPYKEFGATVELLSFLP SDFFPSVRDLLDTAAALYRD ALESPEHCSPHHTALRQAIL CWGDLSTLATWVGTNMEDPA SRDLVVSYVNTNVGLKFRQL LWFHISCLTFGRETVLEYLV SFGVWIRTPPACRPPCAPIL STLPETTVV	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAACTGCTGTCTTTCC TGCCGTCTGATTTCTTCCCGTCTGTTCGTGACCTGCTGGACACCGCGGCAGCACTG TACCGTGACGCGCTGGAATCTCCGGAACATTGTTCTCCGCATCACACTGCGCTGC GTCAGGCGATTCTGTGCTGGGGCGACCTGagcACCCTGGCGACTTGGGTTGGCACC AACatgGAAGATCCGGCGTCTCGTGATCTGGTTGTTTCTTACGTTAACACTAACGTT GGTCTGAAATTCCGTCAGCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGGTCG TGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGATTCGTACTCCGCCGG CTTGCCGTCCGCCGTGCGCACCGATCCTGAGCACCCTGCCGGAAACCACTGTTGT GTAATAA
SS4: Y132C-A137C	MDIDPYKEFGATVELLSFLP SDFFPSVRDLLDTAAALYRD ALESPEHCSPHHTALRQAIL CWGDLSTLATWVGTNMEDPA SRDLVVSYVNTNVGLKFRQL LWFHISCLTFGRETVLEYLV SFGVWIRTPPACRPPNCPIL STLPETTVV	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAACTGCTGTCTTTCC TGCCGTCTGATTTCTTCCCGTCTGTTCGTGACCTGCTGGACACCGCGGCAGCACTG TACCGTGACGCGCTGGAATCTCCGGAACATTGTTCTCCGCATCACACTGCGCTGC GTCAGGCGATTCTGTGCTGGGGCGACCTGagcACCCTGGCGACTTGGGTTGGCACC AACatgGAAGATCCGGCGTCTCGTGATCTGGTTGTTTCTTACGTTAACACTAACGTT GGTCTGAAATTCCGTCAGCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGGTCG TGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGATTCGTACTCCGCCGG CTTGCCGTCCGCCGAACTGCCCGATCCTGAGCACCCTGCCGGAAACCACTGTTGT GTAATAA
SS5: R133C-N136C	MDIDPYKEFGATVELLSFLP SDFFPSVRDLLDTAAALYRD ALESPEHCSPHHTALRQAIL CWGDLSTLATWVGTNMEDPA SRDLVVSYVNTNVGLKFRQL LWFHISCLTFGRETVLEYLV SFGVWIRTPPAYCPPCAPIL STLPETTVV	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAACTGCTGTCTTTCC TGCCGTCTGATTTCTTCCCGTCTGTTCGTGACCTGCTGGACACCGCGGCAGCACTG TACCGTGACGCGCTGGAATCTCCGGAACATTGTTCTCCGCATCACACTGCGCTGC GTCAGGCGATTCTGTGCTGGGGCGACCTGagcACCCTGGCGACTTGGGTTGGCACC AACatgGAAGATCCGGCGTCTCGTGATCTGGTTGTTTCTTACGTTAACACTAACGTT GGTCTGAAATTCCGTCAGCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGGTCG TGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGATTCGTACTCCGCCGG CTTACTGCCCGCCGTGCGCACCGATCCTGAGCACCCTGCCGGAAACCACTGTTGT GTAATAA

SS6: R133C-A137C	MDIDPYKEFGATVELLSFLP SDFFPSVRDLLDTAAALYRD ALESPEHCSPHHTALRQAIL CWGDLSTLATWVGTNMEDPA SRDLVVSYVNTNVGLKFRQL LWFHISCLTFGRETVLEYLV SFGVWIRTPPAYCPPNCPIL STLPETTVV	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAACTGCTGTCTTTCC TGCCGTCTGATTTCTTCCCGTCTGTTCGTGACCTGCTGGACACCGCGGCAGCACTG TACCGTGACGCGCTGGAATCTCCGGAACATTGTTCTCCGCATCACACTGCGCTGC GTCAGGCGATTCTGTGCTGGGGCGACCTGagcACCCTGGCGACTTGGGTTGGCACC AACatgGAAGATCCGGCGTCTCGTGATCTGGTTGTTTCTTACGTTAACACTAACGTT GGTCTGAAATTCCGTCAGCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGGTCG TGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGATTCGTACTCCGCCGG CTTACTGCCCGCCGAACTGCCCGATCCTGAGCACCCTGCCGGAAACCACTGTTGT GTAATAA
