(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
31 July 2014 (31.07.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/116730 A2

(51) International Patent Classification:
H04L 12/18 (2006.01)

(21) International Application Number:
PCT/US2014/012586

(22) International Filing Date:
22 January 2014 (22.01.2014)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/755,850 23 January 2013 (23.01.2013) US
61/901,243 7 November 2013 (07.11.2013) US

(71) Applicant: **THE BOARD OF TRUSTEES OF THE LELAND STANFORD JUNIOR UNIVERSITY** [US/US]; 1705 El Camino Real, Palo Alto, California 94306-1106 (US).

(72) Inventors: **LU, Yuan**; 265 W. Charleston Road, Palo Alto, California 94306 (US). **SWARTZ, James Robert**; 1860 White Oak Drive, Menlo Park, California 94025 (US).

(74) Agent: **SHERWOOD, Pamela J.**; Bozicevic, Field & Francis LLP, 1900 University Avenue, Suite 200, East Palo Alto, California 94303 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *of inventorship (Rule 4.17(iv))*

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

— *with sequence listing part of description (Rule 5.2(a))*

WO 2014/116730 A2

(54) Title: STABILIZED HEPATITIS B CORE POLYPEPTIDE

(57) Abstract: Genetically modified HBc polypeptides are provided.

STABILIZED HEPATITIS B CORE POLYPEPTIDE

BACKGROUND OF THE INVENTION

[0001] Virus-like particles (VLPs) are non-infectious, have repetitive surfaces that can display molecules with a high surface density, and have comparable cellular uptake and intracellular trafficking compared to natural virus. All of these functional attributes make them attractive as the assembly core for vaccines, diagnostics, and therapeutics. They can potentially serve as polyvalent scaffolds for the display of nucleic acids, proteins, and other chemical moieties. VLPs are particularly attractive as vaccines as they offer in vivo stability, trafficking to lymph nodes, and stimulation of B and T cell responses by the displayed epitopes. They can also be filled with cargo to serve as delivery vehicles.

[0002] Cell-free protein synthesis (CFPS) can be an effective method for producing VLPs, for example those comprising Hepatitis B core protein (HBc), MS2 bacteriophage coat protein, and Qβ bacteriophage coat protein, and the like. CFPS also provides a facile means for introducing non-natural amino acids (nnAAs) into proteins, which allows for the direct protein-protein coupling of antigens to VLPs using Cu(I)-catalyzed [3 + 2] cycloaddition click chemistry.

[0003] Among different types of VLPs, the HBc VLP is a flexible and promising model for knowledge-based display of foreign peptide sequences. The HBc particle was first reported as a promising VLP carrier in 1986. Being one the first VLP candidates and the first icosahedral VLP carrier, HBc VLP has been well characterized and widely used as a carrier for over 100 different foreign sequences. The HBc capsid protein is 183 to 185 amino acids long. The arginine-rich C-terminus of HBc protein is dispensable for VLP assembly, so the HBc protein truncated at amino acid 149 is widely used. The truncated HBc (1-149) proteins can self-assemble into the particle with an average diameter of 28 to 30 nm and a dominant icosahedral symmetry of T=4.

[0004] However, in current applications of HBc VLPs, there is a serious problem. The VLP is not stable during click chemistry conjugations, and can disassemble after conjugation with functional molecules. Two truncated HBc monomers (16.7 kDa) form a dimer (33.5 kDa) by an intradimer C61-C61 disulfide bond. Then 120 dimers self-assemble into one VLP by hydrophobic interactions. Because the interdimer contacts are weak, conjugation of functional molecules onto the VLP surface can disturb the interactions between VLP dimers, resulting in VLP instability. The second problem is molecules with negative charges conjugate poorly to VLPs. At physiological pH, the surface of the HBc VLP is negatively charged. Because like charges repel, molecules with negative charges cannot get close to HBc VLP, and therefore the click chemistry conjugation cannot proceed effectively.

[0005] The present invention addresses these two problems, and provides stabilized HBc VLPs and a modified VLP surface.

Relevant literature

[0006] Methods of introducing unnatural amino acids during CFPS are described in patent publication US 2010-0093024 A1. Methods of directly linking antigens and other polypeptides to a virus-like particle through unnatural amino acids are described in patent application US-2010-0168402-A1. Methods of encapsidating virus-like particles produced by CFPS are described in patent publication US-2010-0167981-A1. Each of these documents are herein specifically incorporated by reference.

SUMMARY OF THE INVENTION

[0007] Genetically modified hepatitis B core (HBc) proteins are provided, which proteins comprise sequence modification that enhance the stability and/or utility of the protein. In some embodiments of the invention, at least two amino acids of the native sequence are substituted with cysteine residues that provide for intermolecular disulfide bonding. In some embodiments four amino acids are substituted with cysteines. The HBc protein is thus stabilized and is maintained as a virus-like particle (VLP) under conditions otherwise unfavorable to retention of the quaternary structure, for example during click chemistry reactions.

[0008] In some embodiments, at least two amino acids present in an HBc monomer, usually two amino acids or four amino acids, for example a sequence set forth in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:34, SEQ ID NO:35, or a comparable HBc polypeptide, are replaced with cysteine, which replacements stabilize the protein assembly. Exemplary pairs of amino acid substitutions include, without limitation and using relative to SEQ ID NO:1, 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; and SS10: P135C, N136C. In some embodiments the amino acid substitutions are D29C and R127C. In other embodiments the amino acid substitutions are P134C and N136C. In some embodiments the amino acid substitutions are D29C, R127C, P134C and N136C.

[0009] In other embodiments, the HBc protein alternatively or in addition comprises a set of amino acid substitutions that reduces the negative charge on the “spike tip” of the protein, i.e. , i.e. the region of residues 73-81, relative to SEQ ID NO:1 or SEQ ID NO:2. In some embodiments the set of amino acid substitutions is derived from a naturally occurring virus genotype with a reduced charge. In some embodiments, the set of amino acid changes, relative to SEQ ID NO:1 or SEQ ID NO:2 are I57V, L60S, G63R, D64E, L65V, M66T, T67D, L68F, A69G, T70D, T74N, L76M, E77Q, P79Q, S81A, S87N, T91A, V93I, and F97I. In some

embodiments the set of amino acid changes is T74N, L76M, E77Q, P79Q, and S81A. In some embodiments, the amino acid sequence of the HBc protein with a reduced negative charge is SEQ ID NO:34 or SEQ ID NO:35.

[0010] The HBc protein, either cysteine stabilized, charge reduced, or both, can further comprise one or more unnatural amino acids at a pre-determined site. Unnatural amino acids of interest include without limitation azidohomoalanine, *p*-acetyl-phenylalanine, *p*-ethynyl-phenylalanine, *p*-propargyloxyphenylalanine, *p*-azido-phenylalanine, *etc*. The unnatural amino acid(s) may be positioned at the spike of the HBc protein. Sites of interest include, for example, N75, T74, L76, Q77, D78, Q79 and A80. In some embodiments the unnatural amino acid replaces D78. In some embodiments the unnatural amino acid is azidohomoalanine.

[0011] HBc polypeptides, or VLP generated therefrom may comprise a conjugated moiety other than an HBc polypeptide, where such a moiety is conjugated to the HBc at the introduced unnatural amino acid, e.g. by click chemistry. Suitable moieties include polypeptides, nucleic acids, polysaccharides, therapeutic drugs, imaging moieties, and the like. In a related embodiment, a method is provided, where the unnatural amino acid in HBc is utilized in a click chemistry reaction to join an additional moiety to the HBc of the invention, or a VLP comprising HBc of the invention.

[0012] The HBc polypeptides of this invention can be made by transforming host cells with nucleic acid encoding the polypeptide, culturing the host cell and recovering the polypeptide from the culture, or alternatively by generating a nucleic acid construct encoding the HBc polypeptides and producing the polypeptide by cell free synthesis, which synthesis may include coupled transcription and translation reactions. Also provided are vectors and polynucleotides encoding the HBc polypeptides. In some embodiments a VLP comprising polypeptides of the invention is provided.

[0013] In one embodiment of the invention, a method is provided for the cell-free protein synthesis (CFPS) of the protein of the invention. In some embodiments the CFPS product is synthesized; and may further be assembled into a VLP, in a reducing environment. The CFPS product may be dialyzed in a solution of from about 1M to about 2 M salt, e.g. about 1.5 M salt, e.g. NaCl, *etc*. The assembled VLP may be isolated in a reducing environment. Following synthesis and assembly into a VLP, the VLP may be switched to an oxidizing environment to generate stabilizing disulfide bonds.

BRIEF DESCRIPTION OF THE DRAWINGS

[0014] Fig. 1 The structure of HBc VLP.

[0015] Fig. 2 The structures of the pentamer and the hexamer.

[0016] Fig. 3 Selection of possible S-S bond positions at both 5-fold units and 6-fold units. (a) Introduction of S-S bonds between side chains of dimers; (b) Introduction of S-S bonds between C-terminal regions of monomers.

[0017] Fig. 4 The procedure for the formation and the assessment of correct S-S bonds. After the sucrose gradient centrifugation, VLP samples were also assessed by SDS-PAGE analysis (non-reducing).

[0018] Fig. 5 CFPS yields (a) of orginal HBc protein and mutants, and the autoradiogram analysis (b) of non-reducing SDS-PAGE. O: Original HBc; 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.

[0019] Fig. 6 SEC analysis of original HBc and mutants.

[0020] Fig. 7 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 the oxidants. The three bands that appear between MWs of 79 and 98 kDa represent contamination by the three polypeptides of the pyruvate dehydrogenase complex.

[0021] Fig. 8 The sucrose gradient centrifugation analysis of original HBc VLP and oxidized SS1, SS7, SS8, SS9 and SS10 VLPs.

[0022] Fig. 9 The reducing SDS-PAGE analysis of click-reaction products. Cu(I) was not added in the reaction labeled “no Cu” as the control. HBc VLPs were radioactive. Flagellin and GM-CSF were not radioactive.

[0023] Fig. 10. Illustration for the creation of (HBc(ST) and HBc(HP)). The spike regions of natural mutant Q8B6N7 were transplanted to native protein.

[0024] Fig. 11. CFPS yields of mutants HBc(D78M), HBc(ST) and HBc(HP). HBc(D78M), HBc(ST), and HBc(HP) were stabilized by introducing new disulfide bridges (D29C-R127C).

[0025] Fig. 12. Size-exclusion chromatography (SEC) analysis after dialysis against buffer with 0.5 M NaCl (A) or 1.5 M NaCl (B).

[0026] Fig. 13. Selection of six different nnAA sites (N75AHA, L76AHA, Q77AHA, D78AHA, Q79AHA, and A80AHA) on the spike tip of mutant HBc(HP).

[0027] Fig. 14. Soluble yields after CFPS and dialysis of six different mutants (N75AHA, L76AHA, Q77AHA, D78AHA, Q79AHA, and A80AHA). The dialysis buffer was 10 mM Tris-HCl (pH 7.4), 1.5 M NaCl.

[0028] Fig. 15. Size-exclusion chromatography (SEC) analysis after dialysis against buffer with 1.5 M NaCl.

[0029] Fig. 16. The click-reaction analysis of HBc VLP mutants. (A) The surface charge distribution of four different molecules at physiological pH and the position of non-natural amino acid (nnAA) with an alkyne moiety. These four molecules are flagellin protein, GMCSF

protein, IM9-ScFv protein and CpG DNA. (B) The reducing SDS-PAGE autoradiogram analysis of click-reaction products.

[0030] Fig. 17. Illustration of Hepatitis B core protein (HBc) VLP assembly.

[0031] Fig. 18. Illustration for the introduction of artificial disulfide bond (S-S) network.

[0032] Fig. 19. Purification of VLPs by size-exclusion chromatography (SEC) using buffer with 1.5 M NaCl.

[0033] Fig. 20. 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.

DETAILED DESCRIPTION OF THE EMBODIMENTS

[0034] Genetically modified hepatitis B core (HBc) proteins are provided, which proteins comprise sequence modification that enhance the stability and/or utility of the protein. In some embodiments HBc polypeptides are provided in which the structure is stabilized by disulfide bonds. The substituted HBc protein is thus stabilized and is maintained as a VLP under conditions otherwise unfavorable to retention of the quaternary structure. In other embodiments, the amino acid sequence of an HBc protein comprises a set of amino acid substitutions that reduces the negative charge on the "spike tip" of the protein, i.e. the region of residues 73-81, relative to SEQ ID NO:1 or SEQ ID NO:2. In other embodiments an HBc protein comprises one or more unnatural amino acids at a pre-determined site, for example, N75, T74, L76, Q77, D78, Q79 and A80. In certain embodiments, an HBc protein of the invention comprises all three classes of modification: disulfide bond stabilization, negative charge reduction, and an unnatural amino acid at a pre-determined site.

[0035] The HBc polypeptides of the invention find particular use as a component of a VLP, and particularly a VLP designed for conjugation to one or more additional moieties through, for example, click chemistry. In some embodiments the unnatural amino acid is used to link the HBc protein to the additional moiety(s).

[0036] In some embodiments, the invention provides a use of a conjugate, compound, or composition herein in the manufacture of a medicament. In an embodiment, the invention provides a use of a conjugate, compound, or composition herein in the manufacture of a medicament, e.g. a vaccine, for the prevention or treatment of an infection. In some embodiments, the invention provides a use of a conjugate, compound, or composition herein for the prevention or treatment of an infection.

Definitions

[0037] It is to be understood that this invention is not limited to the particular methodology, protocols, cell lines, animal species or genera, and reagents described, as such may vary. It

is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention, which will be limited only by the appended claims.

[0038] As used herein the singular forms "a", "and", and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a cell" includes a plurality of such cells and reference to "the culture" includes reference to one or more cultures and equivalents thereof known to those skilled in the art, and so forth. All technical and scientific terms used herein have the same meaning as commonly understood to one of ordinary skill in the art to which this invention belongs unless clearly indicated otherwise.

[0039] As used herein the singular forms "a", "and", and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a cell" includes a plurality of such cells and reference to "the culture" includes reference to one or more cultures and equivalents thereof known to those skilled in the art, and so forth. All technical and scientific terms used herein have the same meaning as commonly understood to one of ordinary skill in the art to which this invention belongs unless clearly indicated otherwise.

[0040] The term "HBc" refers to the amino acid peptide sequence of the Hepatitis B core protein, or to a truncated version thereof as set forth in SEQ ID NO:1 or SEQ ID NO:2, or a comparable protein, for example as set forth in any one of SEQ ID NO:3-SEQ ID NO:52. One of skill in the art will understand that minor amino acid changes can be made in the sequence without altering the function of the protein, e.g. changes of 1, 2, 3, 4, 5, 6, 7, 8, 9, or up to about 10 amino acids, and that a full-length protein may be substitutued for the truncated versions exemplified herein. HBc is functionally capable of self-assembling to form an icosahedral virus like particle. The HBc polypeptides of the invention comprise amino acid substitutions as described above, which include one or more of: (a) introducing one or more pairs of cysteine residues capable of forming intermolecular disulfide bonds when assembled into a VLP; (b) one or more unnatural amino acids at a predetermined site, preferably those capable of participating in a click chemistry reaction; and (c) one or more amino acid substitutions to decrease the negative charge of the proteins.

[0041] As used herein, the terms “purified” and “isolated” when used in the context of a polypeptide that is substantially free of contaminating materials from the material from which it was obtained, e.g. cellular materials, such as but not limited to cell debris, cell wall materials, membranes, organelles, the bulk of the nucleic acids, carbohydrates, proteins, and/or lipids present in cells. Thus, a polypeptide that is isolated includes preparations of a polypeptide having less than about 30%, 20%, 10%, 5%, 2%, or 1% (by dry weight) of cellular materials and/or contaminating materials. As used herein, the terms “purified” and “isolated” when used in the context of a polypeptide that is chemically synthesized refers to a polypeptide which is

substantially free of chemical precursors or other chemicals which are involved in the syntheses of the polypeptide.

[0042] The term "polypeptide," "peptide," "oligopeptide," and "protein," are used interchangeably herein, and refer to a polymeric form of amino acids of any length, which can include coded and non-coded amino acids, chemically, or biochemically modified or derivatized amino acids, and polypeptides having modified peptide backbones.

[0043] The polypeptides may be isolated and purified in accordance with conventional methods of recombinant synthesis or cell free protein synthesis. Exemplary coding sequences are provided, however one of skill in the art can readily design a suitable coding sequence based on the provided amino acid sequences. Methods which are well known to those skilled in the art can be used to construct expression vectors containing coding sequences and appropriate transcriptional/translational control signals. These methods include, for example, *in vitro* recombinant DNA techniques, synthetic techniques and *in vivo* recombination/genetic recombination. Alternatively, RNA capable of encoding the polypeptides of interest may be chemically synthesized. One of skill in the art can readily utilize well-known codon usage tables and synthetic methods to provide a suitable coding sequence for any of the polypeptides of the invention. The nucleic acids may be isolated and obtained in substantial purity. Usually, the nucleic acids, either as DNA or RNA, will be obtained substantially free of other naturally-occurring nucleic acid sequences, generally being at least about 50%, usually at least about 90% pure and are typically “recombinant,” *e.g.*, flanked by one or more nucleotides with which it is not normally associated on a naturally occurring chromosome. The nucleic acids of the invention can be provided as a linear molecule or within a circular molecule, and can be provided within autonomously replicating molecules (vectors) or within molecules without replication sequences. Expression of the nucleic acids can be regulated by their own or by other regulatory sequences known in the art. The nucleic acids of the invention can be introduced into suitable host cells using a variety of techniques available in the art.

[0044] As used herein, the term “virus like particle” refers to a stable macromolecular assembly of one or more virus proteins, usually viral coat proteins. The number of separate protein chains in a VLP will usually be at least about 60 proteins, about 80 proteins, at least about 120 proteins, or more, depending on the specific viral geometry. In the methods of the invention, the HBc is maintained in conditions permissive for self-assembly into the capsid structure, particularly reducing conditions. The methods of the invention provide for synthesis of the coat protein in the absence of the virus polynucleotide genome, and thus the capsid may be empty, or contain non-viral components, *e.g.* mRNA fragments, etc.

[0045] A stable VLP maintains the association of proteins in a capsid structure under physiological conditions for extended periods of time, e.g. for at least about 24 hrs, at least about 1 week, at least about 1 month, or more. Once assembled, the VLP can have a stability commensurate with the native virus particle, *e.g.* upon exposure to pH changes, heat, freezing, ionic changes, *etc*. Additional components of VLPs, as known in the art, can be included within or disposed on the VLP. VLPs do not contain intact viral nucleic acids, and they are non-infectious. In some embodiments there is sufficient viral surface envelope glycoprotein and/or adjuvant molecules on the surface of the VLP so that when a VLP preparation is formulated into an immunogenic composition and administered to an animal or human, an immune response (cell-mediated or humoral) is raised.

[0046] An "effective amount" or a "sufficient amount" of a substance is that amount sufficient to cause a desired biological effect, such as beneficial results, including clinical results, and, as such, an "effective amount" depends upon the context in which it is being applied. In the context of this invention, an example of an effective amount of a vaccine is an amount sufficient to induce an immune response (e.g., antibody production) in an individual. An effective amount can be administered in one or more administrations.

[0047] Folding, as used herein, refers to the process of forming the three-dimensional structure of polypeptides and proteins, where interactions between amino acid residues act to stabilize the structure. Non-covalent interactions are important in determining structure, and the effect of membrane contacts with the protein may be important for the correct structure. For naturally occurring proteins and polypeptides or derivatives and variants thereof, the result of proper folding is typically the arrangement that results in optimal biological activity, and can conveniently be monitored by assays for activity, *e.g.* ligand binding, enzymatic activity, *etc*.

[0048] In some instances, for example where the desired product is of synthetic origin, assays based on biological activity will be less meaningful. The proper folding of such molecules may be determined on the basis of physical properties, energetic considerations, modeling studies, and the like.

[0049] Separation procedures of interest include affinity chromatography. Affinity chromatography makes use of the highly specific binding sites usually present in biological macromolecules, separating molecules on their ability to bind a particular ligand. Covalent bonds attach the ligand to an insoluble, porous support medium in a manner that overtly presents the ligand to the protein sample, thereby using natural biospecific binding of one molecular species to separate and purify a second species from a mixture. Antibodies are commonly used in affinity chromatography. Preferably a microsphere or matrix is used as the support for affinity chromatography. Such supports are known in the art and are commercially available, and include activated supports that can be combined to the linker molecules. For

example, Affi-Gel supports, based on agarose or polyacrylamide are low pressure gels suitable for most laboratory-scale purifications with a peristaltic pump or gravity flow elution. Affi-Prep supports, based on a pressure-stable macroporous polymer, are suitable for preparative and process scale applications.

[0050] Proteins may also be separated by ion exchange chromatography, and/or concentrated, filtered, dialyzed, *etc.*, using methods known in the art. The methods of the present invention provide for proteins containing unnatural amino acids that have biological activity comparable to the native protein. One may determine the specific activity of a protein in a composition by determining the level of activity in a functional assay, quantitating the amount of protein present in a non-functional assay, *e.g.* immunostaining, ELISA, quantitation on coomassie or silver stained gel, *etc.*, and determining the ratio of biologically active protein to total protein. Generally, the specific activity as thus defined will be at least about 5% that of the native protein, usually at least about 10% that of the native protein, and may be about 25%, about 50%, about 90% or greater.

[0051] A modified HBc protein of the invention will usually comprise at least one unnatural amino acid at a pre-determined site, and may comprise or contain 1, 2, 3, 4, 5 or more unnatural amino acids. If present at two or more sites in the polypeptide, the unnatural amino acids can be the same or different. Where the unnatural amino acids are different, an orthogonal tRNA and cognate tRNA synthetase will be present for each unnatural amino acid.

[0052] Examples of unnatural amino acids that can be used in the methods of the invention include: an unnatural analogue of a tyrosine amino acid; an unnatural analogue of a glutamine amino acid; an unnatural analogue of a phenylalanine amino acid; an unnatural analogue of a serine amino acid; an unnatural analogue of a threonine amino acid; an alkyl, aryl, acyl, azido, cyano, halo, hydrazine, hydrazide, hydroxyl, alkenyl, alkynl, ether, thiol, sulfonyl, seleno, ester, thioacid, borate, boronate, phospho, phosphono, phosphine, heterocyclic, enone, imine, aldehyde, hydroxylamine, keto, or amino substituted amino acid, or any combination thereof; an amino acid with a photoactivatable cross-linker; a spin-labeled amino acid; a fluorescent amino acid; an amino acid with a novel functional group; an amino acid that covalently or noncovalently interacts with another molecule; a metal binding amino acid; a metal-containing amino acid; a radioactive amino acid; a photocaged and/or photoisomerizable amino acid; a biotin or biotin-analogue containing amino acid; a glycosylated or carbohydrate modified amino acid; a keto containing amino acid; amino acids comprising polyethylene glycol or polyether; a heavy atom substituted amino acid; a chemically cleavable or photocleavable amino acid; an amino acid with an elongated side chain; an amino acid containing a toxic group; a sugar substituted amino acid, e.g., a sugar substituted serine or the like; a carbon-linked sugar-containing amino acid; a redox-active amino acid; an α-hydroxy containing acid;

an amino thio acid containing amino acid; an α,α disubstituted amino acid; a β-amino acid; a cyclic amino acid other than proline, *etc.*

[0053] Unnatural amino acids of interest include, without limitation, amino acids that provide a reactant group for CLICK chemistry reactions (see *Click Chemistry: Diverse Chemical Function from a Few Good Reactions* Hartmuth C. Kolb, M. G. Finn, K. Barry Sharpless Angewandte Chemie International Edition Volume 40, 2001, P. 2004, herein specifically incorporated by reference). For example, the amino acids azidohomoalanine, *p*-acetyl-L-phenylalanine and *p*-azido-L-phenylalanine are of interest.

[0054] In some embodiments, the unnatural amino acid is introduced by global replacement of methionine on the protein, e.g. methionine can be left out of a cell-free reaction mixture, and substituted by from 0.25 – 2.5 mM azidohomoalanine (AHA). In such embodiments it is preferred to substitute natural methionines, e.g. M66, with a different amino acid.

