(12) STANDARD PATENT (11) Application No. **AU 2021392894 B2**
(19) AUSTRALIAN PATENT OFFICE

(54) Title
DONOR STRAND COMPLEMENTED FIMH

(51) International Patent Classification(s)
C07K 14/245 (2006.01) ***A61K 39/108*** (2006.01)
A61K 35/74 (2015.01) ***A61P 31/04*** (2006.01)
A61K 39/00 (2006.01)

(21) Application No: **2021392894** (22) Date of Filing: **2021.11.30**

(87) WIPO No: **WO22/117595**

(30) Priority Data

(31) Number	(32) Date	(33) Country
20211337.9	**2020.12.02**	**EP**
20214942.3	**2020.12.17**	**EP**

(43) Publication Date: **2022.06.09**
(44) Accepted Journal Date: **2024.09.19**

(71) Applicant(s)
GLAXOSMITHKLINE BIOLOGICALS SA

(72) Inventor(s)
ADAMO, Roberto;COZZI, Roberta;FANTONI, Adele;PHOGAT, Sanjay;ROSINI, Roberto;SCARSELLI, Maria;WAHOME, Newton

(74) Agent / Attorney
AJ PARK, PO Box 949, Wellington, 6140, NZ

(56) Related Art
WO 01/04148 A2
CHOUDHURY D., et al., 1999, 'X-ray structure of the FimC-FimH chaperone-adhesin complex from uropathogenic Escherichia coli', Science, 285(5430), pages 1061-1066
US 2005/0232938 A1

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
09 June 2022 (09.06.2022)

WIPO | PCT

(10) International Publication Number
WO 2022/117595 A3

(51) International Patent Classification:
C07K 14/245 (2006.01) *A61K 35/74* (2015.01)
A61K 39/00 (2006.01) *A61P 31/04* (2006.01)
A61K 39/108 (2006.01)

(21) International Application Number:
PCT/EP2021/083659

(22) International Filing Date:
30 November 2021 (30.11.2021)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
20211337.9 02 December 2020 (02.12.2020) EP
20214942.3 17 December 2020 (17.12.2020) EP

(71) Applicant: **GLAXOSMITHKLINE BIOLOGICALS SA** [BE/BE]; Rue de l'Institut 89, B-1330 Rixensart (BE).

(72) Inventors: **ADAMO, Roberto**; GSK Vaccines srl, Via Fiorentina 1, 53100 Siena (IT). **COZZI, Roberta**; GSK Vaccines srl, Via Fiorentina 1, 53100 Siena (IT). **FANTONI, Adele**; GSK Vaccines srl, Via Fiorentina 1, 53100 Siena (IT). **PHOGAT, Sanjay**; GSK Vaccines srl, Via Fiorentina 1, 53100 Siena (IT). **ROSINI, Roberto**; GSK Vaccines srl, Via Fiorentina 1, 53100 Siena (IT). **SCARSELLI, Maria**; GSK Vaccines srl, Via Fiorentina 1, 53100 Siena (IT). **WAHOME, Newton**; GlaxoSmithKline, 14200 Shady Grove Road, Rockville, Maryland 20850 (US).

(74) Agent: **HITCHCOCK, Lucy**; GlaxoSmithKline, Global Patents (CN925.1), 980 Great West Road, Brentford Middlesex TW8 9GS (GB).

(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, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, 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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, 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, ST, 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).

Published:
— *with international search report (Art. 21(3))*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*
— *with sequence listing part of description (Rule 5.2(a))*
— *in black and white; the international application as filed contained color or greyscale and is available for download from PATENTSCOPE*

(88) Date of publication of the international search report:
07 July 2022 (07.07.2022)

(54) Title: DONOR STRAND COMPLEMENTED FIMH

(57) Abstract: The present invention is directed to novel, modified FimH polypeptides, nucleic acids encoding them, and the use of the polypeptides and nucleic acids in the treatment and/or prevention of disease, in particular, urinary tract infection (UTI).

WO 2022/117595 A3

29 Aug 2024
2021392894

DONOR STRAND COMPLEMENTED FIMH

SEQUENCE LISTING

The instant application contains an electronically submitted Sequence Listing in ASCII text file format (VB67013 FF Seq List_ST25.txt; Size: 356.838 bytes; and Date of Creation: 27th October 2021) which is hereby incorporated by reference in its entirety.

Technical field of the invention

The present invention is directed to novel, modified FimH polypeptides, nucleic acids encoding them, and the use of the polypeptides and nucleic acids in the treatment and/or prevention of disease, in particular, urinary tract infection (UTI).

Background

Uropathogenic Escherichia coli (UPEC) account for approximately 85% of all urinary tract infections (UTIs) (A. R. Ronald, Urinary tract infection in adults: Research priorities and strategies. Int. J. Antimicrob. Agents 17, 343–348; 2001). The tip-localized adhesin FimH of the type 1 pili allows UPEC to colonize the bladder epithelium during UTIs by binding to mannosylated receptors on the urothelial surface (M. A. Mulvey, Induction and evasion of host defences by type 1-piliated uropathogenic Escherichia coli. Science 282, 1494–1497; 1998).

FimH is phase variable and environmental signals influence its expression, allowing bacteria to attach and avoid being eliminated by micturition (Infect. Immun. 1998, 66, 3303). Anti-FimH IgGs are known to inhibit bacterial adhesion to the bladder in mice and monkeys and the protective effect was associated with the presence of anti-FimH IgGs in the urine (Langermann S, et al. Science. 1997 Apr 25;276(5312):607-11; Langermann S, et al. J Infect Dis. 2000 Feb;181(2):774-8). Transudation of serum functional IgGs in the urogenital tract seems responsible for inhibiting bacterial adhesion.

FimH protein is composed of an N-terminal lectin domain ($FimH_L$), which binds mannose via a pocket formed by three loops, a 5–amino acids linker and the C-terminal pilin domain ($FimH_P$) that attaches FimH to the pilus.

Crystal structures of FimH in different stages of pilus assembly showed that $FimH_P$ is constituted by an incomplete immunoglobulin (Ig)–like fold which is stabilized via a donor strand complementation interaction with the chaperone FimC in the periplasm, and with FimG when the pilus assembles. $FimH_P$ adopts a single conformation, but $FimH_L$ can assume at least two conformational states with different affinities for mannose

2021392894 29 Aug 2024

– the high-affinity conformation, the relaxed (R) state, and the low-affinity conformation, the tense (T) state (D. Choudhury, X-ray structure of the FimC-FimH chaperone-adhesin complex from uropathogenic Escherichia coli. Science 285, 1061–1066 (1999); C.-S. Hung, Structural basis of tropism of Escherichia coli to the bladder during urinary tract infection. Mol. Microbiol. 44, 903–915 (2002); I. Le Trong, Structural basis for mechanical force regulation of the adhesin FimH via finger trap-like β sheet twisting. Cell 141, 645–655 (2010); G. Phan, Crystal structure of the FimD usher bound to its cognate FimC–FimH substrate. Nature 474, 49–53 (2011); S. Geibel, Structural and energetic basis of folded-protein transport by the FimD usher. Nature 496, 243–246 (2013)).

When FimH binds to FimC, FimH adopts an elongated conformation in which $FimH_L$ and $FimH_P$ do not interact with each other, and $FimH_L$ is in a high-affinity mannose-binding state. When FimH is bound to FimG, FimH adopts a compact conformation, wherein $FimH_L$ and $FimH_P$ interact closely and $FimH_L$ adopts a low-affinity mannose-binding state. $FimH_P$ can allosterically decrease the ability of $FimH_L$ to bind mannose through interactions with the base of $FimH_L$; while mannose binding to $FimH_L$ induces $FimH_L$ conformations that do not interact with $FimH_P$.

Previously, it has been reported the monoclonal antibodies against $FimH_L$ in the low affinity conformation lead to a better inhibition of adhesion to the bladder compared to monoclonal antibodies against the mannose post-binding form (Tchesnokova *et* al., 2011, 'Type 1 Fimbrial Adhesin FimH Elicits an Immune Response That Enhances Cell Adhesion of Escherichia coli' *Infect. Immun.* 79(10): 3895-3904).

FimH with its non-complemented pilin domain is unstable and tends to aggregate. Of note, FimH has been typically used as antigen in complex with the periplasmic protein FimC. The FimC component did not directly contribute to reduction of bacterial colonization in mice, but rather in FimH stabilization, protecting it from degradation (*Science* 1997, 276, 607; *FEMS Microbiol. Lett*. 2000, 188, 147). To produce a stable FimH protein FimG donor strand peptide (FimG residues 1−14) has been added in vitro to displace the pilus assembly chaperone FimC from FimH. (Sauer MM, et al. Nat Commun. 2016 Mar 7;7:10738.). A low affinity conformation of FimHL has also been obtained inserting a disulphide bridge, locking the mannose pocket (Kisiela DI, et al. Proc Natl Acad Sci U S A. 2013 Nov 19;110(47):19089-94).

Use of FimHC complexes include significant production burdens – i.e., production of two polypeptides, which must then be complexed together, presenting an unwelcome complication and a significant storage problem since, for the antigens to be effective, stability of the complexes must be maintained during storage. The immunogenicity of $FimH_L$ with disulphide bridge is variable due to low molecular weight of the portion, and full FimH with a disulphide bridge in the $FimH_L$ domain proved difficult to express.

Accordingly, there remains an outstanding need for ExPEC antigens that are both immunologically effective and viable for production at scale. It is an object of the present invention to go some way to meeting this need; and/or to at least provide the public with a useful choice.

29 Aug 2024

2021392894

In this specification where reference has been made to patent specifications, other external documents, or other sources of information, this is generally for the purpose of providing a context for discussing the features of the invention. Unless specifically stated otherwise, reference to such external documents is not to be construed as an admission that such documents, or such sources of information, in any jurisdiction, are prior art, or form part of the common general knowledge in the art.

Summary of the invention

The invention provides a polypeptide comprising or consisting of an amino acid sequence X-(a)-L-(b)-Y, wherein "(a)" is a FimH polypeptide; "L" is a first linker; "(b)" is a donor-strand complementing amino acid sequence, "X" is an optional N-terminal amino acid sequence; "Y is an optional C-terminal amino acid sequence, wherein "Y" is not derived from FimC or FimH or a fragment thereof,

wherein "(b)" comprises or consists of an amino acid sequence according to SEQ ID NO: 5, and

wherein the first linker or "L" comprises or consists of PGDGN **[SEQ ID NO: 7]**.

The invention also provides a nanoparticle comprising a polypeptide monomer comprising or consisting of an amino acid sequence that has:

(i) at least 80% sequence identity, for example at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to the amino acid sequence SEQ ID NO: 16 and has one or more mutations from the group consisting of: glycine (G) at the position that aligns to residue 34 of SEQ ID NO: 16 (T34G mutation), aspartic acid (D) at the position that aligns to residue 70 of SEQ ID NO: 16 (N70D mutation), isoleucine (I) at the position that aligns to residue 72 of SEQ ID NO: 16 (V72I mutation) and alanine (A) at the position that aligns to residue 124 of SEQ ID NO: 16 (S124A mutation),

(ii) an amino acid sequence that has at least 80% sequence identity, for example at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to the amino acid sequence SEQ ID NO: 16 and has glycine (G) at the position that aligns to residue 34 of SEQ ID NO: 16 (T34G mutation), aspartic acid (D) at the position that aligns to residue 70 of SEQ ID NO: 16 (N70D mutation), isoleucine (I) at the position that aligns to residue 72 of SEQ ID NO: 16 (V72I mutation) and alanine (A) at the position that aligns to residue 124 of SEQ ID NO: 16 (S124A mutation), optionally wherein the polypeptide monomer comprises an amino acid sequence that has at least 80% sequence identity,

(iii) has at least 80% sequence identity, for example at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to the amino acid sequence SEQ ID NO: 16 and has

glycine (G) at the position that aligns to residue 34 of SEQ ID NO: 16 (T34G mutation), isoleucine (I) at the position that aligns to residue 72 of SEQ ID NO: 16 (V72I mutation) and alanine (A) at the position that aligns to residue 124 of SEQ ID NO: 16 (S124A mutation), optionally wherein the polypeptide monomer comprises an amino acid sequence that has at least 80% sequence identity,

(iv) has at least 80% sequence identity, for example at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to the amino acid sequence SEQ ID NO: 16 and has glycine (G) at the position that aligns to residue 34 of SEQ ID NO: 16 (T34G mutation), and alanine (A) at the position that aligns to residue 124 of SEQ ID NO: 16 (S124A mutation), optionally wherein the polypeptide monomer comprises an amino acid sequence that has at least 80% sequence identity,

(v) has at least 80% sequence identity, for example at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to the amino acid sequence SEQ ID NO: 16 and has glycine (G) at the position that aligns to residue 34 of SEQ ID NO: 16 (T34G mutation), optionally wherein the polypeptide monomer comprises an amino acid sequence that has at least 80% sequence identity, and

wherein the exterior surface structure or interior surface structure of the nanoparticle carries one or more antigen comprising or consisting of the polypeptide of the invention.

The invention also provides a nucleic acid encoding a polypeptide according to the invention or a polypeptide of a nanoparticle according to the invention.

The invention also provides a vector comprising the nucleic acid of the invention.

The invention also provides a cell comprising a nucleic acid of the invention or a vector of the invention.

The invention also provides a pharmaceutical composition comprising the polypeptide or nanoparticle of the invention, the nucleic acid of the invention, or a vector of the invention.

The invention also provides a vaccine comprising the polypeptide or nanoparticle of the invention, a nucleic acid of the invention or a vector of the invention.

The invention also provides a polypeptide or nanoparticle of the invention, a nucleic acid of the invention, a vector of the invention, or a vaccine of the invention for use in medicine.

The invention also provides a polypeptide or nanoparticle of the invention, a nucleic acid of the invention, a vector of the invention, or a vaccine of the invention for use in raising an immune response in a mammal, for example, for treating and/or preventing one or more disease.

2021392894 29 Aug 2024

The invention also provides a use of a polypeptide or nanoparticle of the invention, a nucleic acid of the invention, a vector of the invention, or a vaccine of the invention for the manufacture of a medicament for raising an immune response in a mammal, for example, for treating and/or preventing one or more disease.

The invention also provides a method of raising an immune response in a mammal, the method comprising or consisting of administering the mammal with an effective amount of a polypeptide or nanoparticle of the invention, a nucleic acid of the invention, a vector of the invention, or a vaccine of the invention.

In the description in this specification reference may be made to subject matter which is not within the scope of the claims of the current application. That subject matter should be readily identifiable by a person skilled in the art and may assist in putting into practice the invention as defined in the claims of this application.

The term "comprising" as used in this specification and claims means "consisting at least in part of". When interpreting statements in this specification and claims which include the term "comprising", other features besides the features prefaced by this term in each statement can also be present. Related terms such as "comprise" and "comprised" are to be interpreted in similar manner.

Description

Importantly, FimC stabilizes FimH in its extended post-binding-like form (*Nat. Commun.* 2016, 7, 10738). The present inventors have surprisingly found that, by a structure-guided design, it is possible to stabilize the pre-binding form of FimH in absence of FimC and/or improve the capacity of generated anti -FimH antibodies to inhibit bacterial adhesion to uroepithelial cells.

Accordingly, a first aspect of the invention provides a polypeptide having an amino acid sequence comprising or consisting of:

(a) FimH; or a variant, fragment and/or fusion of FimH, and

(b) a donor-strand complementing amino acid sequence,

wherein (b) is downstream of (a).

By "**downstream**" we mean or include an amino acid sequence that, within the primary amino acid sequence of a polypeptide, is located closer to the C-terminus of the polypeptide respective to a reference sequence.

Alternatively or additionally, the polypeptide of the invention comprises or consists of an amino acid sequence X-(a)-L-(b)-Y, wherein "(a)" is a FimH polypeptide; or a variant, fragment and/or fusion of FimH; "L" is an optional first linker; "(b)" is a donor-strand complementing amino acid sequence, "X" is an optional N-terminal amino acid sequence; "Y is an optional C-terminal amino acid sequence, wherein "Y" is not derived from FimC or FimH or a fragment thereof.

2021392894 29 Aug 2024

By '**a donor-strand complementing amino acid sequence**' we mean an amino acid sequence capable of maintaining FimH in (a) the high-affinity conformation, relaxed (R) state, or (b) the low-affinity conformation, the tense (T) state. In one preferred embodiment, the donor strand complementing amino acid sequence is capable of maintaining FimH in the low affinity conformation, i.e. the tense (T) state.

By '**the high-affinity conformation, relaxed (R) state**' we mean or include with mannose binding affinity closer to that of FimH in the high-affinity conformation than the low-affinity conformation (in particular, FimH from which the polypeptide of the invention was derived or principally derived, especially where complexed with FimC) e.g., at least 51%, 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or 100% of the mannose binding affinity of FimH in the high-affinity conformation, for example, K_d < 1.2 μM as disclosed in Kisiela DI, et al. Proc Natl Acad Sci U S A. 2013 Nov 19;110(47):19089-94.

By '**the low-affinity conformation, the tense (T) state**' we mean or include with mannose binding affinity closer to that of FimH in the low-affinity conformation than the high-affinity conformation (in particular, FimH from which the polypeptide of the invention was derived or principally derived, especially where complexed with FimC) e.g., less than 50%, 40%, 30%, 20%, 15%, 10%, 5%, 4%, 3%, 2% or 1% of the mannose binding affinity of FimH in the high-affinity conformation, for example, K_d ≈ 300 μM or higher (i.e. has no detectable mannose binding affinity), as disclosed in Kisiela DI, et al. Proc Natl Acad Sci U S A. 2013 Nov 19;110(47):19089-94. In one embodiment, the polypeptide of the invention is in the low-affinity conformation, for example has a mannose binding affinity of K_d of about, 100 μM, 200 μM, 300 μM, 400 μM, 500 μM, 600 μM, 700 μM, 800 μM, 900 μM, or 1 mM or has no detectable mannose binding affinity.

Mannose binding can be determined using any suitable means known in the art, for example, surface plasmon resonance (SPR) may be used t_o verify binding, binding specificity and binding constants of FimH constructs with mannosylated bovine serum albumin (Man-BSA) and glucosylated bovine serum albumin (Glc-BSA) (negative control), see, for example *Rabani et al.*, 2018, 'Conformational switch of the bacterial adhesin FimH in the absence of the regulatory domain: Engineering a minimalistic allosteric system' *J. Biol. Chem.*, 293(5):1835-1849, and Bouckaert J, et al. Mol Microbiol. 2005 Jan;55(2):441-55 which are incorporated by reference herein.

The conformation of FimH can also be assessed by measuring the binding of conformational antibodies, using any suitable means known in the art, for example, surface plasmon resonance and as described in the Examples. Exemplary antibodies are capable of recognising epitopes differently overlapping the mannose-binding pocket of FimH, for example antibodies binding to epitopes overlapping with the mannose binding pocket, for example epitopes limited to just one loop of the mannose-binding pocket. Exemplary antibodies are those disclosed in WO2016/183501, or in Kisiela DI, et al. Proc Natl Acad Sci U S A. 2013 Nov 19;110(47):19089-94, Kisiela DI, et al. PLoS Pathog. 2015 May 14;11(5):e1004857and which are incorporated by reference herein. In one embodiment, the conformational antibody has a variable heavy chain (VH)

29 Aug 2024
2021392894

sequence of SEQ ID NO: 125 and a variable light chain (VL) sequence of SEQ ID NO: 126. In one embodiment, the conformational antibody has a variable heavy chain (VH) sequence of SEQ ID NO: 127 and a variable light chain (VL) sequence of SEQ ID NO: 128.

QVQLQQSGAELATPGASVKMSCKASGYTSTNYWIHWVKQRPGQGLEWIGYINPTSGYTEYNQNFKDKATLTAD KSSSTAYMQLTSLTSEDSAVYYCARGVIRDFWGQGTTLTVSSAKTTAPSVYPLAPVCGDTTGSSVTLGCLVKG YFPEPVTLTWNSGSLSSGVHTFPAVLQSDLYTLSSSVTVTSS **[SEQ ID NO: 125] – VH of mAb 926**

DVLMTQTPLSLPVSLGDQASISCRSSQNIVHNNGNTYLEWYLQSPGQSPKLLIYKVSNRFSGVPDRFSGSGSG TDFTLKISRVEAEDLGVYYCFQGSHVPFTFGSGTKLEIK **[SEQ ID NO: 126] – VL (kappa) of mAb 926**

QVQLQQSGAELVRPGSSVKISCKASGYAFSSYWMNWVKQRPGQGLEWIGQIYPRDGDTNYNGKFMDKVTLTAD KSSNTAYMQLSSLTSEDSAVYFCEVGRGFYGMDYWGQGTSVTVSSAKTTAPSVYPLAPVCGDTTGSSVTLGCL VKGYFPEPVTLTWNSGSLSSGVHTFPAVLQSDLYTLSSSVTVTSS **[SEQ ID NO: 127] – VH of mAb 475**

DIVMTQSPKFMSTSVGDRVSVTCKASQNVSNVAWYQQKPGQSPKAMIYSASYRYSGVPGRFTGSGSGTDFTLT INNVQSEDLATYFCQQNSSFPFTFGGGTKLEIK [SEQ ID NO: 128] – **VL (kappa) of mAb 475**

The term '**amino acid**' as used herein includes the standard twenty genetically-encoded amino acids and their corresponding stereoisomers in the 'D' form (as compared to the natural 'L' form), omega-amino acids and other naturally-occurring amino acids, unconventional amino acids (e.g. α,α-disubstituted amino acids, N-alkyl amino acids, etc.) and chemically derivatised amino acids (see below).

Thus, when an amino acid is being specifically enumerated, such as 'alanine' or 'Ala' or 'A', the term refers to both L-alanine and D-alanine unless explicitly stated otherwise. Other unconventional amino acids may also be suitable components for polypeptides of the present invention, as long as the desired functional property is retained by the polypeptide. For the peptides shown, each encoded amino acid residue, where appropriate, is represented by a single letter designation, corresponding to the trivial name of the conventional amino acid.

By '**isolated**' we mean that the feature (e.g., the polypeptide) of the invention is provided in a context other than that in which it may be found naturally. One of skill in the art would understand that 'isolated' means altered 'by the hand of man' from its natural state, i.e., if it occurs in nature, it has been changed or removed from its original environment, or both. For example, a polynucleotide or a polypeptide naturally present in a living organism is not 'isolated' when in such living organism, but the same polynucleotide or polypeptide separated from the coexisting materials of its natural state is 'isolated' as the term is used in this disclosure. Further, a polynucleotide or polypeptide that is introduced into an organism by transformation, genetic manipulation or by any other recombinant method would be understood to be 'isolated' even if it is still present in said organism, which organism may be living or non-living, except where such transformation, genetic manipulation or other recombinant method produces an organism that is otherwise indistinguishable from the naturally-occurring organism.

By '**polypeptide'** we mean or include polypeptides and proteins.

2021392894 29 Aug 2024

By '**variant**' of the polypeptide we include insertions, deletions and/or substitutions, either conservative or non-conservative. In particular, the variant polypeptide may be a non-naturally occurring variant (i.e., does not, or is not known to, occur in nature). Variants may have at least 50% sequence identify with the/a reference sequence, for example, at least 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99% or 99.5%.

'**Sequence identity**' or '**identity**' can be determined by the Smith Waterman homology search algorithm as implemented in the MPSRCH program (Oxford Molecular), using an affine gap search with parameters gap open penalty=12 and gap extension penalty=1, or by the Needleman-Wunsch global alignment algorithm (see e.g. Rubin (2000) Pediatric. Clin. North Am. 47:269-285), using default parameters (e.g. with Gap opening penalty = 10.0, and with Gap extension penalty = 0.5, using the EBLOSUM62 scoring matrix). This algorithm is conveniently implemented in the needle tool in the EMBOSS package. Unless specified otherwise, where the application refers to sequence identity to a particular reference sequence, the identity is intended to be calculated over the entire length of that reference sequence. Alternatively, percent identity can be determined by methods well known in the art, for example using the LALIGN program (Huang and Miller, Adv. Appl. Math. (1991) 12:337-357, the disclosures of which are incorporated herein by reference) at the ExPASy facility website www.ch.embnet.org/software/LALIGN_form.html using as parameters the global alignment option, scoring matrix BLOSUM62, opening gap penalty –14, extending gap penalty –4. Alternatively, the percent sequence identity between two polypeptides may be determined using suitable computer programs, for example AlignX, Vector NTI Advance 10 (from Invitrogen Corporation) or the GAP program (from the University of Wisconsin Genetic Computing Group).

It will be appreciated that percent identity is calculated in relation to polymers (e.g., polypeptide or polynucleotide) whose sequence has been aligned.

Fragments and variants may be made using the methods of protein engineering and site-directed mutagenesis well known in the art (for example, see Molecular Cloning: a Laboratory Manual, 3rd edition, Sambrook & Russell, 2001, Cold Spring Harbor Laboratory Press, the disclosures of which are incorporated herein by reference).

It will be appreciated by skilled persons that the polypeptide of the invention, or fragment, variant or fusion thereof, may comprise one or more amino acids that are modified or derivatised.

Chemical derivatives of one or more amino acids may be achieved by reaction with a functional side group. Such derivatised molecules include, for example, those molecules in which free amino groups have been derivatised to form amine hydrochlorides, p-toluene sulphonyl groups, carboxybenzoxy groups, t-butyloxycarbonyl groups, chloroacetyl groups or formyl groups. Free carboxyl groups may be derivatised to form salts, methyl and ethyl esters or other types of esters and hydrazides. Free hydroxyl groups may be derivatised to form O-acyl or O-alkyl derivatives. Also included as chemical derivatives are those peptides which contain naturally occurring amino acid derivatives of the twenty standard amino acids. For example:

29 Aug 2024 2021392894

4-hydroxyproline may be substituted for proline; 5-hydroxylysine may be substituted for lysine; 3-methylhistidine may be substituted for histidine; homoserine may be substituted for serine and ornithine for lysine. Derivatives also include peptides containing one or more additions or deletions as long as the requisite activity is maintained. Other included modifications are amidation, amino terminal acylation (e.g. acetylation or thioglycolic acid amidation), terminal carboxylamidation (e.g. with ammonia or methylamine), and the like terminal modifications.

It will be further appreciated by persons skilled in the art that peptidomimetic compounds may also be useful. Thus, by 'polypeptide' we include peptidomimetic compounds which exhibit endolysin activity. The term 'peptidomimetic' refers to a compound that mimics the conformation and desirable features of a particular polypeptide as a therapeutic agent.

For example, the polypeptides described herein include not only molecules in which amino acid residues are joined by peptide (-CO-NH-) linkages but also molecules in which the peptide bond is reversed. Such retro-inverso peptidomimetics may be made using methods known in the art, for example such as those described in Meziere et al. (1997) J. Immunol. 159, 3230-3237, the disclosures of which are incorporated herein by reference. Such retro-inverse peptides, which contain NH-CO bonds instead of CO-NH peptide bonds, are much more resistant to proteolysis. Alternatively, the polypeptide of the invention may be a peptidomimetic compound wherein one or more of the amino acid residues are linked by a -γ(CH2NH)- bond in place of the conventional amide linkage.

It will be appreciated that the polypeptide may conveniently be blocked at its N- or C-terminus so as to help reduce susceptibility to exoproteolytic digestion, e.g., by amidation.

As discussed herein, a variety of uncoded or modified amino acids such as D-amino acids and N-methyl amino acids may be used to modify polypeptides of the invention. In addition, a presumed bioactive conformation may be stabilised by a covalent modification, such as cyclisation or by incorporation of lactam, disulphide or other types of bridges. Methods of synthesis of cyclic homodetic peptides and cyclic heterodetic peptides, including disulphide, sulphide and alkylene bridges, are disclosed in US 5,643,872. Other examples of cyclisation methods are discussed and disclosed in US 6,008,058, the relevant disclosures in which documents are hereby incorporated by reference. A further approach to the synthesis of cyclic stabilised peptidomimetic compounds is ring-closing metathesis (RCM).

By '**fusion'** of a polypeptide we include a polypeptide which is fused to any other polypeptide. For example, the polypeptide may comprise one or more additional amino acids, inserted internally and/or at the N- and/or C-termini of the amino acid sequence the polypeptides of the invention.

Thus, as described herein, in one embodiment the polypeptide of the first aspect of the invention comprises a polypeptide of the invention to which is fused an enzymatic domain from a different source (e.g., from a

29 Aug 2024
2021392894

source other than the polypeptide of the first aspect of the invention). Examples of suitable enzymatic domains include: L-alanoyl-D-glutamate endopeptidase; D-glutamyl-m-DAP endopeptidase; interpeptide bridge-specific endopeptidase; N-acetyl-β-D-glucosaminidase (=muramoylhydrolase); N-acetyl-β-D-muramidase (=lysozyme); lytic transglycosylase. Also, N-acetylmuramoyl-L-alanine amidase from other sources could be utilised (see Loessner, 2005, Current Opinion in Microbiology 8: 480-487, the disclosures of which are incorporated herein by reference).

For example, the said polypeptide may be fused to a polypeptide such as glutathione-S-transferase (GST) or protein A in order to facilitate purification of said polypeptide. Examples of such GST fusions are well known to those skilled in the art. Similarly, the said polypeptide may be fused to an oligo-histidine tag such as His6 or to an epitope recognised by an antibody such as the well-known Myc tag epitope. Fusions to any fragment, variant or derivative of said polypeptide are also included in the scope of the invention. It will be appreciated that fusions (or variants or derivatives thereof) which retain desirable properties, e.g., antigenic activity, are preferred. It is also particularly preferred if the fusions are ones which are suitable for use in the methods described herein.

For example, the fusion may comprise a further portion which confers a desirable feature on the said polypeptide of the invention; for example, the portion may be useful in detecting or isolating the polypeptide, promoting cellular uptake of the polypeptide, or directing secretion of the protein from a cell. The portion may be, for example, a biotin moiety, a radioactive moiety, a fluorescent moiety, for example a small fluorophore or a green fluorescent protein (GFP) fluorophore, as well known to those skilled in the art. The moiety may be an immunogenic tag, for example a Myc tag, as known to those skilled in the art or may be a lipophilic molecule or polypeptide domain that is capable of promoting cellular uptake of the polypeptide, as known to those skilled in the art.

It will be appreciated by persons skilled in the art that the polypeptides of the invention also include pharmaceutically acceptable acid or base addition salts of the herein described polypeptides. The acids which are used to prepare the pharmaceutically acceptable acid addition salts of the aforementioned base compounds useful in this invention are those which form non-toxic acid addition salts, i.e., salts containing pharmacologically acceptable anions, such as the hydrochloride, hydrobromide, hydroiodide, nitrate, sulphate, bisulphate, phosphate, acid phosphate, acetate, lactate, citrate, acid citrate, tartrate, bitartrate, succinate, maleate, fumarate, gluconate, saccharate, benzoate, methanesulphonate, ethanesulphonate, benzenesulphonate, p-toluenesulphonate and pamoate [i.e. 1,1'-methylene-bis-(2-hydroxy-3 naphthoate)] salts, among others.

Pharmaceutically acceptable base addition salts may also be used to produce pharmaceutically acceptable salt forms of the polypeptides. The chemical bases that may be used as reagents to prepare pharmaceutically acceptable base salts of the present compounds that are acidic in nature are those that form non-toxic base

salts with such compounds. Such non-toxic base salts include, but are not limited to those derived from such pharmacologically acceptable cations such as alkali metal cations (e.g. potassium and sodium) and alkaline earth metal cations (e.g. calcium and magnesium), ammonium or water-soluble amine addition salts such as N-methylglucamine-(meglumine), and the lower alkanolammonium and other base salts of pharmaceutically acceptable organic amines, among others.

The polypeptide, or fragment, variant, fusion or derivative thereof, may also be lyophilised for storage and reconstituted in a suitable carrier prior to use. Any suitable lyophilisation method (e.g. spray drying, cake drying) and/or reconstitution techniques can be employed. It will be appreciated by those skilled in the art that lyophilisation and reconstitution can lead to varying degrees of activity loss and that use levels may have to be adjusted upward to compensate. Preferably, the lyophilised (freeze dried) polypeptide loses no more than about 20%, or no more than about 25%, or no more than about 30%, or no more than about 35%, or no more than about 40%, or no more than about 45%, or no more than about 50% of its activity (prior to lyophilisation) when rehydrated.

Polypeptides of the invention are preferably provided in purified or substantially purified form i.e., substantially free from other polypeptides (e.g. free from naturally-occurring polypeptides), particularly from other *E. coli* or host cell polypeptides, and are generally at least about 50% pure (by weight), for example at least 70%, 80%, 90%, 95%, 96%, 97%, 98% 99%, 99.5%, 99.5% or 100% pure by weight (i.e., less than 50% of a composition is made up of other expressed polypeptides). Thus, the antigens in the compositions are separated from the whole organism with which the antigen molecule is expressed.

The FimH of (a) may be of any *Escherichia coli* or *Klebsiella pneumoniae* species (or a variant, fragment and/or fusion thereof) but, alternatively or additionally, (a) comprises or consists of:

(A) the amino acid sequence of SEQ ID NO: 1 (Genbank Accession no: ELL41155.1 (FimH of E. coli J96)), SEQ ID NO: 2, SEQ ID NO: 100 (Genbank Accession no: ABG72591.1 (FimH of UPEC 536)) , SEQ ID NO: 101, SEQ ID NO: 102 (Genbank Accession no: AAN83822.1 (FimH of CFT073)), SEQ ID NO: 103, SEQ ID NO: 104 (Genbank Accession no: AJE58925.1 (FimH of E. coli 789)), SEQ ID NO: 105, SEQ ID NO: 106 (Genbank Accession No. AAC35864.1, corresponding to nucleic acid sequence AF089840.1 (FimH of IHE3034), or SEQ ID NO: 107,

(B) an amino acid sequence comprising from 1 to 10 single amino acid alterations compared to SEQ ID NO: 1 (Genbank Accession no: ELL41155.1 (FimH of E. coli J96)), SEQ ID NO: 2 , SEQ ID NO: 100 (Genbank Accession no: ABG72591.1 (FimH of UPEC 536)) , SEQ ID NO: 101, SEQ ID NO: 102 (Genbank Accession no: AAN83822.1 (FimH of CFT073)), SEQ ID NO: 103, SEQ ID NO: 104 (Genbank Accession no: AJE58925.1 (FimH of E. coli 789)), SEQ ID NO: 105, SEQ ID NO: 106

(Genbank Accession No. AAC35864.1, corresponding to nucleic acid sequence AF089840.1 (FimH of IHE3034), or SEQ ID NO: 107, for example, 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 single amino acid alterations,

(C) an amino acid sequence with at least 70% sequence identity with SEQ ID NO: 1 (Genbank Accession no: ELL41155.1 (FimH of E. coli J96)), SEQ ID NO: 2 , SEQ ID NO: 100 (Genbank Accession no: ABG72591.1 (FimH of UPEC 536)) , SEQ ID NO: 101, SEQ ID NO: 102 (Genbank Accession no: AAN83822.1 (FimH of CFT073)), SEQ ID NO: 103, SEQ ID NO: 104 (Genbank Accession no: AJE58925.1 (FimH of E. coli 789)), SEQ ID NO: 105, SEQ ID NO: 106 (Genbank Accession No. AAC35864.1, corresponding to nucleic acid sequence AF089840.1 (FimH of IHE3034), or SEQ ID NO: 107, for example, 80%, 85%, 90%, 91%. 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% sequence identity, and/or

(D) a fragment of at least 10 consecutive amino acids from SEQ ID NO: 1 (Genbank Accession no: ELL41155.1 (FimH of E. coli J96)), SEQ ID NO: 2 , SEQ ID NO: 100 (Genbank Accession no: ABG72591.1 (FimH of UPEC 536)) , SEQ ID NO: 101, SEQ ID NO: 102 (Genbank Accession no: AAN83822.1 (FimH of CFT073)), SEQ ID NO: 103, SEQ ID NO: 104 (Genbank Accession no: AJE58925.1 (FimH of E. coli 789)), SEQ ID NO: 105, SEQ ID NO: 106 (Genbank Accession No. AAC35864.1, corresponding to nucleic acid sequence AF089840.1 (FimH of IHE3034), or SEQ ID NO: 107, for example, at least 20, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 250, 275, 280, 290 or 300 consecutive amino acids.

MKRVITLFAVLLMGWSVNAWSFACKTANGTAIPIGGGSANVYVNLAPVVNVGQNLVVDLSTQIFCHNDYPETI
TDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSL
IAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVYCAKSQNLGYYLS
GTTADAGNSIFTNTASFSPAQGVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTANYARTGGQVTAGNVQSI
IGVTFVYQ

[SEQ ID NO: 1] - GenBank: ELL41155.1 (signal peptide underlined)

FACKTANGTAIPIGGGSANVYVNLAPVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSG
TVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFVW
NIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVYCAKSQNLGYYLSGTTADAGNSIFTNTASFSPAQ
GVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTANYARTGGQVTAGNVQSIIGVTFVYQ

[SEQ ID NO: 2] - GenBank: ELL41155.1 minus 21aa signal peptide

MIVMKRVITLFAVLLMGWSVNAWSFACKTANGTAIPIGGGSANVYVNLAPAVNVGQNLVVDLSTQIFCHNDYP
ETITDYVTLQRGSAYGGVLSSFSGTVKYNGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKA

29 Aug 2024
2021392894

GSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVYCAKSQNLGY
YLSGTTADAGNSIFTNTASFSPAQGVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTANYARTGGQVTAGNV
QSIIGVTFVYQ

[SEQ ID NO: 100] - Genbank Accession no: ABG72591.1 (FimH of UPEC 536)) (signal peptide underlined)

FACKTANGTAIPIGGGSANVYVNLAPAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSSFSG
TVKYNGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFVW
NIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVYCAKSQNLGYYLSGTTADAGNSIFTNTASFSPAQ
GVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTANYARTGGQVTAGNVQSIIGVTFVYQ

[SEQ ID NO: 101] - Genbank Accession no: ABG72591.1 (FimH of UPEC 536) minus signal peptide

MIVMKRVITLFAVLLMGWSVNAWSFACKTANGTAIPIGGGSANVYVNLAPAVNVGQNLVVDLSTQIFCHNDYP
ETITDYVTLQRGSAYGGVLSSFSGTVKYNGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKA
GSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDASARDVTVTLPDYPGSVPIPLTVYCAKSQNLGY
YLSGTTADAGNSIFTNTASFSPAQGVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTANYARTGGQVTAGNV
QSIIGVTFVYQ

[SEQ ID NO: 102] - Genbank Accession no: AAN83822.1 (FimH of CFT073) (signal peptide underlined)

FACKTANGTAIPIGGGSANVYVNLAPAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSSFSG
TVKYNGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFVW
NIYANNDVVVPTGGCDASARDVTVTLPDYPGSVPIPLTVYCAKSQNLGYYLSGTTADAGNSIFTNTASFSPAQ
GVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTANYARTGGQVTAGNVQSIIGVTFVYQ

[SEQ ID NO: 103] - Genbank Accession no: AAN83822.1 (FimH of CFT073) minus signal peptide

MIVMKRVITLFAVLLMGWSVNAWSFACKTANGTAIPIGGGSANVYVNLAPVVNVGQNLVVDLSTQIFCHNDYP
ETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKA
GSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVYCAKSQNLGY
YLSGTTADAGNSIFTNTASFSPAQGVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTANYARTGGQVTAGNV
QSIIGVTFVYQ

[SEQ ID NO: 104] Genbank Accession no: AJE58925.1 (FimH of E. coli 789) (signal peptide underlined)

FACKTANGTAIPIGGGSANVYVNLAPVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSG
TVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFVW
NIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVYCAKSQNLGYYLSGTTADAGNSIFTNTASFSPAQ
GVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTANYARTGGQVTAGNVQSIIGVTFVYQ

[SEQ ID NO: 105] Genbank Accession no: AJE58925.1 (FimH of E. coli 789) minus signal peptide

29 Aug 2024

2021392894

MKRVITLFAVLLMGWSVNAWSFACKTANGTAIPIGGGSANVYVNLAPAVNVGQNLVVDLSTQIFCHNDYPETI
TDYVTLQRGAAYGGVLSSFSGTVKYNGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSL
IAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVYCAKSQNLGYYLS
GTTADAGNSIFTNTASFSPAQGVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTANYARTGGQVTAGNVQSI
IGVTFVYQ

[SEQ ID NO: 106] Genbank Accession No. AAC35864.1, (FimH of IHE3034), (signal peptide underlined)

FACKTANGTAIPIGGGSANVYVNLAPAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGAAYGGVLSSFSG
TVKYNGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFVW
NIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVYCAKSQNLGYYLSGTTADAGNSIFTNTASFSPAQ
GVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTANYARTGGQVTAGNVQSIIGVTFVYQ

[SEQ ID NO: 107] Genbank Accession No. AAC35864.1, (FimH of IHE3034), minus signal peptide

Alternatively or additionally, the polypeptide is a fragment, variant, fusion and/or derivative capable of inducing a specific immune response to a polypeptide selected from the group consisting of SEQ ID NO: 1, (Genbank Accession no: ELL41155.1 (FimH of E. coli J96)), SEQ ID NO: 2 , SEQ ID NO: 100 (Genbank Accession no: ABG72591.1 (FimH of UPEC 536)) , SEQ ID NO: 101, SEQ ID NO: 102 (Genbank Accession no: AAN83822.1 (FimH of CFT073)), SEQ ID NO: 103, SEQ ID NO: 104 (Genbank Accession no: AJE58925.1 (FimH of E. coli 789)), SEQ ID NO: 105, or SEQ ID NO: 106 (Genbank Accession No. AAC35864.1, corresponding to nucleic acid sequence AF089840.1 (FimH of IHE3034), SEQ ID NO: 107

By "**specific immune response**" we mean or include the capability to induce an immune response in a subject that generates (e.g., stimulates the release of) antibody capable of binding to an amino acid sequence specified. It is preferred that the antibody is capable of binding *in vivo*, i.e., under the physiological conditions in which the amino acid sequence or polypeptide exists on or inside of a subject's body. Such binding specificity may be determined by methods well known in the art, such as e.g. ELISA, immunohistochemistry, immunoprecipitation, Western blots and flow cytometry using transfected cells expressing the/a polypeptide of the invention.

Alternatively, or additionally, the immune response is an immune-activating response, for example, a protective immune response. The polypeptide may be capable of eliciting an *in vitro* protective immune response and/or an *in vivo* protective immune response when administered to a subject.

In the presence of co-stimulatory signals, T cells differentiate into specific phenotypic subtypes. Several of these subtypes are involved in suppressing or terminating natural inflammatory signals. By "immune-activating response" we mean and/or include that polypeptide induces or is capable of inducing an immune response in a subject that does not result in suppressing or terminating inflammation or inflammatory signals and, preferably, results in the activation or enhancement of inflammation or inflammatory signals (e.g., cytokines).

2021392894 29 Aug 2024

The *in vivo* protective immune response may be elicited in a mammal. Alternatively or additionally, the mammal is selected from the group consisting of armadillo (*dasypus novemcinctus*), baboon (*papio anubis; papio cynocephalus*), camel (*camelus bactrianus, camelus dromedarius, camelus ferus*), cat (*felis catus*), dog (*canis lupus familiaris*), horse (*equus ferus caballus*), ferret (*mustela putorius furo*), goat (*capra aegagrus hircus*), guinea pig (*cavia porcellus*), golden hamster (*mesocricetus auratus*), kangeroo (*macropus rufus*), llama (*lama glama*), mouse (*mus musculus*), pig (*sus scrofa domesticus*), rabbit (*oryctolagus cuniculus*), rat (*rattus norvegicus*), rhesus macaque (*macaca mulatta*), sheep (*ovis aries*), non-human primates, and human (*Homo sapiens*).

Alternatively or additionally, two glycine residues of the linker connecting $FimH_L$ to $FimH_p$ can be deleted to reduce the flexibility of $FimH_L$ and reduce mannose binding. For example, glycine residues 196 and 197 of polypeptide portion (a), relative to SEQ ID NO: 1, glycine residues 180 and 181 of polypeptide portion (a), relative to SEQ ID NO: 1, glycine residues 183 and 184 of polypeptide portion (a), relative to SEQ ID NO: 100, glycine residues 183 and 184 of polypeptide portion (a), relative to SEQ ID NO: 102, glycine residues 183 and 184 of polypeptide portion (a), relative to SEQ ID NO: 104, are:

(i) present; or

(ii) deleted.

Alternatively or additionally, one or more amino acids of the polypeptide known or predicted to be N-glycosylated or O-glycosylated are substituted with amino acids unsusceptible or less susceptible to glycosylation, e.g., serine (S), aspartic acid (D), alanine (A) or glutamine (Q). Alternatively or additionally, only polypeptide portion (a) includes amino acid substitutions to reduce or abolish N- and/or O-glycosylation.

N- and/or O-glycosylation can be determined using any suitable means known in the art, for example, using the NetNGlyc 1.0 and NetOGlyc 4.0 Server (accessible at http://www.cbs.dtu.dk/services/NetOGlyc/ and http://www.cbs.dtu.dk/services/NetOGlyc/) using default settings.

Alternatively or additionally, polypeptide portion (a) includes one or more of the following amino acid substitutions relative to SEQ ID NO: 2: N28S, N91D, N249D, N256D, or at the positions of SEQ ID NO: 101, 103 and 105 corresponding those positions of SEQ ID NO:2, for example, one, two, three or four of the amino acid substitutions.

Alternatively or additionally, the donor-strand complementing amino acid sequence (b) comprises or consists of:

(i) 6-28 amino acids of SEQ ID NO: 3; or a fragment and/or variant thereof, or

(ii) 8-36 amino acids of SEQ ID NO: 4; or a fragment and/or variant thereof,

ASATIQAADVTITVNGKVVAKPCTVSTT

29 Aug 2024 2021392894

[SEQ ID NO: 3] – FimG donor strand and flanking region (donor strand underlined)

PSMDKSKLTENTLQLAIISRIKLYYRPAKLALPPDQ

[SEQ ID NO: 4] – FimC donor strand and flanking region (donor strand underlined)

Alternatively or additionally, portion (b) comprises or consists of 6-28 amino acids of SEQ ID NO: 3 (or a fragment and/or variant thereof), which amino acids are selected from the group consisting of:

(i) amino acids 1-28 of SEQ ID NO: 3; or a fragment and/or variant thereof,

(ii) amino acids 2-27 of SEQ ID NO: 3; or a fragment and/or variant thereof,

(iii) amino acids 3-26 of SEQ ID NO: 3; or a fragment and/or variant thereof,

(iv) amino acids 4-25 of SEQ ID NO: 3; or a fragment and/or variant thereof,

(v) amino acids 5-24 of SEQ ID NO: 3; or a fragment and/or variant thereof,

(vi) amino acids 6-23 of SEQ ID NO: 3; or a fragment and/or variant thereof,

(vii) amino acids 7-22 of SEQ ID NO: 3; or a fragment and/or variant thereof,

(viii) amino acids 8-21 of SEQ ID NO: 3; or a fragment and/or variant thereof,

(ix) amino acids 9-20 of SEQ ID NO: 3; or a fragment and/or variant thereof,

(x) amino acids 10-19 of SEQ ID NO: 3; or a fragment and/or variant thereof,

(xi) amino acids 11-18 of SEQ ID NO: 3; or a fragment and/or variant thereof, and

(xii) amino acids 12-17 of SEQ ID NO: 3; or a fragment and/or variant thereof.

Alternatively or additionally, portion (b) comprises or consists of 8-36 amino acids of SEQ ID NO: 4 (or a fragment and/or variant thereof), which amino acids are selected from the group consisting of:

(i) amino acids 1-36 of SEQ ID NO: 4; or a fragment and/or variant thereof,

(ii) amino acids 2-35 of SEQ ID NO: 4; or a fragment and/or variant thereof,

(iii) amino acids 3-34 of SEQ ID NO: 4; or a fragment and/or variant thereof,

(iv) amino acids 4-33 of SEQ ID NO: 4; or a fragment and/or variant thereof,

(v) amino acids 5-32 of SEQ ID NO: 4; or a fragment and/or variant thereof,

(vi) amino acids 6-31 of SEQ ID NO: 4; or a fragment and/or variant thereof,

(vii) amino acids 7-30 of SEQ ID NO: 4; or a fragment and/or variant thereof,

(viii) amino acids 8-29 of SEQ ID NO: 4; or a fragment and/or variant thereof,

(ix) amino acids 9-28 of SEQ ID NO: 4; or a fragment and/or variant thereof,

(x) amino acids 10-27 of SEQ ID NO: 4; or a fragment and/or variant thereof,

(xi) amino acids 11-26 of SEQ ID NO: 4; or a fragment and/or variant thereof,

(xii) amino acids 12-25 of SEQ ID NO: 4; or a fragment and/or variant thereof,

(xiii) amino acids 13-24 of SEQ ID NO: 4; or a fragment and/or variant thereof,

2021392894 29 Aug 2024

(xiv) amino acids 14-23 of SEQ ID NO: 4; or a fragment and/or variant thereof,

(xv) amino acids 15-24 of SEQ ID NO: 4; or a fragment and/or variant thereof, and

(xvi) amino acids 16-23 of SEQ ID NO: 4; or a fragment and/or variant thereof.

Alternatively or additionally, the donor-strand complementing amino acid sequence (b) comprises or consists of:

(A) the amino acid sequence of SEQ ID NO: 5 or SEQ ID NO: 6,

(B) an amino acid sequence comprising from 1 to 10 single amino acid alterations compared to SEQ ID NO: 5 or SEQ ID NO: 6, for example, 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 single amino acid alterations,

(C) a fragment of at least 7 consecutive amino acids from SEQ ID NO: 5, for example, at least 8, 9, 10, 11, 12, or 13 consecutive amino acids from SEQ ID NO: 5, and/or

(D) a fragment of at least 7 consecutive amino acids from SEQ ID NO: 6, for example, at least 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, or 18 consecutive amino acids from SEQ ID NO: 6.

ADVTITVNGKVVAK **[SEQ ID NO: 5] – FimG donor strand**

ENTLQLAIISRIKLYYRP **[SEQ ID NO: 6] – FimC donor strand**

In one preferred embodiment, the donor-strand complementing amino acid sequence (b) comprises or consists of SEQ ID NO: 5. Alternatively or additionally, the donor-strand complementing amino acid sequence (b) comprises or consists of SEQ ID NO: 6.

Alternatively or additionally, the donor-strand complementing amino acid sequence (b) is:

(i) directly joined to the C-terminus of (a), or

(ii) joined to the C-terminus of (a) via a first linker.

Alternatively or additionally, the first linker (or "L") comprises or consists of 2-20 amino acids, for example, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20 amino acids. Alternatively or additionally, the first linker begins with proline. In a preferred embodiment, the first linker begins with proline. Alternatively or additionally, the first linker comprises or consists of polar amino acids, for example, wherein the first linker is entirely comprised of polar amino acids or, if the first linker begins with proline, the remainder of the amino acids are polar. Alternatively or additionally, the first linker comprises or consists of:

(i) PGDGN **[SEQ ID NO: 7]**, or a variant or fusion thereof, or

(ii) DNKQ **[SEQ ID NO: 8]**, or a variant or fusion thereof.

In one preferred embodiment the first linker (or "L") comprises or consists of SEQ ID NO: 7.

2021392894 29 Aug 2024

Alternatively or additionally, the polypeptide comprises a protein purification affinity tag at the N-terminus, C-terminus and/or internally, for example, 6, 7, 8, 9 or 10 consecutive histidines.

Alternatively or additionally, "X" comprises a cell secretion leader sequence. Alternatively or additionally, the polypeptide comprises a cell secretion leader sequence:

(i) upstream of (a), or

(ii) at the N-terminus of the polypeptide.

Alternatively or additionally, the cell secretion leader sequence is selected from the group consisting of:

(i) METDTLLLWVLLLWVPGSTGD **[SEQ ID NO: 9]**, or a variant or fusion thereof,

(ii) METDTLLLWVLLLWVPGSTGDAAQPARRARRTKLAL **[SEQ ID NO: 10]**, or a variant or fusion thereof,

(iii) MRLLAKIICLMLWAICVA **[SEQ ID NO: 11]**, or a variant or fusion thereof,

(iv) MGWSCIILFLVATATGVHS **[SEQ ID NO: 12]**, or a variant or fusion thereof,

(v) METPAELLFLLLLWLPDTTG **[SEQ ID NO: 13]**, or a variant or fusion thereof,

(vi) METDTLLLWVLLLWVPGSTG **[SEQ ID NO: 108]**, or a variant or fusion thereof or

(vii) MEFGLSWVFLVAILEGVHC **[SEQ ID NO: 14]**, or a variant or fusion thereof.

Alternatively, or additionally, "X" is a methionine (M) residue, particularly when the polypeptide is expressed in *E. coli* host cells.

Alternatively or additionally, the polypeptide comprises a nanoparticle domain at the N-terminus or C-terminus. Thus, in one embodiment "X" comprises a nanoparticle domain or "Y" comprises a nanoparticle domain. By '**nanoparticle domain**' we mean or include amino acid sequences capable of self-assembly to form protein complexes, in particular, globular protein complexes. By '**self-assembly'** we mean or include assembly with nanoparticle domains of the same type (e.g., if the nanoparticle domain is a ferritin domain, capable of assembling with other ferritin domains to form protein complexes, such as globular protein complexes). In particular, the nanoparticle domains of the invention are capable of self-assembly when they form a portion of the/a polypeptide of the invention.

Alternatively or additionally, the nanoparticle domain is selected from the group consisting of:

(a) ferritin (for example, **[SEQ ID NO: 15]** or [SEQ ID NO: 109] (*Helicobacter pylori*),**[SEQ ID NO: 16]** (*Escherichia coli*)), or any one of [SEQ ID NO: 149]- [SEQ ID NO: 152] (stabilized *Escherichia coli*)), or a variant and/or fragment thereof,

(b) iMX313 (for example **[SEQ ID NO: 17]**), or a variant and/or fragment thereof,

(c) mI3 (for example **[SEQ ID NO: 18]**), or a variant and/or fragment thereof,

2021392894 29 Aug 2024

(d) encapsulin (for example **[SEQ ID NO: 19]**), or a variant and/or fragment thereof, or

(e) Self-assembling viral coat proteins, such as *Acinetobacter* phage AP205 coat protein (NCBI Reference Sequence: NP_085472.1), Hepatitis B virus core protein (HBc) **[SEQ ID NO: 110]**, or bacteriophage Qβ **[SEQ ID NO: 111]**, or a variant and/or fragment thereof.

DIIKLLNEQVNKEMNSSNLYMSMSSWCYTHSLDGAGLFLFDHAAEEYEHAKKLIIFLNENNVPVQLTSISAPE HKFEGLTQIFQKAYEHEQHISESINNIVDHAIKSKDHATFNFLQWYVAEQHEEEVLFKDILDKIELIGNENHG LYLADQYVKGIAKSRK **[SEQ ID NO: 15]** – *H. pylori* ferritin

DIIKLLNEQVNKEMNSSNLYMSMSSWCYTHSLDGAGLFLFDHAAEEYEHAKKLIIFLNENNVPVQLTSISAPE HKFEGLTQIFQKAYEHEQHISESINNIVDHAIKSKDHATFNFLQWYVAEQHEEEVLFKDILDKIELIGNENHG LYLADQYVKGIAKSRKS **[SEQ ID NO: 109]** – *H. pylori* ferritin (with terminal S)

LKPEMIEKLNEQMNLELYSSLLYQQMSAWCSYHTFEGAAAFLRRHAQEEMTHMQRLFDYLTDTGNLPRINTVE SPFAEYSSLDELFQETYKHEQLITQKINELAHAAMTNQDYPTFNFLQWYVSEQHEEEKLFKSIIDKLSLAGKS GEGLYFIDKELSTLDTQN **[SEQ ID NO: 16]** – *E. coli* ferritin

KKQGDADVCGEVAYIQSVVSDCHVPTAELRTLLEIRKLFLEIQKLKVELQGLSKEG **[SEQ ID NO: 17]** iMX313

MKMEELFKKHKIVAVLRANSVEEAKKKALAVFLGGVHLIEITFTVPDADTVIKELSFLKEMGAIIGAGTVTSV EQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKG PFPNVKFVPTGGVNLDNVCEWFKAGVLAVGVGSALVKGTPVEVAEKAKAFVEKIRGCTE **[SEQ ID NO: 18]** mi3

MEFLKRSFAPLTEKQWQEIDNRAREIFKTQLYGRKFVDVEGPYGWEYAAHPLGEVEVLSDENEVVKWGLRKSL PLIELRATFTLDLWELDNLERGKPNVDLSSLEETVRKVAEFEDEVIFRGCEKSGVKGLLSFEERKIECGSTPK DLLEAIVRALSIFSKDGIEGPYTLVINTDRWINFLKEEAGHYPLEKRVEECLRGGKIITTPRIEDALVVSERG GDFKLILGQDLSIGYEDREKDAVRLFITETFTFQVVNPEALILLKF **[SEQ ID NO: 19]** – encapsulin

MDIDPYKEFGATVELLSFLPSDFFPSVRDLLDTASALYREALESPEHCSPHHTALRQAILCWGELMTLATWVG NNLEDASRDLVVNYVNTNMGLKIRQLLWFHISCLTFGRETVLEYLVSFGVWIRTPPAYRPPNAPILSTLPETT VV [SEQ ID NO: 110] HBC

MAKLETVTLGNIGKDGKQTLVLNPRGVNPTNGVASLSQAGAVPALEKRVTVSVSQPSRNRKNYKVQVKIQNPT ACTANGSCDPSVTRQAYADVTFSFTQYSTDEERAFVRTELAALLASPLLIDAIDQLNPAY [SEQ ID NO: 111] - Qbeta

LKPEMIEKLNEQMNLELYSSLLYQQMSAWCSYHGFEGAAAFLRRHAQEEMTHMQRLFDYLTDTGNLPRIDTIP SPFAEYSSLDELFQETYKHEQLITQKINELAHAAMTNQDYPTFNFLQWYVAEQHEEEKLFKSIIDKLSLAGKS GEGLYFIDKELSTLDTQN [SEQ ID NO: 149] - **1EUM_0_5 - stabilized E. coli ferritin**

LKPEMIEKLNEQMNLELYSSLLYQQMSAWCSYHGFEGAAAFLRRHAQEEMTHMQRLFDYLTDTGNLPRINTIP SPFAEYSSLDELFQETYKHEQLITQKINELAHAAMTNQDYPTFNFLQWYVAEQHEEEKLFKSIIDKLSLAGKS GEGLYFIDKELSTLDTQN [SEQ ID NO: 150] - **1EUM_2 - stabilized E. coli ferritin**

LKPEMIEKLNEQMNLELYSSLLYQQMSAWCSYHGFEGAAAFLRRHAQEEMTHMQRLFDYLTDTGNLPRINTVP SPFAEYSSLDELFQETYKHEQLITQKINELAHAAMTNQDYPTFNFLQWYVAEQHEEEKLFKSIIDKLSLAGKS GEGLYFIDKELSTLDTQN [SEQ ID NO: 151] - **1EUM_2_5 - stabilized E. coli ferritin**

LKPEMIEKLNEQMNLELYSSLLYQQMSAWCSYHGFEGAAAFLRRHAQEEMTHMQRLFDYLTDTGNLPRINTVE SPFAEYSSLDELFQETYKHEQLITQKINELAHAAMTNQDYPTFNFLQWYVSEQHEEEKLFKSIIDKLSLAGKS GEGLYFIDKELSTLDTQN [SEQ ID NO: 152] - **1EUM_6- stabilized E. coli ferritin**

2021392894 29 Aug 2024

Alternatively or additionally, the nanoparticle domain is:

(i) directly joined to the polypeptide, or

(ii) joined to the polypeptide via a second linker.

Alternatively or additionally, the second linker comprises or consists of between 2-20 amino acids, for example, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20 amino acids. Alternatively or additionally, the second linker comprises or consists glycines (G) and/or serines (S), or comprises at least 50% glycines (G) and/or serines (S), for example, at least 60%, 70%, 80&, 90% or 95% glycines (G) and/or serines (S).

Alternatively or additionally, the second linker is selected from the group consisting of:

(a) GSSGSGSGS [SEQ ID NO: 112] or a variant or fusion thereof,

(b) GGSGS [SEQ ID NO: 113] or a variant or fusion thereof,

(c) GGS or a variant or fusion thereof,

(d) SGSHHHHHHHHGGS [SEQ ID NO: 114], or a variant or fusion thereof,

(e) AKFVAAWTLKAAA [SEQ ID NO: 115] or a variant or a fusion thereof,

(f) GGGGSLVPRGSGGGGS [SEQ ID NO: 116], or a variant or a fusion thereof,

(g) EAAAKEAAAKEAAAKA [SEQ ID NO: 117], or a variant or a fusion thereof,

(h) SGSFVAAWTLKAAAGGS [SEQ ID NO: 118] or a variant or a fusion thereof, and

(i) SGSGSGGGGGGS [SEQ ID NO: 119] or a variant or a fusion thereof.

The linker AKFVAAWTLKAAA [SEQ ID NO: 115], also known as Pan HLA DR-binding epitope (PADRE) is a peptide that activates antigen specific-CD4+ T cells, which has been proposed as a carrier epitope suitable for use in the development of synthetic and recombinant vaccines, as disclosed in "Linear PADRE T Helper Epitope and Carbohydrate B Cell Epitope Conjugates Induce Specific High Titer IgG Antibody Responses" 10.4049/jimmunol.164.3.1625 whose disclosure is incorporated by reference herein. The linkers GGGGSLVPRGSGGGGS [SEQ ID NO: 116] and EAAAKEAAAKEAAAKA [SEQ ID NO: 117] are rigid linkers which are not capable of folding into an alpha helix.

Alternatively or additionally, the nanoparticle domain is:

(a) upstream of (a),

(b) at the N-terminus of the polypeptide,

(c) downstream of (b), or

(d) at the C-terminus of the polypeptide.

29 Aug 2024
2021392894

In a further aspect, it is provided a designed and *de novo* polypeptide monomer (and the nucleic acid molecules encoding them) capable of self assembling into nanoparticles (i.e., protein nanoparticles). Host cells, vectors or constructs, and method for making or using such polypeptide monomers and protein nanoparticles are also provided. The present invention further relates to nanoparticles (NPs) that have a surface structure comprising, or consisting of, at least one such polypeptide monomer and that, optionally, carries one or more antigen molecule.

The polypeptide monomer of the invention is mutated as compared to its wild type counterpart monomer (i.e., the *E. coli* bacterial ferritin [SEQ ID NO: 16]), and may thereby have an increased stability, such as an improved thermal stability or folding stability in kcal/mol as compared to its wild type counterpart monomer, which may thereby form a self-assembled nanoparticle with an improved thermal stability or folding stability in kcal/mol as compared to its wild type counterpart nanoparticle.

"**Increased stability**" means the molecule has a lower rate of unfolding, decreased misfolding, reduced protein domain movements, reduced protein domain rearrangements, increased half-life (*in-vitro* or *in-vivo*), increased shelf life, increased melting temperature (Tm) (meaning an increase in at least one melting temperature, if the molecule has two or more), lower folding free energy value (kcal/mol), lower binding free energy value (as in the case of a subunit that binds other subunits to form a macromolecule), or a combination thereof; as compared to a control molecule or its wild type counterpart under comparable or the same conditions (*e.g.,* temperature and/or pH). For clarity of the example, if the stability of a molecule is increased via one or more mutations ("stabilizing mutations" such as one or more amino acid mutations), a "**control molecule"** or its "**wild type counterpart"** means a molecule that *does not* comprise the one or more stabilizing mutations. With respect to the present invention, a monomer or nanoparticle may be described as having an increased stability (e.g., increased thermostability and/or increased folding stability and/or increased binding stability) as compared to its wildtype counterpart molecule under comparable (or the same) conditions. "**Conditions**" as used herein includes experimental and physiological conditions. *See, e.g.,* U.S. Pub. No. 2011/0229507; Clapp *et al.*, 2011 J. Pharm. Sci. 100(2): 388-401, discussing increased stability via adjuvants and assessing antigen stability in altered pH, hydration, and temperature conditions; and Rossi *et al.*, 2016 Infect. Immun. 84(6): 1735-1742. For clarity, "stability" may be specified as "**thermostability**" which means the molecule's resistance to unfolding at a particular temperature and which is usually conveyed in the field by the molecule's melting temperature(s), specifically an increase in the molecule's melting temperature (of which there may be more than one melting temperature for oligomeric proteins such as dimers or trimers), *see* Kumar *et al.* 2000 Prot. Eng. Des. Sel. "*Factors enhancing protein thermostability"* 13(3): 179-191; and Miotto *et al.* 2018 bioRxiv doi: 10.1101/354266 "*Insights on protein thermal stability: a graph representation of molecule interactions"*). As the context requires, the thermostability of two or more molecules (such as two or more *modified* molecules that each comprise one

29 Aug 2024 2021392894

or more stabilizing mutation) may be compared and one may be said to be more thermostable than the other (*i.e.,* have an enhanced or increased thermostability as compared to the other). Stability, especially thermostability, herein may be provided by the delta stability (dStability or dS) scoring method, which is the computationally-determined difference between the relative thermostability of an *in-silico* mutant protein and that of a control or its wild type counterpart (*i.e.,* non-stabilized-mutant) protein. Methods of determining dStability are known (WO 2020/079586 (PCT/IB2019/058777), MALITO *et al.*) and may include the use of tools such as Molecular Operating Environment (MOE) software (REF: Molecular Operating Environment (MOE) software; Chemical Computing Group Inc., available at WorldWideWeb(www).chemcomp.com). dS is measured by kcal/mol. Lower dS values indicate higher protein stability, while higher dS values indicate lower protein stability. It may be specified that the mutant polypeptides of the present invention have a higher relative thermostability (in kcal/mol) as compared to a non-mutant polypeptide under the same or comparable experimental conditions. It may be further specified that the mutant polypeptides of the present invention have a lower dS value than a non-mutant polypeptide under the same or comparable experimental conditions. It will be understood from the present invention that a mutant polypeptide having a lower dS value as compared to a non-mutant polypeptide under the same or comparable experimental conditions *is more stable* than the non-mutant polypeptide. The stability enhancement can be assessed using differential scanning calorimetry (DSC) as discussed in Bruylants *et al.* 2005 Curr. Med. Chem. 12: 2011-2020 and Calorimetry Sciences Corporation's "Characterizing Protein stability by DSC" (Life Sciences Application Note, Doc. No. 20211021306 February 2006) or by differential scanning fluorimetry (DSF). An increase in (thermo)stability may be characterized as an at least about 2°C increase in thermal transition midpoint (T_m), as assessed by DSC or DSF. *See, for example,* Thomas *et al.*, 2013 Hum. Vaccin. Immunother. 9(4): 744-752. A "significant" increase in, or enhancement of, thermostability is defined as an increase of at least 5°C in the calculated Tm of a complex (calculated by, for example, the protocol provided at Example 4.7 of WO 2020/079586 (PCT/IB2019/058777), MALITO *et al.*). For clarity, "stability" herein may be specified as "**folding stability**" which refers to the molecule's folding free energy (reported in kilocalories per mole (kcal/mol)) and which may be determined using a variety of known techniques (*see* the Examples section herein as well as*, e.g.,* Zhang *et al.* 2012 Bioinformatics 28(5): 664-671). As the context requires, the folding stability of two or more molecules may be compared and one may be said to be *more stable* than the other because it has a *lower* folding free energy value (in kcal/mol). It may be specified that a monomer or nanoparticle of the present invention has a higher/increased folding stability as compared to a control molecule or its wild type counterpart under the same or comparable conditions (*e.g.,* experimental conditions). A "significant" increase in, or enhancement of, folding stability is defined as a folding free energy change value that is at least 100 kcal/mol lower than the folding free energy change value (in kcal/mol) of the comparison molecule in comparable or the same conditions.

2021392894 29 Aug 2024

In one embodiment, the polypeptide monomer of the invention comprises an amino acid sequence that has at least 80% sequence identity, for example at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to the amino acid sequence SEQ ID NO: 16 and has one or more mutations from the group consisting of: glycine (G) at the position that aligns to residue 34 of SEQ ID NO: 16 (T34G mutation), aspartic acid (D) at the position that aligns to residue 70 of SEQ ID NO: 16 (N70D mutation), isoleucine (I) at the position that aligns to residue 72 of SEQ ID NO: 16 (V72I mutation) and alanine (A) at the position that aligns to residue 124 of SEQ ID NO: 16 (S124A mutation).

In one preferred embodiment, the polypeptide monomer comprises an amino acid sequence that has at least 80% sequence identity, for example at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to the amino acid sequence SEQ ID NO: 16 and has glycine (G) at the position that aligns to residue 34 of SEQ ID NO: 16 (T34G mutation), aspartic acid (D) at the position that aligns to residue 70 of SEQ ID NO: 16 (N70D mutation), isoleucine (I) at the position that aligns to residue 72 of SEQ ID NO: 16 (V72I mutation) and alanine (A) at the position that aligns to residue 124 of SEQ ID NO: 16 (S124A mutation). In one embodiment, the polypeptide monomer comprises an amino acid sequence that has at least 80% sequence identity, for example at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to the amino acid sequence SEQ ID NO: 149. In one embodiment, the polypeptide monomer comprises the amino acid sequence of SEQ ID NO: 149.

In one embodiment, the polypeptide monomer comprises an amino acid sequence that has at least 80% sequence identity, for example at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to the amino acid sequence SEQ ID NO: 16 and has glycine (G) at the position that aligns to residue 34 of SEQ ID NO: 16 (T34G mutation), isoleucine (I) at the position that aligns to residue 72 of SEQ ID NO: 16 (V72I mutation) and alanine (A) at the position that aligns to residue 124 of SEQ ID NO: 16 (S124A mutation). In one embodiment, the polypeptide monomer comprises an amino acid sequence that has at least 80% sequence identity, for example at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to the amino acid sequence SEQ ID NO: 150. In one embodiment, the polypeptide monomer comprises the amino acid sequence of SEQ ID NO: 150.

In one embodiment, the polypeptide monomer comprises an amino acid sequence that has at least 80% sequence identity, for example at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to the amino acid sequence SEQ ID NO: 16 and has glycine (G) at the position that aligns to residue 34 of SEQ ID NO: 16 (T34G mutation), and alanine (A) at the position that aligns to residue 124 of SEQ ID NO: 16 (S124A mutation). In one embodiment, the polypeptide monomer comprises an amino acid sequence that has at least 80% sequence identity, for example at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to the amino acid sequence SEQ ID NO: 151. In one embodiment, the polypeptide monomer comprises the amino acid sequence of SEQ ID NO: 151.

2021392894 29 Aug 2024

In one embodiment, the polypeptide monomer comprises an amino acid sequence that has at least 80% sequence identity, for example at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to the amino acid sequence SEQ ID NO: 16 and has glycine (G) at the position that aligns to residue 34 of SEQ ID NO: 16 (T34G mutation). In one embodiment, the polypeptide monomer comprises an amino acid sequence that has at least 80% sequence identity, for example at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to the amino acid sequence SEQ ID NO: 152. In one embodiment, the polypeptide monomer comprises the amino acid sequence of SEQ ID NO: 152.

The designed and *de novo* polypeptide monomers of the present invention are capable of self-assembly into approximately spherical nanoparticles (e.g., with an exterior surface structure diameter of about 5nm to about 30nm, preferably of about 15 to 20 nm). The polypeptide monomers of the present invention may therefore be used for providing self-assembled protein nanoparticles and, optionally, wherein the self-assembled protein nanoparticle carries (e.g., displays) at least one antigen molecule, at least one immunostimulant molecule, or at least one antigen molecule and at least one immunostimulant molecule. In one embodiment, the nanoparticles of the present invention (e.g., approximately spherical nanoparticles of the present invention) consist of 24 monomer subunits (e.g., wherein at least one monomer subunit is a polypeptide monomer of the present invention) and have an underlying geometry that is octahedral symmetry.

Nanoparticles (naturally occurring and recombinant nanoparticles, e.g., computationally-designed nanoparticles), methods of making them, and their use as, for example, scaffolds (or "carriers") of one or more antigens or immunostimulants (i.e., "pharmaceutically acceptable nanoparticles") are known in the art.

As would be recognized by the art (see, e.g., Ueda et al. 2020 eLife 9: e57659; Pan et al. 2020 Adv. Mater. 32:2002940), protein nanoparticle of the present invention may be used as a "scaffold" by which it carries (through conjugation, i.e., connection, attachment, linkage, fusion, bond or ligation to the exterior surface structure of the nanoparticle) an antigen, an immunostimulant, multiple copies of the same antigen, multiple copies of the same immunostimulant, a mixture of two or more antigens (e.g., two, three, four, or five antigens; i.e., antigen bi-, tri-, quadra-, or pentavalent), a mixture of two or more immunostimulants (e.g., two, three, four, or five immunostimulants; i.e., immunostimulant bi-, tri-, quadra-, or pentavalent), or a mixture of one or more antigen(s) with one or more immunostimulant(s).

In certain embodiments, the self-assembly of polypeptide monomers places their N-termini at the outer/external surface of the nanoparticle and their C-termini at the inner/core/interior surface of the nanoparticle. In this way, an antigen or immunostimulant that is linked to the N-terminus of a polypeptide monomer is displayed at the exterior surface of the assembled nanoparticle. In other embodiments, the self-assembly of polypeptide monomers places their C-termini at the outer/external surface of the nanoparticle and their N-termini at the inner/core/interior surface of the nanoparticle. In this way, an antigen or

29 Aug 2024
2021392894

immunostimulant that is linked to the C-terminus of a polypeptide monomer is displayed at the exterior surface of the assembled nanoparticle. In certain other embodiments, an antigen or immunostimulant is linked to the N-terminus of a polypeptide monomer and an antigen or immunostimulant is linked to the C-terminus of that polypeptide monomer (antigen(s) and/or immunostimulant(s) being the same or different) such that an antigen or immunostimulant is displayed on the exterior surface and carried on the interior surface of the assembled nanoparticle.

So an embodiment of the present invention provides a nanoparticle carrying one or more molecule(s) (e.g., wherein the molecule(s) is/are heterologous as compared to one or more (e.g., all) of the nanoparticle monomers) and optionally wherein the one or more molecule(s) is/are displayed on the exterior surface of the nanoparticle. Where said one or more displayed molecules (e.g., antigen(s) and/or immunostimulant(s)) are proteins (e.g., are all proteins), they may be expressed as part of the polypeptide monomers (i.e., as fusion protein monomers), such that self-assembly of the nanoparticle results in display of the proteins on the nanoparticle exterior surface. Alternatively, a protein display molecule may be attached to the assembled nanoparticle, for example, by chemical or biological conjugation as discussed herein and as known in the art. In a further embodiment of the present invention, the display molecule is a poly- or oligo-saccharide (such as a bacterial capsular polysaccharide); the saccharide may be linked to a nanoparticle to provide a "glycoconjugate". see Polonskaya et al. 2017 J. Clin. Invest. 127(4):1492-1504; Pan et al. 2020 Adv. Mater. 32:2002940.

In one embodiment, the antigen is a polypeptide having an amino acid sequence comprising or consisting of (a) FimH; or a variant, fragment and/or fusion of FimH, and (b) a donor-strand complementing amino acid sequence, wherein (b) is downstream of (a), or as otherwise described herein. In one embodiment, the antigen comprises or consists of an amino acid sequence that has at least 80% sequence identity, for example at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to the amino acid sequence SEQ ID NO: 124. In one embodiment, the antigen comprises or consists of the amino acid sequence of SEQ ID NO: 124.

In one embodiment, the nanoparticle of an amino acid sequence that has at least 80% sequence identity, for example at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to the amino acid sequence SEQ ID NO: 130 or 153. In one embodiment, the nanoparticle comprises or consists of the amino acid sequence of SEQ ID NO: 130 or 153.

Therefore, certain embodiments of the present invention provide polypeptides that are capable of self-assembling into a nanoparticle (i.e., polypeptide monomers) as well as the nucleic acid molecules that encode such polypeptides. An amino acid sequence herein may comprise, or further comprise, a tag (e.g., a purification tag such as a histidine (e.g., 6xHis tag), enterokinase tag, or myc tag), and a linker between the polypeptide monomer and the one or more molecule (e.g. antigen) being carried by the nanoparticle.

29 Aug 2024
2021392894

Further, a nucleic acid sequence herein may encode an amino acid sequence that comprises a tag and/or a linker.

Alternatively, or additionally, the polypeptide includes a phenylalanine (Phe, F) residue at the N-terminus of the FimH polypeptide. Alternatively or additionally, when the polypeptide comprises a nanoparticle domain at the C-terminus or at the N-terminus, the polypeptide includes a phenylalanine (Phe, F) or an aspartic acid (Asp, D) residue at the N-terminus of the mature polypeptide, i.e., after cleavage or removal of a leader sequence, if present. The presence of an aspartic acid (Asp, D) residue at the N-terminus of the mature polypeptide, which comprises a nanoparticle domain at the C-terminus or at the N-terminus, is associated with an improved secretion of the polypeptide when expressed by a mammalian host cell.

Alternatively or additionally, the polypeptide comprises or consists of an amino acid sequence corresponding to:

(a) SEQ ID NO: 20, or a variant and/or fragment thereof,
(b) SEQ ID NO: 21, or a variant and/or fragment thereof,
(c) SEQ ID NO: 22, or a variant and/or fragment thereof,
(d) SEQ ID NO: 23, or a variant and/or fragment thereof,
(e) SEQ ID NO: 24, or a variant and/or fragment thereof,
(f) SEQ ID NO: 25, or a variant and/or fragment thereof,
(g) SEQ ID NO: 26, or a variant and/or fragment thereof,
(h) SEQ ID NO: 27, or a variant and/or fragment thereof,
(i) SEQ ID NO: 28, or a variant and/or fragment thereof,
(j) SEQ ID NO: 29, or a variant and/or fragment thereof,
(k) SEQ ID NO: 30, or a variant and/or fragment thereof,
(l) SEQ ID NO: 31, or a variant and/or fragment thereof,
(m) SEQ ID NO: 32, or a variant and/or fragment thereof,
(n) SEQ ID NO: 33, or a variant and/or fragment thereof,
(o) SEQ ID NO: 34, or a variant and/or fragment thereof,
(p) SEQ ID NO: 35, or a variant and/or fragment thereof,
(q) SEQ ID NO: 36, or a variant and/or fragment thereof,
(r) SEQ ID NO: 37, or a variant and/or fragment thereof,
(s) SEQ ID NO: 38, or a variant and/or fragment thereof,
(t) SEQ ID NO: 39, or a variant and/or fragment thereof,

2021392894 29 Aug 2024

(u) SEQ ID NO: 40, or a variant and/or fragment thereof,

(v) SEQ ID NO: 41, or a variant and/or fragment thereof,

(w) SEQ ID NO: 42, or a variant and/or fragment thereof,

(x) SEQ ID NO: 43, or a variant and/or fragment thereof,

(y) SEQ ID NO: 44, or a variant and/or fragment thereof,

(z) SEQ ID NO: 79, or a variant and/or fragment thereof,

(aa) SEQ ID NO: 80, or a variant and/or fragment thereof,

(bb) SEQ ID NO: 81, or a variant and/or fragment thereof,

(cc) SEQ ID NO: 82, or a variant and/or fragment thereof,

(dd) SEQ ID NO: 83, or a variant and/or fragment thereof,

(ee) SEQ ID NO: 84, or a variant and/or fragment thereof,

(ff) SEQ ID NO: 85, or a variant and/or fragment thereof,

(gg) SEQ ID NO: 86, or a variant and/or fragment thereof,

(hh) SEQ ID NO: 87, or a variant and/or fragment thereof,

(ii) SEQ ID NO: 88, or a variant and/or fragment thereof,

(jj) SEQ ID NO: 89, or a variant and/or fragment thereof,

(kk) any one of SEQ ID NO: 120-124, SEQ ID NO: 129-143 and 153 or a variant and/or a fragment thereof.

In one preferred embodiment, the polypeptide comprises or consists of an amino acid sequence corresponding to SEQ ID NO: 123 or SEQ ID NO:124. In one embodiment, the polypeptide comprises or consists of an amino acid sequence with at least 70% sequence identity to SEQ ID NO: 123 or SEQ ID NO:124, for example, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% sequence identity to SEQ ID NO: 123 or SEQ ID NO:124.

Alternatively or additionally, the mannose binding of the polypeptide is at least 20% lower than that of native FimH complexed with native FimC (FimHC complex), e.g., at least 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100% lower.

Mannose binding can be determined using any suitable means known in the art, for example, surface plasmon resonance may be used to verify binding, binding specificity and binding constants of FimH constructs with Man-BSA and Glc-BSA (negative control), see, for example *Rabani et al.*, 2018, 'Conformational switch of the bacterial adhesin FimH in the absence of the regulatory domain: Engineering a minimalistic allosteric system' *J. Biol. Chem.*, 293(5):1835-1849, which is incorporated by reference herein.

By '**native FimH**' we mean or include wild-type FimH, in particular, wild-type FimH from which domain (a) of the polypeptide of the invention was derived (optionally, with the native N-terminal secretory sequence

2021392894 29 Aug 2024

removed). Alternatively or additionally, we mean or include *E. coli* J96 FimH (e.g., SEQ ID NO: 1 or SEQ ID NO: 2), FimH of *E. coli* UPEC 536 (e.g., SEQ ID NO: 100 or SEQ ID NO: 101), FimH of *E. coli* CFT073 (e.g., SEQ ID NO: 102 or SEQ ID NO: 103), FimH of *E. coli* 789 (e.g., SEQ ID NO: 104 or SEQ ID NO: 105), FimH of *E. coli* IHE3034 (e.g., SEQ ID NO: 106 or SEQ ID NO: 107). In particular, we include FimH in the high-affinity conformation, relaxed (R) state (see above).

By '**native FimC'** we mean or include wild-type FimC (optionally, with the native N-terminal secretory sequence removed). Alternatively or additionally, we mean or include *E. coli* J96 FimC, FimC of UPEC 536 , FimC of *E. coli* CFT073, FimC of *E. coli* 789, FimC of *E. coli* IHE3034.

By '**FimH complexed with native FimC'** and '**FimHC complex**' we mean or include FimH bound to FimC as seen in the periplasm of bacteria naturally expressing FimH and FimC, in the manner and/or conditions taught in the present examples section, in the manner and/or conditions taught in (a) D. Choudhury, X-ray structure of the FimC-FimH chaperone-adhesin complex from uropathogenic Escherichia coli. Science 285, 1061–1066 (1999), (b) C.-S. Hung, Structural basis of tropism of Escherichia coli to the bladder during urinary tract infection. Mol. Microbiol. 44, 903–915 (2002), (c) I. Le Trong, Structural basis for mechanical force regulation of the adhesin FimH via finger trap-like β sheet twisting. Cell 141, 645–655 (2010), (d) G. Phan, Crystal structure of the FimD usher bound to its cognate FimC–FimH substrate. Nature 474, 49–53 (2011), or (e) S. Geibel, Structural and energetic basis of folded-protein transport by the FimD usher. Nature 496, 243–246 (2013).

Alternatively or additionally, the anti-FimH immunogenicity of the polypeptide is at least 20% higher than that of native FimH complexed with native FimC (in particular, we include FimH in the high-affinity conformation, relaxed (R) state (see above).), e.g., at least 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 150%, 200%, 300%, 400% or 500% higher. Immunogenicity can be determined by any suitable means known in the art for example, ELISA or Luminex (see Examples section).

Alternatively or additionally, the auto-aggregation induced by the polypeptide is at least 20% lower than that of native FimH, e.g., at least 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100% lower.

By '**the auto-aggregation induced by the polypeptide is at least X% lower than that of native FimH'** (wherein 'X' is a number between 20 and 100) we mean or include that the polypeptide, when expressed by bacteria instead of native FimH, induces at least X% less bacterial auto-aggregation than otherwise equivalent bacteria expressing the equivalent native FimH. By '**equivalent native FimH'** we mean or include the FimH native to the bacteria being used in the test, the native FimH from which the polypeptide of the invention was derived, and/or the native FimH with which the polypeptide of the invention has the highest sequence identity with. Any suitable means known in the art for determining auto-aggregation may be used but in one embodiment, the method used is that described in Schembri, Christiansen and Klemm, 2001, 'FimH-mediated autoaggregation of Escherichia coli' *Molecular Microbiology*, 41(6), 1419–1430, which is incorporated by

20213928894 29 Aug 2024

reference herein; or Thomas *et al*., 2002, 'Bacterial adhesion to target cells enhanced by shear force' *Cell*, 109(7):913-23, which is incorporated by reference herein; or Hartman *et al*., 2012, 'Inhibition of bacterial adhesion to live human cells: Activity and cytotoxicity of synthetic mannosides' *FEBS Letters*, 586(10): 1459-1465, which is incorporated by reference herein; or Falk *et al*., 1995, 'Chapter 9: Bacterial Adhesion and Colonization Assays' *Meth. Cell, Biol.*, 45:165-192, which is incorporated by reference herein; or Zanaboni *et al*., 2016, 'A novel high-throughput assay to quantify the vaccine-induced inhibition of Bordetella pertussis adhesion to airway epithelia' *BMC Microbiol.*, 16:a215, which is incorporated by reference herein. Alternatively or additionally, bacterial adhesion is (in brief) measured with the BAI assay as follows: UPEC strains engineered to express the mCherry fluorescent marker, are incubated for 30 minutes with monolayers of SV-HUC-1 in 96 well plates in the presence of specific sera against FimH derivatives or positive/negative controls. After adhesion, cells are washed extensively to remove unbound bacteria and fixed with formaldehyde. Finally, the specific fluorescent signal associated with the adhered bacteria is recorded by the use of an automated high content screening microscope (Opera Phenix) and quantified with the Harmony software.

Alternatively or additionally, the polypeptide is capable of inhibiting bacterial adhesion by at least 20%, e.g., by at least 30%, 40%, 50%, 60%, 70%, 80%, 90%, or by 100%.

By '**inhibiting bacterial adhesion'** we mean or include adhesion measured by proxy via bacterial motility or via the bacterial adhesion assay(s) described above (for example with the BAI assay) and/or in the present Examples section.

Alternatively or additionally, the polypeptide is capable of inhibiting hemagglutination of guinea pig red blood cells by at least 2-fold, e.g. by at least 3-fold, 4-fold, 5-fold, 10-fold, 15-fold, 20-fold, 25-fold, 30-fold, 40-fold, 50-fold, 60-fold, 70-fold, 80-fold, 90-fold, or 100-fold.

By '**inhibiting hemagglutination'** we mean or include inhibition of hemagglutination as measured by the hemagglutination inhibition assay (HAI) described in Hultgren et al, Infect Immun 1986, 54, 613-620 and Jarvis C et al, ChemMedChem 2016, 11, 367-373 and/or in the Examples section.

Alternatively or additionally, the polypeptide is soluble by which we mean or include that at least 50% of the polypeptide w/w (e.g., present in a mixture and/or expressed by the/a cell) is in soluble form, for example at least 60%, 70%, 80&, 90%, 95% or 100% of the polypeptide is in soluble form.

A second aspect of the invention provides a nucleic acid encoding a polypeptide according to the first aspect, for example, DNA or RNA.

Alternatively or additionally, the nucleic acid has been codon optimised for expression in a selected prokaryotic or eukaryotic cell, for example, a yeast cell (e.g., *Saccharomyces cerevisiae*, *Pichia pastoris*), an

29 Aug 2024

2021392894

insect cell (e.g., *Spodoptera frugiperda* Sf21 cells, or Sf9 cells), or a mammalian cell (Expi293, Expi293GNTI (Life Technologies), Chinese hamster ovary (CHO) cell, and Human embryonic kidney 293 cells (HEK 293)). By "**codon optimized**" is intended modification with respect to codon usage that may increase translation efficacy and/or half-life of the nucleic acid. Codon usage/optimization tables for many organisms are well known and publicly available (as provided by, e.g., Athey et al. 2017 BMC Bioinf. 18:391). Codon optimisation can be performed using any suitable means known in the art, for example, the method operated by GeneArt.

Alternatively or additionally, the nucleic comprises or consists of a nucleic acid sequence corresponding to:

(1) SEQ ID NO: 45, or a variant and/or fragment thereof,

(2) SEQ ID NO: 46, or a variant and/or fragment thereof,

(3) SEQ ID NO: 47, or a variant and/or fragment thereof,

(4) SEQ ID NO: 48, or a variant and/or fragment thereof,

(5) SEQ ID NO: 49, or a variant and/or fragment thereof,

(6) SEQ ID NO: 50, or a variant and/or fragment thereof,

(7) SEQ ID NO: 51, or a variant and/or fragment thereof,

(8) SEQ ID NO: 52, or a variant and/or fragment thereof,

(9) SEQ ID NO: 53, or a variant and/or fragment thereof,

(10) SEQ ID NO: 54, or a variant and/or fragment thereof,

(11) SEQ ID NO: 55, or a variant and/or fragment thereof,

(12) SEQ ID NO: 56, or a variant and/or fragment thereof,

(13) SEQ ID NO: 57, or a variant and/or fragment thereof,

(14) SEQ ID NO: 58, or a variant and/or fragment thereof,

(15) SEQ ID NO: 59, or a variant and/or fragment thereof,

(16) SEQ ID NO: 60, or a variant and/or fragment thereof,

(17) SEQ ID NO: 61, or a variant and/or fragment thereof,

(18) SEQ ID NO: 62, or a variant and/or fragment thereof,

(19) SEQ ID NO: 63, or a variant and/or fragment thereof,

(20) SEQ ID NO: 64, or a variant and/or fragment thereof,

(21) SEQ ID NO: 65, or a variant and/or fragment thereof,

(22) SEQ ID NO: 66, or a variant and/or fragment thereof,

(23) SEQ ID NO: 67, or a variant and/or fragment thereof,

(24) SEQ ID NO: 68, or a variant and/or fragment thereof,

(25) SEQ ID NO: 69, or a variant and/or fragment thereof,

(26) SEQ ID NO: 70, or a variant and/or fragment thereof,

2021392894 29 Aug 2024

(27) SEQ ID NO: 71, or a variant and/or fragment thereof,

(28) SEQ ID NO: 72, or a variant and/or fragment thereof,

(29) SEQ ID NO: 73, or a variant and/or fragment thereof,

(30) SEQ ID NO: 74, or a variant and/or fragment thereof,

(31) SEQ ID NO: 75, or a variant and/or fragment thereof,

(32) SEQ ID NO: 76, or a variant and/or fragment thereof,

(33) SEQ ID NO: 77, or a variant and/or fragment thereof,

(34) SEQ ID NO: 90, or a variant and/or fragment thereof,

(35) SEQ ID NO: 91, or a variant and/or fragment thereof,

(36) SEQ ID NO: 92, or a variant and/or fragment thereof,

(37) SEQ ID NO: 93, or a variant and/or fragment thereof,

(38) SEQ ID NO: 94, or a variant and/or fragment thereof,

(39) SEQ ID NO: 95, or a variant and/or fragment thereof,

(40) SEQ ID NO: 96, or a variant and/or fragment thereof,

(41) SEQ ID NO: 97, or a variant and/or fragment thereof,

(42) SEQ ID NO: 98, or a variant and/or fragment thereof, and

(43) SEQ ID NO: 99, or a variant and/or fragment thereof.

The skilled person will immediately appreciate that, where the nucleic acid of the invention is an RNA, T is replaced with U in the nucleic acid sequences of the invention (e.g., SEQ ID NOs: 45-99).

A third aspect of the invention provides a vector comprising the nucleic acid of the second aspect. Alternatively or additionally, the vector is a plasmid, for example, an expression plasmid. Alternatively or additionally, the plasmid is selected from the group consisting of pCDNA3.1 (Life Technologies), pCDNA3.4 (Life Technologies), pFUSE, pBROAD, pSEC, pCMV, pDSG-IBA, and pHEK293 Ultra, and the like. Alternatively or additionally, the plasmid is suitable for expression in bacterial host cells and in selected from the group consisting of pACYCDuet-1, pTrcHis2A, pET21, pET15TEV, pET22b+, pET303/CT-HIS, PET303/CT, pBAD/Myc-His A, pET303, pET24b(+), and the like.

Alternatively or additionally, the vector is a viral vector, for example, an RNA viral vector. Alternatively or additionally, the viral vector is selected from the group consisting of Adenoviral vectors, and CHAD.

A fourth aspect of the invention provides a cell, for example a host cell, comprising a nucleic acid of the second or a vector of the fourth aspect.

2021392894 29 Aug 2024

Suitable mammalian host cells are known in the art. Alternatively or additionally, the cell does not have N-acetylglucosaminyltransferase I (GnTI) activity. Alternatively or additionally, the cell is selected from the group consisting of Expi293, Expi293GNTI (Life Technologies), Chinese hamster ovary (CHO) cell, NIH-3T3 cells, 293-T cells, Vero cells, HeLa cells, PERC.6 cells (ECACC deposit number 96022940), Hep G2 cells, MRC-5 (ATCC CCL-171), WI-38 (ATCC CCL-75), fetal rhesus lung cells (ATCC CL-160), Madin-Darby bovine kidney ("MDBK") cells, Madin-Darby canine kidney ("MDCK") cells (e.g., MDCK (NBL2), ATCC CCL34; or MDCK 33016, DSM ACC 2219), baby hamster kidney (BHK) cells, such as BHK21-F, HKCC cells, Human embryonic kidney 293 cells (HEK 293), and the like.

Suitable bacterial host cells are known in the art. Exemplary bacterial host cells include any of the following and derivatives thereof: *Escherichia coli* from strains BL21(DE3), HMS174 (DE3), Origami 2 (DE3), BL21DE3T1r or T7shuffle express.

A fifth aspect of the invention provides a method of producing a polypeptide defined in the first aspect by expressing the protein in a cell as defined in the fourth aspect.

A sixth aspect of the invention provides a vaccine comprising the polypeptide defined in the first aspect, a nucleic acid defined in the second aspect, and/or a vector defined in the third aspect. Alternatively or additionally, the vaccine comprises an adjuvant.

In one embodiment, the vaccine of the invention comprises the polypeptide defined in the first aspect and an adjuvant comprising any one of: 3D-MPL, QS21 and liposomes, for example liposomes comprising cholesterol. In one embodiment, the vaccine of the invention comprises the polypeptide defined in the first aspect and an adjuvant comprising 3D-MPL, QS21 and liposomes comprising cholesterol.