SS7: P134C-P135C	MDIDPYKEFGATVELLSFLP SDFFPSVRDLLDTAAALYRD ALESPEHCSPHHTALRQAIL CWGDLSTLATWVGTNMEDPA SRDLVVSYVNTNVGLKFRQL LWFHISCLTFGRETVLEYLV SFGVWIRTPPAYRCCNAPIL STLPETTVV	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAACTGCTGTCTTTCC TGCCGTCTGATTTCTTCCCGTCTGTTCGTGACCTGCTGGACACCGCGGCAGCACTG TACCGTGACGCGCTGGAATCTCCGGAACATTGTTCTCCGCATCACACTGCGCTGC GTCAGGCGATTCTGTGCTGGGGCGACCTGagcACCCTGGCGACTTGGGTTGGCACC AACatgGAAGATCCGGCGTCTCGTGATCTGGTTGTTTCTTACGTTAACACTAACGTT GGTCTGAAATTCCGTCAGCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGGTCG TGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGATTCGTACTCCGCCGG CTTACCGTTGCTGCAACGCACCGATCCTGAGCACCCTGCCGGAAACCACTGTTGT GTAATAA
SS8: P134C-N136C	MDIDPYKEFGATVELLSFLP SDFFPSVRDLLDTAAALYRD ALESPEHCSPHHTALRQAIL CWGDLSTLATWVGTNMEDPA SRDLVVSYVNTNVGLKFRQL LWFHISCLTFGRETVLEYLV SFGVWIRTPPAYRCPCAPIL STLPETTVV	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAACTGCTGTCTTTCC TGCCGTCTGATTTCTTCCCGTCTGTTCGTGACCTGCTGGACACCGCGGCAGCACTG TACCGTGACGCGCTGGAATCTCCGGAACATTGTTCTCCGCATCACACTGCGCTGC GTCAGGCGATTCTGTGCTGGGGCGACCTGagcACCCTGGCGACTTGGGTTGGCACC AACatgGAAGATCCGGCGTCTCGTGATCTGGTTGTTTCTTACGTTAACACTAACGTT GGTCTGAAATTCCGTCAGCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGGTCG TGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGATTCGTACTCCGCCGG CTTACCGTTGCCCGTGCGCACCGATCCTGAGCACCCTGCCGGAAACCACTGTTGT GTAATAA
SS9: P134C-A137C	MDIDPYKEFGATVELLSFLP SDFFPSVRDLLDTAAALYRD ALESPEHCSPHHTALRQAIL CWGDLSTLATWVGTNMEDPA SRDLVVSYVNTNVGLKFRQL LWFHISCLTFGRETVLEYLV SFGVWIRTPPAYRCPNCPIL STLPETTVV	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAACTGCTGTCTTTCC TGCCGTCTGATTTCTTCCCGTCTGTTCGTGACCTGCTGGACACCGCGGCAGCACTG TACCGTGACGCGCTGGAATCTCCGGAACATTGTTCTCCGCATCACACTGCGCTGC GTCAGGCGATTCTGTGCTGGGGCGACCTGagcACCCTGGCGACTTGGGTTGGCACC AACatgGAAGATCCGGCGTCTCGTGATCTGGTTGTTTCTTACGTTAACACTAACGTT GGTCTGAAATTCCGTCAGCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGGTCG TGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGATTCGTACTCCGCCGG CTTACCGTTGCCCGAACTGCCCGATCCTGAGCACCCTGCCGGAAACCACTGTTGT GTAATAA