[0055] Alternatively the unnatural amino acid is introduced by orthogonal components. Orthogonal components include a tRNA aminoacylated with an unnatural amino acid, where the orthogonal tRNA base pairs with a codon that is not normally associated with an amino acid, *e.g.* a stop codon; a 4 bp codon, *etc.* The reaction mixture may further comprise a tRNA synthetase capable of aminoacylating (with an unnatural amino acid) the cognate orthogonal tRNA. Such components are known in the art, for example as described in U.S. Patent no. 7,045,337, issued May 16, 2006. The orthogonal tRNA recognizes a selector codon, which may be nonsense codons, such as, stop codons, e.g., amber, ochre, and opal codons; four or more base codons; codons derived from natural or unnatural base pairs and the like. The orthogonal tRNA anticodon loop recognizes the selector codon on the mRNA and incorporates the unnatural amino acid at this site in the polypeptide.

[0056] Orthogonal tRNA synthetase can be synthesized exogenously, purified and added to the reaction mix of the invention, usually in a defined quantity, of at least about 10 μg/ml, at least about 20 μg/ml, at least about 30 μg/ml, and not more than about 200 μg/ml. The protein may be synthesized in bacterial or eukaryotic cells and purified, *e.g.* by affinity chromatography, PAGE, gel exclusion chromatography, reverse phase chromatography, and the like, as known in the art.

[0057] The terms "*conjugation partner*" or "*selected additional moiety(s)*" are used interchangeably and refer generally to any moiety, for example a peptide or protein, nucleic acid, polysaccharide, label, etc. that is conjugated to a HBc polypeptide of the invention. The conjugation partner may comprise a complementary active group for CLICK chemistry conjugation to the HBc polypeptide of the invention. For example, it may be synthesized with one or more unnatural amino acids, which allow for the conjugation to the unnatural amino acid present on the HBc protein. One of skill in the art will understand that the chemistry for

conjugation is well-known and can be readily applied to a variety of groups, e.g. CpG, detectable label, antigen, polypeptide, *etc.*

[0058] In some embodiments the conjugation partner is a structural protein, e.g. a collagen, keratin, actin, myosin, elastin, fibrillin, lamin, etc. In some embodiments the conjugation partner is an immunogen, e.g. a pathogen protein useful in immunization, including without limitation influenza proteins such as hemagglutinin. Virus coat proteins of interest include any of the known virus types, e.g. dsDNA viruses, such as smallpox (variola); vaccinia; herpesviruses including varicella-zoster; HSV1, HSV2, KSVH, CMV, EBV; adenovirus; hepatitis B virus; SV40; T even phages such as T4 phage, T2 phage; lambda phage; *etc.* Single stranded DNA viruses include phiX-174; adeno-associated virus, *etc.* Negative-stranded RNA viruses include measles virus; mumps virus; respiratory syncytial virus (RSV); parainfluenza viruses (PIV); metapneumovirus; rabies virus; Ebola virus; influenza virus; etc. Positive-stranded RNA viruses include polioviruses; rhinoviruses; coronaviruses; rubella; yellow fever virus; West Nile virus; dengue fever viruses; equine encephalitis viruses; hepatitis A and hepatitis C viruses; tobacco mosaic virus (TMV); *etc.* Double-stranded RNA viruses include reovirus; *etc.* Retroviruses include rous sarcoma virus; lentivirus such as HIV-1 and HIV-2; etc.

[0059] Examples of polypeptides suitable as conjugation partner include, but are not limited to, antigenic proteins such as tumor antigens, viral proteins, bacterial proteins, including tuberculosis antigens, protozoan proteins, including malarial proteins, renin; growth hormones, including human growth hormone; bovine growth hormone; growth hormone releasing factor; parathyroid hormone; thyroid stimulating hormone; lipoproteins; alpha-1-antitrypsin; insulin A-chain; insulin; proinsulin; follicle stimulating hormone; calcitonin; luteinizing hormone; glucagon; clotting factors such as factor VIIIC, factor IX, tissue factor, and von Willebrands factor; anti-clotting factors such as Protein C; atrial natriuretic factor; lung surfactant; a plasminogen activator, such as urokinase or human urine or tissue-type plasminogen activator (t-PA); bombesin; thrombin; hemopoietic growth factor; tumor necrosis factor-alpha and -beta; enkephalinase; RANTES and other chemokines; human macrophage inflammatory protein (MIP-1α); a serum albumin such as human serum albumin; mullerian-inhibiting substance; relaxin A-chain; relaxin B-chain; prorelaxin; mouse gonadotropin-associated peptide; a microbial protein, such as beta-lactamase; DNase; inhibin; activin; vascular endothelial growth factor (VEGF); receptors for hormones or growth factors; integrin; protein A or D; rheumatoid factors; a neurotrophic factor such as bone-derived neurotrophic factor (BDNF), neurotrophin-3,-4, -5, or -6 (NT-3, NT-4, NT-5, or NT-6), or a nerve growth factor such as NGF-β; platelet-derived growth factor (PDGF); fibroblast growth factor such as αFGF and βFGF; epidermal growth factor (EGF); transforming growth factor (TGF) such as TGF-α and TGF-β, including TGF-β1, TGF-β2, TGF-β3, TGF-β4, or TGF-β5; insulin-like growth factor-I and -II (IGF-I and

IGF-II); des(1-3)-IGF-I (brain IGF-I), insulin-like growth factor binding proteins; CD proteins such as CD-3, CD-4, CD-8, and CD-19; erythropoietin; osteoinductive factors; immunotoxins; a bone morphogenetic protein (BMP); an interferon such as interferon-α, -β, and -γ; colony stimulating factors (CSFs), e.g., M-CSF, GM-CSF, and G-CSF; interleukins (ILs), e.g., IL-1 to IL-18; superoxide dismutase; T-cell receptors; surface membrane proteins; decay accelerating factor; viral antigen such as, for example, a portion of the AIDS envelope; transport proteins; homing receptors; addressins; regulatory proteins; antibodies particularly single chain Fv antibodies; and fragments of any of the above-listed polypeptides.

[0060] *Cell free protein synthesis*, as used herein, refers to the cell-free synthesis of polypeptides in a reaction mix comprising biological extracts and/or defined reagents. The reaction mix will comprise a template for production of the macromolecule, *e.g.* DNA, mRNA, *etc.*; monomers for the macromolecule to be synthesized, *e.g.* amino acids, nucleotides, *etc.*, and such co-factors, enzymes and other reagents that are necessary for the synthesis, *e.g.* ribosomes, tRNA, polymerases, transcriptional factors, *etc.* Such synthetic reaction systems are well-known in the art, and have been described in the literature. The cell free synthesis reaction may be performed as batch, continuous flow, or semi-continuous flow, as known in the art.

[0061] The CFPS and other subsequent steps may be performed under reducing conditions, *e.g.* in the presence of 1 mM DTT or the equivalent. Following assembly of the VLP the conditions may be changed to an oxidizing environment, e.g. by dialysis to remove the reducing agent, optionally in the presence of a salt, e.g. up to about 1M salt, up to about 1.5M salt, up to about 2 M salt, e.g. NaCl, etc., then oxidizing to form disulfide bonds by adding 5-10 mM H_2O_2, 5-10 mM diamide, or the equivalent.

[0062] In some embodiments of the invention, cell free synthesis is performed in a reaction where oxidative phosphorylation is activated, *e.g.* the CYTOMIM™ system. The activation of the respiratory chain and oxidative phosphorylation is evidenced by an increase of polypeptide synthesis in the presence of O_2. In reactions where oxidative phosphorylation is activated, the overall polypeptide synthesis in presence of O_2 is reduced by at least about 40% in the presence of a specific electron transport chain inhibitor, such as HQNO, or in the absence of O_2. The reaction chemistry may be as described in international patent application WO 2004/016778, herein incorporated by reference.

[0063] The CYTOMIM™ environment for synthesis utilizes cell extracts derived from bacterial cells grown in medium containing glucose and phosphate, where the glucose is present initially at a concentration of at least about 0.25% (weight/volume), more usually at least about 1%; and usually not more than about 4%, more usually not more than about 2%. An example

of such media is 2YTPG medium, however one of skill in the art will appreciate that many culture media can be adapted for this purpose, as there are many published media suitable for the growth of bacteria such as *E. coli*, using both defined and undefined sources of nutrients (see Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. Molecular Cloning: A Laboratory Manual, 2nd edition. Cold Spring Harbor University Press, Cold Spring Harbor, NY for examples of glucose containing media). Alternatively, the culture may be grown using a protocol in which the glucose is continually fed as required to maintain a high growth rate in either a defined or complex growth medium. The reaction mixture may be supplemented by the inclusion of vesicles, *e.g.* an inner membrane vesicle solution. Where provided, such vesicles may comprise from about 0 to about 0.5 volumes, usually from about 0.1 to about 0.4 volumes.

[0064] In some embodiments, PEG will be present in not more than trace amounts, for example less than 0.1%, and may be less than 0.01%. Reactions that are substantially free of PEG contain sufficiently low levels of PEG that, for example, oxidative phosphorylation is not PEG-inhibited. The molecules spermidine and putrescine may be used in the place of PEG. Spermine or spermidine is present at a concentration of at least about 0.5 mM, usually at least about 1 mM, preferably about 1.5 mM, and not more than about 2.5 mM. Putrescine is present at a concentration of at least about 0.5 mM, preferably at least about 1 mM, preferably about 1.5 mM, and not more than about 2.5 mM. The spermidine and/or putrescine may be present in the initial cell extract or may be separately added.

[0065] The concentration of magnesium in the reaction mixture affects the overall synthesis. Often there is magnesium present in the cell extracts, which may then be adjusted with additional magnesium to optimize the concentration. Sources of magnesium salts useful in such methods are known in the art. In one embodiment of the invention, the source of magnesium is magnesium glutamate. A preferred concentration of magnesium is at least about 5 mM, usually at least about 10 mM, and preferably a least about 12 mM; and at a concentration of not more than about 25 mM, usually not more than about 20 mM. Other changes that may enhance synthesis or reduce cost include the omission of HEPES buffer and phosphoenol pyruvate from the reaction mixture.

[0066] The system can be run under aerobic and anaerobic conditions. Oxygen may be supplied, particularly for reactions larger than 15µl, in order to increase synthesis yields. The headspace of the reaction chamber can be filled with oxygen; oxygen may be infused into the reaction mixture; *etc.* Oxygen can be supplied continuously or the headspace of the reaction chamber can be refilled during the course of protein expression for longer reaction times. Other electron acceptors, such as nitrate, sulfate, or fumarate may also be supplied in conjunction with preparing cell extracts so that the required enzymes are active in the cell extract.

[0067] It is not necessary to add exogenous cofactors for activation of oxidative phosphorylation. Compounds such as nicotinamide adenine dinucleotide (NADH), NAD^+, or acetyl-coenzyme A may be used to supplement protein synthesis yields but are not required. Addition of oxalic acid, a metabolic inhibitor of phosphoenolpyruvate synthetase (Pps), may be beneficial in increasing protein yields, but is not necessary.

[0068] The template for cell-free protein synthesis can be either mRNA or DNA, preferably a combined system continuously generates mRNA from a DNA template with a recognizable promoter. Either an endogenous RNA polymerase is used, or an exogenous phage RNA polymerase, typically T7 or SP6, is added directly to the reaction mixture. Alternatively, mRNA can be continually amplified by inserting the message into a template for QB replicase, an RNA dependent RNA polymerase. Purified mRNA is generally stabilized by chemical modification before it is added to the reaction mixture. Nucleases can be removed from extracts to help stabilize mRNA levels. The template can encode for any particular gene of interest.

[01] Other salts, particularly those that are biologically relevant, such as manganese, may also be added. Potassium is generally present at a concentration of at least about 50 mM, and not more than about 250 mM. Ammonium may be present, usually at a concentration of not more than 200 mM, more usually at a concentration of not more than about 100 mM. Usually, the reaction is maintained in the range of about pH 5-10 and a temperature of about 20°-50° C; more usually, in the range of about pH 6-9 and a temperature of about 25°-40° C. These ranges may be extended for specific conditions of interest.

[02] Metabolic inhibitors to undesirable enzymatic activity may be added to the reaction mixture. Alternatively, enzymes or factors that are responsible for undesirable activity may be removed directly from the extract or the gene encoding the undesirable enzyme may be inactivated or deleted from the chromosome.

POLYPEPTIDES

[0069] HBc polypeptides are provided in which the quaternary structure is stabilized by the introduction of cysteine residues that form intermolecular disulfide bonds. Polypeptides of the invention comprise an HBc sequence, for example with reference to SEQ ID NO:1 or SEQ ID NO:2, wherein at least one pair of amino acids are substituted with cysteine, including substitutions of two pairs, three pairs, etc. In some embodiments the amino acid substitutions are selected from [D29C, R127C]; [T109C, V120C]; [Y132C, N136C]; [Y132C, A137C]; [R133C, N136C]; [R133C, A137C]; [P134C, P135C]; [P134C, N136C]; [P134C, A137C]; and [P135C, N136C]. In some embodiments the amino acid substitutions are D29C, R127C; P134C and N136C; or D29C, R127C, P134C and N136C. Amino acid sequences of interest include those set forth in the Examples, e.g. SEQ ID NO:3-SEQ ID NO:52.

[0070] In some embodiments HBc polypeptide comprises at least one unnatural amino acid at a pre-determined site, usually in combination with the introduction of cysteine residues as described above. The unnatural amino acid(s) may be positioned at the spike of the HBc protein. Sites of interest include, for example, N75, T74, L76, Q77, D78, Q79 and A80. In some embodiments the unnatural amino acid replaces D78. In some embodiments the unnatural amino acid is azidohomoalanine. In some embodiments the naturally occurring methionine at residue 66 is replaced with serine, M66S, and the unnatural amino acid is introduced by methionine substitution.

[0071] In other embodiments, the HBc protein alternatively or in addition comprises a set of amino acid substitutions that reduces the negative charge on the "spike tip" of the protein, i.e. , i.e. the region of residues 73-81, relative to SEQ ID NO:1 or SEQ ID NO:2. In some embodiments the set of amino acid substitutions is derived from a naturally occurring virus genotype with a reduced charge. Viral genotypes of interest for this purpose include Uniprot accession number Q8B6N7, provided herein for reference as SEQ ID NO:32.

[0072] In some embodiments, the set of amino acid changes, relative to SEQ ID NO:1 or SEQ ID NO:2 are I57V, L60S, G63R, D64E, L65V, M66T, T67D, L68F, A69G, T70D, T74N, L76M, E77Q, P79Q, S81A, S87N, T91A, V93I, and F97I. In some embodiments the set of amino acid changes is T74N, L76M, E77Q, P79Q, and S81A. In some embodiments, the amino acid sequence of the HBc protein with a reduced negative charge is SEQ ID NO:34 or SEQ ID NO:35, which sequences also comprise the optional substitutions of M66S, and D29C, R127C.

[0073] HBc polypeptides of interest include, without limitation, those comprising the set of amino acid substitutions, which may be made relative to SEQ ID NO:1, SEQ ID NO:2, etc.:

- {D29C, R127C}; {P134C ,N136C}; or {D29C, R127C, P134C and N136C};
- M66S
- N75AHA, T74AHA, L76AHA, Q77AHA, D78AHA, Q79AHA or A80AHA
- {I57V, L60S, G63R, D64E, L65V, M66T, T67D, L68F, A69G, T70D, T74N, L76M, E77Q, P79Q, S81A, S87N, T91A, V93I, F97I}; or {T74N, L76M, E77Q, P79Q, S81A}.

[0074] In certain embodiments the HBc polypeptides comprise the set of amino acid substitutions:

- {D29C, R127C, P134C and N136C};
- M66S
- D78AHA
- {I57V, L60S, G63R, D64E, L65V, M66T, T67D, L68F, A69G, T70D, T74N, L76M, E77Q, P79Q, S81A, S87N, T91A, V93I, F97I}.

[0075] In certain embodiments the HBc polypeptide is one of SEQ ID NO:3-SEQ ID NO:31, SEQ ID NO:34-SEQ ID NO:52. In some embodiments the HBc polypeptide is one of SEQ ID NO:39, SEQ ID NO:42, or SEQ ID NO:52.

[0076] In some embodiments of the invention, a monomeric form of the HBc polypeptide of the invention is provided. In other embodiments a dimeric form of the HBc polypeptide of the invention is provided. In some embodiments the HBc polypeptide is assembled into a VLP, that can be stabilized by intermolecular disulfide bonds upon oxidation.

[0077] If desired, various groups may be introduced into the peptide during synthesis or during expression, which allow for linking to other molecules or to a surface. The introduced groups need not be included in the HBc domain itself, but may be introduced as a tag or fusion C-terminal or N-terminal to the HBc domain. Thus cysteines can be used to make thioethers, histidines for linking to a metal ion complex, carboxyl groups for forming amides or esters, amino groups for forming amides, and the like. In some embodiments an unnatural amino acid is included at one or more defined sites in the protein, including without limitation.

[0078] The HBc polypeptides of the invention may include an unnatural amino acid for the control of direct attachment to a conjugation partner. Conjugation partners may have an active group for conjugation to the unnatural amino acid(s) on the HBc polypeptide. In some embodiments the conjugation partner is modified to comprise an unnatural amino acid, are reacted with a HBc polypeptide, usually a HBc polypeptide that also comprises an unnatural amino acid and that is assembled in a disulfide stabilized VLP. The unnatural amino acid on the conjugation partner is different from, and reactive with, the unnatural amino acid present on the HBc polypeptide(s).

[0079] Where the active groups for conjugation are reactive azide and alkyne groups, the reaction between HBc and partner may by catalyzed with a copper(I) catalyst at a concentration sufficient for catalysis, e.g. at least about 1 μM, at least about 0.1 mM, at least about 1 mM, *etc.*, as is known in the art. The reaction can be performed using commercial sources of copper(I) such as cuprous bromide or iodide or a compound such as tetrakis(acetonitrile)copper(I)hexafluorophosphate as long as the reaction is performed under anaerobic conditions. The reaction can be run in a variety of solvents, and mixtures of water and a variety of (partially) miscible organic solvents including alcohols, DMSO, DMF, tBuOH and acetone work well. The reaction will proceed at room temperature, and is allowed to proceed to the desired level of completion, e.g. at least about 15 minutes, at least about one hour, at least about 4 hours, at least about 8 hours, or more.

[0080] The invention further provides nucleic acids encoding the HBc polypeptides of the invention. As will be appreciated by those in the art, due to the degeneracy of the genetic code, an extremely large number of nucleic acids may be made, all of which encode the HBc

polypeptides of the present invention. Thus, having identified a particular amino acid sequence, those skilled in the art could make any number of different nucleic acids, by simply modifying the sequence of one or more codons in a way that does not change the amino acid sequence.

[0081] Using the nucleic acids of the present invention that encode a HBc polypeptide, a variety of expression constructs can be made. The expression constructs may be self-replicating extrachromosomal vectors or vectors which integrate into a host genome. Alternatively, for purposes of cell-free expression the construct may include those elements required for transcription and translation of the desired polypeptide, but may not include such elements as an origin of replication, selectable marker, *etc*. Cell-free constructs may be replicated *in vitro, e.g.* by PCR, and may comprise terminal sequences optimized for amplification reactions.

[0082] Generally, expression constructs include transcriptional and translational regulatory nucleic acid operably linked to the nucleic acid encoding the fusion protein. The term "control sequences" refers to DNA sequences necessary for the expression of an operably linked coding sequence in a particular expression system, e.g. mammalian cell, bacterial cell, cell-free synthesis, *etc*. The control sequences that are suitable for prokaryote systems, for example, include a promoter, optionally an operator sequence, and a ribosome binding site. Eukaryotic cell systems may utilize promoters, polyadenylation signals, and enhancers.

[0083] A nucleic acid is "operably linked" when it is placed into a functional relationship with another nucleic acid sequence. For example, DNA for a presequence or secretory leader is operably linked to DNA for a polypeptide if it is expressed as a preprotein that participates in the secretion of the polypeptide; a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the sequence; or a ribosome binding site is operably linked to a coding sequence if it is positioned so as to facilitate the initiation of translation. Generally, "operably linked" means that the DNA sequences being linked are contiguous, and, in the case of a secretory leader, contiguous and in reading phase. Linking is accomplished by ligation or through amplification reactions. Synthetic oligonucleotide adaptors or linkers may be used for linking sequences in accordance with conventional practice.

[0084] In general, the transcriptional and translational regulatory sequences may include, but are not limited to, promoter sequences, ribosomal binding sites, transcriptional start and stop sequences, translational start and stop sequences, and enhancer or activator sequences. In a preferred embodiment, the regulatory sequences include a promoter and transcriptional start and stop sequences.

[0085] Promoter sequences encode either constitutive or inducible promoters. The promoters may be either naturally occurring promoters or hybrid promoters. Hybrid promoters, which combine elements of more than one promoter, are also known in the art, and are useful in the

present invention. In a preferred embodiment, the promoters are strong promoters, allowing high expression in *in vitro* expression systems, such as the T7 promoter.

[0086] In addition, the expression construct may comprise additional elements. For example, the expression vector may have one or two replication systems, thus allowing it to be maintained in organisms, for example in mammalian or insect cells for expression and in a prokaryotic host for cloning and amplification. In addition the expression construct may contain a selectable marker gene to allow the selection of transformed host cells. Selection genes are well known in the art and will vary with the host cell used.

FORMULATIONS AND USES

[0087] The HBc polypeptides, including VLP comprised of HBc; and monomers, dimers or VLPs comprising one or more conjugated moieties, may be provided in a pharmaceutically acceptable excipient, and may be in various formulations. As is well known in the art, a pharmaceutically acceptable excipient is a relatively inert substance that facilitates administration of a pharmacologically effective substance. For example, an excipient can give form or consistency, or act as a diluent. Suitable excipients include but are not limited to stabilizing agents, wetting and emulsifying agents, salts for varying osmolarity, encapsulating agents, buffers, and skin penetration enhancers. Excipients as well as formulations for parenteral and nonparenteral drug delivery are set forth in Remington's Pharmaceutical Sciences 19th Ed. Mack Publishing (1995).

[0088] Other formulations include suitable delivery forms known in the art including, but not limited to, carriers such as liposomes. Mahato et al. (1997) Pharm. Res. 14:853-859. Liposomal preparations include, but are not limited to, cytofectins, multilamellar vesicles and unilamellar vesicles.

[0089] Generally, these compositions are formulated for administration by injection or inhalation, *e.g.*, intraperitoneally, intravenously, subcutaneously, intramuscularly, *etc.* Accordingly, these compositions are preferably combined with pharmaceutically acceptable vehicles such as saline, Ringer's solution, dextrose solution, and the like. The particular dosage regimen, i.e., dose, timing and repetition, will depend on the particular individual and that individual's medical history.

[0090] It is to be understood that this invention is not limited to the particular methodology, protocols, cell lines, animal species or genera, constructs, and reagents described, as such may, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention, which will be limited only by the appended claims.

[0091] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood to one of ordinary skill in the art to which this invention belongs. Although any methods, devices and materials similar or equivalent to those described herein can be used in the practice or testing of the invention, the preferred methods, devices and materials are now described.

[0092] All publications mentioned herein are incorporated herein by reference for the purpose of describing and disclosing, for example, the reagents, cells, constructs, and methodologies that are described in the publications, and which might be used in connection with the presently described invention. The publications discussed above and throughout the text are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the inventors are not entitled to antedate such disclosure by virtue of prior invention.

[0093] The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how to make and use the subject invention, and are not intended to limit the scope of what is regarded as the invention. Efforts have been made to ensure accuracy with respect to the numbers used (*e.g.* amounts, temperature, concentrations, *etc.*) but some experimental errors and deviations should be allowed for. Unless otherwise indicated, parts are parts by weight, molecular weight is average molecular weight, temperature is in degrees centigrade; and pressure is at or near atmospheric.

EXPERIMENTAL

EXAMPLE 1

Stabilization and modification of Hepatitis B core (HBc) virus-like particles (VLPs)

Materials and methods

[0094] Plasmid construction. The sequence encoding the human Hepatitis B core (HBc) capsid monomer of subtype adyw (Pasek et al., 1979) 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).