The inventors have surprisingly found that vaccines comprising an adjuvant comprising 3D-MPL, QS21 and liposomes comprising cholesterol, such as the AS01 adjuvant, may elicit an improved immune response. By "**improved immune response**" we mean or include an increased level of immunoglobulin G (IgG) in the serum and/or in the urine of an animal, such as a mice, immunized with said vaccine respective to the level of IgG in the serum and/or in the urine of an animal, such as a mice, immunized with a reference or vaccine. For "increased level of IgG in the serum and/or in the urine" we mean or include cells by at least 2-fold, e.g. by at least 3-fold, 4-fold, 5-fold, 10-fold, 15-fold, 20-fold, 25-fold, 30-fold, 40-fold, or 50-fold. Said reference or control vaccine does not comprise an adjuvant comprising 3D-MPL, QS21 and liposomes comprising cholesterol; for example, said reference or control vaccine comprises the PHAD adjuvant.

2021392894 29 Aug 2024

The inventors have also surprisingly found that vaccines comprising the polypeptide defined in the first aspect and an adjuvant comprising 3D-MPL, QS21 and liposomes comprising cholesterol, such as the AS01 adjuvant, are capable of eliciting a protective immune response after one or two doses.

Immunogenic compositions (e.g., vaccines) will be pharmaceutically acceptable. They will usually include components in addition to the antigens e.g. they typically include one or more pharmaceutical carrier(s), excipient(s) and/or adjuvant(s). A thorough discussion of carriers and excipients is available in Current Protocols in Molecular Biology (F.M. Ausubel et al., eds., 1987) Supplement 30, which is incorporated by reference herein. Thorough discussions of vaccine adjuvants are available in Vaccine Design: The Subunit and Adjuvant Approach (eds. Powell & Newman) Plenum Press 1995 (ISBN 0-306-44867-X); and Vaccine Adjuvants: Preparation Methods and Research Protocols (Volume 42 of Methods in Molecular Medicine series), ISBN: 1-59259-083-7. Ed. O'Hagan which are incorporated by reference herein.

Compositions will generally be administered to a mammal in aqueous form. Prior to administration, however, the composition may have been in a non-aqueous form. For instance, although some vaccines are manufactured in aqueous form, then filled and distributed and administered also in aqueous form, other vaccines are lyophilized during manufacture and are reconstituted into an aqueous form at the time of use. Thus, a composition of the invention may be dried, such as a lyophilized formulation. The composition may include preservatives such as thiomersal or 2-phenoxyethanol. It is preferred, however, that the vaccine should be substantially free from (i.e. less than 5μg/ml) mercurial material e.g. thiomersal-free. Vaccines containing no mercury are more preferred. Preservative-free vaccines are particularly preferred. To improve thermal stability, a composition may include a temperature protective agent.

To control tonicity, it is preferred to include a physiological salt, such as a sodium salt. Sodium chloride (NaCl) is preferred, which may be present at between 1 and 20 mg/ml e.g. about 10±2mg/ml NaCl. Other salts that may be present include potassium chloride, potassium dihydrogen phosphate, disodium phosphate dehydrate, magnesium chloride, calcium chloride, etc.

Compositions will generally have an osmolality of between 200 mOsm/kg and 400 mOsm/kg, preferably between 240-360 mOsm/kg, and will more preferably fall within the range of 290-310 mOsm/kg.

Compositions may include one or more buffers. Typical buffers include: a phosphate buffer; a Tris buffer; a borate buffer; a succinate buffer; a histidine buffer (particularly with an aluminium hydroxide adjuvant); or a citrate buffer. Buffers will typically be included in the 5-20mM range.

The pH of a composition will generally be between 5.0 and 8.1, and more typically between 6.0 and 8.0 e.g., 6.5 and 7.5, or between 7.0 and 7.8.

2021392894 29 Aug 2024

The composition is preferably sterile. The composition is preferably non-pyrogenic e.g. containing <1 EU (endotoxin unit, a standard measure) per dose, and preferably <0.1 EU per dose. The composition is preferably gluten free.

The composition may include material for a single immunisation, or may include material for multiple immunizations (i.e. a 'multidose' kit). The inclusion of a preservative is preferred in multidose arrangements. As an alternative (or in addition) to including a preservative in multidose compositions, the compositions may be contained in a container having an aseptic adaptor for removal of material.

Human vaccines are typically administered in a dosage volume of about 0.5ml, although a half dose (i.e. about 0.25ml) may be administered to children.

Immunogenic compositions of the invention may also comprise one or more immunoregulatory agents. Preferably, one or more of the immunoregulatory agents include one or more adjuvants.

Adjuvants

Vaccines and immunogenic compositions of the invention may also comprise an adjuvant in addition to the antigen. Adjuvants are used in vaccines in order to enhance and modulate the immune response to the antigen. The adjuvants described herein may be combined with any of the antigen(s) herein described.

The adjuvant may be any adjuvant known to the skilled person, but adjuvants include (but are not limited to) oil-in-water emulsions (for example MF59 or AS03), liposomes, saponins, TLR2 agonists, TLR3 agonists, TLR4 agonists, TLR5 agonists, TLR6 agonists, TLR7 agonists, TLR8 agonists, TLR9 agonists, aluminium salts, nanoparticles, microparticles, Immune stimulating complexes (ISCOMS), calcium fluoride and organic compound composites or combinations thereof.

Oil-in-water emulsions

In an embodiment of the present invention, there is provided a vaccine or immunogenic composition for use in the invention comprising an oil-in-water emulsion. Oil-in-water emulsions of the present invention comprise a metabolisable oil and an emulsifying agent. In order for any oil-in-water composition to be suitable for human administration, the oil phase of the emulsion system has to comprise a metabolisable oil. The meaning of the term metabolisable oil is well known in the art. Metabolisable can be defined as "being capable of being transformed by metabolism" (Dorland's Illustrated Medical Dictionary, W.B. Sanders Company, 25th edition, 1974). A particularly suitable metabolisable oil is squalene. Squalene (2,6,10,15,19,23-Hexamethyl-2,6,10,14,18,22-tetracosahexaene) is an unsaturated oil which is found in large quantities in shark-liver oil, and in lower quantities in olive oil, wheat germ oil, rice bran oil, and yeast, and is a particularly preferred oil for use in an oil-in-water emulsion of the invention. Squalene is a metabolisable oil by virtue of the fact that it is an intermediate in the biosynthesis of cholesterol (Merck index, 10th Edition, entry no. 8619). In some embodiments, wherein the vaccines or immunogenic compositions of the invention

2021392894 29 Aug 2024

comprise an oil-in-water emulsion, the metabolisable oil is present in the vaccine or in the immunogenic composition in an amount of 0.5% to 10% (v/v) of the total volume of the composition. The oil-in-water emulsion further comprises an emulsifying agent. The emulsifying agent may suitably be polyoxyethylene sorbitan monooleate (POLYSORBATE 80). Further, said emulsifying agent is suitably present in the vaccine or immunogenic composition in an amount of 0.125 to 4% (v/v) of the total volume of the composition. The oil-in-water emulsion may optionally comprise a tocol. Tocols are well known in the art and are described in EP0382271 B1. Suitably, the tocol may be alpha-tocopherol or a derivative thereof such as alpha-tocopherol succinate (also known as vitamin E succinate). Said tocol is suitably present in the adjuvant composition in an amount of 0.25% to 10% (v/v) of the total volume of the immunogenic composition. The oil-in-water emulsion may also optionally comprise sorbitan trioleate (SPAN 85).

The method of producing oil-in-water emulsions is well known to the person skilled in the art. Commonly, the method comprises mixing the oil phase (optionally comprising a tocol) with a surfactant such as a PBS/TWEEN80 solution, followed by homogenisation using a homogenizer; it would be clear to a person skilled in the art that a method comprising passing the mixture twice through a syringe needle is suitable for homogenising small volumes of liquid. Equally, the emulsification process in microfluidiser (e.g., M110S Microfluidics machine, maximum of 50 passes, for a period of 2 minutes at maximum pressure input of 6 bar (output pressure of about 850 bar)) could be adapted by the person skilled in the art to produce smaller or larger volumes of emulsion. The adaptation could be achieved by routine experimentation comprising the measurement of the resultant emulsion until a preparation was achieved with oil droplets of the required diameter.

In an oil-in-water emulsion, the oil and emulsifier should be in an aqueous carrier. The aqueous carrier may be, for example, phosphate buffered saline or citrate. In particular, the oil-in-water emulsion systems used in the present invention have a small oil droplet size in the sub-micron range. Suitably the droplet sizes will be in the range 120 to 750 nm, more particularly sizes from 120 to 600 nm in diameter. Even more particularly, the oil-in water emulsion contains oil droplets of which at least 70% by intensity are less than 500 nm in diameter, more particular at least 80% by intensity are less than 300 nm in diameter, more particular at least 90% by intensity are in the range of 120 to 200 nm in diameter.

The oil droplet size, *i.e.* diameter, according to the present invention is given by intensity. There are several ways of determining the diameter of the oil droplet size by intensity. Intensity is measured by use of a sizing instrument, suitably by dynamic light scattering such as the Malvern Zetasizer 4000 or preferably the Malvern Zetasizer 3000HS. A first possibility is to determine the z average diameter ZAD by dynamic light scattering (PCS-Photon correlation spectroscopy); this method additionally gives the polydispersity index (PDI), and both the ZAD and PDI are calculated with the cumulants algorithm. These values do not require the knowledge of the particle refractive index. A second means is to calculate the diameter of the oil droplet by

29 Aug 2024 2021392894

determining the whole particle size distribution by another algorithm, either the Contin, or NNLS, or the automatic "Malvern" one (the default algorithm provided for by the sizing instrument). Most of the time, as the particle refractive index of a complex composition is unknown, only the intensity distribution is taken into consideration, and if necessary the intensity mean originating from this distribution.

ISCOMs

In some embodiments of the present invention, there are provided vaccines or immunogenic compositions of the invention comprising ISCOMs. ISCOMs are well known in the art (see Kersten & Crommelin, 1995, *Biochimica et Biophysica Acta* 1241: 117-138). ISCOMs comprise a saponin, cholesterol and phospholipids and form an open-cage-like structure of typically about 40 nm in size. ISCOMs result from the interaction of saponins, cholesterol and further phospholipids. A typical reaction mixture for the preparation of ISCOM is 5 mg/ml saponin and 1 mg/ml each for cholesterol and phospholipid. Phospholipids suitable for use in ISCOMs include, but are not limited, to phosphocholine (didecanoyl-L-α-phosphatidylcholine [DDPC], dilauroylphosphatidylcholine [DLPC], dimyristoylphosphatidylcholine [DMPC], dipalmitoyl phosphatidylcholine [DPPC], Distearoyl phosphatidylcholine [DSPC], Dioleoyl phosphatidylcholine [DOPC], 1-palmitoyl, 2-oleoylphosphatidylcholine [POPC], Dielaidoyl phosphatidylcholine [DEPC]), phosphoglycerol (1,2-Dimyristoyl-sn-glycero-3-phosphoglycerol [DMPG], 1,2-dipalmitoyl-sn-glycero-3-phosphoglycerol [DPPG], 1,2-distearoyl-sn-glycero-3-phosphoglycerol [DSPG], 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoglycerol [POPG]), phosphatidic acid (1,2-dimyristoyl-sn-glycero-3-phosphatidic acid [DMPA], dipalmitoyl phosphatidic acid [DPPA], distearoyl-phosphatidic acid [DSPA]), phosphoethanolamine (1,2-dimyristoyl-*sn*-glycero-3-phosphoethanolamine [DMPE], 1,2-Dipalmitoyl-sn-glycero-3-phosphoethanolamine [DPPE], 1,2-distearoyl-sn-glycero-3-phosphoethanolamine DSPE 1,2-Dioleoyl-sn-Glycero-3-Phosphoethanolamine [DOPE]), phoshoserine, polyethylene glycol [PEG] phospholipid (mPEG-phospholipid, polyglycerin-phospholipid, functionalized-phospholipid, terminal activated-phosholipid). In particular embodiments, ISCOMs comprise 1-palmitoyl-2-oleoyl-glycero-3-phosphoethanolamine. In further particular embodiments, highly purified phosphatidylcholine is used and can be selected from the group consisting of: Phosphatidylcholine (from egg), Phosphatidylcholine Hydrogenated (from egg) Phosphatidylcholine (from soy), Phosphatidylcholine Hydrogenated (from soy). In further particular embodiments, ISCOMs comprise phosphatidylethanolamine [POPE] or a derivative thereof. A number of saponins are suitable for use in ISCOMs. The adjuvant and haemolytic activity of individual saponins has been extensively studied in the art. For example, Quil A (derived from the bark of the South American tree Quillaja Saponaria Molina), and fractions thereof, are described in US 5,057,540 and "Saponins as vaccine adjuvants", Kensil, C. R., *Crit. Rev. Ther. Drug. Carrier Syst.*, 1996, 12 (1-2): 1-55; and EP0362279 B1. ISCOMs comprising fractions of Quil A have been used in the manufacture of vaccines (EP0109942 B1). These structures have been reported to have adjuvant activity (EP0109942 B1; WO 96/11711). Fractions of QuilA, derivatives of

2021392894 29 Aug 2024

QuilA and/or combinations thereof are suitable saponin preparations for use in ISCOMs. The haemolytic saponins QS21 and QS17 (HPLC purified fractions of Quil A) have been described as potent adjuvants, and the method of their production is disclosed in US 5,057,540 and EP0362279 B1. Also described in these references is the use of QS7 (a non-haemolytic fraction of Quil-A) which acts as a potent adjuvant for systemic vaccines. Use of QS21 is further described in Kensil *et al.* (1991. J. Immunology vol 146, 431-437). Combinations of QS21 and polysorbate or cyclodextrin are also known (WO 99/10008). Particulate adjuvant systems comprising fractions of QuilA, such as QS21 and QS7 are described in WO 96/33739 and WO 96/11711 and these are incorporated herein. Other particular QuilA fractions designated QH-A, QH-B, QH-C and a mixture of QH-A and QH-C designated QH-703 are disclosed in WO 96/011711 in the form of ISCOMs and are incorporated herein.

Microparticles

In some embodiments of the present invention, there is provided a vaccine or immunogenic composition of the invention comprising microparticles. Microparticles, compositions comprising microparticles, and methods of producing microparticles are well known in the art (see Singh *et al.* [2007 *Expert Rev. Vaccines* 6(5): 797-808] and WO 98/033487). The term "microparticle" as used herein, refers to a particle of about 10 nm to about 10,000 µm in diameter or length, derived from polymeric materials which have a variety of molecular weights and, in the case of the copolymers such as PLG, a variety of lactide:glycolide ratios. In particular, the microparticles will be of a diameter that permits parenteral administration to a subject without occluding the administrating device and/or the subject's capillaries. Microparticles are also known as microspheres. Microparticle size is readily determined by techniques well known in the art, such as photon correlation spectroscopy, laser diffractometry and/or scanning electron microscopy. Microparticles for use herein will be formed from materials that are sterilizable, non-toxic and biodegradable. Such materials include, without limitation, poly(a-hydroxy acid), polyhydroxybutyric acid, polycaprolactone, polyorthoester, polyanhydride.

Liposomes

In some embodiments of the present invention, there is provided a vaccine or immunogenic composition of the invention comprising liposomes. The term "liposomes" generally refers to uni- or multilamellar (particularly 2, 3, 4, 5, 6, 7, 8, 9, or 10 lamellar depending on the number of lipid membranes formed) lipid structures enclosing an aqueous interior. Liposomes and liposome formulations are well known in the art. Lipids, which are capable of forming liposomes, include all substances having fatty or fat-like properties. Lipids which can make up the lipids in the liposomes can be selected from the group comprising of glycerides, glycerophospholipides, glycerophosphinolipids, glycerophosphonolipids, sulfolipids, sphingolipids, phospholipids, isoprenolides, steroids, stearines, sterols, archeolipids, synthetic cationic lipids and carbohydrate containing lipids. Liposome size may vary from 30 nm to several µm depending on the

2021392894 29 Aug 2024

phospholipid composition and the method used for their preparation. In particular embodiments of the invention, the liposome size will be in the range of 50 nm to 500 nm, and in further embodiments, 50 nm to 200 nm. Dynamic laser light scattering is a method used to measure the size of liposomes well known to those skilled in the art. The liposomes suitably contain a neutral lipid, for example phosphatidylcholine, which is suitably non-crystalline at room temperature, for example egg yolk phosphatidylcholine, dioleoyl phosphatidylcholine (DOPC) or dilauryl phosphatidylcholine. In a particular embodiment, the liposomes of the present invention contain DOPC. The liposomes may also contain a charged lipid which increases the stability of the liposome-saponin structure for liposomes composed of saturated lipids. In these cases, the amount of charged lipid is suitably 1 to 20% (w/w), preferably 5 to 10%. The ratio of sterol to phospholipid is 1 to 50% (mol/mol), suitably 20 to 25% (mol/mol).

Saponins

In some embodiments of the invention, the vaccine or immunogenic composition of the invention comprises a saponin. A particularly suitable saponin for use in the present invention is Quil A and its derivatives. Quil A is a saponin preparation isolated from the South American tree *Quillaja Saponaria Molina* and was first described by Dalsgaard *et al.* in 1974 ("Saponin adjuvants", *Archiv. für die gesamte Virusforschung*, Vol. 44, Springer Verlag, Berlin, p243-254) to have adjuvant activity. Purified fragments of Quil A have been isolated by HPLC which retain adjuvant activity without the toxicity associated with Quil A (EP0362278), for example QS7 and QS21 (also known as QA7 and QA21). QS-21 is a natural saponin derived from the bark of *Quillaja saponaria* Molina, which induces CD8+ cytotoxic T cells (CTLs), Th1 cells and a predominant IgG2a antibody response and is a particular saponin in the context of the present invention. The saponin adjuvant within the immunogenic compositions of the invention in particular are immunologically active fractions of Quil A, such as QS-7 or QS-21, suitably QS-21. In particular embodiments, the vaccines and/or immunogenic compositions of the invention contain the immunologically active saponin fraction in substantially pure form. In particular, the vaccines or immunogenic compositions of the invention contain QS21 in substantially pure form, that is to say, the QS21 is at least 75%, 80%, 85%, 90% pure, for example at least 95% pure, or at least 98% pure.

In a particular embodiment, QS21 is provided with an exogenous sterol, such as cholesterol for example. Suitable sterols include β-sitosterol, stigmasterol, ergosterol, ergocalciferol and cholesterol. In a further particular embodiment, the adjuvant composition comprises cholesterol as sterol. These sterols are well known in the art, for example cholesterol is disclosed in the Merck Index, 11th Edition, page 341, as a naturally occurring sterol found in animal fat.

In one embodiment, the liposomes of the invention that comprise a saponin suitably contain a neutral lipid, for example phosphatidylcholine, which is suitably non-crystalline at room temperature, for example egg yolk phosphatidylcholine, dioleoyl phosphatidylcholine (DOPC) or dilauryl phosphatidylcholine. The liposomes may also contain a charged lipid which increases the stability of the liposome-QS21 structure for

2021392894 29 Aug 2024

liposomes composed of saturated lipids. In these cases the amount of charged lipid is suitably 1 to 20% (w/w), particularly 5 to 10% (w/w). The ratio of sterol to phospholipid is 1 to 50% (mol/mol), suitably 20 to 25% (mol/mol).

Where the active saponin fraction is QS21, the ratio of QS21:sterol will typically be in the order of 1:100 to 1:1 (w/w), suitably between 1:10 to 1:1 (w/w), and preferably 1:5 to 1:1 (w/w). Suitably, excess sterol is present, the ratio of QS21:sterol being at least 1:2 (w/w). In one embodiment, the ratio of QS21:sterol is 1:5 (w/w). The sterol is suitably cholesterol.

Other useful saponins are derived from the plants *Aesculus hippocastanum* or *Gyophilla struthium.* Other saponins which have been described in the literature include Escin, which has been described in the Merck index (12th Edition: entry 3737) as a mixture of saponins occurring in the seed of the horse chestnut tree, Lat: *Aesculus hippocastanum*. Its isolation is described by chromatography and purification (Fiedler, *Arzneimittel-Forsch*. 4, 213 (1953)), and by ion-exchange resins (Erbring *et al*., US 3,238,190). Fractions of Escin have been purified and shown to be biologically active (Yoshikawa *et al.,* 1996, *Chem Pharm Bull* (Tokyo), 44(8): 1454-1464). Sapoalbin from *Gypsophilla struthium* (R. Vochten *et al*., 1968, *J. Pharm. Belg*. 42: p213-226) has also been described in relation to ISCOM production for example.

A saponin, such as QS21, can be used at amounts between 1 and 100µg per human dose of the adjuvant composition. QS21 may be used at a level of about 50µg, for example between 40 to 60 µg, suitably between 45 to 55 µg or between 49 and 51 µg or 50µg. In a further embodiment, the human dose of the adjuvant composition comprises QS21 at a level of about 25µg, for example between 20 to 30µg, suitably between 21 to 29µg or between 22 to 28µg or between 28 and 27µg or between 24 and 26µg, or 25µg.

TLR4 agonist

In some embodiments, the vaccine or immunogenic composition of the invention comprises a TLR4 agonist. By "TLR agonist" it is meant a component which is capable of causing a signaling response through a TLR signaling pathway, either as a direct ligand or indirectly through generation of endogenous or exogenous ligand (Sabroe *et al*, 2003, *JI* p1630-5). A TLR4 agonist is capable of causing a signaling response through a TLR-4 signaling pathway. A suitable example of a TLR-4 agonist is a lipopolysaccharide, suitably a non-toxic derivative of lipid A, particularly monophosphoryl lipid A or more particularly 3-Deacylated monophoshoryl lipid A (3D – MPL).

3D-MPL is sold under the name MPL by GlaxoSmithKline Biologicals and is referred throughout the document as MPL or 3D-MPL. See, for example, US 4,436,727; US 4,877,611; US 4,866,034 and US 4,912,094. 3D-MPL primarily promotes CD4+ T cell responses with an IFN-gamma (Th1) phenotype. 3D-MPL can be produced according to the methods disclosed in GB 2 220 211 A. Chemically, it is a mixture of 3-deacylated monophosphoryl lipid A with 4, 5 or 6 acylated chains. In the compositions of the present invention, small

2021392894 29 Aug 2024

particle 3D-MPL may be used to prepare the aqueous adjuvant composition. Small particle 3D-MPL has a particle size such that it may be sterile-filtered through a 0.22μm filter. Such preparations are described in WO 94/21292. Preferably, powdered 3D-MPL is used to prepare the aqueous adjuvant compositions of the present invention.

Other TLR-4 agonists which can be used are alkyl glucosaminide phosphates (AGPs) such as those disclosed in WO 98/50399 or US 6,303, 347 (processes for preparation of AGPs are also disclosed), suitably RC527 or RC529 or pharmaceutically acceptable salts of AGPs as disclosed in US 6,764,840.

Other suitable TLR-4 agonists are as described in WO 03/011223 and in WO 03/099195, such as compound I, compound II and compound III disclosed on pages 4-5 of WO 03/011223 or on pages 3 to 4 of WO 03/099195 and in particular those compounds disclosed in WO 03/011223, as ER803022, ER803058, ER803732, ER804053, ER804057m ER804058, ER804059, ER804442, ER804680 and ER804764. For example, one suitable TLR-4 agonist is ER804057.

A TLR-4 agonist, such as a lipopolysaccharide, such as 3D-MPL, can be used at amounts between 1 and 100 μg per human dose of the adjuvant composition. 3D-MPL may be used at a level of about 50 μg, for example between 40 to 60 μg, suitably between 45 to 55 μg or between 49 to 51 μg or 50 μg per human dose. In a further embodiment, the human dose of the adjuvant composition comprises 3D-MPL at a level of about 25 μg, for example between 20 to 30 μg, suitably between 21 to 29 μg or between 22 to 28 μg or between 28 to 27 μg or between 24 to 26 μg, or 25 μg.

Synthetic derivatives of lipid A are known and thought to be TLR 4 agonists including, but not limited to:

OM174 (2-deoxy-6-o-[2-deoxy-2-[(R)-3-dodecanoyloxytetra-decanoylamino]-4-o-phosphono-β-D-glucopyranosyl]-2-[(R)-3-hydroxytetradecanoylamino]-α-D-glucopyranosyldihydrogenphosphate), (WO 95/14026)

OM294 DP (3S, 9 R) –3--[(R)-dodecanoyloxytetradecanoylamino]-4-oxo-5-aza-9(R)-[(R)-3-hydroxytetradecanoylamino]decan-1,10-diol,1,10-bis(dihydrogenophosphate) (WO 99/64301 and WO 00/0462)

OM197 MP-Ac DP (3S-, 9R) -3-[(R)-dodecanoyloxytetradecanoylamino]-4-oxo-5-aza-9-[(R)-3-hydroxytetradecanoylamino]decan-1,10-diol,1-dihydrogenophosphate 10-(6-aminohexanoate) (WO 01/46127).

PHAD (phosphorylated hexa-acyl disaccharide).

Other suitable TLR-4 ligands, capable of causing a signalling response through TLR-4 (Sabroe *et al*, JI 2003 p1630-5) are, for example, lipopolysaccharide from gram-negative bacteria and its derivatives, or fragments thereof, in particular a non-toxic derivative of LPS (such as 3D-MPL). Other suitable TLR agonist are: heat

2021392894 29 Aug 2024

shock protein (HSP) 10, 60, 65, 70, 75 or 90; surfactant Protein A, hyaluronan oligosaccharides, heparan sulphate fragments, fibronectin fragments, fibrinogen peptides and b-defensin-2, muramyl dipeptide (MDP) or F protein of respiratory syncytial virus (RSV). In one embodiment, the TLR agonist is HSP 60, 70 or 90.

TLR agonists

Rather than a TLR4 agonist, other natural or synthetic agonists of TLR molecules may be used in vaccines or immunogenic composition of the invention. These include, but are not limited to, agonists for TLR2, TLR3, TLR5, TLR6, TLR7, TLR8 and TLR9.

In one embodiment of the present invention, a TLR agonist is used that is capable of causing a signalling response through TLR-1 (Sabroe *et al*, JI 2003 p1630-5). Suitably, the TLR agonist capable of causing a signalling response through TLR-1 is selected from: Tri-acylated lipopeptides (LPs); phenol-soluble modulin; Mycobacterium tuberculosis LP; S-(2,3-bis(palmitoyloxy)-(2-RS)-propyl)-N-palmitoyl-(R)-Cys-(S)-Ser-(S)-Lys(4)-OH, trihydrochloride (Pam3Cys) LP which mimics the acetylated amino terminus of a bacterial lipoprotein and OspA LP from *Borrelia burgdorferi*.

In a further embodiment, a TLR agonist is used that is capable of causing a signalling response through TLR-2 (Sabroe *et al*, JI 2003 p1630-5). Suitably, the TLR agonist capable of causing a signalling response through TLR-2 is one or more of a lipoprotein, a peptidoglycan, a bacterial lipopeptide from *M. tuberculosis*, *B. burgdorferi*, *T. pallidum*, peptidoglycans from species including *Staphylococcus aureus*, lipoteichoic acids, mannuronic acids, Neisseria porins, bacterial fimbriae, Yersinia virulence factors, CMV virions, measles haemagglutinin, and zymosan from yeast.

In a further embodiment, a TLR agonist is used that is capable of causing a signalling response through TLR-3 (Sabroe *et al*, JI 2003 p1630-5). Suitably, the TLR agonist capable of causing a signalling response through TLR-3 is double stranded RNA (dsRNA), or polyinosinic-polycytidylic acid (Poly IC), a molecular nucleic acid pattern associated with viral infection.

In an alternative embodiment, a TLR agonist is used that is capable of causing a signalling response through TLR-5 (Sabroe *et al*, JI 2003 p1630-5). Suitably, the TLR agonist capable of causing a signalling response through TLR-5 is bacterial flagellin. Said TLR-5 agonist may be flagellin or may be a fragment of flagellin which retains TLR-5 agonist activity. The flagellin can include a polypeptide selected from the group consisting of *H. pylori, S. typhimurium, V. cholera, S. marcescens, S. flexneri, T. pallidum, L. pneumophilia, B. burgdorferi; C. difficile, R. meliloti, A. tumefaciens; R. lupine; B. clarridgeiae, P. mirabilis, B. subtilis, L. moncytogenes, P. aeruginosa* and *E. coli.*

In a particular embodiment, the flagellin is selected from the group consisting of *S. typhimurium* flagellin B (Genbank Accession number AF045151), a fragment of *S. typhimurium* flagellin B, *E. coli* FliC. (Genbank

29 Aug 2024 2021392894

Accession number AB028476); fragment of *E. coli* FliC; *S. typhimurium* flagellin FliC (ATCC14028) and a fragment of *S. typhimurium* flagellin FliC

In a further particular embodiment, said TLR-5 agonist is a truncated flagellin, as described in WO 09/156405 *i.e.* one in which the hypervariable domain has been deleted. In one aspect of this embodiment, said TLR-5 agonist is selected from the group consisting of: $FliC_{\Delta 174-400}$; $FliC_{\Delta 161-405}$ and $FliC_{\Delta 138-405}$.

In a further particular embodiment, said TLR-5 agonist is a flagellin, as described in WO 09/128950. In a further embodiment, a TLR agonist is used that is capable of causing a signalling response through TLR-6 (Sabroe *et al*, JI 2003 p1630-5). Suitably, the TLR agonist capable of causing a signalling response through TLR-6 is mycobacterial lipoprotein, di-acylated LP, and phenol-soluble modulin. Further TLR6 agonists are described in WO 03/043572.

In a further embodiment, a TLR agonist is used that is capable of causing a signalling response through TLR-7 (Sabroe *et al*, JI 2003 p1630-5). Suitably, the TLR agonist capable of causing a signalling response through TLR-7 is a single stranded RNA (ssRNA), loxoribine, a guanosine analogue at positions N7 and C8, or an imidazoquinoline compound, or derivative thereof. In a particular embodiment, the TLR agonist is imiquimod. Further TLR7 agonists are described in WO 02/085905.

In a further embodiment, a TLR agonist is used that is capable of causing a signalling response through TLR-8 (Sabroe *et al*, JI 2003 p1630-5). Suitably, the TLR agonist capable of causing a signalling response through TLR-8 is a single stranded RNA (ssRNA), an imidazoquinoline molecule with anti-viral activity, for example resiquimod (R848); resiquimod is also capable of recognition by TLR-7. Other TLR-8 agonists which may be used include those described in WO 04/071459.

In a further embodiment, a TLR agonist is used that is capable of causing a signalling response, such as one that comprises a CpG motif. The term "immunostimulatory oligonucleotide" is used herein to mean an oligonucleotide that is capable of activating a component of the immune system. In one embodiment, the immunostimulatory oligonucleotide comprises one or more unmethylated cytosine-guanosine (CpG) motifs. In a further embodiment, the immunostimulatory oligonucleotide comprises one or more unmethylated thymidine-guanosine (TG) motif or may be T-rich. By T-rich, it is meant that the nucleotide composition of the oligonucleotide comprises greater than 50, 60, 70 or 80% thymidine. In one embodiment, the oligonucleotide is not an immunostimulatory oligonucleotide and does not comprise an unmethylated CpG motif. In a further embodiment the immunostimulatory oligonucleotide is not T-rich and/or does not comprise an unmethylated TG motif.

The oligonucleotide may be modified in order to improve *in vitro* and/or *in vivo* stability. For example, in one embodiment, the oligonucleotides are modified so as to comprise a phosphorothioate backbone, *i.e.* internucleotide linkages. Other suitable modifications including diphosphorothioate, phosphoroamidate and

29 Aug 2024
2021392894

methylphosphonate modifications as well as alternative internucleotide linkages to oligonucleotides are well known to those skilled in the art and are encompassed by the invention.

In another embodiment, the vaccines or immunogenic compositions of the invention further comprise an immunostimulant selected from the group consisting of: a TLR-1 agonist, a TLR-2 agonist, TLR-3 agonist, a TLR-4 agonist, a TLR-5 agonist, a TLR-6 agonist, a TLR-7 agonist, a TLR-8 agonist, TLR-9 agonist, or a combination thereof.

Calcium composites

In some embodiments, the vaccine or immunogenic composition of the invention comprises a calcium fluoride composite, the composite comprising Ca, F, and Z. "Z" as used herein refers to an organic molecule. As used herein, a "composite" is a material that exists as a solid when dry, and that is insoluble, or poorly soluble, in pure water. In some aspects, Z comprises a functional group that forms an anion when ionized. Such functional groups include without limitation one or more functional groups selected from the group consisting of: hydroxyl, hydroxylate, hydroxo, oxo, N-hydroxylate, hydroaxamate, N-oxide, bicarbonate, carbonate, carboxylate, fatty acid, thiolate, organic phosphate, dihydrogenophosphate, monohydrogenophosphate, monoesters of phosphoric acid, diesters of phosphoric acid, esters of phospholipid, phosphorothioate, sulphates, hydrogen sulphates, enolate, ascorbate, phosphoascorbate, phenolate, and imine-olates.

In some aspects, the calcium fluoride composites herein described comprise Z, where Z is an anionic organic molecule possessing an affinity for calcium and forming a water insoluble composite with calcium and fluoride. In further aspects, the calcium fluoride composites herein described comprise Z, where Z may be categorized as comprising a member of a chemical category selected from the group consisting of: hydroxyl, hydroxylates, hydroxo, oxo, N-hydroxylate, hydroaxamate, N-oxide, bicarbonates, carbonates, carboxylates and dicarboxylate, salts of carboxylic-acids, salts of QS21, extract of bark of *Quillaja saponaria*, extract of immunological active saponin, salts of saturated or unsaturated fatty acid, salts of oleic acid, salts of amino-
25 acids, thiolates, thiolactate, salt of thiol-compounds, salts of cysteine, salts of N-acetyl-cysteine, L-2-Oxo-4-thiazolidinecarboxylate, phosphates, dihydrogenophosphates, monohydrogenophosphate, salts of phosphoric-acids, monoesters of phosphoric acids and their salts, diesters of phosphoric acids and their salts, esters of 3-O-desacyl-4'-monophophoryl lipid A, esters of 3D-MLA, MPL, esters of phospholipids, DOPC, dioleolyphosphatidic derivatives, phosphates from CpG motifs, phosphorothioates from CpG family,
30 sulphates, hydrogen sulphates, salts of sulphuric acids, enolates, ascorbates, phosphoascorbate, phenolate, α-tocopherol, imine-olates, cytosine, methyl-cytosine, uracyl, thymine, barbituric acid, hypoxanthine, inosine, guanine, guanosine, 8-oxo-adenine, xanthine, uric acid, pteroic acid, pteroylglutamic acid, folic acid, riboflavin, and lumiflavin. In further aspects, the calcium fluoride composites herein described comprise Z, where Z is selected from the group consisting of: N-acetyl cysteine; thiolactate; adipate; carbonate; folic acid;

2021392894 29 Aug 2024

glutathione; and uric acid. In some aspects, the calcium fluoride composites herein comprise Z, where Z is selected from the group consisting of: N-acetyl cysteine; adipate; carbonate; and folic acid. In further aspects, the calcium fluoride composites herein comprise Z, where Z is N-acetyl cysteine, and the composite comprises between 51% Ca, 48% F, no more than 1% N-acetyl cysteine (w/w) and 37% Ca, 26% F, and 37% N-acetyl cysteine (w/w). In further aspects, the calcium fluoride composites herein comprise Z, where Z is thiolactate, and the composite comprises between 51% Ca, 48% F, no more than 1% thiolactate (w/w) and 42% Ca, 30% F, 28% thiolactate (w/w). In further aspects, the calcium fluoride composites herein comprise Z, where Z is adipate, and the composite comprises between 51% Ca, 48% F, no more than 1% adipate (w/w) and 38% Ca, 27% F, 35% adipate (w/w). In further aspects, the calcium fluoride composites herein comprise Z, where Z is carbonate, and the composite comprises between 51% Ca, 48% F, no more than 1% carbonate (w/w) and 48% Ca, 34% F, 18% carbonate (w/w). In further aspects, the calcium fluoride composites herein comprise Z, where Z is folic acid, and the composite comprises between 51% Ca, 48% F, no more than 1% folic acid (w/w) and 22% Ca, 16% F, 62% folic acid (w/w). In further aspects, the calcium fluoride composites herein comprise Z, where Z is glutathione, and the composite comprises between 51% Ca, 48% F, no more than 1% glutathione (w/w) and 28% Ca, 20% F, 52% glutathione (w/w). In further aspects, the calcium fluoride composites herein comprise Z, where Z is uric acid, and the composite comprises between 51% Ca, 48% F, and no more than 1% uric acid (w/w) and 36% Ca, 26% F, and 38% uric acid (w/w).

Aluminium salts

In one embodiment, the vaccine or immunogenic composition of the invention comprises an aluminium salt. Suitable aluminium salt adjuvants are well known to the skilled person and include but are not limited to aluminium phosphate, aluminium hydroxide or a combination thereof. Suitable aluminium salt adjuvants include but are not limited to REHYDRAGEL HS, ALHYDROGEL 85, REHYDRAGEL PM, REHYDRAGEL AB, REHYDRAGEL HPA, REHYDRAGEL LV, ALHYDROGEL or a combination thereof.

In particular, the aluminium salts may have a protein adsorption capacity of between 2.5 and 3.5, 2.6 and 3.4, 2.7 and 3.3 or 2.9 and 3.2, 2.5 and 3.7, 2.6 and 3.6, 2.7 and 3.5, or 2.8 and 3.4 protein (BSA)/ml aluminium salt. In a particular embodiment of the invention, the aluminium salt has a protein adsorption capacity of between 2.9 and 3.2 mg BSA/mg aluminium salt. Protein adsorption capacity of the aluminium salt can be measured by any means known to the skilled person. The protein adsorption capacity of the aluminium salt may be measured using the method as described in Example 1 of WO 12/136823 (which utilises BSA) or variations thereof.

Aluminium salts described herein (*i.e.* having the protein adsorption capacity described herein) may have a crystal size of between 2.8 and 5.7nm as measured by X-ray diffraction, for example 2.9 to 5.6nm, 2.8 to 3.5nm, 2.9 to 3.4nm or 3.4 to 5.6nm or 3.3 and 5.7nm as measured by X-ray diffraction. X-ray diffraction is

29 Aug 2024

2021392894

well known to the skilled person. In a particular embodiment of the invention the crystal size is measured using the method described in Example 1 of WO 12/136823 or variations thereof.

The polypeptide(s) and/or nucleic acid(s) described herein may be administered to a subject by any route of administration, for example, orally, nasally, sublingually, intravenously, intramuscularly, intradermally (*e.g.* a skin patch with microprojections) or transdermally (*e.g.* an ointment or cream).

A seventh aspect of the invention provides a polypeptide defined in the first aspect, a nucleic acid defined in the second aspect, a vector defined in the third aspect, and/or a vaccine of the sixth aspect for use in medicine.

An eighth aspect provides a polypeptide defined in the first aspect, a nucleic acid defined in the second aspect, a vector defined in the third aspect, and/or a vaccine of the sixth aspect for use in raising an immune response in a mammal, for example, for treating and/or preventing one or more disease.

A ninth aspect provides the use of a polypeptide defined in the first aspect, a nucleic acid defined in the second aspect, a vector defined in the third aspect, and/or a vaccine of the sixth aspect for raising an immune response in a mammal, for example, for treating and/or preventing one or more disease.

A tenth aspect provides the use of a polypeptide defined in the first aspect, a nucleic acid defined in the second aspect, a vector defined in the third aspect, and/or a vaccine of the sixth aspect for the manufacture of a medicament for raising an immune response in a mammal, for example, for treating and/or preventing one or more disease.

An eleventh aspect provides a method of raising an immune response in a mammal, the method comprising or consisting of administering the mammal with an effective amount of a polypeptide defined in the first aspect, a nucleic acid defined in the second aspect, a vector defined in the third aspect, and/or a vaccine of the sixth aspect.

The use or method of any one of the seventh to eleventh aspects wherein the one or more disease is urinary tract infection (UTI). Alternatively or additionally, the UTI is caused by one or more bacterium of a genus selected from the group consisting of *Escherichia* and *Klebsiella*. Alternatively or additionally, the one or more bacterium is selected from the group consisting of *Escherichia coli* and *Klebsiella pneumoniae*.

2021392894 29 Aug 2024

Alternatively or additionally, the *Escherichia coli* is a UroPathogenic *Escherichia coli* (UPEC). Alternatively or additionally, the one or more bacterium is selected from the group consisting of *E. coli* J96, *E. coli* UPEC 536, *E. coli* CFT073, *E. coli* UMN026, *E. coli* CLONE Di14, *E. coli* CLONE Di2, *E. coli* CFT073; *E. coli* IA139, *E. coli* 536, *E. coli* NA114, and *E. coli* UTI89. Alternatively or additionally, the one or more bacterium is selected from the group consisting of the following *K. pneumoniae* strains: C3091, 3824, 3857, 3858, 3859, 3860, 3861, 3928, 3950, 3951, 4041, 4121, 4133, sp3, sp7, sp10, sp13, sp14, sp15, sp19, sp20, sp22, sp25, sp28, sp29, sp30, sp31, sp32, sp33, sp34, sp37, sp39, sp41, cas119, cas120, cas121, cas122, cas123, cas124, cas125, cas126, cas127, cas128, cas663, cas664, cas665, cas666, cas667, cas668, cas669, cas670, cas671, cas672, cas673, cas674, cas675, cas676, cas677, cas678, cas679, cas680, cas681, cas682, Kp342 and MGH78578.

Preferred, non-limiting examples which embody certain aspects of the invention will now be described, with reference to the following tables and figures.

Fig. 1A and Fig. 1B. Schematic representation of FimH constructs.

A) Fig 1A. Structure of stabilized FimH (PDB: 4XO9). Cartoon representation of FimH stabilized by FimGFimG donor strand (in blue – indicated by the arrows). Domain $FimH_L$ is in yellow (top portion) while $FimH_P$ in Red (bottom portion). Glycines natural linker between domains is represented in green sticks.

B) Fig. 1B Structure of FimH_DG_PGDGN_Ferritin. Aminoacidic sequence of FimH_DG_PGDGN (light blue) fused to ferritin (red). A linker composed by SGS-8H-GSG- is connecting FimH to ferritin molecule. IgK leader sequence for expression in mammalian cells and secretion into medium is in yellow followed by the extra N-terminal charged residues. Model of the 3D structure obtained with Rosetta common software. Cartoon representation of FimH_DG displayed on Ferritin surface. 24 FimH subunits are present and coloured in yellow\blue while ferritin in red.

Fig. 2A and Fig. 2B. *E. coli* expression of FimH Nanoparticles results in inclusion Bodies formation:

A) Fig. 2A. SDS-PAGE analysis of *E. coli* cytoplasmic expression boiled and reduced samples of FimH_DG_(GSG4)-Ferritin, FimHL cys-cys_QBeta, FimHL cys-cys_mI3 and FimHL-NOcys-MI3. Constructs are expressed but can be detected only in the insoluble fraction (Urea 8M, U8M) and not in the soluble fraction (sol). The proteins cannot be detected in the total lysate fraction (Tot), due to insolubility; an accumulation of insoluble material can be detected in the upper part of the gel. Anti-His western blotting of *E. coli* cytoplasmic expression boiled and reduced samples of FimHL-Nocys-MI3.The mutation of the internal disulphide bridge in FimHL domain did not improve solubility as in the soluble fraction only a faint band can be detected.

20213928894 29 Aug 2024

B) SDS-PAGE analysis of *E. coli* periplasmic expression of FimHL-MI3 and cytoplasmic FimHL-ferritin. Bands corresponding to FimHL-MI3 and Ferritin fusions were detected in the Total lysate and in the Insoluble fraction (U8M).

Fig. 3. Prediction of N-Glycosylation FimH sites using NetNGly prediction software.

Fig. 4. Expression of Stabilized FimH constructs (FimH_ΔGG_PGDGN_DG: 930SI; FimH_DNKQ_DG: 931SI; FimH_PGDGN_DG: 932SI) and FimHC complex in mammalian cells.

Fig. 5. Western blot analysis of mammalian expressed constructs containing N-terminal extra amino acids.

(A) FIG. 5A: A band corresponding to FIMH nanoparticle was detected only for FIMH_DG_PGDGN-ferritin(995SI) after 3 days and 6 days post transfection.

(B) FIG. 5B: Cartoon representation of FimH from strain 536, the three different residues compared to J96 are highlighted and represented in sticks.

(C) FIG. 5C: PNGase treatment of FIMH_DG_PGDGN_IMX313 and FIMH_DG_PGDGN_ferritin from strain J96. After treatment a shift of the FIMH_DG_PGDGN_IMX313 at the correct MW was obtained, suggesting that the protein is glycosylated in mammalian cells. FIMH_DG_PGDGN_ferritin from strain J96, was not detected in both untreated and treated PNGase samples, suggesting that this protein degrades.

Fig. 6. MS-Spec peptide mapping.

Fig. 7. Expression of candidates not containing extra N-terminal amino acids by Western blot.

Fig. 8. Cryo-EM NS-EM (negative stain) of candidates with extra or without extra AA at N-Term.

Fig. 9. Cryo-EM NS-EM (negative stain) of candidates without extra AA at N-Term.

A) FIG. 9A: Negative staining microscopy images of 1095SI FIMHL- ferritin (strain 536), NO extra amino acids.

B) FIG. 9B: Negative staining microscopy images of FIMHL -MI3 (strain J96) NO extra amino acids.

C) FIG. 9C: Negative staining microscopy images of 1184SI FIMH_DG_PGDGN_536-encapsuline, NO extra amino acids.

Fig. 10. 3D map shows the presence of three “anchor-like” appendages on the 3-fold axis.

Fig. 11. IgG titers measure by ELISA assay. Mice sera were tested at 21 (Post I, green), 35 (post II, blue), and 45 (post III, red) days post-vaccination. FimHL produced from E. coli was used as ELISA plate coating.

Fig. 12. Bacterial inhibition assay (BAI) on SV-HUC cells. Bacterial adhesion measured by microscopy analysis (OPERA Phenix) and SV-HUC (ATCC) cells were used. The Fluorescence Volume or Area of adherent bacteria

29 Aug 2024
20213928894

(µm3 or 2) was used as readout. Pool of sera raised against recombinant protein FimHC, FimHL-cys (purified from E. coli) were used as control. Pool of sera raised against recombinant protein purified from ExPIGnti expression mammalian system FimH_PGDGN_DG(932SI), FimH_DNKQ_DG(931SI), FimH_DNKQ_DG_Deglyc(951SI) and FimH_PGDGN_DG-Ferritin(995SI were used to measure their ability to inhibit the bacterial binding to the SV-HUC cells. Pool of sera raised against AS01 were used as negative control.

Fig. 13. Biochemical characterization of purified FimH_PGDGN_DG by SDS-PAGE, SE-UPLC and RP-UPLC.

Fig. 14. Biochemical characterization of purified FimH_DNKQ_DG by SDS-PAGE, SE-UPLC and RP-UPLC.

Fig. 15. Biochemical characterization of purified FimH_DNKQ_DG_deglycosylated by SDS-PAGE, SE-UPLC and RP-UPLC.

Fig. 16. Biochemical characterization of purified FIMH_DG_PGDGN_ferritin (sequence from UPEC 536 strain) with extra AA at N-Term.

Fig. 17: FimH-specific total IgG (ELISA). Fig. 17 A) Anti-FimH IgG titers in mice sera at post 3 plotted as a function of MPL dose. Fig. 17B) **Anti-FimH IgG** titers in mice urine measured after 1st, 2nd and 3rd vaccine dose. Pre-immune serum was used as negative control. FimHC was immunized in combination with the adjuvants using 1.6 µg of protein content.

Fig. 18. FimH-specific total IgG (ELISA): comparison of bacterial and mammalian expression systems in sera and urine. Fig. 18 A) The antibody titers were assumed to be lognormally distributed and geometric mean titers (GMTs) and their two-sided 95% CIs were computed. For comparison of groups, an ANOVA model was fitted on log10 titers with groups, timepoints and their interaction as fixed factors and a repeated statement for timepoints. Heterogeneity of variances was considered between groups. Geometric mean ratios and their 95% CIs were derived from this model. Antibodies response to each formulation was evaluated against FimHDG used for ELISA plate coating. All statistical analyses were performed using SAS 9.4**. Fig. 18 B)** FimH-specific total urine IgGs.

Fig. 19. FimH-specific total IgGs. ELISA post dose I results: The antibody titers were assumed to be lognormally distributed and geometric mean titers (GMTs) and their two-sided 95% CIs were computed. For comparison of groups, an ANOVA model was fitted on log10 titers with groups, timepoints and their interaction as fixed factors and a repeated statement for timepoints. Heterogeneity of variances was considered between groups. Geometric mean ratios and their 95% CIs were derived from this model. Antibodies response to each formulation was evaluated against FimHDG used for ELISA plate coating. All statistical analyses were performed using SAS 9.4.

Fig.20 FimH-DG elicits a functional immune response. Bacterial inhibition assay of selected constructs in comparison to FimHC. Relative potency is calculated as reported in the examples.

20213928894 29 Aug 2024

Fig. 21. Antibody ability of FimHDG antibodies to inhibit ExPEC adhesion using a bacterial inhibition assay (BAI). All candidates were formulated with AS01.

Fig. 22. SPR analysis of FimH samples and mAb926 interaction (Sensorgrams).

To study the interaction of FimH candidates and mAb926 a SPR analysis was performed resulting in a sensorgram representing a plot of response (ordinates) against time (abscissae) showing the progress of the interaction. Response was measured in Resonance Units (RU) which is directly proportional to the concentration of the molecules on the sensor chip surface. Each sensorgram is composed of two parts, corresponding to the association and dissociation phases of an interaction. The association is the first phase in a biomolecular interaction, during which the binding occurs, when analyte and ligand collide due to diffusion and when the collision has the correct orientation and enough energy. The dissociation is the phase in which the ligand-analyte complex dissociates; the profile of the dissociation can give information about the complex stability: the slower the dissociation, the higher the complex stability and vice versa.

Fig. 23. Fig. 23 A: SDS page analysis of culture supernatant expressing FimHDG tagless in mammalian cells. SDS_Page analysis and SEC-UPLC analysis of purified FimHDG tagless from Expi293 cells and ExpiCHO cells. Fig. 23B: Nano-DSF profiles and melting temperatures values obtained for FimHDG tagless purified from Expi293 and ExpiCHO cells compared to the FimHDG containing the C-terminal His tag. Fig.23 C: SPR binding analysis of mAbs 926 and 475 to FimHDG tagless compared to FimHDG His. SPR analysis of mannose binding to FimHDG tagless compared to FimHDG His. Fig.23 D: SDS-Page analysis of supernatants of FimHDG-ferritin constructs containing different linkers and containing or not the initial Asp residue. Western blotting analysis of pellet from mammalian cells using anti-FimH specific mice serum.

Fig. 24. PROSS-based calculations of a symmetric monomer (relative to other 23 chains) in the octahedral *E. coli* nanoparticle (PDB 1EUM) to introduce stabilizing mutations with increased affinity or stability (bottom left of chart).

Fig. 25. Fig. 25 A: SDS page analysis of total (T) and soluble (S) extracts of WT *E. coli* ferritin and different mutants. Fig. 25 B SEC profile of mutant 0.5. All constructs had a profile with a strong peak (arrow) in the dead volume, which is compatible with the formation of a nanoparticle.

Fig. 26. NS-EM (negative stain) analysis of *E. coli* ferritin WT and different mutants (0.5, 2, 2.5, 6).

Fig. 27. Differential Scanning Fluorimetry analysis of ferritin constructs with thermal profiles. Graph on the left shows the derivate of fluorescence intensity *vs*. temperature. The circle, on the table on the right, indicates the mutant (0.5) with the highest T_m.

Fig. 28. On the left, Western Blotting analysis using anti-His antibody of supernatant expressing different nanoparticles constructs of FimH. The star indicates the *E. coli* nanoparticles FimHDG-ferritin (mutant 0.5). On the right, TEM analysis show the presence of correctly formed ferritin nanoparticles.

20213928894 29 Aug 2024

EXAMPLES

The inventors designed a stable un-complexed (in absence of FimC) variant of full-length FimH in which FimG donor strand peptide [SEQ ID NO: 5] was genetically fused through a linker of 4 or 5 residues (DNKQ [SEQ ID NO: 8] or PGDGN [SEQ ID NO: 7]) to the C-terminus of FimHp, obtaining a "FimH_DG" protein with structural and functional properties of FimH in the assembled pilus. Linkers were designed by choosing highly polar charged residues (DNKQ) or inserting a Proline residue (PGDGN linker) as first residue of the linker that is predicted to support the turn in the secondary structure and to promote the correct protein architecture. In addition, a construct in which two glycines present in the linker that connects $FimH_L$ to FimHp were deleted, to further reduce the flexibility of $FimH_L$ and reduce mannose binding (Fig. 1A).

Moreover, a nanoparticle design for FimH can be utilized to expose multiple copies of stabilized FimH and further increase its immunogenicity as enabler for a 1-2 dose vaccine.

Virus-like particle (VLP) and protein Nanoparticles (NPs) are display platforms for other antigens with potential to induce effective B- and T-cell responses. They have intrinsic ability to self-assemble into highly symmetric stable and organized structures. Several chimeric VLPs/NPs are under investigation in preclinical and clinical research worldwide. Particularly, ferritin scaffold has been genetically fused with viral hemagglutinin to obtain particles that were more immunogenic, in presence of adjuvant, at one dose compared to a seasonal flu vaccine (Nature 2013, 49, 104). The same approach has been used in preclinical research for many other antigens (Chen Y, et al. Vaccine. 2020 Jul 31;38(35):5647-5652). The challenge is not only to engineer a correctly assembled particle presenting the antigens of interest, but also to obtain it manufacturable and scalable. To explore the potential of self-assembling NPs and VLPs to display FimH candidates, different chimeras have been designed through genetic fusions and tested.

Helicobacter pylori ferritin nanoparticle is composed of 24 subunits, a total of eight trimers of the desired antigen can be display in the highly symmetrical octahedral cage structure of ferritin nanoparticles (Fig. 1B). Recently, protein i301 nanocage, a 60-mer NP based on the *Thermotoga maritima* 2-keto-3-deoxy-
25 phosphogluconate (KDPG) aldolase have been computationally designed (Hsia Y, et al. Nature. 2016 Jul 7;535(7610):136-9.). i301 stability has been further improved by mutating two cysteines (mI3) (Bruun TUJ, et al. ACS Nano. 2018 Sep 25;12(9):8855-8866) and by fusing SpyCatcher to the N-terminus of the protein.

We constructed recombinant plasmids to genetically fuse ferritin, mI3 or encapsulin to FimH_DG_PGDGN stabilized antigen or FIMHL and FIMHLCys antigens. In order to separate the displayed antigen and the NP, a
30 linker was added between the two sequences.

The linkers tested contain repetition of Gly and Ser residues but could also contain internal 8xHis tag in order to allow protein purification. In order to increase protein expression and solubility in the *E. coli* cytoplasmic

20213928894 29 Aug 2024

space of FimH NPs, FIMHL constructs mutated of the internal S_S bridge (C24SC65S) were also fused to Ferritin and mI3 and tested for expression and solubility.

Materials and methods

Cloning and E. coli expression

The FimH-NP bacterial constructs were synthesized by Geneart as DNA Strings and cloned directly into the pET15-tev, pET21 or pET22 (see table 1) with the Takara infusion cloning kit. Other constructs were purchased as synthetic genes from Geneart, with the protein of interest directly cloned into the expression vector (pTRC-HIS2A from Life Technologies). All synthetic genes were optimized for *E. coli* expression and contained N terminal, C-terminal or internal HIS tag to allow protein affinity purification. Proteins were expressed in BL21DE3T1r (NEB) or in T7shuffle express using HTMC medium and IPTG induction at 20°C for 24h.

After pellet recovery, it was resuspended in the lysis buffer cell lytic express (Merk) or B-Per solution (Pierce) for 1h at 25°C. After centrifugation a visible inclusion bodies (IB) pellet was present, and it was resolubilized in Urea 8M (U8M). Protein expression and solubility was assessed by SDS-page of the samples collected from soluble fraction (S) and insoluble fraction (IB).

Recombinant proteins production in mammalian cells.

The FimH-NP mammalian constructs (See table 2) were synthesized by Geneart as synthetic genes in pCDNA3.1 or pCDNA3.4 (Life Technologies) vectors. All sequences were codon optimized for expression in mammalian cells and contained an N-terminal leader sequence for secretion into the cells medium. This sequence is the IgK murine leader sequence METDTLLLWVLLLWVPGSTGD **[SEQ ID NO: 9]**, or the IgK murine leader sequence followed by 15 additional charged residues AAQPARRARRTKLAL **[SEQ ID NO: 78]**. (Fig. 1B)

To produce recombinant FimH-NPs, the expression vectors were transfected into Expi293GNTI cells according to the manufacturer instructions (Life Technologies). The Expi293F GnTI- cell line is derived from engineered Expi293F cells that do not have N-acetylglucosaminyltransferase I (GnTI) activity and therefore lack complex N-glycans leading to homogeneously glycosylated recombinant proteins.

Briefly, 30 µg of pCDNA-FimH-NPs-expressing vectors were transfected into 30 ml culture containing 75 x 106 Expi293F cells using ExpiFectamine 293 Reagent. Cells were incubated at 37°C, 120 rpm, 8% CO2 and after 24 h, ExpiFectamine 293 Transfection enhancer 1 and 2 were added. Cells were further incubated at 37°C for 144 h. Aliquots of cultures were harvested every 24 h and analyzed for NA expression by SDS-PAGE and Western Blot (WB). Seventy-two and 144 h after transfection, cell cultures were centrifuged at 1000 rpm for 7 min and the supernatants were harvested, pooled, clarified by centrifugation, filtered through a 0.22 µm filter, and stored at -20°C until purification.

20213928894 29 Aug 2024

PNGase F Proteomics Grade, (P7367, sigma) was used to check glycosylation of mammalian expressed antigens according to manufacturer protocol.

Western blotting was performed using a standard protocol with anti-his-HRP antibodies by sigma diluted 1:1000 or with anti FimHL-cys antibodies raised in mice using the bacterial FimHL-cys purified protein and secondary anti-mouse-HRP antibodies.

Affinity chromatography with Ni2+ was used to purify NPs from culture supernatants. Fractions of interest were pooled and were concentrated by using 100 kDa cut-off spin concentrator (Millipore Amicon Ultra); sodium dodecyl sulphate-poly-acrylamide gel electrophoresis (SDS-PAGE) was performed to check protein purity. Recombinant FimH-NPs and FimH-DG antigens were purified by preparative size exclusion chromatography (SEC) equilibrate with PBS buffer.

All the collected fractions were checked for FimH-NPs or FimH-DG protein content by SDS-PAGE and interested fractions were pooled, filtered at 22 µm, aliquoted and stored at -20°C.

To assess protein size and purity, analytical SEC-HPLC and reverse phase RP-UPLC were performed. Moreover FimH-NPs were analysed by Dynamic light scattering in order to further determine the molecular weight and nanoparticle assembly and the proteins sequence identity was assessed by LC-MS.

Immunisation

Twelve CD1 mice (female) per groups were immunised with 15 micrograms of candidates expressed in mammalian or bacterial systems were adjuvanted with ASO1. All the mice were inoculated by subcutaneous injection (SC) with 200 µl (PBS dilution) of antigen mixture or adjuvant alone for three times. Blood was collected through the tail vein at 0 (preimmune), 21 (Post I), 35 (post II), and 45 or 49 (post III) days post-vaccination.

Analysis of FimH-specific antibody

Serum FimH-specific lgG were measured by enzyme linked immunosorbent assay (ELISA). Briefly, 96-well microtiter plates were coated with 100µl antigen (1µg/ml) to each well of a 96 well Nunc Maxsorp plate and incubated overnight at 4°C. 250µl of (PVP) saturation buffer was added to each well and the plates incubated for 2 hours at 37°C. Wells were washed three times with PBT. Next, 100µl of diluted sera were added to each well and the plates incubated for 2 hours at 37°C. Wells were washed three times with PBT. 100µl of Alkaline phosphatase-conjugated secondary antibody serum diluted 1:2000 in dilution buffer were added to each well and the plates were incubated for 90 minutes at 37°C.

Wells were washed three times with PBT buffer. 100µl of substrate p-nitrophenyl phosphate were added to each well and the plates were left at room temperature for 30 minutes. 100µl 4N NaOH was added to each

20213928894 29 Aug 2024

well and OD 405/620-630nm was followed. The antibody titres were quantified as the dilution of serum that gives an absorbance of 0,4 OD using a multimode microplate reader.

BAI assay

Bacteria (UTI89 wt_mCherry clone2) cultivated in 3 passages of static liquid culture: the growth condition for inducing FimH expression. BAI assay performed with selected conditions: bacterial density of 0,012 OD/ml and incubation time of 30min. Bacterial adhesion measured by microscopy analysis (OPERA Phenix). SV-HUC (ATCC) cells were cultivated in SV-HUC complete medium: F12K (Thermo Scientific) supplemented with 10% FBS and antibiotics. Pre-infection medium: complete media w/o antibiotics.

Tested sera (Heat Inactivated):

Serum ID
Anti-FimHL-cys
Anti AS01
FimH_PGDGN_DG (mammalian)
FimH_DNKQ_DG (mammalian)
FimH_DNKQ_DG Deglyc (mammalian)
FimH_PGDGN_DG Fer, 15ug (mammalian)
FimH_PGDGN_DG Fer, 3ug (mammalian)

3 x T75 flasks of SV-HUC cells (3 x 106 cells/ml, 95% vitality) were trypsinized (x5min, at 37°C). Cells were seeded in 96-well plates, seed 60 wells/plate with 3.5 x 104 cells/well (VF=200ul/well) and incubated at 37°C, 5% CO2. Bacteria preparation consists in three passages of static liquid culture: UTI89 strains are inoculated in 20ml LB (125-ml flask) from plate and are incubate at 37°C, O/N, in static condition. This dilution/incubation passage was repeated three times.

The medium of SV-HUC cells was exchanged with pre-infection medium w/o antibiotics (200ul/well).

2x solutions of sera were prepared in U-bottom 96-well plate with F12K medium or F12K+10%FBS, as indicated below and further diluted with serial dilutions.

20213928894 29 Aug 2024

1 ml of Passage3 Bacterial culture UTI 89 mcherry Clone2 were transferred into single tubes and centrifuged at 4500 x g for 5min at room temperature. Bacteria were washed with PBS and pelleted. Finally, the bacterial pellet was resuspended at 0.5 OD600 /ml with infection medium.

Infection was performed as follows: in each plate medium was sucked off and 50ul/sample of 2x serum/mannose (20% D-(+)-Mannose) solutions or infection medium (ctrl positive & negative) were added followed by 50ul/sample of 2x inoculum or infection medium (ctrl negative). Plates were incubated for 30min and serum dilution from 15% to 0.06% was added. Plates were incubated at 37°C, 5% CO_2, for 30min and the medium was removed and the plates wells were washed with PBS for three times. Bacteria were fixed using 4% formaldehyde (200ul/well) solution. After incubation for 20min, fixation solution was removed, and samples were washed 3 times with PBS (200ul/well). DAPI (62248, ThermoScientific) solution was diluted 1:5000 in PBS and 100ul were added to each well. Samples were incubated for 10 min at room temperature (in the dark). DAPI solution was removed, and PBS was added in each well (200ul/well). Samples were stored at 4°C in the dark and 3h at RT before imaging with OPERA Phenix. Whole well area was acquired with a 10x air objective using the Alexafluor488 setting. For each field a Z-stack (4 planes) was acquired. Data were analysed with Harmony software. Total bacterial fluorescence area (single object ≤ 100 µm2) was calculated as a value of adherence.

Results

FimH stabilized as monomeric antigens as well as FimH-stabilized nanoparticles are secreted as soluble proteins in mammalian expression system and can be easily purified by IMAC

As a first attempt, several FimH NPs constructs have been generated and tested in different conditions. T7 and pTac promoter, of pETvectors and pTrcHIs2A vector respectively have been used to test and solubility of the candidate antigens in *E. coli*. Moreover, both cytoplasmic and periplasmatic expression have been tested, as well as different *E. coli* strains optimized for disulphide bridges formation into the cytoplasmic space as the T7 Shuffle express. In order to increase protein expression and solubility in the *E. coli* cytoplasmic space of FimH NPs, FimHL constructs mutated of the internal S_S bridge were also fused to Ferritin and mI3 and tested for expression and solubility.

However, none of the constructs resulted in soluble protein expression suggesting that the *E. coli* expression system could be not optimal for obtaining FimH nanoparticles. The mutation of the internal disulphide bridge in FimHL domain did not improve significantly solubility as in the soluble fraction only a faint band was detected by western blotting analysis.

E. coli is a prokaryotic expression system that is strongly preferred for low-cost fermentation and easy process. However, the production of proteins by *E. coli* could results in recombinant proteins mainly

29 Aug 2024
20213928894

expressed as inclusion bodies, which are insoluble and inactive, and may require complex refolding process *in vitro* (Fig. 2).

To overcome the problem of insolubility in *E. coli*, the inventors decided to switch to the mammalian EXPI293F expression system. First, the FimH sequence was analysed for N- and O-Glyco sites possibly responsible of glycosylation. Fig. 3 reports the position of putative N-glycosylation sites. O-Glyco site were not detected (data not shown).

In order to express bacterial protein in mammalian cells, reducing as much as possible the glycosylation which occurs in this system compared to the *E. coli* system, the inventors used a genetic mutated EXPI293F cell line called Expi293F GnTI (Thermofisher). This cell line is derived from engineered Expi293F cells but does not have N-acetylglucosaminyltransferase I (GnTI) activity and therefore lacks complex N-glycans leading to homogeneously glycosylated recombinant proteins.

The full length FimH proteins stabilized with the FimG donor strand (FimH-DG) from *E. coli* from strain 536 and/or J96 containing a secretion murine Ig-κ chain leader sequence (plus extra amino acids at the N-terminus of the $FimH_L$ domain (in some of the constructs; Table 1) alone or fused to protein NPs (ferritin, mI3, IMX313, encapsulin and HBc) were used to transfect EXPI293 GNTI cells. The accumulation of secreted recombinant protein was characterized by measuring their expression in culture supernatants at 72 h and 144 h post-transfection by WB and SDS-PAGE. Both analyses revealed that FimH soluble expression could be obtained at high level for several constructs, while others could not be obtained. Expressed and soluble FimH-DG stabilized proteins and FimH-NPs containing the C-term 6xHis tag or internal 8XHis tag were purified from 72 h and 144 h pooled culture media using ion metal immobilized chromatography and preparative SEC chromatography. SDS-page analysis of proteins produced in mammalian expression system revealed that they run at a higher MW compared to the corresponding bacterial proteins, suggesting that they were glycosylated. Consequently, two constructs lacking the putative residues involved in N-glycosylation have been mutated, FimH_DNKQ_DG_deglyc and FimH_PGDGN_DG_deglyc, containing the extra amino acids N-Term and the following mutations N28S, N91D, N249D, N256D, were produced (Table 1).

Western blot analysis of supernatants of mammalian expressed constructs containing N-terminal extra amino acids revealed an expression band corresponding to FimH_DNKQ_DG, FimH_PGDGN_DG and FimHC complex. On the contrary, the FimH_ΔGG_PGDGN_DG (deletion of Gly resides connecting FimHL and FimHP) was not detected after 3 days and 6 days post-transfection (Fig. 4). Protein characterization of the purified products are reported in Fig. 13-16.

The constructs FimH_PGDGN_DG_Ferritin (strain 536; 995SI), containing N-terminal extra AA were successfully expressed and purified. On the contrary, all FimH non-FimG donor stand stabilized constructs (936SI)-FimH-IMX313 j96; (935SI)-FimH_mi3 j96; (929SI)-FimHL-HIS-mI3 j96, all containing N-terminal extra amino acids, were not detected in culture supernatants.

20213928894 29 Aug 2024

PNGase treatment of FimH_DG_PGDGN_IMX313 and FimH_DG_PGDGN_ferritin from strain J96 revealed a shift of the FimH_DG_PGDGN_IMX313 at the correct MW in the treated sample, suggesting that the protein was glycosylated in mammalian cells. FIMH_DG_PGDGN_ferritin from strain J96, was not detected in both untreated and treated PNGase samples, suggesting that this protein was degraded. FimH_PGDGN_DG_Ferritin (strain J96) (1000SI), was not purified from collected supernatants from 3 days and 6 days, even if the protein was detected immediately after sample collection, due to degradation and this construct was not obtained (Fig. 5AFig. -C).

In addition, the predicted N-glycosites reported in Fig. 3 were muted in serine or aspartic acid. The resulting FimH_DNKQ_DGDeglyc candidate showed a higher peptide mapping coverage in comparison to the WT sequence. This result indicates that a possible glycosylation might occur in correspondence of these specific mutated amino acids (Fig. 6)

Moreover, representative constructs reported in Fig. 7 were expressed removing the extra N-terminal amino acids (short leader). The FimH-DG_PDGDN_ferritin (strain 536, extra N-terminal AA) was obtained with a purity of 88% by RP-UPLC. (998SI) FimH_PGDGN_DG-HIS-IMX313 j96 was also well expressed and was successfully purified.

All these constructs were expressed as secreted soluble proteins in cells medium and were further purified as previously described. Western blotting analysis of supernatants with anti-FimHL-cys antibodies raised with bacterial stabilized protein recognized all mammalian expressed tested NPs (Fig. 7).

To confirm that FimH-Nps were correctly assembled, the purified proteins were examined by analytical SE-HPLC and DLS analysis. In SE-HPLC they eluted in a single large not-sharp peak. Based on the comparisons of the elution volumes (Ev) of ferritin NPs with the Ev of molecular weight (MW) standards run in the same conditions, the calculated MW of the FimH-DG-PGDGN-ferritin NPs is consistent with NPs composed by 24 subunits, as confirmed by DLS analysis.

The construct FimH-DG_PDGDN_ferritin SL (sequence from strain 536 or J96, lacking the extra N-terminal AA) resulted in highly expressed with final purity estimated by RP-HPLC.

FimHL-NPs constructs were also successfully purified, and the biochemical characterization confirmed the formation of NPs composed by 24 subunits, for (1095SI) FimHL-HIS-Fer 536 and by 60 subunits for (1096SI) FimHL-HIS-Mi3 J96.

Visualization of generated FIMH-DG NPs

An additional confirmation that recombinant FimH-DG_PDGDN_ferritin extra AA (Fig. 9A-B) fusion protein FIMH_DG_PGDGN-HIS-Ferritin 536 short leader and FimH_PGDGN_DG_HIS-Ferritn j96 produced in

20213928894 29 Aug 2024

mammalian expression system, form stable correctly assembled NPs was obtained by visualizing the purified proteins using negative stain electron microscopy TEM. As shown in Fig. 8B, (995SI) FimH_PGDGN_DG_Ferritin 536, containing N-terminal extra AA sample appeared as differentially oriented homogeneous population of octahedral particles decorated by spikes. Naked ferritin particles showed a diameter of 13nm while spiky ferritin presented a diameter of 30-32nm. The difference in diameter (8.5nm) corresponds to the length of the FimH (calculated on the FimH model). Also, (1142SI) FimH_DG_PGDGN-HIS-Ferritin 536 is correctly folded and decorated by eight spikes of FimH trimers. No naked ferritin particles were present in the sample. Particles showed a diameter of 30-32 nm. (1042SI) FimH_PGDGN_DG_HIS-Ferritn j96 sample presented a mixed population of NPs, with individual or aggregated proteins, correctly folded spiky NPs presenting eight spikes, presence of folded NPs with multiple spikes and NPs non correctly folded. No naked ferritin particles were detected (Fig. 8D).

Cryo-EM NS-EM (negative stain) of (1095SI) FimHL-HIS-Fer 536 and (1096SI) FimHL (J96)-mI3-his showed that that NPs expressed in mammalian system were fully assembled. FimHL-HIS-Mi3 J96(1096SI) presented correctly folded nanoparticles with an icosahedral shape, highly symmetrical of 40nm and decorated by spikes, with few aggregates (Fig. 9A and FIG. 9B). In addition, 1185SI and 1184SI FIMH_DG_PGDGN_536-encapsuline both short leader at the N-term were correctly assembled (Fig. 9 C and D). The constructs containing the stabilized FimH fused to IMX313 were also successfully purified (1043SI) FimH_DG_PGDGN_IMX313_HIS J96 and (998SI) FimH_PGDGN_DG-HIS-IMX313 j96 and the biochemical characterization confirmed the formation of high molecular weight (HMW) species. However, TEM analysis of these constructs showed the presence of only aggregated protein (data not shown).

Structural features in 3D reconstructions of recombinant FimH-DG-Ferritin NPs

Single particle reconstruction method was applied to TEM images in order to generate the three-dimensional structure of the assembled octahedral particles of (995SI) FimH_PGDGN_DG_Ferritin (FimH sequence from strain 536). Single boxed FimH-DG_PDGDN_ferritin nanoparticle (box size 64x64 pixel) were firstly band pass filtered in order to increase the signal-to noise ratio, then rotationally and translationally aligned, and finally centred before undergoing MSA for classification. Fig. 10A shows a selection of FimH-DG_PDGDN_ferritin most abundant 2D class averages, representative of the different orientations of the particle on the carbon film support. The 3D-EM structure (Fig. 10B) of the soluble FimH-DG_PDGDN_ferritin generated confirmed this structure to be composed of a highly symmetrical octahedral cage structure with the presence of three "anchor-like" appendages on the 3fold.

FIMH-DG stabilized proteins and FIMH-DG_PGDGN-Ferritin NPs are highly immunogenic in mice

To assess the immunogenicity of candidates expressed in the mammalian system (FimH_PGDGN_DG, FimH_DNKQ_DG, FimH_DNKQ_DGDeglyc and FimH_PGDGN_DG_Ferritin), single sera from immunised mice were analysed by ELISA assay using the FimH lectin domain (FimHL) expressed in *E. coli* as coating. Overall,

20213928894 29 Aug 2024

all the candidates elicited an IgG response. FimH_PGDGN_DG and FimH_PGDGN_DG_Ferritin showed similar IgG tites, however the NP candidate showed a more homogeneous and compact response at post II. This result suggested that NP generated an earlier efficacious response in comparison to the candidate expressed as recombinant protein. The FimH_PGDGN_DG_Ferritin immunised at two different doses (15 ug and 3 ug) showed that the lower dose (3ug) was comparable to the higher (15ug) dose in terms of total IgG response indicating that that the ferritin nanoparticles carrying recombinant FimH_DG_PGDGN protein had good immunogenicity even at the lower tested dose of 3ug. In addition, the ferritin form led to a less scattered immune response at the second dose as compared to the other candidates, including the otherwise corresponding FimH construct lacking a nanoparticle domain (Fig. 11).

FimH-DG stabilised candidates (produced in mammalian system) indicate a stronger capability to inhibit bacterial adhesion with respect to recombinant (bacterial produced) form

The ability of sera raised against FimH stabilised candidates to impede bacterial adhesion of human bladder cells was tested using an *in vitro* bacterial inhibition assay. Antibodies against vaccine candidates FimH_PGDGN_DG and FimH_PGDGN_DG_Ferritin were more efficacious than FimH_DNKQ_DG or bacterial-produced FimHL-cys candidate in inhibiting the bacterial adhesion to the urothelial cells. These results indicate that the FimH-based stabilised vaccine candidates expressed into a mammalian system have a great potential for further vaccine development. Furthermore, the linker used for FimH stabilisation played a crucial role for the functionality of the generated antibodies (Fig.12), with constructs having the PGDGN linker being associated with improved results in terms of inhibition of bacterial adhesion.

Conclusions

Our study investigated novel FimH candidates stabilised with the donor stand strategy. Vaccine candidates were produced as single recombinant proteins or assembled into nanoparticles carrying FimH subunits. As expression in *E. coli* resulted in insoluble products, soluble antigen expression has been achieved by using a mammalian expression system, through transient transfection of EXPI293-GNTI cells. To our knowledge, the usage of this expression system has never been used before to produce bacterial proteins. In this case, mammalian expression system improved protein solubility, since FimH expressed from in *E. coli* was insoluble in all tested conditions. This expression system has allowed to produce stabilized FimH_DG antigens all in soluble form, as well as different FIMH nanoparticles (FimHL-mI3, FimHL-Ferritin and FimHH_DG_PGDGN-ferritin). On the contrary, when unstabilised FimH was fused to NP, no expression was detected, demonstrating that stabilisation through FimG complementing strand is necessary to produce full length isolated FimH protein in mammalian cells and to display the antigen on ferritin NPs. The deletion of the two glycine residues, which are the natural linker between FimHL and FimHP resulted in no expression of FimH_ΔGG_PGDGN_DG with extra N-terminal AA, suggesting that this deletion was detrimental for protein stability.

29 Aug 2024
2021392894

SDS-PAGE comparison of MW of bacterial insoluble proteins and the corresponding mammalian expressed protein show that they have different molecular weights, suggesting that the mammalian proteins are glycosylated, as confirmed by PNGase treatment. All constructs with the leader sequence (the IgK murine leader sequence alone or with extra amino acids) were successful in secreting FimH constructs into the expression medium. However, none of the constructs with extra amino acids resulted in more homogenous nanoparticles (Fig. 7 and Fig. 9), and no naked ferritin NPs were observed.

Structural data confirmed that all nanoparticles were correctly assembled, and FimH spikes were detected on the surface of ferritin (24 spikes) and mI3 Nanoparticles (60 spikes).

Our data suggested that FimH stabilised candidates expressed into a mammalian system were immunogenic and the raised antibody were able to inhibit the bacterial adhesion to urothelial cells.

Effect of AS01 adjuvant: improvement over FimHC and FimH-DG

To evaluate the contribution of PHAD and AS01 adjuvants systems to the humoral response, FimHC protein complex was used as model antigen and was expressed as described in Langermann S, et al. Science. 1997 Apr 25;276(5312):607-11. IgGs antibodies raised after vaccination were determined, and relative titers were plotted as a function of MPL amount contained in the PHAD and AS01 formulations. Overall, AS01 induced a higher total IgG response than PHAD in mice sera (post-3) and urine (post-2 and -3). Moreover, AS01B used at 5µg-MPL showed the same IgG level in comparison of PHAD containing 12.5µg-MPL (Fig.17 A and Fig. 17 B)

Improved antigen design and adjuvanted formulation elicit functional immune response after 2 doses (instead of 3)

To evaluate the immune response of FimHC complex and FimH-stabilized his-tagged forms (FimHDG, i.e. FimH-PGDGN-DG, wherein DG stands for donor strand complementing peptide from FimG, FimHDG-Ferritin, i.e. FimH-PGDGN-DG-linker(with His-tag)-Ferritin (from H. pylori)), different antigen doses (0.55µg or 1.6µg) adjuvanted with PHAD or AS01 were used for mice immunization. The FimHC complex was expressed as described in Langermann S, et al. Science. 1997 Apr 25;276(5312):607-11 Protein were expressed in bacterial or mammalian systems. FimH-specific total IgG titers (measured by ELISA) in the sera and urine of immunized mice, measured after second and third vaccine injection. IgG titers of post-2 and post-3 sera raised against different forms of FimHDG candidate, formulated with AS01, were determined (Fig.18 A). IgG values were compared with those induced by vaccination with FimHC used in combination with the same AS01 adjuvant and PHAD (with MPL amounts comparable to those present in the AS01). As shown in Figure 18 A, at 0.55 µg antigen dose it there was a clear enhancement of the antibody response with AS01 over PHAD for FimHC after the second and third administration. Also, a better immune response of FimHDG-

20213928894 29 Aug 2024

HisTag, expressed and purified from a bacterial system, was observed in comparison to FimHC benchmark adjuvanted with PHAD.

Finally, both stabilized FimH candidates purified in mammalian cell (FimHDG-HisTag mammalian and FimHDG-His Tag Ferritin) showed higher response than FimHDG expressed in *E. coli*. Both 1,6 and 0,55 µg of mammalian FimHDG constructs induced IgG levels that plateaued after the 2nd and 3rd immunization. Furthermore, FimHDG at the second administration raised a higher response in compared to 3 doses of FimHC-PHAD at both the tested protein doses (Geometric mean ratio of 9.7 and 3 respectively) (Fig. 18 A). The antibody response against FimHDG was evaluated in urine collected by immunized groups with higher protein dose after 1st, 2nd and 3rd dose. As observed in tested sera, higher IgG titers were measured for mice vaccinated with mammalian FimHDG formulations (Fig.18 B).

For selected immunized groups the total IgG response at post-I was also determined. At post dose I, the FimHDG-Ferritin nanoparticle induced twice higher GMTs than FimHDG without Ferritin (at any Ag dose) although variability was higher than in the post-2 and post-3 responses (lead to big 95% CI including 1).

As compared to bacterial derived antigen, the mammalian form adjuvanted with ASO1 induced higher IgG responses at post dose I and II (observed GMRs ranged from 7.1 to 60.8 with all lower limits of 95% CI above 1), while the response was similar after the 3rd dose (observed GMRs around 1.5-fold) (Fig. 19)

Comparison of different linkers of constructs (mammalian/bacterial) in terms of Relative potency

To investigate the effect of different linkers FimHDG candidates expressed both in a bacterial and mammalian systems were compared to FimHC in terms of bacterial inhibition of the adhesion to uroepithelial cells (BAI). Figure 20 showed that all FimHDG constructs were more functional than FimHC independently from the expression system used for the expression (bacterial or mammalian). Interestingly, FimHDG constructs harboring PGDGN linker were more effective in comparison to the DNKQ constructs. These data suggest the likers can stabilize FimH in different conformation consequently raising a different functional antibody response.The BAI assay has been conceived as a multiple dilution assay, where the tested samples together with a reference pool of sera are plated at different concentrations to estimate the dose-response curves. The signal is normalized between 0% and 100% before titer computation. The titer is express as Relative Potency (RP) of the tested sample against the reference pool, comparing the corresponding dose-response curve. In details, the RP is computed considering the dilution in logarithmic scale and fitting a 4 parameters logistic (4PL) constrained model (described in the Eur.Ph. chapter 5.3) where the standard and tested samples slope-factor, upper asymptote and lower asymptote are constrained to be equal. The RP is computed as the ratio between the Reference and the sample EC_{50}. The EC_{50} are calculated from the 4PL constrained inflection point and back transformed (antilog). The model requires that the curves of the reference and the samples have the same slope-factor (parallelism) and the same maximum and minimum response level at the extreme parts (linearity). The suitability of the assumption of parallelism and linearity is assessed for each session

20213928894 29 Aug 2024

evaluating the P-value to test deviations from parallelism, the P-value to test deviations from linearity and the slope ration between reference and sample.

FimH-DG elicits a functional immune response in term of BAI, HAI and conformational mAb binding

Further, the antibody ability of anti-FimHDG antibodies to inhibit ExPEC adhesion using a bacterial inhibition assay (BAI) was assessed. Data on antibodies functionality revealed that sera raised against both bacterial and mammalian FimHDG constructs showed higher ability than FimHC benchmark sera to inhibit bacterial adhesion. Among the candidate tested, FimHDG-ferritin showed at least 10-fold higher functionality compared to the congener FimHDG construct (Fig.21). Similar results were obtained performing an HAI assay. The FimHC complex was expressed as described in Langermann S, et al. Science. 1997 Apr 25;276(5312):607-11. "FimHDG" refers to FimH-PGDGN-DG, wherein DG stands for donor strand complementing peptide from FimG, "FimHDG-Ferritin" refers to FimH-PGDGN-DG-linker(with His-tag)-Ferritin (from H. pylori). The BAI assay and relative potency calculation were performed as described in the previous example.

FimHDG and mAb962 binding

To study the interaction of FimHDG (i.e., FimH-PGDGN-DG, wherein DG stands for donor strand complementing peptide from FimG) and mAb926 (Dagmara I. Kisiela et al (2015) a SPR analysis was performed. FimHDG monomeric forms (i.e., FimH-PGDGN-DG, wherein DG stands for donor strand complementing peptide from FimG) obtained from bacterial or mammalian system showed similar binding to the mAb with slightly differences in the association and dissociation profiles. FimHDG-Ferritin (FimH-PGDGN-DG-linker(with His-tag)-Ferritin (from H. pylori)) resulted in a more stable interaction compared to the monomeric forms possibly due to the multimerization effect with increased avidity. By contrast, the lower interaction of mAb926 with FimHC in comparison with FimHDG suggested that the latter was stabilised in a pre-binding conformation as expected. In fact, mAb926 was generated against a FimH stabilized lectin domain with significantly reduced mannose binding capability (pre-binding conformation) (Dagmara I. Kisiela et al., (2013) while FimC stabilizes FimH in its extended post-binding-like form (Sauer et al., (2016), Nature Communications volume 7, Article number: 10738) (Figure 22). The FimHC complex was expressed as described in Langermann S, et al. Science. 1997 Apr 25;276(5312):607-11.

Evaluation of new linkers

Recombinant proteins production in mammalian cells.

In order to produce FimHDG (i.e., FimH-PGDGN-DG, wherein DG stands for donor strand complementing peptide from FimG) as well as FimHDG-nanoparticles not containing internal or C-terminal repeated His residues, new constructs have been designed inserting different linkers spacing the FimH_DG gene and the nanoparticle (NP) monomer. The FimH-NP mammalian constructs were synthesized by Geneart or Twist as synthetic genes in pCDNA3.4 (LifeTechnologies) vector. All sequences were codon optimized for expression

29 Aug 2024
2021392894

in mammalian cells and contained an N-terminal leader sequence for secretion into the cells medium. This sequence is the IgK murine leader sequence METDTLLLWVLLLWVPGSTG, or the IgK murine leader sequence followed by and aspartic residue METDTLLLWVLLLWVPGSTGD in order to evaluate the contribute of this residue to efficient protein secretion. To produce recombinant FimH-NPs, the expression vectors were transfected into Expi293 cells and\or ExpiCHO cells according to the manufacturer instructions (Life Technologies) and culture supernatants were collected after 5 days of transfection. Protein purification was achieved by an ion exchange chromatography followed by a preparative SEC purification step.

NanoDSF analysis

To assess the fluorescence-monitored unfolding of the FimHDG constructs a nano-DSF analysis was performed. Samples were manually loaded into nano-DSF grade standard capillaries in triplicates and transferred to a Prometheus NT.48 nano-DSF device. For intrinsic tryptophan fluorescence measurements, the excitation wavelength of 280 nm was used, and the emission of tryptophan fluorescence was measured at 330 nm, 350 nm, and their ratios (350 nm/330 nm). Data were analyzed with Prometheus PR. Control software (NanoTemper Technologies) and plotted using the fluorescence ratio against the temperature.

SPR analysis

The FimHDG constructs were diluted with running buffer HBS-EP+ (0.01 M HEPES, 0.15 M NaCl, 0.003 M EDTA and 0.05% v/v Surfactant P20) and captured on the surface of a sensor chip NTA that was previously activated by injecting a 0.5 mM solution of Ni2+ ions and washed with 3mM of EDTA. mAbs were captured at concentration of 20 ug/ml on the surface of a CM5 sensor coated with secondary anti-mouse IgG Fc. A 50 nM fixed concentration of each sample was injected on the surface of the sensor chip for 180 sec. The dissociation followed for 600 sec. Finally, the sensor chip was regenerated using 10 mM Glycine-HCl pH 1.7. The experiments were performed using a Biacore T200 Instrument (GE Healthcare) and analysed with Biacore T200 Evaluation software 3.0 (GE Healthcare).

Results:

The full length FimH-DG stabilized tagless protein containing a secretion murine Ig-κ chain leader sequence alone and fused to protein NPs (ferritin) were used to transfect EXPI293 and ExpiCHO cells. The accumulation of secreted recombinant protein was characterized by measuring their expression in culture supernatants 5 days post-transfection by SDS-PAGE. The analysis revealed that FimHDG soluble expression could be obtained at high level for the tagless construct (figure A) in Expi293 cells and ExpiCHO. The proteins were further purified from the culture supernatants and biochemically characterized in comparison with the previously purified His-Tagged FimHDG and bacterial refolded FimHDG. FimHDG tagless was obtained with good purity level in SDS-Page and SE-UPLC from both EXpi293 and ExpiCHO cells. The proteins run with a higher molecular

29 Aug 2024 2021392894

weight in SDS-Page (around 42 kD) vs the theoretical one (31 kD) due to glycosylation occurring in mammalian cells, compared to bacterial cells (Fig. 23 A).

The folding of the Tagless purified FimDG was analysed by nano-DSF and melting temperature were obtained and compared with the one obtained for FimHDG-HIS. FimH-DG showed a good thermal stability in Nano-DSF with 2 thermal transitions, relative to lectin (Tm1) and pilin (Tm2) domains, while the His-tag FimHDF molecule shows only one transition, probably due to a different folding. Moreover, tag-less proteins showed higher stability (higher melting temperatures values) of pilin domain transition respect the His-tag molecules. Fig. 23 B. This different folding in the His tagged construct compared to the tagless FimH DG is probably due to the absence of both N-terminal aspartic residue and C-terminal His-Tag.

SPR analysis (Fig. 23 C) of mammalian produced FimHDG tagless constructs show that mAb 926 can bind to the constructs with differences in the binding profile compared to the his-tag FimHDG protein. Moreover, the tagless FimHDG proteins show weak interactions with mAb VH_475 and mannose, on the contrary of the His-tagged FimHDG, in agreement with the different folding observed for the tagless construct in comparison with the His-tagged protein.

For the production of tagless FimHDG-ferritin NPs, the His tag has been replaced by different linkers to separate the FimHDG molecule and the nanoparticle monomer sequence. The linkers designed and tested are made of flexible residues like glycine and serine so that the connected protein domains are free to move relative to one another. We tested different length of linkers, longer linkers can ensure that two adjacent domains do not sterically interfere with one another, but culd be more susceptible to degradation. The linker AKFVAAWTLKAAA, also known as Pan HLA DR-binding epitope (PADRE) is a peptide that activates antigen specific-CD4+ T cells, which has been proposed as a carrier epitope suitable for use in the development of synthetic and recombinant vaccines. The linkers GGGGSLVPRGSGGGGS and EAAAKEAAAKEAAAKA are rigid linkers. The linker AEAAAKEAAAKEAAAKA stabilized by Glu-Lys salt bridges, forms an alfa helix structure (Marqusee & Baldwin, 1987). As the tagless FimHDG and His-tagged FimHDG differ also for the initial aspartic residue, some of the linkers were also tested in absence and in presence of N-terminal aspartic residue. The plasmids coding for the different constructs were used for Expi293 transfection. After 5 days of transfection only constructs starting with N-terminal Aspartic residue (D) (tagless or his-tagged) show a band of secreted protein in the supernatant visible by SDS-page (fig. 23 D). Constructs FimHDG_HIS_Ferritin 1619SI and 1042SI have the same sequence except for initial aspartic residue, but only the construct 1042SI is secreted and present in the culture supernatants of EXPI, confirming the importance of this residue at the N-terminus of FimHDG to achieve efficient FimHDG-ferritin nanoparticles secretion. Among the different linkers tested, only the constructs FimHDG-ferritin tagless 1623SI and 1627SI, from E.coli strain J96 and 536, which have initial aspartic residue, resulted to be secreted. A wester blotting analysis was also performed in order to assess

20213928894 29 Aug 2024

the expression of the tagless FimHDG-ferritin 1433SI in absence of initial Asp residue, confirming that the protein is expressed in the pellet fraction while is not present only in the culture supernatant.

E. Coli Ferritin In Silico Stability Studies

Material and Methods

Evolutionary Constraints for the Sequence and Structure-Based Design of E. coli ferritin

The goal of this research was to perform the design of symmetric systems such as self-assembling protein nanoparticles. This approach introduced stabilizing mutations using a combination of computational physics-based algorithms and evolutionary bioinformatics. To achieve this aim, consensus sequence design was performed on the asymmetric unit or monomer of *E. coli* ferritin (PDB: 1EUM WorldWideWeb(www).rcsb.org/structure/1EUM), using the Rosetta suite (Alford RF, et al. J Chem Theory Comput. 2017 Jun 13;13(6):3031-3048) for thermodynamic design, and non-redundant evolutionary homologs (PSI-BLAST, Altschul SF, et al. Nucleic Acids Res. 1997 Sep 1;25(17):3389-402) to limit mutation space. The designed models were constrained within a symmetric framework (DiMaio F, et al PLoS One. 2011;6(6):e20450), in order to optimize the energetics of protein subunits at geometric interfaces. This symmetry-based pipeline was then implemented in a modified version of the structural bioinformatics tool, PROSS (Goldenzweig A, et al. Mol Cell. 2016 Jul 21;63(2):337-346), yielding a list of *in silico* stabilized sequences (SEQ ID NO: 149-152 and **Fig. 24**).

Protein expression and purification

Genes coding for the different mutants of *E. coli* stabilized ferritin and wild type ferritin were cloned into the pET15TEV vector, which contained an N-terminal 6XHis-tag and a TEV cleavage site. Plasmids encoding for the different constructs were transformed into *E. coli* BL21DE3t1r competent cells. For protein expression, the cells were grown at 20°C in HTMC ON and induced at 20°C with 1mM IPTG for 24 hours. The soluble proteins were extracted by chemical lysis using CeLytic™ Express (Sigma Aldrich) and purified by a nickel chelating column, followed by preparative size exclusion chromatography using a Superdex 200® Increase 10/300 GL column (Cytiva), with purity confirmed by SDS-PAGE (**Fig. 25**).

Transmission Electron Microscopy (TEM) analysis

For negative staining: 5 µl of samples (diluted in 20 ng/microliter) were loaded for 30 seconds onto a glow-discharged copper 300-square mesh grid. After blotting the excess, the grid was negatively stained using Nano-W stain (Ted Pella, Inc) for 30 seconds. The samples were analyzed using a Tecnai G2 spirit and images were acquired using a Veleta CCD (**Fig. 26**).

ThermoFluor Assay

20213928894 29 Aug 2024

The ThermoFluor assay is a quick, temperature-based assay to assess the stability of proteins. In this method, each sample is diluted to a final concentration of 0.2 mg/ml, with an additional 4 µl of SYPRO Orange dye 1000X (Molecular Probes) to reach a final volume of 40 µl using buffer solution. This mix was pipetted into the wells of a 96-well thin-wall PCR plate (Bio-Rad), with water added to control samples. Each sample was analyzed in triplicate. The melting point (T_m) of each protein was determined by ramping from 25°C to 100°C with a scan-rate increment of 1°C per min, taking a fluorescence measurement at each 1°C step. The unfolding profile and melting temperature were monitored by a quantitative PCR thermo cycler (Stratagene). All DSF experiments were performed in triplicate. The derivates of fluorescence intensities were plotted as a function of temperature and the reported T_m is the inflection point of the sigmoid curve determined using GraphPad Prism software (**Fig. 27**).