SS10: P135C-N136C	MDIDPYKEFGATVELLSFLP SDFFPSVRDLLDTAAALYRD ALESPEHCSPHHTALRQAIL CWGDLSTLATWVGTNMEDPA SRDLVVSYVNTNVGLKFRQL LWFHISCLTFGRETVLEYLV SFGVWIRTPPAYRPCCAPIL STLPETTVV	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAACTGCTGTCTTTCC TGCCGTCTGATTTCTTCCCGTCTGTTCGTGACCTGCTGGACACCGCGGCAGCACTG TACCGTGACGCGCTGGAATCTCCGGAACATTGTTCTCCGCATCACACTGCGCTGC GTCAGGCGATTCTGTGCTGGGGCGACCTGagcACCCTGGCGACTTGGGTTGGCACC AACatgGAAGATCCGGCGTCTCGTGATCTGGTTGTTTCTTACGTTAACACTAACGTT GGTCTGAAATTCCGTCAGCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGGTCG TGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGATTCGTACTCCGCCGG CTTACCGTCCGTGCTGCGCACCGATCCTGAGCACCCTGCCGGAAACCACTGTTGT GTAATAA

Table S2. The sequences of 15 different HBc variants with different AHA sites or different negative charge mutations.

Variants	DNA encoding sequence
A: L76AHA	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAACTGCTGTCTTTCCTGCCGTCTGATTTCTTCCCGTC TGTTCGTTGCCTGCTGGACACCGCGGCAGCACTGTACCGTGACGCGCTGGAATCTCCGGAACATTGTTCTCCGCAT CACACTGCGCTGCGTCAGGCGATTCTGTGCTGGGGCGACCTGagcACCCTGGCGACTTGGGTTGGCACCAACatgGA AGATCCGGCGTCTCGTGATCTGGTTGTTTCTTACGTTAACACTAACGTTGGTCTGAAATTCCGTCAGCTGCTGTGGT TCCACATCTCTTGCCTGACCTTCGGTCGTGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGATTTGTACT CCGCCGGCTTACCGTCCGCCGAACGCACCGATCCTGAGCACCCTGCCGGAAACCACTGTTGTGTAATAA
B1: D78AHA	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAACTGCTGTCTTTCCTGCCGTCTGATTTCTTCCCGTC TGTTCGTTGCCTGCTGGACACCGCGGCAGCACTGTACCGTGACGCGCTGGAATCTCCGGAACATTGTTCTCCGCAT CACACTGCGCTGCGTCAGGCGATTCTGTGCTGGGGCGACCTGagcACCCTGGCGACTTGGGTTGGCACCAACCTGG AAatgCCGGCGTCTCGTGATCTGGTTGTTTCTTACGTTAACACTAACGTTGGTCTGAAATTCCGTCAGCTGCTGTGGT TCCACATCTCTTGCCTGACCTTCGGTCGTGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGATTTGTACT CCGCCGGCTTACCGTCCGCCGAACGCACCGATCCTGAGCACCCTGCCGGAAACCACTGTTGTGTAATAA
B2: D78AHA+E77S	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAACTGCTGTCTTTCCTGCCGTCTGATTTCTTCCCGTC TGTTCGTTGCCTGCTGGACACCGCGGCAGCACTGTACCGTGACGCGCTGGAATCTCCGGAACATTGTTCTCCGCAT CACACTGCGCTGCGTCAGGCGATTCTGTGCTGGGGCGACCTGagcACCCTGGCGACTTGGGTTGGCACCAACCTGag catgCCGGCGTCTCGTGATCTGGTTGTTTCTTACGTTAACACTAACGTTGGTCTGAAATTCCGTCAGCTGCTGTGGTT CCACATCTCTTGCCTGACCTTCGGTCGTGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGATTTGTACTC CGCCGGCTTACCGTCCGCCGAACGCACCGATCCTGAGCACCCTGCCGGAAACCACTGTTGTGTAATAA