[0095] Sequences of HBc protein variants. The sequences of 10 different variants with different cysteine mutations intended to stabilize the VLPs:

Variants	Protein sequence	DNA encoding sequence
Wild-type (HBc149)	MDIDPYKEFGATVELLS FLPSDFFPSVRDLLDTA AALYRDALESPEHCSPH HTALRQAILCWGDLMTL ATWVGTNLEDPASRDL VVSYVNTNVGLKFRQLL WFHISCLTFGRETVLEY LVSFGVWIRTPPAYRPP NAPILSTLPETTVV (SEQ ID NO:1)	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCG TTGAACTGCTGTCTTTCCTGCCGTCTGATTTCTTCCCG TCTGTTCGTGACCTGCTGGACACCGCGGCAGCACTGT ACCGTGACGCGCTGGAATCTCCGGAACATTGTTCTCC GCATCACACTGCGCTGCGTCAGGCGATTCTGTGCTGG GGCGACCTGATGACCCTGGCGACTTGGGTTGGCACCA ACCTGGAAGATCCGGCGTCTCGTGATCTGGTTGTTTCT TACGTTAACACTAACGTTGGTCTGAAATTCCGTCAGCT GCTGTGGTTCCACATCTCTTGCCTGACCTTCGGTCGTG AAACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGG ATTCGTACTCCGCCGGCTTACCGTCCGCCGAACGCAC CGATCCTGAGCACCCTGCCGGAAACCACTGTTGTGTA ATAA (SEQ ID NO:53)
Original (M66S-L76M)	MDIDPYKEFGATVELLS FLPSDFFPSVRDLLDTA AALYRDALESPEHCSPH HTALRQAILCWGDLSTL ATWVGTNMEDPASRDL VVSYVNTNVGLKFRQLL WFHISCLTFGRETVLEY LVSFGVWIRTPPAYRPP NAPILSTLPETTVV (SEQ ID NO:2)	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCG TTGAACTGCTGTCTTTCCTGCCGTCTGATTTCTTCCCG TCTGTTCGTGACCTGCTGGACACCGCGGCAGCACTGT ACCGTGACGCGCTGGAATCTCCGGAACATTGTTCTCC GCATCACACTGCGCTGCGTCAGGCGATTCTGTGCTGG GGCGACCTGagcACCCTGGCGACTTGGGTTGGCACCA ACatgGAAGATCCGGCGTCTCGTGATCTGGTTGTTTCTT ACGTTAACACTAACGTTGGTCTGAAATTCCGTCAGCTG CTGTGGTTCCACATCTCTTGCCTGACCTTCGGTCGTGA AACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGA TTCGTACTCCGCCGGCTTACCGTCCGCCGAACGCACC GATCCTGAGCACCCTGCCGGAAACCACTGTTGTGTAAT AA (SEQ ID NO:54)
SS1: D29C-R127C	MDIDPYKEFGATVELLS FLPSDFFPSVRCLLDTA AALYRDALESPEHCSPH HTALRQAILCWGDLSTL ATWVGTNMEDPASRDL VVSYVNTNVGLKFRQLL WFHISCLTFGRETVLEY LVSFGVWICTPPAYRPP NAPILSTLPETTVV (SEQ ID NO:3)	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCG TTGAACTGCTGTCTTTCCTGCCGTCTGATTTCTTCCCG TCTGTTCGTTGCCTGCTGGACACCGCGGCAGCACTGT ACCGTGACGCGCTGGAATCTCCGGAACATTGTTCTCC GCATCACACTGCGCTGCGTCAGGCGATTCTGTGCTGG GGCGACCTGagcACCCTGGCGACTTGGGTTGGCACCA ACatgGAAGATCCGGCGTCTCGTGATCTGGTTGTTTCTT ACGTTAACACTAACGTTGGTCTGAAATTCCGTCAGCTG CTGTGGTTCCACATCTCTTGCCTGACCTTCGGTCGTGA AACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGA TTTGTACTCCGCCGGCTTACCGTCCGCCGAACGCACC GATCCTGAGCACCCTGCCGGAAACCACTGTTGTGTAAT AA (SEQ ID NO:55)
SS2: T109C-V120C	MDIDPYKEFGATVELLS FLPSDFFPSVRDLLDTA AALYRDALESPEHCSPH HTALRQAILCWGDLSTL ATWVGTNMEDPASRDL VVSYVNTNVGLKFRQLL WFHISCLCFGRETVLEY LCSFGVWIRTPPAYRPP NAPILSTLPETTVV (SEQ ID NO:4)	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCG TTGAACTGCTGTCTTTCCTGCCGTCTGATTTCTTCCCG TCTGTTCGTGACCTGCTGGACACCGCGGCAGCACTGT ACCGTGACGCGCTGGAATCTCCGGAACATTGTTCTCC GCATCACACTGCGCTGCGTCAGGCGATTCTGTGCTGG GGCGACCTGagcACCCTGGCGACTTGGGTTGGCACCA ACatgGAAGATCCGGCGTCTCGTGATCTGGTTGTTTCTT ACGTTAACACTAACGTTGGTCTGAAATTCCGTCAGCTG CTGTGGTTCCACATCTCTTGCCTGTGCTTCGGTCGTGA AACCGTTCTGGAATACCTGTGTTCTTTTGGTGTTTGGA TTCGTACTCCGCCGGCTTACCGTCCGCCGAACGCACC GATCCTGAGCACCCTGCCGGAAACCACTGTTGTGTAAT AA (SEQ ID NO:56)

SS3: Y132C-N136C	MDIDPYKEFGATVELLS FLPSDFFPSVRDLLDTA AALYRDALESPEHCSPH HTALRQAILCWGDLSTL ATWVGTNMEDPASRDL VVSYVNTNVGLKFRQLL WFHISCLTFGRETVLEY LVSFGVWIRTPPACRPP CAPILSTLPETTVV (SEQ ID NO:5)	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCG TTGAACTGCTGTCTTTCCTGCCGTCTGATTTCTTCCCG TCTGTTCGTGACCTGCTGGACACCGCGGCAGCACTGT ACCGTGACGCGCTGGAATCTCCGGAACATTGTTCTCC GCATCACACTGCGCTGCGTCAGGCGATTCTGTGCTGG GGCGACCTGagcACCCTGGCGACTTGGGTTGGCACCA ACatgGAAGATCCGGCGTCTCGTGATCTGGTTGTTTCTT ACGTTAACACTAACGTTGGTCTGAAATTCCGTCAGCTG CTGTGGTTCCACATCTCTTGCCTGACCTTCGGTCGTGA AACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGA TTCGTACTCCGCCGGCTTGCCGTCCGCCGTGCGCACC GATCCTGAGCACCCTGCCGGAAACCACTGTTGTGTAAT AA (SEQ ID NO:57)
SS4: Y132C-A137C	MDIDPYKEFGATVELLS FLPSDFFPSVRDLLDTA AALYRDALESPEHCSPH HTALRQAILCWGDLSTL ATWVGTNMEDPASRDL VVSYVNTNVGLKFRQLL WFHISCLTFGRETVLEY LVSFGVWIRTPPACRPP NCPILSTLPETTVV SEQ ID NO:6)	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCG TTGAACTGCTGTCTTTCCTGCCGTCTGATTTCTTCCCG TCTGTTCGTGACCTGCTGGACACCGCGGCAGCACTGT ACCGTGACGCGCTGGAATCTCCGGAACATTGTTCTCC GCATCACACTGCGCTGCGTCAGGCGATTCTGTGCTGG GGCGACCTGagcACCCTGGCGACTTGGGTTGGCACCA ACatgGAAGATCCGGCGTCTCGTGATCTGGTTGTTTCTT ACGTTAACACTAACGTTGGTCTGAAATTCCGTCAGCTG CTGTGGTTCCACATCTCTTGCCTGACCTTCGGTCGTGA AACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGA TTCGTACTCCGCCGGCTTGCCGTCCGCCGAACTGCCC GATCCTGAGCACCCTGCCGGAAACCACTGTTGTGTAAT AA (SEQ ID NO:58)
SS5: R133C-N136C	MDIDPYKEFGATVELLS FLPSDFFPSVRDLLDTA AALYRDALESPEHCSPH HTALRQAILCWGDLSTL ATWVGTNMEDPASRDL VVSYVNTNVGLKFRQLL WFHISCLTFGRETVLEY LVSFGVWIRTPPAYCPP CAPILSTLPETTVV (SEQ ID NO:7)	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCG TTGAACTGCTGTCTTTCCTGCCGTCTGATTTCTTCCCG TCTGTTCGTGACCTGCTGGACACCGCGGCAGCACTGT ACCGTGACGCGCTGGAATCTCCGGAACATTGTTCTCC GCATCACACTGCGCTGCGTCAGGCGATTCTGTGCTGG GGCGACCTGagcACCCTGGCGACTTGGGTTGGCACCA ACatgGAAGATCCGGCGTCTCGTGATCTGGTTGTTTCTT ACGTTAACACTAACGTTGGTCTGAAATTCCGTCAGCTG CTGTGGTTCCACATCTCTTGCCTGACCTTCGGTCGTGA AACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGA TTCGTACTCCGCCGGCTTACTGCCCGCCGTGCGCACC GATCCTGAGCACCCTGCCGGAAACCACTGTTGTGTAAT AA (SEQ ID NO:59)
SS6: R133C-A137C	MDIDPYKEFGATVELLS FLPSDFFPSVRDLLDTA AALYRDALESPEHCSPH HTALRQAILCWGDLSTL ATWVGTNMEDPASRDL VVSYVNTNVGLKFRQLL WFHISCLTFGRETVLEY LVSFGVWIRTPPAYCPP NCPILSTLPETTVV (SEQ ID NO:8)	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCG TTGAACTGCTGTCTTTCCTGCCGTCTGATTTCTTCCCG TCTGTTCGTGACCTGCTGGACACCGCGGCAGCACTGT ACCGTGACGCGCTGGAATCTCCGGAACATTGTTCTCC GCATCACACTGCGCTGCGTCAGGCGATTCTGTGCTGG GGCGACCTGagcACCCTGGCGACTTGGGTTGGCACCA ACatgGAAGATCCGGCGTCTCGTGATCTGGTTGTTTCTT ACGTTAACACTAACGTTGGTCTGAAATTCCGTCAGCTG CTGTGGTTCCACATCTCTTGCCTGACCTTCGGTCGTGA AACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGA TTCGTACTCCGCCGGCTTACTGCCCGCCGAACTGCCC GATCCTGAGCACCCTGCCGGAAACCACTGTTGTGTAAT AA (SEQ ID NO:60)

SS7: P134C-P135C	MDIDPYKEFGATVELLS FLPSDFFPSVRDLLDTA AALYRDALESPEHCSPH HTALRQAILCWGDLSTL ATWVGTNMEDPASRDL VVSYVNTNVGLKFRQLL WFHISCLTFGRETVLEY LVSFGVWIRTPPAYRCC NAPILSTLPETTVV (SEQ ID NO:9)	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCG TTGAACTGCTGTCTTTCCTGCCGTCTGATTTCTTCCCG TCTGTTCGTGACCTGCTGGACACCGCGGCAGCACTGT ACCGTGACGCGCTGGAATCTCCGGAACATTGTTCTCC GCATCACACTGCGCTGCGTCAGGCGATTCTGTGCTGG GGCGACCTGagcACCCTGGCGACTTGGGTTGGCACCA ACatgGAAGATCCGGCGTCTCGTGATCTGGTTGTTTCTT ACGTTAACACTAACGTTGGTCTGAAATTCCGTCAGCTG CTGTGGTTCCACATCTCTTGCCTGACCTTCGGTCGTGA AACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGA TTCGTACTCCGCCGGCTTACCGTTGCTGCAACGCACC GATCCTGAGCACCCTGCCGGAAACCACTGTTGTGTAAT AA (SEQ ID NO:61)
SS8: P134C-N136C	MDIDPYKEFGATVELLS FLPSDFFPSVRDLLDTA AALYRDALESPEHCSPH HTALRQAILCWGDLSTL ATWVGTNMEDPASRDL VVSYVNTNVGLKFRQLL WFHISCLTFGRETVLEY LVSFGVWIRTPPAYRCP CAPILSTLPETTVV (SEQ ID NO:10)	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCG TTGAACTGCTGTCTTTCCTGCCGTCTGATTTCTTCCCG TCTGTTCGTGACCTGCTGGACACCGCGGCAGCACTGT ACCGTGACGCGCTGGAATCTCCGGAACATTGTTCTCC GCATCACACTGCGCTGCGTCAGGCGATTCTGTGCTGG GGCGACCTGagcACCCTGGCGACTTGGGTTGGCACCA ACatgGAAGATCCGGCGTCTCGTGATCTGGTTGTTTCTT ACGTTAACACTAACGTTGGTCTGAAATTCCGTCAGCTG CTGTGGTTCCACATCTCTTGCCTGACCTTCGGTCGTGA AACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGA TTCGTACTCCGCCGGCTTACCGTTGCCCGTGCGCACC GATCCTGAGCACCCTGCCGGAAACCACTGTTGTGTAAT AA (SEQ ID NO:62)
SS9: P134C-A137C	MDIDPYKEFGATVELLS FLPSDFFPSVRDLLDTA AALYRDALESPEHCSPH HTALRQAILCWGDLSTL ATWVGTNMEDPASRDL VVSYVNTNVGLKFRQLL WFHISCLTFGRETVLEY LVSFGVWIRTPPAYRCP NCPILSTLPETTVV (SEQ ID NO:11)	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCG TTGAACTGCTGTCTTTCCTGCCGTCTGATTTCTTCCCG TCTGTTCGTGACCTGCTGGACACCGCGGCAGCACTGT ACCGTGACGCGCTGGAATCTCCGGAACATTGTTCTCC GCATCACACTGCGCTGCGTCAGGCGATTCTGTGCTGG GGCGACCTGagcACCCTGGCGACTTGGGTTGGCACCA ACatgGAAGATCCGGCGTCTCGTGATCTGGTTGTTTCTT ACGTTAACACTAACGTTGGTCTGAAATTCCGTCAGCTG CTGTGGTTCCACATCTCTTGCCTGACCTTCGGTCGTGA AACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGA TTCGTACTCCGCCGGCTTACCGTTGCCCGAACTGCCC GATCCTGAGCACCCTGCCGGAAACCACTGTTGTGTAAT AA (SEQ ID NO:63)
SS10: P135C-N136C	MDIDPYKEFGATVELLS FLPSDFFPSVRDLLDTA AALYRDALESPEHCSPH HTALRQAILCWGDLSTL ATWVGTNMEDPASRDL VVSYVNTNVGLKFRQLL WFHISCLTFGRETVLEY LVSFGVWIRTPPAYRPC CAPILSTLPETTVV (SEQ ID NO:12)	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCG TTGAACTGCTGTCTTTCCTGCCGTCTGATTTCTTCCCG TCTGTTCGTGACCTGCTGGACACCGCGGCAGCACTGT ACCGTGACGCGCTGGAATCTCCGGAACATTGTTCTCC GCATCACACTGCGCTGCGTCAGGCGATTCTGTGCTGG GGCGACCTGagcACCCTGGCGACTTGGGTTGGCACCA ACatgGAAGATCCGGCGTCTCGTGATCTGGTTGTTTCTT ACGTTAACACTAACGTTGGTCTGAAATTCCGTCAGCTG CTGTGGTTCCACATCTCTTGCCTGACCTTCGGTCGTGA AACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGA TTCGTACTCCGCCGGCTTACCGTCCGTGCTGCGCACC GATCCTGAGCACCCTGCCGGAAACCACTGTTGTGTAAT AA (SEQ ID NO:64)

[0096] The sequences of 17 different variants with different AHA sites or different negative surface charge mutations intended to change surface charge and improve conjugation

efficiency without affecting expression and assembly yields. All sequences include the SS1 mutations (red codons) and the M66S mutation (blue codon). AHA signifies azidohomoalanine, the non-natural amino acid tested.

Variants	Protein sequence	DNA encoding sequence
00. Original (L76AHA)	MDIDPYKEFGATVE LLSFLPSDFFPSVR DLLDTAAALYRDAL ESPEHCSPHHTAL RQAILCWGDLSTLA TWVGTN(AHA)EDP ASRDLVVSYVNTN VGLKFRQLLWFHIS CLTFGRETVLEYLV SFGVWIRTPPAYR PPNAPILSTLPETT VV (SEQ ID NO:13)	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACC GTTGAACTGCTGTCTTTCCTGCCGTCTGATTTCTTCCC GTCTGTTCGTGACCTGCTGGACACCGCGGCAGCACT GTACCGTGACGCGCTGGAATCTCCGGAACATTGTTCT CCGCATCACACTGCGCTGCGTCAGGCGATTCTGTGC TGGGGCGACCTGagcACCCTGGCGACTTGGGTTGGC ACCAACatgGAAGATCCGGCGTCTCGTGATCTGGTTGT TTCTTACGTTAACACTAACGTTGGTCTGAAATTCCGTC AGCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGG TCGTGAAACCGTTCTGGAATACCTGGTTTCTTTTGGT GTTTGGATTCGTACTCCGCCGGCTTACCGTCCGCCG AACGCACCGATCCTGAGCACCCTGCCGGAAACCACT GTTGTGTAATAA (SEQ ID NO:65)
0. SS1 (L76AHA)	MDIDPYKEFGATVE LLSFLPSDFFPSVR CLLDTAAALYRDAL ESPEHCSPHHTAL RQAILCWGDLSTLA TWVGTN(AHA)EDP ASRDLVVSYVNTN VGLKFRQLLWFHIS CLTFGRETVLEYLV SFGVWICTPPAYR PPNAPILSTLPETT VV (SEQ ID NO:14)	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACC GTTGAACTGCTGTCTTTCCTGCCGTCTGATTTCTTCCC GTCTGTTCGTTGCCTGCTGGACACCGCGGCAGCACT GTACCGTGACGCGCTGGAATCTCCGGAACATTGTTCT CCGCATCACACTGCGCTGCGTCAGGCGATTCTGTGC TGGGGCGACCTGagcACCCTGGCGACTTGGGTTGGC ACCAACatgGAAGATCCGGCGTCTCGTGATCTGGTTGT TTCTTACGTTAACACTAACGTTGGTCTGAAATTCCGTC AGCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGG TCGTGAAACCGTTCTGGAATACCTGGTTTCTTTTGGT GTTTGGATTTGTACTCCGCCGGCTTACCGTCCGCCGA ACGCACCGATCCTGAGCACCCTGCCGGAAACCACTG TTGTGTAATAA (SEQ ID NO:66)
1. T74AHA	MDIDPYKEFGATVE LLSFLPSDFFPSVR CLLDTAAALYRDAL ESPEHCSPHHTAL RQAILCWGDLSTLA TWVG(AHA)NLEDP ASRDLVVSYVNTN VGLKFRQLLWFHIS CLTFGRETVLEYLV SFGVWICTPPAYR PPNAPILSTLPETT VV (SEQ ID NO:15)	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACC GTTGAACTGCTGTCTTTCCTGCCGTCTGATTTCTTCCC GTCTGTTCGTTGCCTGCTGGACACCGCGGCAGCACT GTACCGTGACGCGCTGGAATCTCCGGAACATTGTTCT CCGCATCACACTGCGCTGCGTCAGGCGATTCTGTGC TGGGGCGACCTGagcACCCTGGCGACTTGGGTTGGCa tgAACCTGGAAGATCCGGCGTCTCGTGATCTGGTTGTT TCTTACGTTAACACTAACGTTGGTCTGAAATTCCGTCA GCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGGT CGTGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTG TTTGGATTTGTACTCCGCCGGCTTACCGTCCGCCGAA CGCACCGATCCTGAGCACCCTGCCGGAAACCACTGT TGTGTAATAA (SEQ ID NO:67)

2. D78AHA	MDIDPYKEFGATVE LLSFLPSDFFPSVR CLLDTAAALYRDAL ESPEHCSPHHTAL RQAILCWGDLSTLA TWVGTNLE(AHA)P ASRDLVVSYVNTN VGLKFRQLLWFHIS CLTFGRETVLEYLV SFGVWICTPPAYR PPNAPILSTLPETT VV (SEQ ID NO:16)	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACC GTTGAACTGCTGTCTTTCCTGCCGTCTGATTTCTTCCC GTCTGTTCGTTGCCTGCTGGACACCGCGGCAGCACT GTACCGTGACGCGCTGGAATCTCCGGAACATTGTTCT CCGCATCACACTGCGCTGCGTCAGGCGATTCTGTGC TGGGGCGACCTGagcACCCTGGCGACTTGGGTTGGC ACCAACCTGGAAatgCCGGCGTCTCGTGATCTGGTTGT TTCTTACGTTAACACTAACGTTGGTCTGAAATTCCGTC AGCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGG TCGTGAAACCGTTCTGGAATACCTGGTTTCTTTTGGT GTTTGGATTTGTACTCCGCCGGCTTACCGTCCGCCGA ACGCACCGATCCTGAGCACCCTGCCGGAAACCACTG TTGTGTAATAA (SEQ ID NO:68)
3. D78AHA+E77S	MDIDPYKEFGATVE LLSFLPSDFFPSVR CLLDTAAALYRDAL ESPEHCSPHHTAL RQAILCWGDLSTLA TWVGTNLS(AHA)P ASRDLVVSYVNTN VGLKFRQLLWFHIS CLTFGRETVLEYLV SFGVWICTPPAYR PPNAPILSTLPETT VV (SEQ ID NO:17)	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACC GTTGAACTGCTGTCTTTCCTGCCGTCTGATTTCTTCCC GTCTGTTCGTTGCCTGCTGGACACCGCGGCAGCACT GTACCGTGACGCGCTGGAATCTCCGGAACATTGTTCT CCGCATCACACTGCGCTGCGTCAGGCGATTCTGTGC TGGGGCGACCTGagcACCCTGGCGACTTGGGTTGGC ACCAACCTGagcatgCCGGCGTCTCGTGATCTGGTTGT TTCTTACGTTAACACTAACGTTGGTCTGAAATTCCGTC AGCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGG TCGTGAAACCGTTCTGGAATACCTGGTTTCTTTTGGT GTTTGGATTTGTACTCCGCCGGCTTACCGTCCGCCGA ACGCACCGATCCTGAGCACCCTGCCGGAAACCACTG TTGTGTAATAA (SEQ ID NO:69)
4. D78AHA+L76S	MDIDPYKEFGATVE LLSFLPSDFFPSVR CLLDTAAALYRDAL ESPEHCSPHHTAL RQAILCWGDLSTLA TWVGTNSE(AHA)P ASRDLVVSYVNTN VGLKFRQLLWFHIS CLTFGRETVLEYLV SFGVWICTPPAYR PPNAPILSTLPETT VV (SEQ ID NO:18)	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACC GTTGAACTGCTGTCTTTCCTGCCGTCTGATTTCTTCCC GTCTGTTCGTTGCCTGCTGGACACCGCGGCAGCACT GTACCGTGACGCGCTGGAATCTCCGGAACATTGTTCT CCGCATCACACTGCGCTGCGTCAGGCGATTCTGTGC TGGGGCGACCTGagcACCCTGGCGACTTGGGTTGGC ACCAACagcGAAatgCCGGCGTCTCGTGATCTGGTTGT TTCTTACGTTAACACTAACGTTGGTCTGAAATTCCGTC AGCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGG TCGTGAAACCGTTCTGGAATACCTGGTTTCTTTTGGT GTTTGGATTTGTACTCCGCCGGCTTACCGTCCGCCGA ACGCACCGATCCTGAGCACCCTGCCGGAAACCACTG TTGTGTAATAA (SEQ ID NO:70)
5. D78AHA+L76S+E77A	MDIDPYKEFGATVE LLSFLPSDFFPSVR CLLDTAAALYRDAL ESPEHCSPHHTAL RQAILCWGDLSTLA TWVGTNSA(AHA)P ASRDLVVSYVNTN VGLKFRQLLWFHIS CLTFGRETVLEYLV SFGVWICTPPAYR PPNAPILSTLPETT VV (SEQ ID NO:19)	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACC GTTGAACTGCTGTCTTTCCTGCCGTCTGATTTCTTCCC GTCTGTTCGTTGCCTGCTGGACACCGCGGCAGCACT GTACCGTGACGCGCTGGAATCTCCGGAACATTGTTCT CCGCATCACACTGCGCTGCGTCAGGCGATTCTGTGC TGGGGCGACCTGagcACCCTGGCGACTTGGGTTGGC ACCAACagcgcagcatgCCGGCGTCTCGTGATCTGGTTGTT TCTTACGTTAACACTAACGTTGGTCTGAAATTCCGTCA GCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGGT CGTGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTG TTTGGATTTGTACTCCGCCGGCTTACCGTCCGCCGAA CGCACCGATCCTGAGCACCCTGCCGGAAACCACTGT TGTGTAATAA (SEQ ID NO:71)