Results

Recombinant production of an in silico stabilized E. coli ferritin nanoparticle

To obtain a stabilized nanoparticle from *E. coli* that is presenting an *E. coli* stabilized specific antigen (FimHDG, i.e., FimH-PGDGN-DG, wherein DG stands for donor strand complementing peptide from FimG), a native ferritin scaffold for the repetitive display of FimH was selected and computationally optimized. The Rosetta-based design approach maintained octahedral symmetry and focused on the interface between the monomer and the 23 other chains in the symmetric system (**Fig. 24**). This strategy of having a stabilized ferritin from *E. coli* that is presenting an *E. coli* specific antigen (FimH), is a rational approach for maintaining species or genus-specific designs by using a native scaffold for the repetitive display of an antigen.

The *E. coli* WT ferritin and four of the mutants, representative of all the *in silico* stabilized sequences generated by PROSS (SEQ ID NO: 149-152), were highly expressed and soluble when produced as recombinant His-tagged proteins in an *E. coli* cell line (**Fig. 25**). The constructs were successfully purified with an affinity purification step, followed by preparative size exclusion chromatography. The peak corresponding to the high molecular weight fraction was collected for all the constructs and further analyzed with electron microscopy to assess the correct formation of homogeneous and well-structured nanoparticles. From the TEM analysis, all the samples resulted in correctly folded ferritin nanoparticles, except for mutant 2.5 which had a non-uniform morphology (**Fig. 26**).

To identify the most stable *E. coli* ferritin nanoparticle, the thermal stability of recombinant ferritin constructs (WT, 0.5, 2, 6) was assessed by differential scanning fluorimetry (DSF) using Sypro Orange, which binds to hydrophobic residues and detects their exposure during protein unfolding. The ferritin proteins showed very high thermal stability, as expected for a protein nanocage, with the first unfolding transition being detected around 74°C-76°C. This DSF analysis demonstrated that the *E. coli* mutant (0.5) protein exhibited the highest shift in thermal unfolding, leading to its selection as the preferred construct to be fused with the FimHDG antigen, based on this increase in stability.

20213928894 29 Aug 2024

Mammalian production of *E. coli* stabilized ferritin displaying the FimHDG antigen

To test if the stabilized and native ferritin nanoparticle could be used as a scaffold for the display of FimHDG antigen (i.e., FimH-PGDGN-DG, wherein DG stands for donor strand complementing peptide from FimG), and as an alternative to *H. pylori* ferritin, the sequence of FimHDG (containing the secretion sequence Igk) was genetically fused to the gene of the stabilized ferritin (mutant 0.5). The two molecules were separated by a linker containing a repeated histidine sequence to allow for affinity purification of the recombinant secreted nanoparticles in mammalian cell culture supernatant. This construct was used for transfection of Expi293 Gnti cells, and the accumulation of secreted recombinant protein was characterized by assessing the expression in culture supernatants 5 days post-transfection by western blotting analysis, using anti-His antibody. The analysis revealed that FimHDG-ferritin (mutant 0.5) nanoparticles were successfully secreted in the cell supernatant. The purified FimHDG-ferritin (mutant 0.5) nanoparticles were visualized by transmission electron microscopy, confirming the correct morphology of the ferritin stabilized nanoparticles and the surface display of the FimHDG antigen with a size of around 20 nm (**Fig. 28**).

This data indicates that a stabilized *E. coli* ferritin nanoparticle displaying FimHDG can be successfully produced in mammalian cells, indicating that it is possible to design nanoparticles with antigens and scaffolds that are both native to the target pathogen.

2021392894 29 Aug 2024

Table 1(A): Bacterial tested FimH-NP

RIMS Code	Protein Name	Vector Name	Expected AA sequence	Tag
1097SI	FIMH_DG_PGDGN_HIS-Ferritin536	pTrcHis2A	MFACKTANGTAIPIGGGSANVYVNLAPAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSSFSGTVKYNGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVYCAKSQNLGYYLSGTTADAGNSIFTNTASFSPAQGVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTANYARTGGQVTAGNVQSIIGVTFVYQPGDGNADVTITVNGKVVAKSGSHHHHHHHHGGSDIIKLLNEQVNKEMNSSNLYMSMSSWCYTHSLDGAGLFLFDHAAEEYEHAKKLIIFLNENNVPVQLTSISAPEHKFEGLTQIFQKAYEHEQHISESINNIVDHAIKSKDHATFNFLQWYVAEQHEEEVLFKDILDKIELIGNENHGLYLADQYVKGIAKSRK **[SEQ ID NO: 20]**	internal-His
1064SI	LS-FIMHL-IMX313-HIS	pET21	MKYLLPTAAAGLLLLAAQPAMAFacktangtaipigggsanvyvnlapvvnvgqnlvvdlstqifchndypetitdyvtlqrgsayggvlsnfsgtvkysgssypfpttsetprvvynsrtdkpwpvalyltpvssaggvaikagsliavlilrqtnnynsddfqfvwniyanndvvvptggSSGSGSGSKKQGDADVCGEVAYIQSVVSDCHVPTAELRTLLEIRKLFLEIQKLKVELQGLSKEGGGSGSHHHHHHHH **[SEQ ID NO: 21]**	C-His
955SI	FIMHL-S24S65-IMX313	pTrcHis2A	MfaSktangtaipigggsanvyvnlapvvnvgqnlvvdlstqifShndypetitdyvtlqrgsayggvlsnfsgtvkysgssypfpttsetprvvynsrtdkpwpvalyltpvssaggvaikagsliavlilrqtnnynsddfqfvwniyanndvvvptggSSGSGSGSKKQGDADVCGEVAYIQSVVSDCHVPTAELRTLLEIRKLFLEIQKLKVELQGLSKEGGGSGSHHHHHH **[SEQ ID NO: 22]**	C-His
954SI	FIMHL-S24S65-foldon-ferritin	pTrcHis2A	MFASKTANGTAIPIGGGSANVYVNLAPAVNVGQNLVVDLSTQIFSHNDYPETITDYVTLQRGSAYGGVLSSFSGTVKYNGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGSGYIPEAPRDGQAYVRKDGEWVLLSTFLGSGHHHHHHGSGDIIKLLNEQVNKEMNSSNLYMSMSSWCYTHSLDGAGLFLFDHAAEEYEHAKKLIIFLNENNVPVQLTSISAPEHKFEGLTQIFQKAYEHEQHISESINNIVDHAIKSKDHATFNFLQWYVAEQHEEEVLFKDILDKIELIGNENHGLYLADQYVKGIAKSRK **[SEQ ID NO: 23]**	internal
940SI	FIMHL-S24S65-Mi3	pTrcHis2A	MFASKTANGTAIPIGGGSANVYVNLAPAVNVGQNLVVDLSTQIFSHNDYPETITDYVTLQRGSAYGGVLSSFSGTVKYNGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGSGGSGGSMKMEELFKKHKIVAVLRANSVEEAKKKALAVFLGGVHLIEITFTVPDADTVIKELSFLKEMGAIIGAGTVTSVEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVNLDNVCEWFKAGVLAVGVGSALVKGTPVEVAEKAKAFVEKIRGCTEGSGSGSGSGSHHHHHH **[SEQ ID NO: 24]**	C-His
939SI	FIMHL-mI3	pTrcHis2A	MFACKTANGTAIPIGGGSANVYVNLAPAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSSFSGTVKYNGSSYPFPTTSETPRVVYNSR	C-His

2021392894 29 Aug 2024

			TDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFVWNI YANNDVVVPTGGSGGSGGSMKMEELFKKHKIVAVLRANSVEEAKKKAL AVFLGGVHLIEITFTVPDADTVIKELSFLKEMGAIIGAGTVTSVEQARKAVE SGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTELVKAMKLGHTILKLF PGEVVGPQFVKAMKGPFPNVKFVPTGGVNLDNVCEWFKAGVLAVGV GSALVKGTPVEVAEKAKAFVEKIRGCTEGSGSGSGSGSHHHHHH **[SEQ ID NO: 25]**	
913SI	FimHL-NOCYS-MI3	pET21	MfaSktangtaipigggsanvyvnlapavnvgqnlvvdlstqifShndypetitdyvtlqr gsayggvlsnfsgtvkysgssypfpttsetprvvynsrtdkpwpvalyltpvssaggvaika gsliavlilrqtnnynsddfqfvwniyanndvvvptggGGSGGSGGSGGSMKMEE LFKKHKIVAVLRANSVEEAKKKALAVFLGGVHLIEITFTVPDADTVIKELSFL KEMGAIIGAGTVTSVEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYM PGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPT GGVNLDNVCEWFKAGVLAVGVGSALVKGTPVEVAEKAKAFVEKIRGCT EGSGSGSGSGSHHHHHH **[SEQ ID NO: 26]**	C-His
904SI	FimHdeltaGG_PGDGN_DG_mi3	pET21	Mfacktangtaipigggsanvyvnlapvvnvgqnlvvdlstqifchndypetitdyvtlqr gsayggvlsnfsgtvkysgssypfpttsetprvvynsrtdkpwpvalyltpvssaggvaika gsliavlilrqtnnynsddfqfvwniyanndvvvptcdvsardvtvtlpdypgsvpipltvy caksqnlgyylsgttadagnsiftntasfspaqgvgvqltrngtiipanntvslgavgtsavsl gltanyartggqvtagnvqsiigvtfvyqPGDGNADVTITVNGKVVAKGSGGGG MKMEELFKKHKIVAVLRANSVEEAKKKALAVFLGGVHLIEITFTVPDADT VIKELSFLKEMGAIIGAGTVTSVEQARKAVESGAEFIVSPHLDEEISQFAKE KGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPN VKFVPTGGVNLDNVCEWFKAGVLAVGVGSALVKGTPVEVAEKAKAFVE KIRGCTEGSGSGSGSGSHHHHHH **[SEQ ID NO: 27]**	C-His
888SI	FimHL-GSG4-Ferritin	pET15 TEV	MGSSHHHHHHENLYFQGFACKTANGTAIPIGGGSANVYVNLAPAVNV GQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSSFSGTVKYNGS SYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILR QTNNYNSDDFQFVWNIYANNDVVVPTGSGGGGDIIKLLNEQVNKEMN SSNLYMSMSSWCYTHSLDGAGLFLFDHAAEEYEHAKKLIIFLNENNVPV QLTSISAPEHKFEGLTQIFQKAYEHEQHISESINNIVDHAIKSKDHATFNFL QWYVAEQHEEEVLFKDILDKIELIGNENHGLYLADQYVKGIAKSRK **[SEQ ID NO: 28]**	N-His
887SI	pelBLS-FimHL-mI3	pET22	MKYLLPTAAAGLLLLAAQPAMAFacktangtaipigggsanvyvnlapvvnvgq nlvvdlstqifchndypetitdyvtlqrgsayggvlsnfsgtvkysgssypfpttsetprvvy nsrtdkpwpvalyltpvssaggvaikagsliavlilrqtnnynsddfqfvwniyanndvvv ptggGSGMKMEELFKKHKIVAVLRANSVEEAK**K**KALAVF**L**GGVHLIEITFT VPDADTVIKELSFLKE**M**GAIIGAGTVTSVEQ**A**RKAVESGAEFIVSPHLDEE ISQF**A**KEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKA MKGPFPNVKFVPTGGVNLDNVCEWFKAGVLAVGVGSALVKGTP**V**EV**A** EKAKAFVEKIRGCTEGSGSGSGS**HHHHHH [SEQ ID NO: 29]**	C-His
837SI	FimH_DG_Ferrit	pET15 TEV	MGSSHHHHHHENLYFQGDVVVPTGGCDVSARDVTVTLPDYPGSVPIPL TVYCAKSQNLGYYLSGTTADAGNSIFTNTASFSPAQGVGVQLTRNGTIIP	N-His

2021392894 29 Aug 2024

	in (GSGGGG)		ANNTVSLGAVGTSAVSLGLTANYARTGGQVTAGNVQSIIGVTFVYQPGD GNADVTITVNGKVVAKGSGGGGDIIKLLNEQVNKEMNSSNLYMSMSS WCYTHSLDGAGLFLFDHAAEEYEHAKKLIIFLNENNVPVQLTSISAPEHKF EGLTQIFQKAYEHEQHISESINNIVDHAIKSKDHATFNFLQWYVAEQHEE EVLFKDILDKIELIGNENHGLYLADQYVKGIAKSRK **[SEQ ID NO: 30]**	
836SI	FimHL-C-C-MI3	pET21	MfacktangtaipigggsanvyvnlapvCnvgqnCvvdlstqifchndypetitdyvtlq rgsayggvlsnfsgtvkysgssypfpttsetprvvynsrtdkpwpvalyltpvssaggvaik agsliavlilrqtnnynsddfqfvwniyanndvvvptggGGSGGSGGSGGSMKME ELFKKHKIVAVLRANSVEEAKKKALAVFLGGVHLIEITFTVPDADTVIKELSF LKEMGAIIGAGTVTSVEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFY MPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVP TGGVNLDNVCEWFKAGVLAVGVGSALVKGTPVEVAEKAKAFVEKIRGC TEGSGSGSGSGSHHHHHH **[SEQ ID NO: 31]**	C-His
835SI	FimHL-C-C-qBeta	pET21	MfacktangtaipigggsanvyvnlapvCnvgqnCvvdlstqifchndypetitdyvtlq rgsayggvlsnfsgtvkysgssypfpttsetprvvynsrtdkpwpvalyltpvssaggvaik agsliavlilrqtnnynsddfqfvwniyanndvvvptgg**GGSGGSGGSGGS**AKLET VTLGNIGKDGKQTLVLNPRGVNPTNGVASLSQAGAVPALEKRVTVSVSQ PSRNRKNYKVQVKIQNPTACTANGSCDPSVTRQAYADVTFSFTQYSTDE ERAFVRTELAALLASPLLIDAIDQLNPAY **[SEQ ID NO: 32]**	Tagless

Table 1(B): FimH espressed as single recombinant protein in *E. coli*

RIMS Code	**Protein Name**	**Vector Name**	**Expected AA sequence**	**Tag**
1023SI	FimH_DNKQ_DG citopl pET22b+	pET22b+	MFACKTANGTAIPIGGGSANVYVNLAPVVNVGQNLVVDLSTQIFCHND YPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSR TDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFVWNI YANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVYCAKSQNLGYYL SGTTADAGNSIFTNTASFSPAQGVGVQLTRNGTIIPANNTVSLGAVGTSA VSLGLTANYARTGGQVTAGNVQSIIGVTFVYQDNKQADVTITVNGKVV AKGSGHHHHHH **[SEQ ID NO: 120]**	C-His
1024SI	FimH_PGDGN_DG citopl pET22b+	pET22b+	MFACKTANGTAIPIGGGSANVYVNLAPVVNVGQNLVVDLSTQIFCHND YPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSR TDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFVWNI YANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVYCAKSQNLGYYL SGTTADAGNSIFTNTASFSPAQGVGVQLTRNGTIIPANNTVSLGAVGTSA VSLGLTANYARTGGQVTAGNVQSIIGVTFVYQPGDGNADVTITVNGKV VAKGSGHHHHHH **[SEQ ID NO: 121]**	C-His
1025SI	FimH_DGG_PGDGN_DG citopl	pET22b+	MFACKTANGTAIPIGGGSANVYVNLAPVVNVGQNLVVDLSTQIFCHND YPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSR TDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFVWNI YANNDVVVPTCDVSARDVTVTLPDYPGSVPIPLTVYCAKSQNLGYYLSGT	C-His

2021392894 29 Aug 2024

	pET22b+		TADAGNSIFTNTASFSPAQGVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTANYARTGGQVTAGNVQSIIGVTFVYQPGDGNADVTITVNGKVVAKGSGHHHHHH **[SEQ ID NO: 122]**	
1122SI	FimH_PGDGN_DG citopl PET24 Tagless	pET24b(+)	MFACKTANGTAIPIGGGSANVYVNLAPVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVYCAKSQNLGYYLSGTTADAGNSIFTNTASFSPAQGVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTANYARTGGQVTAGNVQSIIGVTFVYQPGDGNADVTITVNGKVVAK **[SEQ ID NO: 123]**	Tagless
	FimH_PGDGN_DG citopl tagless no Met		FACKTANGTAIPIGGGSANVYVNLAPVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVYCAKSQNLGYYLSGTTADAGNSIFTNTASFSPAQGVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTANYARTGGQVTAGNVQSIIGVTFVYQPGDGNADVTITVNGKVVAK **[SEQ ID NO: 124]**	tagless

Table 2: Mammalian-expressed FimH as single recombinant proteins and Nanoparticles:

RIMS Code	**Protein Name**		**Expected AA sequence**	**expression**	**Tag**
1096 SI	FIMHL-HIS-Mi3 J96	short leader	METDTLLLWVLLLWVPGSTGDFACKTANGTAIPIGGGSANVYVNLAPVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGGSGSHHHHHHHHGGSMKMEELFKKHKIVAVLRANSVEEAKKKALAVFLGGVHLIEITFTVPDADTVIKELSFLKEMGAIIGAGTVTSVEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVNLDNVCEWFKAGVLAVGVGSALVKGTPVEVAEKAKAFVEKIRGCTE **[SEQ ID NO: 33]**	yes	internal HIS
1095 SI	FIMHL-HIS-Fer 536	short leader	METDTLLLWVLLLWVPGSTGDFACKTANGTAIPIGGGSANVYVNLAPAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSSFSGTVKYNGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGGSGSHHHHHHHGGSDIIKLLNEQVNKEMNSSNLYMSMSSWCYTHSLDGAGLFLFDHAAEEYEHAKKLIIFLNENNVPVQLTSISAPEHKFEGLTQIFQKAYEHEQHISESINNIVDHAIKSKDHATFNFLQWYVAEQHEEEVLFKDILDKIELIGNENHGLYLADQYVKGIAKSRK **[SEQ ID NO: 34]**	yes	internal HIS

2021392894 29 Aug 2024

1043 SI	FIMH_DG_PGDGN_IMX313_HISJ96	short leader	METDTLLLWVLLLWVPGSTGDFACKTANGTAIPIGGGSANVYVNLA PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFS GTVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAI KAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSAR DVTVTLPDYPGSVPIPLTVYCAKSQNLGYYLSGTTADAGNSIFTNTASF SPAQGVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTANYARTGG QVTAGNVQSIIGVTFVYQPGDGNADVTITVNGKVVAKGSSGSGSGS KKQGDADVCGEVAYIQSVVSDCHVPTAELRTLLEIRKLFLEIQKLKVEL QGLSKEGGGSGSHHHHHH **[SEQ ID NO: 35]**	yes	C-His
1042 SI	FimH_PGDGN_DG_HIS-Ferritnj96	short leader	METDTLLLWVLLLWVPGSTGDFACKTANGTAIPIGGGSANVYVNLA PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFS GTVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAI KAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSAR DVTVTLPDYPGSVPIPLTVYCAKSQNLGYYLSGTTADAGNSIFTNTASF SPAQGVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTANYARTGG QVTAGNVQSIIGVTFVYQPGDGNADVTITVNGKVVAKSGSHHHHH HHHGGSDIIKLLNEQVNKEMNSSNLYMSMSSWCYTHSLDGAGLFLF DHAAEEYEHAKKLIIFLNENNVPVQLTSISAPEHKFEGLTQIFQKAYEH EQHISESINNIVDHAIKSKDHATFNFLQWYVAEQHEEEVLFKDILDKIE LIGNENHGLYLADQYVKGIAKSRK **[SEQ ID NO: 36]**	yes	internal HIS
1142 SI	FIMH_DG_PGDGN-HIS-Ferritin536	short leader	METDTLLLWVLLLWVPGSTGDFACKTANGTAIPIGGGSANVYVNLA PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSSFS GTVKYNGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAI KAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSAR DVTVTLPDYPGSVPIPLTVYCAKSQNLGYYLSGTTADAGNSIFTNTASF SPAQGVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTANYARTGG QVTAGNVQSIIGVTFVYQPGDGNADVTITVNGKVVAKSGSHHHHH HHHGGSDIIKLLNEQVNKEMNSSNLYMSMSSWCYTHSLDGAGLFLF DHAAEEYEHAKKLIIFLNENNVPVQLTSISAPEHKFEGLTQIFQKAYEH EQHISESINNIVDHAIKSKDHATFNFLQWYVAEQHEEEVLFKDILDKIE LIGNENHGLYLADQYVKGIAKSRK **[SEQ ID NO: 37]**	yes	internal HIS
1000 SI	FimH_PGDGN_DG-HIS-Ferritin J96	EXTRA AA	METDTLLLWVLLLWVPGSTGDAAQPARRARRTKLALFACKTANGTA IPIGGGSANVYVNLAPVVNVGQNLVVDLSTQIFCHNDYPETITDYVTL QRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKPWPV ALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFVWNIYANN DVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVYCAKSQNLGYYLSG TTADAGNSIFTNTASFSPAQGVGVQLTRNGTIIPANNTVSLGAVGTS AVSLGLTANYARTGGQVTAGNVQSIIGVTFVYQPGDGNADVTITVN GKVVAKSGSHHHHHHHHGGSDIIKLLNEQVNKEMNSSNLYMSMSS WCYTHSLDGAGLFLFDHAAEEYEHAKKLIIFLNENNVPVQLTSISAPE HKFEGLTQIFQKAYEHEQHISESINNIVDHAIKSKDHATFNFLQWYVA EQHEEEVLFKDILDKIELIGNENHGLYLADQYVKGIAKSRK **[SEQ ID NO: 38]**	NO	internal HIS

2021392894 29 Aug 2024

999SI	FimH_PGDGN_DG-HIS-MI3 j96	EXTRAAA	METDTLLLWVLLLWVPGSTGDAAQPARRARRTKLALFACKTANGTAIPIGGGSANVYVNLAPVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVYCAKSQNLGYYLSGTTADAGNSIFTNTASFSPAQGVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTANYARTGGQVTAGNVQSIIGVTFVYQPGDGNADVTITVNGKVVAKSGSHHHHHHHHGGSMKMEELFKKHKIVAVLRANSVEEAKKKALAVFLGGVHLIEITFTVPDADTVIKELSFLKEMGAIIGAGTVTSVEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVNLDNVCEWFKAGVLAVGVGSALVKGTPVEVAEKAKAFVEKIRGCTE **[SEQ ID NO: 39]**	NO	internal HIS
998SI	FimH_PGDGN_DG-HIS-IMX313 j96	EXTRAAA	METDTLLLWVLLLWVPGSTGDAAQPARRARRTKLALFACKTANGTAIPIGGGSANVYVNLAPVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVYCAKSQNLGYYLSGTTADAGNSIFTNTASFSPAQGVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTANYARTGGQVTAGNVQSIIGVTFVYQPGDGNADVTITVNGKVVAKGSSGSGSGSKKQGDADVCGEVAYIQSVVSDCHVPTAELRTLLEIRKLFLEIQKLKVELQGLSKEGGGSGSHHHHHH **[SEQ ID NO: 40]**	yes	C_his
995SI	FimH_PGDGN_DG_Ferritin (536)	EXTRAAA	METDTLLLWVLLLWVPGSTGDAAQPARRARRTKLALFACKTANGTAIPIGGGSANVYVNLAPAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSSFSGTVKYNGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVYCAKSQNLGYYLSGTTADAGNSIFTNTASFSPAQGVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTANYARTGGQVTAGNVQSIIGVTFVYQPGDGNADVTITVNGKVVAKSGSHHHHHHHHGGSDIIKLLNEQVNKEMNSSNLYMSMSSWCYTHSLDGAGLFLFDHAAEEYEHAKKLIIFLNENNVPVQLTSISAPEHKFEGLTQIFQKAYEHEQHISESINNIVDHAIKSKDHATFNFLQWYVAEQHEEEVLFKDILDKIELIGNENHGLYLADQYVKGIAKSRK **[SEQ ID NO: 41]**	yes	internal HIS
936SI	FimH-IMX313 j96	EXTRAAA	METDTLLLWVLLLWVPGSTGDAAQPARRARRTKLALfacktangtaipigggsanvyvnlapvvnvgqnlvvdlstqifchndypetitdyvtlqrgsayggvlsnfsgtvkysgssypfpttsetprvvynsrtdkpwpvalyltpvssaggvaikagsliavlilrqtnnynsddfqfvwniyanndvvvptcdvsardvtvtlpdypgsvpipltvycaksqnlgyylsgttadagnsiftntasfspaqgvgvqltrngtiipanntvslgavgtsavslgltanyartggqvtagnvqsiigvtfvyqGSSGSGSGSKKQGDADVCGEVAYIQSVVSDCHVPTAELRTLLEIRKLFLEIQKLKVELQGLSKEGGGSGSHHHHHH **[SEQ ID NO: 42]**	NO	C-His

2021392894 29 Aug 2024

935SI	FimH_mi3 j96	EXTRAAA	METDTLLLWVLLLWVPGSTGDAAQPARRARRTKLALfacktangtaipigggsanvyvnlapvvnvgqnlvvdlstqifchndypetitdyvtlqrgsayggvlsnfsgtvkysgssypfpttsetprvvynsrtdkpwpvalyltpvssaggvaikagsliavlilrqtnnynsddfqfvwniyanndvvvptcdvsardvtvtlpdypgsvpipltvycaksqnlgyylsgttadagnsiftntasfspaqgvgvqltrngtiipanntvslgavgtsavslgltanyartggqvtagnvqsiigvtfvyqGSGGGGMKMEELFKKHKIVAVLRANSVEEAKKKALAVFLGGVHLIEITFTVPDADTVIKELSFLKEMGAIIGAGTVTSVEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVNLDNVCEWFKAGVLAVGVGSALVKGTPVEVAEKAKAFVEKIRGCTEGSGSGSGSGSHHHHHHHH **[SEQ ID NO: 43]**	NO	C-His
929SI	FIMHL-HIS-mI3 j96	EXTRAAA	METDTLLLWVLLLWVPGSTGDAAQPARRARRTKLALFACKTANGTAIPIGGGSANVYVNLAPVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGSGGHHHHHHHHGSGSMKMEELFKKHKIVAVLRANSVEEAKKKALAVFLGGVHLIEITFTVPDADTVIKELSFLKEMGAIIGAGTVTSVEQARKAVESGAEFIVSPHLDEEISQFAKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVNLDNVCEWFKAGVLAVGVGSALVKGTPVEVAEKAKAFVEKIRGCTE **[SEQ ID NO: 44]**	NO	internal HIS
951SI	FimH_DNKQ_DG_deglyc	EXTRAAA	METDTLLLWVLLLWVPGSTGDAAQPARRARRTKLALFACKTASGTAIPIGGGSANVYVNLAPVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSDFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVYCAKSQNLGYYLSGTTADAGNSIFTNTASFSPAQGVGVQLTRDGTIIPADNTVSLGAVGTSAVSLGLTANYARTGGQVTAGNVQSIIGVTFVYQDNKQADVTITVNGKVVAKGSGHHHHHH* **[SEQ ID NO: 79]**	yes	C-His
932SI	FimH_PGDGN_DG	EXTRAAA	METDTLLLWVLLLWVPGSTGDAAQPARRARRTKLALFACKTANGTAIPIGGGSANVYVNLAPVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVYCAKSQNLGYYLSGTTADAGNSIFTNTASFSPAQGVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTANYARTGGQVTAGNVQSIIGVTFVYQPGDGNADVTITVNGKVVAKGSGHHHHHH* **[SEQ ID NO: 80]**	yes	C-His
931SI	FimH_DNKQ_DG	EXTRAAA	METDTLLLWVLLLWVPGSTGDAAQPARRARRTKLALFACKTANGTAIPIGGGSANVYVNLAPVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVYCAKSQNLGYYLSGTTADAGNSIFTNTASFSPAQGVGVQLTRNGTIIPANNTVSLGAVGTS	yes	C-His

2021392894 29 Aug 2024

			AVSLGLTANYARTGGQVTAGNVQSIIGVTFVYQDNKQADVTITVNG KVVAKGSGHHHHHH***[SEQ ID NO: 81]**		
930SI	FimH_DeltaGG_PGDGN_DG	EXTRAAA	METDTLLLWVLLLWVPGSTGDAAQPARRARRTKLALFACKTANGTA IPIGGGSANVYVNLAPVVNVGQNLVVDLSTQIFCHNDYPETITDYVTL QRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKPWPV ALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFVWNIYANN DVVVPTCDVSARDVTVTLPDYPGSVPIPLTVYCAKSQNLGYYLSGTTA DAGNSIFTNTASFSPAQGVGVQLTRNGTIIPANNTVSLGAVGTSAVS LGLTANYARTGGQVTAGNVQSIIGVTFVYQPGDGNADVTITVNGKV VAKGSGHHHHHH***[SEQ ID NO: 82]**	no	C-His
989SI	FimH_DGG_sl	short leader	METDTLLLWVLLLWVPGSTGDFACKTANGTAIPIGGGSANVYVNLA PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFS GTVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAI KAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTCDVSARDV TVTLPDYPGSVPIPLTVYCAKSQNLGYYLSGTTADAGNSIFTNTASFSP AQGVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTANYARTGGQV TAGNVQSIIGVTFVYQPGDGNADVTITVNGKVVAKGSGHHHHHH* ***[SEQ ID NO: 83]**	yes	C-His
988SI	FimH_PGDGN_sl	short leader	METDTLLLWVLLLWVPGSTGDFACKTANGTAIPIGGGSANVYVNLA PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFS GTVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAI KAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSAR DVTVTLPDYPGSVPIPLTVYCAKSQNLGYYLSGTTADAGNSIFTNTASF SPAQGVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTANYARTGG QVTAGNVQSIIGVTFVYQPGDGNADVTITVNGKVVAKGSGHHHHH H***[SEQ ID NO: 84]**	yes	C-His
987SI	FimH_DNKQ_sl	short leader	METDTLLLWVLLLWVPGSTGDFACKTANGTAIPIGGGSANVYVNLA PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFS GTVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAI KAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSAR DVTVTLPDYPGSVPIPLTVYCAKSQNLGYYLSGTTADAGNSIFTNTASF SPAQGVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTANYARTGG QVTAGNVQSIIGVTFVYQDNKQADVTITVNGKVVAKGSGHHHHHH ***[SEQ ID NO: 85]**	yes	C-His
1183SI	FIMH_DG_PGDGN_536-MI3	short leader	METDTLLLWVLLLWVPGSTGDFACKTANGTAIPIGGGSANVYVNLA PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSSFS GTVKYNGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAI KAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSAR DVTVTLPDYPGSVPIPLTVYCAKSQNLGYYLSGTTADAGNSIFTNTASF	yes	internal HIS

2021392894 29 Aug 2024

			SPAQGVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTANYARTGG QVTAGNVQSIIGVTFVYQPGDGNADVTITVNGKVVAKSGSHHHHH HHHGGSMKMEELFKKHKIVAVLRANSVEEAKKKALAVFLGGVHLIEI TFTVPDADTVIKELSFLKEMGAIIGAGTVTSVEQARKAVESGAEFIVSP HLDEEISQFAKEKGVFYMPGVMTPTELVKAMKLGHTILKLFPGEVVG PQFVKAMKGPFPNVKFVPTGGVNLDNVCEWFKAGVLAVGVGSALV KGTPVEVAEKAKAFVEKIRGCTE***[SEQ ID NO: 86]**		
1184 SI	FIMH_DG_PGDGN_536-encapsuline	short leader	METDTLLLWVLLLWVPGSTGDFACKTANGTAIPIGGGSANVYVNLA PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSSFS GTVKYNGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAI KAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSAR DVTVTLPDYPGSVPIPLTVYCAKSQNLGYYLSGTTADAGNSIFTNTASF SPAQGVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTANYARTGG QVTAGNVQSIIGVTFVYQPGDGNADVTITVNGKVVAKSGSHHHHH HHHGGSMEFLKRSFAPLTEKQWQEIDNRAREIFKTQLYGRKFVDVE GPYGWEYAAHPLGEVEVLSDENEVVKWGLRKSLPLIELRATFTLDLW ELDNLERGKPNVDLSSLEETVRKVAEFEDEVIFRGCEKSGVKGLLSFEE RKIECGSTPKDLLEAIVRALSIFSKDGIEGPYTLVINTDRWINFLKEEAG HYPLEKRVEECLRGGKIITTPRIEDALVVSERGGDFKLILGQDLSIGYED REKDAVRLFITETFTFQVVNPEALILLKF***[SEQ ID NO: 87]**	yes	internal HIS
1127 SI	HBcFIMHLJ96	short leader	METDTLLLWVLLLWVPGSTGDDIDPYKEFGASVELLSFLPSDFFPSIR DLLDTASALYREALESPEHCSPHHTALRQAILCWGELMNLATWVGS NLEDPGSGGGGFACKTANGTAIPIGGGSANVYVNLAPVVNVGQNL VVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSY PFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLIL RQTNNYNSDDFQFVWNIYANNDVVVPTGGGSGGASRELVVSYVNV NMGLKIRQLLWFHISCLTFGRETVLEYLVSFGVWIRTPPAYRPPNAPIL STLPETTVVGSGGGGHHHHHH***[SEQ ID NO: 88]**	no	C-His
1126 SI	HBcFIMHDGJ96	short leader	METDTLLLWVLLLWVPGSTGDDIDPYKEFGASVELLSFLPSDFFPSIR DLLDTASALYREALESPEHCSPHHTALRQAILCWGELMNLATWVGS NLEDPGSGGGGFACKTANGTAIPIGGGSANVYVNLAPVVNVGQNL VVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSY PFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLIL RQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYP GSVPIPLTVYCAKSQNLGYYLSGTTADAGNSIFTNTASFSPAQGVGV QLTRNGTIIPANNTVSLGAVGTSAVSLGLTANYARTGGQVTAGNVQ SIIGVTFVYQPGDGNADVTITVNGKVVAKGSGGGGASRELVVSYVNV NMGLKIRQLLWFHISCLTFGRETVLEYLVSFGVWIRTPPAYRPPNAPIL STLPETTVVGSGGGGHHHHHH**[SEQ ID NO: 89]**	no	C-His
	D_FimHDG_Fer_GSG4	GSGGGGG linker	**M**ETDTLLLWVLLLWVPGSTGDFACKTANGTAIPIGGGSANVYVNLA PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFS GTVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAI KAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSAR DVTVTLPDYPGSVPIPLTVYCAKSQNLGYYLSGTTADAGNSIFTNTASF		

2021392894 29 Aug 2024

			SPAQGVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTANYARTGG QVTAGNVQSIIGVTFVYQPGDGNADVTITVNGKVVAKGSGGGGDII KLLNEQVNKE**M**NSSNLY**M**S**M**SSWCYTHSLDGAGLFLFDHAAEEYE HAKKLIIFLNENNVPVQLTSISAPEHKFEGLTQIFQKAYEHEQHISESIN NIVDHAIKSKDHATFNFLQWYVAEQHEEEVLFKDILDKIELIGNENHG LYLADQYVKGIAKSRK--**[SEQ ID NO: 129]**		
	D_FimH DG_Fer 0.5_degl yc_tagle ss	E.coli ferritin 0.5 tagless	**M**ETDTLLLWVLLLWVPGSTGDFACKTAQGTAIPIGGGSANVYVNLA PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSQFS GTVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAI KAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSAR DVTVTLPDYPGSVPIPLTVYCAKSQNLGYYLSGTTADAGNSIFTNTASF SPAQGVGVQLTRQGTIIPAQNTVSLGAVGTSAVSLGLTANYARTGG QVTAGNVQSIIGVTFVYQPGDGNADVTITVNGKVVAKGSGGGGGG **M**LKPE**M**IEKLNEQ**M**NLELYSSLLYQQ**M**SAWCSYHGFEGAAAFLRR HAQEE**M**TH**M**QRLFDYLTDTGNLPRIDTIPSPFAEYSSLDELFQETYKH EQLITQKINELAHAA**M**TNQDYPTFNFLQWYVAEQHEEEKLFKSIIDKL SLAGKSGEGLYFIDKELSTLDTQN----**[SEQ ID NO: 130]**		
	FimHDG _deglyc _tagless	N-->Q mutati on for avoidi ng glycosi lation	**M**ETDTLLLWVLLLWVPGSTGFACKTAQGTAIPIGGGSANVYVNLAP VVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSQFSG TVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIK AGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARD VTVTLPDYPGSVPIPLTVYCAKSQNLGYYLSGTTADAGNSIFTNTASFS PAQGVGVQLTRQGTIIPAQNTVSLGAVGTSAVSLGLTANYARTGGQ VTAGNVQSIIGVTFVYQPGDGNADVTITVNGKVVAK----**[SEQ ID NO: 131]**		
	D_FimH DG_degl yc_tagle ss	Initial D; N-- >Q mutati on for avoidi ng glycosi lation	**M**ETDTLLLWVLLLWVPGSTGDFACKTAQGTAIPIGGGSANVYVNLA PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSQFS GTVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAI KAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSAR DVTVTLPDYPGSVPIPLTVYCAKSQNLGYYLSGTTADAGNSIFTNTASF SPAQGVGVQLTRQGTIIPAQNTVSLGAVGTSAVSLGLTANYARTGG QVTAGNVQSIIGVTFVYQPGDGNADVTITVNGKVVAK----**[SEQ ID NO: 132]**		
	FimHDG _N7Q_t agless		**M**ETDTLLLWVLLLWVPGSTGFACKTAQGTAIPIGGGSANVYVNLAP VVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSG TVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIK AGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARD VTVTLPDYPGSVPIPLTVYCAKSQNLGYYLSGTTADAGNSIFTNTASFS PAQGVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTANYARTGGQ VTAGNVQSIIGVTFVYQPGDGNADVTITVNGKVVAK----**[SEQ ID NO: 133]**		

2021392894 29 Aug 2024

	D_FimHDG_Fer_deglyc_tagless	N-->Q mutation for avoiding glycosilation	**M**ETDTLLLWVLLLWVPGSTGDFACKTAQGTAIPIGGGSANVYVNLAPVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSQFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVYCAKSQNLGYYLSGTTADAGNSIFTNTASFSPAQGVGVQLTRQGTIIPAQNTVSLGAVGTSAVSLGLTANYARTGGQVTAGNVQSIIGVTFVYQPGDGNADVTITVNGKVVAKSGSGSGGGGGGSDIIKLLNEQVNKE**M**QSSNLY**M**S**M**SSWCYTHSLDGAGLFLFDHAAEEYEHAKKLIIFLNENNVPVQLTSISAPEHKFEGLTQIFQKAYEHEQHISESINNIVDHAIKSKDHATFNFLQWYVAEQHEEEVLFKDILDKIELIGNENHGLYLADQYVKGIAKSRK----**[SEQ ID NO: 134]**		
	FimH_PGDGN_DG_GGS4-Ferritnj96	Tagless FimHDG	METDTLLLWVLLLWVPGSTGFACKTANGTAIPIGGGSANVYVNLAPVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVYCAKSQNLGYYLSGTTADAGNSIFTNTASFSPAQGVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTANYARTGGQVTAGNVQSIIGVTFVYQPGDGNADVTITVNGKVVAKGGSGGSGGSGGSDIIKLLNEQVNKEMNSSNLYMSMSSWCYTHSLDGAGLFLFDHAAEEYEHAKKLIIFLNENNVPVQLTSISAPEHKFEGLTQIFQKAYEHEQHISESINNIVDHAIKSKDHATFNFLQWYVAEQHEEEVLFKDILDKIELIGNENHGLYLADQYVKGIAKSRKS* **[SEQ ID NO: 135]**		
1397 SI	FimH_PGDGN_DG j96	Tagless FimHDG	METDTLLLWVLLLWVPGSTGFACKTANGTAIPIGGGSANVYVNLAPVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVYCAKSQNLGYYLSGTTADAGNSIFTNTASFSPAQGVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTANYARTGGQVTAGNVQSIIGVTFVYQPGDGNADVTITVNGKVVAK **[SEQ ID NO: 136]**		
	FIMHDG_PADRE-encapsuline536	linker PADRE	METDTLLLWVLLLWVPGSTGFACKTANGTAIPIGGGSANVYVNLAPAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSSFSGTVKYNGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVYCAKSQNLGYYLSGTTADAGNSIFTNTASFSPAQGVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTANYARTGGQVTAGNVQSIIGVTFVYQPGDGNADVTITVNGKVVAKFVAAWTLKAAAMEFLKRSFAPLTEKQWQEIDNRAREIFKTQLYGRKFVDVEGPYGWEYAAHPLGEVEVLSDENEVVKWGLRKSLPLIELRATFTLDLWELDNLERGKPNVDLSSLEETVRKVAEFEDEVIFRGCEKSGVKGLLSFEERKIECGSTPKDLLEAIVRALSIFSKDGIEGPYTLVINTDRWINFLKEEAGHYPLEKRVEECLRGGKIITTPRIEDALVVSERGGDFKLILGQDLSIGYEDREKDAVRLFITETFTFQVVNPEALILLKF-**[SEQ ID NO: 137]**		

2021392894 29 Aug 2024

	FIMH_NOD-FERRITN_J96	mixed linker with GS and H	METDTLLLWVLLLWVPGSTGFACKTANGTAIPIGGGSANVYVNLAP VVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSG TVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIK AGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARD VTVTLPDYPGSVPIPLTVYCAKSQNLGYYLSGTTADAGNSIFTNTASFS PAQGVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTANYARTGGQ VTAGNVQSIIGVTFVYQPGDGNADVTITVNGKVVAKSGSHHHGSGG GGGSDIIKLLNEQVNKEMNSSNLYMSMSSWCYTHSLDGAGLFLFDH AAEEYEHAKKLIIFLNENNVPVQLTSISAPEHKFEGLTQIFQKAYEHEQ HISESINNIVDHAIKSKDHATFNFLQWYVAEQHEEEVLFKDILDKIELIG NENHGLYLADQYVKGIAKSRK **[SEQ ID NO: 138]**		
	FIMH_DG_NOALFA-ferritin_J96	rigid linker	METDTLLLWVLLLWVPGSTGFACKTANGTAIPIGGGSANVYVNLAP VVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSG TVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIK AGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARD VTVTLPDYPGSVPIPLTVYCAKSQNLGYYLSGTTADAGNSIFTNTASFS PAQGVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTANYARTGGQ VTAGNVQSIIGVTFVYQPGDGNADVTITVNGKVVAKGGGGSLVPRG SGGGGSDIIKLLNEQVNKEMNSSNLYMSMSSWCYTHSLDGAGLFLF DHAAEEYEHAKKLIIFLNENNVPVQLTSISAPEHKFEGLTQIFQKAYEH EQHISESINNIVDHAIKSKDHATFNFLQWYVAEQHEEEVLFKDILDKIE LIGNENHGLYLADQYVKGIAKSRKS-- **[SEQ ID NO: 139]**		
	FIMH_NOD_S_HIS_FERRITN_J96	HIS linker-terminal SRKS	METDTLLLWVLLLWVPGSTGFACKTANGTAIPIGGGSANVYVNLAP VVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSG TVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIK AGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARD VTVTLPDYPGSVPIPLTVYCAKSQNLGYYLSGTTADAGNSIFTNTASFS PAQGVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTANYARTGGQ VTAGNVQSIIGVTFVYQPGDGNADVTITVNGKVVAKSGSHHHHHH HHGGSDIIKLLNEQVNKEMNSSNLYMSMSSWCYTHSLDGAGLFLFD HAAEEYEHAKKLIIFLNENNVPVQLTSISAPEHKFEGLTQIFQKAYEHE QHISESINNIVDHAIKSKDHATFNFLQWYVAEQHEEEVLFKDILDKIELI GNENHGLYLADQYVKGIAKSRKS--**[SEQ ID NO: 140]**		
	FIMH_DG_PGDGN-PADRE-Ferritin536		METDTLLLWVLLLWVPGSTGFACKTANGTAIPIGGGSANVYVNLAP AVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSSFSG TVKYNGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIK AGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARD VTVTLPDYPGSVPIPLTVYCAKSQNLGYYLSGTTADAGNSIFTNTASFS PAQGVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTANYARTGGQ VTAGNVQSIIGVTFVYQPGDGNADVTITVNGKVVAKSGSFVAAWTL KAAAGGSDIIKLLNEQVNKEMNSSNLYMSMSSWCYTHSLDGAGLFL FDHAAEEYEHAKKLIIFLNENNVPVQLTSISAPEHKFEGLTQIFQKAYE HEQHISESINNIVDHAIKSKDHATFNFLQWYVAEQHEEEVLFKDILDKI ELIGNENHGLYLADQYVKGIAKSRKS- **[SEQ ID NO: 141]**		

2021392894 29 Aug 2024

	fimh-DG_ferritina-linkerAlpha	Rigid linker	METDTLLLWVLLLWVPGSTGFACKTANGTAIPIGGGSANVYVNLAP VVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSG TVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIK AGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARD VTVTLPDYPGSVPIPLTVYCAKSQNLGYYLSGTTADAGNSIFTNTASFS PAQGVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTANYARTGGQ VTAGNVQSIIGVTFVYQPGDGNADVTITVNGKVVAKAEAAAKEAAA KEAAAKADIIKLLNEQVNKEMNSSNLYMSMSSWCYTHSLDGAGLFL FDHAAEEYEHAKKLIIFLNENNVPVQLTSISAPEHKFEGLTQIFQKAYE HEQHISESINNIVDHAIKSKDHATFNFLQWYVAEQHEEEVLFKDILDKI ELIGNENHGLYLADQYVKGIAKSRKS-- **[SEQ ID NO: 142]**		
	FIMH-PADRE-Ferritin536_noSpaces	Linker PADRE	METDTLLLWVLLLWVPGSTGFACKTANGTAIPIGGGSANVYVNLAP AVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSSFSG TVKYNGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIK AGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARD VTVTLPDYPGSVPIPLTVYCAKSQNLGYYLSGTTADAGNSIFTNTASFS PAQGVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTANYARTGGQ VTAGNVQSIIGVTFVYQPGDGNADVTITVNGKVVAKFVAAWTLKAA ADIIKLLNEQVNKEMNSSNLYMSMSSWCYTHSLDGAGLFLFDHAAE EYEHAKKLIIFLNENNVPVQLTSISAPEHKFEGLTQIFQKAYEHEQHISE SINNIVDHAIKSKDHATFNFLQWYVAEQHEEEVLFKDILDKIELIGNEN HGLYLADQYVKGIAKSRKS--**[SEQ ID NO: 143]**		
	FIMH_DG_GSG4-ferritin_J96		METDTLLLWVLLLWVPGSTGDFACKTANGTAIPIGGGSANVYVNLA PVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFS GTVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAI KAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSAR DVTVTLPDYPGSVPIPLTVYCAKSQNLGYYLSGTTADAGNSIFTNTASF SPAQGVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTANYARTGG QVTAGNVQSIIGVTFVYQPGDGNADVTITVNGKVVAKSGSGSGGGG GGSDIIKLLNEQVNKEMNSSNLYMSMSSWCYTHSLDGAGLFLFDHA AEEYEHAKKLIIFLNENNVPVQLTSISAPEHKFEGLTQIFQKAYEHEQHI SESINNIVDHAIKSKDHATFNFLQWYVAEQHEEEVLFKDILDKIELIGN ENHGLYLADQYVKGIAKSRK-- **[SEQ ID NO: 144]**		
	FIMH_DG_PGDGN-SGS_PADRE-Ferritinj96		METDTLLLWVLLLWVPGSTGFACKTANGTAIPIGGGSANVYVNLAP VVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSG TVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIK AGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARD VTVTLPDYPGSVPIPLTVYCAKSQNLGYYLSGTTADAGNSIFTNTASFS PAQGVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTANYARTGGQ VTAGNVQSIIGVTFVYQPGDGNADVTITVNGKVVAKSGSFVAAWTL KAAAGGSDIIKLLNEQVNKEMNSSNLYMSMSSWCYTHSLDGAGLFL FDHAAEEYEHAKKLIIFLNENNVPVQLTSISAPEHKFEGLTQIFQKAYE HEQHISESINNIVDHAIKSKDHATFNFLQWYVAEQHEEEVLFKDILDKI ELIGNENHGLYLADQYVKGIAKSRKS- **[SEQ ID NO: 145]**		

2021392894 29 Aug 2024

	fimh-ferritina-linkerNONa	Rigid linker	METDTLLLWVLLLWVPGSTGFACKTANGTAIPIGGGSANVYVNLAP VVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSG TVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIK AGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARD VTVTLPDYPGSVPIPLTVYCAKSQNLGYYLSGTTADAGNSIFTNTASFS PAQGVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTANYARTGGQ VTAGNVQSIIGVTFVYQPGDGNADVTITVNGKVVAKGGGGSLVPRG SGGGGSDIIKLLNEQVNKEMNSSNLYMSMSSWCYTHSLDGAGLFLF DHAAEEYEHAKKLIIFLNENNVPVQLTSISAPEHKFEGLTQIFQKAYEH EQHISESINNIVDHAIKSKDHATFNFLQWYVAEQHEEEVLFKDILDKIE LIGNENHGLYLADQYVKGIAKSRKS--**[SEQ ID NO: 146]**		
	FIMH-Ferritinj96_noSpaces	Linker PADRE	METDTLLLWVLLLWVPGSTGFACKTANGTAIPIGGGSANVYVNLAP VVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSNFSG TVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIK AGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSARD VTVTLPDYPGSVPIPLTVYCAKSQNLGYYLSGTTADAGNSIFTNTASFS PAQGVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTANYARTGGQ VTAGNVQSIIGVTFVYQPGDGNADVTITVNGKVVAKFVAAWTLKAA ADIIKLLNEQVNKEMNSSNLYMSMSSWCYTHSLDGAGLFLFDHAAE EYEHAKKLIIFLNENNVPVQLTSISAPEHKFEGLTQIFQKAYEHEQHISE SINNIVDHAIKSKDHATFNFLQWYVAEQHEEEVLFKDILDKIELIGNEN HGLYLADQYVKGIAKSRKS--**[SEQ ID NO: 147]**		
	FIMH_DG_PGDGN-GGS4-Ferritin536	With N-terminal D	METDTLLLWVLLLWVPGSTGDFACKTANGTAIPIGGGSANVYVNLA PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSSFS GTVKYNGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAI KAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSAR DVTVTLPDYPGSVPIPLTVYCAKSQNLGYYLSGTTADAGNSIFTNTASF SPAQGVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTANYARTGG QVTAGNVQSIIGVTFVYQPGDGNADVTITVNGKVVAKGGSGGSGGS GGSDIIKLLNEQVNKEMNSSNLYMSMSSWCYTHSLDGAGLFLFDHA AEEYEHAKKLIIFLNENNVPVQLTSISAPEHKFEGLTQIFQKAYEHEQHI SESINNIVDHAIKSKDHATFNFLQWYVAEQHEEEVLFKDILDKIELIGN ENHGLYLADQYVKGIAKSRK-**[SEQ ID NO: 148]**		
	Fusion of FimH_DG to E.Coli Ferritin stabilized 0.5		METDTLLLWVLLLWVPGSTGDFACKTANGTAIPIGGGSANVYVNLA PAVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQRGSAYGGVLSSFS GTVKYNGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAI KAGSLIAVLILRQTNNYNSDDFQFVWNIYANNDVVVPTGGCDVSAR DVTVTLPDYPGSVPIPLTVYCAKSQNLGYYLSGTTADAGNSIFTNTASF SPAQGVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTANYARTGG QVTAGNVQSIIGVTFVYQPGDGNADVTITVNGKVVAKSGSHHHHH HHHGGSMLKPEMIEKLNEQMNLELYSSLLYQQMSAWCSYHGFEGA AAFLRRHAQEEMTHMQRLFDYLTDTGNLPRIDTIPSPFAEYSSLDELF QETYKHEQLITQKINELAHAAMTNQDYPTFNFLQWYVAEQHEEEKL FKSIIDKLSLAGKSGEGLYFIDKELSTLDTQN -**[SEQ ID NO: 153]**		

2021392894 29 Aug 2024

Table 3 – construct nucleic acid sequences

FIMH_DG_536-1EUM_0_5	ATGGAAACCGATACACTGCTGCTGTGGGTGCTGTTGCTCTGGGTTCCAGGCTCTACAGGCGATTTTGC CTGCAAGACCGCCAACGGCACAGCCATTCCTATTGGCGGAGGCAGCGCCAATGTGTACGTTAACCTGG CTCCTGCCGTGAACGTGGGCCAGAATCTGGTGGTGGATCTGAGCACCCAGATCTTTTGCCACAACGAC TACCCCGAGACAATCACCGACTACGTGACACTGCAGAGAGGCTCTGCTTACGGCGGCGTGCTGTCTAG CTTTAGCGGCACCGTGAAGTACAACGGCAGCAGCTACCCATTTCCTACCACCAGCGAGACACCCAGAG TGGTGTACAACAGCAGAACCGACAAGCCCTGGCCTGTGGCTCTGTACCTGACACCTGTTAGTTCTGCC GGCGGAGTGGCCATTAAGGCCGGATCTCTGATTGCCGTGCTGATCCTGCGGCAGACCAACAACTACA ACAGCGACGACTTCCAGTTCGTGTGGAACATCTACGCCAACAACGACGTGGTGGTGCCTACAGGCGG ATGTGATGTGTCCGCCAGAGATGTGACAGTGACCCTGCCTGATTACCCCGGCTCTGTGCCTATTCCTCT GACCGTGTACTGCGCCAAGAGCCAGAACCTGGGCTACTACCTGTCTGGCACAACAGCCGATGCCGGC AACAGCATCTTTACCAACACCGCCAGCTTCAGCCCTGCTCAAGGTGTTGGAGTGCAGCTGACCCGGAA CGGAACAATCATCCCCGCCAACAATACCGTGTCTCTGGGAGCTGTGGGCACCTCTGCTGTGTCTCTTGG CCTGACAGCCAACTATGCCAGAACAGGCGGACAAGTGACAGCCGGCAATGTGCAGTCTATCATCGGC GTGACCTTCGTGTATCAGCCTGGCGACGGAAACGCCGATGTGACCATCACAGTGAATGGCAAGGTGG TGGCCAAGAGCGGAAGCCACCACCATCATCACCATCACCACGGCGGCTCTATGCTGAAGCCCGAGATG ATCGAGAAGCTGAACGAGCAGATGAACCTGGAACTGTACAGCTCCCTGCTGTACCAGCAGATGAGCG CCTGGTGTAGCTATCACGGATTTGAGGGCGCTGCCGCCTTTCTGAGAAGGCACGCCCAAGAGGAAAT GACCCACATGCAGCGGCTGTTCGACTACCTGACCGATACCGGCAATCTGCCCAGAATCGACACAATCC CATCTCCATTCGCCGAGTACAGCAGCCTGGACGAGCTGTTCCAAGAAACCTACAAGCACGAGCAGCTG ATCACCCAGAAGATCAACGAACTGGCCCATGCCGCCATGACCAACCAGGACTACCCTACCTTCAACTTC CTGCAGTGGTACGTGGCCGAGCAGCACGAGGAAGAGAAGCTGTTCAAGAGCATCATCGACAAGCTGA GCCTGGCCGGAAAGTCTGGCGAGGGCCTGTACTTTATCGACAAAGAGCTGAGCACACTGGATACCCA GAACTGA **[SEQ ID NO: 45]**
FIMH_DG_PGDGN_536-MI3	ATGGAAACCGATACACTGCTGCTGTGGGTGCTGTTGCTCTGGGTTCCAGGCTCTACAGGCGATTTTGC CTGCAAGACCGCCAACGGCACAGCCATTCCTATTGGCGGAGGCAGCGCCAATGTGTACGTTAACCTGG CTCCTGCCGTGAACGTGGGCCAGAATCTGGTGGTGGATCTGAGCACCCAGATCTTTTGCCACAACGAC TACCCCGAGACAATCACCGACTACGTGACACTGCAGAGAGGCTCTGCTTACGGCGGCGTGCTGTCTAG CTTTAGCGGCACCGTGAAGTACAACGGCAGCAGCTACCCATTTCCTACCACCAGCGAGACACCCAGAG TGGTGTACAACAGCAGAACCGACAAGCCCTGGCCTGTGGCTCTGTACCTGACACCTGTTAGTTCTGCC GGCGGAGTGGCCATTAAGGCCGGATCTCTGATTGCCGTGCTGATCCTGCGGCAGACCAACAACTACA ACAGCGACGACTTCCAGTTCGTGTGGAACATCTACGCCAACAACGACGTGGTGGTGCCTACAGGCGG ATGTGATGTGTCCGCCAGAGATGTGACAGTGACCCTGCCTGATTACCCCGGCTCTGTGCCTATTCCTCT GACCGTGTACTGCGCCAAGAGCCAGAACCTGGGCTACTACCTGTCTGGCACAACAGCCGATGCCGGC AACAGCATCTTTACCAACACCGCCAGCTTCAGCCCTGCTCAAGGTGTTGGAGTGCAGCTGACCCGGAA CGGAACAATCATCCCCGCCAACAATACCGTGTCTCTGGGAGCTGTGGGCACCTCTGCTGTGTCTCTTGG CCTGACAGCCAACTATGCCAGAACAGGCGGACAAGTGACAGCCGGCAATGTGCAGTCTATCATCGGC GTGACCTTCGTGTATCAGCCTGGCGACGGAAACGCCGATGTGACCATCACAGTGAATGGCAAGGTGG TGGCCAAGAGCGGAAGCCACCACCATCATCACCATCACCACGGCGGCAGCATGAAGATGGAAGAACT GTTCAAGAAGCACAAGATCGTCGCCGTGCTGCGGGCCAATTCTGTGGAAGAGGCCAAAAAAAAGGCC CTGGCCGTGTTTCTTGGCGGAGTGCACCTGATCGAGATCACCTTTACCGTGCCTGACGCCGACACCGT GATCAAAGAGCTGAGCTTCCTGAAAGAGATGGGCGCCATCATCGGAGCCGGCACAGTGACATCTGTT GAGCAGGCCAGAAAGGCCGTGGAATCTGGCGCCGAGTTTATCGTGTCCCCTCACCTGGATGAGGAAA TCAGCCAGTTCGCCAAAGAAAAGGGCGTGTTCTACATGCCCGGCGTGATGACACCTACAGAGCTGGTC

2021392894 29 Aug 2024

	AAAGCCATGAAGCTGGGCCACACCATCCTGAAGCTGTTTCCAGGCGAAGTCGTGGGCCCTCAGTTCGT GAAAGCTATGAAGGGCCCATTTCCAAACGTGAAGTTCGTGCCCACTGGCGGCGTGAACCTGGATAAT GTGTGCGAGTGGTTCAAGGCTGGCGTGCTGGCTGTTGGAGTTGGCTCTGCTCTGGTCAAGGGCACAC CTGTGGAAGTGGCTGAGAAGGCCAAGGCCTTCGTGGAAAAGATCAGAGGCTGCACCGAGTGA **[SEQ ID NO: 46]**
FIMH_DG_PGDGN_536-encapsuline	ATGGAAACCGATACACTGCTGCTGTGGGTGCTGTTGCTCTGGGTTCCAGGCTCTACAGGCGATTTTGC CTGCAAGACCGCCAACGGCACAGCCATTCCTATTGGCGGAGGCAGCGCCAATGTGTACGTTAACCTGG CTCCTGCCGTGAACGTGGGCCAGAATCTGGTGGTGGATCTGAGCACCCAGATCTTTTGCCACAACGAC TACCCCGAGACAATCACCGACTACGTGACACTGCAGAGAGGCTCTGCTTACGGCGGCGTGCTGTCTAG CTTTAGCGGCACCGTGAAGTACAACGGCAGCAGCTACCCATTTCCTACCACCAGCGAGACACCCAGAG TGGTGTACAACAGCAGAACCGACAAGCCCTGGCCTGTGGCTCTGTACCTGACACCTGTTAGTTCTGCC GGCGGAGTGGCCATTAAGGCCGGATCTCTGATTGCCGTGCTGATCCTGCGGCAGACCAACAACTACA ACAGCGACGACTTCCAGTTCGTGTGGAACATCTACGCCAACAACGACGTGGTGGTGCCTACAGGCGG ATGTGATGTGTCCGCCAGAGATGTGACAGTGACCCTGCCTGATTACCCCGGCTCTGTGCCTATTCCTCT GACCGTGTACTGCGCCAAGAGCCAGAACCTGGGCTACTACCTGTCTGGCACAACAGCCGATGCCGGC AACAGCATCTTTACCAACACCGCCAGCTTCAGCCCTGCTCAAGGTGTTGGAGTGCAGCTGACCCGGAA CGGAACAATCATCCCCGCCAACAATACCGTGTCTCTGGGAGCTGTGGGCACCTCTGCTGTGTCTCTTGG CCTGACAGCCAACTATGCCAGAACAGGCGGACAAGTGACAGCCGGCAATGTGCAGTCTATCATCGGC GTGACCTTCGTGTATCAGCCTGGCGACGGAAACGCCGATGTGACCATCACAGTGAATGGCAAGGTGG TGGCCAAGAGCGGAAGCCACCACCATCATCACCATCACCACGGCGGCAGCATGGAATTTCTGAAGAG AAGCTTCGCCCCACTGACCGAGAAGCAGTGGCAAGAGATCGACAACCGGGCCAGAGAGATCTTCAAG ACCCAGCTGTACGGCCGGAAGTTCGTGGATGTGGAAGGCCCTTATGGCTGGGAGTATGCCGCTCATC CTCTGGGCGAAGTGGAAGTGCTGAGCGACGAGAATGAGGTCGTGAAGTGGGGCCTGAGAAAGAGCC TGCCTCTGATCGAGCTGAGAGCCACCTTCACACTGGACCTGTGGGAACTCGACAACCTGGAAAGGGG CAAGCCCAATGTGGACCTGAGCAGCCTGGAAGAGACAGTGCGGAAGGTGGCCGAGTTCGAGGACGA AGTGATCTTCAGAGGCTGCGAGAAGTCTGGCGTGAAGGGCCTGCTGAGCTTCGAGGAACGGAAGATC GAGTGTGGCAGCACCCCTAAGGATCTGCTGGAAGCCATCGTGCGGGCCCTGAGCATCTTCTCTAAGGA TGGCATCGAGGGCCCCTACACACTGGTCATCAACACCGACCGGTGGATCAACTTCCTGAAAGAGGAA GCCGGCCACTATCCTCTGGAAAAGCGCGTGGAAGAGTGCCTGAGAGGCGGCAAGATCATCACAACCC CTAGAATCGAGGACGCCCTGGTGGTTTCTGAGAGAGGCGGAGACTTCAAGCTGATCCTTGGCCAGGA CCTGTCCATCGGCTACGAGGACAGAGAAAAAGACGCCGTGCGGCTGTTCATCACCGAAACCTTCACCT TCCAAGTGGTCAACCCCGAGGCTCTGATTCTGCTGAAGTTCTGA **[SEQ ID NO: 47]**
HBcFIMHLJ96	ATGGAAACCGATACACTGCTGCTGTGGGTGCTGTTGCTCTGGGTTCCAGGATCTACCGGCGACGACAT CGACCCCTACAAAGAGTTTGGCGCCAGCGTCGAGCTGCTGAGCTTCCTGCCTAGCGACTTCTTCCCTTC CATCCGGGATCTGCTGGATACCGCTAGCGCCCTGTATAGAGAGGCCCTGGAAAGCCCTGAGCACTGCT CTCCACATCACACAGCCCTGAGACAGGCCATCCTGTGTTGGGGCGAACTGATGAATCTGGCCACCTGG GTCGGAAGCAACCTGGAAGATCCTGGTTCTGGCGGCGGAGGCTTTGCCTGTAAAACAGCCAATGGCA CCGCCATTCCTATCGGAGGCGGCAGCGCCAATGTGTACGTTAACCTGGCTCCTGTGGTCAACGTGGGC CAGAATCTGGTGGTGGACCTGAGCACCCAGATCTTTTGCCACAACGACTACCCCGAGACAATCACCGA CTACGTGACACTGCAGAGAGGCTCTGCTTACGGCGGCGTGCTGAGCAATTTTTCCGGCACAGTGAAGT ACAGCGGCAGCAGCTACCCATTTCCTACCACCAGCGAGACACCCAGAGTGGTGTACAACAGCAGAACC GACAAGCCCTGGCCTGTGGCTCTGTACCTGACACCTGTTAGTTCTGCTGGCGGAGTGGCCATCAAGGC CGGATCTCTGATTGCCGTGCTGATCCTGCGGCAGACCAACAACTACAACAGCGACGACTTCCAGTTCG TGTGGAACATCTACGCCAACAACGACGTGGTGGTGCCTACAGGCGGAGGATCTGGCGGAGCTTCTAG

2021392894 29 Aug 2024

	AGAACTGGTCGTGTCCTACGTGAACGTGAACATGGGCCTGAAGATCCGGCAGCTGCTCTGGTTTCACA TCAGCTGTCTGACCTTCGGCCGGGAAACCGTGCTGGAATACCTGGTGTCCTTCGGCGTGTGGATCAGA ACCCCTCCTGCCTATAGACCTCCTAACGCTCCCATCCTGAGCACACTGCCTGAGACAACAGTTGTTGGA AGCGGAGGCGGAGGCCACCACCATCACCATCAT **[SEQ ID NO: 48]**
HBcFIM HDGJ96	ATGGAGACCGACACCCTGCTGCTGTGGGTGCTGCTGCTGTGGGTGCCCGGCAGCACCGGCGACGACA TCGACCCCTACAAGGAGTTCGGCGCCAGCGTGGAGCTGCTGAGCTTCCTGCCCAGCGACTTCTTCCCC AGCATCCGGGACCTGCTGGACACCGCCAGCGCCCTGTACCGGGAGGCCCTGGAGAGCCCCGAGCACT GCAGCCCCCACCACACCGCCCTGCGGCAGGCCATCCTGTGCTGGGGCGAGCTGATGAACCTGGCCAC CTGGGTGGGCAGCAACCTGGAGGACCCCGGCAGCGGCGGCGGCGGCTTCGCCTGCAAGACCGCCAA CGGCACCGCCATCCCCATCGGCGGCGGCAGCGCCAACGTGTACGTGAACCTGGCCCCCGTGGTGAAC GTGGGCCAGAACCTGGTGGTGGACCTGAGCACCCAGATCTTCTGCCACAACGACTACCCCGAGACCAT CACCGACTACGTGACCCTGCAGCGGGGCAGCGCCTACGGCGGCGTGCTGAGCAACTTCAGCGGCACC GTGAAGTACAGCGGCAGCAGCTACCCCTTCCCCACCACCAGCGAGACCCCCCGGGTGGTGTACAACA GCCGGACCGACAAGCCCTGGCCCGTGGCCCTGTACCTGACCCCCGTGAGCAGCGCCGGCGGCGTGGC CATCAAGGCCGGCAGCCTGATCGCCGTGCTGATCCTGCGGCAGACCAACAACTACAACAGCGACGACT TCCAGTTCGTGTGGAACATCTACGCCAACAACGACGTGGTGGTGCCCACCGGCGGCTGCGACGTGAG CGCCCGGGACGTGACCGTGACCCTGCCCGACTACCCCGGCAGCGTGCCCATCCCCCTGACCGTGTACT GCGCCAAGAGCCAGAACCTGGGCTACTACCTGAGCGGCACCACCGCCGACGCCGGCAACAGCATCTT CACCAACACCGCCAGCTTCAGCCCCGCCCAGGGCGTGGGCGTGCAGCTGACCCGGAACGGCACCATC ATCCCCGCCAACAACACCGTGAGCCTGGGCGCCGTGGGCACCAGCGCCGTGAGCCTGGGCCTGACCG CCAACTACGCCCGGACCGGCGGCCAGGTGACCGCCGGCAACGTGCAGAGCATCATCGGCGTGACCTT CGTGTACCAGCCCGGCGACGGCAACGCCGACGTGACCATCACCGTGAACGGCAAGGTGGTGGCCAA GGGCAGCGGCGGCGGCGGCGCCAGCCGGGAGCTGGTGGTGAGCTACGTGAACGTGAACATGGGCCT GAAGATCCGGCAGCTGCTGTGGTTCCACATCAGCTGCCTGACCTTCGGCCGGGAGACCGTGCTGGAG TACCTGGTGAGCTTCGGCGTGTGGATCCGGACCCCCCCCGCCTACCGGCCCCCCAACGCCCCCATCCTG AGCACCCTGCCCGAGACCACCGTGGTGGGCAGCGGCGGCGGCGGCCACCACCACCACCACCAC **[SEQ ID NO: 49]**
FIMHL-HIS-Mi3J96	ATGGAAACCGATACACTGCTGCTGTGGGTGCTGTTGCTCTGGGTTCCAGGCTCTACAGGCGATTTTGC CTGCAAGACCGCCAACGGCACAGCCATTCCTATTGGCGGAGGCAGCGCCAATGTGTACGTGAACCTG GCTCCTGTGGTCAACGTGGGCCAGAATCTGGTGGTGGACCTGAGCACCCAGATCTTTTGCCACAACGA CTACCCCGAGACAATCACCGACTACGTGACACTGCAGAGAGGCTCTGCTTACGGCGGCGTGCTGAGC AATTTTTCCGGCACAGTGAAGTACAGCGGCAGCAGCTACCCATTTCCTACCACCAGCGAGACACCCAG AGTGGTGTACAACAGCAGAACCGACAAGCCCTGGCCTGTGGCTCTGTACCTGACACCTGTTAGTTCTG CCGGCGGAGTGGCCATTAAGGCCGGATCTCTGATTGCCGTGCTGATCCTGCGGCAGACCAACAACTAC AACAGCGACGACTTCCAGTTCGTGTGGAACATCTACGCCAACAACGACGTGGTGGTGCCTACAAGCG GCAGCCACCACCATCATCACCATCACCACGGCGGCAGCATGAAGATGGAAGAACTGTTCAAGAAGCA CAAGATCGTCGCCGTGCTGCGGGCCAATTCTGTGGAAGAGGCCAAAAAAAAGGCCCTGGCCGTGTTT CTTGGCGGAGTGCACCTGATCGAGATCACCTTTACCGTGCCTGACGCCGACACCGTGATCAAAGAGCT GAGCTTCCTGAAAGAGATGGGCGCCATCATCGGAGCCGGAACCGTGACATCTGTTGAGCAGGCCAGA AAGGCCGTGGAATCTGGCGCCGAGTTTATCGTGTCCCCTCACCTGGATGAGGAAATCAGCCAGTTCGC CAAAGAAAAGGGCGTGTTCTACATGCCCGGCGTGATGACACCTACAGAGCTGGTCAAAGCCATGAAG CTGGGCCACACCATCCTGAAGCTGTTTCCAGGCGAAGTCGTGGGCCCTCAGTTCGTGAAAGCTATGAA GGGCCCATTTCCAAACGTGAAGTTCGTGCCCACTGGCGGCGTCAACCTGGATAATGTGTGCGAGTGGT

2021392894 29 Aug 2024

	TCAAGGCTGGCGTGCTGGCTGTTGGAGTGGGATCTGCTCTGGTCAAGGGCACACCTGTGGAAGTGGC TGAGAAGGCCAAGGCCTTCGTGGAAAAGATCAGAGGCTGCACCGAGTGA **[SEQ ID NO: 50]**
FIMHL-HIS-Fer536	ATGGAAACCGATACACTGCTGCTGTGGGTGCTGTTGCTCTGGGTTCCAGGCTCTACAGGCGATTTTGC CTGCAAGACCGCCAACGGCACAGCCATTCCTATTGGCGGAGGCAGCGCCAATGTGTACGTTAACCTGG CTCCTGCCGTGAACGTGGGCCAGAATCTGGTGGTGGATCTGAGCACCCAGATCTTTTGCCACAACGAC TACCCCGAGACAATCACCGACTACGTGACACTGCAGAGAGGCTCTGCTTACGGCGGCGTGCTGTCTAG CTTTAGCGGCACCGTGAAGTACAACGGCAGCAGCTACCCATTTCCTACCACCAGCGAGACACCCAGAG TGGTGTACAACAGCAGAACCGACAAGCCCTGGCCTGTGGCTCTGTACCTGACACCTGTTAGTTCTGCC GGCGGAGTGGCCATTAAGGCCGGATCTCTGATTGCCGTGCTGATCCTGCGGCAGACCAACAACTACA ACAGCGACGACTTCCAGTTCGTGTGGAACATCTACGCCAACAACGACGTGGTGGTGCCTACAAGCGG CAGCCACCACCATCATCACCATCACCACGGCGGCAGCGACATCATCAAGCTGCTGAACGAGCAAGTGA ACAAAGAGATGAACAGCAGCAACCTGTACATGAGCATGAGCAGCTGGTGCTACACACACAGCCTGGA TGGCGCCGGACTGTTCCTGTTTGATCATGCCGCCGAGGAATACGAGCACGCCAAGAAGCTGATCATCT TCCTGAACGAGAACAACGTGCCCGTGCAGCTGACATCTATCAGCGCCCCTGAGCACAAGTTCGAGGGC CTGACACAGATCTTCCAGAAGGCCTACGAACACGAGCAGCACATCAGCGAGAGCATCAACAACATCGT GGACCACGCCATCAAGAGCAAGGATCACGCCACCTTCAACTTTCTGCAGTGGTACGTGGCCGAACAGC ACGAGGAAGAGGTGCTGTTCAAGGACATCCTGGACAAGATCGAGCTGATCGGCAACGAGAACCACG GCCTGTACCTGGCCGATCAGTATGTGAAGGGAATCGCCAAGAGCCGGAAGTGA **[SEQ ID NO: 51]**
FIMH_DG_PGDGN_IMX313_HISJ96	ATGGAAACCGATACACTGCTGCTGTGGGTGCTGTTGCTCTGGGTTCCAGGATCTACAGGGGATTTTGC CTGCAAGACCGCCAATGGCACAGCCATTCCTATTGGCGGCGGAAGCGCCAATGTGTACGTGAACCTG GCTCCTGTGGTCAACGTGGGCCAGAATCTGGTGGTGGACCTGAGCACCCAGATCTTTTGCCACAACGA CTACCCCGAGACAATCACCGACTACGTGACACTGCAGAGAGGCTCTGCTTACGGCGGCGTGCTGAGC AATTTTTCCGGCACAGTGAAGTACAGCGGCAGCAGCTACCCATTTCCTACCACCAGCGAGACACCCAG AGTGGTGTACAACAGCAGAACCGACAAGCCCTGGCCTGTGGCTCTGTACCTGACACCTGTTAGTTCTG CCGGCGGAGTGGCCATTAAGGCCGGATCTCTGATTGCCGTGCTGATCCTGCGGCAGACCAACAACTAC AACAGCGACGACTTCCAGTTCGTGTGGAACATCTACGCCAACAACGACGTGGTGGTGCCTACAGGCG GATGTGATGTGTCCGCCAGAGATGTGACAGTGACCCTGCCTGATTACCCCGGCTCTGTGCCTATTCCTC TGACCGTGTACTGCGCCAAGTCTCAGAACCTGGGCTACTACCTGAGCGGCACAACAGCCGATGCCGGC AACAGCATCTTTACCAACACCGCCAGCTTCAGCCCTGCTCAAGGTGTTGGAGTGCAGCTGACCCGGAA CGGAACAATCATCCCCGCCAACAATACCGTGTCTCTGGGAGCTGTGGGCACCTCTGCTGTGTCTCTTGG CCTGACAGCCAACTATGCCAGAACAGGCGGACAAGTGACAGCCGGCAATGTGCAGTCTATCATCGGC GTGACCTTCGTGTATCAGCCTGGCGACGGAAATGCCGACGTGACCATCACAGTGAATGGCAAGGTGG TGGCCAAGGGCAGCTCAGGCTCTGGCTCTGGATCTAAAAAACAGGGCGACGCCGATGTGTGTGGCGA GGTGGCATATATCCAGAGCGTGGTGTCCGATTGTCACGTGCCAACCGCCGAGCTGAGAACCCTGCTG GAAATCCGGAAGCTGTTCCTCGAAATTCAGAAGCTGAAGGTCGAGCTGCAGGGCCTGTCTAAAGAAG GCGGAGGAAGCGGATCTCACCACCACCATCACCACTGATGA **[SEQ ID NO: 52]**
FimH_PGDGN_DG_HIS-Ferritnj96	ATGGAAACCGATACACTGCTGCTGTGGGTGCTGTTGCTCTGGGTTCCAGGATCTACAGGGGATTTTGC CTGCAAGACCGCCAATGGCACAGCCATTCCTATTGGCGGCGGAAGCGCCAATGTGTACGTGAACCTG GCTCCTGTGGTCAACGTGGGCCAGAATCTGGTGGTGGACCTGAGCACCCAGATCTTTTGCCACAACGA CTACCCCGAGACAATCACCGACTACGTGACACTGCAGAGAGGCTCTGCTTACGGCGGCGTGCTGAGC AATTTTTCCGGCACAGTGAAGTACAGCGGCAGCAGCTACCCATTTCCTACCACCAGCGAGACACCCAG AGTGGTGTACAACAGCAGAACCGACAAGCCCTGGCCTGTGGCTCTGTACCTGACACCTGTTAGTTCTG CCGGCGGAGTGGCCATTAAGGCCGGATCTCTGATTGCCGTGCTGATCCTGCGGCAGACCAACAACTAC AACAGCGACGACTTCCAGTTCGTGTGGAACATCTACGCCAACAACGACGTGGTGGTGCCTACAGGCG

2021392894 29 Aug 2024

	GATGTGATGTGTCCGCCAGAGATGTGACAGTGACCCTGCCTGATTACCCCGGCTCTGTGCCTATTCCTC TGACCGTGTACTGCGCCAAGTCTCAGAACCTGGGCTACTACCTGAGCGGCACAACAGCCGATGCCGGC AACAGCATCTTTACCAACACCGCCAGCTTCAGCCCTGCTCAAGGTGTTGGAGTGCAGCTGACCCGGAA CGGAACAATCATCCCCGCCAACAATACCGTGTCTCTGGGAGCTGTGGGCACCTCTGCTGTGTCTCTTGG CCTGACAGCCAACTATGCCAGAACAGGCGGACAAGTGACAGCCGGCAATGTGCAGTCTATCATCGGC GTGACCTTCGTGTATCAGCCTGGCGACGGAAATGCCGACGTGACCATCACAGTGAATGGCAAGGTGG TGGCCAAGAGCGGAAGCCACCACCATCATCACCATCACCACGGCGGCAGCATGAAGATGGAAGAACT GTTCAAGAAGCACAAGATCGTCGCCGTGCTGCGGGCCAATTCTGTGGAAGAGGCCAAAAAAAAGGCC CTGGCCGTGTTTCTTGGCGGAGTGCACCTGATCGAGATCACCTTTACCGTGCCTGACGCCGACACCGT GATCAAAGAGCTGAGCTTCCTGAAAGAGATGGGCGCCATCATCGGAGCCGGAACCGTGACATCTGTT GAGCAGGCCAGAAAGGCCGTGGAATCTGGCGCCGAGTTTATCGTGTCCCCTCACCTGGATGAGGAAA TCAGCCAGTTCGCCAAAGAAAAGGGCGTGTTCTACATGCCCGGCGTGATGACACCTACAGAGCTGGTC AAAGCCATGAAGCTGGGCCACACCATCCTGAAGCTGTTTCCAGGCGAAGTCGTGGGCCCTCAGTTCGT GAAAGCTATGAAGGGCCCATTTCCAAACGTGAAGTTCGTGCCCACTGGCGGCGTCAACCTGGATAATG TGTGCGAGTGGTTCAAGGCTGGCGTGCTGGCTGTTGGAGTTGGCTCTGCTCTGGTCAAGGGCACACCT GTGGAAGTGGCTGAGAAGGCCAAGGCCTTCGTGGAAAAGATCAGAGGCTGCACCGAGTGATGA **[SEQ ID NO: 53]**
FIMH_DG_PGDGN-HIS-Ferritin536	ATGGAAACCGATACACTGCTGCTGTGGGTGCTGTTGCTCTGGGTTCCAGGATCTACAGGGGATGTTTG CCTGCAAGACCGCCAATGGCACAGCCATTCCTATTGGCGGCGGAAGCGCCAATGTGTACGTTAACCTG GCTCCTGCCGTGAACGTGGGCCAGAATCTGGTGGTGGATCTGAGCACCCAGATCTTTTGCCACAACGA CTACCCCGAGACAATCACCGACTACGTGACACTGCAGAGAGGCTCTGCTTACGGCGGCGTGCTGTCTA GCTTTAGCGGCACCGTGAAGTACAACGGCAGCAGCTACCCATTTCCTACCACCAGCGAGACACCCAGA GTGGTGTACAACAGCAGAACCGACAAGCCCTGGCCTGTGGCTCTGTACCTGACACCTGTTAGTTCTGC CGGCGGAGTGGCCATTAAGGCCGGATCTCTGATTGCCGTGCTGATCCTGCGGCAGACCAACAACTACA ACAGCGACGACTTCCAGTTCGTGTGGAACATCTACGCCAACAACGACGTGGTGGTGCCTACAGGCGG ATGTGATGTGTCCGCCAGAGATGTGACAGTGACCCTGCCTGATTACCCCGGCTCTGTGCCTATTCCTCT GACCGTGTACTGCGCCAAGTCTCAGAACCTGGGCTACTACCTGAGCGGCACAACAGCCGATGCCGGC AACAGCATCTTTACCAACACCGCCAGCTTCAGCCCTGCTCAAGGTGTTGGAGTGCAGCTGACCCGGAA CGGAACAATCATCCCCGCCAACAATACCGTGTCTCTGGGAGCTGTGGGCACCTCTGCTGTGTCTCTTGG CCTGACAGCCAACTATGCCAGAACAGGCGGACAAGTGACAGCCGGCAATGTGCAGTCTATCATCGGC GTGACCTTCGTGTATCAGCCTGGCGACGGAAATGCCGACGTGACCATCACAGTGAATGGCAAGGTGG TGGCCAAGAGCGGAAGCCACCACCATCATCACCATCACCACGGCGGCAGCGACATCATCAAGCTGCTG AACGAGCAAGTGAACAAAGAGATGAACAGCAGCAACCTGTACATGAGCATGAGCAGCTGGTGCTACA CACACAGCCTGGATGGCGCCGGACTGTTCCTGTTTGATCATGCCGCCGAGGAATACGAGCACGCCAA GAAGCTGATCATCTTCCTGAACGAGAACAACGTGCCCGTCCAGCTGACATCTATCAGCGCCCCTGAGC ACAAGTTCGAGGGCCTGACACAGATCTTCCAGAAGGCCTACGAACACGAGCAGCACATCAGCGAGAG CATCAACAACATCGTGGACCACGCCATCAAGAGCAAGGATCACGCCACCTTCAACTTTCTGCAGTGGT ACGTGGCCGAACAGCACGAGGAAGAGGTGCTGTTCAAGGACATCCTGGACAAGATCGAGCTGATCG GCAACGAGAACCACGGCCTGTACCTGGCCGATCAGTATGTGAAGGGAATCGCCAAGAGCCGCAAGTG A **[SEQ ID NO: 54]**
FimH_PGDGN_DG-HIS-	ATGGAAACCGATACACTGCTGCTGTGGGTGCTGTTGCTCTGGGTTCCAGGATCTACAGGGGATGCCGC TCAACCTGCTCGGAGAGCCAGAAGAACAAAGCTGGCCCTGTTTGCCTGCAAGACCGCCAATGGCACA GCCATTCCTATTGGCGGCGGAAGCGCCAATGTGTACGTGAACCTGGCTCCTGTGGTCAACGTGGGCCA GAATCTGGTGGTGGACCTGAGCACCCAGATCTTTTGCCACAACGACTACCCCGAGACAATCACCGACT

2021392894 29 Aug 2024

Ferritin J96	ACGTGACACTGCAGAGAGGCTCTGCTTACGGCGGCGTGCTGAGCAATTTTTCCGGCACAGTGAAGTAC AGCGGCAGCAGCTACCCATTTCCTACCACCAGCGAGACACCCAGAGTGGTGTACAACAGCAGAACCG ACAAGCCCTGGCCTGTGGCTCTGTACCTGACACCTGTTAGTTCTGCCGGCGGAGTGGCCATTAAGGCC GGATCTCTGATTGCCGTGCTGATCCTGCGGCAGACCAACAACTACAACAGCGACGACTTCCAGTTCGT GTGGAACATCTACGCCAACAACGACGTGGTGGTGCCTACAGGCGGATGTGATGTGTCCGCCAGAGAT GTGACAGTGACCCTGCCTGATTACCCCGGCTCTGTGCCTATTCCTCTGACCGTGTACTGCGCCAAGTCT CAGAACCTGGGCTACTACCTGAGCGGCACAACAGCCGATGCCGGCAACAGCATCTTTACCAACACCGC CAGCTTCAGCCCTGCTCAAGGTGTTGGAGTGCAGCTGACCCGGAACGGAACAATCATCCCCGCCAACA ATACCGTGTCTCTGGGAGCTGTGGGCACCTCTGCTGTGTCTCTTGGCCTGACAGCCAACTATGCCAGA ACAGGCGGACAAGTGACAGCCGGCAATGTGCAGTCTATCATCGGCGTGACCTTCGTGTATCAGCCTG GCGACGGAAATGCCGACGTGACCATCACAGTGAATGGCAAGGTGGTGGCCAAGAGCGGAAGCCACC ACCATCATCACCATCACCACGGCGGCAGCGACATCATCAAGCTGCTGAACGAGCAAGTGAACAAAGA GATGAACAGCAGCAACCTGTACATGAGCATGAGCAGCTGGTGCTACACACACAGCCTGGATGGCGCC GGACTGTTCCTGTTTGATCATGCCGCCGAGGAATACGAGCACGCCAAGAAGCTGATCATCTTCCTGAA CGAGAACAACGTGCCCGTCCAGCTGACATCTATCAGCGCCCCTGAGCACAAGTTCGAGGGCCTGACAC AGATCTTCCAGAAGGCCTACGAACACGAGCAGCACATCAGCGAGAGCATCAACAACATCGTGGACCA CGCCATCAAGAGCAAGGATCACGCCACCTTCAACTTTCTGCAGTGGTACGTGGCCGAACAGCACGAG GAAGAGGTGCTGTTCAAGGACATCCTGGACAAGATCGAGCTGATCGGCAACGAGAACCACGGCCTGT ACCTGGCCGATCAGTATGTGAAGGGAATCGCCAAGAGCCGCAAGTGA **[SEQ ID NO: 55]**
FimH_PGDGN_DG-HIS-MI3 j96	ATGGAAACCGATACACTGCTGCTGTGGGTGCTGTTGCTCTGGGTTCCAGGATCTACAGGGGATGCCGC TCAACCTGCTCGGAGAGCCAGAAGAACAAAGCTGGCCCTGTTTGCCTGCAAGACCGCCAATGGCACA GCCATTCCTATTGGCGGCGGAAGCGCCAATGTGTACGTGAACCTGGCTCCTGTGGTCAACGTGGGCCA GAATCTGGTGGTGGACCTGAGCACCCAGATCTTTTGCCACAACGACTACCCCGAGACAATCACCGACT ACGTGACACTGCAGAGAGGCTCTGCTTACGGCGGCGTGCTGAGCAATTTTTCCGGCACAGTGAAGTAC AGCGGCAGCAGCTACCCATTTCCTACCACCAGCGAGACACCCAGAGTGGTGTACAACAGCAGAACCG ACAAGCCCTGGCCTGTGGCTCTGTACCTGACACCTGTTAGTTCTGCCGGCGGAGTGGCCATTAAGGCC GGATCTCTGATTGCCGTGCTGATCCTGCGGCAGACCAACAACTACAACAGCGACGACTTCCAGTTCGT GTGGAACATCTACGCCAACAACGACGTGGTGGTGCCTACAGGCGGATGTGATGTGTCCGCCAGAGAT GTGACAGTGACCCTGCCTGATTACCCCGGCTCTGTGCCTATTCCTCTGACCGTGTACTGCGCCAAGTCT CAGAACCTGGGCTACTACCTGAGCGGCACAACAGCCGATGCCGGCAACAGCATCTTTACCAACACCGC CAGCTTCAGCCCTGCTCAAGGTGTTGGAGTGCAGCTGACCCGGAACGGAACAATCATCCCCGCCAACA ATACCGTGTCTCTGGGAGCTGTGGGCACCTCTGCTGTGTCTCTTGGCCTGACAGCCAACTATGCCAGA ACAGGCGGACAAGTGACAGCCGGCAATGTGCAGTCTATCATCGGCGTGACCTTCGTGTATCAGCCTG GCGACGGAAATGCCGACGTGACCATCACAGTGAATGGCAAGGTGGTGGCCAAGAGCGGAAGCCACC ACCATCATCACCATCACCACGGCGGCAGCATGAAGATGGAAGAACTGTTCAAGAAGCACAAGATCGT CGCCGTGCTGCGGGCCAATTCTGTGGAAGAGGCCAAAAAAAAGGCCCTGGCCGTGTTTCTTGGCGGA GTGCACCTGATCGAGATCACCTTTACCGTGCCTGACGCCGACACCGTGATCAAAGAGCTGAGCTTCCT GAAAGAGATGGGCGCCATCATCGGAGCCGGAACCGTGACATCTGTTGAGCAGGCCAGAAAGGCCGT GGAATCTGGCGCCGAGTTTATCGTGTCCCCTCACCTGGATGAGGAAATCAGCCAGTTCGCCAAAGAAA AGGGCGTGTTCTACATGCCCGGCGTGATGACACCTACAGAGCTGGTCAAAGCCATGAAGCTGGGCCA CACCATCCTGAAGCTGTTTCCAGGCGAAGTCGTGGGCCCTCAGTTCGTGAAAGCTATGAAGGGCCCAT TTCCAAACGTGAAGTTCGTGCCCACTGGCGGCGTCAACCTGGATAATGTGTGCGAGTGGTTCAAGGCT GGCGTGCTGGCTGTTGGAGTTGGCTCTGCTCTGGTCAAGGGCACACCTGTGGAAGTGGCTGAGAAGG CCAAGGCCTTCGTGGAAAAGATCAGAGGCTGCACCGAGTGATGA **[SEQ ID NO: 56]**

2021392894 29 Aug 2024

FimH_PGDGN_DG-HIS-IMX313j96	ATGGAAACCGATACACTGCTGCTGTGGGTGCTGTTGCTCTGGGTTCCAGGATCTACAGGGGATGCCGC TCAACCTGCTCGGAGAGCCAGAAGAACAAAGCTGGCCCTGTTTGCCTGCAAGACCGCCAATGGCACA GCCATTCCTATTGGCGGCGGAAGCGCCAATGTGTACGTGAACCTGGCTCCTGTGGTCAACGTGGGCCA GAATCTGGTGGTGGACCTGAGCACCCAGATCTTTTGCCACAACGACTACCCCGAGACAATCACCGACT ACGTGACACTGCAGAGAGGCTCTGCTTACGGCGGCGTGCTGAGCAATTTTTCCGGCACAGTGAAGTAC AGCGGCAGCAGCTACCCATTTCCTACCACCAGCGAGACACCCAGAGTGGTGTACAACAGCAGAACCG ACAAGCCCTGGCCTGTGGCTCTGTACCTGACACCTGTTAGTTCTGCCGGCGGAGTGGCCATTAAGGCC GGATCTCTGATTGCCGTGCTGATCCTGCGGCAGACCAACAACTACAACAGCGACGACTTCCAGTTCGT GTGGAACATCTACGCCAACAACGACGTGGTGGTGCCTACAGGCGGATGTGATGTGTCCGCCAGAGAT GTGACAGTGACCCTGCCTGATTACCCCGGCTCTGTGCCTATTCCTCTGACCGTGTACTGCGCCAAGTCT CAGAACCTGGGCTACTACCTGAGCGGCACAACAGCCGATGCCGGCAACAGCATCTTTACCAACACCGC CAGCTTCAGCCCTGCTCAAGGTGTTGGAGTGCAGCTGACCCGGAACGGAACAATCATCCCCGCCAACA ATACCGTGTCTCTGGGAGCTGTGGGCACCTCTGCTGTGTCTCTTGGCCTGACAGCCAACTATGCCAGA ACAGGCGGACAAGTGACAGCCGGCAATGTGCAGTCTATCATCGGCGTGACCTTCGTGTATCAGCCTG GCGACGGAAATGCCGACGTGACCATCACAGTGAATGGCAAGGTGGTGGCCAAGGGCAGCTCAGGCT CTGGCTCTGGATCTAAAAAACAGGGCGACGCCGATGTGTGTGGCGAGGTGGCATATATCCAGAGCGT GGTGTCCGATTGTCACGTGCCAACCGCCGAGCTGAGAACCCTGCTGGAAATCCGGAAGCTGTTCCTCG AAATTCAGAAGCTGAAGGTCGAGCTGCAGGGCCTGTCTAAAGAAGGCGGAGGAAGCGGATCTCACC ACCACCATCACCACTGATGAC **[SEQ ID NO: 57]**
FimH_PGDGN_DG_Ferritin (536)	ATGGAAACCGATACACTGCTGCTGTGGGTGCTGTTGCTCTGGGTTCCAGGATCTACAGGGGATGCCGC TCAACCTGCTCGGAGAGCCAGAAGAACAAAGCTGGCCCTGTTTGCCTGCAAGACCGCCAATGGCACA GCCATTCCTATTGGCGGCGGAAGCGCCAATGTGTACGTTAACCTGGCTCCTGCCGTGAACGTGGGCCA GAATCTGGTGGTGGATCTGAGCACCCAGATCTTTTGCCACAACGACTACCCCGAGACAATCACCGACT ACGTGACACTGCAGAGAGGCTCTGCTTACGGCGGCGTGCTGTCTAGCTTTAGCGGCACCGTGAAGTAC AACGGCAGCAGCTACCCATTTCCTACCACCAGCGAGACACCCAGAGTGGTGTACAACAGCAGAACCG ACAAGCCCTGGCCTGTGGCTCTGTACCTGACACCTGTTAGTTCTGCCGGCGGAGTGGCCATTAAGGCC GGATCTCTGATTGCCGTGCTGATCCTGCGGCAGACCAACAACTACAACAGCGACGACTTCCAGTTCGT GTGGAACATCTACGCCAACAACGACGTGGTGGTGCCTACAGGCGGATGTGATGTGTCCGCCAGAGAT GTGACAGTGACCCTGCCTGATTACCCCGGCTCTGTGCCTATTCCTCTGACCGTGTACTGCGCCAAGTCT CAGAACCTGGGCTACTACCTGAGCGGCACAACAGCCGATGCCGGCAACAGCATCTTTACCAACACCGC CAGCTTCAGCCCTGCTCAAGGTGTTGGAGTGCAGCTGACCCGGAACGGAACAATCATCCCCGCCAACA ATACCGTGTCTCTGGGAGCTGTGGGCACCTCTGCTGTGTCTCTTGGCCTGACAGCCAACTATGCCAGA ACAGGCGGACAAGTGACAGCCGGCAATGTGCAGTCTATCATCGGCGTGACCTTCGTGTATCAGCCTG GCGACGGAAATGCCGACGTGACCATCACAGTGAATGGCAAGGTGGTGGCCAAGAGCGGAAGCCACC ACCATCATCACCATCACCACGGCGGCAGCGACATCATCAAGCTGCTGAACGAGCAAGTGAACAAAGA GATGAACAGCAGCAACCTGTACATGAGCATGAGCAGCTGGTGCTACACACACAGCCTGGATGGCGCC GGACTGTTCCTGTTTGATCATGCCGCCGAGGAATACGAGCACGCCAAGAAGCTGATCATCTTCCTGAA CGAGAACAACGTGCCCGTCCAGCTGACATCTATCAGCGCCCCTGAGCACAAGTTCGAGGGCCTGACAC AGATCTTCCAGAAGGCCTACGAACACGAGCAGCACATCAGCGAGAGCATCAACAACATCGTGGACCA CGCCATCAAGAGCAAGGATCACGCCACCTTCAACTTTCTGCAGTGGTACGTGGCCGAACAGCACGAG GAAGAGGTGCTGTTCAAGGACATCCTGGACAAGATCGAGCTGATCGGCAACGAGAACCACGGCCTGT ACCTGGCCGATCAGTATGTGAAGGGAATCGCCAAGAGCCGCAAGTGA **[SEQ ID NO: 58]**