C1: P79AHA	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAACTGCTGTCTTTCCTGCCGTCTGATTTCTTCCCGTC TGTTCGTTGCCTGCTGGACACCGCGGCAGCACTGTACCGTGACGCGCTGGAATCTCCGGAACATTGTTCTCCGCAT CACACTGCGCTGCGTCAGGCGATTCTGTGCTGGGGCGACCTGagcACCCTGGCGACTTGGGTTGGCACCAACCTGG AAGATatgGCGTCTCGTGATCTGGTTGTTTCTTACGTTAACACTAACGTTGGTCTGAAATTCCGTCAGCTGCTGTGGT TCCACATCTCTTGCCTGACCTTCGGTCGTGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGATTTGTACT CCGCCGGCTTACCGTCCGCCGAACGCACCGATCCTGAGCACCCTGCCGGAAACCACTGTTGTGTAATAA
C2: P79AHA+E77K	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAACTGCTGTCTTTCCTGCCGTCTGATTTCTTCCCGTC TGTTCGTTGCCTGCTGGACACCGCGGCAGCACTGTACCGTGACGCGCTGGAATCTCCGGAACATTGTTCTCCGCAT CACACTGCGCTGCGTCAGGCGATTCTGTGCTGGGGCGACCTGagcACCCTGGCGACTTGGGTTGGCACCAAACCTGaaa aGATatgGCGTCTCGTGATCTGGTTGTTTCTTACGTTAACACTAACGTTGGTCTGAAATTCCGTCAGCTGCTGTGGTT CCACATCTCTTGCCTGACCTTCGGTCGTGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGATTTGTACTC CGCCGGCTTACCGTCCGCCGAACGCACCGATCCTGAGCACCCTGCCGGAAACCACTGTTGTGTAATAA
C3: P79AHA+D78K	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAACTGCTGTCTTTCCTGCCGTCTGATTTCTTCCCGTC TGTTCGTTGCCTGCTGGACACCGCGGCAGCACTGTACCGTGACGCGCTGGAATCTCCGGAACATTGTTCTCCGCAT CACACTGCGCTGCGTCAGGCGATTCTGTGCTGGGGCGACCTGagcACCCTGGCGACTTGGGTTGGCACCAACCTGG AAaaaatgGCGTCTCGTGATCTGGTTGTTTCTTACGTTAACACTAACGTTGGTCTGAAATTCCGTCAGCTGCTGTGGTT CCACATCTCTTGCCTGACCTTCGGTCGTGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGATTTGTACTC CGCCGGCTTACCGTCCGCCGAACGCACCGATCCTGAGCACCCTGCCGGAAACCACTGTTGTGTAATAA
C4: P79AHA+E77S	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAACTGCTGTCTTTCCTGCCGTCTGATTTCTTCCCGTC TGTTCGTTGCCTGCTGGACACCGCGGCAGCACTGTACCGTGACGCGCTGGAATCTCCGGAACATTGTTCTCCGCAT CACACTGCGCTGCGTCAGGCGATTCTGTGCTGGGGCGACCTGagcACCCTGGCGACTTGGGTTGGCACCAACCTGag cGATatgGCGTCTCGTGATCTGGTTGTTTCTTACGTTAACACTAACGTTGGTCTGAAATTCCGTCAGCTGCTGTGGTT CCACATCTCTTGCCTGACCTTCGGTCGTGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGATTTGTACTC CGCCGGCTTACCGTCCGCCGAACGCACCGATCCTGAGCACCCTGCCGGAAACCACTGTTGTGTAATAA
C5: P79AHA+D78S	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAACTGCTGTCTTTCCTGCCGTCTGATTTCTTCCCGTC TGTTCGTTGCCTGCTGGACACCGCGGCAGCACTGTACCGTGACGCGCTGGAATCTCCGGAACATTGTTCTCCGCAT CACACTGCGCTGCGTCAGGCGATTCTGTGCTGGGGCGACCTGagcACCCTGGCGACTTGGGTTGGCACCAACCTGG AAagcatgGCGTCTCGTGATCTGGTTGTTTCTTACGTTAACACTAACGTTGGTCTGAAATTCCGTCAGCTGCTGTGGTT CCACATCTCTTGCCTGACCTTCGGTCGTGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGATTTGTACTC CGCCGGCTTACCGTCCGCCGAACGCACCGATCCTGAGCACCCTGCCGGAAACCACTGTTGTGTAATAA