6. P79AHA	MDIDPYKEFGATVE LLSFLPSDFFPSVR CLLDTAAALYRDAL ESPEHCSPHHTAL RQAILCWGDLSTLA TWVGTNLED(AHA) ASRDLVVSYVNTN VGLKFRQLLWFHIS CLTFGRETVLEYLV SFGVWICTPPAYR PPNAPILSTLPETT VV (SEQ ID NO:20)	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACC GTTGAACTGCTGTCTTTCCTGCCGTCTGATTTCTTCCC GTCTGTTCGTTGCCTGCTGGACACCGCGGCAGCACT GTACCGTGACGCGCTGGAATCTCCGGAACATTGTTCT CCGCATCACACTGCGCTGCGTCAGGCGATTCTGTGC TGGGGCGACCTGagcACCCTGGCGACTTGGGTTGGC ACCAACCTGGAAGATatgGCGTCTCGTGATCTGGTTGT TTCTTACGTTAACACTAACGTTGGTCTGAAATTCCGTC AGCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGG TCGTGAAACCGTTCTGGAATACCTGGTTTCTTTTGGT GTTTGGATTTGTACTCCGCCGGCTTACCGTCCGCCGA ACGCACCGATCCTGAGCACCCTGCCGGAAACCACTG TTGTGTAATAA (SEQ ID NO:72)
7. P79AHA+E77K	MDIDPYKEFGATVE LLSFLPSDFFPSVR CLLDTAAALYRDAL ESPEHCSPHHTAL RQAILCWGDLSTLA TWVGTNLKD(AHA) ASRDLVVSYVNTN VGLKFRQLLWFHIS CLTFGRETVLEYLV SFGVWICTPPAYR PPNAPILSTLPETT VV (SEQ ID NO:21)	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACC GTTGAACTGCTGTCTTTCCTGCCGTCTGATTTCTTCCC GTCTGTTCGTTGCCTGCTGGACACCGCGGCAGCACT GTACCGTGACGCGCTGGAATCTCCGGAACATTGTTCT CCGCATCACACTGCGCTGCGTCAGGCGATTCTGTGC TGGGGCGACCTGagcACCCTGGCGACTTGGGTTGGC ACCAACCTGaaaGATatgGCGTCTCGTGATCTGGTTGTT TCTTACGTTAACACTAACGTTGGTCTGAAATTCCGTCA GCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGGT CGTGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTG TTTGGATTTGTACTCCGCCGGCTTACCGTCCGCCGAA CGCACCGATCCTGAGCACCCTGCCGGAAACCACTGT TGTGTAATAA (SEQ ID NO:73)
8. P79AHA+D78K	MDIDPYKEFGATVE LLSFLPSDFFPSVR CLLDTAAALYRDAL ESPEHCSPHHTAL RQAILCWGDLSTLA TWVGTNLEK(AHA) ASRDLVVSYVNTN VGLKFRQLLWFHIS CLTFGRETVLEYLV SFGVWICTPPAYR PPNAPILSTLPETT VV (SEQ ID NO:22)	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACC GTTGAACTGCTGTCTTTCCTGCCGTCTGATTTCTTCCC GTCTGTTCGTTGCCTGCTGGACACCGCGGCAGCACT GTACCGTGACGCGCTGGAATCTCCGGAACATTGTTCT CCGCATCACACTGCGCTGCGTCAGGCGATTCTGTGC TGGGGCGACCTGagcACCCTGGCGACTTGGGTTGGC ACCAACCTGGAAaaaatgGCGTCTCGTGATCTGGTTGT TTCTTACGTTAACACTAACGTTGGTCTGAAATTCCGTC AGCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGG TCGTGAAACCGTTCTGGAATACCTGGTTTCTTTTGGT GTTTGGATTTGTACTCCGCCGGCTTACCGTCCGCCGA ACGCACCGATCCTGAGCACCCTGCCGGAAACCACTG TTGTGTAATAA (SEQ ID NO:74)
9. P79AHA+E77S	MDIDPYKEFGATVE LLSFLPSDFFPSVR CLLDTAAALYRDAL ESPEHCSPHHTAL RQAILCWGDLSTLA TWVGTNLSD(AHA) ASRDLVVSYVNTN VGLKFRQLLWFHIS CLTFGRETVLEYLV SFGVWICTPPAYR PPNAPILSTLPETT VV (SEQ ID NO:23)	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACC GTTGAACTGCTGTCTTTCCTGCCGTCTGATTTCTTCCC GTCTGTTCGTTGCCTGCTGGACACCGCGGCAGCACT GTACCGTGACGCGCTGGAATCTCCGGAACATTGTTCT CCGCATCACACTGCGCTGCGTCAGGCGATTCTGTGC TGGGGCGACCTGagcACCCTGGCGACTTGGGTTGGC ACCAACCTGagcGATatgGCGTCTCGTGATCTGGTTGTT TCTTACGTTAACACTAACGTTGGTCTGAAATTCCGTCA GCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGGT CGTGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTG TTTGGATTTGTACTCCGCCGGCTTACCGTCCGCCGAA CGCACCGATCCTGAGCACCCTGCCGGAAACCACTGT TGTGTAATAA (SEQ ID NO:75)

10. P79AHA+D78S	MDIDPYKEFGATVE LLSFLPSDFFPSVR CLLDTAAALYRDAL ESPEHCSPHHTAL RQAILCWGDLSTLA TWVGTNLES(AHA) ASRDLVVSYVNTN VGLKFRQLLWFHIS CLTFGRETVLEYLV SFGVWICTPPAYR PPNAPILSTLPETT VV (SEQ ID NO:24)	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACC GTTGAACTGCTGTCTTTCCTGCCGTCTGATTTCTTCCC GTCTGTTCGTTGCCTGCTGGACACCGCGGCAGCACT GTACCGTGACGCGCTGGAATCTCCGGAACATTGTTCT CCGCATCACACTGCGCTGCGTCAGGCGATTCTGTGC TGGGGCGACCTGagcACCCTGGCGACTTGGGTTGGC ACCAACCTGGAAagcatgGCGTCTCGTGATCTGGTTGTT TCTTACGTTAACACTAACGTTGGTCTGAAATTCCGTCA GCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGGT CGTGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTG TTTGGATTTGTACTCCGCCGGCTTACCGTCCGCCGAA CGCACCGATCCTGAGCACCCTGCCGGAAACCACTGT TGTGTAATAA (SEQ ID NO:76)
11. P79AHA+E77S+D78S	MDIDPYKEFGATVE LLSFLPSDFFPSVR CLLDTAAALYRDAL ESPEHCSPHHTAL RQAILCWGDLSTLA TWVGTNLSS(AHA) ASRDLVVSYVNTN VGLKFRQLLWFHIS CLTFGRETVLEYLV SFGVWICTPPAYR PPNAPILSTLPETT VV (SEQ ID NO:25)	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACC GTTGAACTGCTGTCTTTCCTGCCGTCTGATTTCTTCCC GTCTGTTCGTTGCCTGCTGGACACCGCGGCAGCACT GTACCGTGACGCGCTGGAATCTCCGGAACATTGTTCT CCGCATCACACTGCGCTGCGTCAGGCGATTCTGTGC TGGGGCGACCTGagcACCCTGGCGACTTGGGTTGGC ACCAACCTGagcagcatgGCGTCTCGTGATCTGGTTGTT TCTTACGTTAACACTAACGTTGGTCTGAAATTCCGTCA GCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGGT CGTGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTG TTTGGATTTGTACTCCGCCGGCTTACCGTCCGCCGAA CGCACCGATCCTGAGCACCCTGCCGGAAACCACTGT TGTGTAATAA (SEQ ID NO:77)
12. A80AHA	MDIDPYKEFGATVE LLSFLPSDFFPSVR CLLDTAAALYRDAL ESPEHCSPHHTAL RQAILCWGDLSTLA TWVGTNLEDP(AH A)SRDLVVSYVNTN VGLKFRQLLWFHIS CLTFGRETVLEYLV SFGVWICTPPAYR PPNAPILSTLPETT VV (SEQ IDNO:26)	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACC GTTGAACTGCTGTCTTTCCTGCCGTCTGATTTCTTCCC GTCTGTTCGTTGCCTGCTGGACACCGCGGCAGCACT GTACCGTGACGCGCTGGAATCTCCGGAACATTGTTCT CCGCATCACACTGCGCTGCGTCAGGCGATTCTGTGC TGGGGCGACCTGagcACCCTGGCGACTTGGGTTGGC ACCAACCTGGAAGATCCGatgTCTCGTGATCTGGTTGT TTCTTACGTTAACACTAACGTTGGTCTGAAATTCCGTC AGCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGG TCGTGAAACCGTTCTGGAATACCTGGTTTCTTTTGGT GTTTGGATTTGTACTCCGCCGGCTTACCGTCCGCCGA ACGCACCGATCCTGAGCACCCTGCCGGAAACCACTG TTGTGTAATAA (SEQ ID NO:78)
13. A80AHA+E77K	MDIDPYKEFGATVE LLSFLPSDFFPSVR CLLDTAAALYRDAL ESPEHCSPHHTAL RQAILCWGDLSTLA TWVGTNLKDP(AH A)SRDLVVSYVNTN VGLKFRQLLWFHIS CLTFGRETVLEYLV SFGVWICTPPAYR PPNAPILSTLPETT VV (SEQ ID NO:27)	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACC GTTGAACTGCTGTCTTTCCTGCCGTCTGATTTCTTCCC GTCTGTTCGTTGCCTGCTGGACACCGCGGCAGCACT GTACCGTGACGCGCTGGAATCTCCGGAACATTGTTCT CCGCATCACACTGCGCTGCGTCAGGCGATTCTGTGC TGGGGCGACCTGagcACCCTGGCGACTTGGGTTGGC ACCAACCTGaaaGATCCGatgTCTCGTGATCTGGTTGTT TCTTACGTTAACACTAACGTTGGTCTGAAATTCCGTCA GCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGGT CGTGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTG TTTGGATTTGTACTCCGCCGGCTTACCGTCCGCCGAA CGCACCGATCCTGAGCACCCTGCCGGAAACCACTGT TGTGTAATAA (SEQ ID NO:79)

14. A80AHA+D78K	MDIDPYKEFGATVE LLSFLPSDFFPSVR CLLDTAAALYRDAL ESPEHCSPHHTAL RQAILCWGDLSTLA TWVGTNLEKPFAH AISRDLVVSYVNTN VGLKFRQLLWFHIS CLTFGRETVLEYLV SFGVWICTPPAYR PPNAPILSTLPETT VV (SEQ ID NO:28)	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACC GTTGAACTGCTGTCTTTCCTGCCGTCTGATTTCTTCCC GTCTGTTCGTTGCCTGCTGGACACCGCGGCAGCACT GTACCGTGACGCGCTGGAATCTCCGGAACATTGTTCT CCGCATCACACTGCGCTGCGTCAGGCGATTCTGTGC TGGGGCGACCTGagcACCCTGGCGACTTGGGTTGGC ACCAACCTGGAAaaaCCGatgTCTCGTGATCTGGTTGTT TCTTACGTTAACACTAACGTTGGTCTGAAATTCCGTCA GCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGGT CGTGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTG TTTGGATTTGTACTCCGCCGGCTTACCGTCCGCCGAA CGCACCGATCCTGAGCACCCTGCCGGAAACCACTGT TGTGTAATAA (SEQ ID NO:80)
15. A80AHA+E77S	MDIDPYKEFGATVE LLSFLPSDFFPSVR CLLDTAAALYRDAL ESPEHCSPHHTAL RQAILCWGDLSTLA TWVGTNLSDPFAH AISRDLVVSYVNTN VGLKFRQLLWFHIS CLTFGRETVLEYLV SFGVWICTPPAYR PPNAPILSTLPETT VV (SEQ ID NO:29)	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACC GTTGAACTGCTGTCTTTCCTGCCGTCTGATTTCTTCCC GTCTGTTCGTTGCCTGCTGGACACCGCGGCAGCACT GTACCGTGACGCGCTGGAATCTCCGGAACATTGTTCT CCGCATCACACTGCGCTGCGTCAGGCGATTCTGTGC TGGGGCGACCTGagcACCCTGGCGACTTGGGTTGGC ACCAACCTGagcGATCCGatgTCTCGTGATCTGGTTGTT TCTTACGTTAACACTAACGTTGGTCTGAAATTCCGTCA GCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGGT CGTGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTG TTTGGATTTGTACTCCGCCGGCTTACCGTCCGCCGAA CGCACCGATCCTGAGCACCCTGCCGGAAACCACTGT TGTGTAATAA (SEQ ID NO:81)
16. A80AHA+D78S	MDIDPYKEFGATVE LLSFLPSDFFPSVR CLLDTAAALYRDAL ESPEHCSPHHTAL RQAILCWGDLSTLA TWVGTNLESPFAH AISRDLVVSYVNTN VGLKFRQLLWFHIS CLTFGRETVLEYLV SFGVWICTPPAYR PPNAPILSTLPETT VV (SEQ ID NO:30)	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACC GTTGAACTGCTGTCTTTCCTGCCGTCTGATTTCTTCCC GTCTGTTCGTTGCCTGCTGGACACCGCGGCAGCACT GTACCGTGACGCGCTGGAATCTCCGGAACATTGTTCT CCGCATCACACTGCGCTGCGTCAGGCGATTCTGTGC TGGGGCGACCTGagcACCCTGGCGACTTGGGTTGGC ACCAACCTGGAAagcCCGatgTCTCGTGATCTGGTTGTT TCTTACGTTAACACTAACGTTGGTCTGAAATTCCGTCA GCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGGT CGTGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTG TTTGGATTTGTACTCCGCCGGCTTACCGTCCGCCGAA CGCACCGATCCTGAGCACCCTGCCGGAAACCACTGT TGTGTAATAA (SEQ ID NO:82)
17. A80AHA+E77S+D78S	MDIDPYKEFGATVE LLSFLPSDFFPSVR CLLDTAAALYRDAL ESPEHCSPHHTAL RQAILCWGDLSTLA TWVGTNLSSPFAH AISRDLVVSYVNTN VGLKFRQLLWFHIS CLTFGRETVLEYLV SFGVWICTPPAYR PPNAPILSTLPETT VV (SEQ ID NO:31)	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACC GTTGAACTGCTGTCTTTCCTGCCGTCTGATTTCTTCCC GTCTGTTCGTTGCCTGCTGGACACCGCGGCAGCACT GTACCGTGACGCGCTGGAATCTCCGGAACATTGTTCT CCGCATCACACTGCGCTGCGTCAGGCGATTCTGTGC TGGGGCGACCTGagcACCCTGGCGACTTGGGTTGGC ACCAACCTGagcagcCCGatgTCTCGTGATCTGGTTGTT TCTTACGTTAACACTAACGTTGGTCTGAAATTCCGTCA GCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGGT CGTGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTG TTTGGATTTGTACTCCGCCGGCTTACCGTCCGCCGAA CGCACCGATCCTGAGCACCCTGCCGGAAACCACTGT TGTGTAATAA (SEQ ID NO:83)

[0097] 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 (Jewett and Swartz 2004) 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 (Goerke and Swartz 2008). For global replacement of methionines in HBc proteins, methionine was left out of cell-free reaction mixtures, and substituted by 1 mM azidohomoalanine (AHA) (Medchem Source LLP, Federal Way, WA), a non-natural amino acid that displays an azide moiety. All reagents were obtained from Sigma–Aldrich (St. Louis, MO) unless otherwise noted.

[0098] 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 and a Beckman LS3801 liquid scintillation counter (Beckman Coulter, Fullerton, CA).

[0099] The production of GM-CSF, IM9-ScFv and CpG DNA with an alkyne moiety were described by Patel *et al*. The production of flagellin with an alkyne moiety was described by Lu *et al*.

[00100] Size-exclusion chromatography (SEC). To remove unincorporated L-[U-14C] leucine, the cell-free product was immediately dialyzed in 6-8000 MWCO Specra/Pro Molecularporous 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 2 buffer exchanges. The dialyzed cell-free reaction product was loaded on an SEC (size-exclusion chromatography) column packed with Sepharose 6 FastFlow resin (GE Healthcare). The running buffer is as the dialysis buffer with 5 mM DTT. The protein concentrations of eluate fractions were determined based on radioactivity.

[00101] Sucrose gradient sedimentation. The isolated VLPs from SEC were firstly dialyzed against the Dialysis Buffer to remove DTT, and were then oxidized to form disulfide bonds by adding 10 mM H_2O_2 or 10 mM diamide and incubating at room temperature for 1 h. The oxidants were removed by dialysis against the Dialysis Buffer with 2 buffer exchanges. The oxidized VLPs were assessed by velocity sedimentation. Ten to forty percentage weight per volume continuous sucrose density gradients were prepared in Dialysis Buffer in Polyallose 16

× 102 mm Centrifuge Tubes (Beckman) with the Gradient Master Ver3.05L Gradient Maker (Biocomp Instruments, Inc., Fredericton, Canada). The VLP product (200 µL) was 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 °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.

[00102] SDS-PAGE and autoradiogram. Protein size was analyzed by SDS-PAGE and autoradiography. NuPAGE Novex precast gels and reagents were purchased from Invitrogen (Carlsbad, CA). For reducing SDS-PAGE, samples were denatured for 10 min at 95 °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 a storage phosphor screen (Molecular Dynamics), which was subsequently scanned using a Typhoon Scanner (GE Healthcare).

[00103] 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 ([(CH3CN)4Cu]PF6 or simply Cu (I) catalyst) (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, IM9-ScFv 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.

Results and discussions

[00104] Stabilization of HBc VLP by introducing covalent disulfide bonds. In this study, HBc capsid protein truncated at V149 was used. The truncated HBc protein self-assembles into VLPs, composed of 240 subunits arranged with T = 4 icosahedral symmetry. The T = 4 icosahedral capsid has 12 regular pentagonal faces and 30 regular hexagonal faces, as shown in Fig. 1. Each pentamer is surrounded by 5 hexamers. A five-fold unit of HBc dimers and a six-fold unit of HBc dimers comprise one pentamer and one hexamer, respectively, as

shown in Fig. 2. The 5-fold unit and the 6-fold unit share one dimer, and are cross-linked to form the VLP.

[00105] There are four types of bonding interactions between "side chains" including: hydrogen bonding, salt bridges, disulfide bonds, and non-polar hydrophobic interactions. The covalent disulfide bonds are the strongest. After analyzing the results of many conjugation reactions, we discovered that the HBc VLP scaffold was not stable during the click conjugation reaction. To address this, we assessed the 3D tertiary structure of the capsid to identify sites where neighboring monomers were close enough to be linked by disulfide bonds. If we could introduce disulfide bridges to cross-link the 5-fold and 6-fold units, the VLP could be stabilized.

[00106] To form stable VLPs, possible disulfide (S-S) bond positions at both 5-fold unit and 6-fold unit were sought. We devised two strategies. The first is looking for two amino acids with the shortest distances between the side chains of dimers, as shown in Fig. 3(a). In the second, we examined possible disulfide bonds between the C-terminal regions of monomers, as shown in Fig. 3(b). The distances suggested by the icosahedral capsid structure are shown in parentheses. Because one original disulfide bond already exists in the dimer, the introduction of new disulfide bonds can greatly stabilize the VLP. A total of 10 positions was selected, including SS1 (D29-R127), SS2 (T109-V120), SS3 (Y132-N136), SS4 (Y132-A137), SS5 (R133-N136), SS6 (R133-A137), SS7 (P134-P135), SS8 (P134-N136), SS9 (P134-A137), and SS10 (P135-N136). The codons for these amino acids were then changed to codons for cysteine.

[00107] To form correct S-S bonds, the HBc proteins were firstly synthesized in the CFPS system in a reducing environment. The HBc protein can self-assemble into VLPs in the CFPS system. The assembled VLPs were then purified using size-exclusion chromatography (SEC) in a reducing environment. The maintenance of the reducing environment is to prevent the formation of incorrect S-S bonds before assembly of the VLPs. The purified VLPs were then oxidized to form S-S bonds by adding oxidants (hydrogen peroxide or diamide). The particle sizes were finally assessed using sucrose gradient centrifugation. The whole procedure is shown in Fig. 4.

[00108] The CFPS results showed that the mutants all have similar total yields as the orginal HBc protein (Fig. 5 (a)). All 10 candidates expressed with good soluble yields (≥ 300μg/ml). The non-reducing SDS-PAGE autoradiograph (Fig. 5(b)) showed that only a small amount of dimers formed in the CFPS system for the mutants. The CFPS products were then purified using SEC (Fig. 6). The mutants SS3, SS4, SS5 and SS6 did not assemble into VLPs. The SEC fractions 9-11 were pooled and oxidized by the addition of 10 mM H_2O_2 or 10 mM diamide. After the oxidization treatment, the fully cross-linked VLPs could not be disassembled with SDS and failed to enter the SDS-PAGE gel (Fig. 7) . We could see some monomers and dimers for mutants SS2, SS7, SS8, SS9 and SS10, which demonstrated that

some S-S bonds were not formed. Only SS1 particles all stayed in the well of SDS-PAGE gel after oxidation with diamide, and have no monomer or dimer, 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 completely cross-linked if treated with H_2O_2. The oxidized particles were then assessed using sucrose gradient centrifugation to indicate proper assembly (Fig. 8). The results showed that the sizes of SS1, SS7, SS8, SS9 and SS10 VLPs were all correct.

[00109] To further verify the functions of the oxidized VLPs, the click chemistry reactions were then tested. Flagellin protein and GM-CSF protein were used as example adducts. CFPS provides a facile means for introducing nnAAs with an alkyne moiety into flagellins and GM-CSFs, and nnAAs with an azide moiety into VLPs (Bundy et al. 2008; Goerke and Swartz 2008; Patel and Swartz 2011). This would allow the direct coupling of flagellin or GM-CSF to VLPs using Cu(I)-catalyzed [3 + 2] cycloaddition click chemistry reaction. The reaction results (Fig. 9) showed that flagellin and GM-CSF could be successfully conjugated to SS1, SS7, SS8, SS9 and SS10 VLPs.

[00110] Based on the results above, mutant SS1 VLP was the most effectively stabilized VLP, and SS8 was nearly as good.

Example 2

[00111] At physiological pH, the surface of the HBc VLP is negatively charged. The surface of the HBc VLP is dominated by 120 dimer spikes that stick out from the surface. These spikes are obvious attachment sites with high steric availability, but they are terminated with four negatively charged amino acids (E77 × 2, D78 × 2). The negative charges on the VLP surface are from these four negatively charged amino acids on the spike tip. Because like charges repel, molecules with negative charges cannot get close to HBc VLP, and therefore the click chemistry conjugation cannot proceed effectively. The removal of the surface negative charges greatly decreased the protein's solubility, and HBc proteins could not self-assemble into VLPs very well. The following example addresses this problem.

[00112] *Transplant of new spike*. Removal of HBc VLP surface negative charges decreased its solubility and self-assembly ability. This led us to seek alternative methods for remodeling the HBc dimer spike in naturally occurring viral mutants. A few natural mutants with only one negatively charged amino acid on the protein monomer tip were found. One natural mutant (UniProt accession number: Q8B6N7) was selected and compared with the native HBc protein (UniProt accession number: P03147). In genotype Q8B6N7, one net negative charge is still at the 78 site, but another one transfers to the 70 site from the original 77 site in genotype P03147.