2021392894 29 Aug 2024

FimH-IMX313 j96	ATGGAAACCGATACACTGCTGCTGTGGGTGCTGTTGCTCTGGGTTCCAGGATCTACAGGGGATGCCGC TCAACCTGCTCGGAGAGCCAGAAGAACAAAGCTGGCCCTGTTTGCCTGCAAGACCGCCAATGGCACA GCCATTCCTATTGGCGGCGGAAGCGCCAATGTGTACGTGAACCTGGCTCCTGTGGTCAACGTGGGCCA GAATCTGGTGGTGGACCTGAGCACCCAGATCTTTTGCCACAACGACTACCCCGAGACAATCACCGACT ACGTGACACTGCAGAGAGGCTCTGCTTACGGCGGCGTGCTGAGCAATTTTTCCGGCACAGTGAAGTAC AGCGGCAGCAGCTACCCATTTCCTACCACCAGCGAGACACCCAGAGTGGTGTACAACAGCAGAACCG ACAAGCCCTGGCCTGTGGCTCTGTACCTGACACCTGTTAGTTCTGCCGGCGGAGTGGCCATTAAGGCC GGATCTCTGATTGCCGTGCTGATCCTGCGGCAGACCAACAACTACAACAGCGACGACTTCCAGTTCGT GTGGAACATCTACGCCAACAACGACGTGGTGGTGCCCACCTGTGATGTGTCCGCTAGAGATGTGACA GTGACCCTGCCTGATTACCCCGGCTCTGTGCCTATTCCTCTGACCGTGTACTGCGCCAAGTCTCAGAAC CTGGGCTACTACCTGAGCGGCACAACAGCCGATGCCGGCAACAGCATCTTTACCAACACCGCCAGCTT CAGCCCTGCTCAAGGTGTTGGAGTGCAGCTGACCCGGAACGGAACAATCATCCCCGCCAACAATACCG TGTCTCTGGGAGCTGTGGGCACATCTGCAGTTTCTCTGGGCCTGACAGCCAACTATGCCAGAACAGGC GGACAAGTGACAGCCGGCAATGTGCAGTCTATCATCGGCGTGACATTCGTGTATCAGGGCAGCTCTG GCAGCGGCTCTGGATCTAAAAAACAGGGCGACGCCGATGTGTGTGGCGAGGTGGCATATATCCAGAG CGTGGTGTCCGATTGTCACGTGCCAACCGCCGAGCTGAGAACCCTGCTGGAAATCCGGAAGCTGTTCC TCGAAATTCAGAAGCTGAAGGTCGAGCTGCAGGGCCTGTCTAAAGAAGGCGGAGGAAGCGGATCTC ACCACCACCATCACCACTGA **[SEQ ID NO: 59]**
FimH_mi3 j96	ATGGAAACCGATACACTGCTGCTGTGGGTGCTGTTGCTCTGGGTTCCAGGATCTACAGGGGATGCCGC TCAACCTGCTCGGAGAGCCAGAAGAACAAAGCTGGCCCTGTTTGCCTGCAAGACCGCCAATGGCACA GCCATTCCTATTGGCGGCGGAAGCGCCAATGTGTACGTGAACCTGGCTCCTGTGGTCAACGTGGGCCA GAATCTGGTGGTGGACCTGAGCACCCAGATCTTTTGCCACAACGACTACCCCGAGACAATCACCGACT ACGTGACACTGCAGAGAGGCTCTGCTTACGGCGGCGTGCTGAGCAATTTTTCCGGCACAGTGAAGTAC AGCGGCAGCAGCTACCCATTTCCTACCACCAGCGAGACACCCAGAGTGGTGTACAACAGCAGAACCG ACAAGCCCTGGCCTGTGGCTCTGTACCTGACACCTGTTAGTTCTGCCGGCGGAGTGGCCATTAAGGCC GGATCTCTGATTGCCGTGCTGATCCTGCGGCAGACCAACAACTACAACAGCGACGACTTCCAGTTCGT GTGGAACATCTACGCCAACAACGACGTGGTGGTGCCCACCTGTGATGTGTCCGCTAGAGATGTGACA GTGACCCTGCCTGATTACCCCGGCTCTGTGCCTATTCCTCTGACCGTGTACTGCGCCAAGTCTCAGAAC CTGGGCTACTACCTGAGCGGCACAACAGCCGATGCCGGCAACAGCATCTTTACCAACACCGCCAGCTT CAGCCCTGCTCAAGGTGTTGGAGTGCAGCTGACCCGGAACGGAACAATCATCCCCGCCAACAATACCG TGTCTCTGGGAGCTGTGGGCACATCTGCAGTTTCTCTGGGCCTGACAGCCAACTATGCCAGAACAGGC GGACAAGTGACAGCCGGCAATGTGCAGTCTATCATCGGCGTGACCTTCGTGTACCAAGGATCTGGCG GAGGCGGCATGAAGATGGAAGAACTGTTCAAGAAACACAAGATCGTGGCCGTGCTGCGGGCCAATTC TGTGGAAGAGGCCAAAAAAAAGGCCCTGGCCGTGTTTCTCGGAGGCGTGCACCTGATCGAGATCACC TTTACCGTGCCTGACGCCGACACCGTGATCAAAGAGCTGAGCTTCCTGAAAGAGATGGGCGCCATCAT CGGAGCCGGAACCGTGACATCTGTTGAGCAGGCCAGAAAGGCCGTGGAATCTGGCGCCGAGTTTATC GTGTCCCCTCACCTGGATGAGGAAATCAGCCAGTTCGCCAAAGAAAAGGGCGTGTTCTACATGCCCGG CGTGATGACACCTACAGAGCTGGTCAAAGCCATGAAGCTGGGCCACACCATCCTGAAGCTGTTTCCAG GCGAAGTCGTGGGCCCTCAGTTCGTGAAAGCTATGAAGGGCCCATTTCCAAACGTGAAGTTCGTGCCC ACTGGCGGAGTGAATCTGGACAACGTGTGCGAGTGGTTCAAGGCTGGCGTGCTGGCTGTTGGAGTTG GCTCTGCTCTGGTCAAGGGCACACCTGTGGAAGTGGCTGAGAAGGCCAAGGCCTTCGTGGAAAAGAT CAGAGGCTGTACCGAAGGCAGCGGCTCTGGAAGCGGATCTGGATCTCACCACCATCATCACCATCACC ACTGA **[SEQ ID NO: 60]**

2021392894 29 Aug 2024

FIMHL-HIS-mI3j96	ATGGAAACCGATACACTGCTGCTGTGGGTGCTGTTGCTCTGGGTTCCAGGATCTACAGGGGATGCCGC TCAACCTGCTCGGAGAGCCAGAAGAACAAAGCTGGCCCTGTTTGCCTGCAAGACCGCCAATGGCACA GCCATTCCTATTGGCGGCGGAAGCGCCAATGTGTACGTGAACCTGGCTCCTGTGGTCAACGTGGGCCA GAATCTGGTGGTGGACCTGAGCACCCAGATCTTTTGCCACAACGACTACCCCGAGACAATCACCGACT ACGTGACACTGCAGAGAGGCTCTGCTTACGGCGGCGTGCTGAGCAATTTTTCCGGCACAGTGAAGTAC AGCGGCAGCAGCTACCCATTTCCTACCACCAGCGAGACACCCAGAGTGGTGTACAACAGCAGAACCG ACAAGCCCTGGCCTGTGGCTCTGTACCTGACACCTGTTAGTTCTGCCGGCGGAGTGGCCATTAAGGCC GGATCTCTGATTGCCGTGCTGATCCTGCGGCAGACCAACAACTACAACAGCGACGACTTCCAGTTCGT GTGGAACATCTACGCCAACAACGACGTGGTGGTGCCTACAGGATCTGGCGGACACCACCATCATCACC ATCACCACGGCAGCGGCTCCATGAAGATGGAAGAACTGTTCAAGAAGCACAAGATCGTCGCCGTGCT GCGGGCCAATTCTGTGGAAGAGGCCAAAAAAAAGGCCCTGGCCGTGTTTCTTGGCGGAGTGCACCTG ATCGAGATCACCTTTACCGTGCCTGACGCCGACACCGTGATCAAAGAGCTGAGCTTCCTGAAAGAGAT GGGCGCCATCATCGGAGCCGGAACCGTGACATCTGTTGAGCAGGCCAGAAAGGCCGTGGAATCTGGC GCCGAGTTTATCGTGTCCCCTCACCTGGATGAGGAAATCAGCCAGTTCGCCAAAGAAAAGGGCGTGTT CTACATGCCCGGCGTGATGACACCTACAGAGCTGGTCAAAGCCATGAAGCTGGGCCACACCATCCTGA AGCTGTTTCCAGGCGAAGTCGTGGGCCCTCAGTTCGTGAAAGCTATGAAGGGCCCATTTCCAAACGTG AAGTTCGTGCCCACTGGCGGCGTCAACCTGGATAATGTGTGCGAGTGGTTCAAGGCTGGCGTGCTGG CTGTTGGAGTGGGATCTGCTCTGGTCAAGGGCACACCTGTGGAAGTGGCTGAGAAGGCCAAGGCCTT CGTGGAAAAGATCAGAGGCTGCACCGAGTGA **[SEQ ID NO: 61]**
FIMH_DG_PGDGN_HIS-Ferritin536	ATGTTTGCATGTAAAACCGCAAATGGCACCGCAATTCCGATTGGTGGTGGTAGCGCAAATGTTTATGT TAATCTGGCACCGGCAGTTAATGTTGGTCAGAATCTGGTTGTTGATCTGAGCACCCAGATTTTTTGCCA TAATGATTATCCGGAAACCATCACCGATTATGTTACCCTGCAGCGTGGTAGTGCATATGGTGGTGTTCT GAGCAGCTTTAGCGGCACCGTGAAATATAACGGTAGCAGCTATCCGTTTCCGACCACCAGTGAAACAC CGCGTGTTGTGTATAATAGCCGTACCGATAAACCGTGGCCTGTTGCACTGTATCTGACACCGGTTAGC AGTGCCGGTGGTGTTGCAATTAAAGCAGGTAGCCTGATTGCAGTTCTGATTCTGCGTCAGACCAATAA CTATAACTCCGATGATTTTCAGTTTGTGTGGAACATCTATGCCAATAATGATGTTGTTGTTCCGACCGGT GGTTGTGATGTTAGCGCACGTGATGTTACCGTTACACTGCCGGATTATCCTGGTAGCGTTCCGATTCCG CTGACCGTTTATTGTGCAAAAAGCCAGAACCTGGGTTATTATCTGAGCGGCACCACCGCAGATGCAGG TAATAGCATTTTTACCAATACCGCCAGCTTTAGTCCGGCACAAGGTGTTGGTGTTCAGCTGACCCGTAA TGGCACCATTATTCCGGCAAATAATACCGTTAGCCTGGGTGCAGTTGGCACCAGCGCAGTGAGCCTGG GTCTGACCGCCAATTATGCACGTACCGGTGGTCAGGTTACCGCAGGTAATGTTCAGAGCATTATTGGT GTTACCTTTGTGTATCAGCCTGGTGATGGTAATGCAGATGTGACCATTACCGTGAATGGTAAAGTTGTT GCCAAAAGCGGTAGTCATCATCACCACCATCATCATCACGGTGGTAGCGATATCATCAAACTGCTGAA TGAACAGGTGAACAAAGAAATGAATAGCAGCAACCTGTATATGAGCATGAGCAGCTGGTGTTATACC CATAGCCTGGATGGTGCAGGTCTGTTTCTGTTTGATCATGCAGCCGAAGAATATGAGCACGCAAAAAA ACTGATCATCTTCCTGAATGAAAATAATGTTCCGGTGCAGCTGACCAGCATTAGCGCTCCGGAACATA AATTTGAAGGTCTGACACAGATTTTTCAGAAAGCCTATGAACATGAACAGCACATTAGCGAAAGCATT AACAACATTGTGGATCACGCCATCAAAAGCAAAGATCATGCAACCTTTAACTTTCTGCAGTGGTATGTT GCAGAACAGCATGAAGAAGAAGTGCTGTTTAAAGACATCCTGGATAAAATTGAACTGATCGGCAACG AAAATCATGGTCTGTATCTGGCAGATCAGTATGTTAAAGGTATTGCGAAAAGCCGCAAATAA **[SEQ ID NO: 62]**
LS-FIMHL-	ATGAAGTATCTGCTGCCGACCGCAGCAGCGGGTCTGCTGCTGCTGGCAGCACAGCCTGCAATGGCATT TGCATGTAAAACCGCAAATGGCACCGCAATTCCGATTGGTGGTGGTAGCGCAAATGTTTATGTTAATC TGGCACCGGTTGTTAATGTTGGTCAGAATCTGGTTGTTGATCTGAGCACCCAGATTTTTTGCCATAATG

2021392894 29 Aug 2024

IMX313-HIS	ATTATCCGGAAACCATCACCGATTATGTTACCCTGCAGCGTGGTAGTGCATATGGTGGTGTTCTGAGC AATTTTAGCGGCACCGTGAAATATAGCGGTAGCAGCTATCCGTTTCCGACCACCAGTGAAACACCGCG TGTTGTGTATAATAGCCGTACCGATAAACCGTGGCCTGTTGCACTGTATCTGACACCGGTGAGCAGTG CCGGTGGTGTTGCAATTAAAGCAGGTAGCCTGATTGCAGTTCTGATTCTGCGTCAGACCAATAACTAT AACTCCGATGATTTTCAGTTTGTGTGGAACATCTATGCCAATAATGATGTTGTTGTTCCGACCGGTGGT AGCAGCGGTAGCGGTTCAGGTAGCAAAAAACAGGGTGATGCAGATGTTTGTGGTGAAGTTGCATATA TTCAGAGCGTTGTTAGCGATTGTCATGTTCCGACAGCAGAACTGCGTACCCTGCTGGAAATTCGTAAA CTGTTTCTGGAAATCCAGAAGCTGAAAGTTGAACTGCAGGGTCTGAGCAAAGAAGGTGGCGGAAGCG GTAGCCATCACCATCACCATCACTGA **[SEQ ID NO: 63]**
FIMHL-S24S65-IMX313	ATGTTTGCAAGCAAAACCGCAAATGGCACCGCAATTCCGATTGGTGGTGGTAGCGCAAATGTTTATGT TAATCTGGCACCGGTTGTTAATGTTGGTCAGAATCTGGTTGTTGATCTGAGCACCCAGATTTTTAGCCA TAATGATTATCCGGAAACCATCACCGATTATGTTACCCTGCAGCGTGGTAGTGCATATGGTGGTGTTCT GAGCAATTTTAGCGGCACCGTGAAATATAGCGGTAGCAGCTATCCGTTTCCGACCACCAGTGAAACAC CGCGTGTTGTGTATAATAGCCGTACCGATAAACCGTGGCCTGTTGCACTGTATCTGACACCGGTTAGC AGTGCCGGTGGTGTTGCAATTAAAGCAGGTAGCCTGATTGCAGTTCTGATTCTGCGTCAGACCAATAA CTATAACTCCGATGATTTTCAGTTTGTGTGGAACATCTATGCCAATAATGATGTTGTTGTTCCGACCGGT GGTAGTAGCGGTAGTGGTAGCGGTTCAAAAAAACAGGGTGATGCAGATGTTTGTGGTGAAGTTGCAT ATATTCAGAGCGTTGTTAGCGATTGTCATGTGCCGACCGCAGAACTGCGTACCCTGCTGGAAATTCGT AAACTGTTTCTGGAAATCCAGAAGCTGAAAGTTGAACTGCAGGGTCTGAGTAAAGAAGGTGGTGGTA GTGGTAGCCATCACCATCATCATCACTAATAA **[SEQ ID NO: 64]**
FIMHL-S24S65-foldon-ferritin	ATGTTTGCAAGCAAAACCGCAAATGGCACCGCAATTCCGATTGGTGGTGGTAGCGCAAATGTTTATGT TAATCTGGCACCGGCAGTTAATGTTGGTCAGAATCTGGTTGTTGATCTGAGCACCCAGATTTTTAGCCA TAATGATTATCCGGAAACCATCACCGATTATGTTACCCTGCAGCGTGGTAGTGCATATGGTGGTGTTCT GAGCAGCTTTAGCGGCACCGTGAAATATAACGGTAGCAGCTATCCGTTTCCGACCACCAGTGAAACAC CGCGTGTTGTGTATAATAGCCGTACCGATAAACCGTGGCCTGTTGCACTGTATCTGACACCGGTTAGC AGTGCCGGTGGTGTTGCAATTAAAGCAGGTAGCCTGATTGCAGTTCTGATTCTGCGTCAGACCAATAA CTATAACTCCGATGATTTTCAGTTTGTGTGGAACATCTATGCCAATAATGATGTTGTTGTTCCGACCGGT TCAGGTTATATTCCGGAAGCACCGCGTGATGGTCAGGCATATGTTCGTAAAGATGGTGAATGGGTTCT GCTGAGCACCTTTTTAGGTAGCGGTCATCATCACCATCATCATGGTAGCGGTGATATCATTAAACTGCT GAATGAACAGGTGAACAAAGAGATGAATAGCAGCAATCTGTATATGAGCATGAGCAGCTGGTGTTAT ACCCATAGCCTGGATGGTGCAGGTCTGTTTCTGTTTGATCATGCAGCCGAAGAATATGAGCACGCAAA AAAACTGATCATCTTCCTGAATGAAAATAATGTTCCGGTTCAGCTGACCAGCATTAGCGCTCCGGAACA TAAATTTGAAGGTCTGACACAGATTTTTCAGAAAGCCTATGAACATGAACAGCACATTAGCGAAAGCA TTAACAACATTGTGGATCACGCCATCAAAAGCAAAGATCATGCAACCTTTAACTTTCTGCAGTGGTATG TTGCAGAACAGCATGAAGAAGAAGTGCTGTTTAAAGACATCCTGGATAAAATTGAACTGATCGGCAAC GAAAATCATGGTCTGTATCTGGCAGATCAGTATGTTAAAGGTATTGCCAAAAGCCGCAAGTAATAA **[SEQ ID NO: 65]**
FIMHL-S24S65-Mi3	ATGTTTGCAAGCAAAACCGCAAATGGCACCGCAATTCCGATTGGTGGTGGTAGCGCAAATGTTTATGT TAATCTGGCACCGGCAGTTAATGTTGGTCAGAATCTGGTTGTTGATCTGAGCACCCAGATTTTTAGCCA TAATGATTATCCGGAAACCATCACCGATTATGTTACCCTGCAGCGTGGTAGTGCATATGGTGGTGTTCT GAGCAGCTTTAGCGGCACCGTGAAATATAACGGTAGCAGCTATCCGTTTCCGACCACCAGTGAAACAC CGCGTGTTGTGTATAATAGCCGTACCGATAAACCGTGGCCTGTTGCACTGTATCTGACACCGGTTAGC AGTGCCGGTGGTGTTGCAATTAAAGCAGGTAGCCTGATTGCAGTTCTGATTCTGCGTCAGACCAATAA CTATAACTCCGATGATTTTCAGTTTGTGTGGAACATCTATGCCAATAATGATGTTGTTGTTCCGACCGGT

2021392894 29 Aug 2024

	GGTAGTGGTGGTTCAGGTGGTAGCATGAAAATGGAAGAACTGTTCAAAAAGCACAAAATTGTTGCCG TTCTGCGTGCAAATAGCGTTGAAGAAGCAAAAAAGAAAGCACTGGCCGTTTTTTTAGGTGGTGTGCAT CTGATTGAAATCACCTTTACCGTTCCGGATGCAGATACCGTTATTAAAGAACTGAGCTTTCTGAAAGAA ATGGGTGCAATTATTGGTGCAGGCACCGTTACCAGCGTTGAACAGGCACGTAAAGCAGTTGAAAGCG GTGCAGAATTTATTGTTAGTCCGCATCTGGATGAAGAAATTAGCCAGTTTGCAAAAGAAAAGGGCGTG TTTTATATGCCTGGTGTTATGACCCCGACCGAACTGGTTAAAGCAATGAAACTGGGTCATACCATCCTG AAACTGTTTCCGGGTGAAGTTGTTGGTCCGCAGTTTGTGAAAGCCATGAAAGGTCCTTTTCCGAACGTT AAATTTGTGCCGACAGGTGGCGTGAATCTGGATAATGTTTGTGAATGGTTTAAAGCCGGTGTTCTGGC CGTTGGTGTTGGTAGTGCCCTGGTGAAAGGTACACCGGTTGAAGTTGCAGAAAAAGCAAAAGCCTTT GTGGAAAAAATTCGTGGTTGTACCGAAGGTAGCGGTAGCGGTTCAGGTAGTGGTAGCCATCACCATC ATCATCACTAATAA **[SEQ ID NO: 66]**
FIMHL-mI3	ATGTTTGCATGTAAAACCGCAAATGGCACCGCAATTCCGATTGGTGGTGGTAGCGCAAATGTTTATGT TAATCTGGCACCGGCAGTTAATGTTGGTCAGAATCTGGTTGTTGATCTGAGCACCCAGATTTTTTGCCA TAATGATTATCCGGAAACCATCACCGATTATGTTACCCTGCAGCGTGGTAGTGCATATGGTGGTGTTCT GAGCAGCTTTAGCGGCACCGTGAAATATAACGGTAGCAGCTATCCGTTTCCGACCACCAGTGAAACAC CGCGTGTTGTGTATAATAGCCGTACCGATAAACCGTGGCCTGTTGCACTGTATCTGACACCGGTTAGC AGTGCCGGTGGTGTTGCAATTAAAGCAGGTAGCCTGATTGCAGTTCTGATTCTGCGTCAGACCAATAA CTATAACTCCGATGATTTTCAGTTTGTGTGGAACATCTATGCCAATAATGATGTTGTTGTTCCGACCGGT GGTAGTGGTGGTTCAGGTGGTAGCATGAAAATGGAAGAACTGTTCAAAAAGCACAAAATTGTTGCCG TTCTGCGTGCAAATAGCGTTGAAGAAGCAAAAAAGAAAGCACTGGCCGTTTTTTTAGGTGGTGTGCAT CTGATTGAAATCACCTTTACCGTTCCGGATGCAGATACCGTTATTAAAGAACTGAGCTTTCTGAAAGAA ATGGGTGCAATTATTGGTGCAGGCACCGTTACCAGCGTTGAACAGGCACGTAAAGCAGTTGAAAGCG GTGCAGAATTTATTGTTAGTCCGCATCTGGATGAAGAAATTAGCCAGTTTGCAAAAGAAAAGGGCGTG TTTTATATGCCTGGTGTTATGACCCCGACCGAACTGGTTAAAGCAATGAAACTGGGTCATACCATCCTG AAACTGTTTCCGGGTGAAGTTGTTGGTCCGCAGTTTGTGAAAGCCATGAAAGGTCCTTTTCCGAACGTT AAATTTGTGCCGACAGGTGGCGTGAATCTGGATAATGTTTGTGAATGGTTTAAAGCCGGTGTTCTGGC CGTTGGTGTTGGTAGTGCCCTGGTGAAAGGTACACCGGTTGAAGTTGCAGAAAAAGCAAAAGCCTTT GTGGAAAAAATTCGTGGTTGTACCGAAGGTAGCGGTAGCGGTTCAGGTAGTGGTAGCCATCACCATC ATCATCACTAATAA **[SEQ ID NO: 67]**
FimHL-NOCYS-MI3	ATGTTCGCAAGCAAAACCGCAAATGGCACCGCAATTCCGATTGGTGGTGGTAGCGCAAATGTTTATGT TAATCTGGCACCGGCAGTTAATGTTGGTCAGAATCTGGTTGTTGATCTGAGCACCCAGATTTTTAGCCA TAATGATTATCCGGAAACCATCACCGATTATGTTACCCTGCAGCGTGGTAGTGCATATGGTGGTGTTCT GAGCAATTTTAGCGGCACCGTGAAATATAGCGGTAGCAGCTATCCGTTTCCGACCACCAGTGAAACAC CGCGTGTTGTGTATAATAGCCGTACCGATAAACCGTGGCCTGTTGCACTGTATCTGACACCGGTTAGC AGTGCCGGTGGTGTTGCAATTAAAGCAGGTAGCCTGATTGCAGTTCTGATTCTGCGTCAGACCAATAA CTATAACTCCGATGATTTTCAGTTTGTGTGGAACATCTATGCCAATAATGATGTTGTTGTTCCGACCGGT GGTGGTGGCAGTGGTGGTTCAGGCGGTAGCGGTGGTAGCATGAAAATGGAAGAACTGTTCAAAAAG CACAAAATTGTTGCCGTTCTGCGTGCAAATAGCGTTGAAGAAGCAAAAAAGAAAGCACTGGCCGTTTT TTTAGGTGGTGTGCATCTGATTGAAATCACCTTTACCGTTCCGGATGCAGATACCGTTATTAAAGAACT GAGCTTTCTGAAAGAAATGGGTGCAATTATTGGTGCAGGCACCGTTACCAGCGTTGAACAGGCACGT AAAGCAGTTGAAAGCGGTGCAGAATTTATTGTTAGTCCGCATCTGGATGAAGAAATTAGCCAGTTTGC AAAAGAAAAGGGCGTGTTTTATATGCCTGGTGTTATGACCCCGACCGAACTGGTTAAAGCAATGAAAC TGGGTCATACCATCCTGAAACTGTTTCCGGGTGAAGTTGTTGGTCCGCAGTTTGTGAAAGCCATGAAA GGTCCTTTTCCGAACGTTAAATTTGTGCCGACAGGTGGCGTGAATCTGGATAATGTTTGTGAATGGTTT

2021392894 29 Aug 2024

	AAAGCCGGTGTTCTGGCAGTTGGTGTTGGTAGTGCCCTGGTGAAAGGTACACCGGTTGAAGTTGCAG AAAAAGCAAAAGCCTTTGTGGAAAAAATTCGTGGTTGTACCGAAGGTAGCGGTAGTGGTAGCGGTTC AGGTAGCCATCACCATCACCATCACTGA **[SEQ ID NO: 68]**
FimHdeltaGG_PGDGN_DG_mi3	ATGTTCGCCTGCAAAACCGCAAATGGCACCGCAATTCCGATTGGTGGTGGTAGCGCAAATGTTTATGT TAATCTGGCACCGGTTGTTAATGTTGGTCAGAATCTGGTTGTTGATCTGAGCACCCAGATTTTTTGCCA TAATGATTATCCGGAAACCATCACCGATTATGTTACCCTGCAGCGTGGTAGTGCATATGGTGGTGTTCT GAGCAATTTTAGCGGCACCGTGAAATATAGCGGTAGCAGCTATCCGTTTCCGACCACCAGTGAAACAC CGCGTGTTGTGTATAATAGCCGTACCGATAAACCGTGGCCTGTTGCACTGTATCTGACACCGGTGAGC AGTGCCGGTGGTGTTGCAATTAAAGCAGGTAGCCTGATTGCAGTTCTGATTCTGCGTCAGACCAATAA CTATAACTCCGATGATTTTCAGTTTGTGTGGAACATCTATGCCAATAATGATGTTGTTGTTCCGACCTGT GATGTTAGCGCACGTGATGTTACCGTTACACTGCCGGATTATCCTGGTAGCGTTCCGATTCCGCTGACC GTTTATTGTGCAAAAAGCCAGAACCTGGGTTATTATCTGAGCGGCACCACCGCAGATGCAGGTAATAG CATTTTTACCAATACCGCAAGCTTTAGTCCGGCACAAGGTGTTGGTGTTCAGCTGACCCGTAATGGCAC CATTATTCCGGCAAATAATACCGTTAGCCTGGGTGCAGTTGGCACCAGCGCAGTGAGCCTGGGTCTGA CCGCCAATTATGCACGTACCGGTGGTCAGGTTACCGCAGGTAATGTTCAGAGCATTATTGGTGTTACCT TTGTGTATCAGCCTGGTGATGGTAATGCAGATGTGACCATTACCGTGAATGGTAAAGTTGTTGCAAAA GGTAGCGGTGGTGGTGGCATGAAAATGGAAGAACTGTTCAAAAAACACAAGATTGTTGCCGTTCTGC GTGCAAATAGCGTTGAAGAAGCAAAAAAGAAAGCACTGGCCGTTTTTTTAGGTGGTGTGCATCTGATT GAAATCACCTTTACCGTTCCGGATGCAGATACCGTTATTAAAGAACTGAGCTTTCTGAAAGAAATGGG TGCAATTATTGGCGCAGGCACCGTTACCAGCGTTGAACAGGCACGTAAAGCAGTTGAAAGCGGTGCA GAATTTATTGTTAGTCCGCATCTGGATGAAGAAATTAGCCAGTTTGCAAAAGAAAAGGGCGTGTTTTA TATGCCTGGTGTTATGACCCCGACCGAACTGGTTAAAGCAATGAAACTGGGTCATACCATCCTGAAAC TGTTTCCGGGTGAAGTTGTTGGTCCGCAGTTTGTGAAAGCCATGAAAGGTCCTTTTCCGAACGTTAAAT TTGTGCCGACCGGTGGCGTGAATCTGGATAATGTTTGTGAATGGTTTAAAGCCGGTGTTCTGGCAGTT GGTGTTGGTAGTGCCCTGGTGAAAGGTACACCGGTTGAAGTTGCAGAAAAAGCAAAAGCCTTTGTGG AAAAAATTCGTGGTTGTACCGAAGGTAGTGGTAGCGGCAGCGGTAGCGGTTCACATCACCATCACCAT CACTGA **[SEQ ID NO: 69]**
FimHL-GSG4-Ferritin	ATGGGCAGCAGCCATCATCATCATCATCACGAACTGTACTTCCAGGGCTTTGCATGTAAAACCGCAAAT GGCACCGCAATTCCGATTGGTGGTGGTAGCGCAAATGTTTATGTTAATCTGGCACCGGCAGTTAATGT TGGTCAGAATCTGGTTGTTGATCTGAGCACCCAGATTTTTTGCCATAATGATTATCCGGAAACCATCAC CGATTATGTTACCCTGCAGCGTGGTAGTGCATATGGTGGTGTTCTGAGCAGCTTTAGCGGCACCGTGA AATATAACGGTAGCAGCTATCCGTTTCCGACCACCAGTGAAACACCGCGTGTTGTGTATAATAGCCGT ACCGATAAACCGTGGCCTGTTGCACTGTATCTGACACCGGTTAGCAGTGCCGGTGGTGTTGCAATTAA AGCAGGTAGCCTGATTGCAGTTCTGATTCTGCGTCAGACCAATAACTATAACTCCGATGATTTTCAGTT TGTGTGGAACATCTATGCCAATAATGATGTTGTTGTTCCGACCGGTAGCGGTGGTGGTGGCGATATTA TCAAACTGCTGAATGAACAGGTGAACAAAGAAATGAATAGCAGCAACCTGTATATGAGCATGAGCAG CTGGTGTTATACCCATAGCCTGGATGGTGCAGGTCTGTTTCTGTTTGATCATGCAGCCGAAGAATATGA GCACGCAAAAAAACTGATCATCTTCCTGAATGAAAATAATGTTCCGGTGCAGCTGACCAGCATTAGCG CTCCGGAACATAAATTTGAAGGTCTGACACAGATTTTTCAGAAAGCCTATGAACATGAACAGCACATT AGCGAAAGCATTAACAACATTGTGGATCACGCCATCAAAAGCAAAGATCATGCAACCTTTAACTTTCTG CAGTGGTATGTTGCAGAACAGCATGAAGAAGAAGTGCTGTTTAAAGACATCCTGGATAAAATTGAACT GATCGGCAATGAAAATCACGGTCTGTATCTGGCAGATCAGTATGTTAAAGGTATTGCCAAAAGCCGCA AATAA **[SEQ ID NO: 70]**

2021392894 29 Aug 2024

pelBLS-FimHL-mI3	ATGAAATACCTGCTGCCGACCGCTGCTGCTGGTCTGCTGCTCCTCGCTGCCCAGCCGGCGATGGCCTTT GCATGTAAAACCGCAAATGGCACCGCAATTCCGATTGGTGGTGGTAGCGCAAATGTTTATGTTAATCT GGCACCGGTTGTTAATGTTGGTCAGAATCTGGTTGTTGATCTGAGCACCCAGATTTTTTGCCATAATGA TTATCCGGAAACCATCACCGATTATGTTACCCTGCAGCGTGGTAGTGCATATGGTGGTGTTCTGAGCA ATTTTAGCGGCACCGTGAAATATAGCGGTAGCAGCTATCCGTTTCCGACCACCAGTGAAACACCGCGT GTTGTGTATAATAGCCGTACCGATAAACCGTGGCCTGTTGCACTGTATCTGACACCGGTGAGCAGTGC CGGTGGTGTTGCAATTAAAGCAGGTAGCCTGATTGCAGTTCTGATTCTGCGTCAGACCAATAACTATA ACTCCGATGATTTTCAGTTTGTGTGGAACATCTATGCCAATAATGATGTTGTTGTTCCGACCGGTGGTG GTTCAGGTATGAAAATGGAAGAACTGTTCAAAAAGCACAAGATTGTTGCCGTTCTGCGTGCAAATAGC GTTGAAGAAGCAAAAAAGAAAGCACTGGCCGTTTTTTTAGGTGGTGTGCATCTGATTGAAATCACCTT TACCGTTCCGGATGCAGATACCGTTATTAAAGAACTGAGCTTTCTGAAAGAAATGGGTGCAATTATTG GCGCAGGCACCGTTACCAGCGTTGAACAGGCACGTAAAGCAGTTGAAAGCGGTGCAGAATTTATTGT TAGTCCGCATCTGGATGAAGAAATTAGCCAGTTTGCAAAAGAAAAGGGCGTGTTTTATATGCCTGGTG TTATGACCCCGACCGAACTGGTTAAAGCAATGAAACTGGGTCATACCATCCTGAAACTGTTTCCGGGT GAAGTTGTTGGTCCGCAGTTTGTGAAAGCCATGAAAGGTCCTTTTCCGAACGTTAAATTTGTGCCGAC AGGTGGCGTGAATCTGGATAATGTTTGTGAATGGTTTAAAGCCGGTGTTCTGGCAGTTGGTGTTGGTA GTGCCCTGGTGAAAGGTACACCGGTTGAAGTTGCAGAAAAAGCAAAAGCCTTTGTGGAAAAAATTCG TGGTTGTACCGAAGGTAGTGGTAGCGGTTCAGGTAGCCACCACCACCACCACTGA **[SEQ ID NO: 71]**
FimH_DG_Ferritin (GSGGGG)	ATGGGCAGCAGCCATCATCATCATCATCACGAACTGTACTTCCAGGGCTTTGCATGTAAAACCGCAAAT GGCACCGCAATTCCGATTGGTGGTGGTAGCGCAAATGTTTATGTTAATCTGGCACCGGCAGTTAATGT TGGTCAGAATCTGGTTGTTGATCTGAGCACCCAGATTTTTTGCCATAATGATTATCCGGAAACCATCAC CGATTATGTTACCCTGCAGCGTGGTAGTGCATATGGTGGTGTTCTGAGCAGCTTTAGCGGCACCGTGA AATATAACGGTAGCAGCTATCCGTTTCCGACCACCAGTGAAACACCGCGTGTTGTGTATAATAGCCGT ACCGATAAACCGTGGCCTGTTGCACTGTATCTGACACCGGTTAGCAGTGCCGGTGGTGTTGCAATTAA AGCAGGTAGCCTGATTGCAGTTCTGATTCTGCGTCAGACCAATAACTATAACTCCGATGATTTTCAGTT TGTGTGGAACATCTATGCCAATAATGATGTTGTTGTTCCGACCGGTGGTTGTGATGTTAGCGCACGTG ATGTTACCGTTACACTGCCGGATTATCCTGGTAGCGTTCCGATTCCGCTGACCGTTTATTGTGCAAAAA GCCAGAACCTGGGTTATTATCTGAGCGGCACCACCGCAGATGCAGGTAATAGCATTTTTACCAATACC GCAAGCTTTAGTCCGGCACAAGGTGTTGGTGTTCAGCTGACCCGTAATGGCACCATTATTCCGGCAAA TAATACCGTTAGCCTGGGTGCAGTTGGCACCAGCGCAGTGAGCCTGGGTCTGACCGCCAATTATGCAC GTACCGGTGGTCAGGTTACCGCAGGTAATGTTCAGAGCATTATTGGTGTTACCTTTGTGTATCAGCCTG GTGATGGTAATGCAGATGTGACCATTACCGTGAATGGTAAAGTTGTTGCAAAAGGTAGCGGTGGTGG TGGCGATATTATCAAACTGCTGAATGAACAGGTGAACAAAGAAATGAATAGCAGCAACCTGTATATGA GCATGAGCAGCTGGTGTTATACCCATAGCCTGGATGGTGCAGGTCTGTTTCTGTTTGATCATGCAGCC GAAGAATATGAGCACGCAAAAAAACTGATCATCTTCCTGAATGAAAATAATGTTCCGGTGCAGCTGAC CAGCATTAGCGCTCCGGAACATAAATTTGAAGGTCTGACACAGATTTTTCAGAAAGCCTATGAACATG AACAGCACATTAGCGAAAGCATTAACAACATTGTGGATCACGCCATCAAAAGCAAAGATCATGCAACC TTTAACTTTCTGCAGTGGTATGTTGCAGAACAGCATGAAGAAGAAGTGCTGTTTAAAGACATCCTGGA TAAAATTGAACTGATCGGCAATGAAAATCACGGTCTGTATCTGGCAGATCAGTATGTTAAAGGTATTG CCAAAAGCCGCAAATAA **[SEQ ID NO: 72]**
FimHL-C-C-MI3	ATGTTCGCCTGCAAAACCGCAAATGGCACCGCAATTCCGATTGGTGGTGGTAGCGCAAATGTTTATGT TAATCTGGCACCGGTTTGTAATGTTGGTCAGAATTGTGTTGTTGATCTGAGCACCCAGATTTTTTGCCA TAATGATTATCCGGAAACCATCACCGATTATGTTACCCTGCAGCGTGGTAGTGCATATGGTGGTGTTCT

2021392894 29 Aug 2024

	GAGCAATTTTAGCGGCACCGTGAAATATAGCGGTAGCAGCTATCCGTTTCCGACCACCAGTGAAACAC CGCGTGTTGTGTATAATAGCCGTACCGATAAACCGTGGCCTGTTGCACTGTATCTGACACCGGTGAGC AGTGCCGGTGGTGTTGCAATTAAAGCAGGTAGCCTGATTGCAGTTCTGATTCTGCGTCAGACCAATAA CTATAACTCCGATGATTTTCAGTTTGTGTGGAACATCTATGCCAATAATGATGTTGTTGTTCCGACCGGT GGTGGTGGCAGTGGTGGTTCAGGCGGTAGCGGTGGTAGCATGAAAATGGAAGAACTGTTCAAAAAG CACAAAATTGTTGCCGTTCTGCGTGCAAATAGCGTTGAAGAAGCAAAAAAGAAAGCACTGGCCGTTTT TTTAGGTGGTGTGCATCTGATTGAAATCACCTTTACCGTTCCGGATGCAGATACCGTTATTAAAGAACT GAGCTTTCTGAAAGAAATGGGTGCAATTATTGGTGCAGGCACCGTTACCAGCGTTGAACAGGCACGT AAAGCAGTTGAAAGCGGTGCAGAATTTATTGTTAGTCCGCATCTGGATGAAGAAATTAGCCAGTTTGC AAAAGAAAAGGGCGTGTTTTATATGCCTGGTGTTATGACCCCGACCGAACTGGTTAAAGCAATGAAAC TGGGTCATACCATCCTGAAACTGTTTCCGGGTGAAGTTGTTGGTCCGCAGTTTGTGAAAGCCATGAAA GGTCCTTTTCCGAACGTTAAATTTGTGCCGACAGGTGGCGTGAATCTGGATAATGTTTGTGAATGGTTT AAAGCCGGTGTTCTGGCAGTTGGTGTTGGTAGTGCCCTGGTGAAAGGTACACCGGTTGAAGTTGCAG AAAAAGCAAAAGCCTTTGTGGAAAAAATTCGTGGTTGTACCGAAGGTAGCGGTAGTGGTAGCGGTTC AGGTAGCCATCACCATCACCATCACTGA **[SEQ ID NO: 73]**
FimHL-C-C-qBeta	ATGTTCGCCTGCAAAACCGCAAATGGCACCGCAATTCCGATTGGTGGTGGTAGCGCAAATGTTTATGT TAATCTGGCACCGGTTTGTAATGTTGGTCAGAATTGTGTTGTTGATCTGAGCACCCAGATTTTTTGCCA TAATGATTATCCGGAAACCATCACCGATTATGTTACCCTGCAGCGTGGTAGTGCATATGGTGGTGTTCT GAGCAATTTTAGCGGCACCGTGAAATATAGCGGTAGCAGCTATCCGTTTCCGACCACCAGTGAAACAC CGCGTGTTGTGTATAATAGCCGTACCGATAAACCGTGGCCTGTTGCACTGTATCTGACACCGGTGAGC AGTGCCGGTGGTGTTGCAATTAAAGCAGGTAGCCTGATTGCAGTTCTGATTCTGCGTCAGACCAATAA CTATAACTCCGATGATTTTCAGTTTGTGTGGAACATCTATGCCAATAATGATGTTGTTGTTCCGACCGGT GGTGGTGGCAGTGGTGGTTCAGGCGGTAGCGGTGGCAGCGCCAAACTGGAAACCGTTACACTGGGT AATATTGGTAAAGATGGTAAACAGACCCTGGTTCTGAATCCGCGTGGTGTTAATCCGACCAATGGTGT TGCCAGCCTGAGCCAGGCAGGCGCAGTTCCGGCACTGGAAAAACGTGTTACCGTTAGCGTTAGCCAG CCGAGCCGTAATCGTAAAAACTATAAAGTTCAGGTGAAAATCCAGAATCCGACCGCATGTACCGCCAA TGGTAGCTGTGATCCGAGCGTTACCCGTCAGGCATATGCAGATGTTACCTTTAGTTTTACCCAGTATAG CACCGATGAAGAACGTGCATTTGTTCGTACCGAACTGGCAGCACTGCTGGCAAGTCCGCTGCTGATTG ATGCAATTGATCAGCTGA **[SEQ ID NO: 74]**
HBcAgNC_fimHLsplitted	ATGGATATCGATCCGTATAAAGAATTTGGTGCAAGCGTTGAACTGCTGAGCTTTCTGCCGAGCGATTTT TTTCCGAGCATTCGTGATCTGCTGGATACCGCAAGCGCACTGTATCGTGAAGCACTGGAAAGTCCGGA ACATTGTAGTCCGCATCATACCGCACTGCGTCAGGCAATTCTGTGTTGGGGTGAACTGATGAATCTGG CAACCTGGGTTGGTAGCAATCTGGAAGATCCGTAGAAGGAGATATACATATGTTTGCATGTAAAACCG CAAATGGCACCGCAATTCCGATTGGTGGTGGTAGCGCAAATGTTTATGTTAATCTGGCACCGGTTGTT AATGTTGGTCAGAATCTGGTTGTTGATCTGAGCACCCAGATTTTTTGCCATAATGATTATCCGGAAACC ATCACCGATTATGTTACCCTGCAGCGTGGTAGTGCATATGGTGGTGTTCTGAGCAATTTTAGCGGCACC GTGAAATATAGCGGTAGCAGCTATCCGTTTCCGACCACCAGTGAAACACCGCGTGTTGTGTATAATAG CCGTACCGATAAACCGTGGCCTGTTGCGCTGTATCTGACACCGGTGAGCAGTGCCGGTGGTGTTGCAA TTAAAGCAGGTAGCCTGATTGCAGTTCTGATTCTGCGTCAGACCAATAACTATAACTCCGATGATTTTC AGTTTGTGTGGAACATCTATGCCAATAATGATGTTGTTGTTCCGACCGGTGGTGGTTCAGGTGCCAGC CGTGAACTGGTTGTTAGCTATGTTAATGTGAATATGGGCCTGAAAATTCGTCAGCTGCTGTGGTTTCAT ATTTCATGTCTGACCTTTGGTCGTGAAACCGTTCTGGAATATCTGGTTAGCTTTGGTGTTTGGATTCGTA CCCCTCCGGCATATCGTCCGCCTAATGCACCGATTCTGAGTACCCTGCCGGAAACAACCGTTGTTTGAG GATCC **[SEQ ID NO: 75]**

2021392894 29 Aug 2024

HBcAgNC_fimHL-LS	ATGAAATATCTGCTGCCGACCGCAGCAGCGGGTCTGCTGCTGCTGGCAGCACAGCCTGCAATGGCAG GTCATCATCACCATCATCATAGCGGTGGTATGGATATTGATCCGTATAAAGAATTTGGTGCCAGCGTTG AACTGCTGAGCTTTCTGCCGAGCGATTTTTTTCCGAGCATTCGTGATCTGCTGGATACCGCAAGCGCAC TGTATCGTGAAGCACTGGAAAGTCCGGAACATTGTAGTCCGCATCATACCGCACTGCGTCAGGCAATT CTGTGTTGGGGTGAACTGATGAATCTGGCAACCTGGGTTGGTAGCAATCTGGAAGATCCGTAGAAGG AGATATACATATGAAATACCTGTTACCGACAGCCGCAGCAGGCCTGTTACTGTTAGCAGCCCAGCCAG CCATGGCATTTGCATGTAAAACCGCAAATGGCACCGCAATTCCGATTGGTGGTGGTAGCGCAAATGTT TATGTTAATCTGGCACCGGTTGTTAATGTTGGTCAGAATCTGGTTGTTGATCTGAGCACCCAGATTTTT TGCCATAATGATTATCCGGAAACCATCACCGATTATGTTACCCTGCAGCGTGGTAGTGCATATGGTGGT GTTCTGAGCAATTTTAGCGGCACCGTGAAATATAGCGGTAGCAGCTATCCGTTTCCGACCACCAGTGA AACACCGCGTGTTGTGTATAATAGCCGTACCGATAAACCGTGGCCTGTTGCGCTGTATCTGACACCGG TGAGCAGTGCCGGTGGTGTTGCAATTAAAGCAGGTAGCCTGATTGCAGTTCTGATTCTGCGTCAGACC AATAACTATAACTCCGATGATTTTCAGTTTGTGTGGAACATCTATGCCAATAATGATGTTGTTGTTCCGA CCGGTGGTGGTTCAGGTGCAAGCCGTGAACTGGTTGTTAGCTATGTTAATGTGAATATGGGCCTGAAA ATTCGTCAGCTGCTGTGGTTTCATATTTCATGTCTGACCTTTGGTCGTGAAACCGTTCTGGAATATCTGG TTAGCTTTGGTGTTTGGATTCGTACCCCTCCGGCATATCGTCCGCCTAATGCACCGATTCTGAGTACCCT GCCGGAAACAACCGTTGTTTGACTCGAG **[SEQ ID NO: 76]**
FIMHL-MI3	ATGTTCGCCTGCAAAACCGCAAATGGCACCGCAATTCCGATTGGTGGTGGTAGCGCAAATGTTTATGT TAATCTGGCACCGGTTGTTAATGTTGGTCAGAATCTGGTTGTTGATCTGAGCACCCAGATTTTTTGCCA TAATGATTATCCGGAAACCATCACCGATTATGTTACCCTGCAGCGTGGTAGTGCATATGGTGGTGTTCT GAGCAATTTTAGCGGCACCGTGAAATATAGCGGTAGCAGCTATCCGTTTCCGACCACCAGTGAAACAC CGCGTGTTGTGTATAATAGCCGTACCGATAAACCGTGGCCTGTTGCACTGTATCTGACACCGGTGAGC AGTGCCGGTGGTGTTGCAATTAAAGCAGGTAGCCTGATTGCAGTTCTGATTCTGCGTCAGACCAATAA CTATAACTCCGATGATTTTCAGTTTGTGTGGAACATCTATGCCAATAATGATGTTGTTGTTCCGACCGGT AGCGGTGGTGGTGGCATGAAAATGGAAGAACTGTTCAAAAAACACAAGATTGTTGCCGTTCTGCGTG CAAATAGCGTTGAAGAAGCAAAAAAGAAAGCACTGGCCGTTTTTTTAGGTGGTGTGCATCTGATTGAA ATCACCTTTACCGTTCCGGATGCAGATACCGTTATTAAAGAACTGAGCTTTCTGAAAGAAATGGGTGC AATTATTGGCGCAGGCACCGTTACCAGCGTTGAACAGGCACGTAAAGCAGTTGAAAGCGGTGCAGAA TTTATTGTTAGTCCGCATCTGGATGAAGAAATTAGCCAGTTTGCAAAAGAAAAGGGCGTGTTTTATAT GCCTGGTGTTATGACCCCGACCGAACTGGTTAAAGCAATGAAACTGGGTCATACCATCCTGAAACTGT TTCCGGGTGAAGTTGTTGGTCCGCAGTTTGTGAAAGCCATGAAAGGTCCTTTTCCGAACGTTAAATTTG TGCCGACCGGTGGCGTGAATCTGGATAATGTTTGTGAATGGTTTAAAGCCGGTGTTCTGGCAGTTGGT GTTGGTAGTGCCCTGGTGAAAGGTACACCGGTTGAAGTTGCAGAAAAAGCAAAAGCCTTTGTGGAAA AAATTCGTGGTTGTACCGAAGGTAGTGGTAGCGGCAGCGGTAGCGGTTCACATCACCATCACCATCAC TGA **[SEQ ID NO: 77]**
FimH_DNKQ_DG_deglyc	ATGGAAACCGATACACTGCTGCTGTGGGTGCTGTTGCTCTGGGTTCCAGGATCTACAGGGGATGCCGC TCAACCTGCTCGGAGAGCCAGAAGAACAAAGCTGGCCCTGTTTGCCTGTAAAACCGCCAGCGGCACA GCCATTCCTATTGGCGGAGGCAGCGCCAATGTGTACGTGAACCTGGCTCCTGTGGTCAACGTGGGCCA GAATCTGGTGGTGGACCTGAGCACCCAGATCTTTTGCCACAACGACTACCCCGAGACAATCACCGACT ACGTGACACTGCAGAGAGGCTCTGCTTACGGCGGCGTGCTGAGCGATTTTTCCGGCACAGTGAAGTA CAGCGGCAGCAGCTACCCATTTCCTACCACCAGCGAGACACCCAGAGTGGTGTACAACAGCAGAACC GACAAGCCCTGGCCTGTGGCTCTGTACCTGACACCTGTTAGTTCTGCCGGCGGAGTGGCCATTAAGGC CGGATCTCTGATTGCCGTGCTGATCCTGCGGCAGACCAACAACTACAACAGCGACGACTTCCAGTTCG TGTGGAACATCTACGCCAACAACGACGTGGTGGTGCCTACAGGCGGATGTGATGTGTCCGCCAGAGA

29 Aug 2024
2021392894

	TGTGACAGTGACCCTGCCTGATTACCCCGGCTCTGTGCCTATTCCTCTGACCGTGTACTGCGCCAAGTC TCAGAACCTGGGCTACTACCTGAGCGGCACAACAGCCGATGCCGGCAACAGCATCTTTACCAACACCG CCAGCTTCAGCCCTGCTCAAGGTGTTGGAGTGCAGCTGACCAGAGATGGCACAATCATCCCCGCCGAC AATACCGTGTCTCTGGGCGCTGTTGGCACATCTGCAGTTTCTCTGGGCCTGACCGCCAACTATGCCAGA ACAGGTGGACAAGTGACCGCCGGCAATGTGCAGTCTATCATCGGCGTGACATTCGTGTATCAGGACA ACAAGCAGGCCGACGTGACCATCACCGTGAATGGCAAAGTGGTGGCCAAAGGCTCTGGCCATCACCA CCACCATCACTG **[SEQ ID NO: 90]**
FimH_P GDGN_ DG	ATGGAAACCGACACACTGCTGCTGTGGGTGCTGCTTTTGTGGGTGCCAGGATCTACAGGGGATGCCG CTCAACCTGCTCGGAGAGCCAGAAGAACAAAGCTGGCCCTGTTTGCCTGCAAGACCGCCAATGGCACA GCCATTCCTATTGGCGGCGGAAGCGCCAATGTGTACGTGAACCTGGCTCCTGTGGTCAACGTGGGCCA GAATCTGGTGGTGGACCTGAGCACCCAGATCTTTTGCCACAACGACTACCCCGAGACAATCACCGACT ACGTGACACTGCAGAGAGGCTCTGCTTACGGCGGCGTGCTGAGCAATTTTTCCGGCACAGTGAAGTAC AGCGGCAGCAGCTACCCATTTCCTACCACCAGCGAGACACCCAGAGTGGTGTACAACAGCAGAACCG ACAAGCCCTGGCCTGTGGCTCTGTACCTGACACCTGTTAGTTCTGCCGGCGGAGTGGCCATTAAGGCC GGATCTCTGATTGCCGTGCTGATCCTGCGGCAGACCAACAACTACAACAGCGACGACTTCCAGTTCGT GTGGAACATCTACGCCAACAACGACGTGGTGGTGCCTACAGGCGGATGTGATGTGTCCGCCAGAGAT GTGACAGTGACCCTGCCTGATTACCCCGGCTCTGTGCCTATTCCTCTGACCGTGTACTGCGCCAAGTCT CAGAACCTGGGCTACTACCTGAGCGGCACAACAGCCGATGCCGGCAACAGCATCTTTACCAACACCGC CAGCTTCAGCCCTGCTCAAGGTGTTGGAGTGCAGCTGACCCGGAACGGAACAATCATCCCCGCCAACA ATACCGTGTCTCTGGGAGCTGTGGGCACCTCTGCTGTTTCTCTGGGCCTGACAGCCAACTATGCCAGA ACAGGCGGACAAGTGACAGCCGGCAATGTGCAGTCTATCATCGGCGTGACATTCGTGTATCAGGACA ACAAGCAGGCCGACGTGACCATCACCGTGAATGGCAAAGTGGTGGCCAAAGGCTCTGGCCAT **[SEQ ID NO: 91]**
FimH_D NKQ_D G	ATGGAAACCGACACACTGCTGCTGTGGGTGCTGCTTTTGTGGGTGCCAGGATCTACAGGGGATGCCG CTCAACCTGCTCGGAGAGCCAGAAGAACAAAGCTGGCCCTGTTTGCCTGCAAGACCGCCAATGGCACA GCCATTCCTATTGGCGGCGGAAGCGCCAATGTGTACGTGAACCTGGCTCCTGTGGTCAACGTGGGCCA GAATCTGGTGGTGGACCTGAGCACCCAGATCTTTTGCCACAACGACTACCCCGAGACAATCACCGACT ACGTGACACTGCAGAGAGGCTCTGCTTACGGCGGCGTGCTGAGCAATTTTTCCGGCACAGTGAAGTAC AGCGGCAGCAGCTACCCATTTCCTACCACCAGCGAGACACCCAGAGTGGTGTACAACAGCAGAACCG ACAAGCCCTGGCCTGTGGCTCTGTACCTGACACCTGTTAGTTCTGCCGGCGGAGTGGCCATTAAGGCC GGATCTCTGATTGCCGTGCTGATCCTGCGGCAGACCAACAACTACAACAGCGACGACTTCCAGTTCGT GTGGAACATCTACGCCAACAACGACGTGGTGGTGCCTACAGGCGGATGTGATGTGTCCGCCAGAGAT GTGACAGTGACCCTGCCTGATTACCCCGGCTCTGTGCCTATTCCTCTGACCGTGTACTGCGCCAAGTCT CAGAACCTGGGCTACTACCTGAGCGGCACAACAGCCGATGCCGGCAACAGCATCTTTACCAACACCGC CAGCTTCAGCCCTGCTCAAGGTGTTGGAGTGCAGCTGACCCGGAACGGAACAATCATCCCCGCCAACA ATACCGTGTCTCTGGGAGCTGTGGGCACCTCTGCTGTTTCTCTGGGCCTGACAGCCAACTATGCCAGA ACAGGCGGACAAGTGACAGCCGGCAATGTGCAGTCTATCATCGGCGTGACCTTCGTGTATCAGCCTG GCGACGGAAATGCCGACGTGACCATCACAGTGAATGGCAAGGTGGTGGCCAAAGGCTCTGGACACCA CCACCATCACCACTG **[SEQ ID NO: 92]**
FimH_D eltaGG_ PGDGN _DG	ATGGAAACCGACACACTGCTGCTGTGGGTGCTGCTTTTGTGGGTGCCAGGATCTACAGGGGATGCCG CTCAACCTGCTCGGAGAGCCAGAAGAACAAAGCTGGCCCTGTTTGCCTGCAAGACCGCCAATGGCACA GCCATTCCTATTGGCGGCGGAAGCGCCAATGTGTACGTGAACCTGGCTCCTGTGGTCAACGTGGGCCA GAATCTGGTGGTGGACCTGAGCACCCAGATCTTTTGCCACAACGACTACCCCGAGACAATCACCGACT ACGTGACACTGCAGAGAGGCTCTGCTTACGGCGGCGTGCTGAGCAATTTTTCCGGCACAGTGAAGTAC

2021392894 29 Aug 2024

	AGCGGCAGCAGCTACCCATTTCCTACCACCAGCGAGACACCCAGAGTGGTGTACAACAGCAGAACCG ACAAGCCCTGGCCTGTGGCTCTGTACCTGACACCTGTTAGTTCTGCCGGCGGAGTGGCCATTAAGGCC GGATCTCTGATTGCCGTGCTGATCCTGCGGCAGACCAACAACTACAACAGCGACGACTTCCAGTTCGT GTGGAACATCTACGCCAACAACGACGTGGTGGTGCCCACCTGTGATGTGTCCGCTAGAGATGTGACA GTGACCCTGCCTGATTACCCCGGCTCTGTGCCTATTCCTCTGACCGTGTACTGCGCCAAGTCTCAGAAC CTGGGCTACTACCTGAGCGGCACAACAGCCGATGCCGGCAACAGCATCTTTACCAACACCGCCAGCTT CAGCCCTGCTCAAGGTGTTGGAGTGCAGCTGACCCGGAACGGAACAATCATCCCCGCCAACAATACCG TGTCTCTGGGAGCTGTGGGCACATCTGCTGTTTCTCTGGGCCTGACAGCCAACTATGCCAGAACAGGC GGACAAGTGACAGCCGGCAATGTGCAGTCTATCATCGGCGTGACCTTCGTGTATCAGCCTGGCGACG GAAATGCCGACGTGACCATCACAGTGAATGGCAAGGTGGTGGCCAAAGGCTCTGGACACCACCACCA TCACCACTG **[SEQ ID NO: 93]**
FimH_DGG_sl	ATGGAAACCGACACACTGCTGCTGTGGGTGCTGCTTTTGTGGGTGCCAGGCTCTACAGGCGATTTTGC CTGCAAGACCGCCAACGGCACAGCCATTCCTATTGGCGGAGGCAGCGCCAATGTGTACGTGAACCTG GCTCCTGTGGTCAACGTGGGCCAGAATCTGGTGGTGGACCTGAGCACCCAGATCTTTTGCCACAACGA CTACCCCGAGACAATCACCGACTACGTGACACTGCAGAGAGGCTCTGCTTACGGCGGCGTGCTGAGC AATTTTTCCGGCACAGTGAAGTACAGCGGCAGCAGCTACCCATTTCCTACCACCAGCGAGACACCCAG AGTGGTGTACAACAGCAGAACCGACAAGCCCTGGCCTGTGGCTCTGTACCTGACACCTGTTAGTTCTG CCGGCGGAGTGGCCATTAAGGCCGGATCTCTGATTGCCGTGCTGATCCTGCGGCAGACCAACAACTAC AACAGCGACGACTTCCAGTTCGTGTGGAACATCTACGCCAACAACGACGTGGTGGTGCCCACCTGTGA TGTGTCCGCTAGAGATGTGACAGTGACCCTGCCTGATTACCCCGGCTCTGTGCCTATTCCTCTGACCGT GTACTGCGCCAAGAGCCAGAACCTGGGCTACTACCTGTCTGGCACAACAGCCGATGCCGGCAACAGC ATCTTTACCAACACCGCCAGCTTCAGCCCTGCTCAAGGTGTTGGAGTGCAGCTGACCCGGAACGGAAC AATCATCCCCGCCAACAATACCGTGTCTCTGGGAGCTGTGGGCACATCTGCTGTTTCTCTGGGCCTGAC CGCCAATTATGCCAGAACAGGCGGACAAGTGACCGCCGGCAATGTGCAGTCTATCATCGGCGTGACC TTCGTGTATCAGCCTGGCGACGGAAACGCCGATGTGACCATCACAGTGAATGGCAAGGTGGTGGCCA AAGGCTCTGGACACCACCACCATCACCACTGA **[SEQ ID NO: 94]**
FimH_PGDGN_sl	ATGGAAACCGACACACTGCTGCTGTGGGTGCTGCTTTTGTGGGTGCCAGGCTCTACAGGCGATTTTGC CTGCAAGACCGCCAACGGCACAGCCATTCCTATTGGCGGAGGCAGCGCCAATGTGTACGTGAACCTG GCTCCTGTGGTCAACGTGGGCCAGAATCTGGTGGTGGACCTGAGCACCCAGATCTTTTGCCACAACGA CTACCCCGAGACAATCACCGACTACGTGACACTGCAGAGAGGCTCTGCTTACGGCGGCGTGCTGAGC AATTTTTCCGGCACAGTGAAGTACAGCGGCAGCAGCTACCCATTTCCTACCACCAGCGAGACACCCAG AGTGGTGTACAACAGCAGAACCGACAAGCCCTGGCCTGTGGCTCTGTACCTGACACCTGTTAGTTCTG CCGGCGGAGTGGCCATTAAGGCCGGATCTCTGATTGCCGTGCTGATCCTGCGGCAGACCAACAACTAC AACAGCGACGACTTCCAGTTCGTGTGGAACATCTACGCCAACAACGACGTGGTGGTGCCTACAGGCG GATGTGATGTGTCCGCCAGAGATGTGACAGTGACCCTGCCTGATTACCCCGGCTCTGTGCCTATTCCTC TGACCGTGTACTGCGCCAAGAGCCAGAACCTGGGCTACTACCTGTCTGGCACAACAGCCGATGCCGGC AACAGCATCTTTACCAACACCGCCAGCTTCAGCCCTGCTCAAGGTGTTGGAGTGCAGCTGACCCGGAA CGGAACAATCATCCCCGCCAACAATACCGTGTCTCTGGGAGCTGTGGGCACCTCTGCTGTTTCTCTGGG CCTGACAGCCAACTATGCCAGAACAGGCGGACAAGTGACAGCCGGCAATGTGCAGTCTATCATCGGC GTGACCTTCGTGTATCAGCCTGGCGACGGAAACGCCGATGTGACCATCACAGTGAATGGCAAGGTGG TGGCCAAAGGCTCTGGACACCACCACCATCACCACTGACTCGAG **[SEQ ID NO: 95]**

2021392894 29 Aug 2024

FimH_DNKQ_sl	ATGGAAACCGACACACTGCTGCTGTGGGTGCTGCTTTTGTGGGTGCCAGGCTCTACAGGCGATTTTGC CTGCAAGACCGCCAACGGCACAGCCATTCCTATTGGCGGAGGCAGCGCCAATGTGTACGTGAACCTG GCTCCTGTGGTCAACGTGGGCCAGAATCTGGTGGTGGACCTGAGCACCCAGATCTTTTGCCACAACGA CTACCCCGAGACAATCACCGACTACGTGACACTGCAGAGAGGCTCTGCTTACGGCGGCGTGCTGAGC AATTTTTCCGGCACAGTGAAGTACAGCGGCAGCAGCTACCCATTTCCTACCACCAGCGAGACACCCAG AGTGGTGTACAACAGCAGAACCGACAAGCCCTGGCCTGTGGCTCTGTACCTGACACCTGTTAGTTCTG CCGGCGGAGTGGCCATTAAGGCCGGATCTCTGATTGCCGTGCTGATCCTGCGGCAGACCAACAACTAC AACAGCGACGACTTCCAGTTCGTGTGGAACATCTACGCCAACAACGACGTGGTGGTGCCTACAGGCG GATGTGATGTGTCCGCCAGAGATGTGACAGTGACCCTGCCTGATTACCCCGGCTCTGTGCCTATTCCTC TGACCGTGTACTGCGCCAAGAGCCAGAACCTGGGCTACTACCTGTCTGGCACAACAGCCGATGCCGGC AACAGCATCTTTACCAACACCGCCAGCTTCAGCCCTGCTCAAGGTGTTGGAGTGCAGCTGACCCGGAA CGGAACAATCATCCCCGCCAACAATACCGTGTCTCTGGGAGCTGTGGGCACCTCTGCTGTTTCTCTGGG CCTGACAGCCAACTATGCCAGAACAGGCGGACAAGTGACAGCCGGCAATGTGCAGTCTATCATCGGC GTGACATTCGTGTATCAGGACAACAAGCAGGCCGACGTGACCATCACCGTGAATGGCAAAGTGGTGG CCAAAGGCTCTGGCCATCACCACCACCATCACTGACTCGAG **[SEQ ID NO: 96]**
FIMH_DG_PGDGN_536-MI3	ATGGAAACCGATACACTGCTGCTGTGGGTGCTGTTGCTCTGGGTTCCAGGCTCTACAGGCGATTTTGC CTGCAAGACCGCCAACGGCACAGCCATTCCTATTGGCGGAGGCAGCGCCAATGTGTACGTTAACCTGG CTCCTGCCGTGAACGTGGGCCAGAATCTGGTGGTGGATCTGAGCACCCAGATCTTTTGCCACAACGAC TACCCCGAGACAATCACCGACTACGTGACACTGCAGAGAGGCTCTGCTTACGGCGGCGTGCTGTCTAG CTTTAGCGGCACCGTGAAGTACAACGGCAGCAGCTACCCATTTCCTACCACCAGCGAGACACCCAGAG TGGTGTACAACAGCAGAACCGACAAGCCCTGGCCTGTGGCTCTGTACCTGACACCTGTTAGTTCTGCC GGCGGAGTGGCCATTAAGGCCGGATCTCTGATTGCCGTGCTGATCCTGCGGCAGACCAACAACTACA ACAGCGACGACTTCCAGTTCGTGTGGAACATCTACGCCAACAACGACGTGGTGGTGCCTACAGGCGG ATGTGATGTGTCCGCCAGAGATGTGACAGTGACCCTGCCTGATTACCCCGGCTCTGTGCCTATTCCTCT GACCGTGTACTGCGCCAAGAGCCAGAACCTGGGCTACTACCTGTCTGGCACAACAGCCGATGCCGGC AACAGCATCTTTACCAACACCGCCAGCTTCAGCCCTGCTCAAGGTGTTGGAGTGCAGCTGACCCGGAA CGGAACAATCATCCCCGCCAACAATACCGTGTCTCTGGGAGCTGTGGGCACCTCTGCTGTGTCTCTTGG CCTGACAGCCAACTATGCCAGAACAGGCGGACAAGTGACAGCCGGCAATGTGCAGTCTATCATCGGC GTGACCTTCGTGTATCAGCCTGGCGACGGAAACGCCGATGTGACCATCACAGTGAATGGCAAGGTGG TGGCCAAGAGCGGAAGCCACCACCATCATCACCATCACCACGGCGGCAGCATGAAGATGGAAGAACT GTTCAAGAAGCACAAGATCGTCGCCGTGCTGCGGGCCAATTCTGTGGAAGAGGCCAAAAAAAAGGCC CTGGCCGTGTTTCTTGGCGGAGTGCACCTGATCGAGATCACCTTTACCGTGCCTGACGCCGACACCGT GATCAAAGAGCTGAGCTTCCTGAAAGAGATGGGCGCCATCATCGGAGCCGGCACAGTGACATCTGTT GAGCAGGCCAGAAAGGCCGTGGAATCTGGCGCCGAGTTTATCGTGTCCCCTCACCTGGATGAGGAAA TCAGCCAGTTCGCCAAAGAAAAGGGCGTGTTCTACATGCCCGGCGTGATGACACCTACAGAGCTGGTC AAAGCCATGAAGCTGGGCCACACCATCCTGAAGCTGTTTCCAGGCGAAGTCGTGGGCCCTCAGTTCGT GAAAGCTATGAAGGGCCCATTTCCAAACGTGAAGTTCGTGCCCACTGGCGGCGTGAACCTGGATAAT GTGTGCGAGTGGTTCAAGGCTGGCGTGCTGGCTGTTGGAGTTGGCTCTGCTCTGGTCAAGGGCACAC CTGTGGAAGTGGCTGAGAAGGCCAAGGCCTTCGTGGAAAAGATCAGAGGCTGCACCGAGTGA **[SEQ ID NO: 97]**
HBcFIMHLJ96	ATGGAAACCGATACACTGCTGCTGTGGGTGCTGTTGCTCTGGGTTCCAGGATCTACCGGCGACGACAT CGACCCCTACAAAGAGTTTGGCGCCAGCGTCGAGCTGCTGAGCTTCCTGCCTAGCGACTTCTTCCCTTC CATCCGGGATCTGCTGGATACCGCTAGCGCCCTGTATAGAGAGGCCCTGGAAAGCCCTGAGCACTGCT CTCCACATCACACAGCCCTGAGACAGGCCATCCTGTGTTGGGGCGAACTGATGAATCTGGCCACCTGG

2021392894 29 Aug 2024

	GTCGGAAGCAACCTGGAAGATCCTGGTTCTGGCGGCGGAGGCTTTGCCTGTAAAACAGCCAATGGCA CCGCCATTCCTATCGGAGGCGGCAGCGCCAATGTGTACGTTAACCTGGCTCCTGTGGTCAACGTGGGC CAGAATCTGGTGGTGGACCTGAGCACCCAGATCTTTTGCCACAACGACTACCCCGAGACAATCACCGA CTACGTGACACTGCAGAGAGGCTCTGCTTACGGCGGCGTGCTGAGCAATTTTTCCGGCACAGTGAAGT ACAGCGGCAGCAGCTACCCATTTCCTACCACCAGCGAGACACCCAGAGTGGTGTACAACAGCAGAACC GACAAGCCCTGGCCTGTGGCTCTGTACCTGACACCTGTTAGTTCTGCTGGCGGAGTGGCCATCAAGGC CGGATCTCTGATTGCCGTGCTGATCCTGCGGCAGACCAACAACTACAACAGCGACGACTTCCAGTTCG TGTGGAACATCTACGCCAACAACGACGTGGTGGTGCCTACAGGCGGAGGATCTGGCGGAGCTTCTAG AGAACTGGTCGTGTCCTACGTGAACGTGAACATGGGCCTGAAGATCCGGCAGCTGCTCTGGTTTCACA TCAGCTGTCTGACCTTCGGCCGGGAAACCGTGCTGGAATACCTGGTGTCCTTCGGCGTGTGGATCAGA ACCCCTCCTGCCTATAGACCTCCTAACGCTCCCATCCTGAGCACACTGCCTGAGACAACAGTTGTTGGA AGCGGAGGCGGAGGCCACCACCATCACCATCAT **[SEQ ID NO: 98]**
HBcFIM HDGJ96	ATGGAGACCGACACCCTGCTGCTGTGGGTGCTGCTGCTGTGGGTGCCCGGCAGCACCGGCGACGACA TCGACCCCTACAAGGAGTTCGGCGCCAGCGTGGAGCTGCTGAGCTTCCTGCCCAGCGACTTCTTCCCC AGCATCCGGGACCTGCTGGACACCGCCAGCGCCCTGTACCGGGAGGCCCTGGAGAGCCCCGAGCACT GCAGCCCCCACCACACCGCCCTGCGGCAGGCCATCCTGTGCTGGGGCGAGCTGATGAACCTGGCCAC CTGGGTGGGCAGCAACCTGGAGGACCCCGGCAGCGGCGGCGGCGGCTTCGCCTGCAAGACCGCCAA CGGCACCGCCATCCCCATCGGCGGCGGCAGCGCCAACGTGTACGTGAACCTGGCCCCCGTGGTGAAC GTGGGCCAGAACCTGGTGGTGGACCTGAGCACCCAGATCTTCTGCCACAACGACTACCCCGAGACCAT CACCGACTACGTGACCCTGCAGCGGGGCAGCGCCTACGGCGGCGTGCTGAGCAACTTCAGCGGCACC GTGAAGTACAGCGGCAGCAGCTACCCCTTCCCCACCACCAGCGAGACCCCCCGGGTGGTGTACAACA GCCGGACCGACAAGCCCTGGCCCGTGGCCCTGTACCTGACCCCCGTGAGCAGCGCCGGCGGCGTGGC CATCAAGGCCGGCAGCCTGATCGCCGTGCTGATCCTGCGGCAGACCAACAACTACAACAGCGACGACT TCCAGTTCGTGTGGAACATCTACGCCAACAACGACGTGGTGGTGCCCACCGGCGGCTGCGACGTGAG CGCCCGGGACGTGACCGTGACCCTGCCCGACTACCCCGGCAGCGTGCCCATCCCCCTGACCGTGTACT GCGCCAAGAGCCAGAACCTGGGCTACTACCTGAGCGGCACCACCGCCGACGCCGGCAACAGCATCTT CACCAACACCGCCAGCTTCAGCCCCGCCCAGGGCGTGGGCGTGCAGCTGACCCGGAACGGCACCATC ATCCCCGCCAACAACACCGTGAGCCTGGGCGCCGTGGGCACCAGCGCCGTGAGCCTGGGCCTGACCG CCAACTACGCCCGGACCGGCGGCCAGGTGACCGCCGGCAACGTGCAGAGCATCATCGGCGTGACCTT CGTGTACCAGCCCGGCGACGGCAACGCCGACGTGACCATCACCGTGAACGGCAAGGTGGTGGCCAA GGGCAGCGGCGGCGGCGGCGCCAGCCGGGAGCTGGTGGTGAGCTACGTGAACGTGAACATGGGCCT GAAGATCCGGCAGCTGCTGTGGTTCCACATCAGCTGCCTGACCTTCGGCCGGGAGACCGTGCTGGAG TACCTGGTGAGCTTCGGCGTGTGGATCCGGACCCCCCCCGCCTACCGGCCCCCCAACGCCCCCATCCTG AGCACCCTGCCCGAGACCACCGTGGTGGGCAGCGGCGGCGGCGGCCACCACCACCACCACCAC **[SEQ ID NO: 99]**

29 Aug 2024
2021392894

THE CLAIMS DEFINING THE INVENTION ARE AS FOLLOWS:

1. A polypeptide comprising or consisting of an amino acid sequence X-(a)-L-(b)-Y, wherein "(a)" is a FimH polypeptide; "L" is a first linker; "(b)" is a donor-strand complementing amino acid sequence, "X" is an optional N-terminal amino acid sequence; "Y is an optional C-terminal amino acid sequence, wherein "Y" is not derived from FimC or FimH or a fragment thereof,

   wherein "(b)" comprises or consists of an amino acid sequence according to SEQ ID NO: 5, and

   wherein the first linker or "L" comprises or consists of PGDGN **[SEQ ID NO: 7]**.

2. The polypeptide of claim 1, wherein one or more amino acids known or predicted to be N-glycosylated or O-glycosylated are substituted with serine (S), alanine (A), aspartic acid (D) or glutamine (Q).

3. The polypeptide of claim 2, wherein (a) includes one or more of the following amino acid substitutions relative to SEQ ID NO: 2: N28S, N91D, N249D, N256D, or at the positions of SEQ ID NO: 101, 103 and 105 corresponding those positions of SEQ ID NO:2, for example, one, two, three or four of the amino acid substitutions.

4. The polypeptide any one of the preceding claims, wherein the first linker comprises or consists of polar amino acids.

5. The polypeptide of any one of the preceding claims, wherein the polypeptide comprises a nanoparticle domain at the N-terminus or C-terminus, optionally wherein "X" or "Y" comprise a nanoparticle domain.

6. The polypeptide of claim 5, wherein the nanoparticle domain is selected from the group consisting of:

   (i) ferritin (for example, **[SEQ ID NO: 15]** or [SEQ ID NO: 109] (*Helicobacter pylori*), **[SEQ ID NO: 16]** (*Escherichia coli*)), or any one of [SEQ ID NO: 149]-[SEQ ID NO: 152] (stabilized *Escherichia coli*)), or a variant and/or fragment thereof,

   (ii) iMX313 (for example **[SEQ ID NO: 17]**), or a variant and/or fragment thereof,

   (iii) mI3 (for example **[SEQ ID NO: 18]**), or a variant and/or fragment thereof,

   (iv) encapsulin (for example **[SEQ ID NO: 19]**), or a variant and/or fragment thereof, or

   (v) Self-assembling viral coat proteins, such as *Acinetobacter* phage AP205 coat protein (NCBI Reference Sequence: NP_085472.1), Hepatitis B virus core protein (HBc) **[SEQ ID NO: 110]**, or bacteriophage Qβ **[SEQ ID NO: 111]**, or a variant and/or fragment thereof.

7. A nanoparticle comprising a polypeptide monomer comprising or consisting of an amino acid sequence that has:

2021392894 29 Aug 2024

(i) at least 80% sequence identity, for example at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to the amino acid sequence SEQ ID NO: 16 and has one or more mutations from the group consisting of: glycine (G) at the position that aligns to residue 34 of SEQ ID NO: 16 (T34G mutation), aspartic acid (D) at the position that aligns to residue 70 of SEQ ID NO: 16 (N70D mutation), isoleucine (I) at the position that aligns to residue 72 of SEQ ID NO: 16 (V72I mutation) and alanine (A) at the position that aligns to residue 124 of SEQ ID NO: 16 (S124A mutation),

(ii) an amino acid sequence that has at least 80% sequence identity, for example at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to the amino acid sequence SEQ ID NO: 16 and has glycine (G) at the position that aligns to residue 34 of SEQ ID NO: 16 (T34G mutation), aspartic acid (D) at the position that aligns to residue 70 of SEQ ID NO: 16 (N70D mutation), isoleucine (I) at the position that aligns to residue 72 of SEQ ID NO: 16 (V72I mutation) and alanine (A) at the position that aligns to residue 124 of SEQ ID NO: 16 (S124A mutation), optionally wherein the polypeptide monomer comprises an amino acid sequence that has at least 80% sequence identity,

(iii) at least 80% sequence identity, for example at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to the amino acid sequence SEQ ID NO: 16 and has glycine (G) at the position that aligns to residue 34 of SEQ ID NO: 16 (T34G mutation), isoleucine (I) at the position that aligns to residue 72 of SEQ ID NO: 16 (V72I mutation) and alanine (A) at the position that aligns to residue 124 of SEQ ID NO: 16 (S124A mutation), optionally wherein the polypeptide monomer comprises an amino acid sequence that has at least 80% sequence identity,

(iv) at least 80% sequence identity, for example at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to the amino acid sequence SEQ ID NO: 16 and has glycine (G) at the position that aligns to residue 34 of SEQ ID NO: 16 (T34G mutation), and alanine (A) at the position that aligns to residue 124 of SEQ ID NO: 16 (S124A mutation), optionally wherein the polypeptide monomer comprises an amino acid sequence that has at least 80% sequence identity,

(v) at least 80% sequence identity, for example at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to the amino acid sequence SEQ ID NO: 16 and has glycine (G) at the position that aligns to residue 34 of SEQ ID NO: 16 (T34G mutation), optionally wherein the polypeptide monomer comprises an amino acid sequence that has at least 80% sequence identity, and

wherein the exterior surface structure or interior surface structure of the nanoparticle carries one or more antigen comprising or consisting of the polypeptide of any one of claims 1 to 6.

2021392894 29 Aug 2024

8. The nanoparticle of claim 7, wherein the polypeptide monomer comprises an amino acid sequence selected from the group consisting of: SEQ ID NO: 149, SEQ ID NO: 150, SEQ ID NO: 151, and SEQ ID NO: 152.

9. The nanoparticle of claim 7 or 8, comprising or consisting of an amino acid sequence that has the amino acid sequence SEQ ID NO: 130 or 153.

10. The polypeptide or nanoparticle of any one of the preceding claims, wherein the polypeptide comprises or consists of an amino acid sequence SEQ ID NO: 123 or SEQ ID NO:124.

11. The polypeptide or nanoparticle of any one of the preceding claims, wherein the mannose binding of the polypeptide is at least 20% lower than that of native FimH complexed with native FimC (FimHC complex), e.g., at least 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100% lower.

12. The polypeptide or nanoparticle of any one of the preceding claims, wherein FimH is in the low affinity conformation, i.e. in the tense (T) state.

13. A nucleic acid encoding a polypeptide according to any one of claims 1-6 and 10-12 or a polypeptide of a nanoparticle according to any one of claims 7-12.

14. A vector comprising the nucleic acid of claim 13.

15. A cell comprising a nucleic acid of claim 13 or a vector of claim 14.

16. A pharmaceutical composition comprising the polypeptide or nanoparticle of any one of claims 1-12, the nucleic acid of claim 13, or a vector of claim 14.

17. A vaccine comprising the polypeptide or nanoparticle of any one of claims 1-12, a nucleic acid of claim 13 or a vector of claim 14.

18. The vaccine of claim 17 further comprising an adjuvant.

19. The vaccine of claim 18, wherein the adjuvant comprises 3D-MPL, QS21 and liposomes comprising cholesterol.

20. The vaccine of any one of claims 17-19, wherein the vaccine elicits a protective immune response after one or two doses.

21. A polypeptide or nanoparticle of any one of claims 1-12, a nucleic acid of claim 13, a vector of claim 14, or a vaccine of any one of claims 17-20 for use in medicine.

22. A polypeptide or nanoparticle of any one of claims 1-12, a nucleic acid of claim 13, a vector of claim 14, or a vaccine of any one of claims 17-20 for use in raising an immune response in a mammal, for example, for treating and/or preventing one or more disease.

23. Use of a polypeptide or nanoparticle of any one of claims 1-12, a nucleic acid of claim 13, a vector of claim 14, or a vaccine of any one of claims 17-20 for the manufacture of a medicament

2021392894 29 Aug 2024

for raising an immune response in a mammal, for example, for treating and/or preventing one or more disease.

24. A method of raising an immune response in a mammal, the method comprising or consisting of administering the mammal with an effective amount of a polypeptide or nanoparticle of any one of claims 1-12, a nucleic acid of claim 13, a vector of claim 14, or a vaccine of any one of claims 17-20.