C6: P79AHA+E77S+D78S	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAACTGCTGTCTTTCCTGCCGTCTGATTTCTTCCCGTC TGTTCGTTGCCTGCTGGACACCGCGGCAGCACTGTACCGTGACGCGCTGGAATCTCCGGAACATTGTTCTCCGCAT CACACTGCGCTGCGTCAGGCGATTCTGTGCTGGGGCGACCTGagcACCCTGGCGACTTGGGTTGGCACCAACCTGag cagcatgGCGTCTCGTGATCTGGTTGTTTCTTACGTTAACACTAACGTTGGTCTGAAATTCCGTCAGCTGCTGTGGTTC CACATCTCTTGCCTGACCTTCGGTCGTGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGATTTGTACTCC GCCGGCTTACCGTCCGCCGAACGCACCGATCCTGAGCACCCTGCCGGAAACCACTGTTGTGTAATAA
D1: A80AHA	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAACTGCTGTCTTTCCTGCCGTCTGATTTCTTCCCGTC TGTTCGTTGCCTGCTGGACACCGCGGCAGCACTGTACCGTGACGCGCTGGAATCTCCGGAACATTGTTCTCCGCAT CACACTGCGCTGCGTCAGGCGATTCTGTGCTGGGGCGACCTGagcACCCTGGCGACTTGGGTTGGCACCAACCTGG AAGATCCGatgTCTCGTGATCTGGTTGTTTCTTACGTTAACACTAACGTTGGTCTGAAATTCCGTCAGCTGCTGTGGT TCCACATCTCTTGCCTGACCTTCGGTCGTGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGATTTGTACT CCGCCGGCTTACCGTCCGCCGAACGCACCGATCCTGAGCACCCTGCCGGAAACCACTGTTGTGTAATAA
D2: A80AHA+E77K	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAACTGCTGTCTTTCCTGCCGTCTGATTTCTTCCCGTC TGTTCGTTGCCTGCTGGACACCGCGGCAGCACTGTACCGTGACGCGCTGGAATCTCCGGAACATTGTTCTCCGCAT CACACTGCGCTGCGTCAGGCGATTCTGTGCTGGGGCGACCTGagcACCCTGGCGACTTGGGTTGGCACCAACCTGaa aGATCCGatgTCTCGTGATCTGGTTGTTTCTTACGTTAACACTAACGTTGGTCTGAAATTCCGTCAGCTGCTGTGGTT CCACATCTCTTGCCTGACCTTCGGTCGTGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGATTTGTACTC CGCCGGCTTACCGTCCGCCGAACGCACCGATCCTGAGCACCCTGCCGGAAACCACTGTTGTGTAATAA
D3: A80AHA+D78K	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAACTGCTGTCTTTCCTGCCGTCTGATTTCTTCCCGTC TGTTCGTTGCCTGCTGGACACCGCGGCAGCACTGTACCGTGACGCGCTGGAATCTCCGGAACATTGTTCTCCGCAT CACACTGCGCTGCGTCAGGCGATTCTGTGCTGGGGCGACCTGagcACCCTGGCGACTTGGGTTGGCACCAACCTGG AAaaaCCGatgTCTCGTGATCTGGTTGTTTCTTACGTTAACACTAACGTTGGTCTGAAATTCCGTCAGCTGCTGTGGTT CCACATCTCTTGCCTGACCTTCGGTCGTGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGATTTGTACTC CGCCGGCTTACCGTCCGCCGAACGCACCGATCCTGAGCACCCTGCCGGAAACCACTGTTGTGTAATAA
D4: A80AHA+E77S	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAACTGCTGTCTTTCCTGCCGTCTGATTTCTTCCCGTC TGTTCGTTGCCTGCTGGACACCGCGGCAGCACTGTACCGTGACGCGCTGGAATCTCCGGAACATTGTTCTCCGCAT CACACTGCGCTGCGTCAGGCGATTCTGTGCTGGGGCGACCTGagcACCCTGGCGACTTGGGTTGGCACCAACCTGag cGATCCGatgTCTCGTGATCTGGTTGTTTCTTACGTTAACACTAACGTTGGTCTGAAATTCCGTCAGCTGCTGTGGTT CCACATCTCTTGCCTGACCTTCGGTCGTGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGATTTGTACTC CGCCGGCTTACCGTCCGCCGAACGCACCGATCCTGAGCACCCTGCCGGAAACCACTGTTGTGTAATAA