[00113] To determine the effects of these sequence changes in Q8B6N7 on the characteristics of the VLP, the spike (the spike tip (ST) or the whole hydrophobic pocket (HP)) in Q8B6N7 was transplanted into native HBc proteins (Fig. 10). Two new mutants HBc(ST) and HBc(HP)

were created and compared with mutant HBc(D78M). HBc(D78M) is the native HBc protein with the removal of one negative charge (D78) on the tip. HBc(D78M), HBc(ST), and HBc(HP) were stabilized by introducing new disulfide bridges (D29C-R127C). After one negative charge on the tip of the native protein was removed, the solubility of the mutant HBc(D78M) was only around 50% (Fig. 11). These two new mutants HBc(ST) and HBc(HP) improved solubility greatly (Fig. 11). HBc(HP) was more soluble than HBc(ST). Native HBc proteins can self-assemble into VLPs in buffer with 0.5 M NaCl, but HBc(ST) and HBc(HP) could not self-assemble into VLPs after dialysis against a buffer with 0.5 M NaCl (Fig. 12), which further suggested that the negative charges on the virus capsid surface are important for the particle assembly. After dialysis against buffer with a higher concentration of NaCl (1.5 M), the modified HBc proteins could assemble into VLPs by SEC analysis. Because HBc(HP) (SEQ ID NO:35) had better solubility than HBc(ST), HBc(HP) was chosen in the subsequent development.

[00114] *Selection of non-natural amino acid (nnAA) sites on the spike tip.* The surface of the HBc VLP is dominated by 120 dimer spikes that stick out from the surface. The tip of these spikes are obvious attachment sites with high steric availability. To identify the best conjugation site, six different nnAA sites (N75AHA, L76AHA, Q77AHA, D78AHA, Q79AHA, and A80AHA) on the spike tip were tried, as shown in Fig. 13. These mutants were all stabilized by introducing new disulfide bridges (D29C-R127C). AHA signifies the nnAA, azidohomoalanine, with an azide moiety. Incorporating the nnAA site at D78 removes the negative charges at the spike tip.

[00115] CFPS results showed that all these mutants were very soluble except mutant D78AHA (Fig. 14). However, after dialysis against buffer with 1.5 M NaCl, the soluble yield of mutant D78AHA could reach 70%. SEC results showed that all these mutants could self-assemble into VLPs (Fig. 15).

[00116] The ability to conjugate ligands onto these mutated VLPs were then tested with click chemistry reactions. We used four molecules with the alkyne functional group, including flagellin protein, GMCSF protein, IM9-ScFv protein, and CpG DNA. IM9-scFv and CpG DNA are characterized by negative charge density near the alkyne functional group. As previously mentioned, the surface of native HBc VLPs were negatively charged, so molecules with negative charges cannot get close to native HBc VLP surface, and therefore the click chemistry conjugation cannot proceed effectively. The mutant HBc(HP) D78AHA had no negative charge on the spike tip. The conjugation reaction results (Fig. 16) showed that the removal of surface negative charges on the HBc VLP did greatly improve the conjugation efficiency of IM9-scFv and CpG DNA, especially for CpG DNA. Based on the results above, mutant HBc(HP) D78AHA was the best VLP in terms of conjugation efficiency.

[00117] Stabilization by the introduction of disulfide bridges. HBc protein truncated at amino acid 149 has been shown to form predominantly (> 95%) the *T*=4 VLP. Two monomers (16.7 kDa) associate to give a compact dimer (33.5 kDa). At the dimer interface, there is a disulfide bridge between the Cys-61 residues of the two monomers further stabilizing the dimer (Fig. 17). Dimers (120 copies) then self-assemble into the *T*=4 VLP by electrostatic interactions, hydrogen bonds and weak hydrophobic interactions. The T = 4 icosahedral capsid has 12 regular pentagonal faces and 30 regular hexagonal faces (Fig. 17). Each pentamer is surrounded by 5 hexamers. The five-fold unit of HBc dimer and the six-fold unit of HBc dimer comprise of one pentamer and one hexamer, respectively. The 5-fold unit and the 6-fold unit share one dimer, and are cross-linked to form the VLP. At the interface of the 5-fold unit and the 6-fold unit, there is an intradimer disulfide bond.

[00118] In order to strengthen the HBc VLP scaffold after conjugation with big molecules or multiple molecules, covalent disulfide bridges that would stabilize both the 5-fold and 6-fold units were introduced (Fig. 18).

[00119] To form stable VLPs, consistent possible disulfide (S-S) bond positions in both 5-fold unit and 6-fold unit were searched. In total, 10 positions were selected, 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). The codons for chosen amino acids were changed to codons for cysteine. After CFPS, the VLPs were purified by SEC, as shown in Fig. 19. The VLP fractions were pooled and oxidized by the addition of 20 mM diamide. After the oxidization treatment, the fully cross-linked VLPs could not be disassembled with SDS and failed to enter the SDS-PAGE gel (Fig. 20). Only SS1 particles stayed completely in the sample addition well of the SDS-PAGE gel, which demonstrated that all S-S bonds were formed in SS1 VLPs. However, SS8 was nearly as stable. Based on these results, mutants SS1 and SS8 had the desired characteristics. To make the disulfide network in the VLP stronger, SS1 and SS8 were introduced in the VLP at the same time. SEC analysis showed that mutation SS1+SS8 did not hinder the VLP assembly. SDS-PAGE analysis also indicated that all S-S bonds were formed in the SS1+SS8 VLPs.

[00120] Based on the results above, mutant HBc(HP) D78AHA SS1+SS8 (SEQ ID NO:52) was the most stabilized VLP and also provides high conjugation efficiency.

Materials and methods

[00121] Plasmid construction. The sequence encoding the human Hepatitis B core (HBc) capsid monomer of subtype adyw (Pasek et al., 1979) 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).

[00122] Sequences of HBc protein variants. The sequences of wild-type, HBc(D78M), HBc(ST) and HBc(HP) were shown in the table below. HBc(D78M), HBc(ST), and HBc(HP) were stabilized by introducing new disulfide bridge SS1(D29C-R127C). All sequences include the SS1 mutations and the M66S mutation (underlined).

Variants	Protein sequence	DNA encoding sequence
Wild-type (UniProt accession number P03147, truncated at 149)	MDIDPYKEFGATVEL LSFLPSDFFPSVRDL LDTAAALYRDALESP EHCSPHHTALRQAIL CWGDLMTLATWVG TNLEDPASRDLVVS YVNTNVGLKFRQLL WFHISCLTFGRETVL EYLVSFGVWIRTPP AYRPPNAPILSTLPE TTVV (SEQ ID NO:32)	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAA CTGCTGTCTTTCCTGCCGTCTGATTTCTTCCCGTCTGTTCGTG ACCTGCTGGACACCGCGGCAGCACTGTACCGTGACGCGCTG GAATCTCCGGAACATTGTTCTCCGCATCACACTGCGCTGCGT CAGGCGATTCTGTGCTGGGGCGACCTGATGACCCTGGCGAC TTGGGTTGGCACCAACCTGGAAGATCCGGCGTCTCGTGATC TGGTTGTTTCTTACGTTAACACTAACGTTGGTCTGAAATTCCG TCAGCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGGTCG TGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGATT CGTACTCCGCCGGCTTACCGTCCGCCGAACGCACCGATCCT GAGCACCCTGCCGGAAACCACTGTTGTGTAATAA
HBc(D78M)	MDIDPYKEFGATVEL LSFLPSDFFPSVRQL LDTAAALYRDALESP EHCSPHHTALRQAIL CWGDLSTLATWVG TNLEMPASRDLVVS YVNTNVGLKFRQLL WFHISCLTFGRETVL EYLVSFGVWIQTPP AYRPPNAPILSTLPE TTVV (SEQ ID NO:33)	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAA CTGCTGTCTTTCCTGCCGTCTGATTTCTTCCCGTCTGTTCGTI GCCTGCTGGACACCGCGGCAGCACTGTACCGTGACGCGCTG GAATCTCCGGAACATTGTTCTCCGCATCACACTGCGCTGCGT CAGGCGATTCTGTGCTGGGGCGACCTGagcACCCTGGCGACT TGGGTTGGCACCAACCTGGAA*atg*CCGGCGTCTCGTGATCTG GTTGTTTCTTACGTTAACACTAACGTTGGTCTGAAATTCCGTC AGCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGGTCGTG AAACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGATTIG IACTCCGCCGGCTTACCGTCCGCCGAACGCACCGATCCTGA GCACCCTGCCGGAAACCACTGTTGTGTAATAA
HBc(ST)	MDIDPYKEFGATVEL LSFLPSDFFPSVRQL LDTAAALYRDALESP EHCSPHHTALRQAIL CWGDLSTLATWVG NNMQDQAARDLVV SYVNTNVGLKFRQL LWFHISCLTFGRETV LEYLVSFGVWIQTPP AYRPPNAPILSTLPE TTVV (SEQ ID NO:34)	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAA CTGCTGTCTTTCCTGCCGTCTGATTTCTTCCCGTCTGTTCGTI GCCTGCTGGACACCGCGGCAGCACTGTACCGTGACGCGCTG GAATCTCCGGAACATTGTTCTCCGCATCACACTGCGCTGCGT CAGGCGATTCTGTGCTGGGGCGACCTGagcACCCTGGCGACT TGGGTTGGC*AACAAC**atg**CAGGATCAGGCGGCG*CGTGATCTG GTTGTTTCTTACGTTAACACTAACGTTGGTCTGAAATTCCGTC AGCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGGTCGTG AAACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGATTIG IACTCCGCCGGCTTACCGTCCGCCGAACGCACCGATCCTGA GCACCCTGCCGGAAACCACTGTTGTGTAATAA

HBc(HP)	MDIDPYKEFGATVEL LSFLPSDFFPSVRCL LDTAAALYRDALESP EHCSPHHTALRQAV SCWREVTDFGDWV GNNMQDQAARDLV VNYVNANIGLKIRQL LWFHISCLTFGRETV LEYLVSFGVWICTPP AYRPPNAPILSTLPE TTVV (SEQ ID NO:35)	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAA CTGCTGTCTTTCCTGCCGTCTGATTTCTTCCCGTCTGTTCGTT GCCTGCTGGACACCGCGGCAGCACTGTACCGTGACGCGCTG GAATCTCCGGAACATTGTTCTCCGCATCACACTGCGCTGCGT CAGGCGGTGAGCTGCTGGCGCGAAGTGACCGATTTTGGCGA TTGGGTGGGCAACAACatgCAGGATCAGGCGGCGCGCGATCT GGTGGTGAACTATGTGAACGCGAACATTGGCCTGAAAATTCG TCAGCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGGTCG TGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGATT TGTACTCCGCCGGCTTACCGTCCGCCGAACGCACCGATCCT GAGCACCCTGCCGGAAACCACTGTTGTGTAATAA

[00123] The sequences of HBc(HP) with different AHA sites are shown in the table below. They were all stabilized by introducing new disulfide bridges SS1(D29C-R127C). AHA signifies azidohomoalanine, the non-natural amino acid tested.

Variants	Protein sequence	DNA encoding sequence
HBc(HP), N75AHA	(AHA)DIDPYKEFGAT VELLSFLPSDFFPSV RCLLDTAAALYRDAL ESPEHCSPHHTALR QAVSCWREVTDFG DWVGN(AHA)LQDQ AARDLVVNYVNANI GLKIRQLLWFHISCL TFGRETVLEYLVSF GVWICTPPAYRPPN APILSTLPETTVV (SEQ ID NO:36)	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAA CTGCTGTCTTTCCTGCCGTCTGATTTCTTCCCGTCTGTTCGTT GCCTGCTGGACACCGCGGCAGCACTGTACCGTGACGCGCTG GAATCTCCGGAACATTGTTCTCCGCATCACACTGCGCTGCGT CAGGCGGTGAGCTGCTGGCGCGAAGTGACCGATTTTGGCGA TTGGGTGGGCAACatgCTGCAGGATCAGGCGGCGCGCGATC TGGTGGTGAACTATGTGAACGCGAACATTGGCCTGAAAATTC GTCAGCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGGTC GTGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGAT TTGTACTCCGCCGGCTTACCGTCCGCCGAACGCACCGATCC TGAGCACCCTGCCGGAAACCACTGTTGTGTAATAA
HBc(HP), L76AHA	(AHA)DIDPYKEFGAT VELLSFLPSDFFPSV RCLLDTAAALYRDAL ESPEHCSPHHTALR QAVSCWREVTDFG DWVGNN(AHA)QDQ AARDLVVNYVNANI GLKIRQLLWFHISCL TFGRETVLEYLVSF GVWICTPPAYRPPN APILSTLPETTVV (SEQ ID NO:37)	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAA CTGCTGTCTTTCCTGCCGTCTGATTTCTTCCCGTCTGTTCGTT GCCTGCTGGACACCGCGGCAGCACTGTACCGTGACGCGCTG GAATCTCCGGAACATTGTTCTCCGCATCACACTGCGCTGCGT CAGGCGGTGAGCTGCTGGCGCGAAGTGACCGATTTTGGCGA TTGGGTGGGCAACAACatgCAGGATCAGGCGGCGCGCGATCT GGTGGTGAACTATGTGAACGCGAACATTGGCCTGAAAATTCG TCAGCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGGTCG TGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGATT TGTACTCCGCCGGCTTACCGTCCGCCGAACGCACCGATCCT GAGCACCCTGCCGGAAACCACTGTTGTGTAATAA
HBc(HP), Q77AHA	(AHA)DIDPYKEFGAT VELLSFLPSDFFPSV RCLLDTAAALYRDAL ESPEHCSPHHTALR QAVSCWREVTDFG DWVGNNL(AHA)DQ AARDLVVNYVNANI GLKIRQLLWFHISCL TFGRETVLEYLVSF GVWICTPPAYRPPN APILSTLPETTVV (SEQ ID NO:38)	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAA CTGCTGTCTTTCCTGCCGTCTGATTTCTTCCCGTCTGTTCGTT GCCTGCTGGACACCGCGGCAGCACTGTACCGTGACGCGCTG GAATCTCCGGAACATTGTTCTCCGCATCACACTGCGCTGCGT CAGGCGGTGAGCTGCTGGCGCGAAGTGACCGATTTTGGCGA TTGGGTGGGCAACAACCTGatgGATCAGGCGGCGCGCGATCT GGTGGTGAACTATGTGAACGCGAACATTGGCCTGAAAATTCG TCAGCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGGTCG TGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGATT TGTACTCCGCCGGCTTACCGTCCGCCGAACGCACCGATCCT GAGCACCCTGCCGGAAACCACTGTTGTGTAATAA

HBc(HP), D78AHA	(AHA)DIDPYKEFGAT VELLSFLPSDFFPSV RCLLDTAAALYRDAL ESPEHCSPHHTALR QAVSCWREVTDFG DWVGNNLQ(AHA)Q AARDLVVNYVNANI GLKIRQLLWFHISCL TFGRETVLEYLVSF GVWICTPPAYRPPN APILSTLPETTVV (SEQ ID NO:39)	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAA CTGCTGTCTTTCCTGCCGTCTGATTTCTTCCCGTCTGTTCGTT GCCTGCTGGACACCGCGGCAGCACTGTACCGTGACGCGCTG GAATCTCCGGAACATTGTTCTCCGCATCACACTGCGCTGCGT CAGGCGGTGAGCTGCTGGCGCGAAGTGACCGATTTTGGCGA TTGGGTGGGCAACAACCTGCAGatgCAGGCGGCGCGCGATC TGGTGGTGAACTATGTGAACGCGAACATTGGCCTGAAAATTC GTCAGCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGGTC GTGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGAT TTGTACTCCGCCGGCTTACCGTCCGCCGAACGCACCGATCC TGAGCACCCTGCCGGAAACCACTGTTGTGTAATAA
HBc(HP), Q79AHA	(AHA)DIDPYKEFGAT VELLSFLPSDFFPSV RCLLDTAAALYRDAL ESPEHCSPHHTALR QAVSCWREVTDFG DWVGNNLQD(AHA) AARDLVVNYVNANI GLKIRQLLWFHISCL TFGRETVLEYLVSF GVWICTPPAYRPPN APILSTLPETTVV (SEQ ID NO:40)	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAA CTGCTGTCTTTCCTGCCGTCTGATTTCTTCCCGTCTGTTCGTT GCCTGCTGGACACCGCGGCAGCACTGTACCGTGACGCGCTG GAATCTCCGGAACATTGTTCTCCGCATCACACTGCGCTGCGT CAGGCGGTGAGCTGCTGGCGCGAAGTGACCGATTTTGGCGA TTGGGTGGGCAACAACCTGCAGGATatgGCGGCGCGCGATCT GGTGGTGAACTATGTGAACGCGAACATTGGCCTGAAAATTCG TCAGCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGGTCG TGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGATT TGTACTCCGCCGGCTTACCGTCCGCCGAACGCACCGATCCT GAGCACCCTGCCGGAAACCACTGTTGTGTAATAA
HBc(HP), A80AHA	(AHA)DIDPYKEFGAT VELLSFLPSDFFPSV RCLLDTAAALYRDAL ESPEHCSPHHTALR QAVSCWREVTDFG DWVGNNLQDQ(AHA)ARDLVVNYVNANIG LKIRQLLWFHISCLT FGRETVLEYLVSFG VWICTPPAYRPPNA PILSTLPETTVV (SEQ ID NO:41)	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAA CTGCTGTCTTTCCTGCCGTCTGATTTCTTCCCGTCTGTTCGTT GCCTGCTGGACACCGCGGCAGCACTGTACCGTGACGCGCTG GAATCTCCGGAACATTGTTCTCCGCATCACACTGCGCTGCGT CAGGCGGTGAGCTGCTGGCGCGAAGTGACCGATTTTGGCGA TTGGGTGGGCAACAACCTGCAGGATCAGatgGCGCGCGATCT GGTGGTGAACTATGTGAACGCGAACATTGGCCTGAAAATTCG TCAGCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGGTCG TGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGATT TGTACTCCGCCGGCTTACCGTCCGCCGAACGCACCGATCCT GAGCACCCTGCCGGAAACCACTGTTGTGTAATAA

[00124] The sequences of HBc(HP) D78AHA with different disulfide bond sites are shown in the table below. AHA signifies azidohomoalanine, the non-natural amino acid tested.

Variants	Protein sequence	DNA encoding sequence
HBc(HP) D78AHA, SS1:D29C-R127C	(AHA)DIDPYKEFGAT VELLSFLPSDFFPSV RCLLDTAAALYRDAL ESPEHCSPHHTALR QAVSCWREVTDFG DWVGNNLQ(AHA)Q AARDLVVNYVNANI GLKIRQLLWFHISCL TFGRETVLEYLVSF GVWICTPPAYRPPN APILSTLPETTVV (SEQ ID NO:42)	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAA CTGCTGTCTTTCCTGCCGTCTGATTTCTTCCCGTCTGTTCGTT GCCTGCTGGACACCGCGGCAGCACTGTACCGTGACGCGCTG GAATCTCCGGAACATTGTTCTCCGCATCACACTGCGCTGCGT CAGGCGGTGAGCTGCTGGCGCGAAGTGACCGATTTTGGCGA TTGGGTGGGCAACAACCTGCAGatgCAGGCGGCGCGCGATC TGGTGGTGAACTATGTGAACGCGAACATTGGCCTGAAAATTC GTCAGCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGGTC GTGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGAT TTGTACTCCGCCGGCTTACCGTCCGCCGAACGCACCGATCC TGAGCACCCTGCCGGAAACCACTGTTGTGTAATAA
HBc(HP) D78AHA, SS2:T109C-V120C	(AHA)DIDPYKEFGAT VELLSFLPSDFFPSV RDLLDTAAALYRDAL ESPEHCSPHHTALR QAVSCWREVTDFG DWVGNNLQ(AHA)Q AARDLVVNYVNANI GLKIRQLLWFHISCL CFGRETVLEYLCSF GVWIRTPPAYRPPN APILSTLPETTVV (SEQ ID NO:43)	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAA CTGCTGTCTTTCCTGCCGTCTGATTTCTTCCCGTCTGTTCGTG ACCTGCTGGACACCGCGGCAGCACTGTACCGTGACGCGCTG GAATCTCCGGAACATTGTTCTCCGCATCACACTGCGCTGCGT CAGGCGGTGAGCTGCTGGCGCGAAGTGACCGATTTTGGCGA TTGGGTGGGCAACAACCTGCAGatgCAGGCGGCGCGCGATC TGGTGGTGAACTATGTGAACGCGAACATTGGCCTGAAAATTC GTCAGCTGCTGTGGTTCCACATCTCTTGCCTGTGCTTCGGTC GTGAAACCGTTCTGGAATACCTGTGTTCTTTTGGTGTTTGGAT TCGTACTCCGCCGGCTTACCGTCCGCCGAACGCACCGATCC TGAGCACCCTGCCGGAAACCACTGTTGTGTAATAA