FIG. 1A

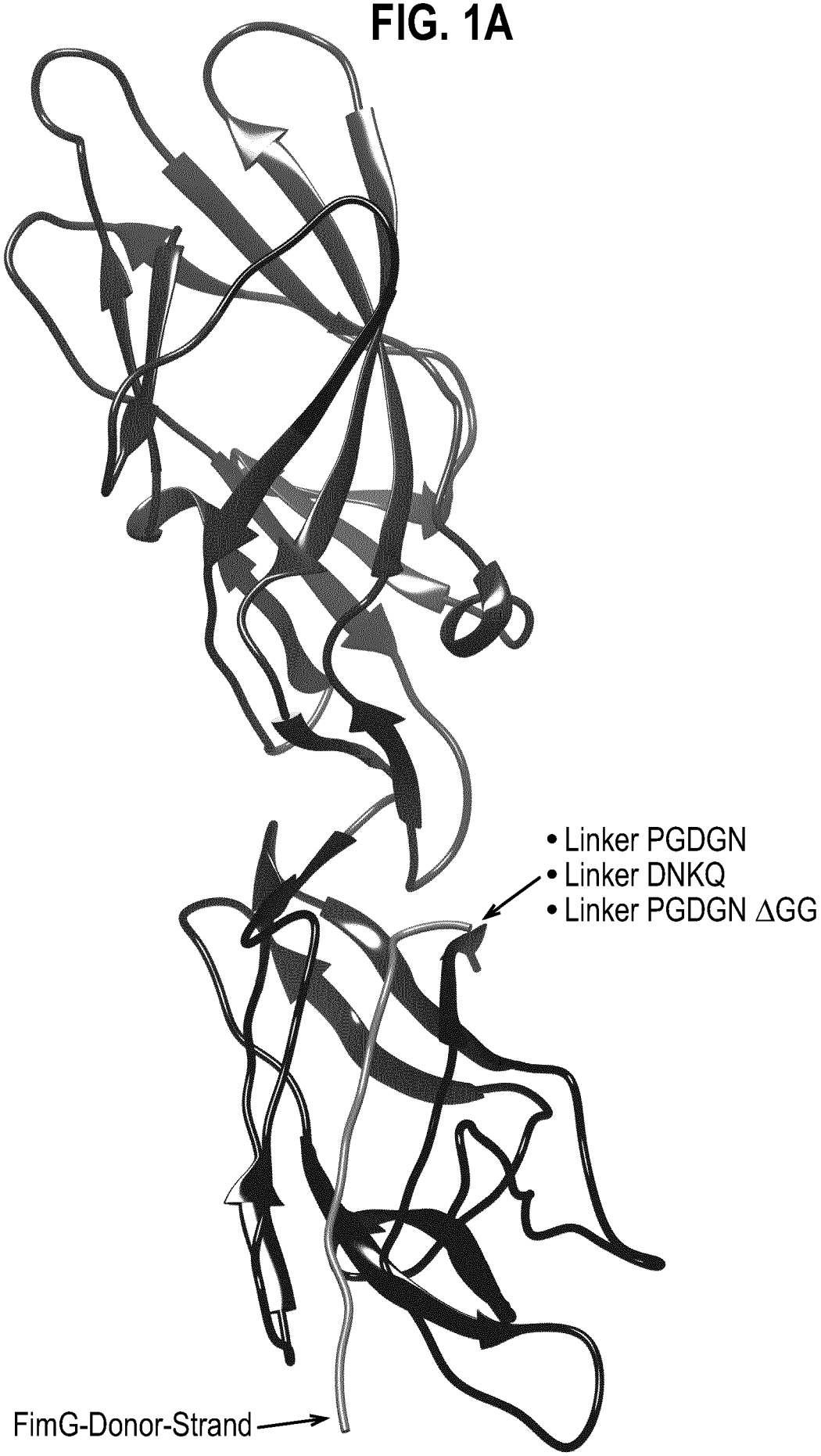

• Linker PGDGN
• Linker DNKQ
• Linker PGDGN ΔGG
FimG-Donor-Strand

FIG. 1B

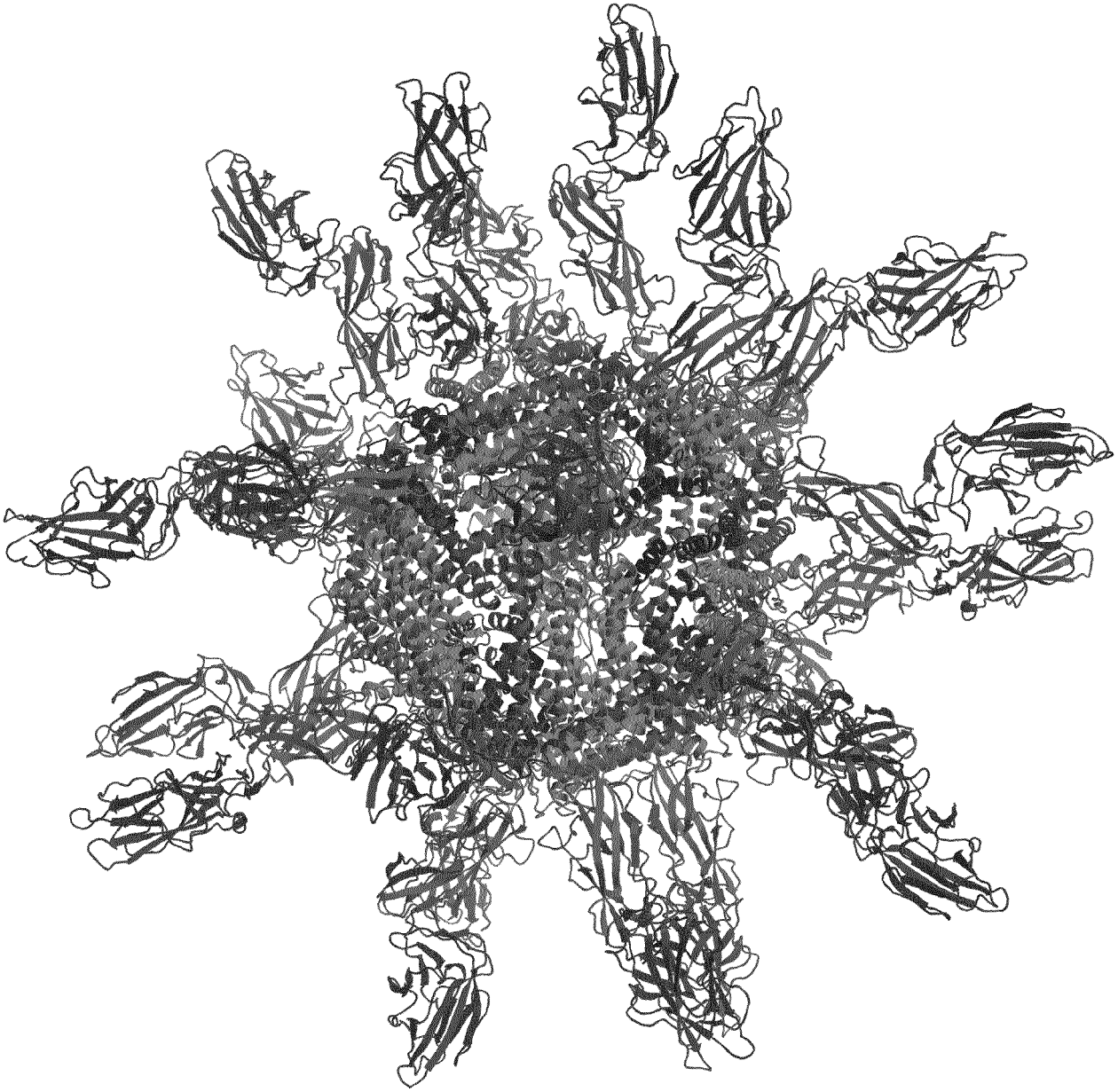


FIG. 2A Cytoplasmic expression

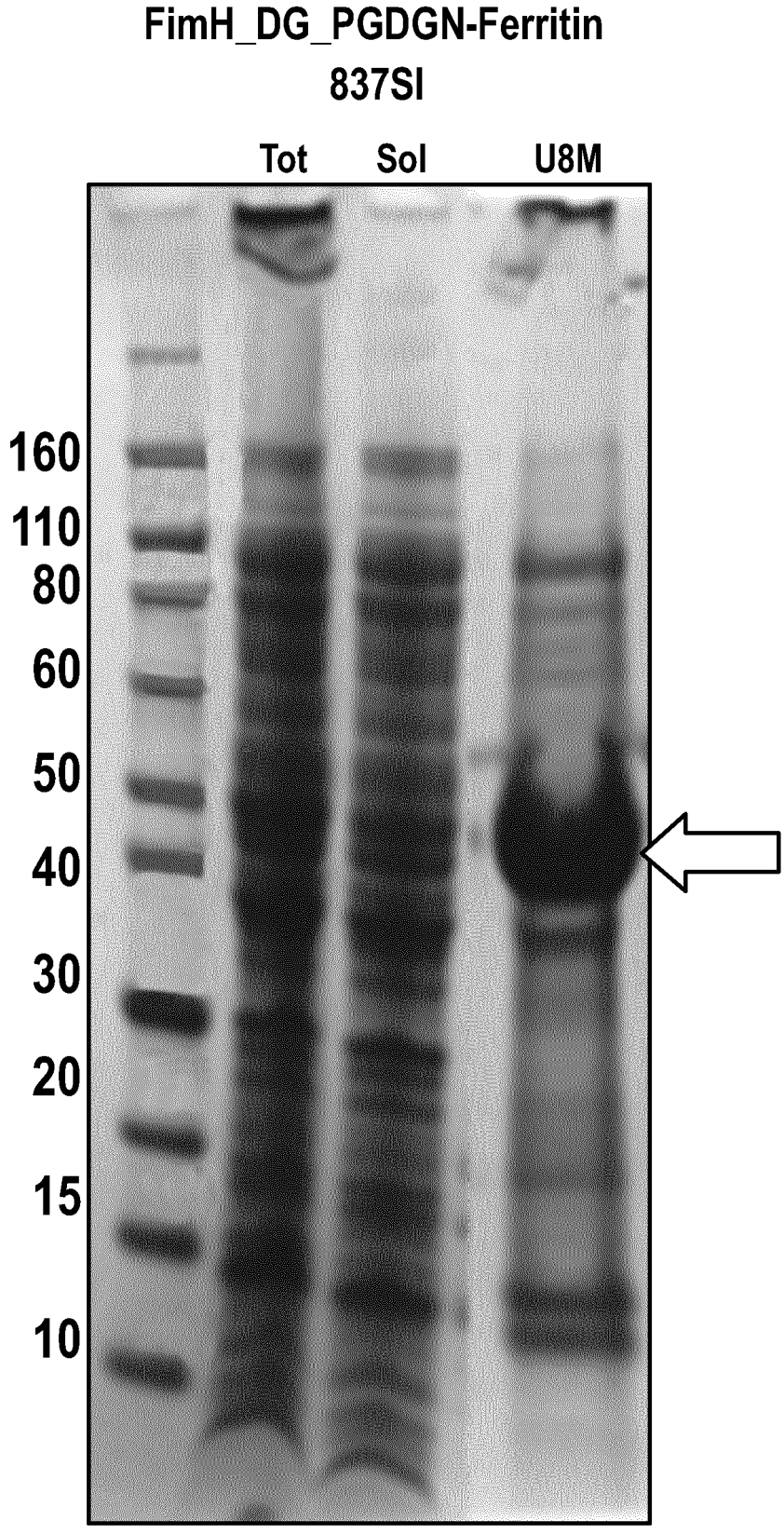


FIG. 2A(contd) Cytoplasmic expression

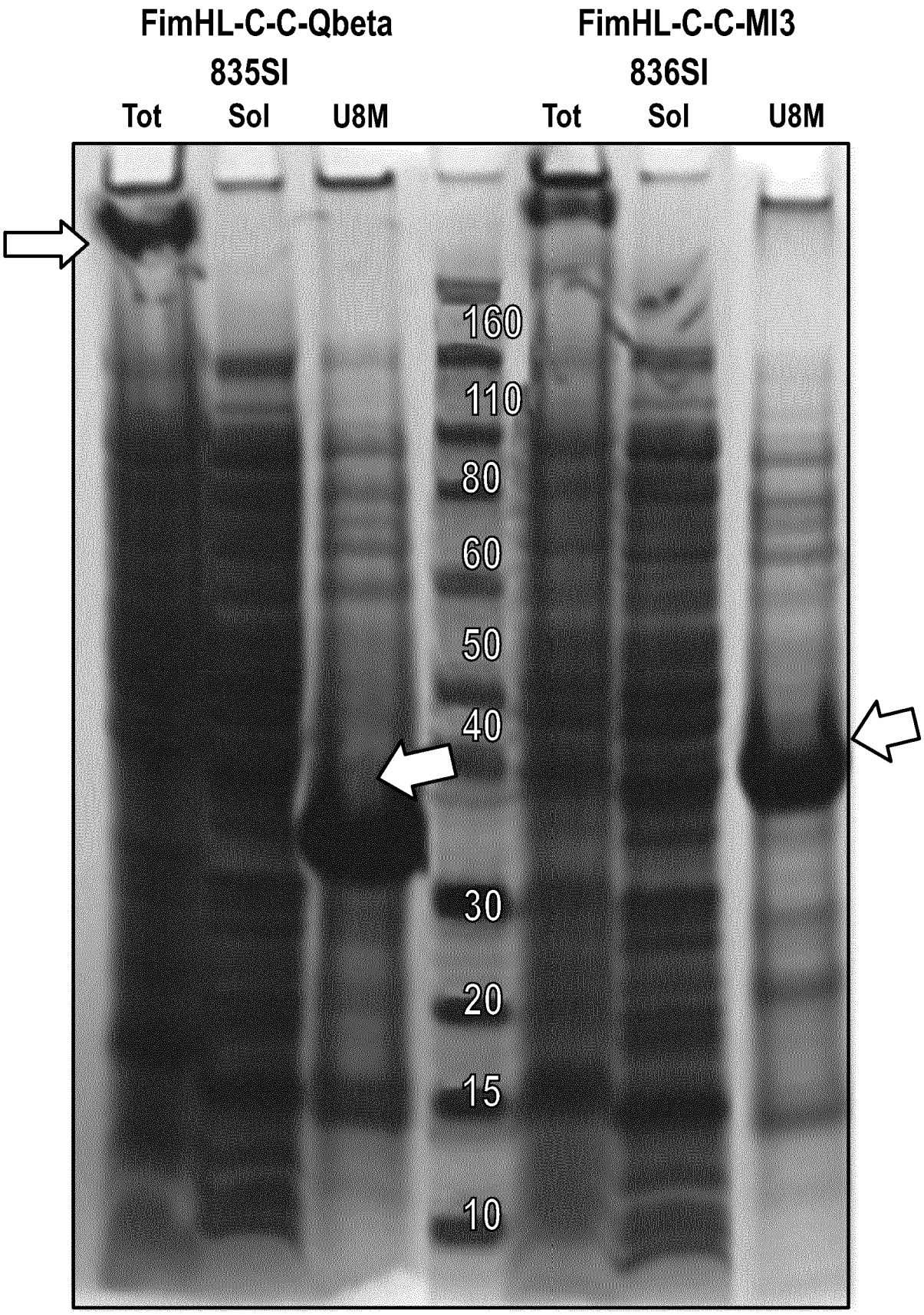


FIG. 2A(contd) Cytoplasmic expression

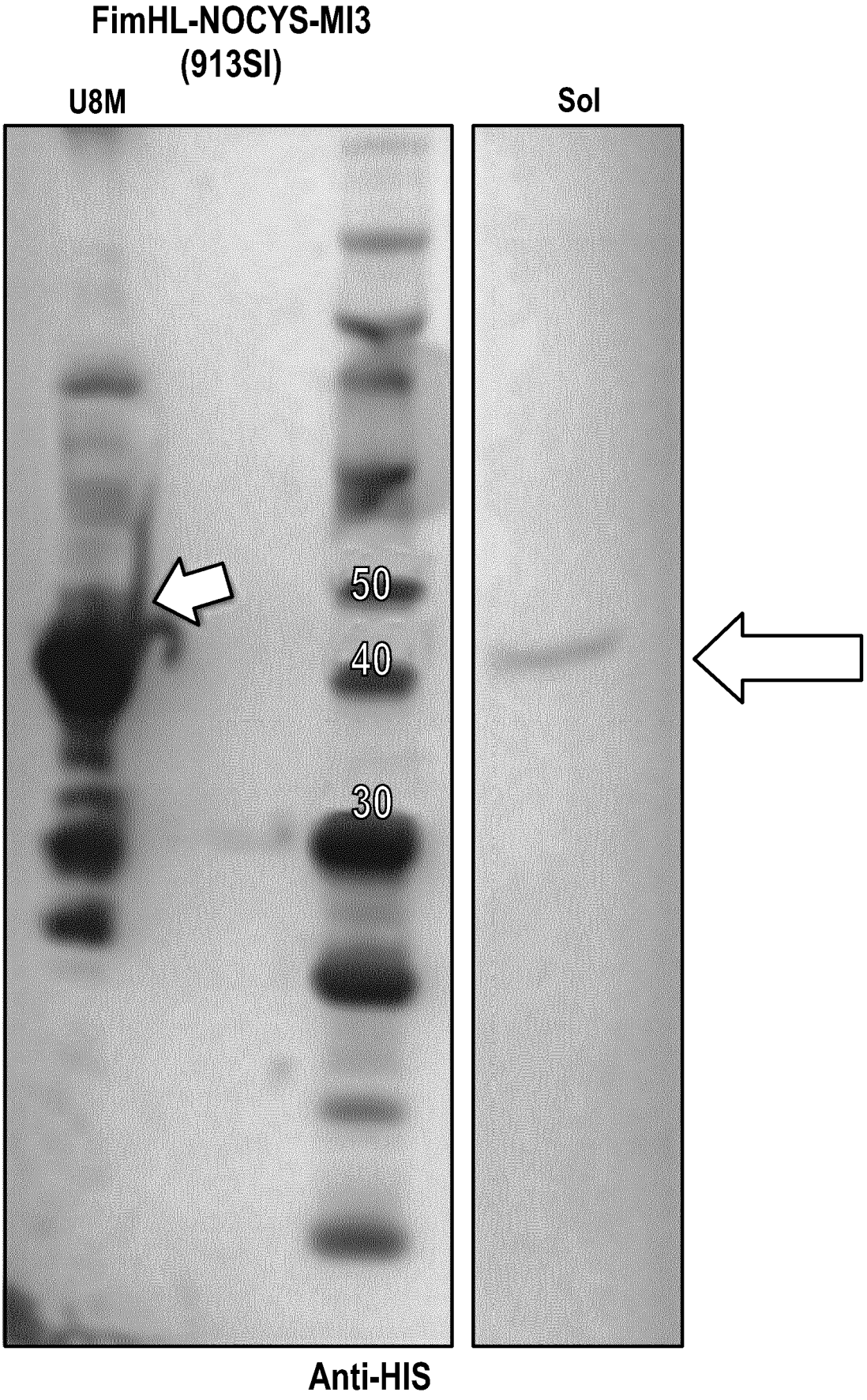


FIG. 2B

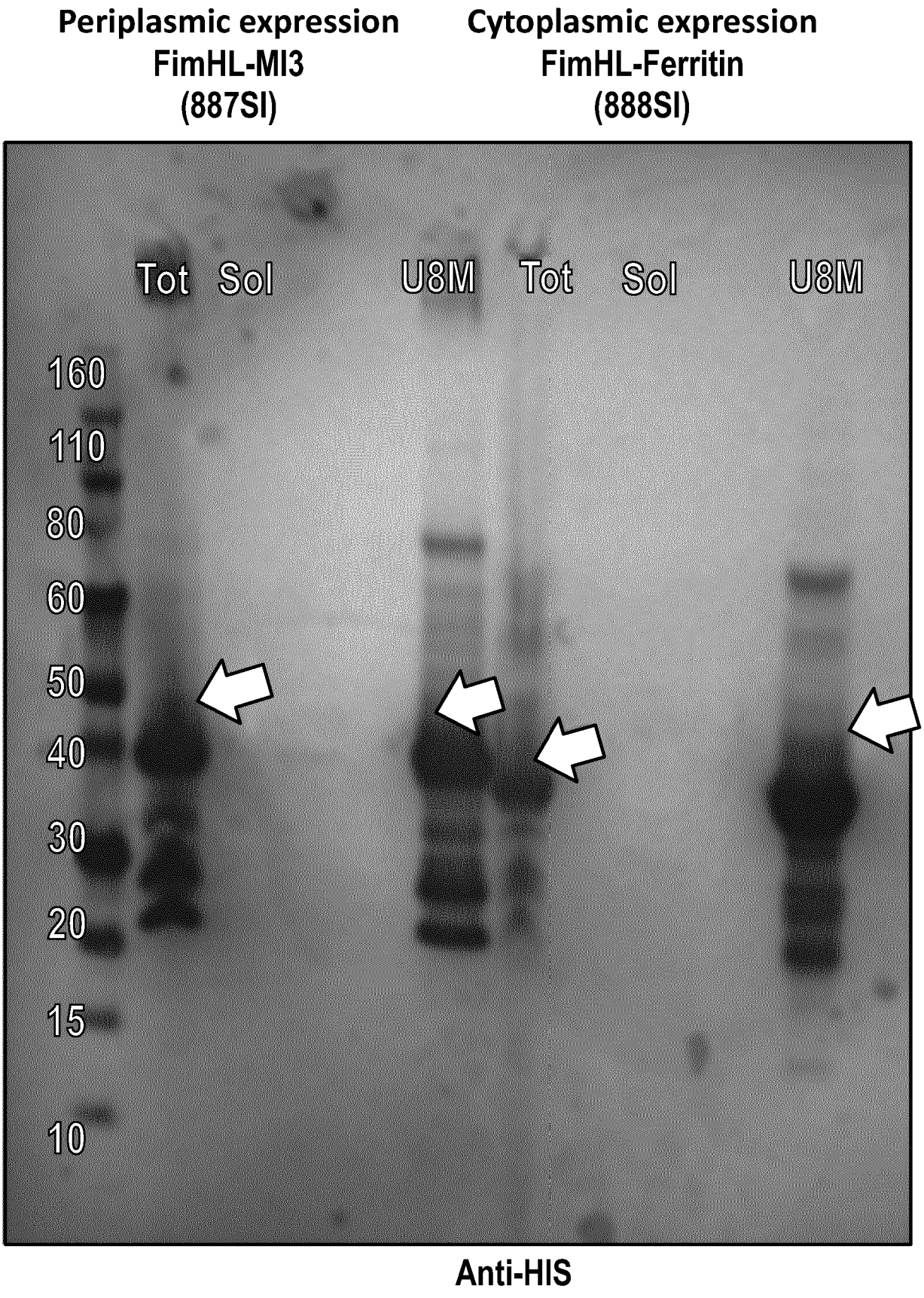

Periplasmic expression
FimHL-MI3
(887SI)
Cytoplasmic expression
FimHL-Ferritin
(888SI)
Tot
Sol
U8M
Tot
Sol
U8M
160
110
80
60
50
40
30
20
15
10
Anti-HIS

FIG. 3

CENTERFORBIOLOGICALSEQUENCEANALYSIS CBS

NetNGlyc 1.0 Server - prediction results

Technical University of Denmark

Asn-Xaa-Ser/Thr sequons in the sequence output below are highlighted in blue.
Asparagines predicted to be N-glycosylated are highlighted in red.

Output for 'Sequence'

```
Name:  Sequence          Length:  300
MKRVITLFAVLLMGWSVNAWSFACKTANGTAIPIGGGSANVYVNLAPVVNVGQNLVVDLSTQIFCHNDYPETITDYVTLQ      80
RGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGSLIAVLILRQTNNYNS     160
DDFQFVWNIYANNDVVVPTGGCDVSARDVTVTLPDYPGSVPIPLTVYCAKSQNLGYYLSGTTADAGNSIFTNTASFSPAQ     240
GVGVQLTRNGTIIPANNTVSLGAVGTSAVSLGLTANYARTGGQVTAGNVQSIIGVTFVYQ
.............................N.....................................................      80
..........N.....................................................................      160
................................................................................      240
........N.....................................................                         320

(Threshold=0.5)
--------------------------------------------------------------------------------
SeqName      Position  Potential   Jury     N-Glyc
                                   agreement result
--------------------------------------------------------------------------------
Sequence       28 NGTA   0.7688     (9/9)    +++
Sequence       91 NFSG   0.6250     (9/9)    ++
Sequence      249 NGTI   0.7063     (9/9)    ++
Sequence      256 NNTV   0.4498     (6/9)    -
--------------------------------------------------------------------------------
```

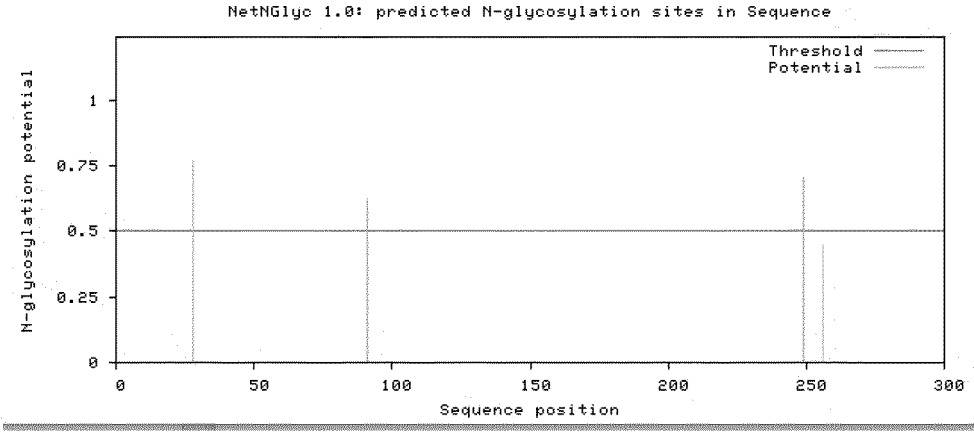


FIG. 4

M	Novex sharp protein marker (Invitrogen LC5800)	MW
1	FimH_ΔGG_PGDGN_DG	34
2	FimH_DNKQ_DG	34
3	FimH_PGDGN_DG	34
4	FimH_C-His	32

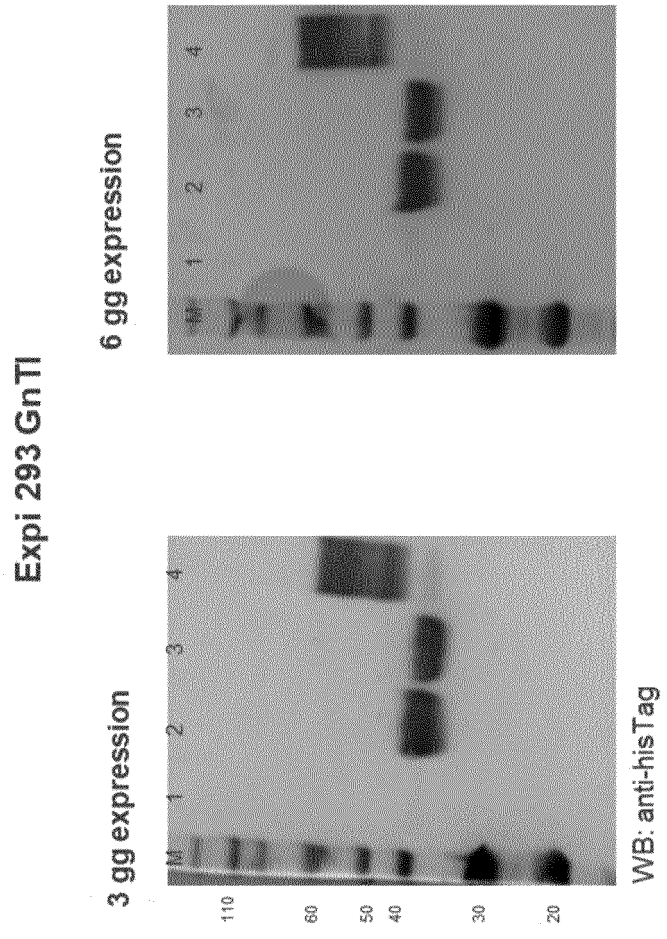


FIG. 5

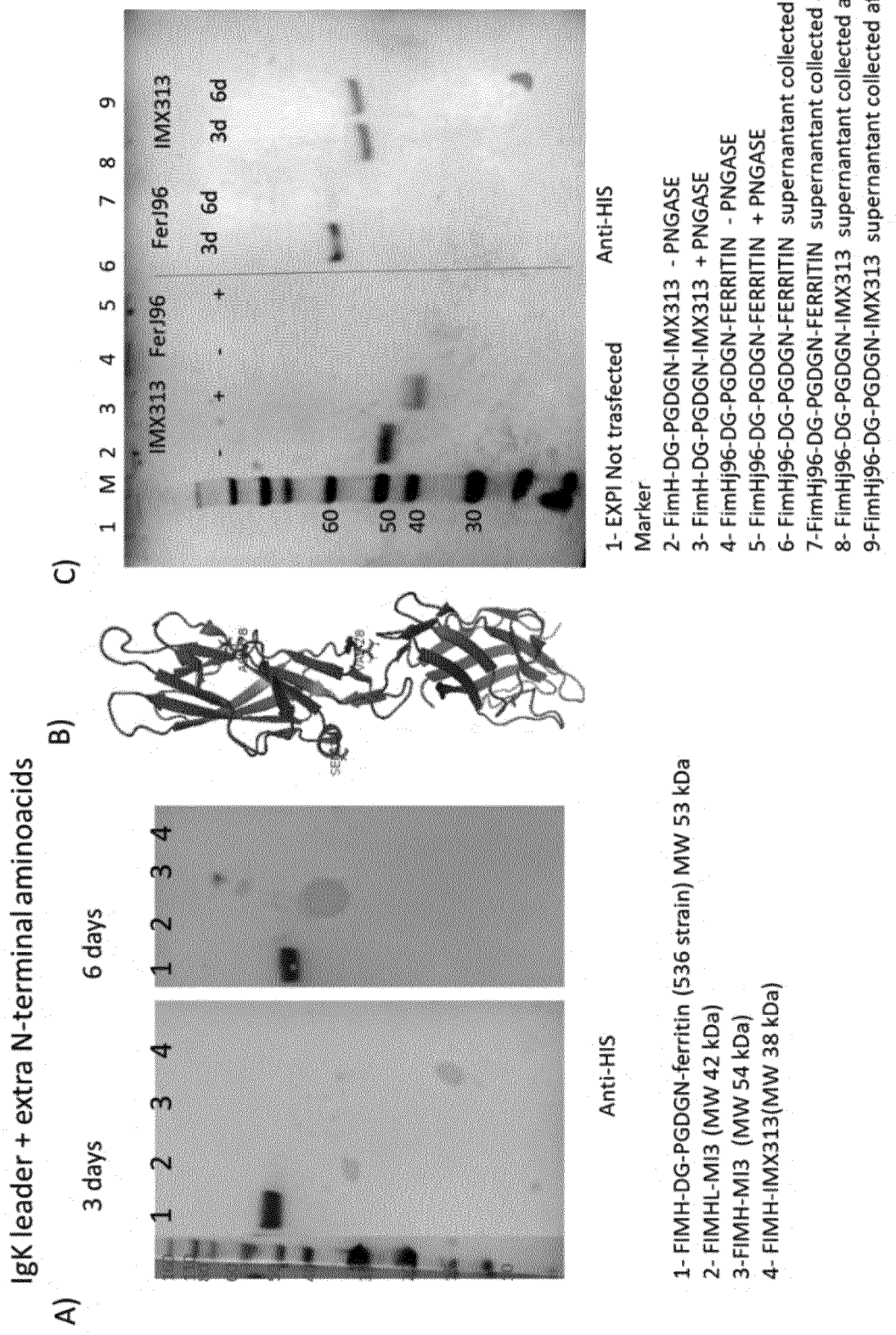
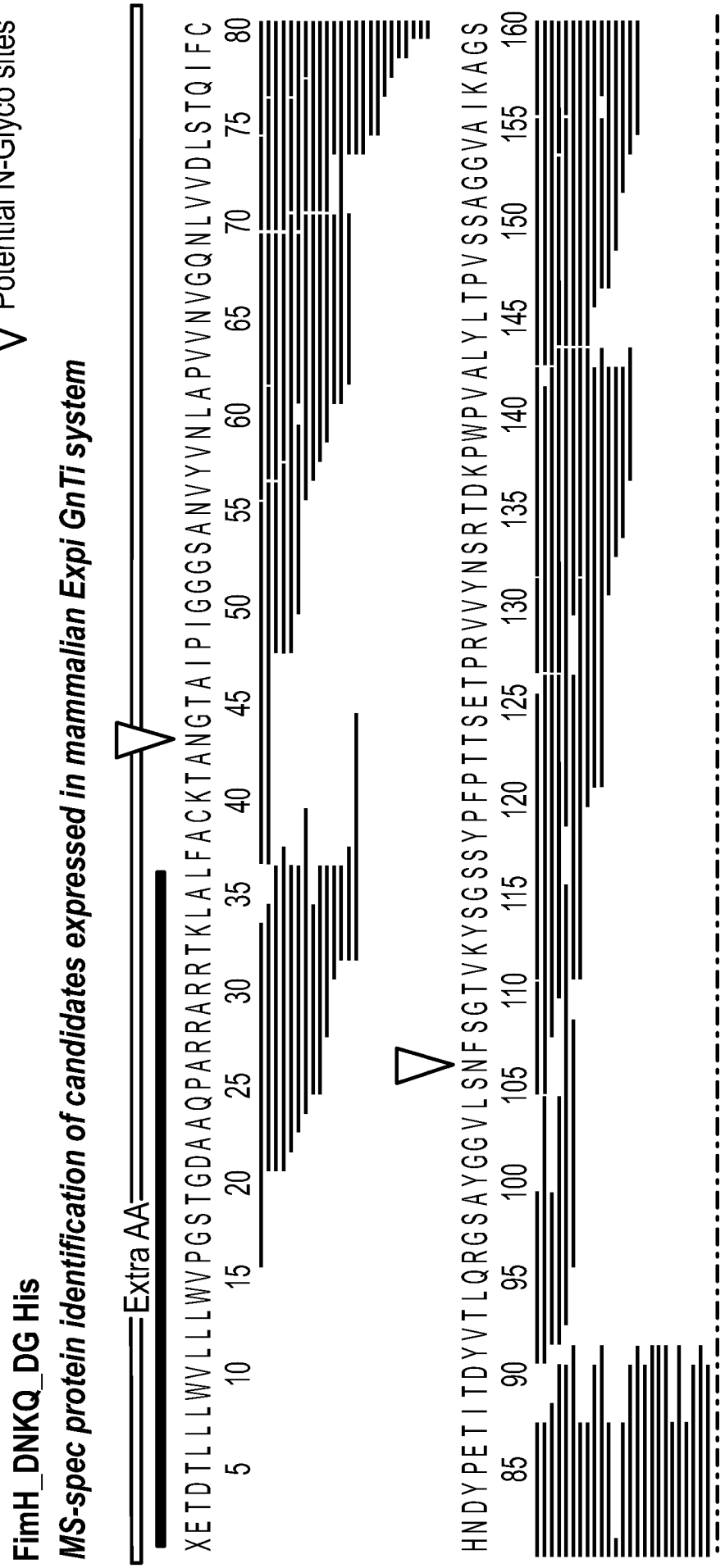

A)
IgK leader + extra N-terminal aminoacids
3 days
6 days
1 2 3 4
1 2 3 4
Anti-HIS
1- FIMH-DG-PGDGN-ferritin (536 strain) MW 53 kDa
2- FIMHL-MI3 (MW 42 kDa)
3-FIMH-MI3 (MW 54 kDa)
4- FIMH-IMX313(MW 38 kDa)
B)
C)
1 M 2 3 4 5 6 7 8 9
IMX313 FerJ96 FerJ96 IMX313
- + - + 3d 6d 3d 6d
60
50
40
30
Anti-HIS
1- EXPI Not trasfected
Marker
2- FimH-DG-PGDGN-IMX313 - PNGASE
3- FimH-DG-PGDGN-IMX313 + PNGASE
4- FimHj96-DG-PGDGN-FERRITIN - PNGASE
5- FimHj96-DG-PGDGN-FERRITIN + PNGASE
6- FimHj96-DG-PGDGN-FERRITIN supernantant collected after 3days
7-FimHj96-DG-PGDGN-FERRITIN supernantant collected after 6days
8- FimHj96-DG-PGDGN-IMX313 supernantant collected after 3days
9-FimHj96-DG-PGDGN-IMX313 supernantant collected after 6days

FIG. 6


FimH_DNKQ_DG His
MS-spec protein identification of candidates expressed in mammalian Expi GnTi system
Potential N-Glyco sites
Extra AA
XETDTLLLWVLLLWVPGSTGDAAQPARRARRTKLALFACKTANGTAIPIGGGSANVYVNLAPVVNVGQNLVVDLSTQIFC
HNDYPETITDYVTLQRGSAYGGVLSNFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGS
5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80
85 90 95 100 105 110 115 120 125 130 135 140 145 150 155 160

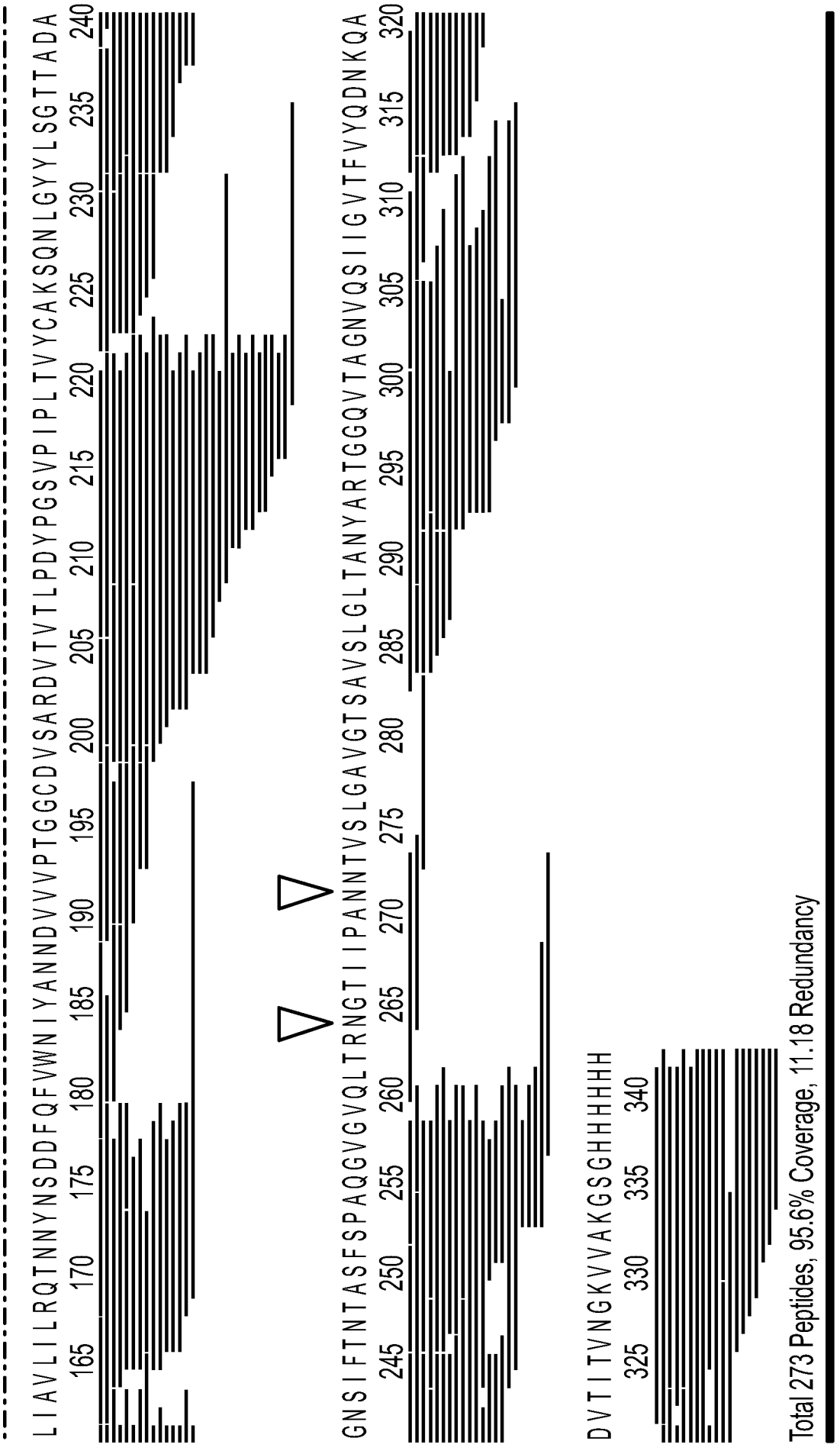
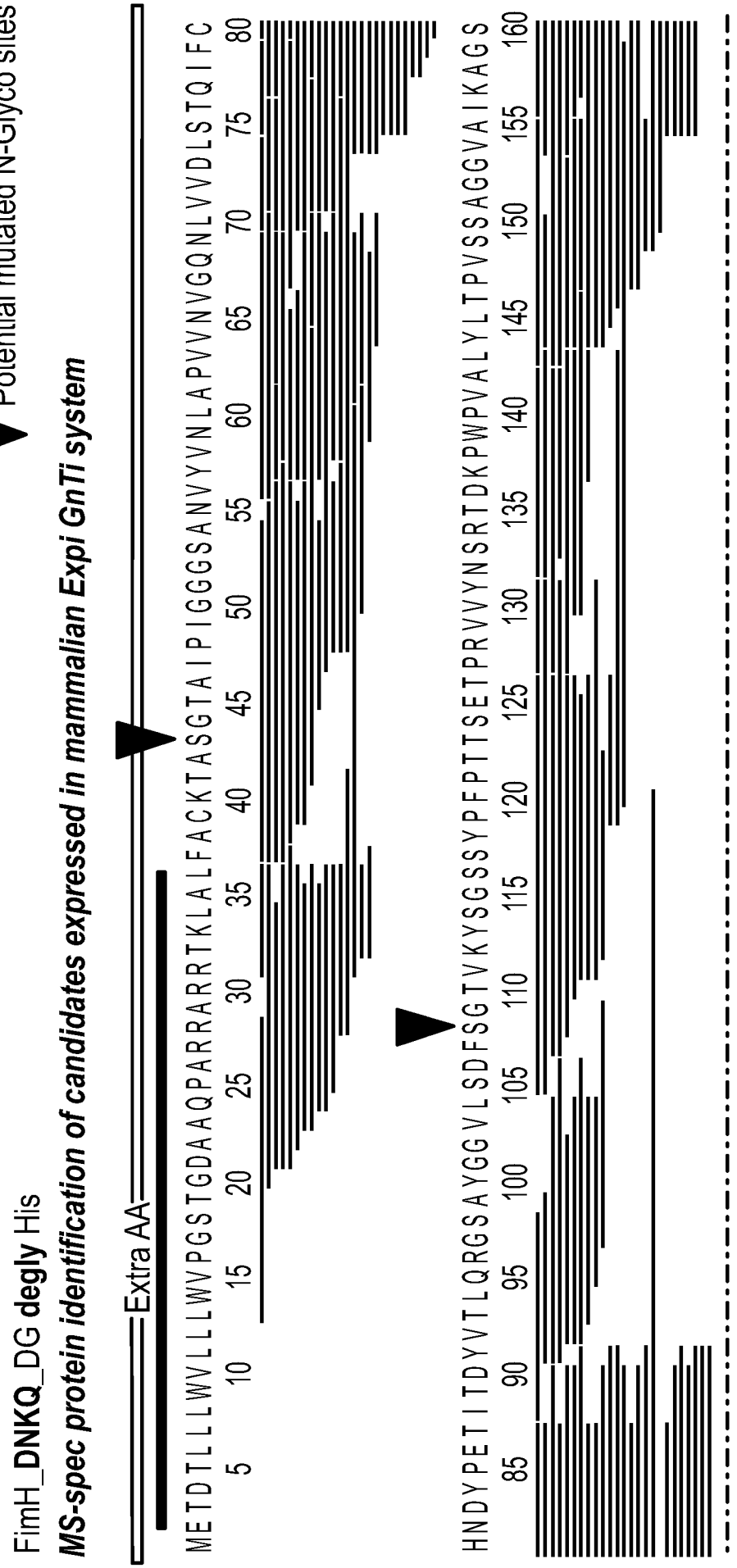

L I A V L I L R Q T N N Y N S D D F Q F V W N I Y A N N D V V V P T G G C D V S A R D V T V T L P D Y P G S V P I P L T V Y C A K S Q N L G Y Y L S G T T A D A
165 170 175 180 185 190 195 200 205 210 215 220 225 230 235 240
G N S I F T N T A S F S P A Q G V G V Q L T R N G T I I P A N N T V S L G A V G T S A V S L G L T A N Y A R T G G Q V T A G N V Q S I I G V T F V Y Q D N K Q A
245 250 255 260 265 270 275 280 285 290 295 300 305 310 315 320
D V T I T V N G K V V A K G S G H H H H H H
325 330 335 340
Total 273 Peptides, 95.6% Coverage, 11.18 Redundancy

FIG. 6(contd)

FIG. 6(contd)


Potential mutated N-Glyco sites
FimH_DNKQ_DG degly His
MS-spec protein identification of candidates expressed in mammalian Expi GnTi system
Extra AA
METDTLLLLWVLLLWVPGSTGDAAQPARRARRTKLALFACKTASGTAIPIGGGSANVYVNLAPVVNVGQNLVVDLSTQIFC
HNDYPETITDYVTLQRGSAYGGVLSDFSGTVKYSGSSYPFPTTSETPRVVYNSRTDKPWPVALYLTPVSSAGGVAIKAGS

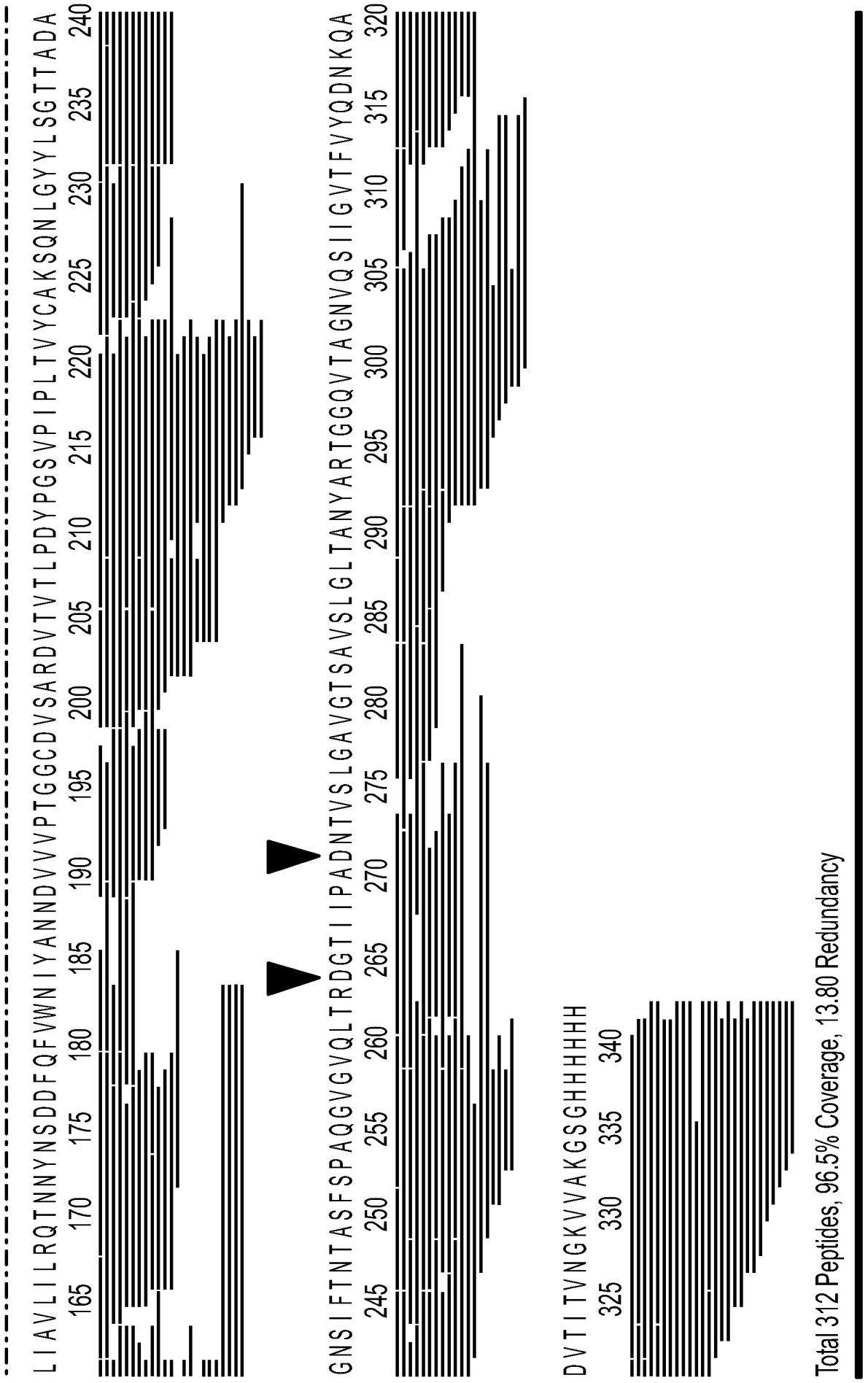
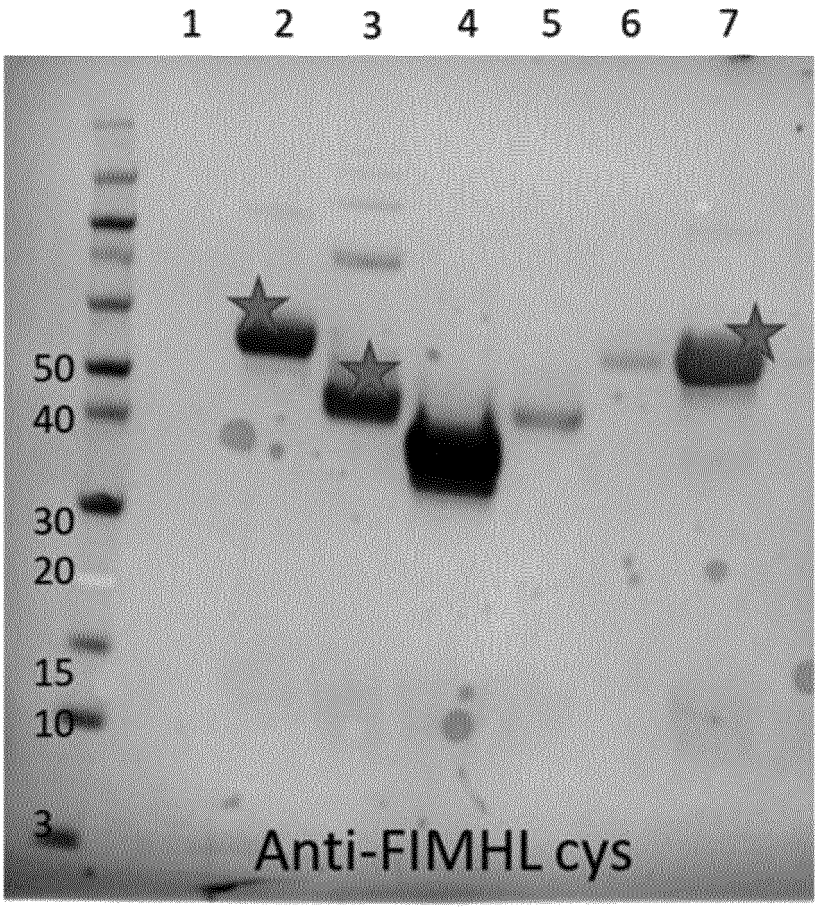

L I A V L I L R Q T N N Y N S D D F Q F V W N I Y A N N D V V V P T G G C D V S A R D V T V T L P D Y P G S V P I P L T V Y C A K S Q N L G Y Y L S G T T A D A
165 170 175 180 185 190 195 200 205 210 215 220 225 230 235 240
G N S I F T N T A S F S P A Q G V G V Q L T R D G T I I P A D N T V S L G A V G T S A V S L G L T A N Y A R T G G Q V T A G N V Q S I I G V T F V Y Q D N K Q A
245 250 255 260 265 270 275 280 285 290 295 300 305 310 315 320
D V T I T V N G K V V A K G S G H H H H H H
325 330 335 340
Total 312 Peptides, 96.5% Coverage, 13.80 Redundancy

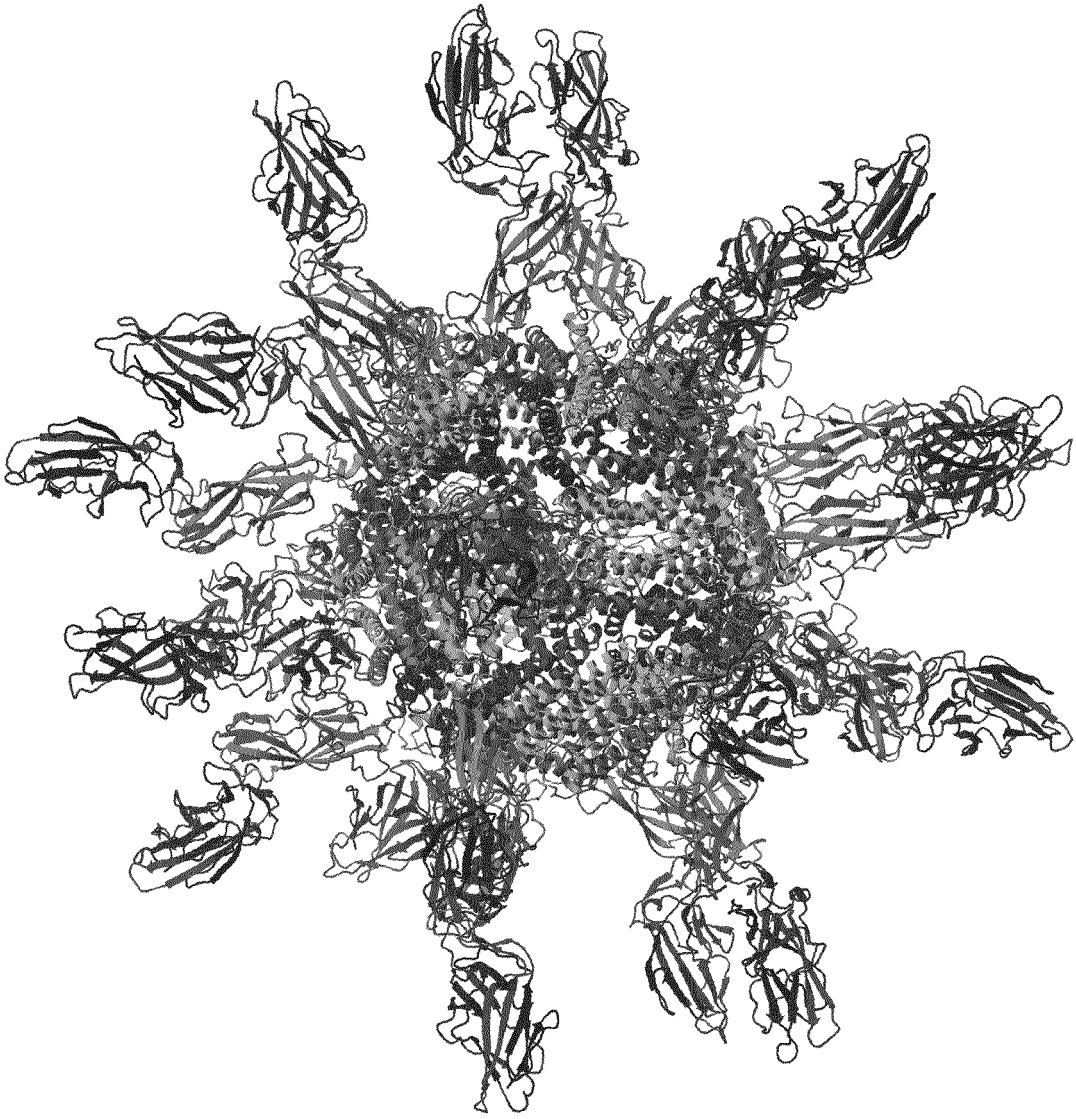
FIG. 6(contd)

FIG. 7



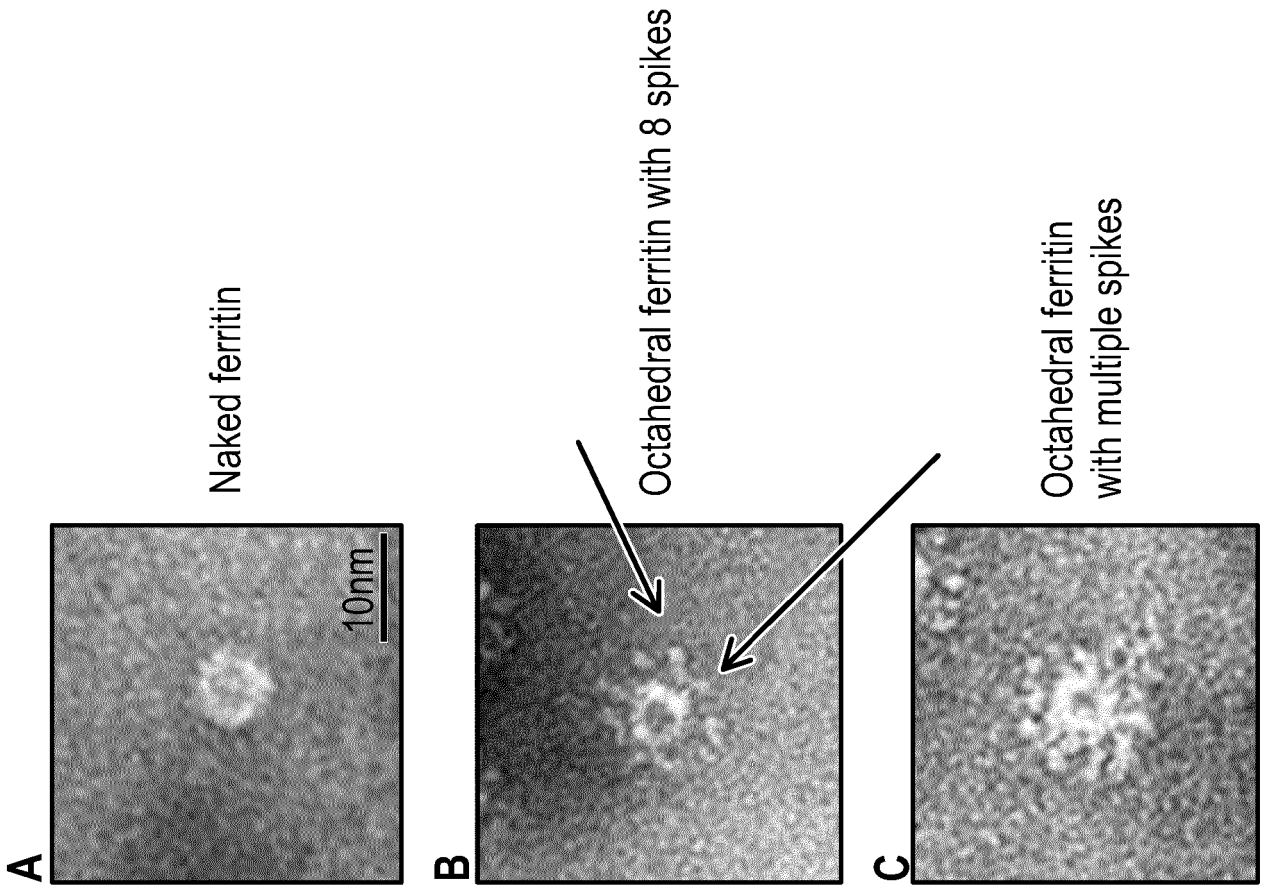
1- Not trasfected GNTI cells
2- Short leader FIMH_DG_PGDGN_Ferritin (j96) (1042SI)
3- Short leader FIMH_DG_PGDGN_IMX313(j96) (1043SI)
4- Short leader FIMH_DG_PGDGN (j96) (988SI)
5- Extra AA FIMH_DG_PGDGN_IMX313 (j96) (993SI)
6- EXTRA AAFIMH_DG_PGDGN_Ferritin (536) (995SI)
7- Short leader FIMH_DG_PGDGN_Ferritin (536) (1142SI)

FIG. 8A



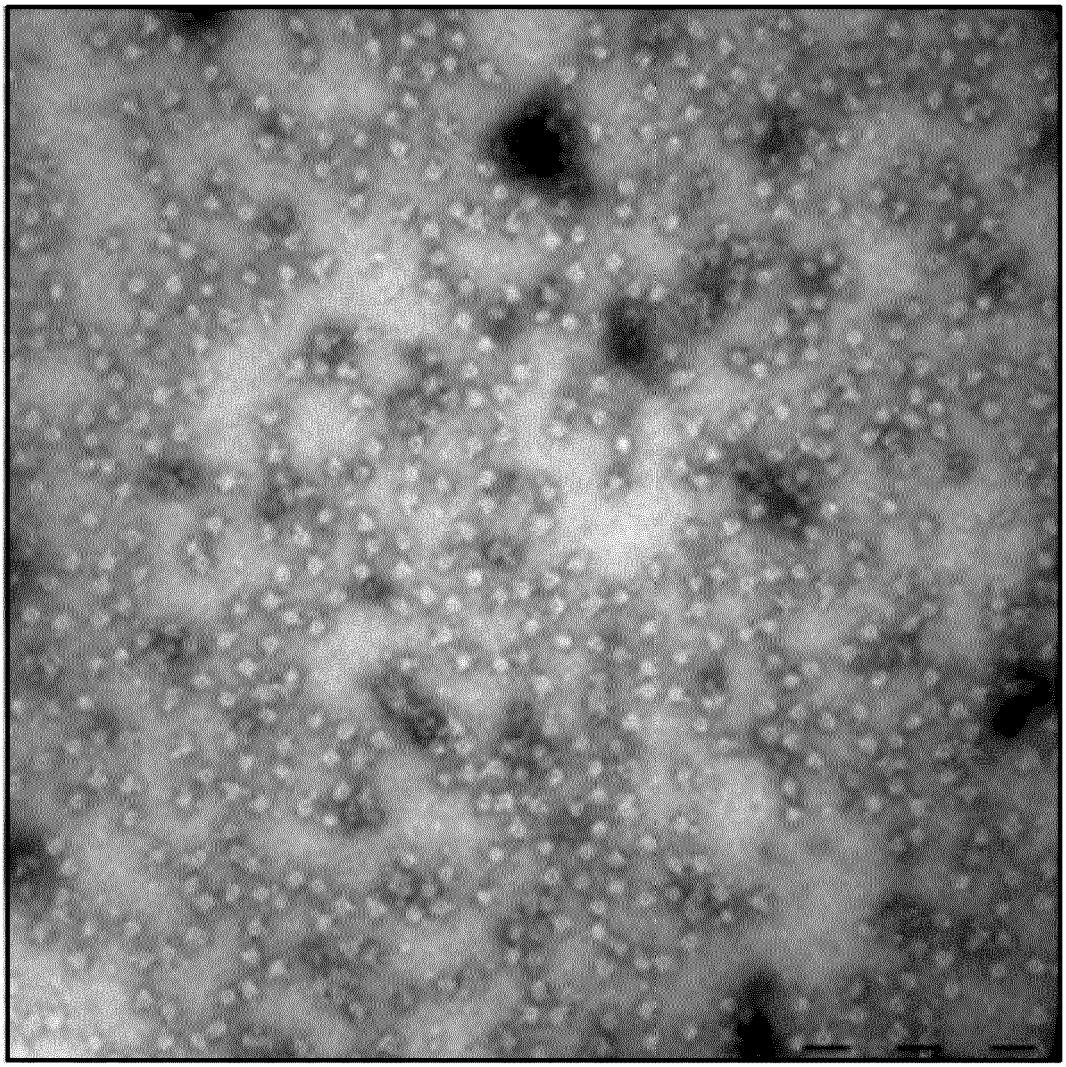
995SI

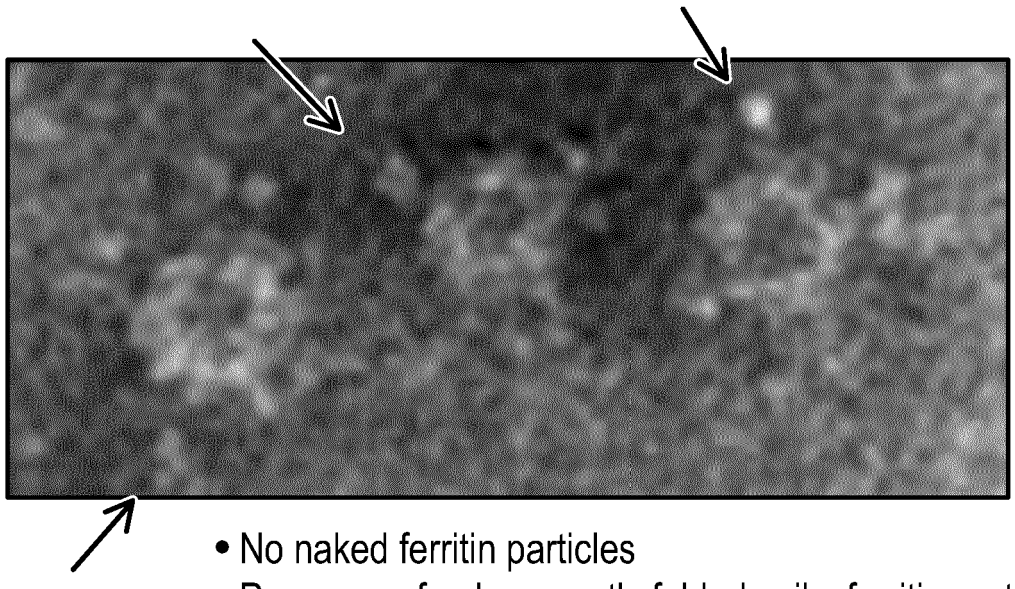
FIG. 8B



- Naked ferritin particles have a diameter of 13nm
- Spiky ferritin have a diameter of 30-32nm

FIG. 8C
1142SI





- No naked ferritin particles
- Presence of only correctly folded spiky ferritin particles
- Diameter of 30-32nm

FIG. 8D

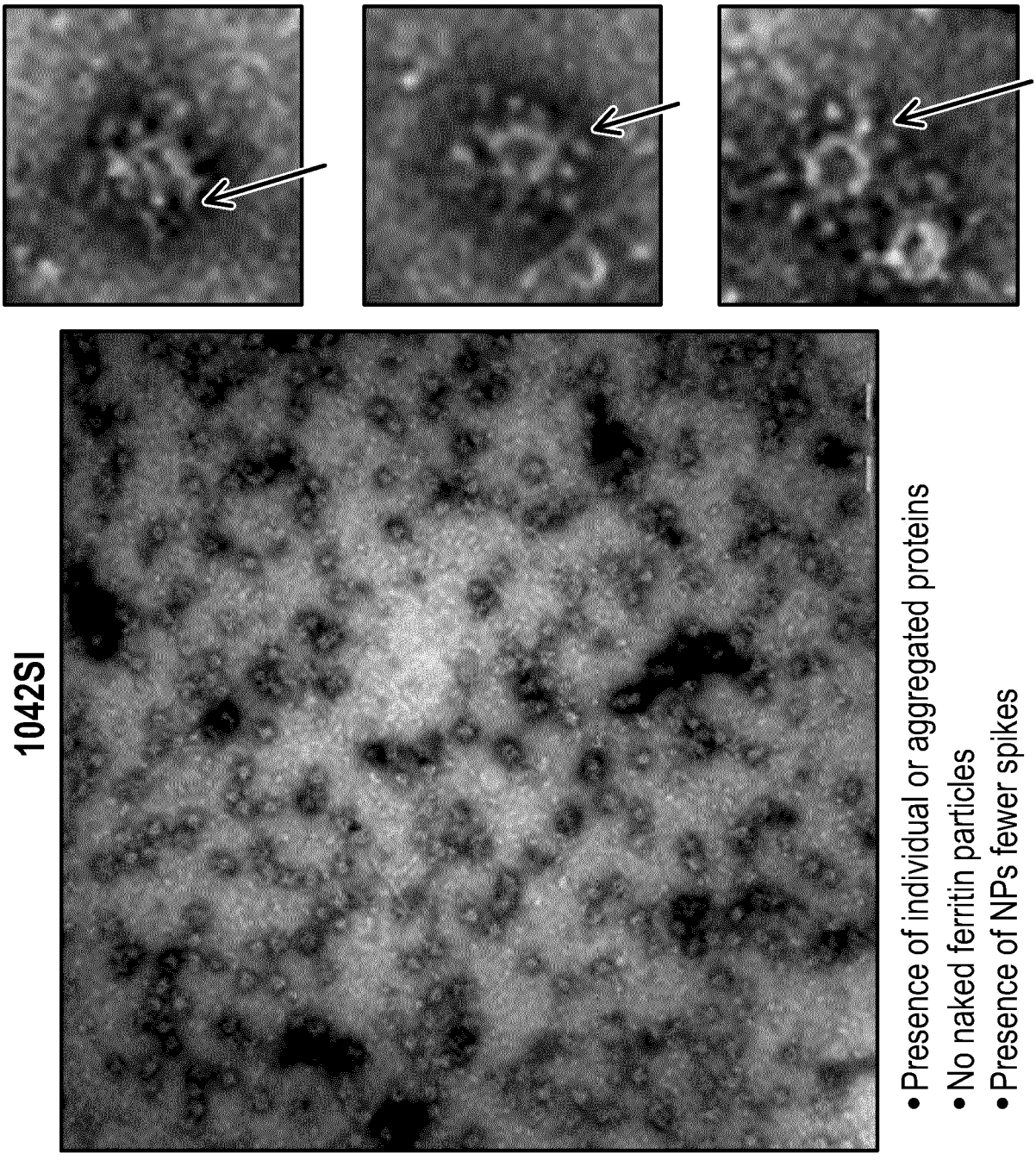

1042SI
Presence of non correctly folded particles
Presence of correctly folded NPs presenting multiple spikes
Presence of correctly folded spiky NPs presenting eight spikes
• Presence of individual or aggregated proteins
• No naked ferritin particles
• Presence of NPs fewer spikes

FIG. 9

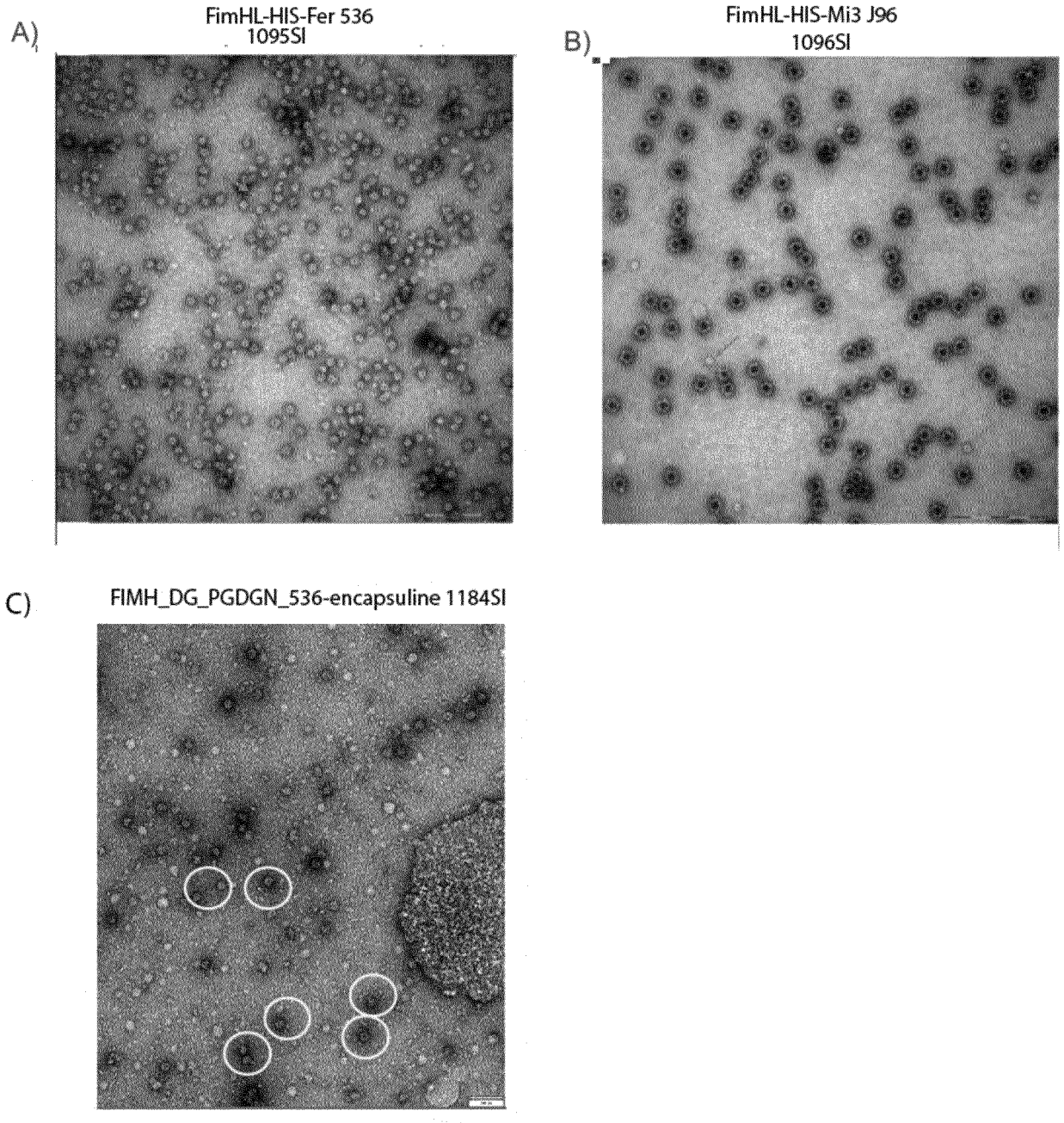

A)
FimHL-HIS-Fer 536
1095SI
B)
FimHL-HIS-Mi3 J96
1096SI
C)
FIMH_DG_PGDGN_536-encapsuline 1184SI

FIG. 10

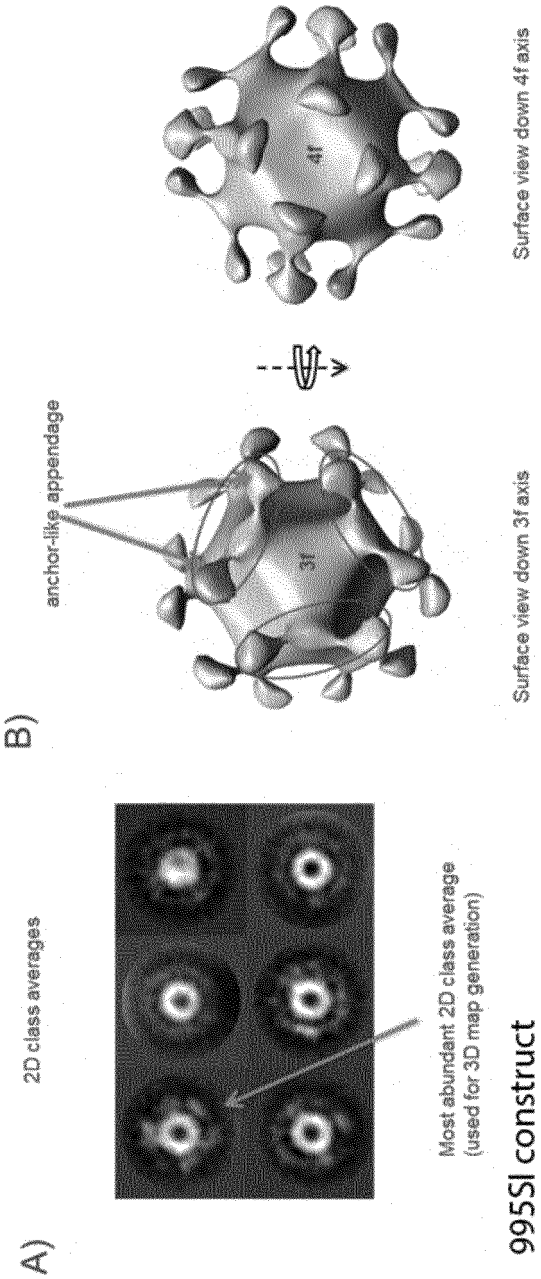

A)
2D class averages
Most abundant 2D class average
(used for 3D map generation)
995SI construct
B)
anchor-like appendage
3f
Surface view down 3f axis
4f
Surface view down 4f axis

FIG. 11

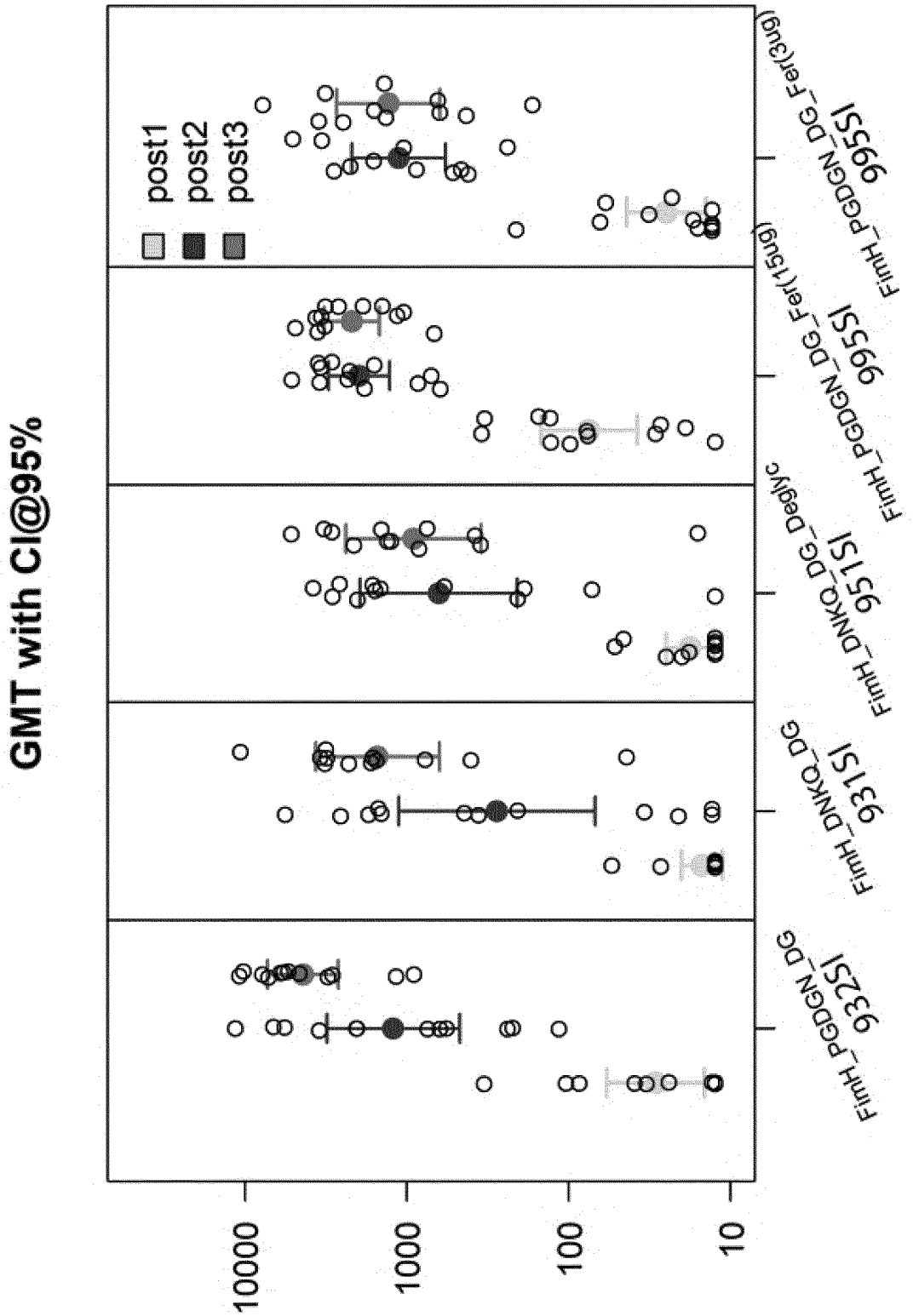

GMT with CI@95%
post1
post2
post3
10000
1000
100
10
FimH_pDGDN_DG 9253
FimH_DNKQ_DG 9313
FimH_DNKQ_DG_DegVc 9515
FimH_pDGDN_DG_Fer(1sug) 9955
FimH_pDGDN_DG_Fer(3sug) 9959

FIG. 12

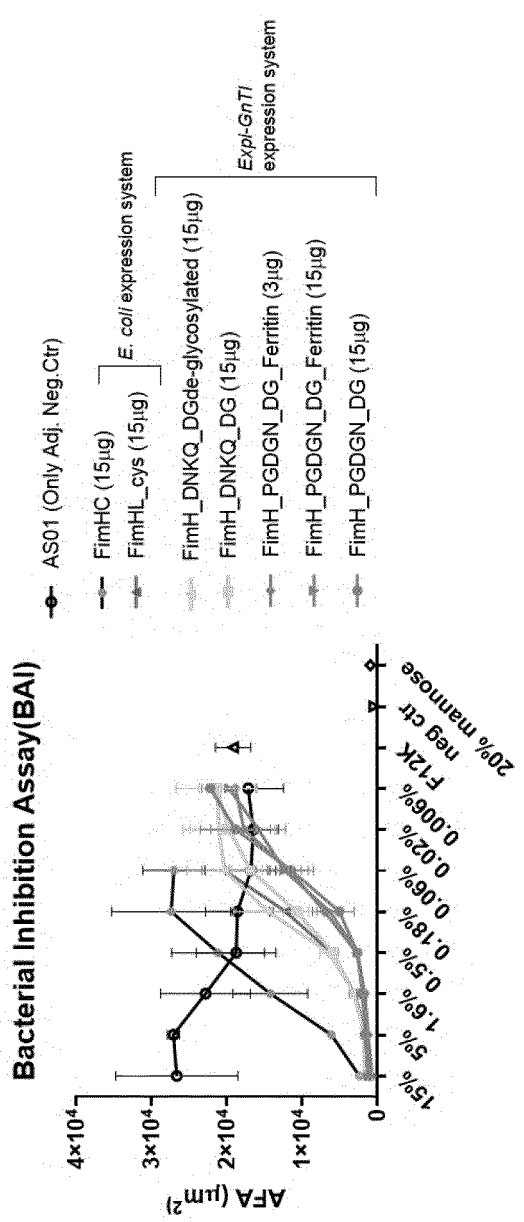

Bacterial Inhibition Assay(BAI)
AS01 (Only Adj. Neg.Ctr)
FimHC (15µg)
FimHL_cys (15µg)
E. coli expression system
FimH_DNKQ_DGde-glycosylated (15µg)
FimH_DNKQ_DG (15µg)
FimH_PGDGN_DG_Ferritin (3µg)
FimH_PGDGN_DG_Ferritin (15µg)
FimH_PGDGN_DG (15µg)
Expi-GnTI expression system
AFA (µm²)
4×10⁴
3×10⁴
2×10⁴
1×10⁴
0
15%
5%
1.6%
0.5%
0.18%
0.06%
0.02%
0.006%
F12K
neg ctr
20% mannose

FIG. 13

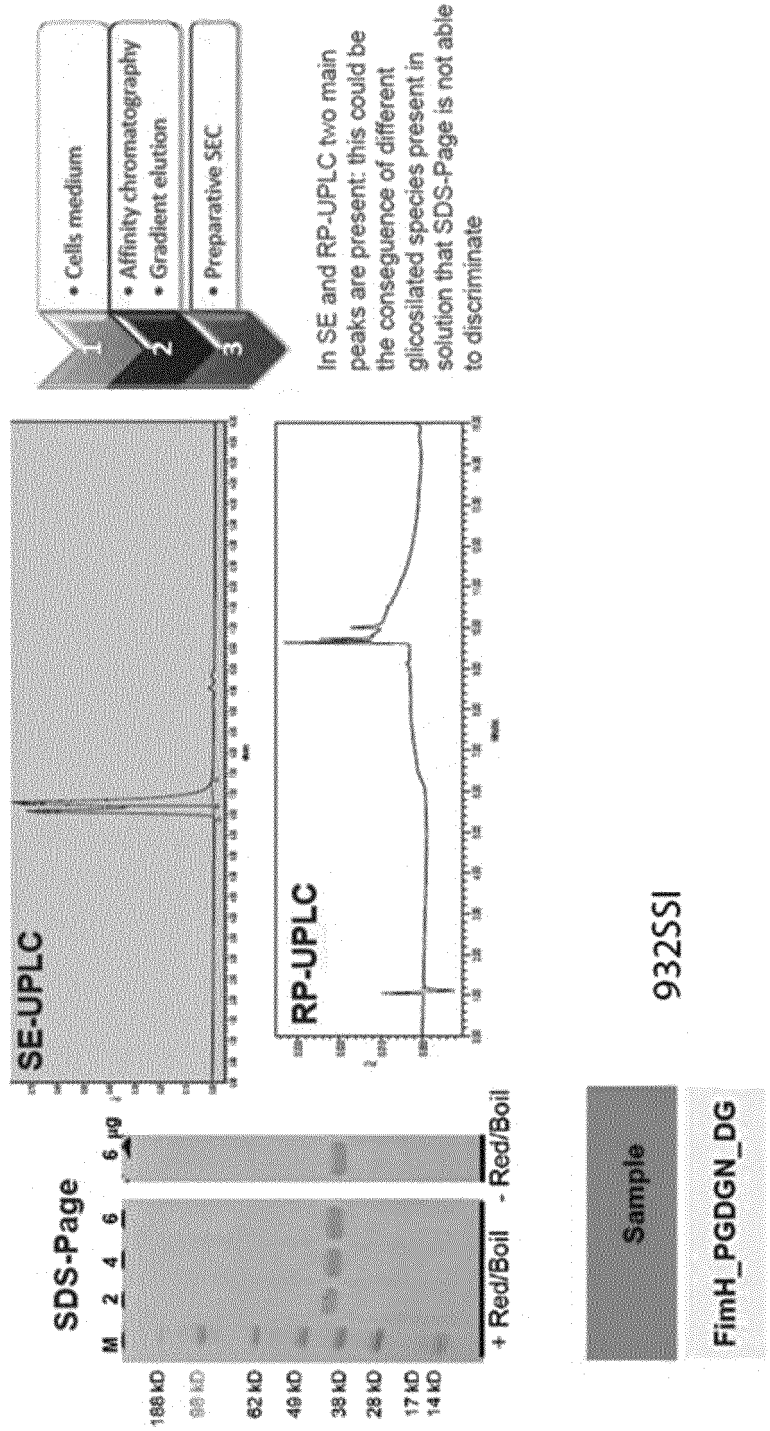

• Cells medium
• Affinity chromatography
• Gradient elution
• Preparative SEC
1
2
3
In SE and RP-UPLC two main peaks are present: this could be the consequence of different glicosilated species present in solution that SDS-Page is not able to discriminate
SE-UPLC
RP-UPLC
SDS-Page
M 2 4 6
6 µg
+ Red/Boil
- Red/Boil
188 kD
98 kD
62 kD
49 kD
38 kD
28 kD
17 kD
14 kD
Sample
932SSI
FimH_PGDGN_DG

FIG. 14

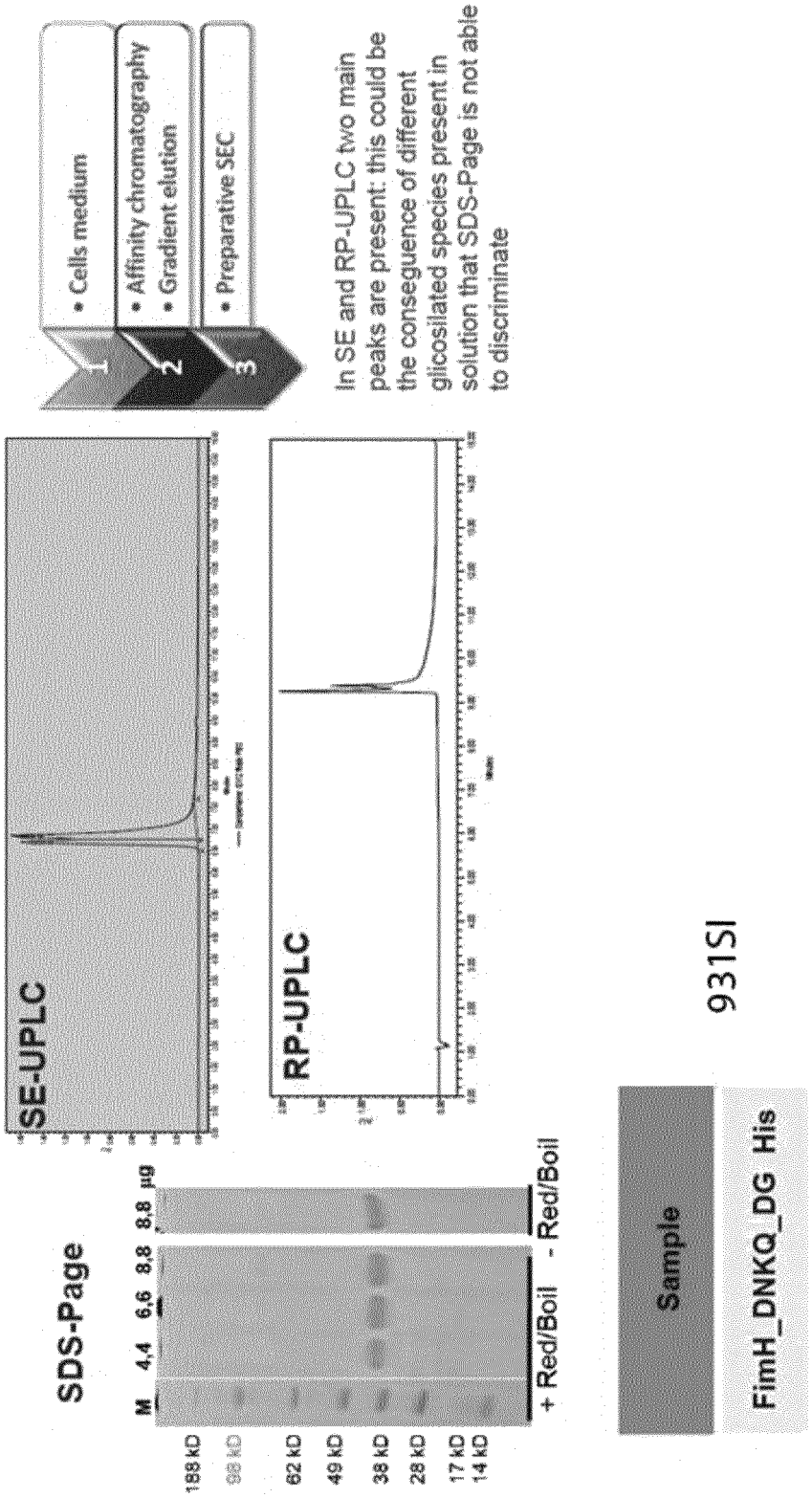

• Cells medium
• Affinity chromatography
• Gradient elution
• Preparative SEC
1
2
3
In SE and RP-UPLC two main peaks are present: this could be the consequence of different glicosilated species present in solution that SDS-Page is not able to discriminate
SE-UPLC
RP-UPLC
SDS-Page
M 4,4 6,6 8,8 8,8 µg
+ Red/Boil - Red/Boil
188 kD
98 kD
62 kD
49 kD
38 kD
28 kD
17 kD
14 kD
Sample
931SI
FimH_DNKQ_DG His

FIG. 15

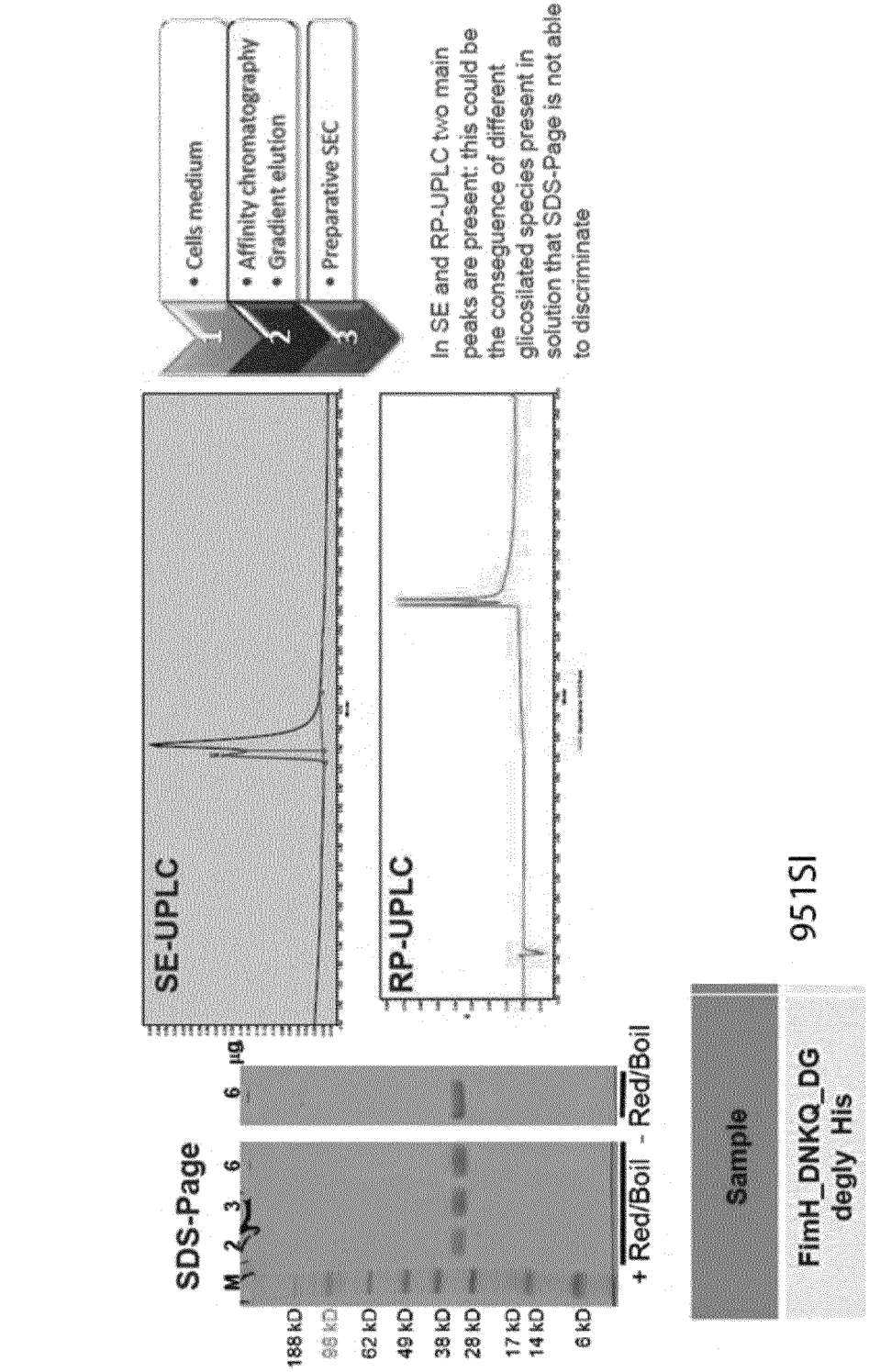

Sample
FimH_DNKQ_DG
degly His
951SI
SDS-Page
M 1 2 3 6 µg
+ Red/Boil - Red/Boil
188 kD
98 kD
62 kD
49 kD
38 kD
28 kD
17 kD
14 kD
6 kD
SE-UPLC
RP-UPLC
In SE and RP-UPLC two main peaks are present: this could be the consequence of different glicosilated species present in solution that SDS-Page is not able to discriminate
• Cells medium
• Affinity chromatography
• Gradient elution
• Preparative SEC
1
2
3

FIG. 16

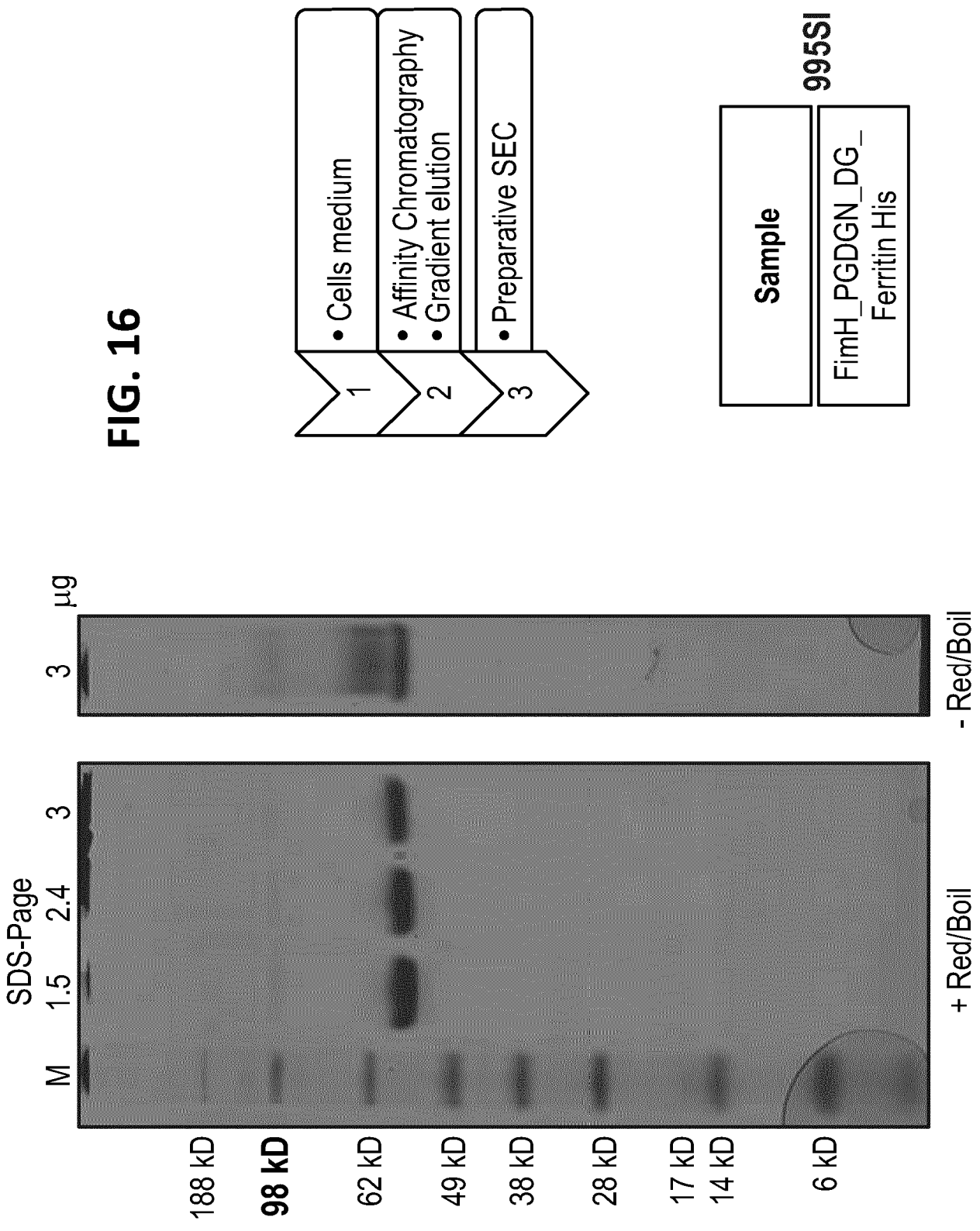

• Cells medium
• Affinity Chromatography
• Gradient elution
• Preparative SEC
1
2
3
Sample
995SI
FimH_PGDGN_DG_
Ferritin His
SDS-Page
M
1.5
2.4
3
3 µg
+ Red/Boil
- Red/Boil
188 kD
98 kD
62 kD
49 kD
38 kD
28 kD
17 kD
14 kD
6 kD

FIG. 16(contd)

SE-HPLC

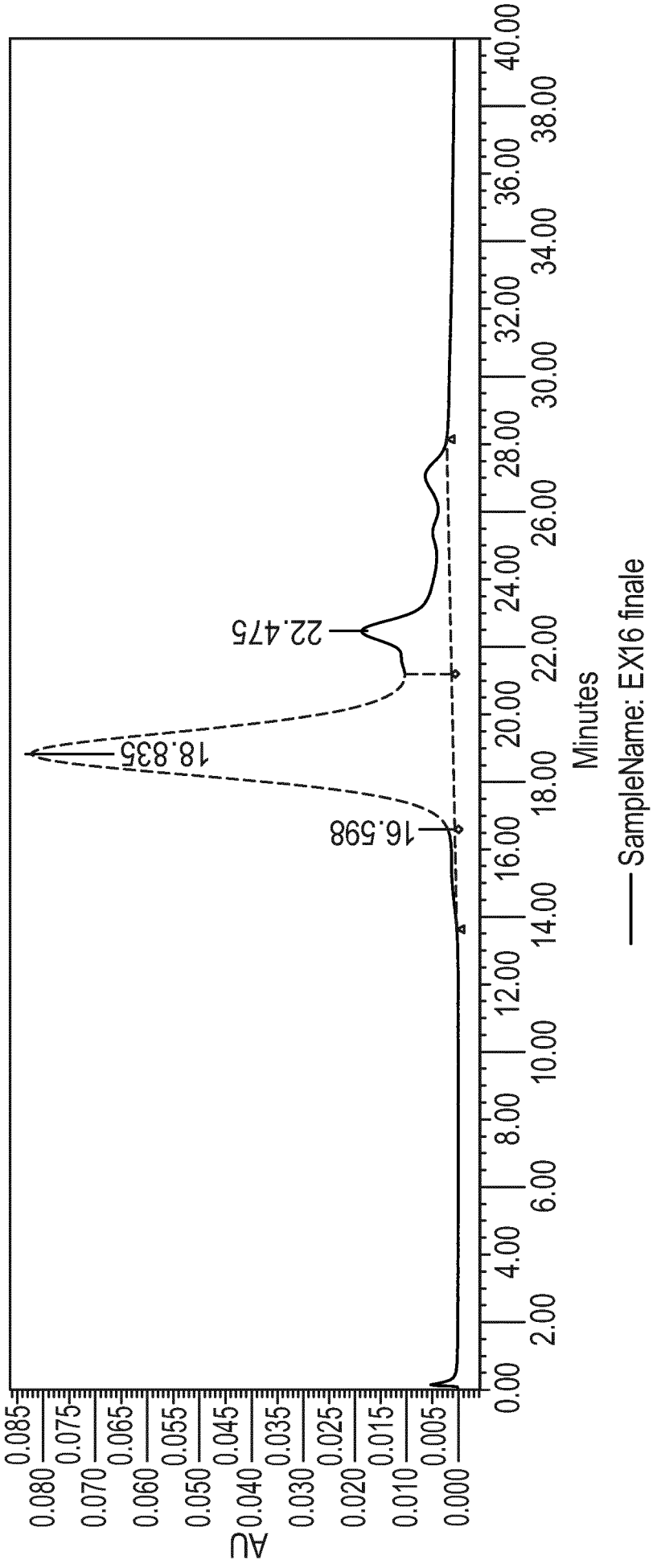


FIG. 16(contd)