D5: A80AHA+D78S	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAACTGCTGTCTTTCCTGCCGTCTGATTTCTTCCCGTC TGTTCGTTGCCTGCTGGACACCGCGGCAGCACTGTACCGTGACGCGCTGGAATCTCCGGAACATTGTTCTCCGCAT CACACTGCGCTGCGTCAGGCGATTCTGTGCTGGGGCGACCTGagcACCCTGGCGACTTGGGTTGGCACCAACCTGG AAagcCCGatgTCTCGTGATCTGGTTGTTTCTTACGTTAACACTAACGTTGGTCTGAAATTCCGTCAGCTGCTGTGGTT CCACATCTCTTGCCTGACCTTCGGTCGTGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGATTTGTACTC CGCCGGCTTACCGTCCGCCGAACGCACCGATCCTGAGCACCCTGCCGGAAACCACTGTTGTGTAATAA

D6: A80AHA+E77S+D78S	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAACTGCTGTCTTTCCTGCCGTCTGATTTCTTCCCGTC TGTTCGTTGCCTGCTGGACACCGCGGCAGCACTGTACCGTGACGCGCTGGAATCTCCGGAACATTGTTCTCCGCAT CACACTGCGCTGCGTCAGGCGATTCTGTGCTGGGGCGACCTGagcACCCTGGCGACTTGGGTTGGCACCAACCTGag cagcCCGatgTCTCGTGATCTGGTTGTTTCTTACGTTAACACTAACGTTGGTCTGAAATTCCGTCAGCTGCTGTGGTTC CACATCTCTTGCCTGACCTTCGGTCGTGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGATTTGTACTCC GCCGGCTTACCGTCCGCCGAACGCACCGATCCTGAGCACCCTGCCGGAAACCACTGTTGTGTAATAA

Table S3. The sequences of SS1(b1), SS1(ST), and SS1(HP).

Variants	Protein sequence (with nnAA AHA)	DNA encoding sequence
SS1(B1)	(AHA)DIDPYKEFGATVELLSFLPS DFFPSVRCLLDTAAALYRDALESPE HCSPHHTALRQAILCWGDLSTLATW VGTNLE(AHA)PASRDLVVSYVNTN VGLKFRQLLWFHISCLTFGRETVLE YLVSFGVWICTPPAYRPPNAPILST LPETTVV	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAACTGCTGTCTTTCC TGCCGTCTGATTTCTTCCCGTCTGTTCGTTGCCTGCTGGACACCGCGGCAGCACTG TACCGTGACGCGCTGGAATCTCCGGAACATTGTTCTCCGCATCACACTGCGCTGC GTCAGGCGATTCTGTGCTGGGGCGACCTGagcACCCTGGCGACTTGGGTTGGCACC AACCTGGAAatgCCGGCGTCTCGTGATCTGGTTGTTTCTTACGTTAACACTAACGTT GGTCTGAAATTCCGTCAGCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGGTCG TGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGATTTGTACTCCGCCGG CTTACCGTCCGCCGAACGCACCGATCCTGAGCACCCTGCCGGAAACCACTGTTGT GTAATAA
SS1(ST)	(AHA)DIDPYKEFGATVELLSFLPS DFFPSVRCLLDTAAALYRDALESPE HCSPHHTALRQAILCWGDLSTLATW VGNN(AHA)QDQAARDLVVSYVNTN VGLKFRQLLWFHISCLTFGRETVLE YLVSFGVWICTPPAYRPPNAPILST LPETTVV	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAACTGCTGTCTTTCC TGCCGTCTGATTTCTTCCCGTCTGTTCGTTGCCTGCTGGACACCGCGGCAGCACTG TACCGTGACGCGCTGGAATCTCCGGAACATTGTTCTCCGCATCACACTGCGCTGC GTCAGGCGATTCTGTGCTGGGGCGACCTGagcACCCTGGCGACTTGGGTTGGCAACC AACatgCAGGATCAGGCGGCGCGTATCTGGTTGTTTCTTACGTTAACACTAACGT TGGTCTGAAATTCCGTCAGCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGGTC GTGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGATTTGTACTCCGCCG GCTTACCGTCCGCCGAACGCACCGATCCTGAGCACCCTGCCGGAAACCACTGTTG TGTAATAA