HBc(HP) D78AHA, SS3: Y132C-N136C	(AHA)DIDPYKEFGAT	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAA
	VELLSFLPSDFFPSV	CTGCTGTCTTTCCTGCCGTCTGATTTCTTCCCGTCTGTTCGTG
	RDLLDTAAALYRDAL	ACCTGCTGGACACCGCGGCAGCACTGTACCGTGACGCGCTG
	ESPEHCSPHHTALR	GAATCTCCGGAACATTGTTCTCCGCATCACACTGCGCTGCGT
	QAVSCWREVTDFG	CAGGCGGTGAGCTGCTGGCGCGAAGTGACCGATTTTGGCGA
	DWVGNNLQ(AHA)Q	TTGGGTGGGCAACAACCTGCAGatgCAGGCGGCGCGCGATC
	AARDLVVNYVNANI	TGGTGGTGAACTATGTGAACGCGAACATTGGCCTGAAAATTC
	GLKIRQLLWFHISCL	GTCAGCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGGTC
	TFGRETVLEYLVSF	GTGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGAT
	GVWIRTPPACRPPC	TCGTACTCCGCCGGCTTGCCGTCCGCCGTGCGCACCGATCC
	APILSTLPETTVV	TGAGCACCCTGCCGGAAACCACTGTTGTGTAATAA
	(SEQ ID NO:44)	
HBc(HP) D78AHA, SS4: Y132C-A137C	(AHA)DIDPYKEFGAT	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAA
	VELLSFLPSDFFPSV	CTGCTGTCTTTCCTGCCGTCTGATTTCTTCCCGTCTGTTCGTG
	RDLLDTAAALYRDAL	ACCTGCTGGACACCGCGGCAGCACTGTACCGTGACGCGCTG
	ESPEHCSPHHTALR	GAATCTCCGGAACATTGTTCTCCGCATCACACTGCGCTGCGT
	QAVSCWREVTDFG	CAGGCGGTGAGCTGCTGGCGCGAAGTGACCGATTTTGGCGA
	DWVGNNLQ(AHA)Q	TTGGGTGGGCAACAACCTGCAGatgCAGGCGGCGCGCGATC
	AARDLVVNYVNANI	TGGTGGTGAACTATGTGAACGCGAACATTGGCCTGAAAATTC
	GLKIRQLLWFHISCL	GTCAGCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGGTC
	TFGRETVLEYLVSF	GTGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGAT
	GVWIRTPPACRPPN	TCGTACTCCGCCGGCTTGCCGTCCGCCGAACTGCCCGATCC
	CPILSTLPETTVV	TGAGCACCCTGCCGGAAACCACTGTTGTGTAATAA
	(SEQ ID NO:45)	
HBc(HP) D78AHA, SS5: R133C-N136C	(AHA)DIDPYKEFGAT	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAA
	VELLSFLPSDFFPSV	CTGCTGTCTTTCCTGCCGTCTGATTTCTTCCCGTCTGTTCGTG
	RDLLDTAAALYRDAL	ACCTGCTGGACACCGCGGCAGCACTGTACCGTGACGCGCTG
	ESPEHCSPHHTALR	GAATCTCCGGAACATTGTTCTCCGCATCACACTGCGCTGCGT
	QAVSCWREVTDFG	CAGGCGGTGAGCTGCTGGCGCGAAGTGACCGATTTTGGCGA
	DWVGNNLQ(AHA)Q	TTGGGTGGGCAACAACCTGCAGatgCAGGCGGCGCGCGATC
	AARDLVVNYVNANI	TGGTGGTGAACTATGTGAACGCGAACATTGGCCTGAAAATTC
	GLKIRQLLWFHISCL	GTCAGCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGGTC
	TFGRETVLEYLVSF	GTGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGAT
	GVWIRTPPAYCPPC	TCGTACTCCGCCGGCTTACTGCCCGCCGTGCGCACCGATCC
	APILSTLPETTVV	TGAGCACCCTGCCGGAAACCACTGTTGTGTAATAA
	(SEQ ID NO:46)	
HBc(HP) D78AHA, SS6: R133C-A137C	(AHA)DIDPYKEFGAT	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAA
	VELLSFLPSDFFPSV	CTGCTGTCTTTCCTGCCGTCTGATTTCTTCCCGTCTGTTCGTG
	RDLLDTAAALYRDAL	ACCTGCTGGACACCGCGGCAGCACTGTACCGTGACGCGCTG
	ESPEHCSPHHTALR	GAATCTCCGGAACATTGTTCTCCGCATCACACTGCGCTGCGT
	QAVSCWREVTDFG	CAGGCGGTGAGCTGCTGGCGCGAAGTGACCGATTTTGGCGA
	DWVGNNLQ(AHA)Q	TTGGGTGGGCAACAACCTGCAGatgCAGGCGGCGCGCGATC
	AARDLVVNYVNANI	TGGTGGTGAACTATGTGAACGCGAACATTGGCCTGAAAATTC
	GLKIRQLLWFHISCL	GTCAGCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGGTC
	TFGRETVLEYLVSF	GTGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGAT
	GVWIRTPPAYCPPN	TCGTACTCCGCCGGCTTACTGCCCGCCGAACTGCCCGATCC
	CPILSTLPETTVV	TGAGCACCCTGCCGGAAACCACTGTTGTGTAATAA
	(SEQ ID NO:47)	
HBc(HP) D78AHA, SS7: P134C-P135C	(AHA)DIDPYKEFGAT	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAA
	VELLSFLPSDFFPSV	CTGCTGTCTTTCCTGCCGTCTGATTTCTTCCCGTCTGTTCGTG
	RDLLDTAAALYRDAL	ACCTGCTGGACACCGCGGCAGCACTGTACCGTGACGCGCTG
	ESPEHCSPHHTALR	GAATCTCCGGAACATTGTTCTCCGCATCACACTGCGCTGCGT
	QAVSCWREVTDFG	CAGGCGGTGAGCTGCTGGCGCGAAGTGACCGATTTTGGCGA
	DWVGNNLQ(AHA)Q	TTGGGTGGGCAACAACCTGCAGatgCAGGCGGCGCGCGATC
	AARDLVVNYVNANI	TGGTGGTGAACTATGTGAACGCGAACATTGGCCTGAAAATTC
	GLKIRQLLWFHISCL	GTCAGCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGGTC
	TFGRETVLEYLVSF	GTGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGAT
	GVWIRTPPAYRCCN	TCGTACTCCGCCGGCTTACCGTTGCTGCAACGCACCGATCCT
	APILSTLPETTVV	GAGCACCCTGCCGGAAACCACTGTTGTGTAATAA
	(SEQ ID NO:48)	

HBc(HP) D78AHA, SS8: P134C-N136C	(AHA)DIDPYKEFGAT	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAA
	VELLSFLPSDFFPSV	CTGCTGTCTTTCCTGCCGTCTGATTTCTTCCCGTCTGTTCGTG
	RDLLDTAAALYRDAL	ACCTGCTGGACACCGCGGCAGCACTGTACCGTGACGCGCTG
	ESPEHCSPHHTALR	GAATCTCCGGAACATTGTTCTCCGCATCACACTGCGCTGCGT
	QAVSCWREVTDFG	CAGGCGGTGAGCTGCTGGCGCGAAGTGACCGATTTTGGCGA
	DWVGNNLQ(AHA)Q	TTGGGTGGGCAACAACCTGCAGatgCAGGCGGCGCGCGATC
	AARDLVVNYVNANI	TGGTGGTGAACTATGTGAACGCGAACATTGGCCTGAAAATTC
	GLKIRQLLWFHISCL	GTCAGCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGGTC
	TFGRETVLEYLVSF	GTGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGAT
	GVWIRTPPAYRCPC	TCGTACTCCGCCGGCTTACCGTTGCCCGTGCGCACCGATCC
	APILSTLPETTVV	TGAGCACCCTGCCGGAAACCACTGTTGTGTAATAA
	(SEQ ID NO:49)	
HBc(HP) D78AHA, SS9: P134C-A137C	(AHA)DIDPYKEFGAT	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAA
	VELLSFLPSDFFPSV	CTGCTGTCTTTCCTGCCGTCTGATTTCTTCCCGTCTGTTCGTG
	RDLLDTAAALYRDAL	ACCTGCTGGACACCGCGGCAGCACTGTACCGTGACGCGCTG
	ESPEHCSPHHTALR	GAATCTCCGGAACATTGTTCTCCGCATCACACTGCGCTGCGT
	QAVSCWREVTDFG	CAGGCGGTGAGCTGCTGGCGCGAAGTGACCGATTTTGGCGA
	DWVGNNLQ(AHA)Q	TTGGGTGGGCAACAACCTGCAGatgCAGGCGGCGCGCGATC
	AARDLVVNYVNANI	TGGTGGTGAACTATGTGAACGCGAACATTGGCCTGAAAATTC
	GLKIRQLLWFHISCL	GTCAGCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGGTC
	TFGRETVLEYLVSF	GTGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGAT
	GVWIRTPPAYRCPN	TCGTACTCCGCCGGCTTACCGTTGCCCGAACTGCCCGATCC
	CPILSTLPETTVV	TGAGCACCCTGCCGGAAACCACTGTTGTGTAATAA
	(SEQ ID NO:50)	
HBc(HP) D78AHA, SS10: P135C-N136C	(AHA)DIDPYKEFGAT	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAA
	VELLSFLPSDFFPSV	CTGCTGTCTTTCCTGCCGTCTGATTTCTTCCCGTCTGTTCGTG
	RDLLDTAAALYRDAL	ACCTGCTGGACACCGCGGCAGCACTGTACCGTGACGCGCTG
	ESPEHCSPHHTALR	GAATCTCCGGAACATTGTTCTCCGCATCACACTGCGCTGCGT
	QAVSCWREVTDFG	CAGGCGGTGAGCTGCTGGCGCGAAGTGACCGATTTTGGCGA
	DWVGNNLQ(AHA)Q	TTGGGTGGGCAACAACCTGCAGatgCAGGCGGCGCGCGATC
	AARDLVVNYVNANI	TGGTGGTGAACTATGTGAACGCGAACATTGGCCTGAAAATTC
	GLKIRQLLWFHISCL	GTCAGCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGGTC
	TFGRETVLEYLVSF	GTGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGAT
	GVWIRTPPAYRPCC	TCGTACTCCGCCGGCTTACCGTCCGTGCTGCGCACCGATCC
	APILSTLPETTVV	TGAGCACCCTGCCGGAAACCACTGTTGTGTAATAA
	(SEQ ID NO:51)	
HBc(HP) D78AHA, SS1+SS8: (D29C-R127C) + (P134C-N136C)	(AHA)DIDPYKEFGAT	ATGGATATCGACCCGTACAAAGAATTCGGCGCGACCGTTGAA
	VELLSFLPSDFFPSV	CTGCTGTCTTTCCTGCCGTCTGATTTCTTCCCGTCTGTTCGTT
	RCLLDTAAALYRDAL	GCCTGCTGGACACCGCGGCAGCACTGTACCGTGACGCGCTG
	ESPEHCSPHHTALR	GAATCTCCGGAACATTGTTCTCCGCATCACACTGCGCTGCGT
	QAVSCWREVTDFG	CAGGCGGTGAGCTGCTGGCGCGAAGTGACCGATTTTGGCGA
	DWVGNNLQ(AHA)Q	TTGGGTGGGCAACAACCTGCAGatgCAGGCGGCGCGCGATC
	AARDLVVNYVNANI	TGGTGGTGAACTATGTGAACGCGAACATTGGCCTGAAAATTC
	GLKIRQLLWFHISCL	GTCAGCTGCTGTGGTTCCACATCTCTTGCCTGACCTTCGGTC
	TFGRETVLEYLVSF	GTGAAACCGTTCTGGAATACCTGGTTTCTTTTGGTGTTTGGAT
	GVWICTPPAYRCPC	TTGTACTCCGCCGGCTTACCGTTGCCCGTGCGCACCGATCC
	APILSTLPETTVV	TGAGCACCCTGCCGGAAACCACTGTTGTGTAATAA
	(SEQ ID NO:52)	

[00125] 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 (Jewett and Swartz 2004) 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 (Goerke and Swartz 2008).

For global replacement of methionines in HBc proteins, methionine was left out of cell-free reaction mixtures, and substituted by 1 mM azidohomoalanine (AHA) (Medchem Source LLP, Federal Way, WA), a non-natural amino acid that displays an azide moiety. All reagents were obtained from Sigma–Aldrich (St. Louis, MO) unless otherwise noted.

[00126] 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 (Calhoun and Swartz 2005) and a Beckman LS3801 liquid scintillation counter (Beckman Coulter, Fullerton, CA).

[00127] The production of GM-CSF, IM9-ScFv and CpG DNA with an alkyne moiety were described by Patel *et al*. The production of flagellin with an alkyne moiety was described by Lu *et al*.

[00128] Size-exclusion chromatography (SEC). To remove unincorporated L-[U-14C] leucine, the cell-free product was immediately dialyzed in 6-8000 MWCO Specra/Pro Molecularporous Membrane Tubing (Spectrum Labs, Rancho Dominguez, CA) against Dialysis Buffer (10 mM Tris-HCl, pH 7.4, 1.5 M NaCl) with 1 mM DTT with 2 buffer exchanges. The dialyzed cell-free reaction product was loaded on an SEC (size-exclusion chromatography) column packed with Sepharose 6 FastFlow resin (GE Healthcare). The running buffer is as the dialysis buffer with 5 mM DTT. The protein concentrations of eluate fractions were determined based on radioactivity.

[00129] Sucrose gradient sedimentation. The isolated VLPs from SEC were firstly dialyzed against the Dialysis Buffer to remove DTT, and were then oxidized to form disulfide bonds by adding 10 mM H_2O_2 or 10 mM diamide and incubating at room temperature for 1 h. The oxidants were removed by dialysis against the Dialysis Buffer with 2 buffer exchanges. The oxidized VLPs were assessed by velocity sedimentation. Ten to forty percentage weight per volume continuous sucrose density gradients were prepared in Dialysis Buffer in Polyallose 16 × 102 mm Centrifuge Tubes (Beckman) with the Gradient Master Ver3.05L Gradient Maker (Biocomp Instruments, Inc., Fredericton, Canada). The VLP product (200 µL) was 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 °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.

[00130] SDS-PAGE and autoradiogram. Protein size was analyzed by SDS-PAGE and autoradiography. NuPAGE Novex precast gels and reagents were purchased from Invitrogen

(Carlsbad, CA). For reducing SDS-PAGE, samples were denatured for 10 min at 95 °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 a storage phosphor screen (Molecular Dynamics), which was subsequently scanned using a Typhoon Scanner (GE Healthcare).

[00131] 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 ([(CH3CN)4Cu]PF6 or simply Cu (I) catalyst) (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, IM9-ScFv 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.

WHAT IS CLAIMED IS:

1. An HBc polypeptide, wherein at least two amino acids of a native HBc polypeptide are substituted with cysteine residues that can form intermolecular disulfide bonds when the HBc is assembled into a virus like particle (VLP).

2. The cysteine-substituted HBc polypeptide of Claim 1, wherein the native HBc polypeptide comprises the amino acid sequence set forth in SEQ ID NO:1 or SEQ ID NO:2.

3. The cysteine-substituted HBc polypeptide of Claim 2, wherein the amino acid substitutions are selected from [D29C, R127C]; [T109C, V120C]; [Y132C, N136C]; [Y132C, A137C]; [R133C, N136C]; [R133C, A137C]; [P134C, P135C]; [P134C, N136C]; [P134C, A137C]; and [P135C, N136C].

4. The cysteine-substituted HBc polypeptide of Claim 3, wherein the amino acid substitutions are [D29C, R127C]; [P134C, N136C]; or [D29C, R127C, P134C, N136C].

5. An HBc polypeptide, wherein the amino acids of the spike region at residues 73-81 are substituted to reduce the negative charge, relative to SEQ ID NO:1.

6. The HBc polypeptide of Claim 5, wherein the amino acids of the hydrophobic pocket at residues 57-81 are substituted.

7. The HBc polypeptide of Claim 5 or Claim 6, wherein the substituted amino acids are derived from a naturally occurring viral genotype.

8. The polypeptide of Claim 7, wherein the amino acid substitutions are one of [I57V, L60S, G63R, D64E, L65V, M66T, T67D, L68F, A69G, T70D, T74N, L76M, E77Q, P79Q, S81A, S87N, T91A, V93I, F97I] or [T74N, L76M, E77Q, P79Q, S81A].

9. The HBc polypeptide of any one of Claims 5-8, further comprising at least one pair of amino acid substitutions selected from [D29C, R127C]; [T109C, V120C]; [Y132C, N136C]; [Y132C, A137C]; [R133C, N136C]; [R133C, A137C]; [P134C, P135C]; [P134C, N136C]; [P134C, A137C]; and [P135C, N136C].

10. The HBc polypeptide of any one of Claims 5-8, further comprising amino acid substitutions selected from [D29C, R127C]; [P134C, N136C]; and [D29C, R127C, P134C, N136C].

11. The HBC polypeptide of any one of Claims 1-10, further comprising the amino acid substitution M66S.

12. The HBc polypeptide of any one of Claims 1-11, wherein the polypeptide is a monomer.

13. The HBc polypeptide of any one of Claims 1-11,, wherein the polypeptide is a dimer.

14. The HBc polypeptide of any one of Claims 1-11, wherein the polypeptide is assembled into a VLP.

15. The HBc polypeptide of any one of Claims 1-11, wherein the polypeptide further comprises at least one unnatural amino acid.

16. The HBc polypeptide of Claim 15, wherein the unnatural amino acid provides a reactant group for Click chemistry reactions.

17. The HBc polypeptide of Claim 16, wherein the unnatural amino acid is azidohomoalanine, *p*-acetyl-L-phenylalanine or *p*-azido-L-phenylalanine.

18. The HBc polypeptide of Claim 16, further comprising one or more additional moieties conjugated to the unnatural amino acid.

19. An HBc polypeptide comprising the amino acid sequence of any one of SEQ ID NO:3-31, 34-52.

20. The HBc polypeptide of Claim 19, comprising the amino acid sequence of SEQ ID NO:39, SEQ ID NO:42, or SEQ ID NO:52

21. A polynucleotide encoding a polypeptide according to any one of Claims 1-20.

22. A vector comprising the polynucleotide of Claim 21.

23. A cell or cell free protein synthesis reaction mixture comprising the vector of Claim 22 or the polynucleotide of Claim 21.

24. A method of producing a polypeptide according to any of Claims 1-20, comprising introducing a polynucleotide sequence encoding the fusion protein operably linked to a promoter, into a cell or cell free protein synthesis reaction mix and incubating for a period of time sufficient to synthesize the protein.

25. The method of Claim 24, wherein the polypeptide is assembled into a VLP under reducing conditions.

26. The method of Claim 25, wherein the VLP is oxidized to generate disulfide bonds prior to further reaction to conjugate an additional polypeptide to the VLP.

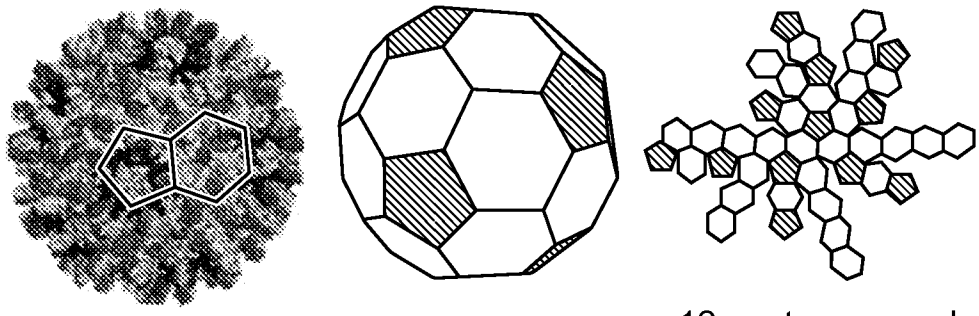

HBc VLP
T = 4 capsid
12 pentamers and
30 hexamers

FIG. 1

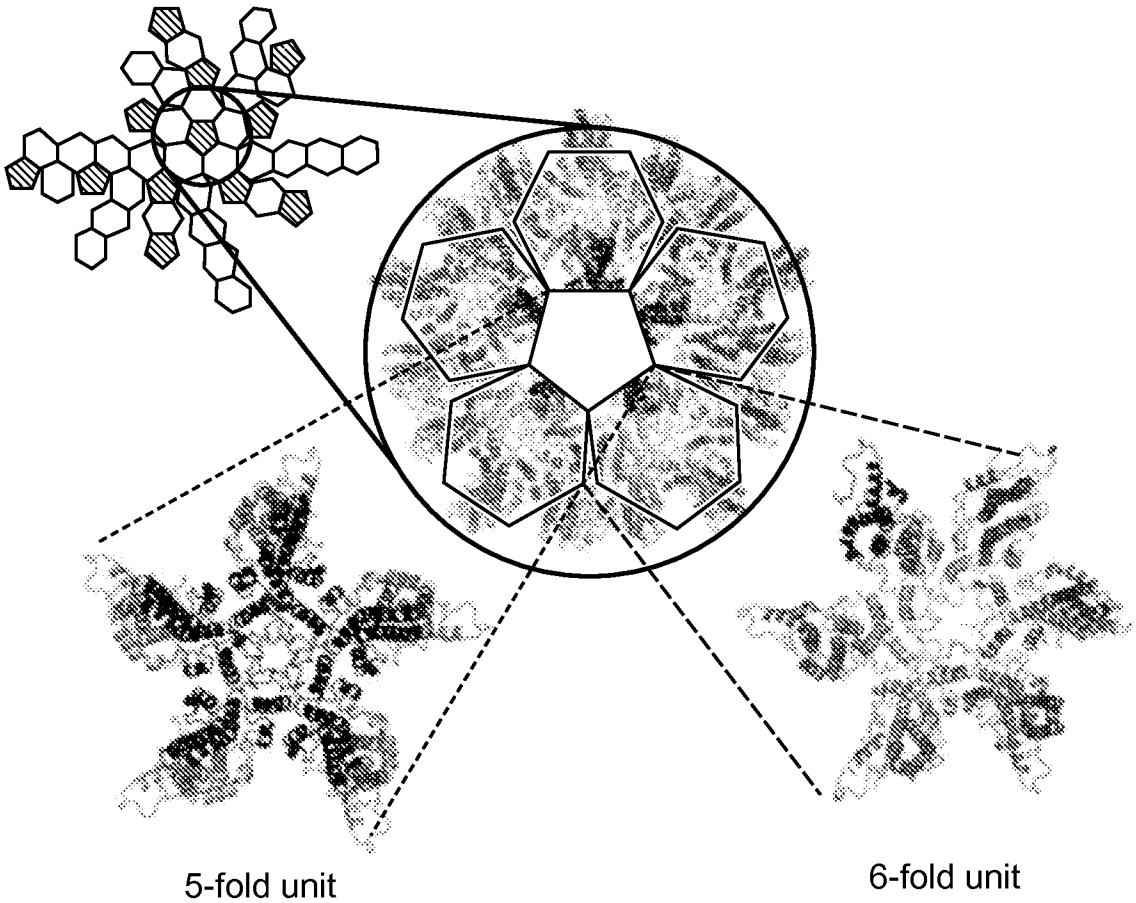
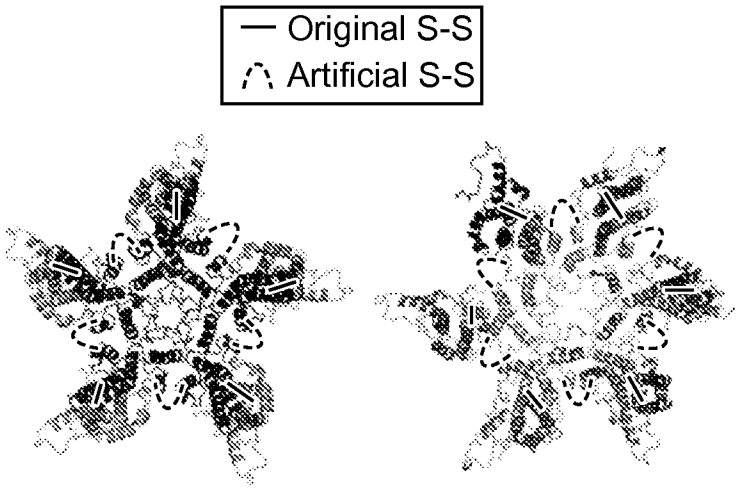
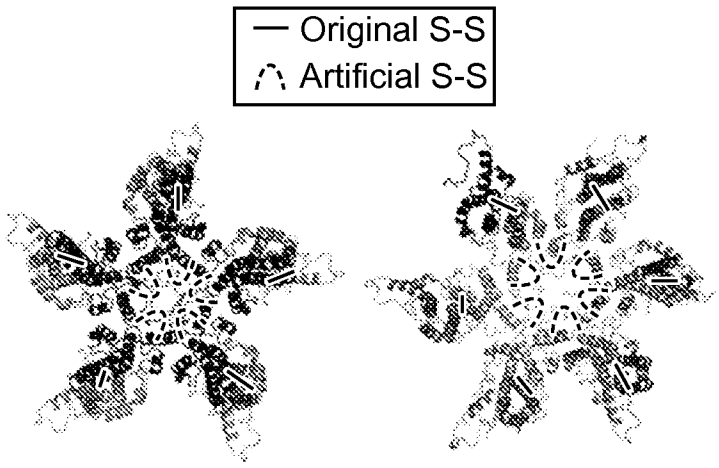

5-fold unit
6-fold unit

FIG. 2



SS1: D29-R127 (~6.6Å)
SS2: T109-V120 (~9.4Å)

FIG. 3A



SS3: Y132-N136 (~7.4Å)
SS4: Y132-A137 (~6.4Å)
SS5: R133-N136 (~6.4Å)
SS6: R133-A137 (~6.3Å)
SS7: P134-P135 (~6.6Å)
SS8: P134-N136 (~4.5Å)
SS9: P134-A137 (~5.5Å)
SS10: P135-N136 (~6.6Å)