RP-UPLC

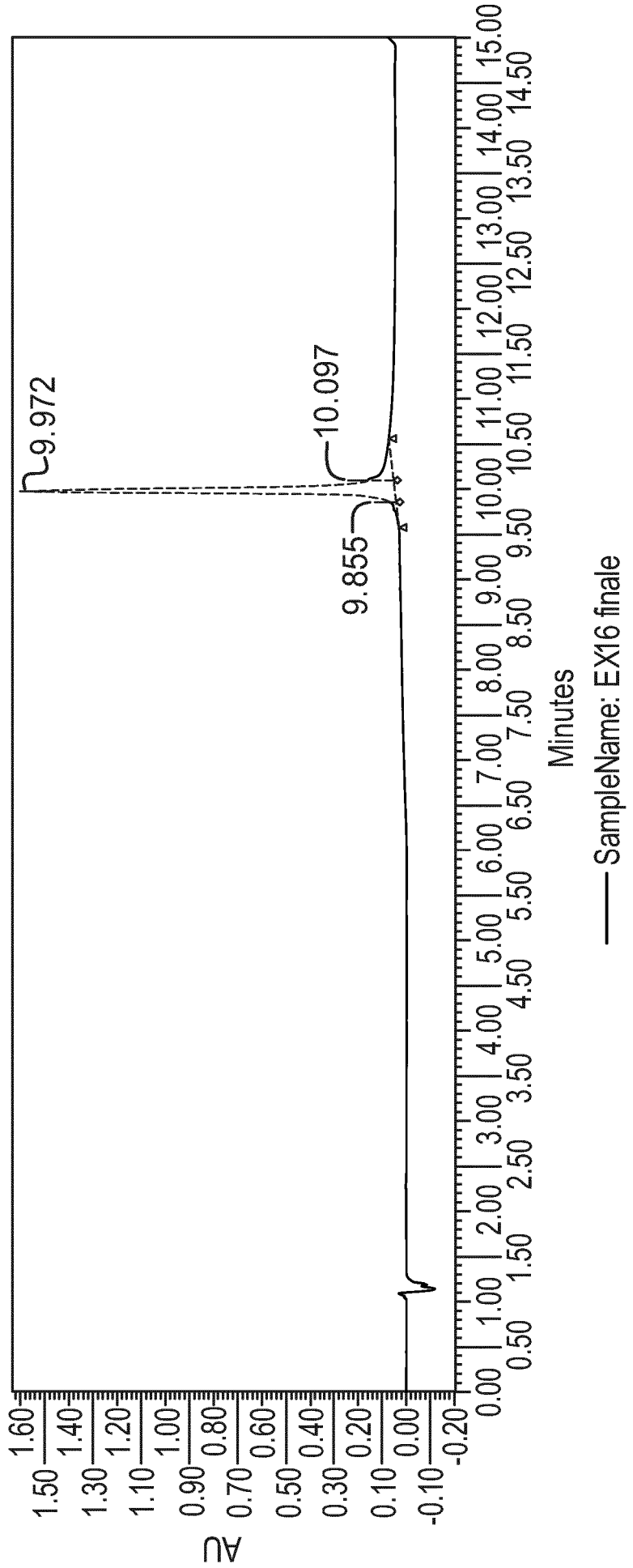
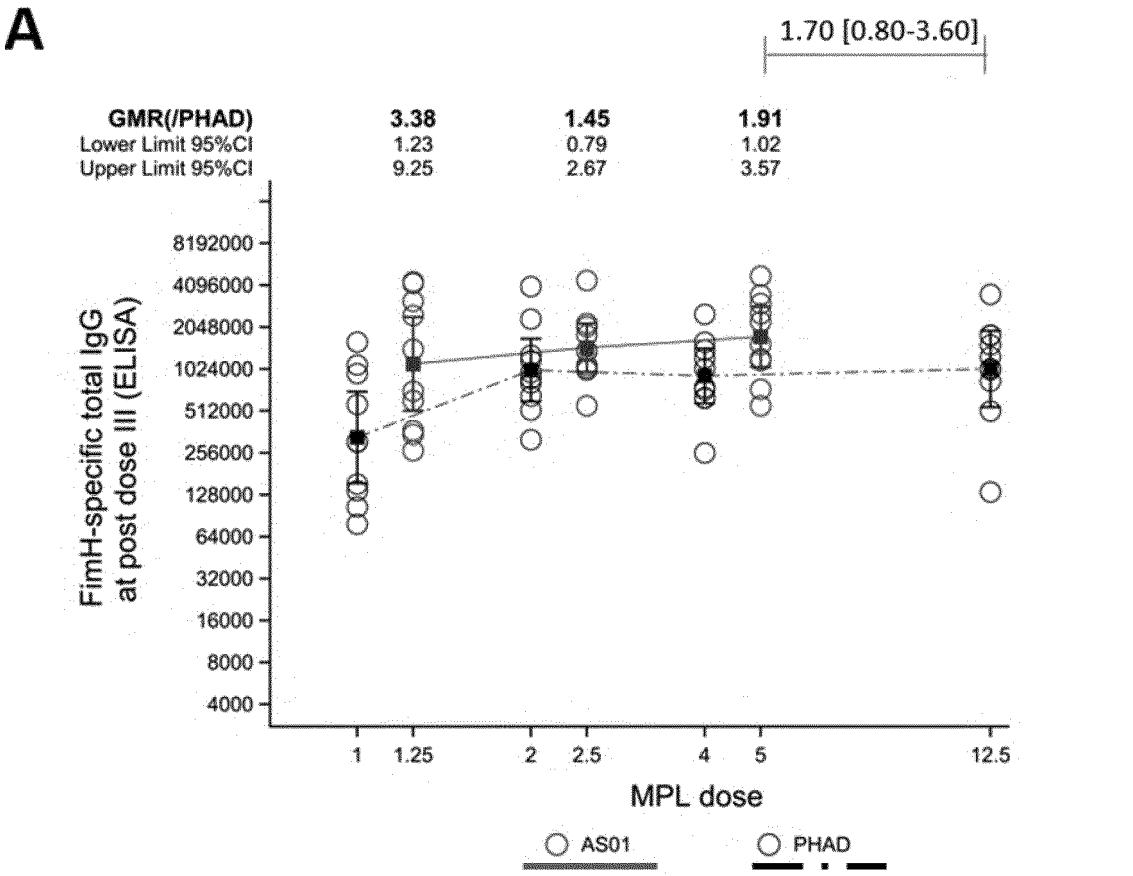


FIG. 17


A
1.70 [0.80-3.60]
GMR(/PHAD) 3.38 1.45 1.91
Lower Limit 95%CI 1.23 0.79 1.02
Upper Limit 95%CI 9.25 2.67 3.57
FimH-specific total IgG at post dose III (ELISA)
8192000
4096000
2048000
1024000
512000
256000
128000
64000
32000
16000
8000
4000
1 1.25 2 2.5 4 5 12.5
MPL dose
AS01
PHAD

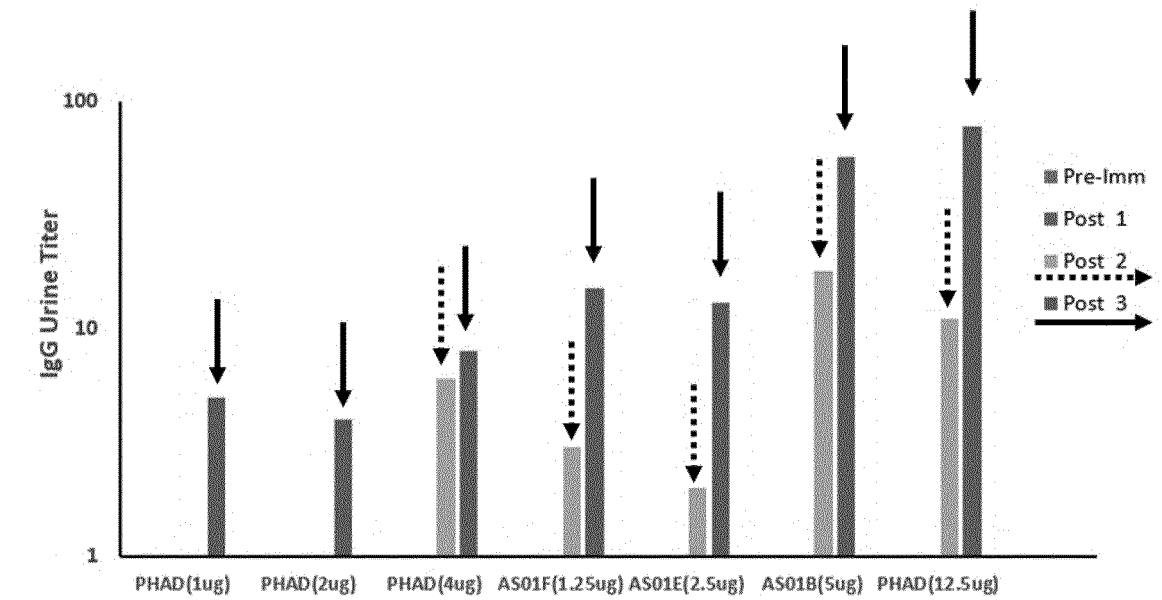

B
IgG Urine Titer
100
10
1
Pre-Imm
Post 1
Post 2
Post 3
PHAD(1ug)
PHAD(2ug)
PHAD(4ug)
AS01F(1.25ug)
AS01E(2.5ug)
AS01B(5ug)
PHAD(12.5ug)

FIG. 18 A

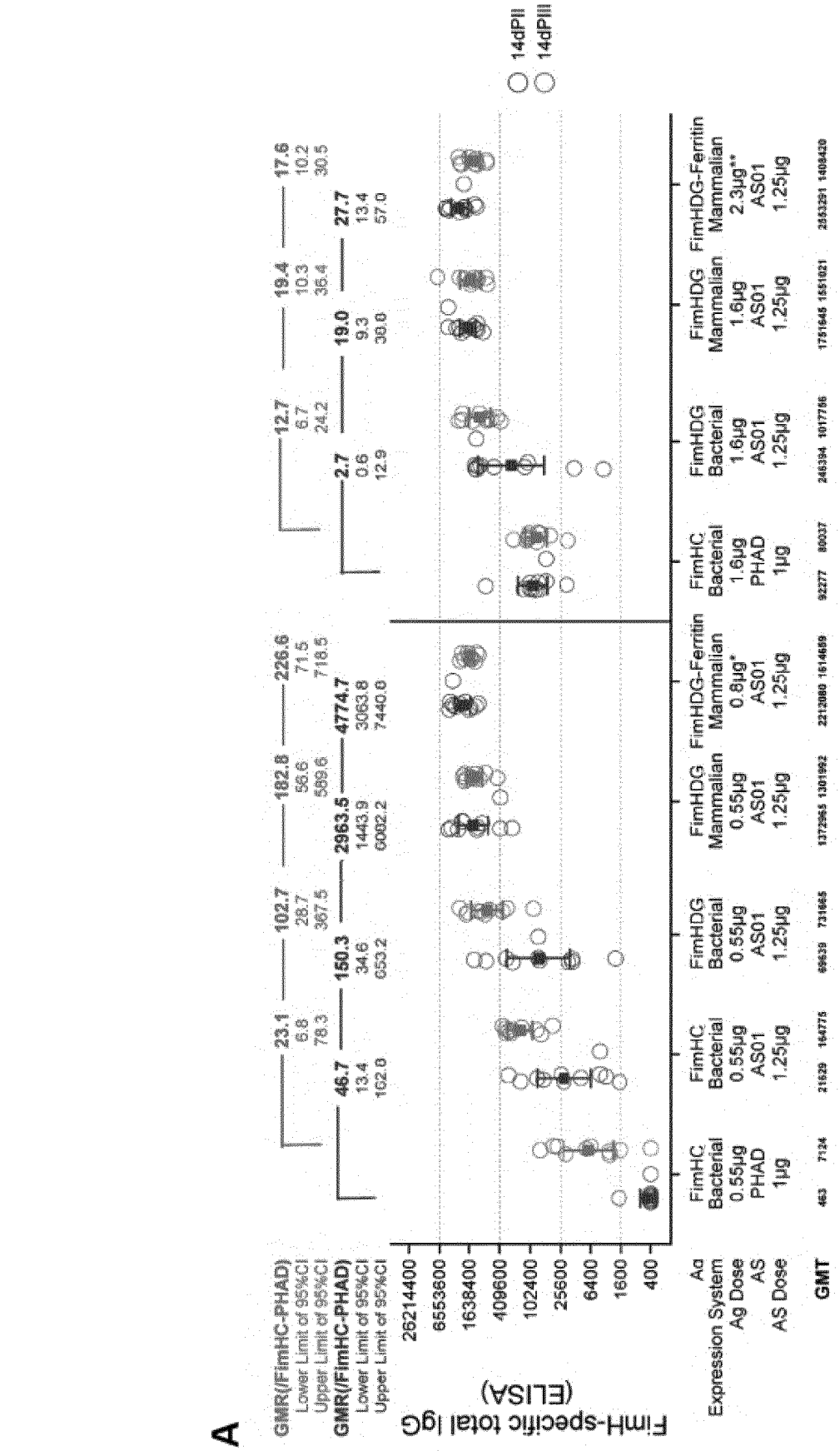

A
FimH-specific total IgG (ELISA)
14dPII
14dPIII
GMR(/FimHC-PHAD)
Lower Limit of 95%CI
Upper Limit of 95%CI
GMR(/FimHC-PHAD)
Lower Limit of 95%CI
Upper Limit of 95%CI
Ag
Expression System
Ag Dose
AS
AS Dose
GMT

FIG. 18B

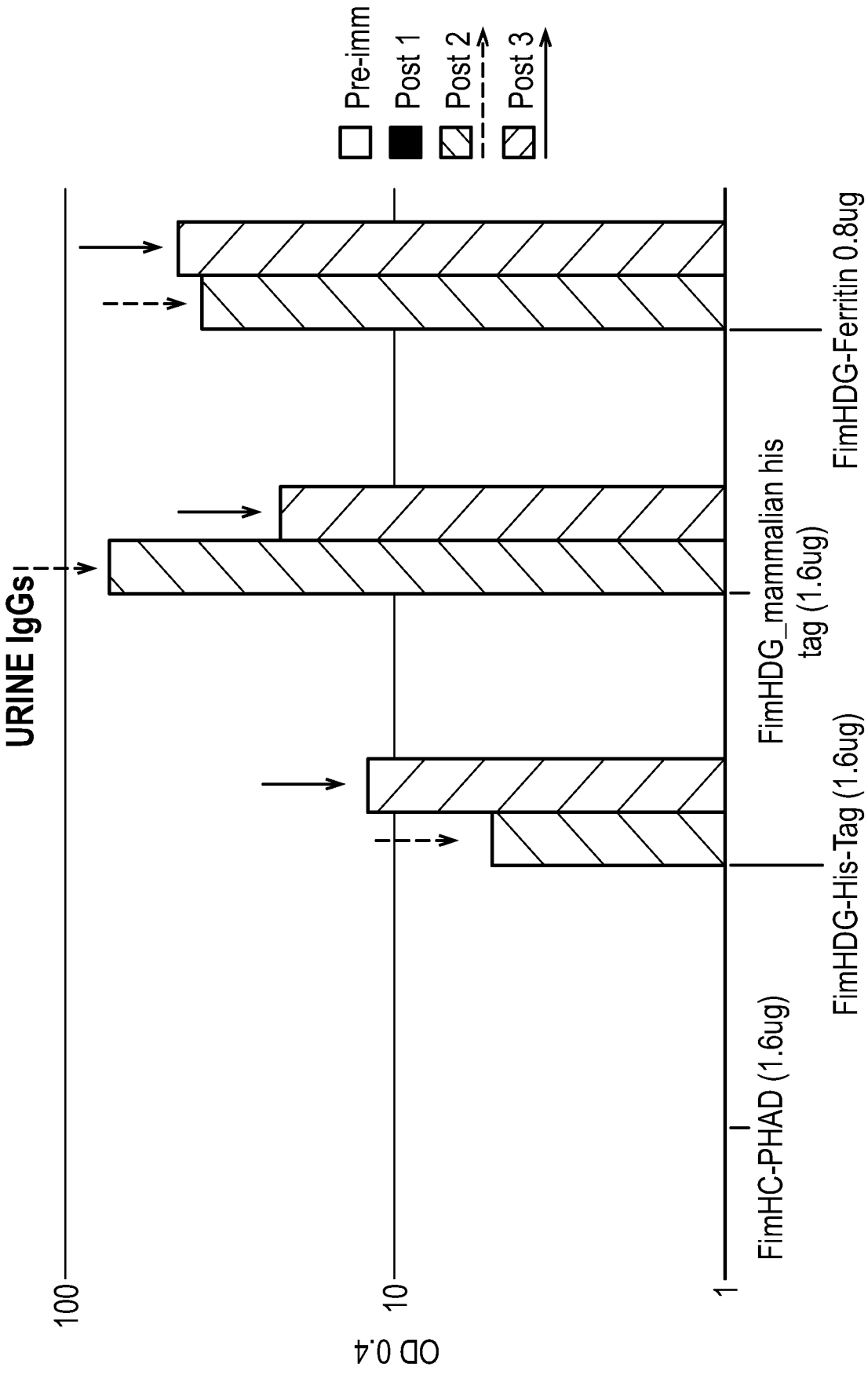

Pre-imm
Post 1
Post 2
Post 3
URINE IgGs
FimHDG-Ferritin 0.8ug
FimHDG_mammalian his tag (1.6ug)
FimHDG-His-Tag (1.6ug)
FimHC-PHAD (1.6ug)
100
10
1
OD 0.4

FIG. 19

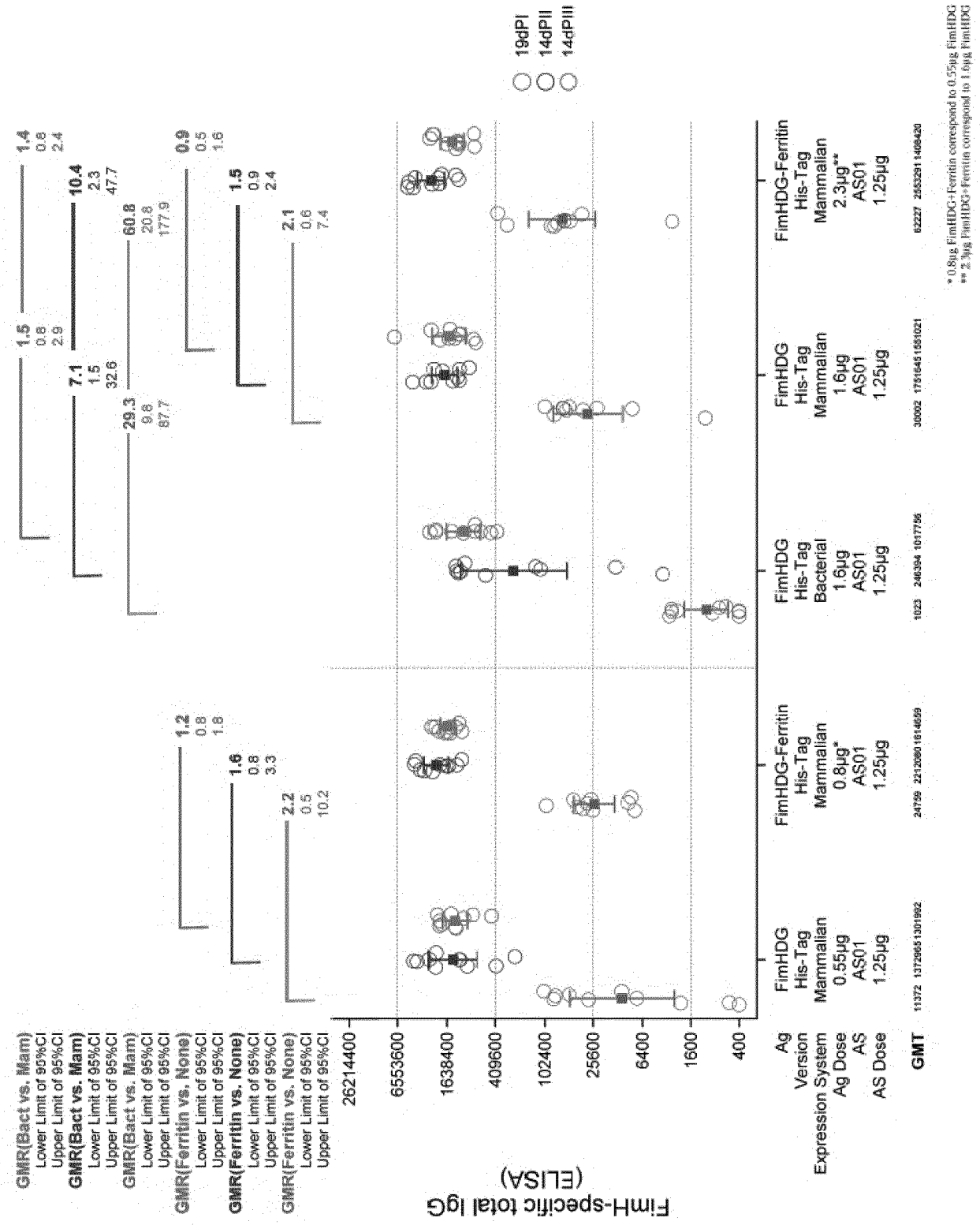

FimH-specific total IgG (ELISA)
19dPI
14dPII
14dPIII
GMR(Bact vs. Mam)
Lower Limit of 95%CI
Upper Limit of 95%CI
GMR(Ferritin vs. None)
Ag
Version
Expression System
Ag Dose
AS
AS Dose
GMT
FimHDG His-Tag Mammalian 0.55µg AS01 1.25µg
FimHDG-Ferritin His-Tag Mammalian 0.8µg* AS01 1.25µg
FimHDG His-Tag Bacterial 1.6µg AS01 1.25µg
FimHDG His-Tag Mammalian 1.6µg AS01 1.25µg
FimHDG-Ferritin His-Tag Mammalian 2.3µg** AS01 1.25µg
* 0.8µg FimHDG+Ferritin correspond to 0.55µg FimHDG
** 2.3µg FimHDG+Ferritin correspond to 1.6µg FimHDG

FIG. 20

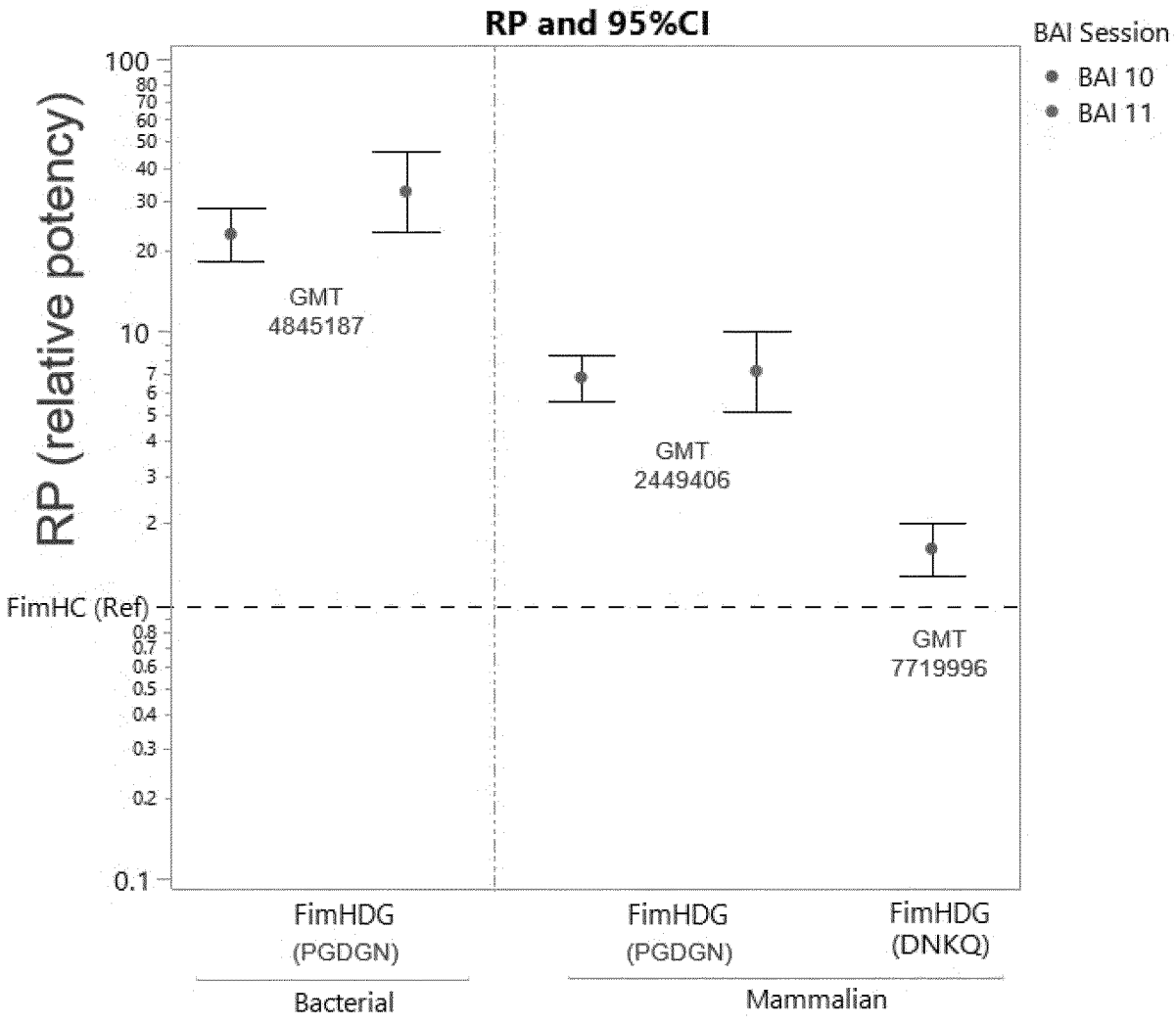

RP and 95%CI
BAI Session
BAI 10
BAI 11
RP (relative potency)
100
80
70
60
50
40
30
20
10
7
6
5
4
3
2
FimHC (Ref)
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
GMT
4845187
GMT
2449406
GMT
7719996
FimHDG
(PGDGN)
FimHDG
(PGDGN)
FimHDG
(DNKQ)
Bacterial
Mammalian

FIG. 21

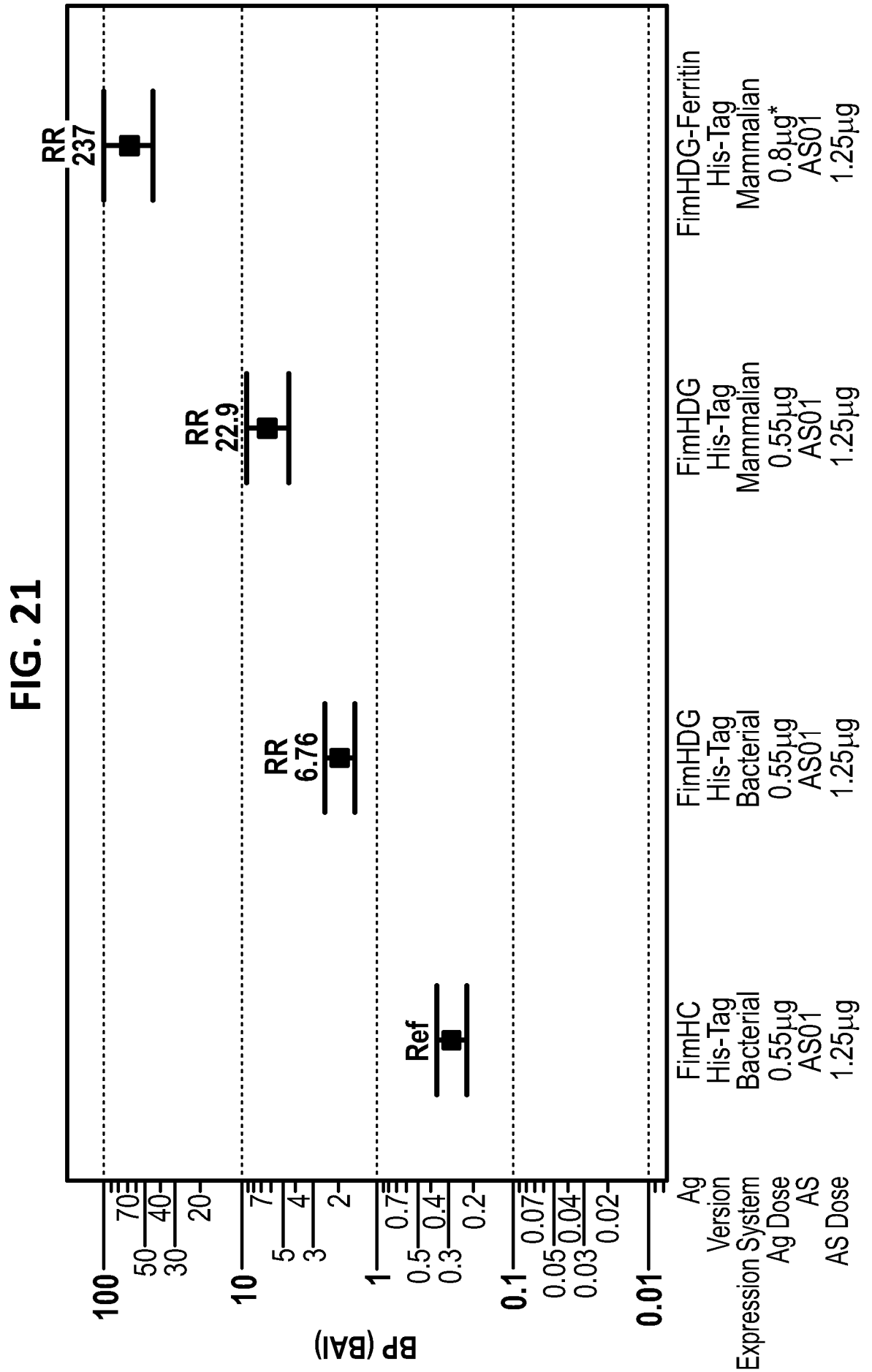
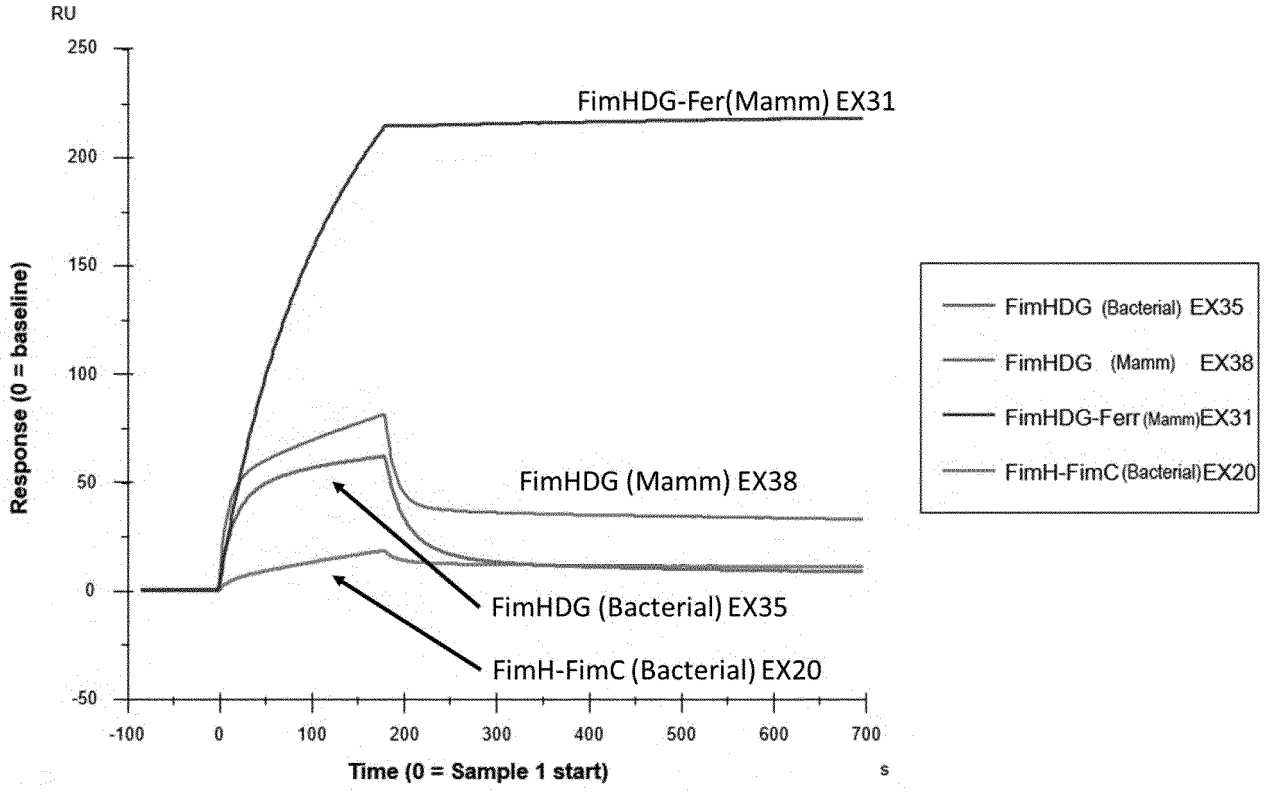

RR
237
RR
22.9
RR
6.76
Ref
BP (BAI)
100
70
50
40
30
20
10
7
5
4
3
2
1
0.7
0.5
0.4
0.3
0.2
0.1
0.07
0.05
0.04
0.03
0.02
0.01
Ag
Version
Expression System
Ag Dose
AS
AS Dose
FimHC
His-Tag
Bacterial
0.55µg
AS01
1.25µg
FimHDG
His-Tag
Bacterial
0.55µg
AS01
1.25µg
FimHDG
His-Tag
Mammalian
0.55µg
AS01
1.25µg
FimHDG-Ferritin
His-Tag
Mammalian
0.8µg*
AS01
1.25µg
RR= relative ratio of RP#/RP-FimHC-PHAD

FIG. 22



Sample	mAb	Binding level [RU]_1Exp	Binding level [RU]_2Exp
FimH-FimC (Bacterial) EX20	**VK_926**	21,4	18,0
FimH-DG (Bacterial) EX35	**VK_926**	63,1	61,8
FimH-DG His (Mamm.) EX38	**VK_926**	81,8	80,3
FimHDG-Ferr(Mamm.)EX31	**VK_926**	200,2	211,0

FIG. 23 A

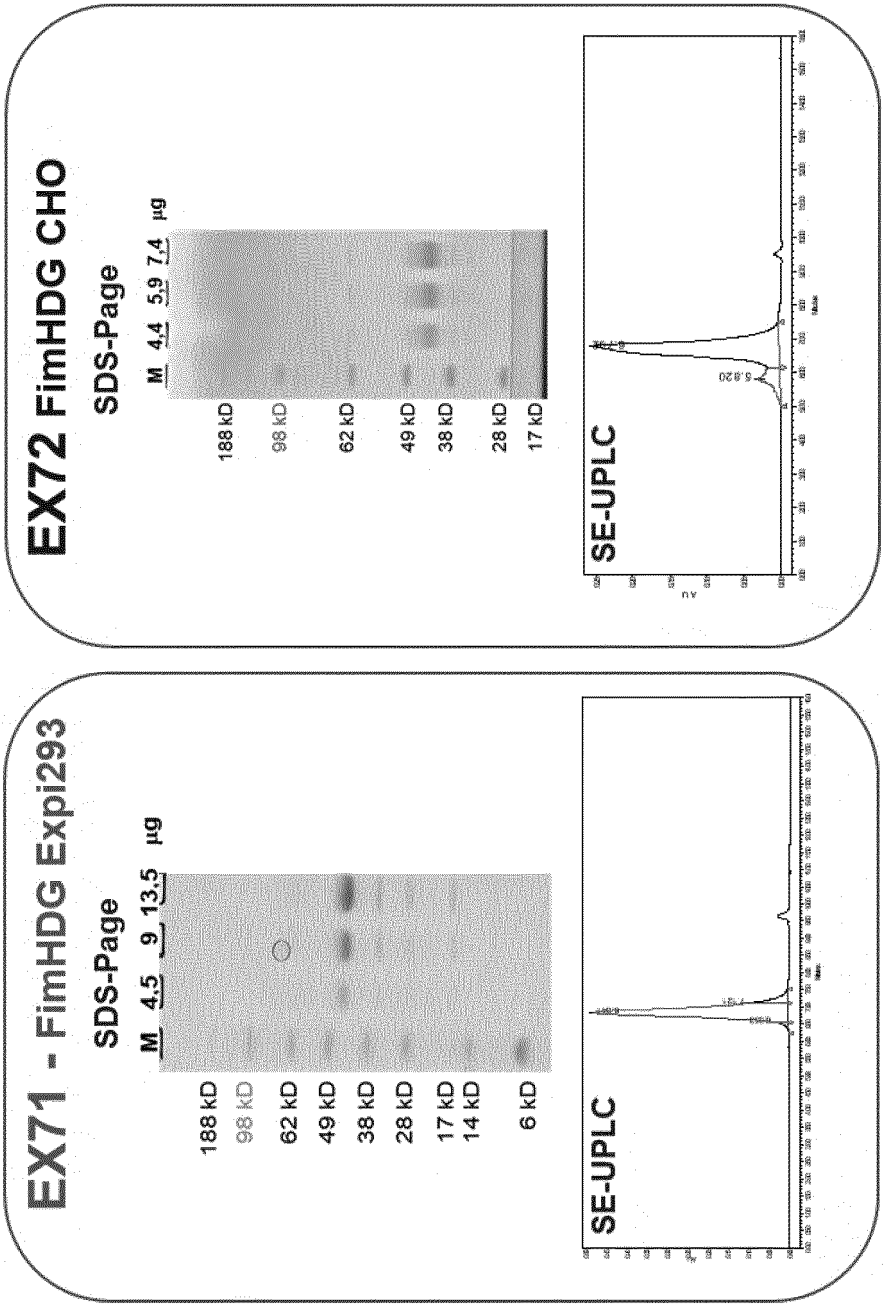
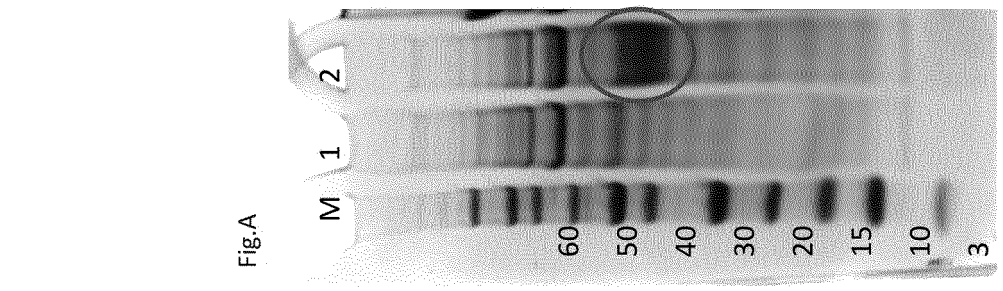


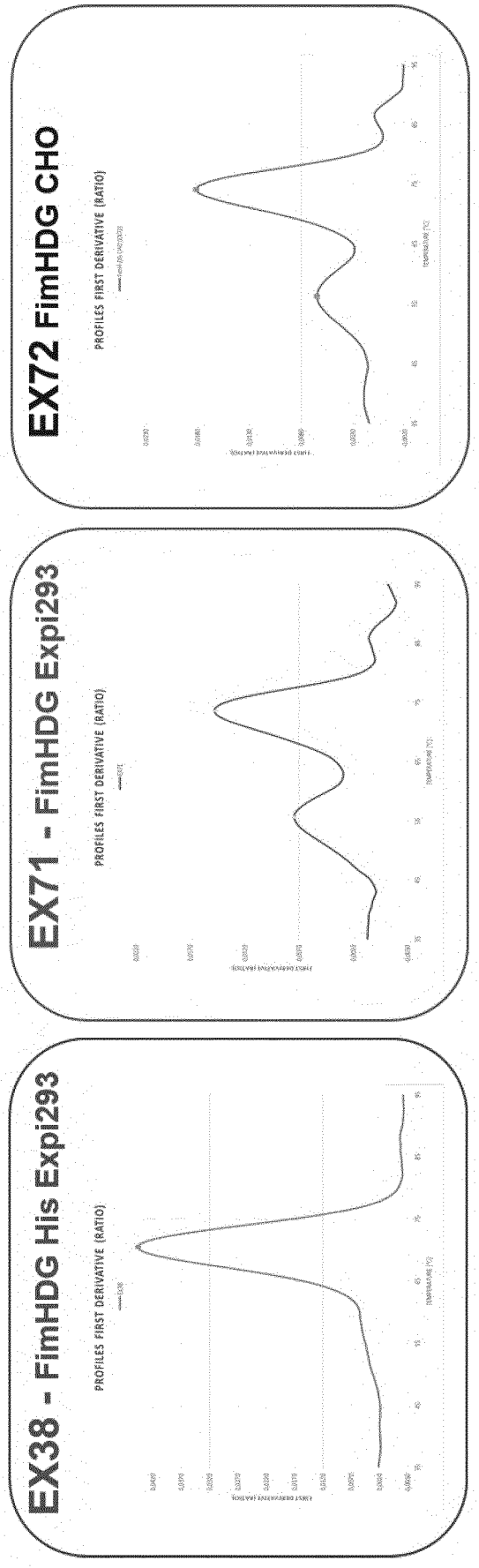
Fig.A



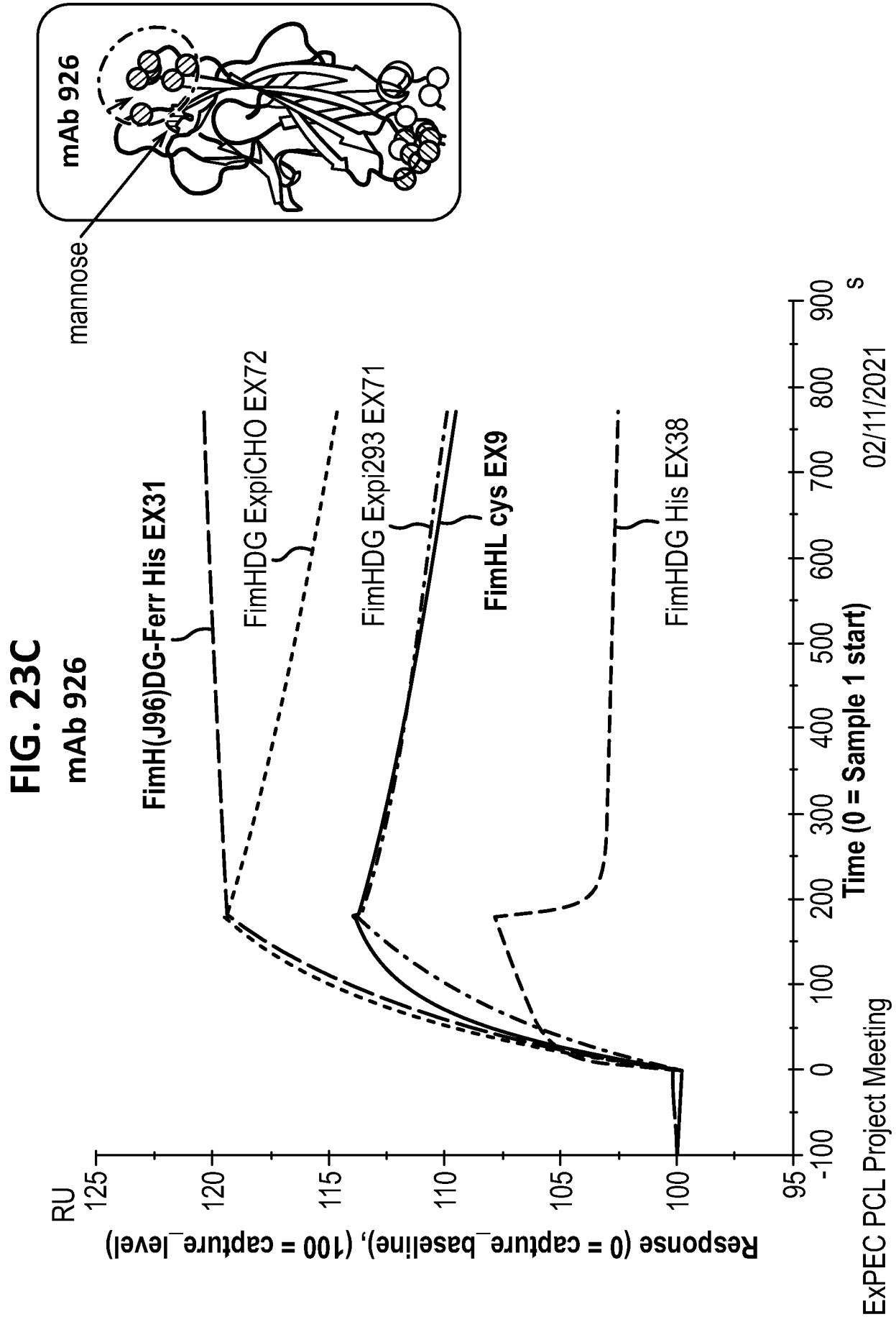
1- Expi293 non transfected

2- Expi293 transfected FimHDG tagless

FIG. 23 B



	EX38 - FimHDG His Expi293	EX71 - FimH-DG Expi293	EX72 FimH-DG CHO
Tm1	n.d.	56°C	56°C
Tm2	71°C	74°C	74°C


mAb 926
mannose
FIG. 23C
mAb 926
FimH(J96)DG-Ferr His EX31
FimHDG ExpiCHO EX72
FimHDG Expi293 EX71
FimHL cys EX9
FimHDG His EX38
RU
125
120
115
110
105
100
95
Response (0 = capture_baseline), (100 = capture_level)
-100
0
100
200
300
400
500
600
700
800
900
s
Time (0 = Sample 1 start)
02/11/2021
ExPEC PCL Project Meeting

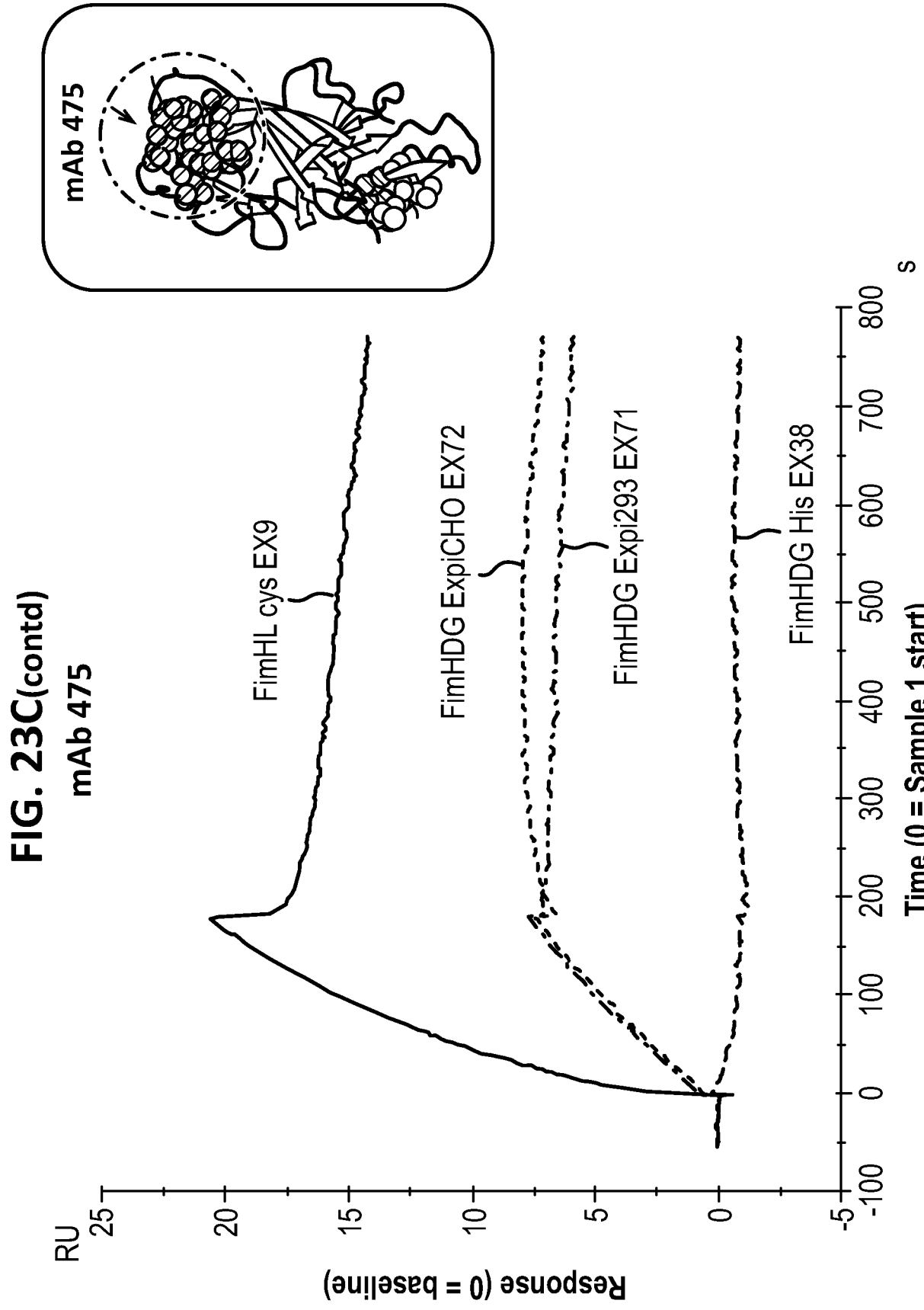

mAb 475
FIG. 23C(contd)
mAb 475
FimHL cys EX9
FimHDG ExpiCHO EX72
FimHDG Expi293 EX71
FimHDG His EX38
RU
25
20
15
10
5
0
-5
Response (0 = baseline)
-100
0
100
200
300
400
500
600
700
800
s
Time (0 = Sample 1 start)

FIG. 23C(contd)
D-Mannose

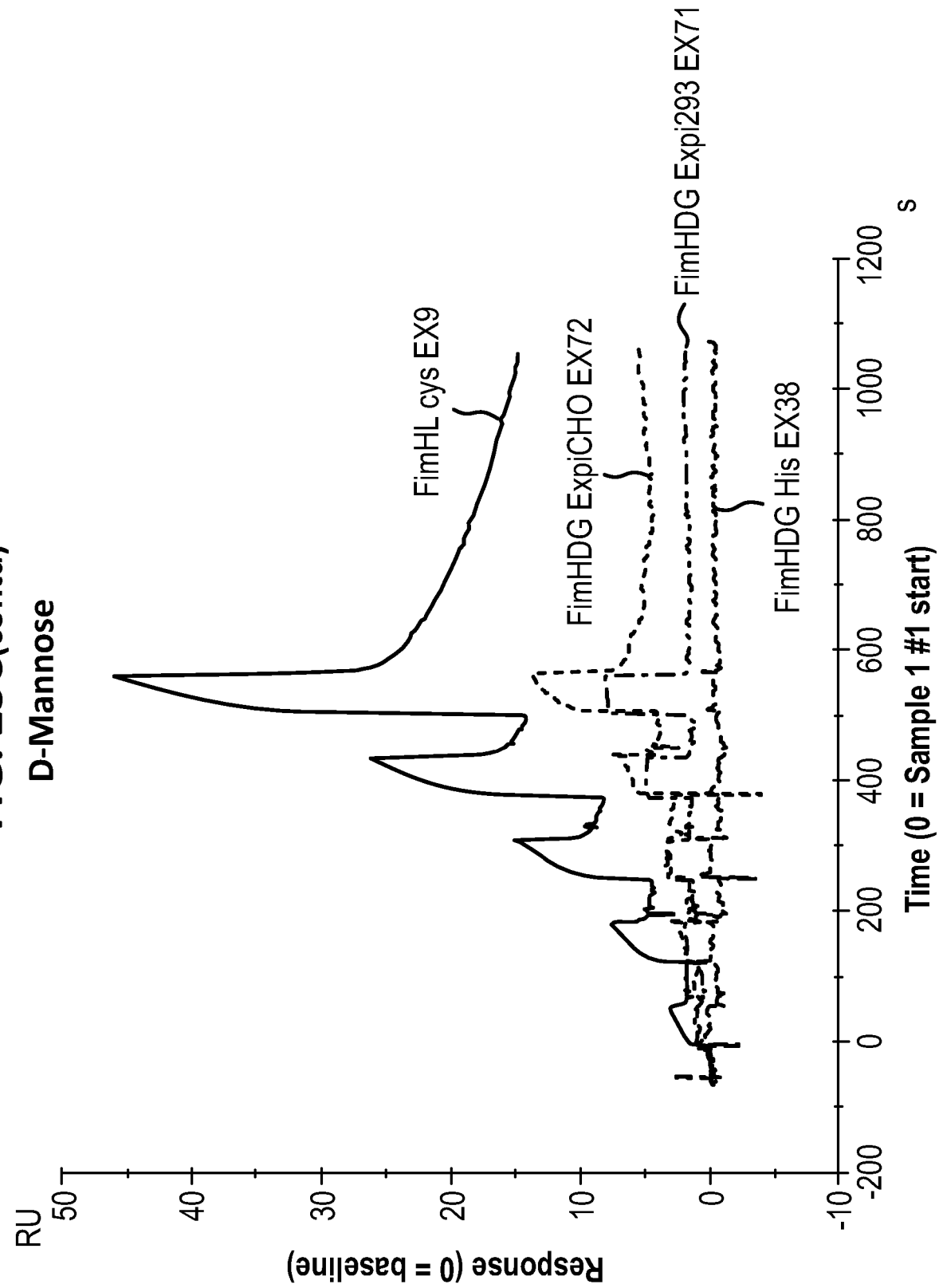
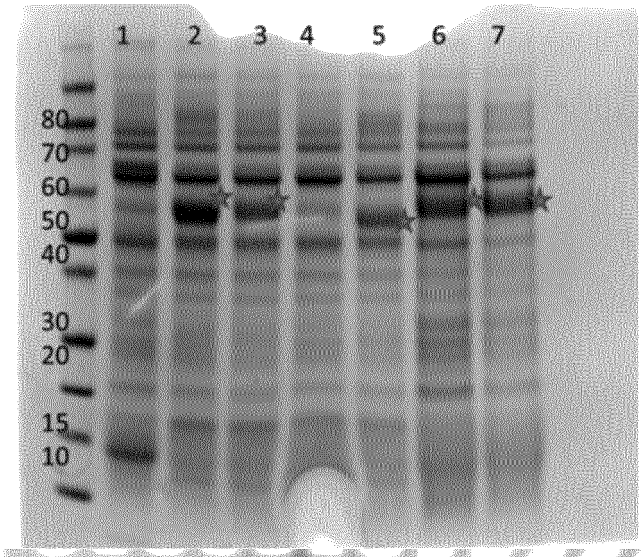
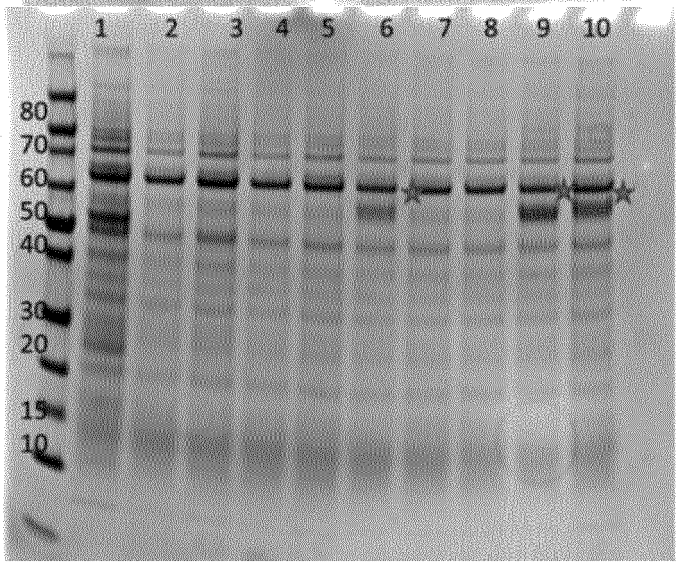


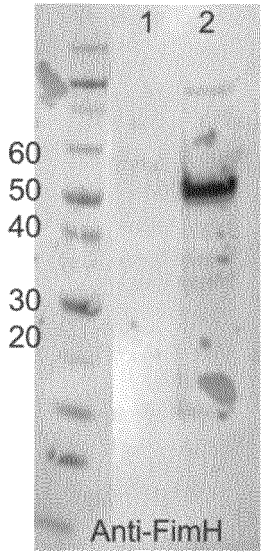
FIG. 23 D



1-expi293 negative controls (mab)
2-1627SI FIMH_DG_PGDGN-GGS4-Ferritin536
3-1623 SI FIMH_DG_GSG4-ferritin_J96
4-1433SI FimH_PGDGN_DG_GGS4-Ferritn j96
5-1185SI FIMHDG_0.5 E.coli ferritin
6-1041SI FIMH_DG_PGDGN-HIS-Ferritin 536
7-1042SI FimH_PGDGN_DG_HIS-Ferritn j96

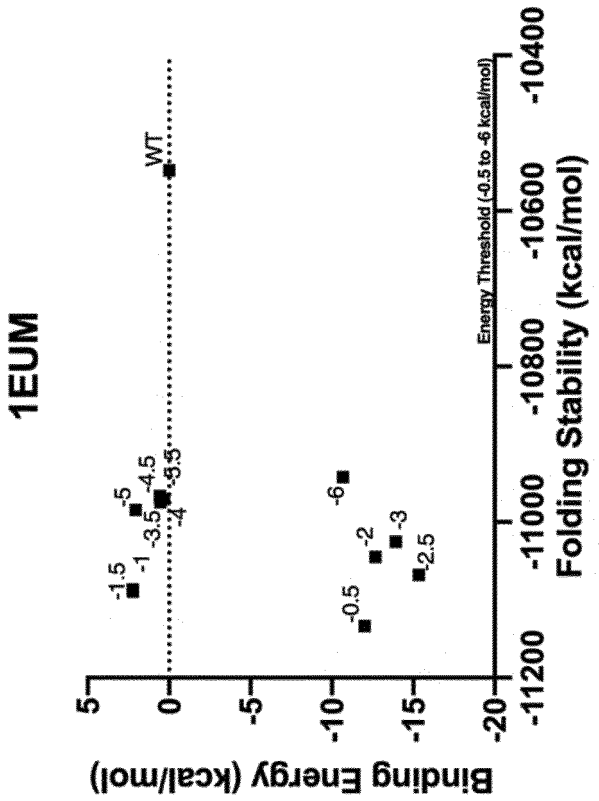


1-Expi293 negative control (mab)
2-1617 SI FimH_NOD-FERRITN_J96
3-1619 SI FimH_NOD_S_HIS_FERRITN_J96
4- 1621 SI fimh-DG_ferritina-linkerAlpha
5-1622 SI FimH-PADRE-Ferritin536_noSpaces
6-1623 SI FimH_DG_GSG4-ferritin_J96
7-1624SI FimH_DG_PGDGN-SGS_PADRE-Ferritinj96
8- 1625SI FimH-ferritina-linkerNONa
9-1627SI FimH_DG_PGDGN-GGS4-Ferritin536
10-1042SI FimH_PGDGN_DG_HIS-Ferritn j96



1-pellet cells not transfected
2-pellet cells FimHDG-Ferritin tagless (RIMS 1433 SI)

FIG. 24


1EUM
WT
-1.5
-1
-3.5
-4
-5
-4.5
-5.5
-6
-0.5
-2
-3
-2.5
5
0
-5
-10
-15
-20
Binding Energy (kcal/mol)
-11200
-11000
-10800
-10600
-10400
Folding Stability (kcal/mol)
Energy Threshold (-0.5 to -6 kcal/mol)

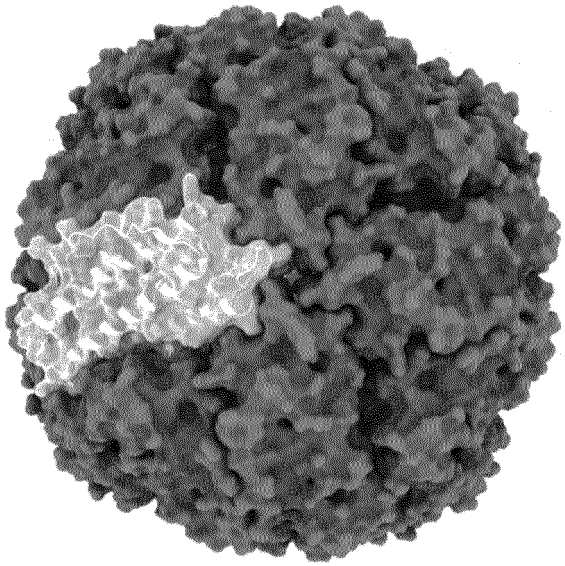


FIG. 25A

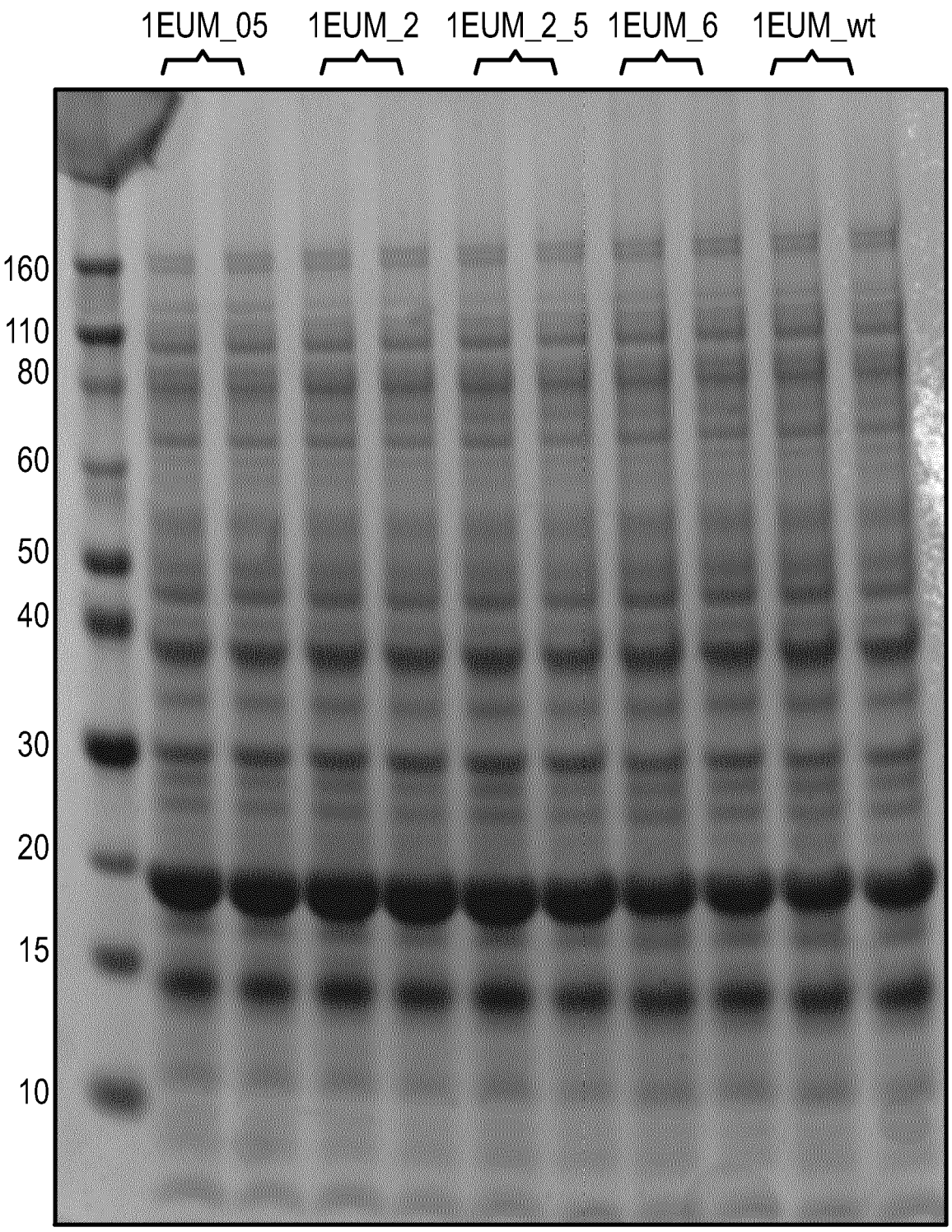

1EUM_05
1EUM_2
1EUM_2_5
1EUM_6
1EUM_wt
160
110
80
60
50
40
30
20
15
10

FIG. 25B

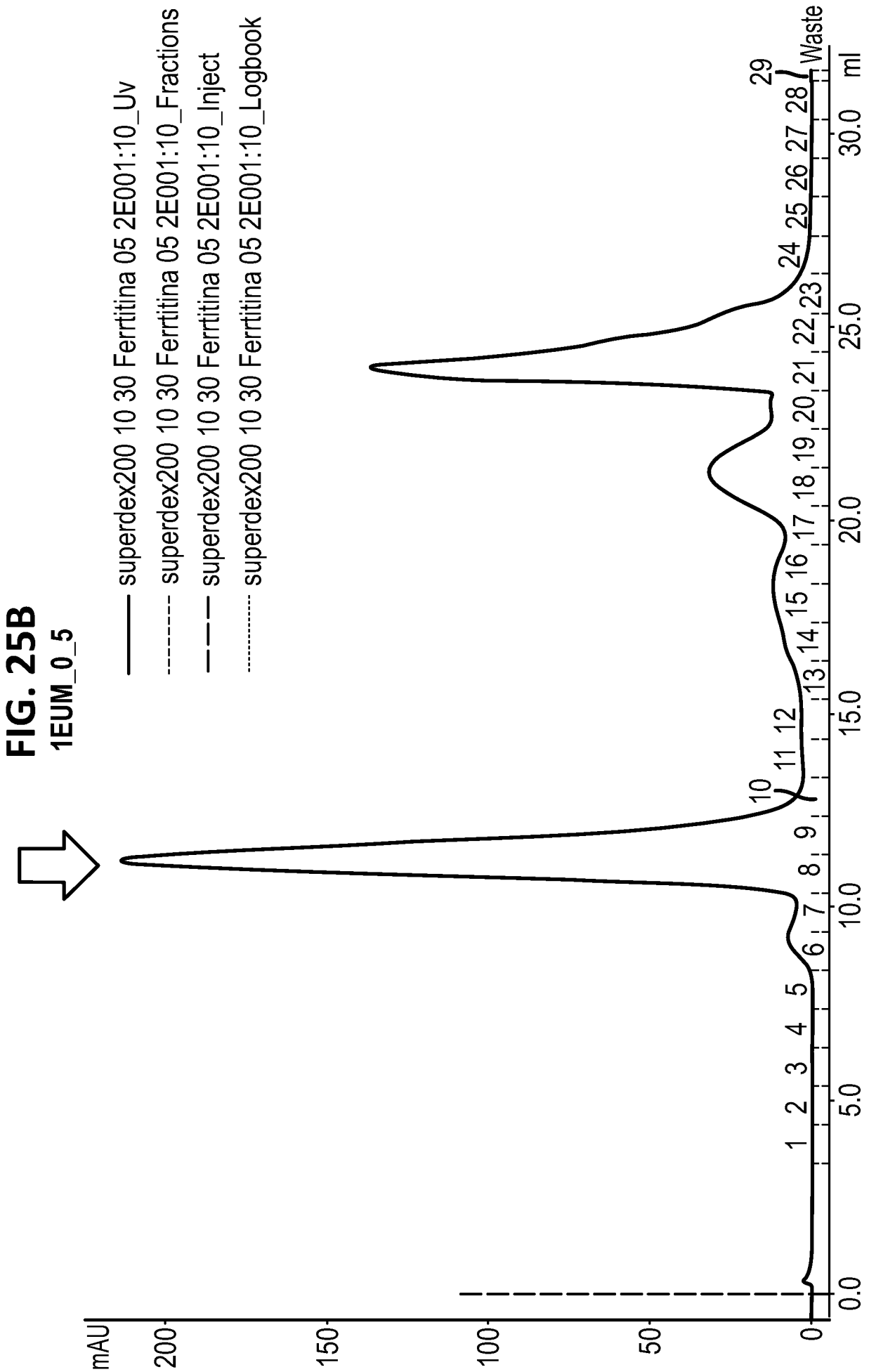

1EUM_0_5
superdex200 10 30 Ferritina 05 2E001:10_Uv
superdex200 10 30 Ferritina 05 2E001:10_Fractions
superdex200 10 30 Ferritina 05 2E001:10_Inject
superdex200 10 30 Ferritina 05 2E001:10_Logbook
mAU
200
150
100
50
0
0.0
5.0
10.0
15.0
20.0
25.0
30.0
ml
Waste
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29

FIG. 26

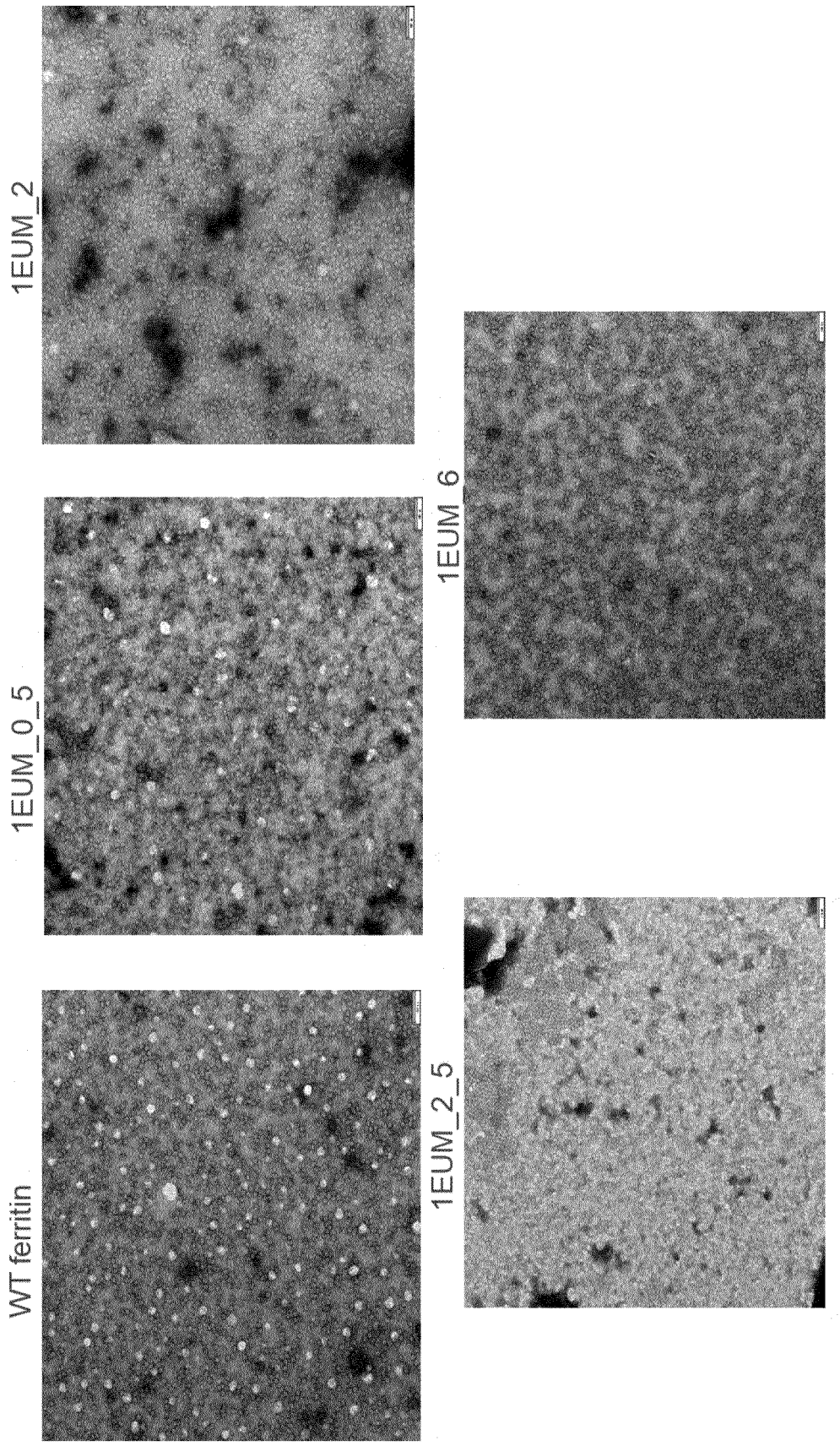


FIG. 27

Ferritin	Tm [°C]
WT	72,5
0.5	76,5
2	75,5
6	74,5

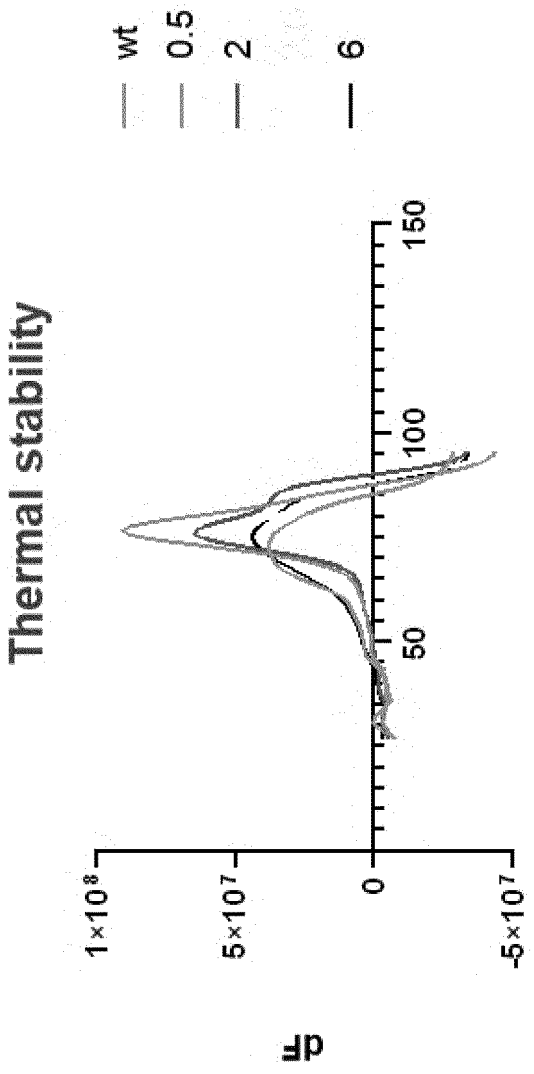
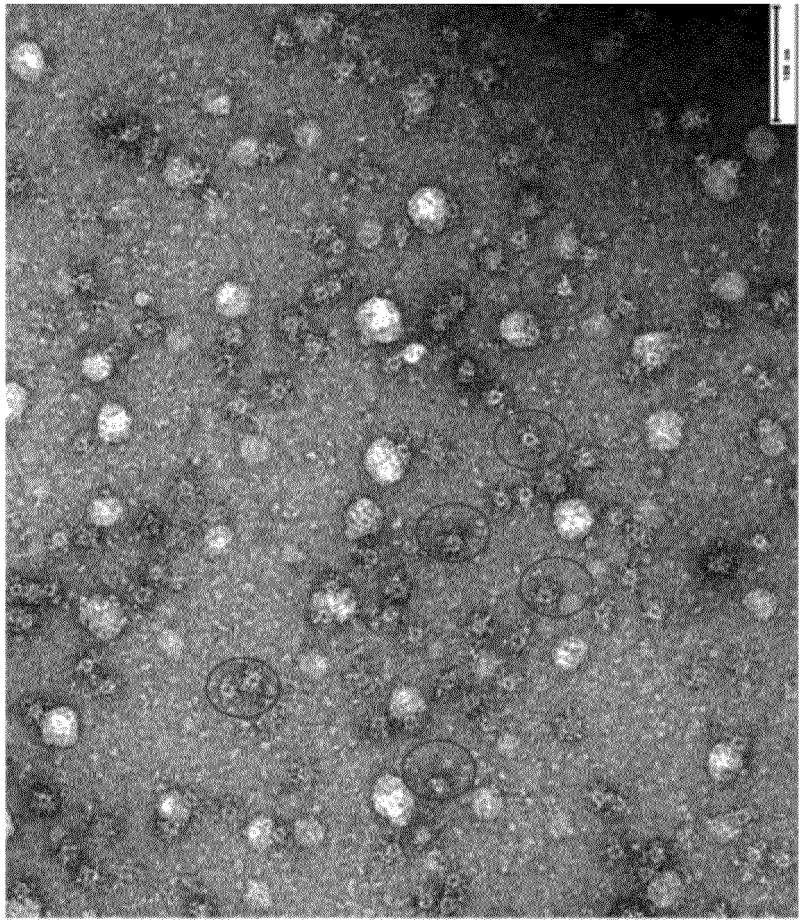
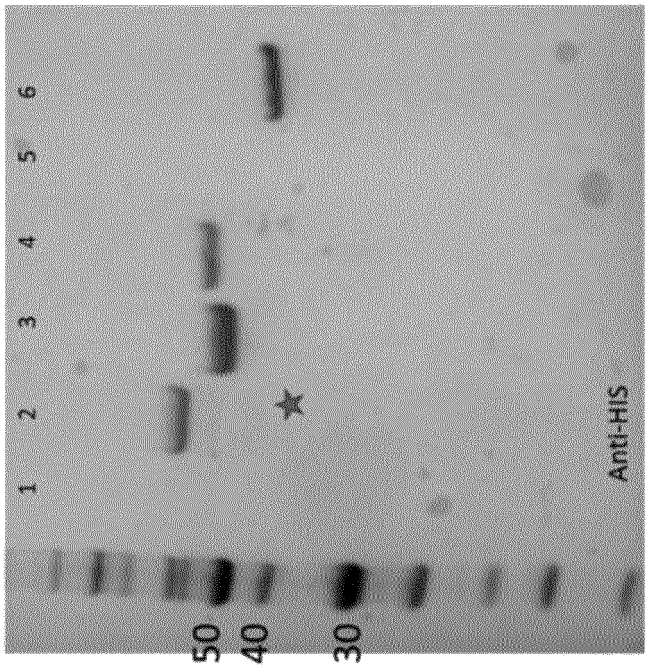


FIG. 28





NPS:
1- Not transfected cells
2- 1184SI FIMH_DG_PGDGN_536-encapsulin
3- 1185SI FIMH_DG_536-1EUM_0_5
4-1183SI FIMH_DG_536-MI3
5-1126SI FIMH_DG_J96-Hbc
6-1096SI FIMHL-MI3

SEQUENCE LISTING

<110> GlaxoSmithKline Biologicals S.A.

<120> Novel antigens

<130> VB67013 FF

<150> EP20211337.9
<151> 2020-12-02

<150> EP20214942.3
<151> 2020-12-17

<160> 153

<170> PatentIn version 3.5

<210> 1
<211> 300
<212> PRT
<213> E. coli

<400> 1

```
Met Lys Arg Val Ile Thr Leu Phe Ala Val Leu Leu Met Gly Trp Ser
1               5                  10                  15


Val Asn Ala Trp Ser Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile
            20                  25                  30


Pro Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val
        35                  40                  45


Val Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe
    50                  55                  60


Cys His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln
65                  70                  75                  80


Arg Gly Ser Ala Tyr Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val
                85                  90                  95


Lys Tyr Ser Gly Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro
```

```
            100                 105                 110

Arg Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu
        115                 120                 125

Tyr Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly
    130                 135                 140

Ser Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser
145                 150                 155                 160

Asp Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val
                165                 170                 175

Val Pro Thr Gly Gly Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr
            180                 185                 190

Leu Pro Asp Tyr Pro Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys
        195                 200                 205

Ala Lys Ser Gln Asn Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp
    210                 215                 220

Ala Gly Asn Ser Ile Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln
225                 230                 235                 240

Gly Val Gly Val Gln Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn
                245                 250                 255

Asn Thr Val Ser Leu Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly
            260                 265                 270

Leu Thr Ala Asn Tyr Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn
        275                 280                 285

Val Gln Ser Ile Ile Gly Val Thr Phe Val Tyr Gln
    290                 295                 300
```

<210> 2
<211> 279
<212> PRT
<213> E. coli

<400> 2

```
Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro Ile Gly Gly Gly
1               5                   10                  15


Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val Val Asn Val Gly Gln
            20                  25                  30


Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys His Asn Asp Tyr
        35                  40                  45


Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg Gly Ser Ala Tyr
    50                  55                  60


Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val Lys Tyr Ser Gly Ser
65                  70                  75                  80


Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg Val Val Tyr Asn
                85                  90                  95


Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr Leu Thr Pro Val
            100                 105                 110


Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser Leu Ile Ala Val
        115                 120                 125


Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp Asp Phe Gln Phe
    130                 135                 140


Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val Pro Thr Gly Gly
145                 150                 155                 160


Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu Pro Asp Tyr Pro
                165                 170                 175
```

```
Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala Lys Ser Gln Asn
            180                 185                 190


Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp Ala Gly Asn Ser Ile
        195                 200                 205


Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly Val Gly Val Gln
    210                 215                 220


Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn Thr Val Ser Leu
225                 230                 235                 240


Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu Thr Ala Asn Tyr
                245                 250                 255


Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val Gln Ser Ile Ile
            260                 265                 270


Gly Val Thr Phe Val Tyr Gln
        275
```

<210> 3
<211> 28
<212> PRT
<213> E. coli

<400> 3

```
Ala Ser Ala Thr Ile Gln Ala Ala Asp Val Thr Ile Thr Val Asn Gly
1               5                   10                  15


Lys Val Val Ala Lys Pro Cys Thr Val Ser Thr Thr
            20                  25
```

<210> 4
<211> 36
<212> PRT
<213> E. coli

<400> 4

Pro Ser Met Asp Lys Ser Lys Leu Thr Glu Asn Thr Leu Gln Leu Ala
1 5 10 15

Ile Ile Ser Arg Ile Lys Leu Tyr Tyr Arg Pro Ala Lys Leu Ala Leu
20 25 30

Pro Pro Asp Gln
35

<210> 5
<211> 14
<212> PRT
<213> E. coli

<400> 5

Ala Asp Val Thr Ile Thr Val Asn Gly Lys Val Val Ala Lys
1 5 10

<210> 6
<211> 18
<212> PRT
<213> E. coli

<400> 6

Glu Asn Thr Leu Gln Leu Ala Ile Ile Ser Arg Ile Lys Leu Tyr Tyr
1 5 10 15

Arg Pro

<210> 7
<211> 5
<212> PRT
<213> Artificial

<220>
<223> Linker

<400> 7

Pro Gly Asp Gly Asn
1   5

<210> 8
<211> 4
<212> PRT
<213> Artificial

<220>
<223> Linker

<400> 8

Asp Asn Lys Gln
1

<210> 9
<211> 21
<212> PRT
<213> Artificial

<220>
<223> Leader

<400> 9

Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1   5   10   15

Gly Ser Thr Gly Asp
            20

<210> 10
<211> 36
<212> PRT
<213> Artificial

<220>
<223> Leader

<400> 10

Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1   5   10   15

Gly Ser Thr Gly Asp Ala Ala Gln Pro Ala Arg Arg Ala Arg Arg Thr
            20   25   30

Lys Leu Ala Leu
        35

<210> 11
<211> 18
<212> PRT
<213> Artificial

<220>
<223> Leader

<400> 11

Met Arg Leu Leu Ala Lys Ile Ile Cys Leu Met Leu Trp Ala Ile Cys
1   5   10   15

Val Ala

<210> 12
<211> 19
<212> PRT
<213> Artificial

<220>
<223> Leader

<400> 12

Met Gly Trp Ser Cys Ile Ile Leu Phe Leu Val Ala Thr Ala Thr Gly
1   5   10   15

Val His Ser

<210> 13
<211> 20
<212> PRT
<213> Artificial

<220>

<223> Leader

<400> 13

Met Glu Thr Pro Ala Glu Leu Leu Phe Leu Leu Leu Leu Trp Leu Pro
1   5   10   15

Asp Thr Thr Gly
            20

<210> 14
<211> 19
<212> PRT
<213> Artificial

<220>
<223> Leader

<400> 14

Met Glu Phe Gly Leu Ser Trp Val Phe Leu Val Ala Ile Leu Glu Gly
1   5   10   15

Val His Cys

<210> 15
<211> 162
<212> PRT
<213> H. pylori

<400> 15

Asp Ile Ile Lys Leu Leu Asn Glu Gln Val Asn Lys Glu Met Asn Ser
1   5   10   15

Ser Asn Leu Tyr Met Ser Met Ser Ser Trp Cys Tyr Thr His Ser Leu
            20   25   30

Asp Gly Ala Gly Leu Phe Leu Phe Asp His Ala Ala Glu Glu Tyr Glu
        35   40   45

His Ala Lys Lys Leu Ile Ile Phe Leu Asn Glu Asn Asn Val Pro Val

```
    50                  55                  60

Gln Leu Thr Ser Ile Ser Ala Pro Glu His Lys Phe Glu Gly Leu Thr
65                  70                  75                  80

Gln Ile Phe Gln Lys Ala Tyr Glu His Glu Gln His Ile Ser Glu Ser
                85                  90                  95

Ile Asn Asn Ile Val Asp His Ala Ile Lys Ser Lys Asp His Ala Thr
            100                 105                 110

Phe Asn Phe Leu Gln Trp Tyr Val Ala Glu Gln His Glu Glu Glu Val
        115                 120                 125

Leu Phe Lys Asp Ile Leu Asp Lys Ile Glu Leu Ile Gly Asn Glu Asn
    130                 135                 140

His Gly Leu Tyr Leu Ala Asp Gln Tyr Val Lys Gly Ile Ala Lys Ser
145                 150                 155                 160

Arg Lys
```

<210> 16
<211> 164
<212> PRT
<213> E. coli

<400> 16

```
Leu Lys Pro Glu Met Ile Glu Lys Leu Asn Glu Gln Met Asn Leu Glu
1               5                   10                  15

Leu Tyr Ser Ser Leu Leu Tyr Gln Gln Met Ser Ala Trp Cys Ser Tyr
            20                  25                  30

His Thr Phe Glu Gly Ala Ala Ala Phe Leu Arg Arg His Ala Gln Glu
        35                  40                  45
```

```
Glu Met Thr His Met Gln Arg Leu Phe Asp Tyr Leu Thr Asp Thr Gly
    50                  55                  60

Asn Leu Pro Arg Ile Asn Thr Val Glu Ser Pro Phe Ala Glu Tyr Ser
65                  70                  75                  80

Ser Leu Asp Glu Leu Phe Gln Glu Thr Tyr Lys His Glu Gln Leu Ile
                85                  90                  95

Thr Gln Lys Ile Asn Glu Leu Ala His Ala Ala Met Thr Asn Gln Asp
            100                 105                 110

Tyr Pro Thr Phe Asn Phe Leu Gln Trp Tyr Val Ser Glu Gln His Glu
        115                 120                 125

Glu Glu Lys Leu Phe Lys Ser Ile Ile Asp Lys Leu Ser Leu Ala Gly
    130                 135                 140

Lys Ser Gly Glu Gly Leu Tyr Phe Ile Asp Lys Glu Leu Ser Thr Leu
145                 150                 155                 160

Asp Thr Gln Asn
```

<210> 17
<211> 56
<212> PRT
<213> Artificial

<220>
<223> iMX313

<400> 17

```
Lys Lys Gln Gly Asp Ala Asp Val Cys Gly Glu Val Ala Tyr Ile Gln
1               5                   10                  15

Ser Val Val Ser Asp Cys His Val Pro Thr Ala Glu Leu Arg Thr Leu
            20                  25                  30
```

```
Leu Glu Ile Arg Lys Leu Phe Leu Glu Ile Gln Lys Leu Lys Val Glu
        35                  40                  45


Leu Gln Gly Leu Ser Lys Glu Gly
    50                  55


<210>  18
<211>  205
<212>  PRT
<213>  Artificial

<220>
<223>  mi3

<400>  18

Met Lys Met Glu Glu Leu Phe Lys Lys His Lys Ile Val Ala Val Leu
1               5                   10                  15


Arg Ala Asn Ser Val Glu Glu Ala Lys Lys Lys Ala Leu Ala Val Phe
            20                  25                  30


Leu Gly Gly Val His Leu Ile Glu Ile Thr Phe Thr Val Pro Asp Ala
        35                  40                  45


Asp Thr Val Ile Lys Glu Leu Ser Phe Leu Lys Glu Met Gly Ala Ile
    50                  55                  60


Ile Gly Ala Gly Thr Val Thr Ser Val Glu Gln Ala Arg Lys Ala Val
65                  70                  75                  80


Glu Ser Gly Ala Glu Phe Ile Val Ser Pro His Leu Asp Glu Glu Ile
                85                  90                  95


Ser Gln Phe Ala Lys Glu Lys Gly Val Phe Tyr Met Pro Gly Val Met
            100                 105                 110


Thr Pro Thr Glu Leu Val Lys Ala Met Lys Leu Gly His Thr Ile Leu
        115                 120                 125
```

```
Lys Leu Phe Pro Gly Glu Val Val Gly Pro Gln Phe Val Lys Ala Met
    130                 135                 140

Lys Gly Pro Phe Pro Asn Val Lys Phe Val Pro Thr Gly Gly Val Asn
145                 150                 155                 160

Leu Asp Asn Val Cys Glu Trp Phe Lys Ala Gly Val Leu Ala Val Gly
                165                 170                 175

Val Gly Ser Ala Leu Val Lys Gly Thr Pro Val Glu Val Ala Glu Lys
            180                 185                 190

Ala Lys Ala Phe Val Glu Lys Ile Arg Gly Cys Thr Glu
        195                 200                 205
```

<210> 19
<211> 265
<212> PRT
<213> Artificial

<220>
<223> encapsulin

<400> 19

```
Met Glu Phe Leu Lys Arg Ser Phe Ala Pro Leu Thr Glu Lys Gln Trp
1               5                   10                  15

Gln Glu Ile Asp Asn Arg Ala Arg Glu Ile Phe Lys Thr Gln Leu Tyr
            20                  25                  30

Gly Arg Lys Phe Val Asp Val Glu Gly Pro Tyr Gly Trp Glu Tyr Ala
        35                  40                  45

Ala His Pro Leu Gly Glu Val Glu Val Leu Ser Asp Glu Asn Glu Val
    50                  55                  60

Val Lys Trp Gly Leu Arg Lys Ser Leu Pro Leu Ile Glu Leu Arg Ala
65                  70                  75                  80
```

```
Thr Phe Thr Leu Asp Leu Trp Glu Leu Asp Asn Leu Glu Arg Gly Lys
                85                  90                  95

Pro Asn Val Asp Leu Ser Ser Leu Glu Glu Thr Val Arg Lys Val Ala
            100                 105                 110

Glu Phe Glu Asp Glu Val Ile Phe Arg Gly Cys Glu Lys Ser Gly Val
        115                 120                 125

Lys Gly Leu Leu Ser Phe Glu Glu Arg Lys Ile Glu Cys Gly Ser Thr
    130                 135                 140

Pro Lys Asp Leu Leu Glu Ala Ile Val Arg Ala Leu Ser Ile Phe Ser
145                 150                 155                 160

Lys Asp Gly Ile Glu Gly Pro Tyr Thr Leu Val Ile Asn Thr Asp Arg
                165                 170                 175

Trp Ile Asn Phe Leu Lys Glu Glu Ala Gly His Tyr Pro Leu Glu Lys
            180                 185                 190

Arg Val Glu Glu Cys Leu Arg Gly Gly Lys Ile Ile Thr Thr Pro Arg
        195                 200                 205

Ile Glu Asp Ala Leu Val Val Ser Glu Arg Gly Gly Asp Phe Lys Leu
    210                 215                 220

Ile Leu Gly Gln Asp Leu Ser Ile Gly Tyr Glu Asp Arg Glu Lys Asp
225                 230                 235                 240

Ala Val Arg Leu Phe Ile Thr Glu Thr Phe Thr Phe Gln Val Val Asn
                245                 250                 255

Pro Glu Ala Leu Ile Leu Leu Lys Phe
            260                 265
```

<210> 20
<211> 475

<212> PRT
<213> Artificial

<220>
<223> FIMH_DG_PGDGN_HIS-Ferritin 536

<400> 20

```
Met Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro Ile Gly Gly
1               5                  10                  15


Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Ala Val Asn Val Gly
            20                  25                  30


Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys His Asn Asp
        35                  40                  45


Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg Gly Ser Ala
    50                  55                  60


Tyr Gly Gly Val Leu Ser Ser Phe Ser Gly Thr Val Lys Tyr Asn Gly
65                  70                  75                  80


Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg Val Val Tyr
                85                  90                  95


Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr Leu Thr Pro
            100                 105                 110


Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser Leu Ile Ala
        115                 120                 125


Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp Asp Phe Gln
    130                 135                 140


Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val Pro Thr Gly
145                 150                 155                 160


Gly Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu Pro Asp Tyr
                165                 170                 175
```

```
Pro Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala Lys Ser Gln
            180                 185                 190

Asn Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp Ala Gly Asn Ser
        195                 200                 205

Ile Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly Val Gly Val
    210                 215                 220

Gln Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn Thr Val Ser
225                 230                 235                 240

Leu Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu Thr Ala Asn
                245                 250                 255

Tyr Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val Gln Ser Ile
            260                 265                 270

Ile Gly Val Thr Phe Val Tyr Gln Pro Gly Asp Gly Asn Ala Asp Val
        275                 280                 285

Thr Ile Thr Val Asn Gly Lys Val Val Ala Lys Ser Gly Ser His His
    290                 295                 300

His His His His His His Gly Gly Ser Asp Ile Ile Lys Leu Leu Asn
305                 310                 315                 320

Glu Gln Val Asn Lys Glu Met Asn Ser Ser Asn Leu Tyr Met Ser Met
                325                 330                 335

Ser Ser Trp Cys Tyr Thr His Ser Leu Asp Gly Ala Gly Leu Phe Leu
            340                 345                 350

Phe Asp His Ala Ala Glu Glu Tyr Glu His Ala Lys Lys Leu Ile Ile
        355                 360                 365
```

```
Phe Leu Asn Glu Asn Asn Val Pro Val Gln Leu Thr Ser Ile Ser Ala
    370                 375                 380

Pro Glu His Lys Phe Glu Gly Leu Thr Gln Ile Phe Gln Lys Ala Tyr
385                 390                 395                 400

Glu His Glu Gln His Ile Ser Glu Ser Ile Asn Asn Ile Val Asp His
                405                 410                 415

Ala Ile Lys Ser Lys Asp His Ala Thr Phe Asn Phe Leu Gln Trp Tyr
            420                 425                 430

Val Ala Glu Gln His Glu Glu Glu Val Leu Phe Lys Asp Ile Leu Asp
        435                 440                 445

Lys Ile Glu Leu Ile Gly Asn Glu Asn His Gly Leu Tyr Leu Ala Asp
    450                 455                 460

Gln Tyr Val Lys Gly Ile Ala Lys Ser Arg Lys
465                 470                 475
```

<210> 21
<211> 259
<212> PRT
<213> Artificial

<220>
<223> LS-FIMHL-IMX313-HIS

<400> 21

```
Met Lys Tyr Leu Leu Pro Thr Ala Ala Ala Gly Leu Leu Leu Leu Ala
1               5                   10                  15

Ala Gln Pro Ala Met Ala Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala
            20                  25                  30

Ile Pro Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro
        35                  40                  45
```

```
Val Val Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile
    50                  55                  60

Phe Cys His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu
65                  70                  75                  80

Gln Arg Gly Ser Ala Tyr Gly Gly Val Leu Ser Asn Phe Ser Gly Thr
                85                  90                  95

Val Lys Tyr Ser Gly Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr
            100                 105                 110

Pro Arg Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala
        115                 120                 125

Leu Tyr Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala
    130                 135                 140

Gly Ser Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn
145                 150                 155                 160

Ser Asp Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val
                165                 170                 175

Val Val Pro Thr Gly Gly Ser Ser Gly Ser Gly Ser Gly Ser Lys Lys
            180                 185                 190

Gln Gly Asp Ala Asp Val Cys Gly Glu Val Ala Tyr Ile Gln Ser Val
        195                 200                 205

Val Ser Asp Cys His Val Pro Thr Ala Glu Leu Arg Thr Leu Leu Glu
    210                 215                 220

Ile Arg Lys Leu Phe Leu Glu Ile Gln Lys Leu Lys Val Glu Leu Gln
225                 230                 235                 240

Gly Leu Ser Lys Glu Gly Gly Gly Ser Gly Ser His His His His His
                245                 250                 255
```