SS1(HP)	(AHA)DIDPYKEFGATVELLSFLPS DFFPSVRCLLDTAAALYRDALESPE HCSPHHTALRQAVSCWREVTDFGDW VGNN(AHA)QDQAARDLVVNYVNAN IGLKIRQLLWFHISCLTFGRETVLE YLVSFGVWICTPPAYRPPNAPILST LPETTVV	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAACTGCTGTCTTTCC TGCCGTCTGATTTCTTCCCGTCTGTTCGTTGCCTGCTGGACACCGCGGCAGCACTG TACCGTGACGCGCTGGAATCTCCGGAACATTGTTCTCCGCATCACACTGCGCTGC GTCAGGCGGTGAGCTGCTGGCGCGAAGTGACCGATTTTGGCGATTGGGTGGGCA ACAACatgCAGGATCAGGCGGCGCGCGATCTGGTGGTGAACTATGTGAACGCGAA CATTGGCCTGAAAATTCGTCAGCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCG GTCGTGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGATTTGTACTCCG CCGGCTTACCGTCCGCCGAACGCACCGATCCTGAGCACCCTGCCGGAAACCACTG TTGTGTAATAA

References

1. Pasek M*, et al.* (1979) Hepatitis-B virus genes and their expression in *Escherichia coli*. *Nature* 282(5739):575-579.
2. Jewett MC & Swartz JR (2004) Mimicking the *Escherichia coli* cytoplasmic environment activates long-lived and efficient cell-free protein synthesis. *Biotechnol Bioeng* 86(1):19-26.
3. Calhoun KA & Swartz JR (2006) Total amino acid stabilization during cell-free protein synthesis reactions. *J Biotechnol* 123(2):193-203.
4. Calhoun KA & Swartz JR (2005) Energizing cell-free protein synthesis with glucose metabolism. *Biotechnol Bioeng* 90(5):606-613.
5. Patel KG & Swartz JR (2011) Surface Functionalization of virus-like particles by direct conjugation using azide-alkyne click chemistry. *Bioconjugate Chem* 22(3):376-387.
6. Lu Y, Welsh JP, Chan W, & Swartz JR (2013) *Escherichia coli*-based cell free production of flagellin and ordered flagellin display on virus-like particles. *Biotechnol Bioeng* 110(8):2073-2085.
7. Zlotnick A*, et al.* (2007) In vitro screening for molecules that affect virus capsid assembly (and other protein association reactions). *Nat Protoc* 2(3):490-498.
8. Petsch D & Anspach FB (2000) Endotoxin removal from protein solutions. *J Biotechnol* 76(2-3):97-119.
9. Aida Y & Pabst MJ (1990) Removal of endotoxin from protein solutions by phase separation using Triton X-114. *J Immunol Methods* 132(2):191-195.
10. Aguilar JC*, et al.* (2004) Development of a nasal vaccine for chronic hepatitis B infection that uses the ability of hepatitis B core antigen to stimulate a strong Th1 response against hepatitis B surface antigen. *Immunol Cell Biol* 82(5):539-546.
11. Kruisbeek AM, Shevach E, & Thornton AM (2004) Proliferative assays for T cell function. *Current protocols in immunology / edited by John E. Coligan ... [et al.]* Chapter 3:Unit 3 12.