FIG. 3B

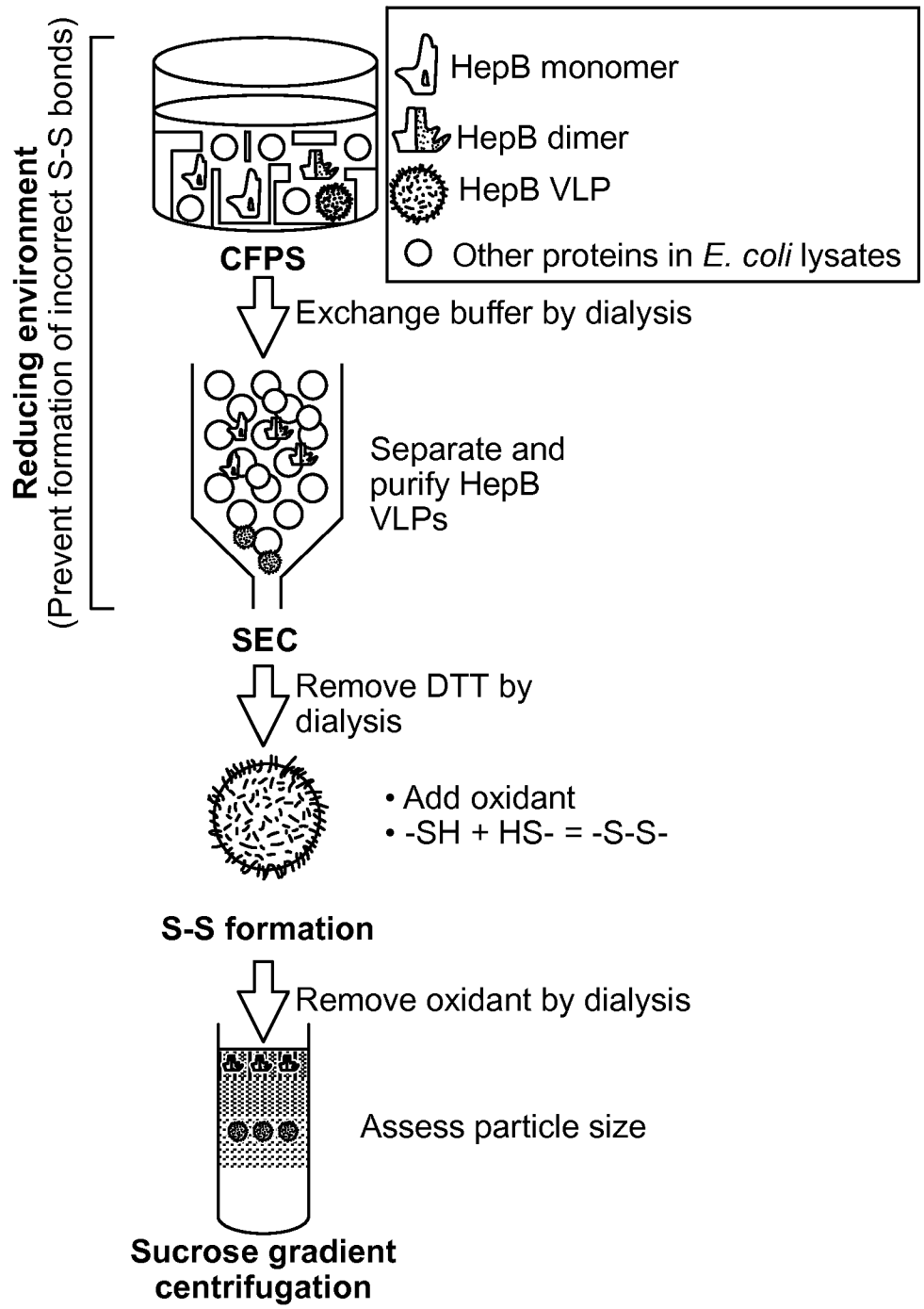

Reducing environment
(Prevent formation of incorrect S-S bonds)
HepB monomer
HepB dimer
HepB VLP
Other proteins in E. coli lysates
CFPS
Exchange buffer by dialysis
Separate and purify HepB VLPs
SEC
Remove DTT by dialysis
• Add oxidant
• -SH + HS- = -S-S-
S-S formation
Remove oxidant by dialysis
Assess particle size
Sucrose gradient centrifugation

FIG. 4

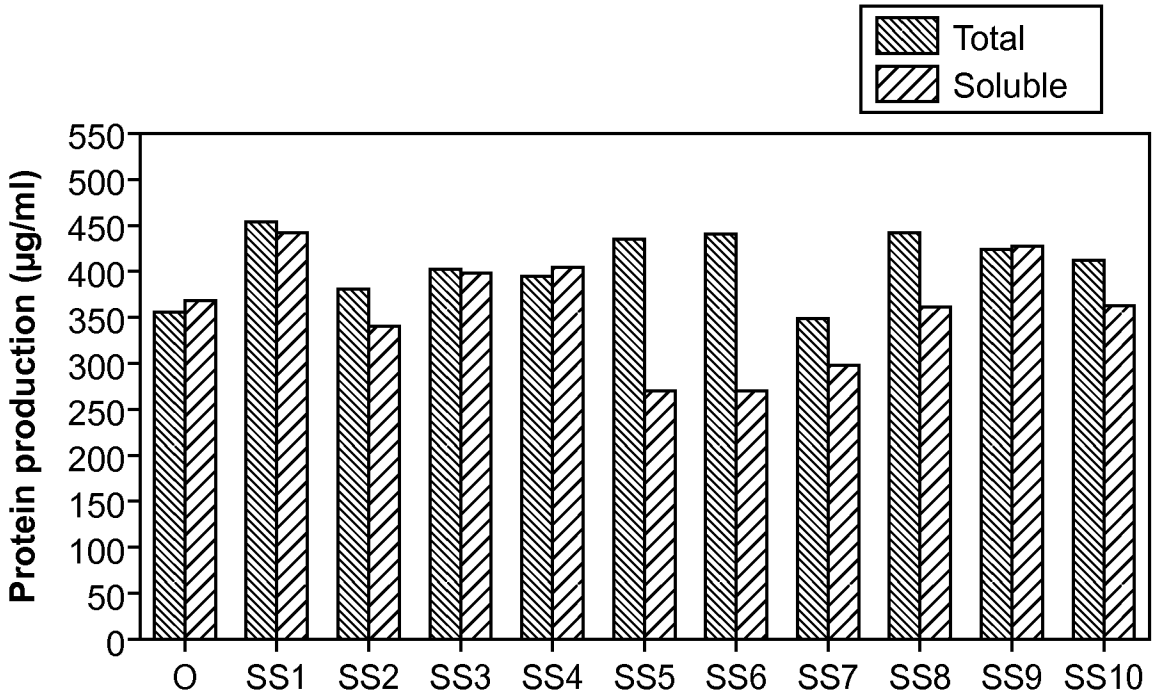

Total
Soluble
Protein production (µg/ml)
550
500
450
400
350
300
250
200
150
100
50
0
O
SS1
SS2
SS3
SS4
SS5
SS6
SS7
SS8
SS9
SS10

FIG. 5A

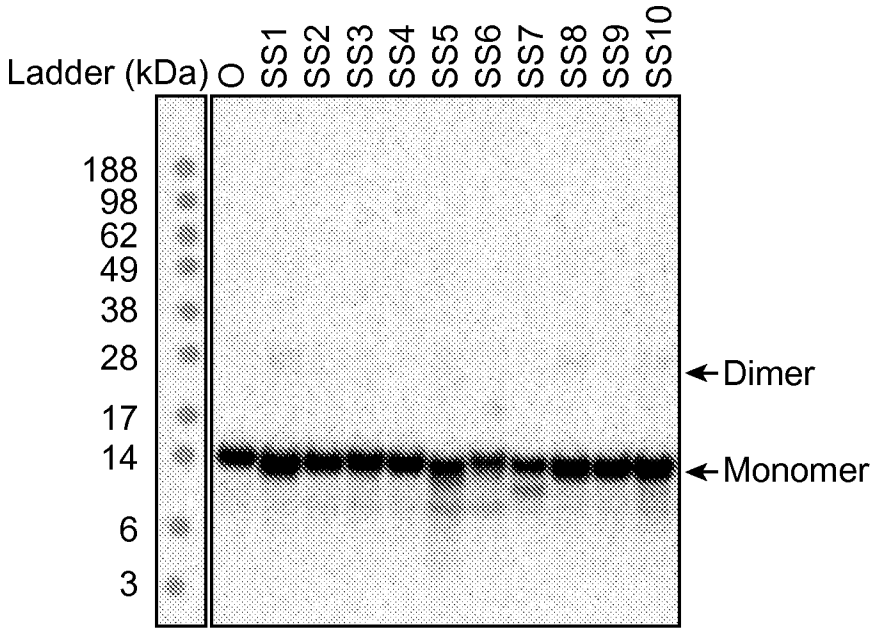

Ladder (kDa)
O
SS1
SS2
SS3
SS4
SS5
SS6
SS7
SS8
SS9
SS10
188
98
62
49
38
28
17
14
6
3
Dimer
Monomer

FIG. 5B

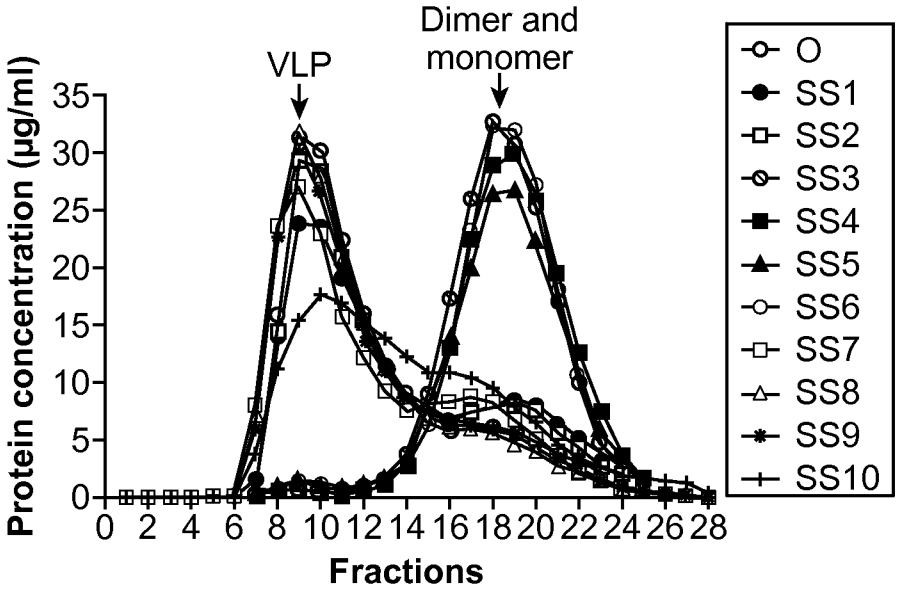

Dimer and monomer
VLP
Protein concentration (µg/ml)
35
30
25
20
15
10
5
0
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28
Fractions
O
SS1
SS2
SS3
SS4
SS5
SS6
SS7
SS8
SS9
SS10

FIG. 6

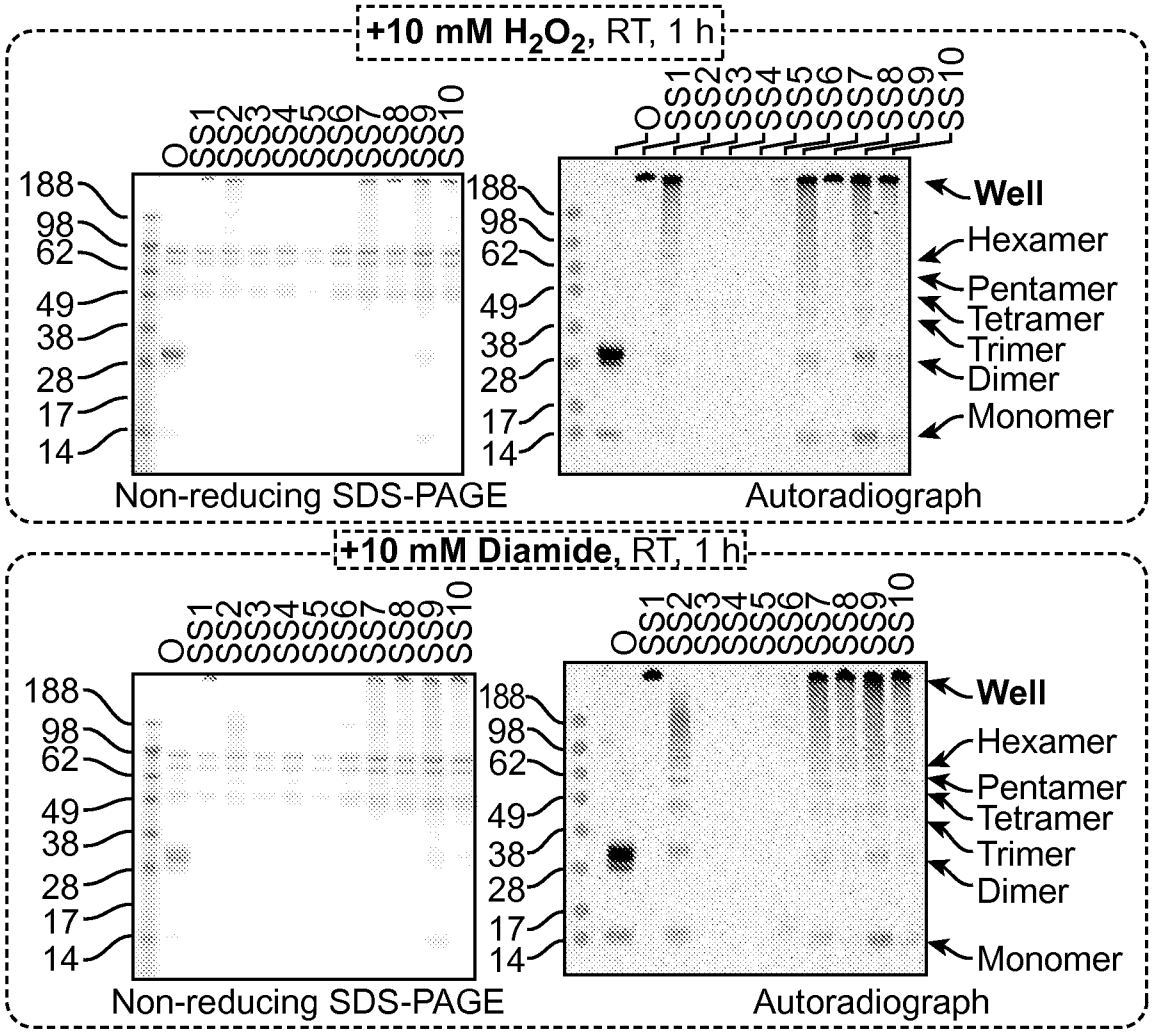

+10 mM H2O2, RT, 1 h
O SS1 SS2 SS3 SS4 SS5 SS6 SS7 SS8 SS9 SS10
188
98
62
49
38
28
17
14
Non-reducing SDS-PAGE
Autoradiograph
Well
Hexamer
Pentamer
Tetramer
Trimer
Dimer
Monomer
+10 mM Diamide, RT, 1 h
O SS1 SS2 SS3 SS4 SS5 SS6 SS7 SS8 SS9 SS10
188
98
62
49
38
28
17
14
Non-reducing SDS-PAGE
Autoradiograph
Well
Hexamer
Pentamer
Tetramer
Trimer
Dimer
Monomer

FIG. 7

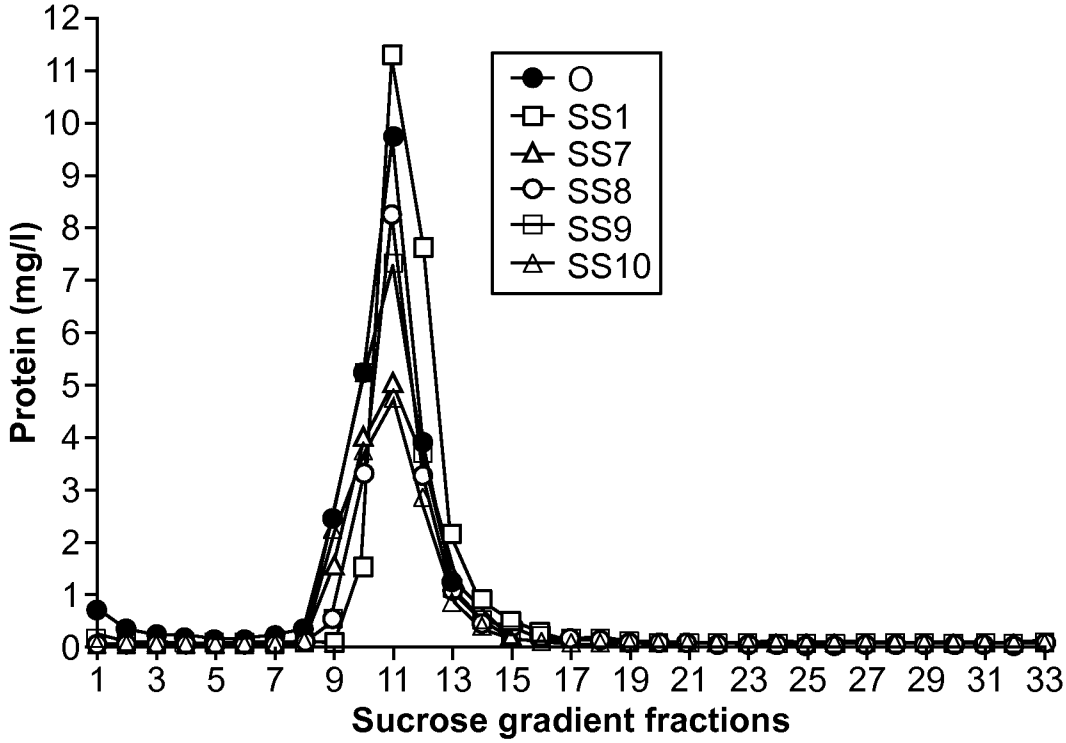

Protein (mg/l)
Sucrose gradient fractions
O
SS1
SS7
SS8
SS9
SS10

FIG. 8

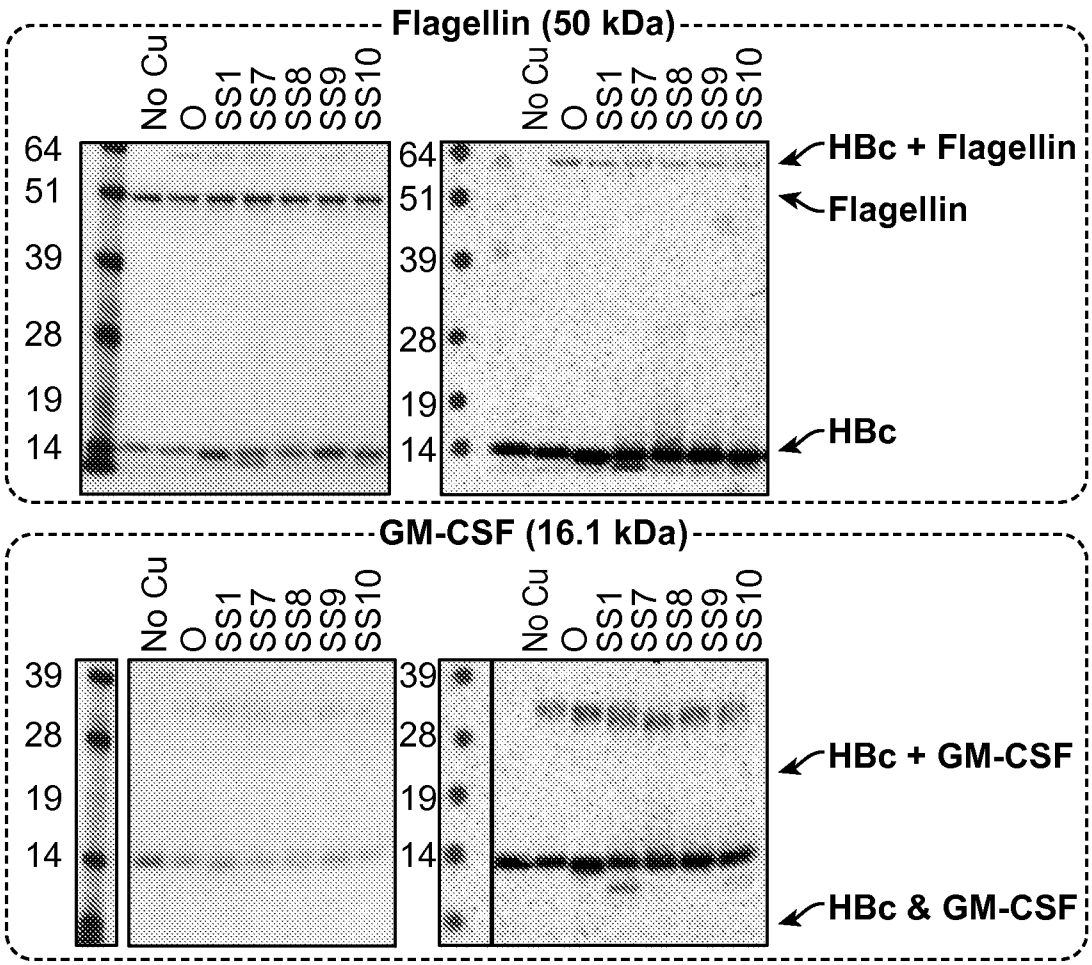

Flagellin (50 kDa)
No Cu
O
SS1
SS7
SS8
SS9
SS10
HBc + Flagellin
Flagellin
HBc
GM-CSF (16.1 kDa)
HBc + GM-CSF
HBc & GM-CSF

FIG. 9

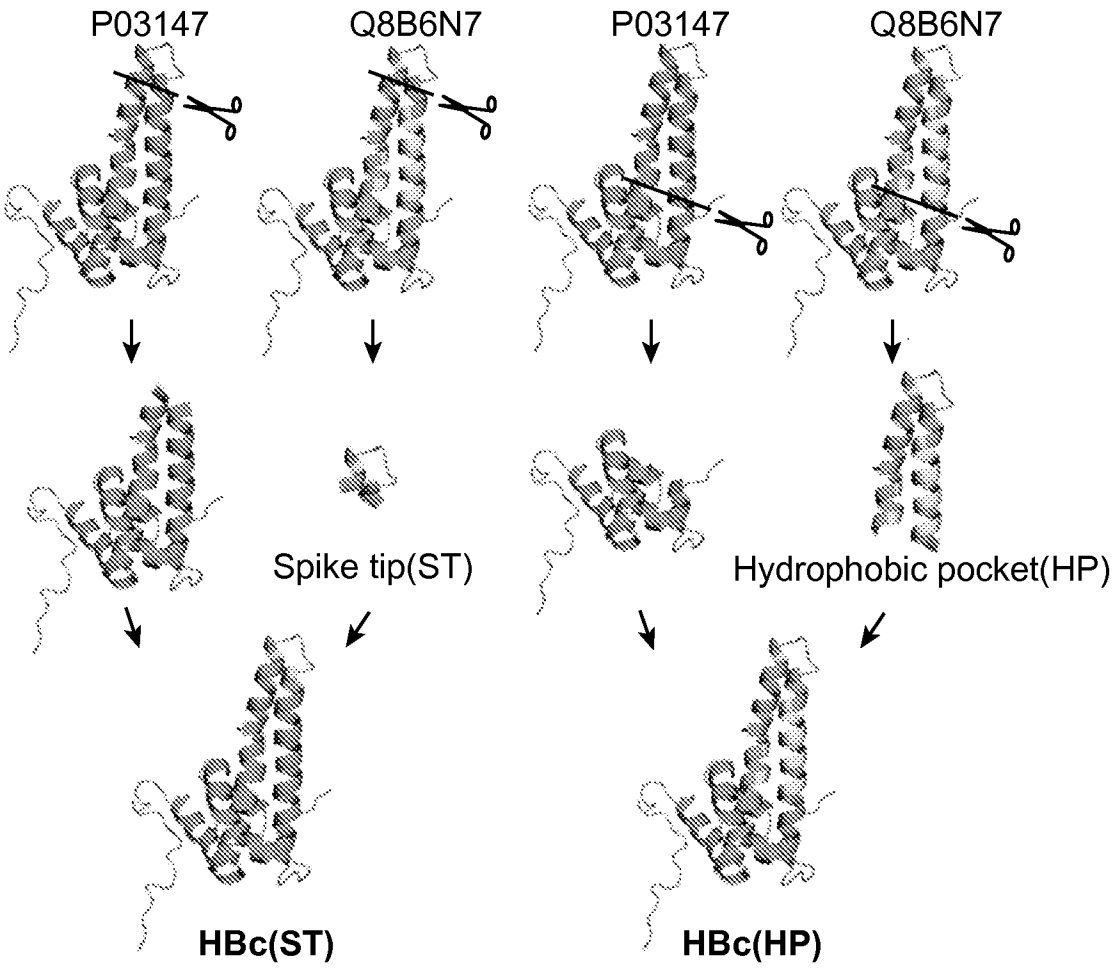

P03147
Q8B6N7
P03147
Q8B6N7
Spike tip(ST)
Hydrophobic pocket(HP)
HBc(ST)
HBc(HP)