His His His

<210> 22
<211> 236
<212> PRT
<213> Artificial

<220>
<223> FIMHL-S24S65- IMX313

<400> 22

```
Met Phe Ala Ser Lys Thr Ala Asn Gly Thr Ala Ile Pro Ile Gly Gly
1               5                   10                  15


Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val Val Asn Val Gly
            20                  25                  30


Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Ser His Asn Asp
        35                  40                  45


Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg Gly Ser Ala
    50                  55                  60


Tyr Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val Lys Tyr Ser Gly
65                  70                  75                  80


Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg Val Val Tyr
                85                  90                  95


Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr Leu Thr Pro
            100                 105                 110


Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser Leu Ile Ala
        115                 120                 125


Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp Asp Phe Gln
    130                 135                 140
```

```
Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val Pro Thr Gly
145                 150                 155                 160


Gly Ser Ser Gly Ser Gly Ser Gly Ser Lys Lys Gln Gly Asp Ala Asp
                165                 170                 175


Val Cys Gly Glu Val Ala Tyr Ile Gln Ser Val Val Ser Asp Cys His
            180                 185                 190


Val Pro Thr Ala Glu Leu Arg Thr Leu Leu Glu Ile Arg Lys Leu Phe
        195                 200                 205


Leu Glu Ile Gln Lys Leu Lys Val Glu Leu Gln Gly Leu Ser Lys Glu
    210                 215                 220


Gly Gly Gly Ser Gly Ser His His His His His His
225                 230                 235


<210>  23
<211>  362
<212>  PRT
<213>  Artificial

<220>
<223>  FIMHL-S24S65-foldon -ferritin

<400>  23

Met Phe Ala Ser Lys Thr Ala Asn Gly Thr Ala Ile Pro Ile Gly Gly
1               5                   10                  15


Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Ala Val Asn Val Gly
            20                  25                  30


Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Ser His Asn Asp
        35                  40                  45


Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg Gly Ser Ala
    50                  55                  60
```

```
Tyr Gly Gly Val Leu Ser Ser Phe Ser Gly Thr Val Lys Tyr Asn Gly
65                  70                  75                  80

Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg Val Val Tyr
                85                  90                  95

Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr Leu Thr Pro
            100                 105                 110

Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser Leu Ile Ala
        115                 120                 125

Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp Asp Phe Gln
    130                 135                 140

Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val Pro Thr Gly
145                 150                 155                 160

Ser Gly Tyr Ile Pro Glu Ala Pro Arg Asp Gly Gln Ala Tyr Val Arg
                165                 170                 175

Lys Asp Gly Glu Trp Val Leu Leu Ser Thr Phe Leu Gly Ser Gly His
            180                 185                 190

His His His His His Gly Ser Gly Asp Ile Ile Lys Leu Leu Asn Glu
        195                 200                 205

Gln Val Asn Lys Glu Met Asn Ser Ser Asn Leu Tyr Met Ser Met Ser
    210                 215                 220

Ser Trp Cys Tyr Thr His Ser Leu Asp Gly Ala Gly Leu Phe Leu Phe
225                 230                 235                 240

Asp His Ala Ala Glu Glu Tyr Glu His Ala Lys Lys Leu Ile Ile Phe
                245                 250                 255
```

```
Leu Asn Glu Asn Asn Val Pro Val Gln Leu Thr Ser Ile Ser Ala Pro
            260                 265                 270

Glu His Lys Phe Glu Gly Leu Thr Gln Ile Phe Gln Lys Ala Tyr Glu
        275                 280                 285

His Glu Gln His Ile Ser Glu Ser Ile Asn Asn Ile Val Asp His Ala
    290                 295                 300

Ile Lys Ser Lys Asp His Ala Thr Phe Asn Phe Leu Gln Trp Tyr Val
305                 310                 315                 320

Ala Glu Gln His Glu Glu Glu Val Leu Phe Lys Asp Ile Leu Asp Lys
                325                 330                 335

Ile Glu Leu Ile Gly Asn Glu Asn His Gly Leu Tyr Leu Ala Asp Gln
            340                 345                 350

Tyr Val Lys Gly Ile Ala Lys Ser Arg Lys
        355                 360


<210>  24
<211>  389
<212>  PRT
<213>  Artificial

<220>
<223>  FIMHL-S24S65-Mi3

<400>  24

Met Phe Ala Ser Lys Thr Ala Asn Gly Thr Ala Ile Pro Ile Gly Gly
1               5                   10                  15

Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Ala Val Asn Val Gly
            20                  25                  30

Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Ser His Asn Asp
        35                  40                  45
```

```
Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg Gly Ser Ala
    50                  55                  60

Tyr Gly Gly Val Leu Ser Ser Phe Ser Gly Thr Val Lys Tyr Asn Gly
65                  70                  75                  80

Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg Val Val Tyr
                85                  90                  95

Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr Leu Thr Pro
            100                 105                 110

Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser Leu Ile Ala
        115                 120                 125

Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp Asp Phe Gln
    130                 135                 140

Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val Pro Thr Gly
145                 150                 155                 160

Gly Ser Gly Gly Ser Gly Gly Ser Met Lys Met Glu Glu Leu Phe Lys
                165                 170                 175

Lys His Lys Ile Val Ala Val Leu Arg Ala Asn Ser Val Glu Glu Ala
            180                 185                 190

Lys Lys Lys Ala Leu Ala Val Phe Leu Gly Gly Val His Leu Ile Glu
        195                 200                 205

Ile Thr Phe Thr Val Pro Asp Ala Asp Thr Val Ile Lys Glu Leu Ser
    210                 215                 220

Phe Leu Lys Glu Met Gly Ala Ile Ile Gly Ala Gly Thr Val Thr Ser
225                 230                 235                 240

Val Glu Gln Ala Arg Lys Ala Val Glu Ser Gly Ala Glu Phe Ile Val
                245                 250                 255
```

```
Ser Pro His Leu Asp Glu Glu Ile Ser Gln Phe Ala Lys Glu Lys Gly
            260                 265                 270


Val Phe Tyr Met Pro Gly Val Met Thr Pro Thr Glu Leu Val Lys Ala
        275                 280                 285


Met Lys Leu Gly His Thr Ile Leu Lys Leu Phe Pro Gly Glu Val Val
    290                 295                 300


Gly Pro Gln Phe Val Lys Ala Met Lys Gly Pro Phe Pro Asn Val Lys
305                 310                 315                 320


Phe Val Pro Thr Gly Gly Val Asn Leu Asp Asn Val Cys Glu Trp Phe
                325                 330                 335


Lys Ala Gly Val Leu Ala Val Gly Val Gly Ser Ala Leu Val Lys Gly
            340                 345                 350


Thr Pro Val Glu Val Ala Glu Lys Ala Lys Ala Phe Val Glu Lys Ile
        355                 360                 365


Arg Gly Cys Thr Glu Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser His
    370                 375                 380


His His His His His
385
```

<210> 25
<211> 389
<212> PRT
<213> Artificial

<220>
<223> FIMHL-mI3

<400> 25

```
Met Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro Ile Gly Gly
1               5                   10                  15
```

```
Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Ala Val Asn Val Gly
            20                  25                  30

Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys His Asn Asp
        35                  40                  45

Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg Gly Ser Ala
    50                  55                  60

Tyr Gly Gly Val Leu Ser Ser Phe Ser Gly Thr Val Lys Tyr Asn Gly
65                  70                  75                  80

Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg Val Val Tyr
                85                  90                  95

Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr Leu Thr Pro
            100                 105                 110

Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser Leu Ile Ala
        115                 120                 125

Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp Asp Phe Gln
    130                 135                 140

Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val Pro Thr Gly
145                 150                 155                 160

Gly Ser Gly Gly Ser Gly Gly Ser Met Lys Met Glu Glu Leu Phe Lys
                165                 170                 175

Lys His Lys Ile Val Ala Val Leu Arg Ala Asn Ser Val Glu Glu Ala
            180                 185                 190

Lys Lys Lys Ala Leu Ala Val Phe Leu Gly Gly Val His Leu Ile Glu
        195                 200                 205
```

```
Ile Thr Phe Thr Val Pro Asp Ala Asp Thr Val Ile Lys Glu Leu Ser
    210                 215                 220

Phe Leu Lys Glu Met Gly Ala Ile Ile Gly Ala Gly Thr Val Thr Ser
225                 230                 235                 240

Val Glu Gln Ala Arg Lys Ala Val Glu Ser Gly Ala Glu Phe Ile Val
                245                 250                 255

Ser Pro His Leu Asp Glu Glu Ile Ser Gln Phe Ala Lys Glu Lys Gly
            260                 265                 270

Val Phe Tyr Met Pro Gly Val Met Thr Pro Thr Glu Leu Val Lys Ala
        275                 280                 285

Met Lys Leu Gly His Thr Ile Leu Lys Leu Phe Pro Gly Glu Val Val
    290                 295                 300

Gly Pro Gln Phe Val Lys Ala Met Lys Gly Pro Phe Pro Asn Val Lys
305                 310                 315                 320

Phe Val Pro Thr Gly Gly Val Asn Leu Asp Asn Val Cys Glu Trp Phe
                325                 330                 335

Lys Ala Gly Val Leu Ala Val Gly Val Gly Ser Ala Leu Val Lys Gly
            340                 345                 350

Thr Pro Val Glu Val Ala Glu Lys Ala Lys Ala Phe Val Glu Lys Ile
        355                 360                 365

Arg Gly Cys Thr Glu Gly Ser Gly Ser Gly Ser Gly Ser Gly Ser His
    370                 375                 380

His His His His His
385
```

<210> 26
<211> 394

<212> PRT
<213> Artificial

<220>
<223> FimHL-NOCYS -MI3

<400> 26

```
Met Phe Ala Ser Lys Thr Ala Asn Gly Thr Ala Ile Pro Ile Gly Gly
1               5                  10                  15


Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Ala Val Asn Val Gly
            20                  25                  30


Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Ser His Asn Asp
        35                  40                  45


Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg Gly Ser Ala
    50                  55                  60


Tyr Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val Lys Tyr Ser Gly
65                  70                  75                  80


Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg Val Val Tyr
                85                  90                  95


Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr Leu Thr Pro
            100                 105                 110


Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser Leu Ile Ala
        115                 120                 125


Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp Asp Phe Gln
    130                 135                 140


Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val Pro Thr Gly
145                 150                 155                 160


Gly Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Met Lys Met
                165                 170                 175
```

```
Glu Glu Leu Phe Lys Lys His Lys Ile Val Ala Val Leu Arg Ala Asn
            180                 185                 190

Ser Val Glu Glu Ala Lys Lys Lys Ala Leu Ala Val Phe Leu Gly Gly
        195                 200                 205

Val His Leu Ile Glu Ile Thr Phe Thr Val Pro Asp Ala Asp Thr Val
    210                 215                 220

Ile Lys Glu Leu Ser Phe Leu Lys Glu Met Gly Ala Ile Ile Gly Ala
225                 230                 235                 240

Gly Thr Val Thr Ser Val Glu Gln Ala Arg Lys Ala Val Glu Ser Gly
                245                 250                 255

Ala Glu Phe Ile Val Ser Pro His Leu Asp Glu Glu Ile Ser Gln Phe
            260                 265                 270

Ala Lys Glu Lys Gly Val Phe Tyr Met Pro Gly Val Met Thr Pro Thr
        275                 280                 285

Glu Leu Val Lys Ala Met Lys Leu Gly His Thr Ile Leu Lys Leu Phe
    290                 295                 300

Pro Gly Glu Val Val Gly Pro Gln Phe Val Lys Ala Met Lys Gly Pro
305                 310                 315                 320

Phe Pro Asn Val Lys Phe Val Pro Thr Gly Gly Val Asn Leu Asp Asn
                325                 330                 335

Val Cys Glu Trp Phe Lys Ala Gly Val Leu Ala Val Gly Val Gly Ser
            340                 345                 350

Ala Leu Val Lys Gly Thr Pro Val Glu Val Ala Glu Lys Ala Lys Ala
        355                 360                 365
```

```
Phe Val Glu Lys Ile Arg Gly Cys Thr Glu Gly Ser Gly Ser Gly Ser
    370                 375                 380


Gly Ser Gly Ser His His His His His His
385                 390


<210>  27
<211>  524
<212>  PRT
<213>  Artificial

<220>
<223>  FimHdeltaGG_PGDGN_DG_mi3

<400>  27

Met Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro Ile Gly Gly
1               5                   10                  15


Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val Val Asn Val Gly
            20                  25                  30


Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys His Asn Asp
        35                  40                  45


Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg Gly Ser Ala
    50                  55                  60


Tyr Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val Lys Tyr Ser Gly
65                  70                  75                  80


Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg Val Val Tyr
                85                  90                  95


Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr Leu Thr Pro
            100                 105                 110


Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser Leu Ile Ala
        115                 120                 125
```

```
Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp Asp Phe Gln
    130                 135                 140

Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val Pro Thr Cys
145                 150                 155                 160

Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu Pro Asp Tyr Pro Gly
                165                 170                 175

Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala Lys Ser Gln Asn Leu
            180                 185                 190

Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp Ala Gly Asn Ser Ile Phe
        195                 200                 205

Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly Val Gly Val Gln Leu
    210                 215                 220

Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn Thr Val Ser Leu Gly
225                 230                 235                 240

Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu Thr Ala Asn Tyr Ala
                245                 250                 255

Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val Gln Ser Ile Ile Gly
            260                 265                 270

Val Thr Phe Val Tyr Gln Pro Gly Asp Gly Asn Ala Asp Val Thr Ile
        275                 280                 285

Thr Val Asn Gly Lys Val Val Ala Lys Gly Ser Gly Gly Gly Gly Met
    290                 295                 300

Lys Met Glu Glu Leu Phe Lys Lys His Lys Ile Val Ala Val Leu Arg
305                 310                 315                 320

Ala Asn Ser Val Glu Glu Ala Lys Lys Lys Ala Leu Ala Val Phe Leu
                325                 330                 335
```

```
Gly Gly Val His Leu Ile Glu Ile Thr Phe Thr Val Pro Asp Ala Asp
            340                 345                 350

Thr Val Ile Lys Glu Leu Ser Phe Leu Lys Glu Met Gly Ala Ile Ile
        355                 360                 365

Gly Ala Gly Thr Val Thr Ser Val Glu Gln Ala Arg Lys Ala Val Glu
    370                 375                 380

Ser Gly Ala Glu Phe Ile Val Ser Pro His Leu Asp Glu Glu Ile Ser
385                 390                 395                 400

Gln Phe Ala Lys Glu Lys Gly Val Phe Tyr Met Pro Gly Val Met Thr
                405                 410                 415

Pro Thr Glu Leu Val Lys Ala Met Lys Leu Gly His Thr Ile Leu Lys
            420                 425                 430

Leu Phe Pro Gly Glu Val Val Gly Pro Gln Phe Val Lys Ala Met Lys
        435                 440                 445

Gly Pro Phe Pro Asn Val Lys Phe Val Pro Thr Gly Gly Val Asn Leu
    450                 455                 460

Asp Asn Val Cys Glu Trp Phe Lys Ala Gly Val Leu Ala Val Gly Val
465                 470                 475                 480

Gly Ser Ala Leu Val Lys Gly Thr Pro Val Glu Val Ala Glu Lys Ala
                485                 490                 495

Lys Ala Phe Val Glu Lys Ile Arg Gly Cys Thr Glu Gly Ser Gly Ser
            500                 505                 510

Gly Ser Gly Ser Gly Ser His His His His His His
        515                 520
```

<210> 28
<211> 343
<212> PRT
<213> Artificial

<220>
<223> FimHL-GSG4- Ferritin

<400> 28

```
Met Gly Ser Ser His His His His His His Glu Asn Leu Tyr Phe Gln
1               5                  10                  15


Gly Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro Ile Gly Gly
            20                  25                  30


Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Ala Val Asn Val Gly
        35                  40                  45


Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys His Asn Asp
    50                  55                  60


Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg Gly Ser Ala
65                  70                  75                  80


Tyr Gly Gly Val Leu Ser Ser Phe Ser Gly Thr Val Lys Tyr Asn Gly
                85                  90                  95


Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg Val Val Tyr
            100                 105                 110


Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr Leu Thr Pro
        115                 120                 125


Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser Leu Ile Ala
    130                 135                 140


Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp Asp Phe Gln
145                 150                 155                 160
```

```
Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val Pro Thr Gly
                165                 170                 175

Ser Gly Gly Gly Gly Asp Ile Ile Lys Leu Leu Asn Glu Gln Val Asn
            180                 185                 190

Lys Glu Met Asn Ser Ser Asn Leu Tyr Met Ser Met Ser Ser Trp Cys
        195                 200                 205

Tyr Thr His Ser Leu Asp Gly Ala Gly Leu Phe Leu Phe Asp His Ala
    210                 215                 220

Ala Glu Glu Tyr Glu His Ala Lys Lys Leu Ile Ile Phe Leu Asn Glu
225                 230                 235                 240

Asn Asn Val Pro Val Gln Leu Thr Ser Ile Ser Ala Pro Glu His Lys
                245                 250                 255

Phe Glu Gly Leu Thr Gln Ile Phe Gln Lys Ala Tyr Glu His Glu Gln
            260                 265                 270

His Ile Ser Glu Ser Ile Asn Asn Ile Val Asp His Ala Ile Lys Ser
        275                 280                 285

Lys Asp His Ala Thr Phe Asn Phe Leu Gln Trp Tyr Val Ala Glu Gln
    290                 295                 300

His Glu Glu Glu Val Leu Phe Lys Asp Ile Leu Asp Lys Ile Glu Leu
305                 310                 315                 320

Ile Gly Asn Glu Asn His Gly Leu Tyr Leu Ala Asp Gln Tyr Val Lys
                325                 330                 335

Gly Ile Ala Lys Ser Arg Lys
            340
```

<210> 29
<211> 404

<212> PRT
<213> Artificial

<220>
<223> pelBLS-FimHL-mI3

<400> 29

```
Met Lys Tyr Leu Leu Pro Thr Ala Ala Ala Gly Leu Leu Leu Leu Ala
1               5                  10                  15


Ala Gln Pro Ala Met Ala Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala
            20                  25                  30


Ile Pro Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro
        35                  40                  45


Val Val Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile
    50                  55                  60


Phe Cys His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu
65                  70                  75                  80


Gln Arg Gly Ser Ala Tyr Gly Gly Val Leu Ser Asn Phe Ser Gly Thr
                85                  90                  95


Val Lys Tyr Ser Gly Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr
            100                 105                 110


Pro Arg Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala
        115                 120                 125


Leu Tyr Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala
    130                 135                 140


Gly Ser Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn
145                 150                 155                 160


Ser Asp Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val
                165                 170                 175
```

```
Val Val Pro Thr Gly Gly Gly Ser Gly Met Lys Met Glu Glu Leu Phe
            180                 185                 190

Lys Lys His Lys Ile Val Ala Val Leu Arg Ala Asn Ser Val Glu Glu
        195                 200                 205

Ala Lys Lys Lys Ala Leu Ala Val Phe Leu Gly Gly Val His Leu Ile
    210                 215                 220

Glu Ile Thr Phe Thr Val Pro Asp Ala Asp Thr Val Ile Lys Glu Leu
225                 230                 235                 240

Ser Phe Leu Lys Glu Met Gly Ala Ile Ile Gly Ala Gly Thr Val Thr
                245                 250                 255

Ser Val Glu Gln Ala Arg Lys Ala Val Glu Ser Gly Ala Glu Phe Ile
            260                 265                 270

Val Ser Pro His Leu Asp Glu Glu Ile Ser Gln Phe Ala Lys Glu Lys
        275                 280                 285

Gly Val Phe Tyr Met Pro Gly Val Met Thr Pro Thr Glu Leu Val Lys
    290                 295                 300

Ala Met Lys Leu Gly His Thr Ile Leu Lys Leu Phe Pro Gly Glu Val
305                 310                 315                 320

Val Gly Pro Gln Phe Val Lys Ala Met Lys Gly Pro Phe Pro Asn Val
                325                 330                 335

Lys Phe Val Pro Thr Gly Gly Val Asn Leu Asp Asn Val Cys Glu Trp
            340                 345                 350

Phe Lys Ala Gly Val Leu Ala Val Gly Val Gly Ser Ala Leu Val Lys
        355                 360                 365
```

```
Gly Thr Pro Val Glu Val Ala Glu Lys Ala Lys Ala Phe Val Glu Lys
    370                 375                 380

Ile Arg Gly Cys Thr Glu Gly Ser Gly Ser Gly Ser Gly Ser His His
385                 390                 395                 400

His His His His
```

<210> 30
<211> 331
<212> PRT
<213> Artificial

<220>
<223> FimH_DG_Ferritin (GSGGGG)

<400> 30

```
Met Gly Ser Ser His His His His His His Glu Asn Leu Tyr Phe Gln
1               5                   10                  15

Gly Asp Val Val Val Pro Thr Gly Gly Cys Asp Val Ser Ala Arg Asp
            20                  25                  30

Val Thr Val Thr Leu Pro Asp Tyr Pro Gly Ser Val Pro Ile Pro Leu
        35                  40                  45

Thr Val Tyr Cys Ala Lys Ser Gln Asn Leu Gly Tyr Tyr Leu Ser Gly
    50                  55                  60

Thr Thr Ala Asp Ala Gly Asn Ser Ile Phe Thr Asn Thr Ala Ser Phe
65                  70                  75                  80

Ser Pro Ala Gln Gly Val Gly Val Gln Leu Thr Arg Asn Gly Thr Ile
                85                  90                  95

Ile Pro Ala Asn Asn Thr Val Ser Leu Gly Ala Val Gly Thr Ser Ala
            100                 105                 110
```

```
Val Ser Leu Gly Leu Thr Ala Asn Tyr Ala Arg Thr Gly Gly Gln Val
        115                 120                 125

Thr Ala Gly Asn Val Gln Ser Ile Ile Gly Val Thr Phe Val Tyr Gln
    130                 135                 140

Pro Gly Asp Gly Asn Ala Asp Val Thr Ile Thr Val Asn Gly Lys Val
145                 150                 155                 160

Val Ala Lys Gly Ser Gly Gly Gly Gly Asp Ile Ile Lys Leu Leu Asn
                165                 170                 175

Glu Gln Val Asn Lys Glu Met Asn Ser Ser Asn Leu Tyr Met Ser Met
            180                 185                 190

Ser Ser Trp Cys Tyr Thr His Ser Leu Asp Gly Ala Gly Leu Phe Leu
        195                 200                 205

Phe Asp His Ala Ala Glu Glu Tyr Glu His Ala Lys Lys Leu Ile Ile
    210                 215                 220

Phe Leu Asn Glu Asn Asn Val Pro Val Gln Leu Thr Ser Ile Ser Ala
225                 230                 235                 240

Pro Glu His Lys Phe Glu Gly Leu Thr Gln Ile Phe Gln Lys Ala Tyr
                245                 250                 255

Glu His Glu Gln His Ile Ser Glu Ser Ile Asn Asn Ile Val Asp His
            260                 265                 270

Ala Ile Lys Ser Lys Asp His Ala Thr Phe Asn Phe Leu Gln Trp Tyr
        275                 280                 285

Val Ala Glu Gln His Glu Glu Glu Val Leu Phe Lys Asp Ile Leu Asp
    290                 295                 300

Lys Ile Glu Leu Ile Gly Asn Glu Asn His Gly Leu Tyr Leu Ala Asp
305                 310                 315                 320
```

Gln Tyr Val Lys Gly Ile Ala Lys Ser Arg Lys
325 330

<210> 31
<211> 394
<212> PRT
<213> Artificial

<220>
<223> FimHL-C-C- MI3

<400> 31

Met Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro Ile Gly Gly
1 5 10 15

Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val Cys Asn Val Gly
20 25 30

Gln Asn Cys Val Val Asp Leu Ser Thr Gln Ile Phe Cys His Asn Asp
35 40 45

Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg Gly Ser Ala
50 55 60

Tyr Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val Lys Tyr Ser Gly
65 70 75 80

Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg Val Val Tyr
85 90 95

Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr Leu Thr Pro
100 105 110

Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser Leu Ile Ala
115 120 125

Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp Asp Phe Gln
130 135 140

```
Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val Pro Thr Gly
145                 150                 155                 160

Gly Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Met Lys Met
                165                 170                 175

Glu Glu Leu Phe Lys Lys His Lys Ile Val Ala Val Leu Arg Ala Asn
            180                 185                 190

Ser Val Glu Glu Ala Lys Lys Lys Ala Leu Ala Val Phe Leu Gly Gly
        195                 200                 205

Val His Leu Ile Glu Ile Thr Phe Thr Val Pro Asp Ala Asp Thr Val
    210                 215                 220

Ile Lys Glu Leu Ser Phe Leu Lys Glu Met Gly Ala Ile Ile Gly Ala
225                 230                 235                 240

Gly Thr Val Thr Ser Val Glu Gln Ala Arg Lys Ala Val Glu Ser Gly
                245                 250                 255

Ala Glu Phe Ile Val Ser Pro His Leu Asp Glu Glu Ile Ser Gln Phe
            260                 265                 270

Ala Lys Glu Lys Gly Val Phe Tyr Met Pro Gly Val Met Thr Pro Thr
        275                 280                 285

Glu Leu Val Lys Ala Met Lys Leu Gly His Thr Ile Leu Lys Leu Phe
    290                 295                 300

Pro Gly Glu Val Val Gly Pro Gln Phe Val Lys Ala Met Lys Gly Pro
305                 310                 315                 320

Phe Pro Asn Val Lys Phe Val Pro Thr Gly Gly Val Asn Leu Asp Asn
                325                 330                 335
```

```
Val Cys Glu Trp Phe Lys Ala Gly Val Leu Ala Val Gly Val Gly Ser
            340                 345                 350


Ala Leu Val Lys Gly Thr Pro Val Glu Val Ala Glu Lys Ala Lys Ala
        355                 360                 365


Phe Val Glu Lys Ile Arg Gly Cys Thr Glu Gly Ser Gly Ser Gly Ser
    370                 375                 380


Gly Ser Gly Ser His His His His His His
385                 390


<210>  32
<211>  305
<212>  PRT
<213>  Artificial

<220>
<223>  FimHL-C-C-qBeta

<400>  32

Met Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro Ile Gly Gly
1               5                   10                  15


Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val Cys Asn Val Gly
            20                  25                  30


Gln Asn Cys Val Val Asp Leu Ser Thr Gln Ile Phe Cys His Asn Asp
        35                  40                  45


Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg Gly Ser Ala
    50                  55                  60


Tyr Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val Lys Tyr Ser Gly
65                  70                  75                  80


Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg Val Val Tyr
                85                  90                  95
```

```
Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr Leu Thr Pro
            100                 105                 110

Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser Leu Ile Ala
        115                 120                 125

Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp Asp Phe Gln
    130                 135                 140

Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val Pro Thr Gly
145                 150                 155                 160

Gly Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Ala Lys Leu
                165                 170                 175

Glu Thr Val Thr Leu Gly Asn Ile Gly Lys Asp Gly Lys Gln Thr Leu
            180                 185                 190

Val Leu Asn Pro Arg Gly Val Asn Pro Thr Asn Gly Val Ala Ser Leu
        195                 200                 205

Ser Gln Ala Gly Ala Val Pro Ala Leu Glu Lys Arg Val Thr Val Ser
    210                 215                 220

Val Ser Gln Pro Ser Arg Asn Arg Lys Asn Tyr Lys Val Gln Val Lys
225                 230                 235                 240

Ile Gln Asn Pro Thr Ala Cys Thr Ala Asn Gly Ser Cys Asp Pro Ser
                245                 250                 255

Val Thr Arg Gln Ala Tyr Ala Asp Val Thr Phe Ser Phe Thr Gln Tyr
            260                 265                 270

Ser Thr Asp Glu Glu Arg Ala Phe Val Arg Thr Glu Leu Ala Ala Leu
        275                 280                 285

Leu Ala Ser Pro Leu Leu Ile Asp Ala Ile Asp Gln Leu Asn Pro Ala
    290                 295                 300
```

Tyr
305

<210> 33
<211> 401
<212> PRT
<213> Artificial

<220>
<223> FIMHL-HIS-Mi3 J96

<400> 33

Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1   5   10   15

Gly Ser Thr Gly Asp Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile
            20   25   30

Pro Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val
        35   40   45

Val Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe
    50   55   60

Cys His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln
65   70   75   80

Arg Gly Ser Ala Tyr Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val
                85   90   95

Lys Tyr Ser Gly Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro
            100   105   110

Arg Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu
        115   120   125

Tyr Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly
    130   135   140

```
Ser Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser
145                 150                 155                 160

Asp Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val
                165                 170                 175

Val Pro Thr Gly Gly Gly Ser Gly Ser His His His His His His His
            180                 185                 190

His Gly Gly Ser Met Lys Met Glu Glu Leu Phe Lys Lys His Lys Ile
        195                 200                 205

Val Ala Val Leu Arg Ala Asn Ser Val Glu Glu Ala Lys Lys Lys Ala
    210                 215                 220

Leu Ala Val Phe Leu Gly Gly Val His Leu Ile Glu Ile Thr Phe Thr
225                 230                 235                 240

Val Pro Asp Ala Asp Thr Val Ile Lys Glu Leu Ser Phe Leu Lys Glu
                245                 250                 255

Met Gly Ala Ile Ile Gly Ala Gly Thr Val Thr Ser Val Glu Gln Ala
            260                 265                 270

Arg Lys Ala Val Glu Ser Gly Ala Glu Phe Ile Val Ser Pro His Leu
        275                 280                 285

Asp Glu Glu Ile Ser Gln Phe Ala Lys Glu Lys Gly Val Phe Tyr Met
    290                 295                 300

Pro Gly Val Met Thr Pro Thr Glu Leu Val Lys Ala Met Lys Leu Gly
305                 310                 315                 320

His Thr Ile Leu Lys Leu Phe Pro Gly Glu Val Val Gly Pro Gln Phe
                325                 330                 335
```

```
Val Lys Ala Met Lys Gly Pro Phe Pro Asn Val Lys Phe Val Pro Thr
            340                 345                 350

Gly Gly Val Asn Leu Asp Asn Val Cys Glu Trp Phe Lys Ala Gly Val
        355                 360                 365

Leu Ala Val Gly Val Gly Ser Ala Leu Val Lys Gly Thr Pro Val Glu
    370                 375                 380

Val Ala Glu Lys Ala Lys Ala Phe Val Glu Lys Ile Arg Gly Cys Thr
385                 390                 395                 400

Glu
```

<210> 34
<211> 358
<212> PRT
<213> Artificial

<220>
<223> FIMHL-HIS-Fer 536

<400> 34

```
Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1               5                   10                  15

Gly Ser Thr Gly Asp Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile
            20                  25                  30

Pro Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Ala
        35                  40                  45

Val Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe
    50                  55                  60

Cys His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln
65                  70                  75                  80
```

```
Arg Gly Ser Ala Tyr Gly Gly Val Leu Ser Ser Phe Ser Gly Thr Val
                85                  90                  95

Lys Tyr Asn Gly Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro
            100                 105                 110

Arg Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu
        115                 120                 125

Tyr Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly
    130                 135                 140

Ser Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser
145                 150                 155                 160

Asp Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val
                165                 170                 175

Val Pro Thr Gly Gly Gly Ser Gly Ser His His His His His His His
            180                 185                 190

His Gly Gly Ser Asp Ile Ile Lys Leu Leu Asn Glu Gln Val Asn Lys
        195                 200                 205

Glu Met Asn Ser Ser Asn Leu Tyr Met Ser Met Ser Ser Trp Cys Tyr
    210                 215                 220

Thr His Ser Leu Asp Gly Ala Gly Leu Phe Leu Phe Asp His Ala Ala
225                 230                 235                 240

Glu Glu Tyr Glu His Ala Lys Lys Leu Ile Ile Phe Leu Asn Glu Asn
                245                 250                 255

Asn Val Pro Val Gln Leu Thr Ser Ile Ser Ala Pro Glu His Lys Phe
            260                 265                 270

Glu Gly Leu Thr Gln Ile Phe Gln Lys Ala Tyr Glu His Glu Gln His
        275                 280                 285
```

```
Ile Ser Glu Ser Ile Asn Asn Ile Val Asp His Ala Ile Lys Ser Lys
    290                 295                 300

Asp His Ala Thr Phe Asn Phe Leu Gln Trp Tyr Val Ala Glu Gln His
305                 310                 315                 320

Glu Glu Glu Val Leu Phe Lys Asp Ile Leu Asp Lys Ile Glu Leu Ile
                325                 330                 335

Gly Asn Glu Asn His Gly Leu Tyr Leu Ala Asp Gln Tyr Val Lys Gly
            340                 345                 350

Ile Ala Lys Ser Arg Lys
        355
```

<210> 35
<211> 395
<212> PRT
<213> Artificial

<220>
<223> FIMH_DG_PGDGN_IMX313_HIS J96

<400> 35

```
Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1               5                   10                  15

Gly Ser Thr Gly Asp Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile
            20                  25                  30

Pro Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val
        35                  40                  45

Val Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe
    50                  55                  60

Cys His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln
65                  70                  75                  80
```

```
Arg Gly Ser Ala Tyr Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val
                85                  90                  95

Lys Tyr Ser Gly Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro
            100                 105                 110

Arg Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu
        115                 120                 125

Tyr Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly
    130                 135                 140

Ser Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser
145                 150                 155                 160

Asp Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val
                165                 170                 175

Val Pro Thr Gly Gly Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr
            180                 185                 190

Leu Pro Asp Tyr Pro Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys
        195                 200                 205

Ala Lys Ser Gln Asn Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp
    210                 215                 220

Ala Gly Asn Ser Ile Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln
225                 230                 235                 240

Gly Val Gly Val Gln Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn
                245                 250                 255

Asn Thr Val Ser Leu Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly
            260                 265                 270
```

```
Leu Thr Ala Asn Tyr Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn
        275                 280                 285


Val Gln Ser Ile Ile Gly Val Thr Phe Val Tyr Gln Pro Gly Asp Gly
    290                 295                 300


Asn Ala Asp Val Thr Ile Thr Val Asn Gly Lys Val Val Ala Lys Gly
305                 310                 315                 320


Ser Ser Gly Ser Gly Ser Gly Ser Lys Lys Gln Gly Asp Ala Asp Val
                325                 330                 335


Cys Gly Glu Val Ala Tyr Ile Gln Ser Val Val Ser Asp Cys His Val
            340                 345                 350


Pro Thr Ala Glu Leu Arg Thr Leu Leu Glu Ile Arg Lys Leu Phe Leu
        355                 360                 365


Glu Ile Gln Lys Leu Lys Val Glu Leu Gln Gly Leu Ser Lys Glu Gly
    370                 375                 380


Gly Gly Ser Gly Ser His His His His His His
385                 390                 395


<210>  36
<211>  495
<212>  PRT
<213>  Artificial

<220>
<223>  FimH_PGDGN_DG_HIS-Ferritn j96

<400>  36

Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1               5                   10                  15


Gly Ser Thr Gly Asp Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile
            20                  25                  30
```

```
Pro Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val
        35                  40                  45

Val Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe
    50                  55                  60

Cys His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln
65                  70                  75                  80

Arg Gly Ser Ala Tyr Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val
                85                  90                  95

Lys Tyr Ser Gly Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro
            100                 105                 110

Arg Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu
        115                 120                 125

Tyr Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly
    130                 135                 140

Ser Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser
145                 150                 155                 160

Asp Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val
                165                 170                 175

Val Pro Thr Gly Gly Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr
            180                 185                 190

Leu Pro Asp Tyr Pro Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys
        195                 200                 205

Ala Lys Ser Gln Asn Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp
    210                 215                 220

Ala Gly Asn Ser Ile Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln
225                 230                 235                 240
```

```
Gly Val Gly Val Gln Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn
                245                 250                 255

Asn Thr Val Ser Leu Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly
            260                 265                 270

Leu Thr Ala Asn Tyr Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn
        275                 280                 285

Val Gln Ser Ile Ile Gly Val Thr Phe Val Tyr Gln Pro Gly Asp Gly
    290                 295                 300

Asn Ala Asp Val Thr Ile Thr Val Asn Gly Lys Val Val Ala Lys Ser
305                 310                 315                 320

Gly Ser His His His His His His His His Gly Gly Ser Asp Ile Ile
                325                 330                 335

Lys Leu Leu Asn Glu Gln Val Asn Lys Glu Met Asn Ser Ser Asn Leu
            340                 345                 350

Tyr Met Ser Met Ser Ser Trp Cys Tyr Thr His Ser Leu Asp Gly Ala
        355                 360                 365

Gly Leu Phe Leu Phe Asp His Ala Ala Glu Glu Tyr Glu His Ala Lys
    370                 375                 380

Lys Leu Ile Ile Phe Leu Asn Glu Asn Asn Val Pro Val Gln Leu Thr
385                 390                 395                 400

Ser Ile Ser Ala Pro Glu His Lys Phe Glu Gly Leu Thr Gln Ile Phe
                405                 410                 415

Gln Lys Ala Tyr Glu His Glu Gln His Ile Ser Glu Ser Ile Asn Asn
            420                 425                 430
```

```
Ile Val Asp His Ala Ile Lys Ser Lys Asp His Ala Thr Phe Asn Phe
        435                 440                 445


Leu Gln Trp Tyr Val Ala Glu Gln His Glu Glu Glu Val Leu Phe Lys
    450                 455                 460


Asp Ile Leu Asp Lys Ile Glu Leu Ile Gly Asn Glu Asn His Gly Leu
465                 470                 475                 480


Tyr Leu Ala Asp Gln Tyr Val Lys Gly Ile Ala Lys Ser Arg Lys
                485                 490                 495


<210>  37
<211>  495
<212>  PRT
<213>  Artificial

<220>
<223>  FIMH_DG_PGDGN-HIS-Ferritin 536

<400>  37

Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1               5                   10                  15


Gly Ser Thr Gly Asp Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile
            20                  25                  30


Pro Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Ala
        35                  40                  45


Val Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe
    50                  55                  60


Cys His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln
65                  70                  75                  80


Arg Gly Ser Ala Tyr Gly Gly Val Leu Ser Ser Phe Ser Gly Thr Val
                85                  90                  95
```

```
Lys Tyr Asn Gly Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro
            100                 105                 110

Arg Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu
        115                 120                 125

Tyr Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly
    130                 135                 140

Ser Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser
145                 150                 155                 160

Asp Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val
                165                 170                 175

Val Pro Thr Gly Gly Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr
            180                 185                 190

Leu Pro Asp Tyr Pro Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys
        195                 200                 205

Ala Lys Ser Gln Asn Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp
    210                 215                 220

Ala Gly Asn Ser Ile Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln
225                 230                 235                 240

Gly Val Gly Val Gln Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn
                245                 250                 255

Asn Thr Val Ser Leu Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly
            260                 265                 270

Leu Thr Ala Asn Tyr Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn
        275                 280                 285

Val Gln Ser Ile Ile Gly Val Thr Phe Val Tyr Gln Pro Gly Asp Gly
    290                 295                 300
```

```
Asn Ala Asp Val Thr Ile Thr Val Asn Gly Lys Val Val Ala Lys Ser
305                 310                 315                 320

Gly Ser His His His His His His His His Gly Gly Ser Asp Ile Ile
                325                 330                 335

Lys Leu Leu Asn Glu Gln Val Asn Lys Glu Met Asn Ser Ser Asn Leu
            340                 345                 350

Tyr Met Ser Met Ser Ser Trp Cys Tyr Thr His Ser Leu Asp Gly Ala
        355                 360                 365

Gly Leu Phe Leu Phe Asp His Ala Ala Glu Glu Tyr Glu His Ala Lys
    370                 375                 380

Lys Leu Ile Ile Phe Leu Asn Glu Asn Asn Val Pro Val Gln Leu Thr
385                 390                 395                 400

Ser Ile Ser Ala Pro Glu His Lys Phe Glu Gly Leu Thr Gln Ile Phe
                405                 410                 415

Gln Lys Ala Tyr Glu His Glu Gln His Ile Ser Glu Ser Ile Asn Asn
            420                 425                 430

Ile Val Asp His Ala Ile Lys Ser Lys Asp His Ala Thr Phe Asn Phe
        435                 440                 445

Leu Gln Trp Tyr Val Ala Glu Gln His Glu Glu Glu Val Leu Phe Lys
    450                 455                 460

Asp Ile Leu Asp Lys Ile Glu Leu Ile Gly Asn Glu Asn His Gly Leu
465                 470                 475                 480

Tyr Leu Ala Asp Gln Tyr Val Lys Gly Ile Ala Lys Ser Arg Lys
                485                 490                 495
```

<210> 38
<211> 510
<212> PRT
<213> Artificial

<220>
<223> FimH_PGDGN_DG-HIS-Ferritin J96

<400> 38

```
Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1               5                  10                  15


Gly Ser Thr Gly Asp Ala Ala Gln Pro Ala Arg Arg Ala Arg Arg Thr
            20                  25                  30


Lys Leu Ala Leu Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro
        35                  40                  45


Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val Val
    50                  55                  60


Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys
65                  70                  75                  80


His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg
                85                  90                  95


Gly Ser Ala Tyr Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val Lys
            100                 105                 110


Tyr Ser Gly Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg
        115                 120                 125


Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr
    130                 135                 140


Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser
145                 150                 155                 160
```

```
Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp
                165                 170                 175

Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val
            180                 185                 190

Pro Thr Gly Gly Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu
        195                 200                 205

Pro Asp Tyr Pro Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala
    210                 215                 220

Lys Ser Gln Asn Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp Ala
225                 230                 235                 240

Gly Asn Ser Ile Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly
                245                 250                 255

Val Gly Val Gln Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn
            260                 265                 270

Thr Val Ser Leu Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu
        275                 280                 285

Thr Ala Asn Tyr Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val
    290                 295                 300

Gln Ser Ile Ile Gly Val Thr Phe Val Tyr Gln Pro Gly Asp Gly Asn
305                 310                 315                 320

Ala Asp Val Thr Ile Thr Val Asn Gly Lys Val Val Ala Lys Ser Gly
                325                 330                 335

Ser His His His His His His His His Gly Gly Ser Asp Ile Ile Lys
            340                 345                 350

Leu Leu Asn Glu Gln Val Asn Lys Glu Met Asn Ser Ser Asn Leu Tyr
        355                 360                 365
```

```
Met Ser Met Ser Ser Trp Cys Tyr Thr His Ser Leu Asp Gly Ala Gly
    370                 375                 380

Leu Phe Leu Phe Asp His Ala Ala Glu Glu Tyr Glu His Ala Lys Lys
385                 390                 395                 400

Leu Ile Ile Phe Leu Asn Glu Asn Asn Val Pro Val Gln Leu Thr Ser
                405                 410                 415

Ile Ser Ala Pro Glu His Lys Phe Glu Gly Leu Thr Gln Ile Phe Gln
            420                 425                 430

Lys Ala Tyr Glu His Glu Gln His Ile Ser Glu Ser Ile Asn Asn Ile
        435                 440                 445

Val Asp His Ala Ile Lys Ser Lys Asp His Ala Thr Phe Asn Phe Leu
    450                 455                 460

Gln Trp Tyr Val Ala Glu Gln His Glu Glu Glu Val Leu Phe Lys Asp
465                 470                 475                 480

Ile Leu Asp Lys Ile Glu Leu Ile Gly Asn Glu Asn His Gly Leu Tyr
                485                 490                 495

Leu Ala Asp Gln Tyr Val Lys Gly Ile Ala Lys Ser Arg Lys
            500                 505                 510
```

<210> 39
<211> 553
<212> PRT
<213> Artificial

<220>
<223> FimH_PGDGN_DG-HIS-MI3 j96

<400> 39

```
Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1               5                   10                  15
```

```
Gly Ser Thr Gly Asp Ala Ala Gln Pro Ala Arg Arg Ala Arg Arg Thr
            20                  25                  30

Lys Leu Ala Leu Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro
        35                  40                  45

Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val Val
    50                  55                  60

Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys
65                  70                  75                  80

His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg
                85                  90                  95

Gly Ser Ala Tyr Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val Lys
            100                 105                 110

Tyr Ser Gly Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg
        115                 120                 125

Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr
    130                 135                 140

Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser
145                 150                 155                 160

Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp
                165                 170                 175

Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val
            180                 185                 190

Pro Thr Gly Gly Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu
        195                 200                 205
```

```
Pro Asp Tyr Pro Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala
    210                 215                 220

Lys Ser Gln Asn Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp Ala
225                 230                 235                 240

Gly Asn Ser Ile Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly
                245                 250                 255

Val Gly Val Gln Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn
            260                 265                 270

Thr Val Ser Leu Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu
        275                 280                 285

Thr Ala Asn Tyr Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val
    290                 295                 300

Gln Ser Ile Ile Gly Val Thr Phe Val Tyr Gln Pro Gly Asp Gly Asn
305                 310                 315                 320

Ala Asp Val Thr Ile Thr Val Asn Gly Lys Val Val Ala Lys Ser Gly
                325                 330                 335

Ser His His His His His His His His Gly Gly Ser Met Lys Met Glu
            340                 345                 350

Glu Leu Phe Lys Lys His Lys Ile Val Ala Val Leu Arg Ala Asn Ser
        355                 360                 365

Val Glu Glu Ala Lys Lys Lys Ala Leu Ala Val Phe Leu Gly Gly Val
    370                 375                 380

His Leu Ile Glu Ile Thr Phe Thr Val Pro Asp Ala Asp Thr Val Ile
385                 390                 395                 400

Lys Glu Leu Ser Phe Leu Lys Glu Met Gly Ala Ile Ile Gly Ala Gly
                405                 410                 415
```

```
Thr Val Thr Ser Val Glu Gln Ala Arg Lys Ala Val Glu Ser Gly Ala
            420                 425                 430

Glu Phe Ile Val Ser Pro His Leu Asp Glu Glu Ile Ser Gln Phe Ala
        435                 440                 445

Lys Glu Lys Gly Val Phe Tyr Met Pro Gly Val Met Thr Pro Thr Glu
    450                 455                 460

Leu Val Lys Ala Met Lys Leu Gly His Thr Ile Leu Lys Leu Phe Pro
465                 470                 475                 480

Gly Glu Val Val Gly Pro Gln Phe Val Lys Ala Met Lys Gly Pro Phe
                485                 490                 495

Pro Asn Val Lys Phe Val Pro Thr Gly Gly Val Asn Leu Asp Asn Val
            500                 505                 510

Cys Glu Trp Phe Lys Ala Gly Val Leu Ala Val Gly Val Gly Ser Ala
        515                 520                 525

Leu Val Lys Gly Thr Pro Val Glu Val Ala Glu Lys Ala Lys Ala Phe
    530                 535                 540

Val Glu Lys Ile Arg Gly Cys Thr Glu
545                 550
```

<210> 40
<211> 410
<212> PRT
<213> Artificial

<220>
<223> FimH_PGDGN_DG-HIS-IMX313 j96

<400> 40

```
Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1               5                   10                  15
```

```
Gly Ser Thr Gly Asp Ala Ala Gln Pro Ala Arg Arg Ala Arg Arg Thr
            20                  25                  30

Lys Leu Ala Leu Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro
        35                  40                  45

Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val Val
    50                  55                  60

Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys
65                  70                  75                  80

His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg
                85                  90                  95

Gly Ser Ala Tyr Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val Lys
            100                 105                 110

Tyr Ser Gly Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg
        115                 120                 125

Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr
    130                 135                 140

Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser
145                 150                 155                 160

Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp
                165                 170                 175

Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val
            180                 185                 190

Pro Thr Gly Gly Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu
        195                 200                 205
```

```
Pro Asp Tyr Pro Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala
    210                 215                 220

Lys Ser Gln Asn Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp Ala
225                 230                 235                 240

Gly Asn Ser Ile Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly
                245                 250                 255

Val Gly Val Gln Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn
            260                 265                 270

Thr Val Ser Leu Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu
        275                 280                 285

Thr Ala Asn Tyr Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val
    290                 295                 300

Gln Ser Ile Ile Gly Val Thr Phe Val Tyr Gln Pro Gly Asp Gly Asn
305                 310                 315                 320

Ala Asp Val Thr Ile Thr Val Asn Gly Lys Val Val Ala Lys Gly Ser
                325                 330                 335

Ser Gly Ser Gly Ser Gly Ser Lys Lys Gln Gly Asp Ala Asp Val Cys
            340                 345                 350

Gly Glu Val Ala Tyr Ile Gln Ser Val Val Ser Asp Cys His Val Pro
        355                 360                 365

Thr Ala Glu Leu Arg Thr Leu Leu Glu Ile Arg Lys Leu Phe Leu Glu
    370                 375                 380

Ile Gln Lys Leu Lys Val Glu Leu Gln Gly Leu Ser Lys Glu Gly Gly
385                 390                 395                 400

Gly Ser Gly Ser His His His His His His
                405                 410
```

<210> 41
<211> 510
<212> PRT
<213> Artificial

<220>
<223> FimH_PGDGN_DG_Ferritin (536)

<400> 41

```
Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1               5                  10                  15


Gly Ser Thr Gly Asp Ala Ala Gln Pro Ala Arg Arg Ala Arg Arg Thr
            20                  25                  30


Lys Leu Ala Leu Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro
        35                  40                  45


Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Ala Val
    50                  55                  60


Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys
65                  70                  75                  80


His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg
                85                  90                  95


Gly Ser Ala Tyr Gly Gly Val Leu Ser Ser Phe Ser Gly Thr Val Lys
            100                 105                 110


Tyr Asn Gly Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg
        115                 120                 125


Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr
    130                 135                 140


Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser
145                 150                 155                 160
```

```
Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp
                165                 170                 175

Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val
            180                 185                 190

Pro Thr Gly Gly Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu
        195                 200                 205

Pro Asp Tyr Pro Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala
    210                 215                 220

Lys Ser Gln Asn Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp Ala
225                 230                 235                 240

Gly Asn Ser Ile Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly
                245                 250                 255

Val Gly Val Gln Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn
            260                 265                 270

Thr Val Ser Leu Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu
        275                 280                 285

Thr Ala Asn Tyr Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val
    290                 295                 300

Gln Ser Ile Ile Gly Val Thr Phe Val Tyr Gln Pro Gly Asp Gly Asn
305                 310                 315                 320

Ala Asp Val Thr Ile Thr Val Asn Gly Lys Val Val Ala Lys Ser Gly
                325                 330                 335

Ser His His His His His His His His Gly Gly Ser Asp Ile Ile Lys
            340                 345                 350
```

```
Leu Leu Asn Glu Gln Val Asn Lys Glu Met Asn Ser Ser Asn Leu Tyr
        355                 360                 365

Met Ser Met Ser Ser Trp Cys Tyr Thr His Ser Leu Asp Gly Ala Gly
    370                 375                 380

Leu Phe Leu Phe Asp His Ala Ala Glu Glu Tyr Glu His Ala Lys Lys
385                 390                 395                 400

Leu Ile Ile Phe Leu Asn Glu Asn Asn Val Pro Val Gln Leu Thr Ser
                405                 410                 415

Ile Ser Ala Pro Glu His Lys Phe Glu Gly Leu Thr Gln Ile Phe Gln
            420                 425                 430

Lys Ala Tyr Glu His Glu Gln His Ile Ser Glu Ser Ile Asn Asn Ile
        435                 440                 445

Val Asp His Ala Ile Lys Ser Lys Asp His Ala Thr Phe Asn Phe Leu
    450                 455                 460

Gln Trp Tyr Val Ala Glu Gln His Glu Glu Glu Val Leu Phe Lys Asp
465                 470                 475                 480

Ile Leu Asp Lys Ile Glu Leu Ile Gly Asn Glu Asn His Gly Leu Tyr
                485                 490                 495

Leu Ala Asp Gln Tyr Val Lys Gly Ile Ala Lys Ser Arg Lys
            500                 505                 510
```

<210> 42
<211> 389
<212> PRT
<213> Artificial

<220>
<223> FimH-IMX313 j96

<400> 42

```
Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1               5                  10                  15

Gly Ser Thr Gly Asp Ala Ala Gln Pro Ala Arg Arg Ala Arg Arg Thr
            20                  25                  30

Lys Leu Ala Leu Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro
        35                  40                  45

Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val Val
    50                  55                  60

Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys
65                  70                  75                  80

His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg
                85                  90                  95

Gly Ser Ala Tyr Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val Lys
            100                 105                 110

Tyr Ser Gly Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg
        115                 120                 125

Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr
    130                 135                 140

Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser
145                 150                 155                 160

Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp
                165                 170                 175

Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val
            180                 185                 190

Pro Thr Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu Pro Asp
        195                 200                 205
```

```
Tyr Pro Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala Lys Ser
    210                 215                 220

Gln Asn Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp Ala Gly Asn
225                 230                 235                 240

Ser Ile Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly Val Gly
                245                 250                 255

Val Gln Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn Thr Val
            260                 265                 270

Ser Leu Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu Thr Ala
        275                 280                 285

Asn Tyr Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val Gln Ser
    290                 295                 300

Ile Ile Gly Val Thr Phe Val Tyr Gln Gly Ser Ser Gly Ser Gly Ser
305                 310                 315                 320

Gly Ser Lys Lys Gln Gly Asp Ala Asp Val Cys Gly Glu Val Ala Tyr
                325                 330                 335

Ile Gln Ser Val Val Ser Asp Cys His Val Pro Thr Ala Glu Leu Arg
            340                 345                 350

Thr Leu Leu Glu Ile Arg Lys Leu Phe Leu Glu Ile Gln Lys Leu Lys
        355                 360                 365

Val Glu Leu Gln Gly Leu Ser Lys Glu Gly Gly Gly Ser Gly Ser His
    370                 375                 380

His His His His His
385
```

<210> 43
<211> 542
<212> PRT
<213> Artificial

<220>
<223> FimH_mi3 j96

<400> 43

```
Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1               5                  10                  15


Gly Ser Thr Gly Asp Ala Ala Gln Pro Ala Arg Arg Ala Arg Arg Thr
            20                  25                  30


Lys Leu Ala Leu Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro
        35                  40                  45


Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val Val
    50                  55                  60


Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys
65                  70                  75                  80


His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg
                85                  90                  95


Gly Ser Ala Tyr Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val Lys
            100                 105                 110


Tyr Ser Gly Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg
        115                 120                 125


Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr
    130                 135                 140


Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser
145                 150                 155                 160
```

```
Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp
                165                 170                 175

Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val
            180                 185                 190

Pro Thr Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu Pro Asp
        195                 200                 205

Tyr Pro Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala Lys Ser
    210                 215                 220

Gln Asn Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp Ala Gly Asn
225                 230                 235                 240

Ser Ile Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly Val Gly
                245                 250                 255

Val Gln Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn Thr Val
            260                 265                 270

Ser Leu Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu Thr Ala
        275                 280                 285

Asn Tyr Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val Gln Ser
    290                 295                 300

Ile Ile Gly Val Thr Phe Val Tyr Gln Gly Ser Gly Gly Gly Gly Met
305                 310                 315                 320

Lys Met Glu Glu Leu Phe Lys Lys His Lys Ile Val Ala Val Leu Arg
                325                 330                 335

Ala Asn Ser Val Glu Glu Ala Lys Lys Lys Ala Leu Ala Val Phe Leu
            340                 345                 350

Gly Gly Val His Leu Ile Glu Ile Thr Phe Thr Val Pro Asp Ala Asp
        355                 360                 365
```

```
Thr Val Ile Lys Glu Leu Ser Phe Leu Lys Glu Met Gly Ala Ile Ile
    370                 375                 380

Gly Ala Gly Thr Val Thr Ser Val Glu Gln Ala Arg Lys Ala Val Glu
385                 390                 395                 400

Ser Gly Ala Glu Phe Ile Val Ser Pro His Leu Asp Glu Glu Ile Ser
                405                 410                 415

Gln Phe Ala Lys Glu Lys Gly Val Phe Tyr Met Pro Gly Val Met Thr
            420                 425                 430

Pro Thr Glu Leu Val Lys Ala Met Lys Leu Gly His Thr Ile Leu Lys
        435                 440                 445

Leu Phe Pro Gly Glu Val Val Gly Pro Gln Phe Val Lys Ala Met Lys
    450                 455                 460

Gly Pro Phe Pro Asn Val Lys Phe Val Pro Thr Gly Gly Val Asn Leu
465                 470                 475                 480

Asp Asn Val Cys Glu Trp Phe Lys Ala Gly Val Leu Ala Val Gly Val
                485                 490                 495

Gly Ser Ala Leu Val Lys Gly Thr Pro Val Glu Val Ala Glu Lys Ala
            500                 505                 510

Lys Ala Phe Val Glu Lys Ile Arg Gly Cys Thr Glu Gly Ser Gly Ser
        515                 520                 525

Gly Ser Gly Ser Gly Ser His His His His His His His His
    530                 535                 540
```

<210> 44
<211> 415
<212> PRT
<213> Artificial

<220>
<223> FIMHL-HIS-mI3 j96

<400> 44

```
Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1               5                  10                  15


Gly Ser Thr Gly Asp Ala Ala Gln Pro Ala Arg Arg Ala Arg Arg Thr
            20                  25                  30


Lys Leu Ala Leu Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro
        35                  40                  45


Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val Val
    50                  55                  60


Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys
65                  70                  75                  80


His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg
                85                  90                  95


Gly Ser Ala Tyr Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val Lys
            100                 105                 110


Tyr Ser Gly Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg
        115                 120                 125


Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr
    130                 135                 140


Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser
145                 150                 155                 160


Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp
                165                 170                 175
```

```
Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val
            180                 185                 190

Pro Thr Gly Ser Gly Gly His His His His His His His His Gly Ser
        195                 200                 205

Gly Ser Met Lys Met Glu Glu Leu Phe Lys Lys His Lys Ile Val Ala
    210                 215                 220

Val Leu Arg Ala Asn Ser Val Glu Glu Ala Lys Lys Lys Ala Leu Ala
225                 230                 235                 240

Val Phe Leu Gly Gly Val His Leu Ile Glu Ile Thr Phe Thr Val Pro
                245                 250                 255

Asp Ala Asp Thr Val Ile Lys Glu Leu Ser Phe Leu Lys Glu Met Gly
            260                 265                 270

Ala Ile Ile Gly Ala Gly Thr Val Thr Ser Val Glu Gln Ala Arg Lys
        275                 280                 285

Ala Val Glu Ser Gly Ala Glu Phe Ile Val Ser Pro His Leu Asp Glu
    290                 295                 300

Glu Ile Ser Gln Phe Ala Lys Glu Lys Gly Val Phe Tyr Met Pro Gly
305                 310                 315                 320

Val Met Thr Pro Thr Glu Leu Val Lys Ala Met Lys Leu Gly His Thr
                325                 330                 335

Ile Leu Lys Leu Phe Pro Gly Glu Val Val Gly Pro Gln Phe Val Lys
            340                 345                 350

Ala Met Lys Gly Pro Phe Pro Asn Val Lys Phe Val Pro Thr Gly Gly
        355                 360                 365

Val Asn Leu Asp Asn Val Cys Glu Trp Phe Lys Ala Gly Val Leu Ala
    370                 375                 380
```

```
Val Gly Val Gly Ser Ala Leu Val Lys Gly Thr Pro Val Glu Val Ala
385                 390                 395                 400


Glu Lys Ala Lys Ala Phe Val Glu Lys Ile Arg Gly Cys Thr Glu
                405                 410                 415


<210>  45
<211>  1497
<212>  DNA
<213>  Artificial

<220>
<223>  FIMH_DG_536-1EUM_0_5

<400>  45
atggaaaccg atacactgct gctgtgggtg ctgttgctct gggttccagg ctctacaggc       60

gattttgcct gcaagaccgc caacggcaca gccattccta ttggcggagg cagcgccaat      120

gtgtacgtta acctggctcc tgccgtgaac gtgggccaga atctggtggt ggatctgagc      180

acccagatct tttgccacaa cgactacccc gagacaatca ccgactacgt gacactgcag      240

agaggctctg cttacggcgg cgtgctgtct agctttagcg gcaccgtgaa gtacaacggc      300

agcagctacc catttcctac caccagcgag acacccagag tggtgtacaa cagcagaacc      360

gacaagccct ggcctgtggc tctgtacctg acacctgtta gttctgccgg cggagtggcc      420

attaaggccg gatctctgat tgccgtgctg atcctgcggc agaccaacaa ctacaacagc      480

gacgacttcc agttcgtgtg gaacatctac gccaacaacg acgtggtggt gcctacaggc      540

ggatgtgatg tgtccgccag agatgtgaca gtgaccctgc ctgattaccc cggctctgtg      600

cctattcctc tgaccgtgta ctgcgccaag agccagaacc tgggctacta cctgtctggc      660

acaacagccg atgccggcaa cagcatcttt accaacaccg ccagcttcag ccctgctcaa      720

ggtgttggag tgcagctgac ccggaacgga acaatcatcc ccgccaacaa taccgtgtct      780

ctgggagctg tgggcacctc tgctgtgtct cttggcctga cagccaacta tgccagaaca      840

ggcggacaag tgacagccgg caatgtgcag tctatcatcg gcgtgacctt cgtgtatcag      900

cctggcgacg gaaacgccga tgtgaccatc acagtgaatg gcaaggtggt ggccaagagc      960
```

ggaagccacc accatcatca ccatcaccac ggcggctcta tgctgaagcc cgagatgatc 1020

gagaagctga acgagcagat gaacctggaa ctgtacagct ccctgctgta ccagcagatg 1080

agcgcctggt gtagctatca cggatttgag ggcgctgccg cctttctgag aaggcacgcc 1140

caagaggaaa tgacccacat gcagcggctg ttcgactacc tgaccgatac cggcaatctg 1200

cccagaatcg acacaatccc atctccattc gccgagtaca gcagcctgga cgagctgttc 1260

caagaaacct acaagcacga gcagctgatc acccagaaga tcaacgaact ggcccatgcc 1320

gccatgacca accaggacta ccctaccttc aacttcctgc agtggtacgt ggccgagcag 1380

cacgaggaag agaagctgtt caagagcatc atcgacaagc tgagcctggc cggaaagtct 1440

ggcgagggcc tgtactttat cgacaaagag ctgagcacac tggataccca gaactga 1497

<210> 46
<211> 1617
<212> DNA
<213> Artificial

<220>
<223> FIMH_DG_PGDGN_536-MI3

<400> 46
atggaaaccg atacactgct gctgtgggtg ctgttgctct gggttccagg ctctacaggc 60

gattttgcct gcaagaccgc caacggcaca gccattccta ttggcggagg cagcgccaat 120

gtgtacgtta acctggctcc tgccgtgaac gtgggccaga atctggtggt ggatctgagc 180

acccagatct tttgccacaa cgactacccc gagacaatca ccgactacgt gacactgcag 240

agaggctctg cttacggcgg cgtgctgtct agctttagcg gcaccgtgaa gtacaacggc 300

agcagctacc catttcctac caccagcgag acacccagag tggtgtacaa cagcagaacc 360

gacaagccct ggcctgtggc tctgtacctg acacctgtta gttctgccgg cggagtggcc 420

attaaggccg gatctctgat tgccgtgctg atcctgcggc agaccaacaa ctacaacagc 480

gacgacttcc agttcgtgtg gaacatctac gccaacaacg acgtggtggt gcctacaggc 540

ggatgtgatg tgtccgccag agatgtgaca gtgaccctgc ctgattaccc cggctctgtg 600

cctattcctc tgaccgtgta ctgcgccaag agccagaacc tgggctacta cctgtctggc 660

acaacagccg atgccggcaa cagcatcttt accaacaccg ccagcttcag ccctgctcaa 720

ggtgttggag tgcagctgac ccggaacgga acaatcatcc ccgccaacaa taccgtgtct 780

ctgggagctg tgggcacctc tgctgtgtct cttggcctga cagccaacta tgccagaaca 840

ggcggacaag tgacagccgg caatgtgcag tctatcatcg gcgtgacctt cgtgtatcag 900

cctggcgacg gaaacgccga tgtgaccatc acagtgaatg gcaaggtggt ggccaagagc 960

ggaagccacc accatcatca ccatcaccac ggcggcagca tgaagatgga agaactgttc 1020

aagaagcaca agatcgtcgc cgtgctgcgg gccaattctg tggaagaggc caaaaaaaag 1080

gccctggccg tgtttcttgg cggagtgcac ctgatcgaga tcacctttac cgtgcctgac 1140

gccgacaccg tgatcaaaga gctgagcttc ctgaaagaga tgggcgccat catcggagcc 1200

ggcacagtga catctgttga gcaggccaga aaggccgtgg aatctggcgc cgagtttatc 1260

gtgtcccctc acctggatga ggaaatcagc cagttcgcca aagaaaaggg cgtgttctac 1320

atgcccggcg tgatgacacc tacagagctg gtcaaagcca tgaagctggg ccacaccatc 1380

ctgaagctgt ttccaggcga agtcgtgggc cctcagttcg tgaaagctat gaagggccca 1440

tttccaaacg tgaagttcgt gcccactggc ggcgtgaacc tggataatgt gtgcgagtgg 1500

ttcaaggctg gcgtgctggc tgttggagtt ggctctgctc tggtcaaggg cacacctgtg 1560

gaagtggctg agaaggccaa ggccttcgtg gaaaagatca gaggctgcac cgagtga 1617

<210> 47
<211> 1797
<212> DNA
<213> Artificial

<220>
<223> FIMH_DG_PGDGN_536-encapsuline

<400> 47
atggaaaccg atacactgct gctgtgggtg ctgttgctct gggttccagg ctctacaggc 60

gattttgcct gcaagaccgc caacggcaca gccattccta ttggcggagg cagcgccaat 120

gtgtacgtta acctggctcc tgccgtgaac gtgggccaga atctggtggt ggatctgagc 180

acccagatct tttgccacaa cgactacccc gagacaatca ccgactacgt gacactgcag 240

agaggctctg cttacggcgg cgtgctgtct agctttagcg gcaccgtgaa gtacaacggc 300

agcagctacc catttcctac caccagcgag acacccagag tggtgtacaa cagcagaacc 360

gacaagccct ggcctgtggc tctgtacctg acacctgtta gttctgccgg cggagtggcc 420

attaaggccg gatctctgat tgccgtgctg atcctgcggc agaccaacaa ctacaacagc 480

gacgacttcc agttcgtgtg gaacatctac gccaacaacg acgtggtggt gcctacaggc 540

ggatgtgatg tgtccgccag agatgtgaca gtgaccctgc ctgattaccc cggctctgtg 600

cctattcctc tgaccgtgta ctgcgccaag agccagaacc tgggctacta cctgtctggc 660

acaacagccg atgccggcaa cagcatcttt accaacaccg ccagcttcag ccctgctcaa 720

ggtgttggag tgcagctgac ccggaacgga acaatcatcc ccgccaacaa taccgtgtct 780

ctgggagctg tgggcacctc tgctgtgtct cttggcctga cagccaacta tgccagaaca 840

ggcggacaag tgacagccgg caatgtgcag tctatcatcg gcgtgacctt cgtgtatcag 900

cctggcgacg gaaacgccga tgtgaccatc acagtgaatg gcaaggtggt ggccaagagc 960

ggaagccacc accatcatca ccatcaccac ggcggcagca tggaatttct gaagagaagc 1020

ttcgccccac tgaccgagaa gcagtggcaa gagatcgaca accgggccag agagatcttc 1080

aagacccagc tgtacggccg gaagttcgtg gatgtggaag gcccttatgg ctgggagtat 1140

gccgctcatc ctctgggcga agtggaagtg ctgagcgacg agaatgaggt cgtgaagtgg 1200

ggcctgagaa agagcctgcc tctgatcgag ctgagagcca ccttcacact ggacctgtgg 1260

gaactcgaca acctggaaag gggcaagccc aatgtggacc tgagcagcct ggaagagaca 1320

gtgcggaagg tggccgagtt cgaggacgaa gtgatcttca gaggctgcga gaagtctggc 1380

gtgaagggcc tgctgagctt cgaggaacgg aagatcgagt gtggcagcac ccctaaggat 1440

ctgctggaag ccatcgtgcg ggccctgagc atcttctcta aggatggcat cgagggcccc 1500

tacacactgg tcatcaacac cgaccggtgg atcaacttcc tgaaagagga agccggccac 1560

tatcctctgg aaaagcgcgt ggaagagtgc ctgagaggcg gcaagatcat cacaacccct 1620

agaatcgagg acgccctggt ggtttctgag agaggcggag acttcaagct gatccttggc 1680

caggacctgt ccatcggcta cgaggacaga gaaaaagacg ccgtgcggct gttcatcacc 1740

gaaaccttca ccttccaagt ggtcaacccc gaggctctga ttctgctgaa gttctga 1797

<210> 48
<211> 1053
<212> DNA
<213> Artificial

<220>
<223> HBcFIMHLJ96

<400> 48
atggaaaccg atacactgct gctgtgggtg ctgttgctct gggttccagg atctaccggc 60

gacgacatcg acccctacaa agagtttggc gccagcgtcg agctgctgag cttcctgcct 120

agcgacttct tcccttccat ccgggatctg ctggataccg ctagcgccct gtatagagag 180

gccctggaaa gccctgagca ctgctctcca catcacacag ccctgagaca ggccatcctg 240

tgttggggcg aactgatgaa tctggccacc tgggtcggaa gcaacctgga agatcctggt 300

tctggcggcg gaggctttgc ctgtaaaaca gccaatggca ccgccattcc tatcggaggc 360

ggcagcgcca atgtgtacgt taacctggct cctgtggtca acgtgggcca gaatctggtg 420

gtggacctga gcacccagat cttttgccac aacgactacc ccgagacaat caccgactac 480

gtgacactgc agagaggctc tgcttacggc ggcgtgctga gcaatttttc cggcacagtg 540

aagtacagcg gcagcagcta cccatttcct accaccagcg agacacccag agtggtgtac 600

aacagcagaa ccgacaagcc ctggcctgtg gctctgtacc tgacacctgt tagttctgct 660

ggcggagtgg ccatcaaggc cggatctctg attgccgtgc tgatcctgcg gcagaccaac 720

aactacaaca gcgacgactt ccagttcgtg tggaacatct acgccaacaa cgacgtggtg 780

gtgcctacag gcggaggatc tggcggagct tctagagaac tggtcgtgtc ctacgtgaac 840

gtgaacatgg gcctgaagat ccggcagctg ctctggtttc acatcagctg tctgaccttc 900

ggccgggaaa ccgtgctgga atacctggtg tccttcggcg tgtggatcag aacccctcct 960

gcctatagac ctcctaacgc tcccatcctg agcacactgc ctgagacaac agttgttgga 1020

agcggaggcg gaggccacca ccatcaccat cat 1053

<210> 49
<211> 1473
<212> DNA
<213> Artificial

<220>
<223> HBcFIMHDGJ96

<400> 49
atggagaccg acaccctgct gctgtgggtg ctgctgctgt gggtgcccgg cagcaccggc 60

gacgacatcg acccctacaa ggagttcggc gccagcgtgg agctgctgag cttcctgccc 120

agcgacttct tccccagcat ccgggacctg ctggacaccg ccagcgccct gtaccgggag 180

gccctggaga gccccgagca ctgcagcccc caccacaccg ccctgcggca ggccatcctg 240

tgctggggcg agctgatgaa cctggccacc tgggtgggca gcaacctgga ggaccccggc 300

agcggcggcg gcggcttcgc ctgcaagacc gccaacggca ccgccatccc catcggcggc 360

ggcagcgcca acgtgtacgt gaacctggcc cccgtggtga acgtgggcca gaacctggtg 420

gtggacctga gcacccagat cttctgccac aacgactacc ccgagaccat caccgactac 480

gtgaccctgc agcggggcag cgcctacggc ggcgtgctga gcaacttcag cggcaccgtg 540

aagtacagcg gcagcagcta ccccttcccc accaccagcg agacccccg ggtggtgtac 600

aacagccgga ccgacaagcc ctggcccgtg gccctgtacc tgaccccgt gagcagcgcc 660

ggcggcgtgg ccatcaaggc cggcagcctg atcgccgtgc tgatcctgcg gcagaccaac 720

aactacaaca gcgacgactt ccagttcgtg tggaacatct acgccaacaa cgacgtggtg 780

gtgcccaccg gcggctgcga cgtgagcgcc cgggacgtga ccgtgaccct gcccgactac 840

cccggcagcg tgcccatccc cctgaccgtg tactgcgcca agagccagaa cctgggctac 900

tacctgagcg gcaccaccgc cgacgccggc aacagcatct tcaccaacac cgccagcttc 960

agccccgccc agggcgtggg cgtgcagctg acccggaacg gcaccatcat ccccgccaac 1020

aacaccgtga gcctgggcgc cgtgggcacc agcgccgtga gcctgggcct gaccgccaac 1080

tacgcccgga ccggcggcca ggtgaccgcc ggcaacgtgc agagcatcat cggcgtgacc 1140

ttcgtgtacc agcccggcga cggcaacgcc gacgtgacca tcaccgtgaa cggcaaggtg 1200

gtggccaagg gcagcggcgg cggcggcgcc agccgggagc tggtggtgag ctacgtgaac 1260

gtgaacatgg gcctgaagat ccggcagctg ctgtggttcc acatcagctg cctgaccttc 1320

ggccgggaga ccgtgctgga gtacctggtg agcttcggcg tgtggatccg gacccccccc 1380

gcctaccggc cccccaacgc ccccatcctg agcaccctgc ccgagaccac cgtggtgggc 1440

agcggcggcg gcggccacca ccaccaccac cac 1473

<210> 50
<211> 1197
<212> DNA
<213> Artificial

<220>
<223> FIMHL-HIS-Mi3 J96

<400> 50
atggaaaccg atacactgct gctgtgggtg ctgttgctct gggttccagg ctctacaggc 60

gattttgcct gcaagaccgc caacggcaca gccattccta ttggcggagg cagcgccaat 120

gtgtacgtga acctggctcc tgtggtcaac gtgggccaga atctggtggt ggacctgagc 180

acccagatct tttgccacaa cgactacccc gagacaatca ccgactacgt gacactgcag 240

agaggctctg cttacggcgg cgtgctgagc aatttttccg gcacagtgaa gtacagcggc 300

agcagctacc catttcctac caccagcgag acacccagag tggtgtacaa cagcagaacc 360

gacaagccct ggcctgtggc tctgtacctg acacctgtta gttctgccgg cggagtggcc 420

attaaggccg gatctctgat tgccgtgctg atcctgcggc agaccaacaa ctacaacagc 480

gacgacttcc agttcgtgtg gaacatctac gccaacaacg acgtggtggt gcctacaagc 540

ggcagccacc accatcatca ccatcaccac ggcggcagca tgaagatgga agaactgttc 600

aagaagcaca agatcgtcgc cgtgctgcgg gccaattctg tggaagaggc caaaaaaaag 660

gccctggccg tgtttcttgg cggagtgcac ctgatcgaga tcacctttac cgtgcctgac 720

gccgacaccg tgatcaaaga gctgagcttc ctgaaagaga tgggcgccat catcggagcc 780

ggaaccgtga catctgttga gcaggccaga aaggccgtgg aatctggcgc cgagtttatc 840

gtgtcccctc acctggatga ggaaatcagc cagttcgcca aagaaaaggg cgtgttctac 900

atgcccggcg tgatgacacc tacagagctg gtcaaagcca tgaagctggg ccacaccatc 960

ctgaagctgt ttccaggcga agtcgtgggc cctcagttcg tgaaagctat gaagggccca 1020

tttccaaacg tgaagttcgt gcccactggc ggcgtcaacc tggataatgt gtgcgagtgg 1080

ttcaaggctg gcgtgctggc tgttggagtg ggatctgctc tggtcaaggg cacacctgtg 1140

gaagtggctg agaaggccaa ggccttcgtg gaaaagatca gaggctgcac cgagtga 1197

<210> 51
<211> 1068
<212> DNA
<213> Artificial

<220>
<223> FIMHL-HIS-Fer 536

<400> 51
atggaaaccg atacactgct gctgtgggtg ctgttgctct gggttccagg ctctacaggc 60

gattttgcct gcaagaccgc caacggcaca gccattccta ttggcggagg cagcgccaat 120

gtgtacgtta acctggctcc tgccgtgaac gtgggccaga atctggtggt ggatctgagc 180

acccagatct tttgccacaa cgactacccc gagacaatca ccgactacgt gacactgcag 240

agaggctctg cttacggcgg cgtgctgtct agctttagcg gcaccgtgaa gtacaacggc 300

agcagctacc catttcctac caccagcgag acacccagag tggtgtacaa cagcagaacc 360

gacaagccct ggcctgtggc tctgtacctg acacctgtta gttctgccgg cggagtggcc 420

attaaggccg gatctctgat tgccgtgctg atcctgcggc agaccaacaa ctacaacagc 480

gacgacttcc agttcgtgtg gaacatctac gccaacaacg acgtggtggt gcctacaagc 540

ggcagccacc accatcatca ccatcaccac ggcggcagcg acatcatcaa gctgctgaac 600

gagcaagtga acaaagagat gaacagcagc aacctgtaca tgagcatgag cagctggtgc 660

tacacacaca gcctggatgg cgccggactg ttcctgtttg atcatgccgc cgaggaatac 720

gagcacgcca agaagctgat catcttcctg aacgagaaca acgtgcccgt gcagctgaca 780

tctatcagcg cccctgagca caagttcgag ggcctgacac agatcttcca gaaggcctac 840

gaacacgagc agcacatcag cgagagcatc aacaacatcg tggaccacgc catcaagagc 900

aaggatcacg ccaccttcaa ctttctgcag tggtacgtgg ccgaacagca cgaggaagag 960

gtgctgttca aggacatcct ggacaagatc gagctgatcg gcaacgagaa ccacggcctg 1020

tacctggccg atcagtatgt gaagggaatc gccaagagcc ggaagtga 1068

<210> 52
<211> 1191
<212> DNA
<213> Artificial

<220>
<223> FIMH_DG_PGDGN_IMX313_HIS J96

<400> 52
atggaaaccg atacactgct gctgtgggtg ctgttgctct gggttccagg atctacaggg 60

gattttgcct gcaagaccgc caatggcaca gccattccta ttggcggcgg aagcgccaat 120

gtgtacgtga acctggctcc tgtggtcaac gtgggccaga atctggtggt ggacctgagc 180

acccagatct tttgccacaa cgactacccc gagacaatca ccgactacgt gacactgcag 240

agaggctctg cttacggcgg cgtgctgagc aatttttccg gcacagtgaa gtacagcggc 300

agcagctacc catttcctac caccagcgag acacccagag tggtgtacaa cagcagaacc 360

gacaagccct ggcctgtggc tctgtacctg acacctgtta gttctgccgg cggagtggcc 420

attaaggccg gatctctgat tgccgtgctg atcctgcggc agaccaacaa ctacaacagc 480

gacgacttcc agttcgtgtg gaacatctac gccaacaacg acgtggtggt gcctacaggc 540

ggatgtgatg tgtccgccag agatgtgaca gtgaccctgc ctgattaccc cggctctgtg 600

cctattcctc tgaccgtgta ctgcgccaag tctcagaacc tgggctacta cctgagcggc 660

acaacagccg atgccggcaa cagcatcttt accaacaccg ccagcttcag ccctgctcaa 720

ggtgttggag tgcagctgac ccggaacgga acaatcatcc ccgccaacaa taccgtgtct 780

ctgggagctg tgggcacctc tgctgtgtct cttggcctga cagccaacta tgccagaaca 840

ggcggacaag tgacagccgg caatgtgcag tctatcatcg gcgtgacctt cgtgtatcag 900

cctggcgacg gaaatgccga cgtgaccatc acagtgaatg gcaaggtggt ggccaagggc 960

agctcaggct ctggctctgg atctaaaaaa cagggcgacg ccgatgtgtg tggcgaggtg 1020

gcatatatcc agagcgtggt gtccgattgt cacgtgccaa ccgccgagct gagaaccctg 1080

ctggaaatcc ggaagctgtt cctcgaaatt cagaagctga aggtcgagct gcagggcctg 1140

tctaaagaag gcggaggaag cggatctcac caccaccatc accactgatg a 1191

<210> 53
<211> 1620
<212> DNA
<213> Artificial

<220>
<223> FimH_PGDGN_DG_HIS-Ferritn j96

<400> 53
atggaaaccg atacactgct gctgtgggtg ctgttgctct gggttccagg atctacaggg 60

gattttgcct gcaagaccgc caatggcaca gccattccta ttggcggcgg aagcgccaat 120

gtgtacgtga acctggctcc tgtggtcaac gtgggccaga atctggtggt ggacctgagc 180

acccagatct tttgccacaa cgactacccc gagacaatca ccgactacgt gacactgcag 240

agaggctctg cttacggcgg cgtgctgagc aatttttccg gcacagtgaa gtacagcggc 300

agcagctacc catttcctac caccagcgag acacccagag tggtgtacaa cagcagaacc 360

gacaagccct ggcctgtggc tctgtacctg acacctgtta gttctgccgg cggagtggcc 420

attaaggccg gatctctgat tgccgtgctg atcctgcggc agaccaacaa ctacaacagc 480

gacgacttcc agttcgtgtg gaacatctac gccaacaacg acgtggtggt gcctacaggc 540

ggatgtgatg tgtccgccag agatgtgaca gtgaccctgc ctgattaccc cggctctgtg 600

cctattcctc tgaccgtgta ctgcgccaag tctcagaacc tgggctacta cctgagcggc 660

acaacagccg atgccggcaa cagcatcttt accaacaccg ccagcttcag ccctgctcaa 720

ggtgttggag tgcagctgac ccggaacgga acaatcatcc ccgccaacaa taccgtgtct 780

ctgggagctg tgggcacctc tgctgtgtct cttggcctga cagccaacta tgccagaaca 840

ggcggacaag tgacagccgg caatgtgcag tctatcatcg gcgtgacctt cgtgtatcag 900

cctggcgacg gaaatgccga cgtgaccatc acagtgaatg gcaaggtggt ggccaagagc 960

ggaagccacc accatcatca ccatcaccac ggcggcagca tgaagatgga agaactgttc 1020

aagaagcaca agatcgtcgc cgtgctgcgg gccaattctg tggaagaggc caaaaaaaag 1080

gccctggccg tgtttcttgg cggagtgcac ctgatcgaga tcacctttac cgtgcctgac 1140

gccgacaccg tgatcaaaga gctgagcttc ctgaaagaga tgggcgccat catcggagcc 1200

ggaaccgtga catctgttga gcaggccaga aaggccgtgg aatctggcgc cgagtttatc 1260

gtgtcccctc acctggatga ggaaatcagc cagttcgcca aagaaaaggg cgtgttctac 1320

atgcccggcg tgatgacacc tacagagctg gtcaaagcca tgaagctggg ccacaccatc 1380

ctgaagctgt ttccaggcga agtcgtgggc cctcagttcg tgaaagctat gaagggccca 1440

tttccaaacg tgaagttcgt gcccactggc ggcgtcaacc tggataatgt gtgcgagtgg 1500

ttcaaggctg gcgtgctggc tgttggagtt ggctctgctc tggtcaaggg cacacctgtg 1560

gaagtggctg agaaggccaa ggccttcgtg gaaaagatca gaggctgcac cgagtgatga 1620

<210> 54
<211> 1489
<212> DNA
<213> Artificial

<220>
<223> FIMH_DG_PGDGN-HIS-Ferritin 536

<400> 54
atggaaaccg atacactgct gctgtgggtg ctgttgctct gggttccagg atctacaggg 60

gatgtttgcc tgcaagaccg ccaatggcac agccattcct attggcggcg gaagcgccaa 120

tgtgtacgtt aacctggctc ctgccgtgaa cgtgggccag aatctggtgg tggatctgag 180

cacccagatc ttttgccaca acgactaccc cgagacaatc accgactacg tgacactgca 240

gagaggctct gcttacggcg gcgtgctgtc tagctttagc ggcaccgtga agtacaacgg 300

cagcagctac ccatttccta ccaccagcga gacacccaga gtggtgtaca acagcagaac 360

cgacaagccc tggcctgtgg ctctgtacct gacacctgtt agttctgccg gcggagtggc 420

cattaaggcc ggatctctga ttgccgtgct gatcctgcgg cagaccaaca actacaacag 480

cgacgacttc cagttcgtgt ggaacatcta cgccaacaac gacgtggtgg tgcctacagg 540

cggatgtgat gtgtccgcca gagatgtgac agtgaccctg cctgattacc ccggctctgt 600

gcctattcct ctgaccgtgt actgcgccaa gtctcagaac ctgggctact acctgagcgg 660

cacaacagcc gatgccggca acagcatctt taccaacacc gccagcttca gccctgctca 720

aggtgttgga gtgcagctga cccggaacgg aacaatcatc cccgccaaca ataccgtgtc 780

tctgggagct gtgggcacct ctgctgtgtc tcttggcctg acagccaact atgccagaac 840

aggcggacaa gtgacagccg gcaatgtgca gtctatcatc ggcgtgacct tcgtgtatca 900

gcctggcgac ggaaatgccg acgtgaccat cacagtgaat ggcaaggtgg tggccaagag 960

cggaagccac caccatcatc accatcacca cggcggcagc gacatcatca agctgctgaa 1020

cgagcaagtg aacaaagaga tgaacagcag caacctgtac atgagcatga gcagctggtg 1080

ctacacacac agcctggatg gcgccggact gttcctgttt gatcatgccg ccgaggaata 1140

cgagcacgcc aagaagctga tcatcttcct gaacgagaac aacgtgcccg tccagctgac 1200

atctatcagc gcccctgagc acaagttcga gggcctgaca cagatcttcc agaaggccta 1260

cgaacacgag cagcacatca gcgagagcat caacaacatc gtggaccacg ccatcaagag 1320

caaggatcac gccaccttca actttctgca gtggtacgtg gccgaacagc acgaggaaga 1380

ggtgctgttc aaggacatcc tggacaagat cgagctgatc ggcaacgaga accacggcct 1440

gtacctggcc gatcagtatg tgaagggaat cgccaagagc cgcaagtga 1489

<210> 55
<211> 1533
<212> DNA
<213> Artificial

<220>
<223> FimH_PGDGN_DG-HIS-Ferritin J96

<400> 55
atggaaaccg atacactgct gctgtgggtg ctgttgctct gggttccagg atctacaggg 60

gatgccgctc aacctgctcg gagagccaga agaacaaagc tggccctgtt tgcctgcaag 120

accgccaatg gcacagccat tcctattggc ggcggaagcg ccaatgtgta cgtgaacctg 180

gctcctgtgg tcaacgtggg ccagaatctg gtggtggacc tgagcaccca gatcttttgc 240

cacaacgact accccgagac aatcaccgac tacgtgacac tgcagagagg ctctgcttac 300

ggcggcgtgc tgagcaattt ttccggcaca gtgaagtaca gcggcagcag ctacccattt 360

cctaccacca gcgagacacc cagagtggtg tacaacagca gaaccgacaa gccctggcct 420

gtggctctgt acctgacacc tgttagttct gccggcggag tggccattaa ggccggatct 480

ctgattgccg tgctgatcct gcggcagacc aacaactaca acagcgacga cttccagttc 540

gtgtggaaca tctacgccaa caacgacgtg gtggtgccta caggcggatg tgatgtgtcc 600

gccagagatg tgacagtgac cctgcctgat taccccggct ctgtgcctat tcctctgacc 660

gtgtactgcg ccaagtctca gaacctgggc tactacctga gcggcacaac agccgatgcc 720

ggcaacagca tctttaccaa caccgccagc ttcagccctg ctcaaggtgt tggagtgcag 780

ctgacccgga acggaacaat catccccgcc aacaataccg tgtctctggg agctgtgggc 840

acctctgctg tgtctcttgg cctgacagcc aactatgcca gaacaggcgg acaagtgaca 900

gccggcaatg tgcagtctat catcggcgtg accttcgtgt atcagcctgg cgacggaaat 960

gccgacgtga ccatcacagt gaatggcaag gtggtggcca agagcggaag ccaccaccat 1020

catcaccatc accacggcgg cagcgacatc atcaagctgc tgaacgagca agtgaacaaa 1080

gagatgaaca gcagcaacct gtacatgagc atgagcagct ggtgctacac acacagcctg 1140

gatggcgccg gactgttcct gtttgatcat gccgccgagg aatacgagca cgccaagaag 1200

ctgatcatct tcctgaacga gaacaacgtg cccgtccagc tgacatctat cagcgcccct 1260

gagcacaagt tcgagggcct gacacagatc ttccagaagg cctacgaaca cgagcagcac 1320

atcagcgaga gcatcaacaa catcgtggac cacgccatca agagcaagga tcacgccacc 1380

ttcaactttc tgcagtggta cgtggccgaa cagcacgagg aagaggtgct gttcaaggac 1440

atcctggaca agatcgagct gatcggcaac gagaaccacg gcctgtacct ggccgatcag 1500

tatgtgaagg gaatcgccaa gagccgcaag tga 1533

<210> 56
<211> 1665
<212> DNA
<213> Artificial

<220>
<223> FimH_PGDGN_DG-HIS-MI3 j96

<400> 56
atggaaaccg atacactgct gctgtgggtg ctgttgctct gggttccagg atctacaggg 60

gatgccgctc aacctgctcg gagagccaga agaacaaagc tggccctgtt tgcctgcaag 120

accgccaatg gcacagccat tcctattggc ggcggaagcg ccaatgtgta cgtgaacctg 180

gctcctgtgg tcaacgtggg ccagaatctg gtggtggacc tgagcaccca gatcttttgc 240

cacaacgact accccgagac aatcaccgac tacgtgacac tgcagagagg ctctgcttac 300

ggcggcgtgc tgagcaattt ttccggcaca gtgaagtaca gcggcagcag ctacccattt 360

cctaccacca gcgagacacc cagagtggtg tacaacagca gaaccgacaa gccctggcct 420

gtggctctgt acctgacacc tgttagttct gccggcggag tggccattaa ggccggatct 480

ctgattgccg tgctgatcct gcggcagacc aacaactaca acagcgacga cttccagttc 540

gtgtggaaca tctacgccaa caacgacgtg gtggtgccta caggcggatg tgatgtgtcc 600

gccagagatg tgacagtgac cctgcctgat taccccggct ctgtgcctat tcctctgacc 660

gtgtactgcg ccaagtctca gaacctgggc tactacctga gcggcacaac agccgatgcc 720

ggcaacagca tctttaccaa caccgccagc ttcagccctg ctcaaggtgt tggagtgcag 780

ctgacccgga acggaacaat catccccgcc aacaataccg tgtctctggg agctgtgggc 840

acctctgctg tgtctcttgg cctgacagcc aactatgcca gaacaggcgg acaagtgaca 900

gccggcaatg tgcagtctat catcggcgtg accttcgtgt atcagcctgg cgacggaaat 960

gccgacgtga ccatcacagt gaatggcaag gtggtggcca agagcggaag ccaccaccat 1020

catcaccatc accacggcgg cagcatgaag atggaagaac tgttcaagaa gcacaagatc 1080

gtcgccgtgc tgcgggccaa ttctgtggaa gaggccaaaa aaaaggccct ggccgtgttt 1140

cttggcggag tgcacctgat cgagatcacc tttaccgtgc ctgacgccga caccgtgatc 1200

aaagagctga gcttcctgaa agagatgggc gccatcatcg gagccggaac cgtgacatct 1260

gttgagcagg ccagaaaggc cgtggaatct ggcgccgagt ttatcgtgtc ccctcacctg 1320

gatgaggaaa tcagccagtt cgccaaagaa aagggcgtgt tctacatgcc cggcgtgatg 1380

acacctacag agctggtcaa agccatgaag ctgggccaca ccatcctgaa gctgtttcca 1440

ggcgaagtcg tgggccctca gttcgtgaaa gctatgaagg gcccatttcc aaacgtgaag 1500

ttcgtgccca ctggcggcgt caacctggat aatgtgtgcg agtggttcaa ggctggcgtg 1560

ctggctgttg gagttggctc tgctctggtc aagggcacac ctgtggaagt ggctgagaag 1620

gccaaggcct tcgtggaaaa gatcagaggc tgcaccgagt gatga 1665

<210> 57
<211> 1237
<212> DNA
<213> Artificial

<220>
<223> FimH_PGDGN_DG-HIS-IMX313 j96

<400> 57
atggaaaccg atacactgct gctgtgggtg ctgttgctct gggttccagg atctacaggg 60

gatgccgctc aacctgctcg gagagccaga agaacaaagc tggccctgtt tgcctgcaag 120

accgccaatg gcacagccat tcctattggc ggcggaagcg ccaatgtgta cgtgaacctg 180

gctcctgtgg tcaacgtggg ccagaatctg gtggtggacc tgagcaccca gatcttttgc 240

cacaacgact accccgagac aatcaccgac tacgtgacac tgcagagagg ctctgcttac 300

ggcggcgtgc tgagcaattt ttccggcaca gtgaagtaca gcggcagcag ctacccattt 360

cctaccacca gcgagacacc cagagtggtg tacaacagca gaaccgacaa gccctggcct 420

gtggctctgt acctgacacc tgttagttct gccggcggag tggccattaa ggccggatct 480

ctgattgccg tgctgatcct gcggcagacc aacaactaca acagcgacga cttccagttc 540

gtgtggaaca tctacgccaa caacgacgtg gtggtgccta caggcggatg tgatgtgtcc 600

gccagagatg tgacagtgac cctgcctgat taccccggct ctgtgcctat tcctctgacc 660

gtgtactgcg ccaagtctca gaacctgggc tactacctga gcggcacaac agccgatgcc 720

ggcaacagca tctttaccaa caccgccagc ttcagccctg ctcaaggtgt tggagtgcag 780

ctgacccgga acggaacaat catccccgcc aacaataccg tgtctctggg agctgtgggc 840

acctctgctg tgtctcttgg cctgacagcc aactatgcca gaacaggcgg acaagtgaca 900

gccggcaatg tgcagtctat catcggcgtg accttcgtgt atcagcctgg cgacggaaat 960

gccgacgtga ccatcacagt gaatggcaag gtggtggcca agggcagctc aggctctggc 1020

tctggatcta aaaaacaggg cgacgccgat gtgtgtggcg aggtggcata tatccagagc 1080

gtggtgtccg attgtcacgt gccaaccgcc gagctgagaa ccctgctgga aatccggaag 1140

ctgttcctcg aaattcagaa gctgaaggtc gagctgcagg gcctgtctaa agaaggcgga 1200

ggaagcggat ctcaccacca ccatcaccac tgatgac 1237

<210> 58
<211> 1533
<212> DNA
<213> Artificial

<220>
<223> FimH_PGDGN_DG_Ferritin (536)