FIG. 10

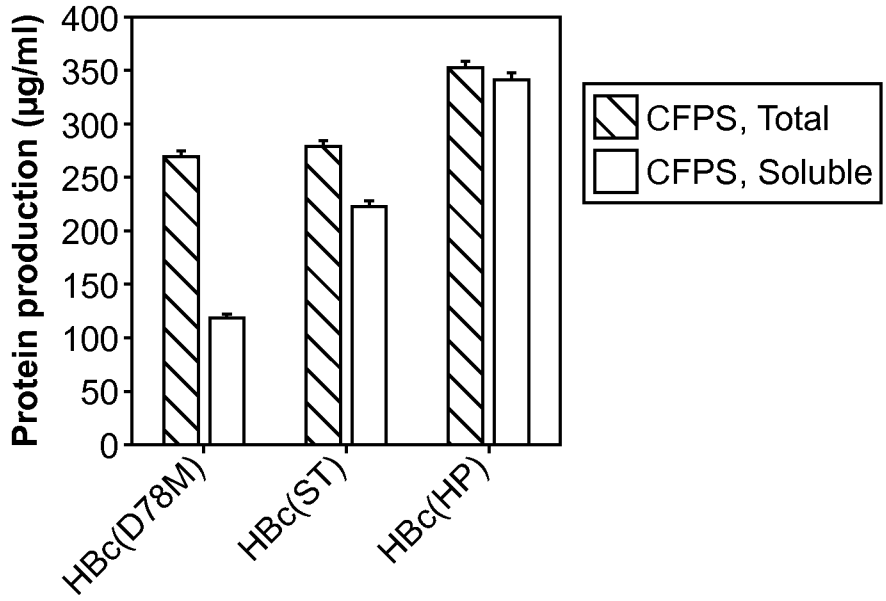


FIG. 11

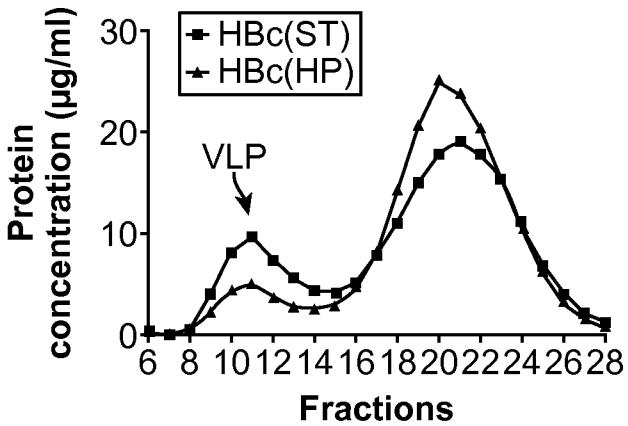


FIG. 12A

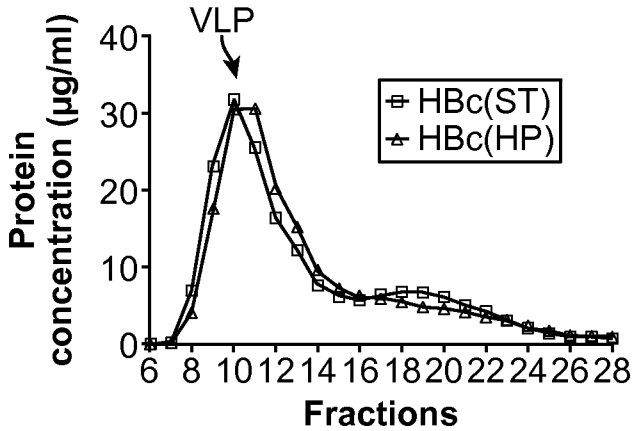
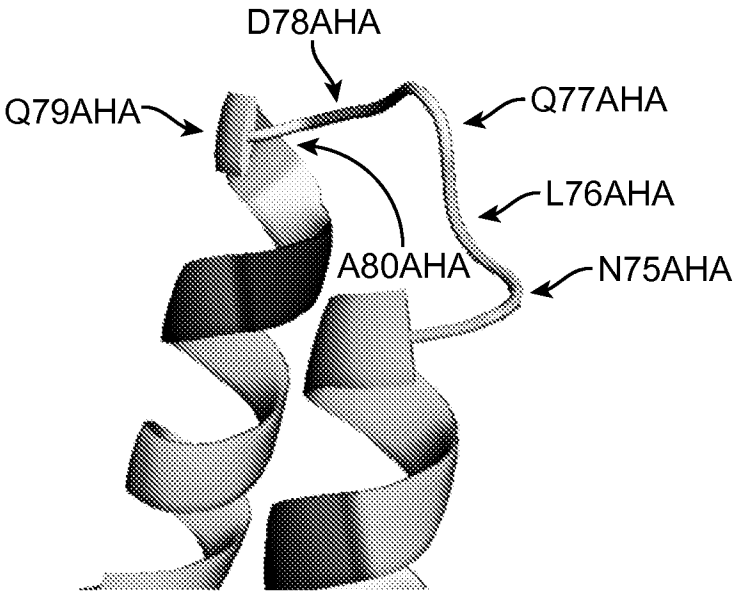


FIG. 12B



Mutant	Negative charges on the tip
N75AHA	1
L76AHA	1
Q77AHA	1
D78AHA	**0**
Q79AHA	1
A80AHA	1

FIG. 13

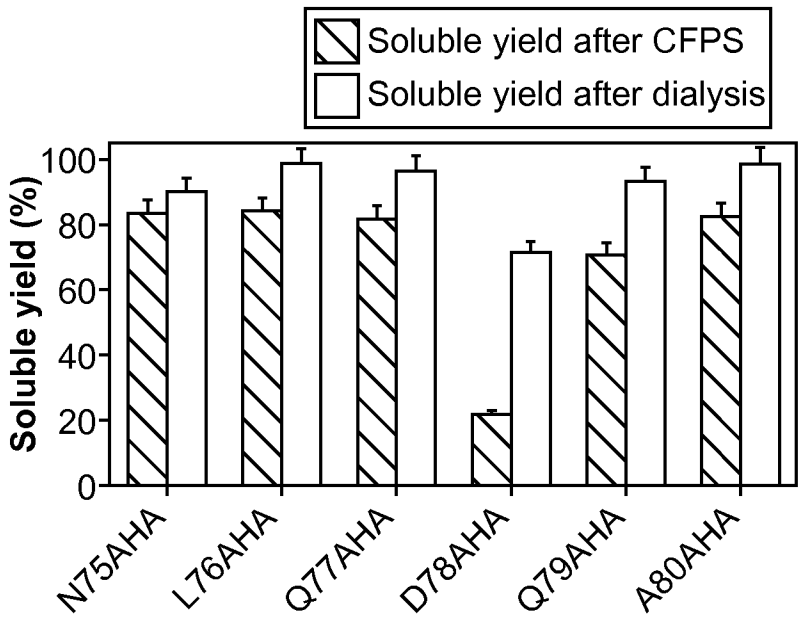


FIG. 14

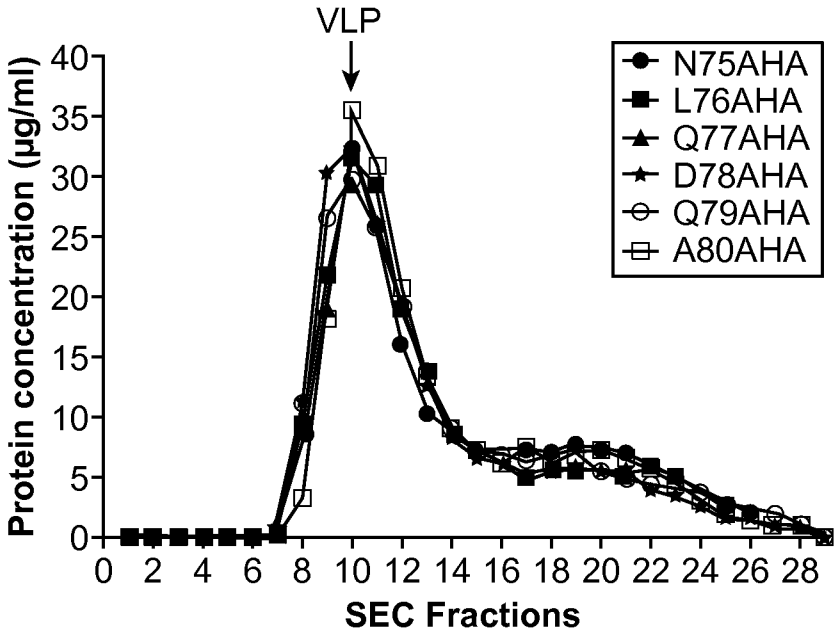


FIG. 15

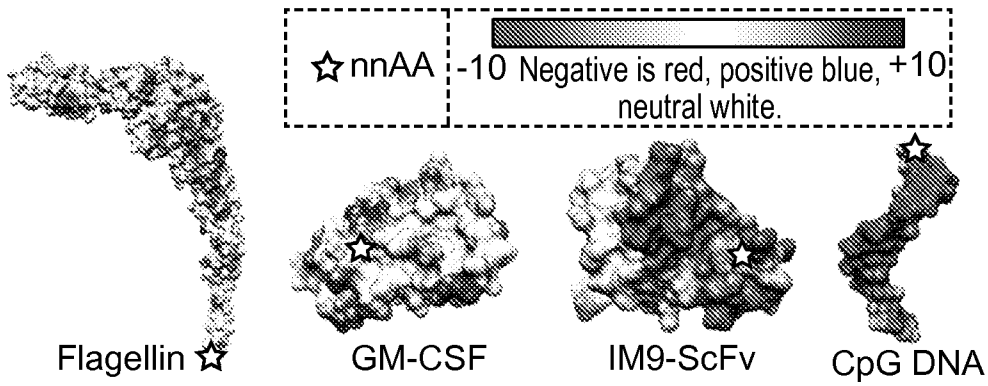


FIG. 16A

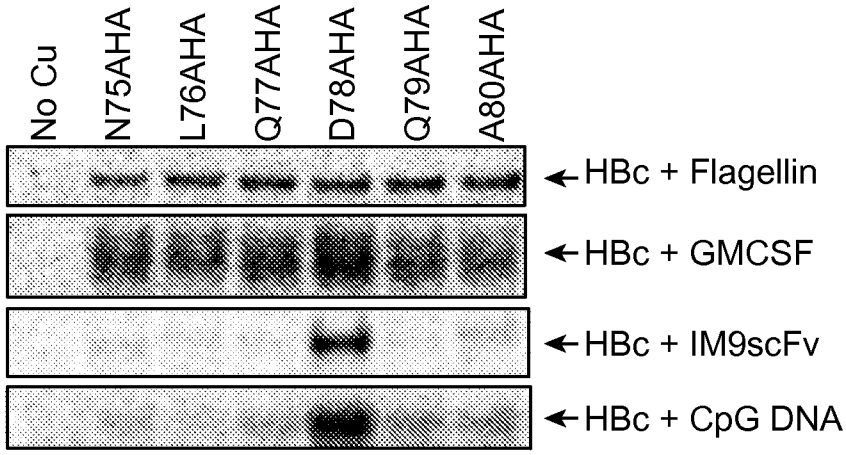


FIG. 16B

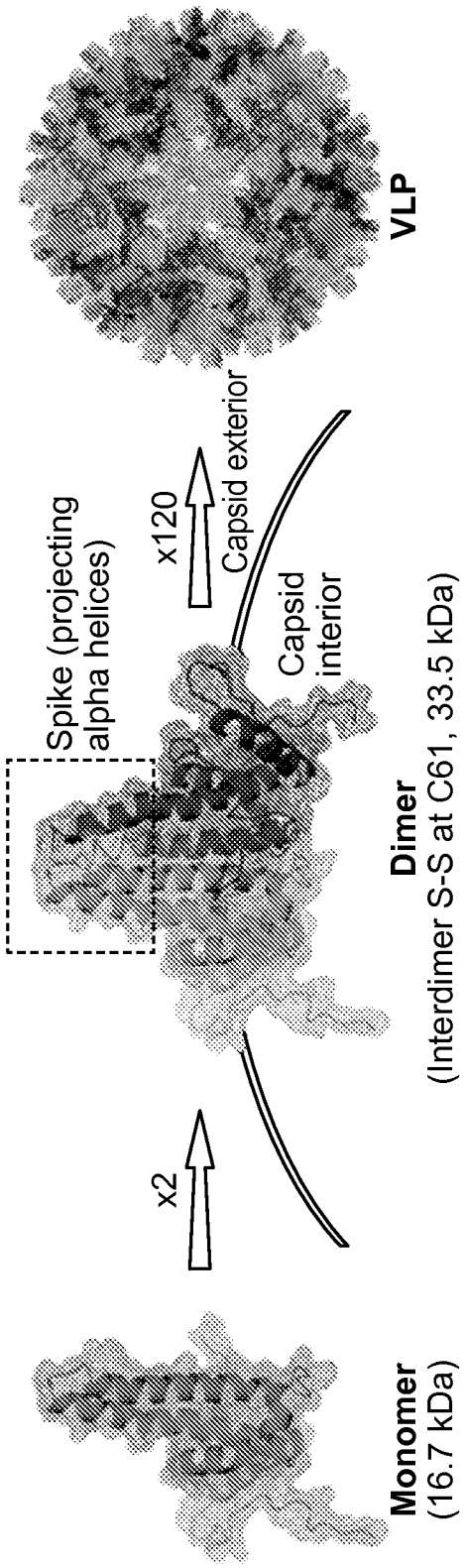

VLP
x120
Capsid exterior
Capsid interior
Spike (projecting alpha helices)
Dimer
(Interdimer S-S at C61, 33.5 kDa)
x2
Monomer
(16.7 kDa)

FIG. 17

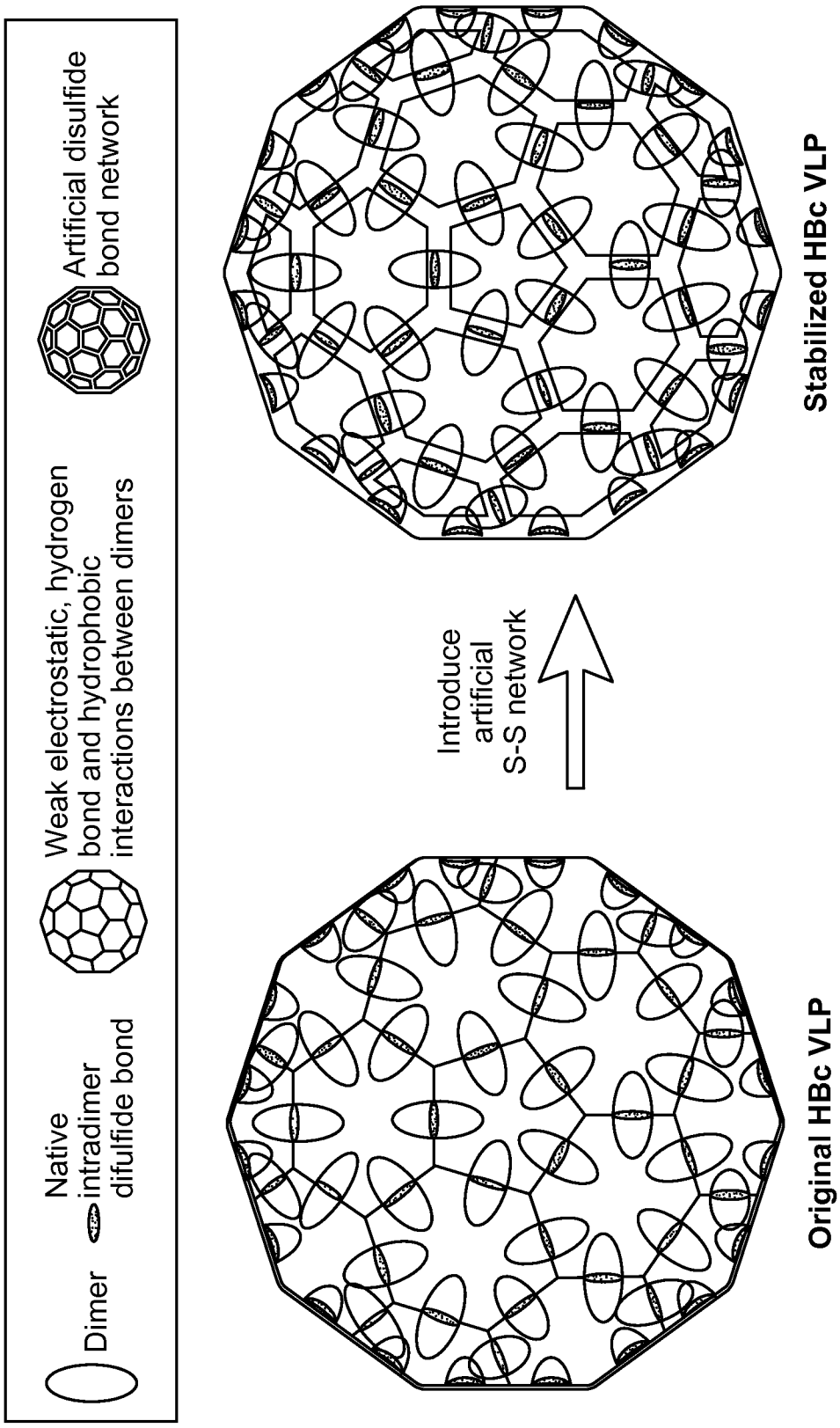

Dimer
Native intradimer difulfide bond
Weak electrostatic, hydrogen bond and hydrophobic interactions between dimers
Artificial disulfide bond network
Introduce artificial S-S network
Original HBc VLP
Stabilized HBc VLP

FIG. 18

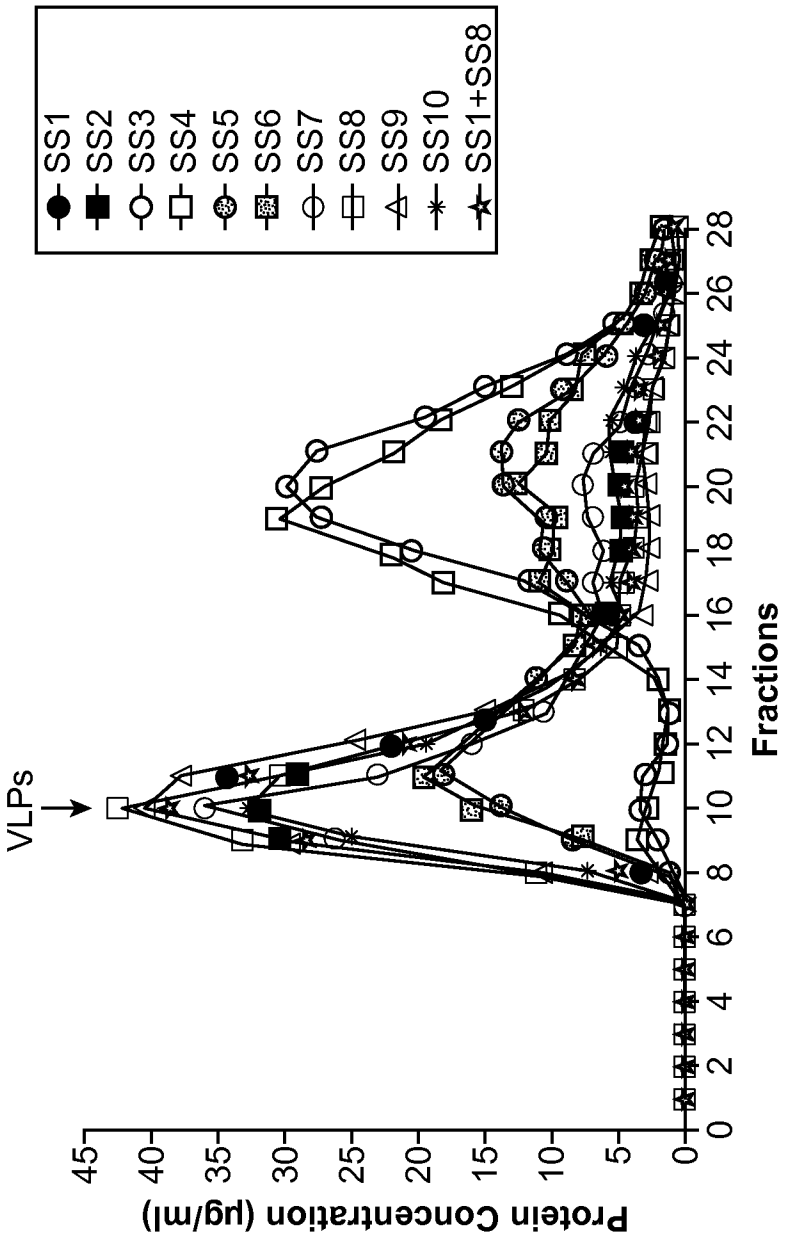

SS1
SS2
SS3
SS4
SS5
SS6
SS7
SS8
SS9
SS10
SS1+SS8
VLPs
Protein Concentration (µg/ml)
Fractions
0
5
10
15
20
25
30
35
40
45
0
2
4
6
8
10
12
14
16
18
20
22
24
26
28

FIG. 19

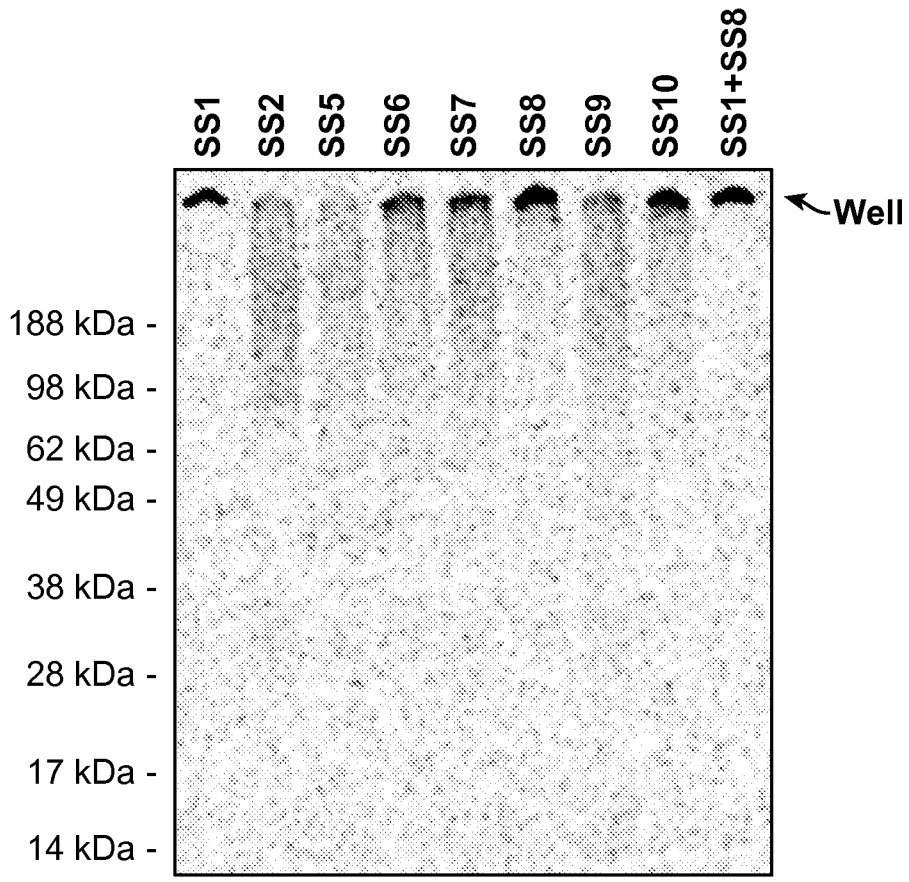

SS1
SS2
SS5
SS6
SS7
SS8
SS9
SS10
SS1+SS8
Well
188 kDa -
98 kDa -
62 kDa -
49 kDa -
38 kDa -
28 kDa -
17 kDa -
14 kDa -

FIG. 20