<400> 58
atggaaaccg atacactgct gctgtgggtg ctgttgctct gggttccagg atctacaggg 60

gatgccgctc aacctgctcg gagagccaga agaacaaagc tggccctgtt tgcctgcaag 120

accgccaatg gcacagccat tcctattggc ggcggaagcg ccaatgtgta cgttaacctg 180

gctcctgccg tgaacgtggg ccagaatctg gtggtggatc tgagcaccca gatcttttgc 240

cacaacgact accccgagac aatcaccgac tacgtgacac tgcagagagg ctctgcttac 300

ggcggcgtgc tgtctagctt tagcggcacc gtgaagtaca acggcagcag ctacccattt 360

cctaccacca gcgagacacc cagagtggtg tacaacagca gaaccgacaa gccctggcct 420

gtggctctgt acctgacacc tgttagttct gccggcggag tggccattaa ggccggatct 480

ctgattgccg tgctgatcct gcggcagacc aacaactaca acagcgacga cttccagttc 540

gtgtggaaca tctacgccaa caacgacgtg gtggtgccta caggcggatg tgatgtgtcc 600

gccagagatg tgacagtgac cctgcctgat taccccggct ctgtgcctat tcctctgacc 660

gtgtactgcg ccaagtctca gaacctgggc tactacctga gcggcacaac agccgatgcc 720

ggcaacagca tctttaccaa caccgccagc ttcagccctg ctcaaggtgt tggagtgcag 780

ctgacccgga acggaacaat catccccgcc aacaataccg tgtctctggg agctgtgggc 840

acctctgctg tgtctcttgg cctgacagcc aactatgcca gaacaggcgg acaagtgaca 900

gccggcaatg tgcagtctat catcggcgtg accttcgtgt atcagcctgg cgacggaaat 960

gccgacgtga ccatcacagt gaatggcaag gtggtggcca agagcggaag ccaccaccat 1020

catcaccatc accacggcgg cagcgacatc atcaagctgc tgaacgagca agtgaacaaa 1080

gagatgaaca gcagcaacct gtacatgagc atgagcagct ggtgctacac acacagcctg 1140

gatggcgccg gactgttcct gtttgatcat gccgccgagg aatacgagca cgccaagaag 1200

ctgatcatct tcctgaacga gaacaacgtg cccgtccagc tgacatctat cagcgcccct 1260

gagcacaagt tcgagggcct gacacagatc ttccagaagg cctacgaaca cgagcagcac 1320

atcagcgaga gcatcaacaa catcgtggac cacgccatca agagcaagga tcacgccacc 1380

ttcaactttc tgcagtggta cgtggccgaa cagcacgagg aagaggtgct gttcaaggac 1440

atcctggaca agatcgagct gatcggcaac gagaaccacg gcctgtacct ggccgatcag 1500

tatgtgaagg gaatcgccaa gagccgcaag tga 1533

<210> 59
<211> 1170
<212> DNA
<213> Artificial

<220>
<223> FimH-IMX313 j96

<400> 59
atggaaaccg atacactgct gctgtgggtg ctgttgctct gggttccagg atctacaggg 60

gatgccgctc aacctgctcg gagagccaga agaacaaagc tggccctgtt tgcctgcaag 120

accgccaatg gcacagccat tcctattggc ggcggaagcg ccaatgtgta cgtgaacctg 180

gctcctgtgg tcaacgtggg ccagaatctg gtggtggacc tgagcaccca gatcttttgc 240

cacaacgact accccgagac aatcaccgac tacgtgacac tgcagagagg ctctgcttac 300

ggcggcgtgc tgagcaattt ttccggcaca gtgaagtaca gcggcagcag ctacccattt 360

cctaccacca gcgagacacc cagagtggtg tacaacagca gaaccgacaa gccctggcct 420

gtggctctgt acctgacacc tgttagttct gccggcggag tggccattaa ggccggatct 480

ctgattgccg tgctgatcct gcggcagacc aacaactaca acagcgacga cttccagttc 540

gtgtggaaca tctacgccaa caacgacgtg gtggtgccca cctgtgatgt gtccgctaga 600

gatgtgacag tgaccctgcc tgattacccc ggctctgtgc ctattcctct gaccgtgtac 660

tgcgccaagt ctcagaacct gggctactac ctgagcggca caacagccga tgccggcaac 720

agcatcttta ccaacaccgc cagcttcagc cctgctcaag gtgttggagt gcagctgacc 780

cggaacggaa caatcatccc cgccaacaat accgtgtctc tgggagctgt gggcacatct 840

gcagtttctc tgggcctgac agccaactat gccagaacag gcggacaagt gacagccggc 900

aatgtgcagt ctatcatcgg cgtgacattc gtgtatcagg gcagctctgg cagcggctct 960

ggatctaaaa aacagggcga cgccgatgtg tgtggcgagg tggcatatat ccagagcgtg 1020

gtgtccgatt gtcacgtgcc aaccgccgag ctgagaaccc tgctggaaat ccggaagctg 1080

ttcctcgaaa ttcagaagct gaaggtcgag ctgcagggcc tgtctaaaga aggcggagga 1140

agcggatctc accaccacca tcaccactga 1170

<210> 60
<211> 1629
<212> DNA
<213> Artificial

<220>
<223> FimH_mi3 j96

<400> 60
atggaaaccg atacactgct gctgtgggtg ctgttgctct gggttccagg atctacaggg 60

gatgccgctc aacctgctcg gagagccaga agaacaaagc tggccctgtt tgcctgcaag 120

accgccaatg gcacagccat tcctattggc ggcggaagcg ccaatgtgta cgtgaacctg 180

gctcctgtgg tcaacgtggg ccagaatctg gtggtggacc tgagcaccca gatcttttgc 240

cacaacgact accccgagac aatcaccgac tacgtgacac tgcagagagg ctctgcttac 300

ggcggcgtgc tgagcaattt ttccggcaca gtgaagtaca gcggcagcag ctacccattt 360

cctaccacca gcgagacacc cagagtggtg tacaacagca gaaccgacaa gccctggcct 420

gtggctctgt acctgacacc tgttagttct gccggcggag tggccattaa ggccggatct 480

ctgattgccg tgctgatcct gcggcagacc aacaactaca acagcgacga cttccagttc 540

gtgtggaaca tctacgccaa caacgacgtg gtggtgccca cctgtgatgt gtccgctaga 600

gatgtgacag tgaccctgcc tgattacccc ggctctgtgc ctattcctct gaccgtgtac 660

tgcgccaagt ctcagaacct gggctactac ctgagcggca caacagccga tgccggcaac 720

agcatcttta ccaacaccgc cagcttcagc cctgctcaag gtgttggagt gcagctgacc 780

cggaacggaa caatcatccc cgccaacaat accgtgtctc tgggagctgt gggcacatct 840

gcagtttctc tgggcctgac agccaactat gccagaacag gcggacaagt gacagccggc 900

aatgtgcagt ctatcatcgg cgtgaccttc gtgtaccaag gatctggcgg aggcggcatg 960

aagatggaag aactgttcaa gaaacacaag atcgtggccg tgctgcgggc caattctgtg 1020

gaagaggcca aaaaaaaggc cctggccgtg tttctcggag gcgtgcacct gatcgagatc 1080

acctttaccg tgcctgacgc cgacaccgtg atcaaagagc tgagcttcct gaaagagatg 1140

ggcgccatca tcggagccgg aaccgtgaca tctgttgagc aggccagaaa ggccgtggaa 1200

tctggcgccg agtttatcgt gtcccctcac ctggatgagg aaatcagcca gttcgccaaa 1260

gaaaagggcg tgttctacat gcccggcgtg atgacaccta cagagctggt caaagccatg 1320

aagctgggcc acaccatcct gaagctgttt ccaggcgaag tcgtgggccc tcagttcgtg 1380

aaagctatga agggcccatt tccaaacgtg aagttcgtgc ccactggcgg agtgaatctg 1440

gacaacgtgt gcgagtggtt caaggctggc gtgctggctg ttggagttgg ctctgctctg 1500

gtcaagggca cacctgtgga agtggctgag aaggccaagg ccttcgtgga aaagatcaga 1560

ggctgtaccg aaggcagcgg ctctggaagc ggatctggat ctcaccacca tcatcaccat 1620

caccactga 1629

<210> 61
<211> 1248
<212> DNA
<213> Artificial

<220>
<223> FIMHL-HIS-mI3 j96

<400> 61
atggaaaccg atacactgct gctgtgggtg ctgttgctct gggttccagg atctacaggg 60

gatgccgctc aacctgctcg gagagccaga agaacaaagc tggccctgtt tgcctgcaag 120

accgccaatg gcacagccat tcctattggc ggcggaagcg ccaatgtgta cgtgaacctg 180

gctcctgtgg tcaacgtggg ccagaatctg gtggtggacc tgagcaccca gatcttttgc 240

cacaacgact accccgagac aatcaccgac tacgtgacac tgcagagagg ctctgcttac 300

ggcggcgtgc tgagcaattt ttccggcaca gtgaagtaca gcggcagcag ctacccattt 360

cctaccacca gcgagacacc cagagtggtg tacaacagca gaaccgacaa gccctggcct 420

gtggctctgt acctgacacc tgttagttct gccggcggag tggccattaa ggccggatct 480

ctgattgccg tgctgatcct gcggcagacc aacaactaca acagcgacga cttccagttc 540

gtgtggaaca tctacgccaa caacgacgtg gtggtgccta caggatctgg cggacaccac 600

catcatcacc atcaccacgg cagcggctcc atgaagatgg aagaactgtt caagaagcac 660

aagatcgtcg ccgtgctgcg ggccaattct gtggaagagg ccaaaaaaaa ggccctggcc 720

gtgtttcttg gcggagtgca cctgatcgag atcaccttta ccgtgcctga cgccgacacc 780

gtgatcaaag agctgagctt cctgaaagag atgggcgcca tcatcggagc cggaaccgtg 840

acatctgttg agcaggccag aaaggccgtg gaatctggcg ccgagtttat cgtgtcccct 900

cacctggatg aggaaatcag ccagttcgcc aaagaaaagg gcgtgttcta catgcccggc 960

gtgatgacac ctacagagct ggtcaaagcc atgaagctgg gccacaccat cctgaagctg 1020

tttccaggcg aagtcgtggg ccctcagttc gtgaaagcta tgaagggccc atttccaaac 1080

gtgaagttcg tgcccactgg cggcgtcaac ctggataatg tgtgcgagtg gttcaaggct 1140

ggcgtgctgg ctgttggagt gggatctgct ctggtcaagg gcacacctgt ggaagtggct 1200

gagaaggcca aggccttcgt ggaaaagatc agaggctgca ccgagtga 1248

<210> 62
<211> 1428
<212> DNA
<213> Artificial

<220>
<223> FIMH_DG_PGDGN_HIS-Ferritin 536

<400> 62
atgtttgcat gtaaaaccgc aaatggcacc gcaattccga ttggtggtgg tagcgcaaat 60

gtttatgtta atctggcacc ggcagttaat gttggtcaga atctggttgt tgatctgagc 120

acccagattt tttgccataa tgattatccg gaaaccatca ccgattatgt taccctgcag 180

cgtggtagtg catatggtgg tgttctgagc agctttagcg gcaccgtgaa atataacggt 240

agcagctatc cgtttccgac caccagtgaa acaccgcgtg ttgtgtataa tagccgtacc 300

gataaaccgt ggcctgttgc actgtatctg acaccggtta gcagtgccgg tggtgttgca 360

attaaagcag gtagcctgat tgcagttctg attctgcgtc agaccaataa ctataactcc 420

gatgattttc agtttgtgtg gaacatctat gccaataatg atgttgttgt tccgaccggt 480

ggttgtgatg ttagcgcacg tgatgttacc gttacactgc cggattatcc tggtagcgtt 540

ccgattccgc tgaccgttta ttgtgcaaaa agccagaacc tgggttatta tctgagcggc 600

accaccgcag atgcaggtaa tagcattttt accaataccg ccagctttag tccggcacaa 660

ggtgttggtg ttcagctgac ccgtaatggc accattattc cggcaaataa taccgttagc 720

ctgggtgcag ttggcaccag cgcagtgagc ctgggtctga ccgccaatta tgcacgtacc 780

ggtggtcagg ttaccgcagg taatgttcag agcattattg gtgttacctt tgtgtatcag 840

cctggtgatg gtaatgcaga tgtgaccatt accgtgaatg gtaaagttgt tgccaaaagc 900

ggtagtcatc atcaccacca tcatcatcac ggtggtagcg atatcatcaa actgctgaat 960

gaacaggtga acaaagaaat gaatagcagc aacctgtata tgagcatgag cagctggtgt 1020

tatacccata gcctggatgg tgcaggtctg tttctgtttg atcatgcagc cgaagaatat 1080

gagcacgcaa aaaaactgat catcttcctg aatgaaaata atgttccggt gcagctgacc 1140

agcattagcg ctccggaaca taaatttgaa ggtctgacac agatttttca gaaagcctat 1200

gaacatgaac agcacattag cgaaagcatt aacaacattg tggatcacgc catcaaaagc 1260

aaagatcatg caacctttaa ctttctgcag tggtatgttg cagaacagca tgaagaagaa 1320

gtgctgttta aagacatcct ggataaaatt gaactgatcg gcaacgaaaa tcatggtctg 1380

tatctggcag atcagtatgt taaaggtatt gcgaaaagcc gcaaataa 1428

<210> 63
<211> 774

<212> DNA
<213> Artificial

<220>
<223> LS-FIMHL-IMX313-HIS

<400> 63
atgaagtatc tgctgccgac cgcagcagcg ggtctgctgc tgctggcagc acagcctgca 60

atggcatttg catgtaaaac cgcaaatggc accgcaattc cgattggtgg tggtagcgca 120

aatgtttatg ttaatctggc accggttgtt aatgttggtc agaatctggt tgttgatctg 180

agcacccaga ttttttgcca taatgattat ccggaaacca tcaccgatta tgttaccctg 240

cagcgtggta gtgcatatgg tggtgttctg agcaatttta gcggcaccgt gaaatatagc 300

ggtagcagct atccgtttcc gaccaccagt gaaacaccgc gtgttgtgta taatagccgt 360

accgataaac cgtggcctgt tgcactgtat ctgacaccgg tgagcagtgc cggtggtgtt 420

gcaattaaag caggtagcct gattgcagtt ctgattctgc gtcagaccaa taactataac 480

tccgatgatt ttcagtttgt gtggaacatc tatgccaata atgatgttgt tgttccgacc 540

ggtggtagca gcggtagcgg ttcaggtagc aaaaaacagg gtgatgcaga tgtttgtggt 600

gaagttgcat atattcagag cgttgttagc gattgtcatg ttccgacagc agaactgcgt 660

accctgctgg aaattcgtaa actgtttctg gaaatccaga agctgaaagt tgaactgcag 720

ggtctgagca aagaaggtgg cggaagcggt agccatcacc atcaccatca ctga 774

<210> 64
<211> 714
<212> DNA
<213> Artificial

<220>
<223> FIMHL-S24S65-IMX313

<400> 64
atgtttgcaa gcaaaaccgc aaatggcacc gcaattccga ttggtggtgg tagcgcaaat 60

gtttatgtta atctggcacc ggttgttaat gttggtcaga atctggttgt tgatctgagc 120

acccagattt ttagccataa tgattatccg gaaaccatca ccgattatgt taccctgcag 180

cgtggtagtg catatggtgg tgttctgagc aattttagcg gcaccgtgaa atatagcggt 240

agcagctatc cgtttccgac caccagtgaa acaccgcgtg ttgtgtataa tagccgtacc 300

gataaaccgt ggcctgttgc actgtatctg acaccggtta gcagtgccgg tggtgttgca 360

attaaagcag gtagcctgat tgcagttctg attctgcgtc agaccaataa ctataactcc 420

gatgattttc agtttgtgtg gaacatctat gccaataatg atgttgttgt tccgaccggt 480

ggtagtagcg gtagtggtag cggttcaaaa aaacagggtg atgcagatgt ttgtggtgaa 540

gttgcatata ttcagagcgt tgttagcgat tgtcatgtgc cgaccgcaga actgcgtacc 600

ctgctggaaa ttcgtaaact gtttctggaa atccagaagc tgaaagttga actgcagggt 660

ctgagtaaag aaggtggtgg tagtggtagc catcaccatc atcatcacta ataa 714

<210> 65
<211> 1092
<212> DNA
<213> Artificial

<220>
<223> FIMHL-S24S65-foldon-ferritin

<400> 65
atgtttgcaa gcaaaaccgc aaatggcacc gcaattccga ttggtggtgg tagcgcaaat 60

gtttatgtta atctggcacc ggcagttaat gttggtcaga atctggttgt tgatctgagc 120

acccagattt ttagccataa tgattatccg gaaaccatca ccgattatgt taccctgcag 180

cgtggtagtg catatggtgg tgttctgagc agctttagcg gcaccgtgaa atataacggt 240

agcagctatc cgtttccgac caccagtgaa acaccgcgtg ttgtgtataa tagccgtacc 300

gataaaccgt ggcctgttgc actgtatctg acaccggtta gcagtgccgg tggtgttgca 360

attaaagcag gtagcctgat tgcagttctg attctgcgtc agaccaataa ctataactcc 420

gatgattttc agtttgtgtg gaacatctat gccaataatg atgttgttgt tccgaccggt 480

tcaggttata ttccggaagc accgcgtgat ggtcaggcat atgttcgtaa agatggtgaa 540

tgggttctgc tgagcacctt tttaggtagc ggtcatcatc accatcatca tggtagcggt 600

gatatcatta aactgctgaa tgaacaggtg aacaaagaga tgaatagcag caatctgtat 660

atgagcatga gcagctggtg ttatacccat agcctggatg gtgcaggtct gtttctgttt 720

gatcatgcag ccgaagaata tgagcacgca aaaaaactga tcatcttcct gaatgaaaat 780

aatgttccgg ttcagctgac cagcattagc gctccggaac ataaatttga aggtctgaca 840

cagatttttc agaaagccta tgaacatgaa cagcacatta gcgaaagcat taacaacatt 900

gtggatcacg ccatcaaaag caaagatcat gcaaccttta actttctgca gtggtatgtt 960

gcagaacagc atgaagaaga agtgctgttt aaagacatcc tggataaaat tgaactgatc 1020

ggcaacgaaa atcatggtct gtatctggca gatcagtatg ttaaaggtat tgccaaaagc 1080

cgcaagtaat aa 1092

<210> 66
<211> 1173
<212> DNA
<213> Artificial

<220>
<223> FIMHL-S24S65-Mi3

<400> 66
atgtttgcaa gcaaaaccgc aaatggcacc gcaattccga ttggtggtgg tagcgcaaat 60

gtttatgtta atctggcacc ggcagttaat gttggtcaga atctggttgt tgatctgagc 120

acccagattt ttagccataa tgattatccg gaaaccatca ccgattatgt taccctgcag 180

cgtggtagtg catatggtgg tgttctgagc agctttagcg gcaccgtgaa atataacggt 240

agcagctatc cgtttccgac caccagtgaa acaccgcgtg ttgtgtataa tagccgtacc 300

gataaaccgt ggcctgttgc actgtatctg acaccggtta gcagtgccgg tggtgttgca 360

attaaagcag gtagcctgat tgcagttctg attctgcgtc agaccaataa ctataactcc 420

gatgattttc agtttgtgtg gaacatctat gccaataatg atgttgttgt tccgaccggt 480

ggtagtggtg gttcaggtgg tagcatgaaa atggaagaac tgttcaaaaa gcacaaaatt 540

gttgccgttc tgcgtgcaaa tagcgttgaa gaagcaaaaa agaaagcact ggccgttttt 600

ttaggtggtg tgcatctgat tgaaatcacc tttaccgttc cggatgcaga taccgttatt 660

aaagaactga gctttctgaa agaaatgggt gcaattattg gtgcaggcac cgttaccagc 720

gttgaacagg cacgtaaagc agttgaaagc ggtgcagaat ttattgttag tccgcatctg 780

gatgaagaaa ttagccagtt tgcaaaagaa aagggcgtgt tttatatgcc tggtgttatg 840

accccgaccg aactggttaa agcaatgaaa ctgggtcata ccatcctgaa actgtttccg 900

ggtgaagttg ttggtccgca gtttgtgaaa gccatgaaag gtccttttcc gaacgttaaa 960

tttgtgccga caggtggcgt gaatctggat aatgtttgtg aatggtttaa agccggtgtt 1020

ctggccgttg gtgttggtag tgccctggtg aaaggtacac cggttgaagt tgcagaaaaa 1080

gcaaaagcct ttgtggaaaa aattcgtggt tgtaccgaag gtagcggtag cggttcaggt 1140

agtggtagcc atcaccatca tcatcactaa taa 1173

<210> 67
<211> 1173
<212> DNA
<213> Artificial

<220>
<223> FIMHL-mI3

<400> 67
atgtttgcat gtaaaaccgc aaatggcacc gcaattccga ttggtggtgg tagcgcaaat 60

gtttatgtta atctggcacc ggcagttaat gttggtcaga atctggttgt tgatctgagc 120

acccagattt tttgccataa tgattatccg gaaaccatca ccgattatgt taccctgcag 180

cgtggtagtg catatggtgg tgttctgagc agctttagcg gcaccgtgaa atataacggt 240

agcagctatc cgtttccgac caccagtgaa acaccgcgtg ttgtgtataa tagccgtacc 300

gataaaccgt ggcctgttgc actgtatctg acaccggtta gcagtgccgg tggtgttgca 360

attaaagcag gtagcctgat tgcagttctg attctgcgtc agaccaataa ctataactcc 420

gatgattttc agtttgtgtg gaacatctat gccaataatg atgttgttgt tccgaccggt 480

ggtagtggtg gttcaggtgg tagcatgaaa atggaagaac tgttcaaaaa gcacaaaatt 540

gttgccgttc tgcgtgcaaa tagcgttgaa gaagcaaaaa agaaagcact ggccgttttt 600

ttaggtggtg tgcatctgat tgaaatcacc tttaccgttc cggatgcaga taccgttatt 660

aaagaactga gctttctgaa agaaatgggt gcaattattg gtgcaggcac cgttaccagc 720

gttgaacagg cacgtaaagc agttgaaagc ggtgcagaat ttattgttag tccgcatctg 780

gatgaagaaa ttagccagtt tgcaaaagaa aagggcgtgt tttatatgcc tggtgttatg 840

accccgaccg aactggttaa agcaatgaaa ctgggtcata ccatcctgaa actgtttccg 900

ggtgaagttg ttggtccgca gtttgtgaaa gccatgaaag gtccttttcc gaacgttaaa 960

tttgtgccga caggtggcgt gaatctggat aatgtttgtg aatggtttaa agccggtgtt 1020

ctggccgttg gtgttggtag tgccctggtg aaaggtacac cggttgaagt tgcagaaaaa 1080

gcaaaagcct ttgtggaaaa aattcgtggt tgtaccgaag gtagcggtag cggttcaggt 1140

agtggtagcc atcaccatca tcatcactaa taa 1173

<210> 68
<211> 1185
<212> DNA
<213> Artificial

<220>
<223> FimHL-NOCYS-MI3

<400> 68
atgttcgcaa gcaaaaccgc aaatggcacc gcaattccga ttggtggtgg tagcgcaaat 60

gtttatgtta atctggcacc ggcagttaat gttggtcaga atctggttgt tgatctgagc 120

acccagattt ttagccataa tgattatccg gaaaccatca ccgattatgt taccctgcag 180

cgtggtagtg catatggtgg tgttctgagc aattttagcg gcaccgtgaa atatagcggt 240

agcagctatc cgtttccgac caccagtgaa acaccgcgtg ttgtgtataa tagccgtacc 300

gataaaccgt ggcctgttgc actgtatctg acaccggtta gcagtgccgg tggtgttgca 360

attaaagcag gtagcctgat tgcagttctg attctgcgtc agaccaataa ctataactcc 420

gatgattttc agtttgtgtg gaacatctat gccaataatg atgttgttgt tccgaccggt 480

ggtggtggca gtggtggttc aggcggtagc ggtggtagca tgaaaatgga agaactgttc 540

aaaaagcaca aaattgttgc cgttctgcgt gcaaatagcg ttgaagaagc aaaaaagaaa 600

gcactggccg ttttttttagg tggtgtgcat ctgattgaaa tcacctttac cgttccggat 660

gcagataccg ttattaaaga actgagcttt ctgaaagaaa tgggtgcaat tattggtgca 720

ggcaccgtta ccagcgttga acaggcacgt aaagcagttg aaagcggtgc agaatttatt 780

gttagtccgc atctggatga agaaattagc cagtttgcaa aagaaaaggg cgtgttttat 840

atgcctggtg ttatgacccc gaccgaactg gttaaagcaa tgaaactggg tcataccatc 900

ctgaaactgt ttccgggtga agttgttggt ccgcagtttg tgaaagccat gaaaggtcct 960

tttccgaacg ttaaatttgt gccgacaggt ggcgtgaatc tggataatgt ttgtgaatgg 1020

tttaaagccg gtgttctggc agttggtgtt ggtagtgccc tggtgaaagg tacaccggtt 1080

gaagttgcag aaaaagcaaa agcctttgtg gaaaaaattc gtggttgtac cgaaggtagc 1140

ggtagtggta gcggttcagg tagccatcac catcaccatc actga 1185

<210> 69
<211> 1575
<212> DNA
<213> Artificial

<220>
<223> FimHdeltaGG_PGDGN_DG_mi3

<400> 69
atgttcgcct gcaaaaccgc aaatggcacc gcaattccga ttggtggtgg tagcgcaaat 60

gtttatgtta atctggcacc ggttgttaat gttggtcaga atctggttgt tgatctgagc 120

acccagattt tttgccataa tgattatccg gaaaccatca ccgattatgt taccctgcag 180

cgtggtagtg catatggtgg tgttctgagc aattttagcg gcaccgtgaa atatagcggt 240

agcagctatc cgtttccgac caccagtgaa acaccgcgtg ttgtgtataa tagccgtacc 300

gataaaccgt ggcctgttgc actgtatctg acaccggtga gcagtgccgg tggtgttgca 360

attaaagcag gtagcctgat tgcagttctg attctgcgtc agaccaataa ctataactcc 420

gatgattttc agtttgtgtg gaacatctat gccaataatg atgttgttgt tccgacctgt 480

gatgttagcg cacgtgatgt taccgttaca ctgccggatt atcctggtag cgttccgatt 540

ccgctgaccg tttattgtgc aaaaagccag aacctgggtt attatctgag cggcaccacc 600

gcagatgcag gtaatagcat ttttaccaat accgcaagct ttagtccggc acaaggtgtt 660

ggtgttcagc tgacccgtaa tggcaccatt attccggcaa ataataccgt tagcctgggt 720

gcagttggca ccagcgcagt gagcctgggt ctgaccgcca attatgcacg taccggtggt 780

caggttaccg caggtaatgt tcagagcatt attggtgtta cctttgtgta tcagcctggt 840

gatggtaatg cagatgtgac cattaccgtg aatggtaaag ttgttgcaaa aggtagcggt 900

ggtggtggca tgaaaatgga agaactgttc aaaaaacaca agattgttgc cgttctgcgt 960

gcaaatagcg ttgaagaagc aaaaaagaaa gcactggccg tttttttagg tggtgtgcat 1020

ctgattgaaa tcacctttac cgttccggat gcagataccg ttattaaaga actgagcttt 1080

ctgaaagaaa tgggtgcaat tattggcgca ggcaccgtta ccagcgttga acaggcacgt 1140

aaagcagttg aaagcggtgc agaatttatt gttagtccgc atctggatga agaaattagc 1200

cagtttgcaa aagaaaaggg cgtgttttat atgcctggtg ttatgacccc gaccgaactg 1260

gttaaagcaa tgaaactggg tcataccatc ctgaaactgt ttccgggtga agttgttggt 1320

ccgcagtttg tgaaagccat gaaaggtcct tttccgaacg ttaaatttgt gccgaccggt 1380

ggcgtgaatc tggataatgt ttgtgaatgg tttaaagccg gtgttctggc agttggtgtt 1440

ggtagtgccc tggtgaaagg tacaccggtt gaagttgcag aaaaagcaaa agcctttgtg 1500

gaaaaaattc gtggttgtac cgaaggtagt ggtagcggca gcggtagcgg ttcacatcac 1560

catcaccatc actga 1575

<210> 70
<211> 1029
<212> DNA
<213> Artificial

<220>
<223> FimHL-GSG4-Ferritin

<400> 70
atgggcagca gccatcatca tcatcatcac gaactgtact tccagggctt tgcatgtaaa 60

accgcaaatg gcaccgcaat tccgattggt ggtggtagcg caaatgttta tgttaatctg 120

gcaccggcag ttaatgttgg tcagaatctg gttgttgatc tgagcaccca gatttttgc 180

cataatgatt atccggaaac catcaccgat tatgttaccc tgcagcgtgg tagtgcatat 240

ggtggtgttc tgagcagctt tagcggcacc gtgaaatata acggtagcag ctatccgttt 300

ccgaccacca gtgaaacacc gcgtgttgtg tataatagcc gtaccgataa accgtggcct 360

gttgcactgt atctgacacc ggttagcagt gccggtggtg ttgcaattaa agcaggtagc 420

ctgattgcag ttctgattct gcgtcagacc aataactata actccgatga ttttcagttt 480

gtgtggaaca tctatgccaa taatgatgtt gttgttccga ccggtagcgg tggtggtggc 540

gatattatca aactgctgaa tgaacaggtg aacaaagaaa tgaatagcag caacctgtat 600

atgagcatga gcagctggtg ttatacccat agcctggatg gtgcaggtct gtttctgttt 660

gatcatgcag ccgaagaata tgagcacgca aaaaaactga tcatcttcct gaatgaaaat 720

aatgttccgg tgcagctgac cagcattagc gctccggaac ataaatttga aggtctgaca 780

cagatttttc agaaagccta tgaacatgaa cagcacatta gcgaaagcat taacaacatt 840

gtggatcacg ccatcaaaag caaagatcat gcaaccttta actttctgca gtggtatgtt 900

gcagaacagc atgaagaaga agtgctgttt aaagacatcc tggataaaat tgaactgatc 960

ggcaatgaaa atcacggtct gtatctggca gatcagtatg ttaaaggtat tgccaaaagc 1020

cgcaaataa 1029

<210> 71
<211> 1215
<212> DNA
<213> Artificial

<220>
<223> pelBLS-FimHL-mI3

<400> 71
atgaaatacc tgctgccgac cgctgctgct ggtctgctgc tcctcgctgc ccagccggcg 60

atggcctttg catgtaaaac cgcaaatggc accgcaattc cgattggtgg tggtagcgca 120

aatgtttatg ttaatctggc accggttgtt aatgttggtc agaatctggt tgttgatctg 180

agcacccaga ttttttgcca taatgattat ccggaaacca tcaccgatta tgttaccctg 240

cagcgtggta gtgcatatgg tggtgttctg agcaatttta gcggcaccgt gaaatatagc 300

ggtagcagct atccgtttcc gaccaccagt gaaacaccgc gtgttgtgta taatagccgt 360

accgataaac cgtggcctgt tgcactgtat ctgacaccgg tgagcagtgc cggtggtgtt 420

gcaattaaag caggtagcct gattgcagtt ctgattctgc gtcagaccaa taactataac 480

tccgatgatt ttcagtttgt gtggaacatc tatgccaata atgatgttgt tgttccgacc 540

ggtggtggtt caggtatgaa aatggaagaa ctgttcaaaa agcacaagat tgttgccgtt 600

ctgcgtgcaa atagcgttga agaagcaaaa aagaaagcac tggccgtttt tttaggtggt 660

gtgcatctga ttgaaatcac ctttaccgtt ccggatgcag ataccgttat taaagaactg 720

agctttctga aagaaatggg tgcaattatt ggcgcaggca ccgttaccag cgttgaacag 780

gcacgtaaag cagttgaaag cggtgcagaa tttattgtta gtccgcatct ggatgaagaa 840

attagccagt ttgcaaaaga aaagggcgtg ttttatatgc ctggtgttat gaccccgacc 900

gaactggtta aagcaatgaa actgggtcat accatcctga aactgtttcc gggtgaagtt 960

gttggtccgc agtttgtgaa agccatgaaa ggtccttttc cgaacgttaa atttgtgccg 1020

acaggtggcg tgaatctgga taatgtttgt gaatggttta aagccggtgt tctggcagtt 1080

ggtgttggta gtgccctggt gaaaggtaca ccggttgaag ttgcagaaaa agcaaaagcc 1140

tttgtggaaa aaattcgtgg ttgtaccgaa ggtagtggta gcggttcagg tagccaccac 1200

caccaccacc actga 1215

<210> 72
<211> 1449
<212> DNA
<213> Artificial

<220>
<223> FimH_DG_Ferritin (GSGGGG)

<400> 72
atgggcagca gccatcatca tcatcatcac gaactgtact tccagggctt tgcatgtaaa 60

accgcaaatg gcaccgcaat tccgattggt ggtggtagcg caaatgttta tgttaatctg 120

gcaccggcag ttaatgttgg tcagaatctg gttgttgatc tgagcaccca gatttttgc 180

cataatgatt atccggaaac catcaccgat tatgttaccc tgcagcgtgg tagtgcatat 240

ggtggtgttc tgagcagctt tagcggcacc gtgaaatata acggtagcag ctatccgttt 300

ccgaccacca gtgaaacacc gcgtgttgtg tataatagcc gtaccgataa accgtggcct 360

gttgcactgt atctgacacc ggttagcagt gccggtggtg ttgcaattaa agcaggtagc 420

ctgattgcag ttctgattct gcgtcagacc aataactata actccgatga ttttcagttt 480

gtgtggaaca tctatgccaa taatgatgtt gttgttccga ccggtggttg tgatgttagc 540

gcacgtgatg ttaccgttac actgccggat tatcctggta gcgttccgat tccgctgacc 600

gtttattgtg caaaaagcca gaacctgggt tattatctga gcggcaccac cgcagatgca 660

ggtaatagca tttttaccaa taccgcaagc tttagtccgg cacaaggtgt tggtgttcag 720

ctgacccgta atggcaccat tattccggca aataataccg ttagcctggg tgcagttggc 780

accagcgcag tgagcctggg tctgaccgcc aattatgcac gtaccggtgg tcaggttacc 840

gcaggtaatg ttcagagcat tattggtgtt acctttgtgt atcagcctgg tgatggtaat 900

gcagatgtga ccattaccgt gaatggtaaa gttgttgcaa aaggtagcgg tggtggtggc 960

gatattatca aactgctgaa tgaacaggtg aacaaagaaa tgaatagcag caacctgtat 1020

atgagcatga gcagctggtg ttatacccat agcctggatg gtgcaggtct gtttctgttt 1080

gatcatgcag ccgaagaata tgagcacgca aaaaaactga tcatcttcct gaatgaaaat 1140

aatgttccgg tgcagctgac cagcattagc gctccggaac ataaatttga aggtctgaca 1200

cagatttttc agaaagccta tgaacatgaa cagcacatta gcgaaagcat taacaacatt 1260

gtggatcacg ccatcaaaag caaagatcat gcaaccttta actttctgca gtggtatgtt 1320

gcagaacagc atgaagaaga agtgctgttt aaagacatcc tggataaaat tgaactgatc 1380

ggcaatgaaa atcacggtct gtatctggca gatcagtatg ttaaaggtat tgccaaaagc 1440

cgcaaataa 1449

<210> 73
<211> 1185
<212> DNA
<213> Artificial

<220>
<223> FimHL-C-C-MI3

<400> 73
atgttcgcct gcaaaaccgc aaatggcacc gcaattccga ttggtggtgg tagcgcaaat 60

gtttatgtta atctggcacc ggtttgtaat gttggtcaga attgtgttgt tgatctgagc 120

acccagattt tttgccataa tgattatccg gaaaccatca ccgattatgt taccctgcag 180

cgtggtagtg catatggtgg tgttctgagc aattttagcg gcaccgtgaa atatagcggt 240

agcagctatc cgtttccgac caccagtgaa acaccgcgtg ttgtgtataa tagccgtacc 300

gataaaccgt ggcctgttgc actgtatctg acaccggtga gcagtgccgg tggtgttgca 360

attaaagcag gtagcctgat tgcagttctg attctgcgtc agaccaataa ctataactcc 420

gatgattttc agtttgtgtg gaacatctat gccaataatg atgttgttgt tccgaccggt 480

ggtggtggca gtggtggttc aggcggtagc ggtggtagca tgaaaatgga agaactgttc 540

aaaaagcaca aaattgttgc cgttctgcgt gcaaatagcg ttgaagaagc aaaaaagaaa 600

gcactggccg ttttttttagg tggtgtgcat ctgattgaaa tcacctttac cgttccggat 660

gcagataccg ttattaaaga actgagcttt ctgaaagaaa tgggtgcaat tattggtgca 720

ggcaccgtta ccagcgttga acaggcacgt aaagcagttg aaagcggtgc agaatttatt 780

gttagtccgc atctggatga agaaattagc cagtttgcaa aagaaaaggg cgtgttttat 840

atgcctggtg ttatgacccc gaccgaactg gttaaagcaa tgaaactggg tcataccatc 900

ctgaaactgt ttccgggtga agttgttggt ccgcagtttg tgaaagccat gaaaggtcct 960

tttccgaacg ttaaatttgt gccgacaggt ggcgtgaatc tggataatgt ttgtgaatgg 1020

tttaaagccg gtgttctggc agttggtgtt ggtagtgccc tggtgaaagg tacaccggtt 1080

gaagttgcag aaaaagcaaa agcctttgtg gaaaaaattc gtggttgtac cgaaggtagc 1140

ggtagtggta gcggttcagg tagccatcac catcaccatc actga 1185

<210> 74
<211> 904
<212> DNA
<213> Artificial

<220>
<223> FimHL-C-C-qBeta

<400> 74
atgttcgcct gcaaaaccgc aaatggcacc gcaattccga ttggtggtgg tagcgcaaat 60

gtttatgtta atctggcacc ggtttgtaat gttggtcaga attgtgttgt tgatctgagc 120

acccagattt tttgccataa tgattatccg gaaaccatca ccgattatgt taccctgcag 180

cgtggtagtg catatggtgg tgttctgagc aattttagcg gcaccgtgaa atatagcggt 240

agcagctatc cgtttccgac caccagtgaa acaccgcgtg ttgtgtataa tagccgtacc 300

gataaaccgt ggcctgttgc actgtatctg acaccggtga gcagtgccgg tggtgttgca 360

attaaagcag gtagcctgat tgcagttctg attctgcgtc agaccaataa ctataactcc 420

gatgattttc agtttgtgtg gaacatctat gccaataatg atgttgttgt tccgaccggt 480

ggtggtggca gtggtggttc aggcggtagc ggtggcagcg ccaaactgga aaccgttaca 540

ctgggtaata ttggtaaaga tggtaaacag accctggttc tgaatccgcg tggtgttaat 600

ccgaccaatg gtgttgccag cctgagccag gcaggcgcag ttccggcact ggaaaaacgt 660

gttaccgtta gcgttagcca gccgagccgt aatcgtaaaa actataaagt tcaggtgaaa 720

atccagaatc cgaccgcatg taccgccaat ggtagctgtg atccgagcgt tacccgtcag 780

gcatatgcag atgttacctt tagttttacc cagtatagca ccgatgaaga acgtgcattt 840

gttcgtaccg aactggcagc actgctggca agtccgctgc tgattgatgc aattgatcag 900

ctga 904

<210> 75
<211> 965
<212> DNA
<213> Artificial

<220>
<223> HBcAgNC_fimHL splitted

<400> 75
atggatatcg atccgtataa agaatttggt gcaagcgttg aactgctgag ctttctgccg 60

agcgattttt ttccgagcat tcgtgatctg ctggataccg caagcgcact gtatcgtgaa 120

gcactggaaa gtccggaaca ttgtagtccg catcataccg cactgcgtca ggcaattctg 180

tgttggggtg aactgatgaa tctggcaacc tgggttggta gcaatctgga agatccgtag 240

aaggagatat acatatgttt gcatgtaaaa ccgcaaatgg caccgcaatt ccgattggtg 300

gtggtagcgc aaatgtttat gttaatctgg caccggttgt taatgttggt cagaatctgg 360

ttgttgatct gagcacccag atttttttgcc ataatgatta tccggaaacc atcaccgatt 420

atgttaccct gcagcgtggt agtgcatatg gtggtgttct gagcaatttt agcggcaccg 480

tgaaatatag cggtagcagc tatccgtttc cgaccaccag tgaaacaccg cgtgttgtgt 540

ataatagccg taccgataaa ccgtggcctg ttgcgctgta tctgacaccg gtgagcagtg 600

ccggtggtgt tgcaattaaa gcaggtagcc tgattgcagt tctgattctg cgtcagacca 660

ataactataa ctccgatgat tttcagtttg tgtggaacat ctatgccaat aatgatgttg 720

ttgttccgac cggtggtggt tcaggtgcca gccgtgaact ggttgttagc tatgttaatg 780

tgaatatggg cctgaaaatt cgtcagctgc tgtggtttca tatttcatgt ctgacctttg 840

gtcgtgaaac cgttctggaa tatctggtta gctttggtgt ttggattcgt acccctccgg 900

catatcgtcc gcctaatgca ccgattctga gtaccctgcc ggaaacaacc gttgtttgag 960

gatcc 965

<210> 76
<211> 1123
<212> DNA
<213> Artificial

<220>
<223> HBcAgNC_fimHL-LS

<400> 76
atgaaatatc tgctgccgac cgcagcagcg ggtctgctgc tgctggcagc acagcctgca 60

atggcaggtc atcatcacca tcatcatagc ggtggtatgg atattgatcc gtataaagaa 120

tttggtgcca gcgttgaact gctgagcttt ctgccgagcg atttttttcc gagcattcgt 180

gatctgctgg ataccgcaag cgcactgtat cgtgaagcac tggaaagtcc ggaacattgt 240

agtccgcatc ataccgcact gcgtcaggca attctgtgtt ggggtgaact gatgaatctg 300

gcaacctggg ttggtagcaa tctggaagat ccgtagaagg agatatacat atgaaatacc 360

tgttaccgac agccgcagca ggcctgttac tgttagcagc ccagccagcc atggcatttg 420

catgtaaaac cgcaaatggc accgcaattc cgattggtgg tggtagcgca aatgtttatg 480

ttaatctggc accggttgtt aatgttggtc agaatctggt tgttgatctg agcacccaga 540

tttttgcca taatgattat ccggaaacca tcaccgatta tgttaccctg cagcgtggta 600

gtgcatatgg tggtgttctg agcaatttta gcggcaccgt gaaatatagc ggtagcagct 660

atccgtttcc gaccaccagt gaaacaccgc gtgttgtgta taatagccgt accgataaac 720

cgtggcctgt tgcgctgtat ctgacaccgg tgagcagtgc cggtggtgtt gcaattaaag 780

caggtagcct gattgcagtt ctgattctgc gtcagaccaa taactataac tccgatgatt 840

ttcagtttgt gtggaacatc tatgccaata atgatgttgt tgttccgacc ggtggtggtt 900

caggtgcaag ccgtgaactg gttgttagct atgttaatgt gaatatgggc ctgaaaattc 960

gtcagctgct gtggtttcat atttcatgtc tgacctttgg tcgtgaaacc gttctggaat 1020

atctggttag ctttggtgtt tggattcgta cccctccggc atatcgtccg cctaatgcac 1080

cgattctgag taccctgccg gaaacaaccg ttgtttgact cga 1123

<210> 77
<211> 1161
<212> DNA
<213> Artificial

<220>
<223> FIMHL-MI3

<400> 77
atgttcgcct gcaaaaccgc aaatggcacc gcaattccga ttggtggtgg tagcgcaaat 60

gtttatgtta atctggcacc ggttgttaat gttggtcaga atctggttgt tgatctgagc 120

acccagattt tttgccataa tgattatccg gaaaccatca ccgattatgt taccctgcag 180

cgtggtagtg catatggtgg tgttctgagc aattttagcg gcaccgtgaa atatagcggt 240

agcagctatc cgtttccgac caccagtgaa acaccgcgtg ttgtgtataa tagccgtacc 300

gataaaccgt ggcctgttgc actgtatctg acaccggtga gcagtgccgg tggtgttgca 360

attaaagcag gtagcctgat tgcagttctg attctgcgtc agaccaataa ctataactcc 420

gatgattttc agtttgtgtg gaacatctat gccaataatg atgttgttgt tccgaccggt 480

agcggtggtg gtggcatgaa aatggaagaa ctgttcaaaa aacacaagat tgttgccgtt 540

ctgcgtgcaa atagcgttga agaagcaaaa aagaaagcac tggccgtttt tttaggtggt 600

gtgcatctga ttgaaatcac ctttaccgtt ccggatgcag ataccgttat taaagaactg 660

agctttctga aagaaatggg tgcaattatt ggcgcaggca ccgttaccag cgttgaacag 720

gcacgtaaag cagttgaaag cggtgcagaa tttattgtta gtccgcatct ggatgaagaa 780

attagccagt ttgcaaaaga aaagggcgtg ttttatatgc ctggtgttat gaccccgacc 840

gaactggtta aagcaatgaa actgggtcat accatcctga aactgtttcc gggtgaagtt 900

gttggtccgc agtttgtgaa agccatgaaa ggtccttttc cgaacgttaa atttgtgccg 960

accggtggcg tgaatctgga taatgtttgt gaatggttta aagccggtgt tctggcagtt 1020

ggtgttggta gtgccctggt gaaaggtaca ccggttgaag ttgcagaaaa agcaaaagcc 1080

tttgtggaaa aaattcgtgg ttgtaccgaa ggtagtggta gcggcagcgg tagcggttca 1140

catcaccatc accatcactg a 1161

<210> 78
<211> 15
<212> PRT
<213> Artificial

<220>
<223> Leader

<400> 78

Ala Ala Gln Pro Ala Arg Arg Ala Arg Arg Thr Lys Leu Ala Leu
1 5 10 15

<210> 79
<211> 342
<212> PRT
<213> Artificial

<220>
<223> FimH_DNKQ_DG_deglyc

<400> 79

Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1 5 10 15

```
Gly Ser Thr Gly Asp Ala Ala Gln Pro Ala Arg Arg Ala Arg Arg Thr
            20                  25                  30

Lys Leu Ala Leu Phe Ala Cys Lys Thr Ala Ser Gly Thr Ala Ile Pro
        35                  40                  45

Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val Val
    50                  55                  60

Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys
65                  70                  75                  80

His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg
                85                  90                  95

Gly Ser Ala Tyr Gly Gly Val Leu Ser Asp Phe Ser Gly Thr Val Lys
            100                 105                 110

Tyr Ser Gly Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg
        115                 120                 125

Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr
    130                 135                 140

Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser
145                 150                 155                 160

Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp
                165                 170                 175

Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val
            180                 185                 190

Pro Thr Gly Gly Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu
        195                 200                 205

Pro Asp Tyr Pro Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala
    210                 215                 220
```

```
Lys Ser Gln Asn Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp Ala
225                 230                 235                 240


Gly Asn Ser Ile Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly
                245                 250                 255


Val Gly Val Gln Leu Thr Arg Asp Gly Thr Ile Ile Pro Ala Asp Asn
            260                 265                 270


Thr Val Ser Leu Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu
        275                 280                 285


Thr Ala Asn Tyr Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val
    290                 295                 300


Gln Ser Ile Ile Gly Val Thr Phe Val Tyr Gln Asp Asn Lys Gln Ala
305                 310                 315                 320


Asp Val Thr Ile Thr Val Asn Gly Lys Val Val Ala Lys Gly Ser Gly
                325                 330                 335


His His His His His His
            340
```

<210> 80
<211> 343
<212> PRT
<213> Artificial

<220>
<223> FimH_PGDGN_DG

<400> 80

```
Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1               5                   10                  15


Gly Ser Thr Gly Asp Ala Ala Gln Pro Ala Arg Arg Ala Arg Arg Thr
            20                  25                  30
```

```
Lys Leu Ala Leu Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro
        35                  40                  45

Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val Val
    50                  55                  60

Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys
65                  70                  75                  80

His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg
                85                  90                  95

Gly Ser Ala Tyr Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val Lys
            100                 105                 110

Tyr Ser Gly Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg
        115                 120                 125

Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr
    130                 135                 140

Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser
145                 150                 155                 160

Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp
                165                 170                 175

Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val
            180                 185                 190

Pro Thr Gly Gly Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu
        195                 200                 205

Pro Asp Tyr Pro Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala
    210                 215                 220
```

```
Lys Ser Gln Asn Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp Ala
225                 230                 235                 240


Gly Asn Ser Ile Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly
                245                 250                 255


Val Gly Val Gln Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn
            260                 265                 270


Thr Val Ser Leu Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu
        275                 280                 285


Thr Ala Asn Tyr Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val
    290                 295                 300


Gln Ser Ile Ile Gly Val Thr Phe Val Tyr Gln Pro Gly Asp Gly Asn
305                 310                 315                 320


Ala Asp Val Thr Ile Thr Val Asn Gly Lys Val Val Ala Lys Gly Ser
                325                 330                 335


Gly His His His His His His
            340


<210>  81
<211>  342
<212>  PRT
<213>  Artificial

<220>
<223>  FimH_DNKQ_DG

<400>  81

Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1               5                   10                  15


Gly Ser Thr Gly Asp Ala Ala Gln Pro Ala Arg Arg Ala Arg Arg Thr
            20                  25                  30
```

```
Lys Leu Ala Leu Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro
        35                  40                  45

Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val Val
    50                  55                  60

Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys
65                  70                  75                  80

His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg
                85                  90                  95

Gly Ser Ala Tyr Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val Lys
            100                 105                 110

Tyr Ser Gly Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg
        115                 120                 125

Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr
    130                 135                 140

Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser
145                 150                 155                 160

Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp
                165                 170                 175

Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val
            180                 185                 190

Pro Thr Gly Gly Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu
        195                 200                 205

Pro Asp Tyr Pro Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala
    210                 215                 220

Lys Ser Gln Asn Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp Ala
225                 230                 235                 240
```

```
Gly Asn Ser Ile Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly
                245                 250                 255


Val Gly Val Gln Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn
            260                 265                 270


Thr Val Ser Leu Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu
        275                 280                 285


Thr Ala Asn Tyr Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val
    290                 295                 300


Gln Ser Ile Ile Gly Val Thr Phe Val Tyr Gln Asp Asn Lys Gln Ala
305                 310                 315                 320


Asp Val Thr Ile Thr Val Asn Gly Lys Val Val Ala Lys Gly Ser Gly
                325                 330                 335


His His His His His His
            340
```

<210> 82
<211> 341
<212> PRT
<213> Artificial

<220>
<223> FimH_DeltaGG_PGDGN_DG

<400> 82

```
Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1               5                   10                  15


Gly Ser Thr Gly Asp Ala Ala Gln Pro Ala Arg Arg Ala Arg Arg Thr
            20                  25                  30


Lys Leu Ala Leu Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro
        35                  40                  45
```

```
Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val Val
    50                  55                  60

Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys
65                  70                  75                  80

His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg
                85                  90                  95

Gly Ser Ala Tyr Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val Lys
            100                 105                 110

Tyr Ser Gly Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg
        115                 120                 125

Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr
    130                 135                 140

Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser
145                 150                 155                 160

Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp
                165                 170                 175

Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val
            180                 185                 190

Pro Thr Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu Pro Asp
        195                 200                 205

Tyr Pro Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala Lys Ser
    210                 215                 220

Gln Asn Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp Ala Gly Asn
225                 230                 235                 240
```

```
Ser Ile Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly Val Gly
                245                 250                 255


Val Gln Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn Thr Val
            260                 265                 270


Ser Leu Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu Thr Ala
        275                 280                 285


Asn Tyr Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val Gln Ser
    290                 295                 300


Ile Ile Gly Val Thr Phe Val Tyr Gln Pro Gly Asp Gly Asn Ala Asp
305                 310                 315                 320


Val Thr Ile Thr Val Asn Gly Lys Val Val Ala Lys Gly Ser Gly His
                325                 330                 335


His His His His His
            340


<210>  83
<211>  326
<212>  PRT
<213>  Artificial

<220>
<223>  FimH_DGG_sl

<400>  83

Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1               5                   10                  15


Gly Ser Thr Gly Asp Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile
            20                  25                  30


Pro Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val
        35                  40                  45
```

```
Val Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe
    50                  55                  60

Cys His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln
65                  70                  75                  80

Arg Gly Ser Ala Tyr Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val
                85                  90                  95

Lys Tyr Ser Gly Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro
            100                 105                 110

Arg Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu
        115                 120                 125

Tyr Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly
    130                 135                 140

Ser Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser
145                 150                 155                 160

Asp Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val
                165                 170                 175

Val Pro Thr Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu Pro
            180                 185                 190

Asp Tyr Pro Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala Lys
        195                 200                 205

Ser Gln Asn Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp Ala Gly
    210                 215                 220

Asn Ser Ile Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly Val
225                 230                 235                 240

Gly Val Gln Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn Thr
                245                 250                 255
```

```
Val Ser Leu Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu Thr
            260                 265                 270


Ala Asn Tyr Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val Gln
        275                 280                 285


Ser Ile Ile Gly Val Thr Phe Val Tyr Gln Pro Gly Asp Gly Asn Ala
    290                 295                 300


Asp Val Thr Ile Thr Val Asn Gly Lys Val Val Ala Lys Gly Ser Gly
305                 310                 315                 320


His His His His His His
                325


<210>  84
<211>  328
<212>  PRT
<213>  Artificial

<220>
<223>  FimH_PGDGN_sl

<400>  84

Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1               5                   10                  15


Gly Ser Thr Gly Asp Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile
            20                  25                  30


Pro Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val
        35                  40                  45


Val Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe
    50                  55                  60


Cys His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln
65                  70                  75                  80
```

```
Arg Gly Ser Ala Tyr Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val
                85                  90                  95

Lys Tyr Ser Gly Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro
            100                 105                 110

Arg Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu
        115                 120                 125

Tyr Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly
    130                 135                 140

Ser Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser
145                 150                 155                 160

Asp Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val
                165                 170                 175

Val Pro Thr Gly Gly Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr
            180                 185                 190

Leu Pro Asp Tyr Pro Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys
        195                 200                 205

Ala Lys Ser Gln Asn Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp
    210                 215                 220

Ala Gly Asn Ser Ile Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln
225                 230                 235                 240

Gly Val Gly Val Gln Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn
                245                 250                 255

Asn Thr Val Ser Leu Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly
            260                 265                 270
```

```
Leu Thr Ala Asn Tyr Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn
        275                 280                 285

Val Gln Ser Ile Ile Gly Val Thr Phe Val Tyr Gln Pro Gly Asp Gly
    290                 295                 300

Asn Ala Asp Val Thr Ile Thr Val Asn Gly Lys Val Val Ala Lys Gly
305                 310                 315                 320

Ser Gly His His His His His His
                325
```

<210> 85
<211> 327
<212> PRT
<213> Artificial

<220>
<223> FimH_DNKQ_sl

<400> 85

```
Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1               5                   10                  15

Gly Ser Thr Gly Asp Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile
            20                  25                  30

Pro Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val
        35                  40                  45

Val Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe
    50                  55                  60

Cys His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln
65                  70                  75                  80

Arg Gly Ser Ala Tyr Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val
                85                  90                  95
```

```
Lys Tyr Ser Gly Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro
            100                 105                 110

Arg Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu
        115                 120                 125

Tyr Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly
    130                 135                 140

Ser Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser
145                 150                 155                 160

Asp Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val
                165                 170                 175

Val Pro Thr Gly Gly Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr
            180                 185                 190

Leu Pro Asp Tyr Pro Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys
        195                 200                 205

Ala Lys Ser Gln Asn Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp
    210                 215                 220

Ala Gly Asn Ser Ile Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln
225                 230                 235                 240

Gly Val Gly Val Gln Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn
                245                 250                 255

Asn Thr Val Ser Leu Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly
            260                 265                 270

Leu Thr Ala Asn Tyr Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn
        275                 280                 285

Val Gln Ser Ile Ile Gly Val Thr Phe Val Tyr Gln Asp Asn Lys Gln
    290                 295                 300
```

```
Ala Asp Val Thr Ile Thr Val Asn Gly Lys Val Val Ala Lys Gly Ser
305                 310                 315                 320


Gly His His His His His His
                325


<210>  86
<211>  538
<212>  PRT
<213>  Artificial

<220>
<223>  FIMH_DG_PGDGN_536-MI3

<400>  86

Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1               5                   10                  15


Gly Ser Thr Gly Asp Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile
            20                  25                  30


Pro Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Ala
        35                  40                  45


Val Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe
    50                  55                  60


Cys His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln
65                  70                  75                  80


Arg Gly Ser Ala Tyr Gly Gly Val Leu Ser Ser Phe Ser Gly Thr Val
                85                  90                  95


Lys Tyr Asn Gly Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro
            100                 105                 110


Arg Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu
        115                 120                 125
```

```
Tyr Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly
    130                 135                 140

Ser Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser
145                 150                 155                 160

Asp Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val
                165                 170                 175

Val Pro Thr Gly Gly Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr
            180                 185                 190

Leu Pro Asp Tyr Pro Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys
        195                 200                 205

Ala Lys Ser Gln Asn Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp
    210                 215                 220

Ala Gly Asn Ser Ile Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln
225                 230                 235                 240

Gly Val Gly Val Gln Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn
                245                 250                 255

Asn Thr Val Ser Leu Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly
            260                 265                 270

Leu Thr Ala Asn Tyr Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn
        275                 280                 285

Val Gln Ser Ile Ile Gly Val Thr Phe Val Tyr Gln Pro Gly Asp Gly
    290                 295                 300

Asn Ala Asp Val Thr Ile Thr Val Asn Gly Lys Val Val Ala Lys Ser
305                 310                 315                 320
```

```
Gly Ser His His His His His His His His Gly Gly Ser Met Lys Met
                325                 330                 335

Glu Glu Leu Phe Lys Lys His Lys Ile Val Ala Val Leu Arg Ala Asn
            340                 345                 350

Ser Val Glu Glu Ala Lys Lys Lys Ala Leu Ala Val Phe Leu Gly Gly
        355                 360                 365

Val His Leu Ile Glu Ile Thr Phe Thr Val Pro Asp Ala Asp Thr Val
    370                 375                 380

Ile Lys Glu Leu Ser Phe Leu Lys Glu Met Gly Ala Ile Ile Gly Ala
385                 390                 395                 400

Gly Thr Val Thr Ser Val Glu Gln Ala Arg Lys Ala Val Glu Ser Gly
                405                 410                 415

Ala Glu Phe Ile Val Ser Pro His Leu Asp Glu Glu Ile Ser Gln Phe
            420                 425                 430

Ala Lys Glu Lys Gly Val Phe Tyr Met Pro Gly Val Met Thr Pro Thr
        435                 440                 445

Glu Leu Val Lys Ala Met Lys Leu Gly His Thr Ile Leu Lys Leu Phe
    450                 455                 460

Pro Gly Glu Val Val Gly Pro Gln Phe Val Lys Ala Met Lys Gly Pro
465                 470                 475                 480

Phe Pro Asn Val Lys Phe Val Pro Thr Gly Gly Val Asn Leu Asp Asn
                485                 490                 495

Val Cys Glu Trp Phe Lys Ala Gly Val Leu Ala Val Gly Val Gly Ser
            500                 505                 510

Ala Leu Val Lys Gly Thr Pro Val Glu Val Ala Glu Lys Ala Lys Ala
        515                 520                 525
```

```
Phe Val Glu Lys Ile Arg Gly Cys Thr Glu
    530                 535


<210>  87
<211>  598
<212>  PRT
<213>  Artificial

<220>
<223>  FIMH_DG_PGDGN_536-encapsuline

<400>  87

Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1               5                   10                  15


Gly Ser Thr Gly Asp Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile
            20                  25                  30


Pro Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Ala
        35                  40                  45


Val Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe
    50                  55                  60


Cys His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln
65                  70                  75                  80


Arg Gly Ser Ala Tyr Gly Gly Val Leu Ser Ser Phe Ser Gly Thr Val
                85                  90                  95


Lys Tyr Asn Gly Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro
            100                 105                 110


Arg Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu
        115                 120                 125


Tyr Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly
    130                 135                 140
```

```
Ser Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser
145                 150                 155                 160

Asp Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val
                165                 170                 175

Val Pro Thr Gly Gly Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr
            180                 185                 190

Leu Pro Asp Tyr Pro Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys
        195                 200                 205

Ala Lys Ser Gln Asn Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp
    210                 215                 220

Ala Gly Asn Ser Ile Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln
225                 230                 235                 240

Gly Val Gly Val Gln Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn
                245                 250                 255

Asn Thr Val Ser Leu Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly
            260                 265                 270

Leu Thr Ala Asn Tyr Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn
        275                 280                 285

Val Gln Ser Ile Ile Gly Val Thr Phe Val Tyr Gln Pro Gly Asp Gly
    290                 295                 300

Asn Ala Asp Val Thr Ile Thr Val Asn Gly Lys Val Val Ala Lys Ser
305                 310                 315                 320

Gly Ser His His His His His His His His Gly Gly Ser Met Glu Phe
                325                 330                 335
```

```
Leu Lys Arg Ser Phe Ala Pro Leu Thr Glu Lys Gln Trp Gln Glu Ile
            340                 345                 350

Asp Asn Arg Ala Arg Glu Ile Phe Lys Thr Gln Leu Tyr Gly Arg Lys
        355                 360                 365

Phe Val Asp Val Glu Gly Pro Tyr Gly Trp Glu Tyr Ala Ala His Pro
    370                 375                 380

Leu Gly Glu Val Glu Val Leu Ser Asp Glu Asn Glu Val Val Lys Trp
385                 390                 395                 400

Gly Leu Arg Lys Ser Leu Pro Leu Ile Glu Leu Arg Ala Thr Phe Thr
                405                 410                 415

Leu Asp Leu Trp Glu Leu Asp Asn Leu Glu Arg Gly Lys Pro Asn Val
            420                 425                 430

Asp Leu Ser Ser Leu Glu Glu Thr Val Arg Lys Val Ala Glu Phe Glu
        435                 440                 445

Asp Glu Val Ile Phe Arg Gly Cys Glu Lys Ser Gly Val Lys Gly Leu
    450                 455                 460

Leu Ser Phe Glu Glu Arg Lys Ile Glu Cys Gly Ser Thr Pro Lys Asp
465                 470                 475                 480

Leu Leu Glu Ala Ile Val Arg Ala Leu Ser Ile Phe Ser Lys Asp Gly
                485                 490                 495

Ile Glu Gly Pro Tyr Thr Leu Val Ile Asn Thr Asp Arg Trp Ile Asn
            500                 505                 510

Phe Leu Lys Glu Glu Ala Gly His Tyr Pro Leu Glu Lys Arg Val Glu
        515                 520                 525

Glu Cys Leu Arg Gly Gly Lys Ile Ile Thr Thr Pro Arg Ile Glu Asp
    530                 535                 540
```

```
Ala Leu Val Val Ser Glu Arg Gly Gly Asp Phe Lys Leu Ile Leu Gly
545                 550                 555                 560


Gln Asp Leu Ser Ile Gly Tyr Glu Asp Arg Glu Lys Asp Ala Val Arg
                565                 570                 575


Leu Phe Ile Thr Glu Thr Phe Thr Phe Gln Val Val Asn Pro Glu Ala
            580                 585                 590


Leu Ile Leu Leu Lys Phe
        595


<210>  88
<211>  351
<212>  PRT
<213>  Artificial

<220>
<223>  HBcFIMHLJ96

<400>  88

Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1               5                   10                  15


Gly Ser Thr Gly Asp Asp Ile Asp Pro Tyr Lys Glu Phe Gly Ala Ser
            20                  25                  30


Val Glu Leu Leu Ser Phe Leu Pro Ser Asp Phe Phe Pro Ser Ile Arg
        35                  40                  45


Asp Leu Leu Asp Thr Ala Ser Ala Leu Tyr Arg Glu Ala Leu Glu Ser
    50                  55                  60


Pro Glu His Cys Ser Pro His His Thr Ala Leu Arg Gln Ala Ile Leu
65                  70                  75                  80


Cys Trp Gly Glu Leu Met Asn Leu Ala Thr Trp Val Gly Ser Asn Leu
                85                  90                  95
```

```
Glu Asp Pro Gly Ser Gly Gly Gly Gly Phe Ala Cys Lys Thr Ala Asn
            100                 105                 110

Gly Thr Ala Ile Pro Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn
        115                 120                 125

Leu Ala Pro Val Val Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser
    130                 135                 140

Thr Gln Ile Phe Cys His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr
145                 150                 155                 160

Val Thr Leu Gln Arg Gly Ser Ala Tyr Gly Gly Val Leu Ser Asn Phe
                165                 170                 175

Ser Gly Thr Val Lys Tyr Ser Gly Ser Ser Tyr Pro Phe Pro Thr Thr
            180                 185                 190

Ser Glu Thr Pro Arg Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp
        195                 200                 205

Pro Val Ala Leu Tyr Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala
    210                 215                 220

Ile Lys Ala Gly Ser Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn
225                 230                 235                 240

Asn Tyr Asn Ser Asp Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn
                245                 250                 255

Asn Asp Val Val Val Pro Thr Gly Gly Gly Ser Gly Gly Ala Ser Arg
            260                 265                 270

Glu Leu Val Val Ser Tyr Val Asn Val Asn Met Gly Leu Lys Ile Arg
        275                 280                 285
```

```
Gln Leu Leu Trp Phe His Ile Ser Cys Leu Thr Phe Gly Arg Glu Thr
    290                 295                 300

Val Leu Glu Tyr Leu Val Ser Phe Gly Val Trp Ile Arg Thr Pro Pro
305                 310                 315                 320

Ala Tyr Arg Pro Pro Asn Ala Pro Ile Leu Ser Thr Leu Pro Glu Thr
                325                 330                 335

Thr Val Val Gly Ser Gly Gly Gly Gly His His His His His His
            340                 345                 350
```

<210> 89
<211> 491
<212> PRT
<213> Artificial

<220>
<223> HBcFIMHDGJ96

<400> 89

```
Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1               5                   10                  15

Gly Ser Thr Gly Asp Asp Ile Asp Pro Tyr Lys Glu Phe Gly Ala Ser
            20                  25                  30

Val Glu Leu Leu Ser Phe Leu Pro Ser Asp Phe Phe Pro Ser Ile Arg
        35                  40                  45

Asp Leu Leu Asp Thr Ala Ser Ala Leu Tyr Arg Glu Ala Leu Glu Ser
    50                  55                  60

Pro Glu His Cys Ser Pro His His Thr Ala Leu Arg Gln Ala Ile Leu
65                  70                  75                  80

Cys Trp Gly Glu Leu Met Asn Leu Ala Thr Trp Val Gly Ser Asn Leu
                85                  90                  95
```

```
Glu Asp Pro Gly Ser Gly Gly Gly Gly Phe Ala Cys Lys Thr Ala Asn
            100                 105                 110

Gly Thr Ala Ile Pro Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn
        115                 120                 125

Leu Ala Pro Val Val Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser
    130                 135                 140

Thr Gln Ile Phe Cys His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr
145                 150                 155                 160

Val Thr Leu Gln Arg Gly Ser Ala Tyr Gly Gly Val Leu Ser Asn Phe
                165                 170                 175

Ser Gly Thr Val Lys Tyr Ser Gly Ser Ser Tyr Pro Phe Pro Thr Thr
            180                 185                 190

Ser Glu Thr Pro Arg Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp
        195                 200                 205

Pro Val Ala Leu Tyr Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala
    210                 215                 220

Ile Lys Ala Gly Ser Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn
225                 230                 235                 240

Asn Tyr Asn Ser Asp Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn
                245                 250                 255

Asn Asp Val Val Val Pro Thr Gly Gly Cys Asp Val Ser Ala Arg Asp
            260                 265                 270

Val Thr Val Thr Leu Pro Asp Tyr Pro Gly Ser Val Pro Ile Pro Leu
        275                 280                 285

Thr Val Tyr Cys Ala Lys Ser Gln Asn Leu Gly Tyr Tyr Leu Ser Gly
    290                 295                 300
```

```
Thr Thr Ala Asp Ala Gly Asn Ser Ile Phe Thr Asn Thr Ala Ser Phe
305                 310                 315                 320

Ser Pro Ala Gln Gly Val Gly Val Gln Leu Thr Arg Asn Gly Thr Ile
                325                 330                 335

Ile Pro Ala Asn Asn Thr Val Ser Leu Gly Ala Val Gly Thr Ser Ala
            340                 345                 350

Val Ser Leu Gly Leu Thr Ala Asn Tyr Ala Arg Thr Gly Gly Gln Val
        355                 360                 365

Thr Ala Gly Asn Val Gln Ser Ile Ile Gly Val Thr Phe Val Tyr Gln
    370                 375                 380

Pro Gly Asp Gly Asn Ala Asp Val Thr Ile Thr Val Asn Gly Lys Val
385                 390                 395                 400

Val Ala Lys Gly Ser Gly Gly Gly Gly Ala Ser Arg Glu Leu Val Val
                405                 410                 415

Ser Tyr Val Asn Val Asn Met Gly Leu Lys Ile Arg Gln Leu Leu Trp
            420                 425                 430

Phe His Ile Ser Cys Leu Thr Phe Gly Arg Glu Thr Val Leu Glu Tyr
        435                 440                 445

Leu Val Ser Phe Gly Val Trp Ile Arg Thr Pro Pro Ala Tyr Arg Pro
    450                 455                 460

Pro Asn Ala Pro Ile Leu Ser Thr Leu Pro Glu Thr Thr Val Val Gly
465                 470                 475                 480

Ser Gly Gly Gly Gly His His His His His His
                485                 490
```

<210> 90
<211> 1028
<212> DNA
<213> Artificial

<220>
<223> FimH_DNKQ_DG_deglyc

<400> 90
atggaaaccg atacactgct gctgtgggtg ctgttgctct gggttccagg atctacaggg 60

gatgccgctc aacctgctcg gagagccaga agaacaaagc tggccctgtt tgcctgtaaa 120

accgccagcg gcacagccat tcctattggc ggaggcagcg ccaatgtgta cgtgaacctg 180

gctcctgtgg tcaacgtggg ccagaatctg gtggtggacc tgagcaccca gatcttttgc 240

cacaacgact accccgagac aatcaccgac tacgtgacac tgcagagagg ctctgcttac 300

ggcggcgtgc tgagcgattt ttccggcaca gtgaagtaca gcggcagcag ctacccattt 360

cctaccacca gcgagacacc cagagtggtg tacaacagca gaaccgacaa gccctggcct 420

gtggctctgt acctgacacc tgttagttct gccggcggag tggccattaa ggccggatct 480

ctgattgccg tgctgatcct gcggcagacc aacaactaca acagcgacga cttccagttc 540

gtgtggaaca tctacgccaa caacgacgtg gtggtgccta caggcggatg tgatgtgtcc 600

gccagagatg tgacagtgac cctgcctgat taccccggct ctgtgcctat tcctctgacc 660

gtgtactgcg ccaagtctca gaacctgggc tactacctga gcggcacaac agccgatgcc 720

ggcaacagca tctttaccaa caccgccagc ttcagccctg ctcaaggtgt tggagtgcag 780

ctgaccagag atggcacaat catccccgcc gacaataccg tgtctctggg cgctgttggc 840

acatctgcag tttctctggg cctgaccgcc aactatgcca gaacaggtgg acaagtgacc 900

gccggcaatg tgcagtctat catcggcgtg acattcgtgt atcaggacaa caagcaggcc 960

gacgtgacca tcaccgtgaa tggcaaagtg gtggccaaag gctctggcca tcaccaccac 1020

catcactg 1028

<210> 91
<211> 1011
<212> DNA
<213> Artificial

<220>
<223> FimH_PGDGN_DG

<400> 91
atggaaaccg acacactgct gctgtgggtg ctgcttttgt gggtgccagg atctacaggg 60

gatgccgctc aacctgctcg gagagccaga agaacaaagc tggccctgtt tgcctgcaag 120

accgccaatg gcacagccat tcctattggc ggcggaagcg ccaatgtgta cgtgaacctg 180

gctcctgtgg tcaacgtggg ccagaatctg gtggtggacc tgagcaccca gatcttttgc 240

cacaacgact accccgagac aatcaccgac tacgtgacac tgcagagagg ctctgcttac 300

ggcggcgtgc tgagcaattt ttccggcaca gtgaagtaca gcggcagcag ctacccattt 360

cctaccacca gcgagacacc cagagtggtg tacaacagca gaaccgacaa gccctggcct 420

gtggctctgt acctgacacc tgttagttct gccggcggag tggccattaa ggccggatct 480

ctgattgccg tgctgatcct gcggcagacc aacaactaca acagcgacga cttccagttc 540

gtgtggaaca tctacgccaa caacgacgtg gtggtgccta caggcggatg tgatgtgtcc 600

gccagagatg tgacagtgac cctgcctgat taccccggct ctgtgcctat tcctctgacc 660

gtgtactgcg ccaagtctca gaacctgggc tactacctga gcggcacaac agccgatgcc 720

ggcaacagca tctttaccaa caccgccagc ttcagccctg ctcaaggtgt tggagtgcag 780

ctgacccgga acggaacaat catccccgcc aacaataccg tgtctctggg agctgtgggc 840

acctctgctg tttctctggg cctgacagcc aactatgcca gaacaggcgg acaagtgaca 900

gccggcaatg tgcagtctat catcggcgtg acattcgtgt atcaggacaa caagcaggcc 960

gacgtgacca tcaccgtgaa tggcaaagtg gtggccaaag gctctggcca t 1011

<210> 92
<211> 1031
<212> DNA
<213> Artificial

<220>
<223> FimH_DNKQ_DG

<400> 92
atggaaaccg acacactgct gctgtgggtg ctgcttttgt gggtgccagg atctacaggg 60

gatgccgctc aacctgctcg gagagccaga agaacaaagc tggccctgtt tgcctgcaag 120

accgccaatg gcacagccat tcctattggc ggcggaagcg ccaatgtgta cgtgaacctg 180

gctcctgtgg tcaacgtggg ccagaatctg gtggtggacc tgagcaccca gatcttttgc 240

cacaacgact accccgagac aatcaccgac tacgtgacac tgcagagagg ctctgcttac 300

ggcggcgtgc tgagcaattt ttccggcaca gtgaagtaca gcggcagcag ctacccattt 360

cctaccacca gcgagacacc cagagtggtg tacaacagca gaaccgacaa gccctggcct 420

gtggctctgt acctgacacc tgttagttct gccggcggag tggccattaa ggccggatct 480

ctgattgccg tgctgatcct gcggcagacc aacaactaca acagcgacga cttccagttc 540

gtgtggaaca tctacgccaa caacgacgtg gtggtgccta caggcggatg tgatgtgtcc 600

gccagagatg tgacagtgac cctgcctgat taccccggct ctgtgcctat tcctctgacc 660

gtgtactgcg ccaagtctca gaacctgggc tactacctga gcggcacaac agccgatgcc 720

ggcaacagca tctttaccaa caccgccagc ttcagccctg ctcaaggtgt tggagtgcag 780

ctgacccgga acggaacaat catccccgcc aacaataccg tgtctctggg agctgtgggc 840

acctctgctg tttctctggg cctgacagcc aactatgcca gaacaggcgg acaagtgaca 900

gccggcaatg tgcagtctat catcggcgtg accttcgtgt atcagcctgg cgacggaaat 960

gccgacgtga ccatcacagt gaatggcaag gtggtggcca aaggctctgg acaccaccac 1020

catcaccact g 1031

<210> 93
<211> 1025
<212> DNA
<213> Artificial

<220>
<223> FimH_DeltaGG_PGDGN_DG

<400> 93
atggaaaccg acacactgct gctgtgggtg ctgcttttgt gggtgccagg atctacaggg 60

gatgccgctc aacctgctcg gagagccaga agaacaaagc tggccctgtt tgcctgcaag 120

accgccaatg gcacagccat tcctattggc ggcggaagcg ccaatgtgta cgtgaacctg 180

gctcctgtgg tcaacgtggg ccagaatctg gtggtggacc tgagcaccca gatcttttgc 240

cacaacgact accccgagac aatcaccgac tacgtgacac tgcagagagg ctctgcttac 300

ggcggcgtgc tgagcaattt ttccggcaca gtgaagtaca gcggcagcag ctacccattt 360

cctaccacca gcgagacacc cagagtggtg tacaacagca gaaccgacaa gccctggcct 420

gtggctctgt acctgacacc tgttagttct gccggcggag tggccattaa ggccggatct 480

ctgattgccg tgctgatcct gcggcagacc aacaactaca acagcgacga cttccagttc 540

gtgtggaaca tctacgccaa caacgacgtg gtggtgccca cctgtgatgt gtccgctaga 600

gatgtgacag tgaccctgcc tgattacccc ggctctgtgc ctattcctct gaccgtgtac 660

tgcgccaagt ctcagaacct gggctactac ctgagcggca caacagccga tgccggcaac 720

agcatcttta ccaacaccgc cagcttcagc cctgctcaag gtgttggagt gcagctgacc 780

cggaacggaa caatcatccc cgccaacaat accgtgtctc tgggagctgt gggcacatct 840

gctgtttctc tgggcctgac agccaactat gccagaacag gcggacaagt gacagccggc 900

aatgtgcagt ctatcatcgg cgtgaccttc gtgtatcagc ctggcgacgg aaatgccgac 960

gtgaccatca cagtgaatgg caaggtggtg gccaaaggct ctggacacca ccaccatcac 1020

cactg 1025

<210> 94
<211> 981
<212> DNA
<213> Artificial

<220>
<223> FimH_DGG_sl

<400> 94
atggaaaccg acacactgct gctgtgggtg ctgcttttgt gggtgccagg ctctacaggc 60

gattttgcct gcaagaccgc caacggcaca gccattccta ttggcggagg cagcgccaat 120

gtgtacgtga acctggctcc tgtggtcaac gtgggccaga atctggtggt ggacctgagc 180

acccagatct tttgccacaa cgactacccc gagacaatca ccgactacgt gacactgcag 240

agaggctctg cttacggcgg cgtgctgagc aatttttccg gcacagtgaa gtacagcggc 300

agcagctacc catttcctac caccagcgag acacccagag tggtgtacaa cagcagaacc 360

gacaagccct ggcctgtggc tctgtacctg acacctgtta gttctgccgg cggagtggcc 420

attaaggccg gatctctgat tgccgtgctg atcctgcggc agaccaacaa ctacaacagc 480

gacgacttcc agttcgtgtg gaacatctac gccaacaacg acgtggtggt gcccacctgt 540

gatgtgtccg ctagagatgt gacagtgacc ctgcctgatt accccggctc tgtgcctatt 600

cctctgaccg tgtactgcgc caagagccag aacctgggct actacctgtc tggcacaaca 660

gccgatgccg gcaacagcat ctttaccaac accgccagct tcagccctgc tcaaggtgtt 720

ggagtgcagc tgacccggaa cggaacaatc atccccgcca acaataccgt gtctctggga 780

gctgtgggca catctgctgt ttctctgggc ctgaccgcca attatgccag aacaggcgga 840

caagtgaccg ccggcaatgt gcagtctatc atcggcgtga ccttcgtgta tcagcctggc 900

gacggaaacg ccgatgtgac catcacagtg aatggcaagg tggtggccaa aggctctgga 960

caccaccacc atcaccactg a 981

<210> 95
<211> 993
<212> DNA
<213> Artificial

<220>
<223> FimH_PGDGN_sl

<400> 95
atggaaaccg acacactgct gctgtgggtg ctgcttttgt gggtgccagg ctctacaggc 60

gattttgcct gcaagaccgc caacggcaca gccattccta ttggcggagg cagcgccaat 120

gtgtacgtga acctggctcc tgtggtcaac gtgggccaga atctggtggt ggacctgagc 180

acccagatct tttgccacaa cgactacccc gagacaatca ccgactacgt gacactgcag 240

agaggctctg cttacggcgg cgtgctgagc aatttttccg gcacagtgaa gtacagcggc 300

agcagctacc catttcctac caccagcgag acacccagag tggtgtacaa cagcagaacc 360

gacaagccct ggcctgtggc tctgtacctg acacctgtta gttctgccgg cggagtggcc 420

attaaggccg gatctctgat tgccgtgctg atcctgcggc agaccaacaa ctacaacagc 480

gacgacttcc agttcgtgtg gaacatctac gccaacaacg acgtggtggt gcctacaggc 540

ggatgtgatg tgtccgccag agatgtgaca gtgaccctgc ctgattaccc cggctctgtg 600

cctattcctc tgaccgtgta ctgcgccaag agccagaacc tgggctacta cctgtctggc 660

acaacagccg atgccggcaa cagcatcttt accaacaccg ccagcttcag ccctgctcaa 720

ggtgttggag tgcagctgac ccggaacgga acaatcatcc ccgccaacaa taccgtgtct 780

ctgggagctg tgggcacctc tgctgtttct ctgggcctga cagccaacta tgccagaaca 840

ggcggacaag tgacagccgg caatgtgcag tctatcatcg gcgtgacctt cgtgtatcag 900

cctggcgacg gaaacgccga tgtgaccatc acagtgaatg gcaaggtggt ggccaaaggc 960

tctggacacc accaccatca ccactgactc gag 993

<210> 96
<211> 990
<212> DNA
<213> Artificial

<220>
<223> FimH_DNKQ_sl

<400> 96
atggaaaccg acacactgct gctgtgggtg ctgcttttgt gggtgccagg ctctacaggc 60

gattttgcct gcaagaccgc caacggcaca gccattccta ttggcggagg cagcgccaat 120

gtgtacgtga acctggctcc tgtggtcaac gtgggccaga atctggtggt ggacctgagc 180

acccagatct tttgccacaa cgactacccc gagacaatca ccgactacgt gacactgcag 240

agaggctctg cttacggcgg cgtgctgagc aatttttccg gcacagtgaa gtacagcggc 300

agcagctacc catttcctac caccagcgag acacccagag tggtgtacaa cagcagaacc 360

gacaagccct ggcctgtggc tctgtacctg acacctgtta gttctgccgg cggagtggcc 420

attaaggccg gatctctgat tgccgtgctg atcctgcggc agaccaacaa ctacaacagc 480

gacgacttcc agttcgtgtg gaacatctac gccaacaacg acgtggtggt gcctacaggc 540

ggatgtgatg tgtccgccag agatgtgaca gtgaccctgc ctgattaccc cggctctgtg 600

cctattcctc tgaccgtgta ctgcgccaag agccagaacc tgggctacta cctgtctggc 660

acaacagccg atgccggcaa cagcatcttt accaacaccg ccagcttcag ccctgctcaa 720

ggtgttggag tgcagctgac ccggaacgga acaatcatcc ccgccaacaa taccgtgtct 780

ctgggagctg tgggcacctc tgctgtttct ctgggcctga cagccaacta tgccagaaca 840

ggcggacaag tgacagccgg caatgtgcag tctatcatcg gcgtgacatt cgtgtatcag 900

gacaacaagc aggccgacgt gaccatcacc gtgaatggca aagtggtggc caaaggctct 960

ggccatcacc accaccatca ctgactcgag 990

<210> 97
<211> 1617
<212> DNA
<213> Artificial

<220>
<223> FIMH_DG_PGDGN_536-MI3

<400> 97
atggaaaccg atacactgct gctgtgggtg ctgttgctct gggttccagg ctctacaggc 60

gattttgcct gcaagaccgc caacggcaca gccattccta ttggcggagg cagcgccaat 120

gtgtacgtta acctggctcc tgccgtgaac gtgggccaga atctggtggt ggatctgagc 180

acccagatct tttgccacaa cgactacccc gagacaatca ccgactacgt gacactgcag 240

agaggctctg cttacggcgg cgtgctgtct agctttagcg gcaccgtgaa gtacaacggc 300

agcagctacc catttcctac caccagcgag acacccagag tggtgtacaa cagcagaacc 360

gacaagccct ggcctgtggc tctgtacctg acacctgtta gttctgccgg cggagtggcc 420

attaaggccg gatctctgat tgccgtgctg atcctgcggc agaccaacaa ctacaacagc 480

gacgacttcc agttcgtgtg gaacatctac gccaacaacg acgtggtggt gcctacaggc 540

ggatgtgatg tgtccgccag agatgtgaca gtgaccctgc ctgattaccc cggctctgtg 600

cctattcctc tgaccgtgta ctgcgccaag agccagaacc tgggctacta cctgtctggc 660

acaacagccg atgccggcaa cagcatcttt accaacaccg ccagcttcag ccctgctcaa 720

ggtgttggag tgcagctgac ccggaacgga acaatcatcc ccgccaacaa taccgtgtct 780

ctgggagctg tgggcacctc tgctgtgtct cttggcctga cagccaacta tgccagaaca 840

ggcggacaag tgacagccgg caatgtgcag tctatcatcg gcgtgacctt cgtgtatcag 900

cctggcgacg gaaacgccga tgtgaccatc acagtgaatg gcaaggtggt ggccaagagc 960

ggaagccacc accatcatca ccatcaccac ggcggcagca tgaagatgga agaactgttc 1020

aagaagcaca agatcgtcgc cgtgctgcgg gccaattctg tggaagaggc caaaaaaaag 1080

gccctggccg tgtttcttgg cggagtgcac ctgatcgaga tcacctttac cgtgcctgac 1140

gccgacaccg tgatcaaaga gctgagcttc ctgaaagaga tgggcgccat catcggagcc 1200

ggcacagtga catctgttga gcaggccaga aaggccgtgg aatctggcgc cgagtttatc 1260

gtgtcccctc acctggatga ggaaatcagc cagttcgcca aagaaaaggg cgtgttctac 1320

atgcccggcg tgatgacacc tacagagctg gtcaaagcca tgaagctggg ccacaccatc 1380

ctgaagctgt ttccaggcga agtcgtgggc cctcagttcg tgaaagctat gaagggccca 1440

tttccaaacg tgaagttcgt gcccactggc ggcgtgaacc tggataatgt gtgcgagtgg 1500

ttcaaggctg gcgtgctggc tgttggagtt ggctctgctc tggtcaaggg cacacctgtg 1560

gaagtggctg agaaggccaa ggccttcgtg gaaaagatca gaggctgcac cgagtga 1617

<210> 98
<211> 1053
<212> DNA
<213> Artificial

<220>
<223> HBcFIMHLJ96

<400> 98
atggaaaccg atacactgct gctgtgggtg ctgttgctct gggttccagg atctaccggc 60

gacgacatcg acccctacaa agagtttggc gccagcgtcg agctgctgag cttcctgcct 120

agcgacttct tcccttccat ccgggatctg ctggataccg ctagcgccct gtatagagag 180

gccctggaaa gccctgagca ctgctctcca catcacacag ccctgagaca ggccatcctg 240

tgttggggcg aactgatgaa tctggccacc tgggtcggaa gcaacctgga agatcctggt 300

tctggcggcg gaggctttgc ctgtaaaaca gccaatggca ccgccattcc tatcggaggc 360

ggcagcgcca atgtgtacgt taacctggct cctgtggtca acgtgggcca gaatctggtg 420

gtggacctga gcacccagat cttttgccac aacgactacc ccgagacaat caccgactac 480

gtgacactgc agagaggctc tgcttacggc ggcgtgctga gcaatttttc cggcacagtg 540

aagtacagcg gcagcagcta cccatttcct accaccagcg agacacccag agtggtgtac 600

aacagcagaa ccgacaagcc ctggcctgtg gctctgtacc tgacacctgt tagttctgct 660

ggcggagtgg ccatcaaggc cggatctctg attgccgtgc tgatcctgcg gcagaccaac 720

aactacaaca gcgacgactt ccagttcgtg tggaacatct acgccaacaa cgacgtggtg 780

gtgcctacag gcggaggatc tggcggagct tctagagaac tggtcgtgtc ctacgtgaac 840

gtgaacatgg gcctgaagat ccggcagctg ctctggtttc acatcagctg tctgaccttc 900

ggccgggaaa ccgtgctgga atacctggtg tccttcggcg tgtggatcag aacccctcct 960

gcctatagac ctcctaacgc tcccatcctg agcacactgc ctgagacaac agttgttgga 1020

agcggaggcg gaggccacca ccatcaccat cat 1053

<210> 99
<211> 1473
<212> DNA
<213> Artificial

<220>
<223> HBcFIMHDGJ96

<400> 99
atggagaccg acaccctgct gctgtgggtg ctgctgctgt gggtgcccgg cagcaccggc 60

gacgacatcg acccctacaa ggagttcggc gccagcgtgg agctgctgag cttcctgccc 120

agcgacttct tccccagcat ccgggacctg ctggacaccg ccagcgccct gtaccgggag 180

gccctggaga gccccgagca ctgcagcccc caccacaccg ccctgcggca ggccatcctg 240

tgctggggcg agctgatgaa cctggccacc tgggtgggca gcaacctgga ggaccccggc 300

agcggcggcg gcggcttcgc ctgcaagacc gccaacggca ccgccatccc catcggcggc 360

ggcagcgcca acgtgtacgt gaacctggcc cccgtggtga acgtgggcca gaacctggtg 420

gtggacctga gcacccagat cttctgccac aacgactacc ccgagaccat caccgactac 480

gtgaccctgc agcggggcag cgcctacggc ggcgtgctga gcaacttcag cggcaccgtg 540

aagtacagcg gcagcagcta ccccttcccc accaccagcg agacccccg ggtggtgtac 600

aacagccgga ccgacaagcc ctggcccgtg gccctgtacc tgacccccgt gagcagcgcc 660

ggcggcgtgg ccatcaaggc cggcagcctg atcgccgtgc tgatcctgcg gcagaccaac 720

aactacaaca gcgacgactt ccagttcgtg tggaacatct acgccaacaa cgacgtggtg 780

gtgcccaccg gcggctgcga cgtgagcgcc cgggacgtga ccgtgaccct gcccgactac 840

cccggcagcg tgcccatccc cctgaccgtg tactgcgcca agagccagaa cctgggctac 900

tacctgagcg gcaccaccgc cgacgccggc aacagcatct tcaccaacac cgccagcttc 960

agccccgccc agggcgtggg cgtgcagctg acccggaacg gcaccatcat ccccgccaac 1020

aacaccgtga gcctgggcgc cgtgggcacc agcgccgtga gcctgggcct gaccgccaac 1080

tacgcccgga ccggcggcca ggtgaccgcc ggcaacgtgc agagcatcat cggcgtgacc 1140

ttcgtgtacc agcccggcga cggcaacgcc gacgtgacca tcaccgtgaa cggcaaggtg 1200

gtggccaagg gcagcggcgg cggcggcgcc agccgggagc tggtggtgag ctacgtgaac 1260

gtgaacatgg gcctgaagat ccggcagctg ctgtggttcc acatcagctg cctgaccttc 1320

ggccgggaga ccgtgctgga gtacctggtg agcttcggcg tgtggatccg gacccccccc 1380

gcctaccggc cccccaacgc ccccatcctg agcaccctgc ccgagaccac cgtggtgggc 1440

agcggcggcg gcggccacca ccaccaccac cac 1473

<210> 100
<211> 303
<212> PRT
<213> E. coli

<400> 100

Met Ile Val Met Lys Arg Val Ile Thr Leu Phe Ala Val Leu Leu Met
1 5 10 15

Gly Trp Ser Val Asn Ala Trp Ser Phe Ala Cys Lys Thr Ala Asn Gly
20 25 30

Thr Ala Ile Pro Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn Leu

```
        35                  40                  45

Ala Pro Ala Val Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser Thr
    50                  55                  60

Gln Ile Phe Cys His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr Val
65                  70                  75                  80

Thr Leu Gln Arg Gly Ser Ala Tyr Gly Gly Val Leu Ser Ser Phe Ser
                85                  90                  95

Gly Thr Val Lys Tyr Asn Gly Ser Ser Tyr Pro Phe Pro Thr Thr Ser
            100                 105                 110

Glu Thr Pro Arg Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp Pro
        115                 120                 125

Val Ala Leu Tyr Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala Ile
    130                 135                 140

Lys Ala Gly Ser Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn Asn
145                 150                 155                 160

Tyr Asn Ser Asp Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn Asn
                165                 170                 175

Asp Val Val Val Pro Thr Gly Gly Cys Asp Val Ser Ala Arg Asp Val
            180                 185                 190

Thr Val Thr Leu Pro Asp Tyr Pro Gly Ser Val Pro Ile Pro Leu Thr
        195                 200                 205

Val Tyr Cys Ala Lys Ser Gln Asn Leu Gly Tyr Tyr Leu Ser Gly Thr
    210                 215                 220

Thr Ala Asp Ala Gly Asn Ser Ile Phe Thr Asn Thr Ala Ser Phe Ser
225                 230                 235                 240
```

```
Pro Ala Gln Gly Val Gly Val Gln Leu Thr Arg Asn Gly Thr Ile Ile
                245                 250                 255


Pro Ala Asn Asn Thr Val Ser Leu Gly Ala Val Gly Thr Ser Ala Val
            260                 265                 270


Ser Leu Gly Leu Thr Ala Asn Tyr Ala Arg Thr Gly Gly Gln Val Thr
        275                 280                 285


Ala Gly Asn Val Gln Ser Ile Ile Gly Val Thr Phe Val Tyr Gln
    290                 295                 300
```

<210> 101
<211> 279
<212> PRT
<213> E. coli

<400> 101

```
Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro Ile Gly Gly Gly
1               5                   10                  15


Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Ala Val Asn Val Gly Gln
            20                  25                  30


Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys His Asn Asp Tyr
        35                  40                  45


Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg Gly Ser Ala Tyr
    50                  55                  60


Gly Gly Val Leu Ser Ser Phe Ser Gly Thr Val Lys Tyr Asn Gly Ser
65                  70                  75                  80


Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg Val Val Tyr Asn
                85                  90                  95


Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr Leu Thr Pro Val
            100                 105                 110
```

```
Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser Leu Ile Ala Val
        115                 120                 125

Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp Asp Phe Gln Phe
    130                 135                 140

Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val Pro Thr Gly Gly
145                 150                 155                 160

Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu Pro Asp Tyr Pro
                165                 170                 175

Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala Lys Ser Gln Asn
            180                 185                 190

Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp Ala Gly Asn Ser Ile
        195                 200                 205

Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly Val Gly Val Gln
    210                 215                 220

Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn Thr Val Ser Leu
225                 230                 235                 240

Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu Thr Ala Asn Tyr
                245                 250                 255

Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val Gln Ser Ile Ile
            260                 265                 270

Gly Val Thr Phe Val Tyr Gln
        275
```

<210> 102
<211> 303
<212> PRT
<213> E. coli

<400> 102

```
Met Ile Val Met Lys Arg Val Ile Thr Leu Phe Ala Val Leu Leu Met
1               5                  10                  15

Gly Trp Ser Val Asn Ala Trp Ser Phe Ala Cys Lys Thr Ala Asn Gly
            20                  25                  30

Thr Ala Ile Pro Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn Leu
        35                  40                  45

Ala Pro Ala Val Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser Thr
    50                  55                  60

Gln Ile Phe Cys His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr Val
65                  70                  75                  80

Thr Leu Gln Arg Gly Ser Ala Tyr Gly Gly Val Leu Ser Ser Phe Ser
                85                  90                  95

Gly Thr Val Lys Tyr Asn Gly Ser Ser Tyr Pro Phe Pro Thr Thr Ser
            100                 105                 110

Glu Thr Pro Arg Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp Pro
        115                 120                 125

Val Ala Leu Tyr Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala Ile
    130                 135                 140

Lys Ala Gly Ser Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn Asn
145                 150                 155                 160

Tyr Asn Ser Asp Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn Asn
                165                 170                 175

Asp Val Val Val Pro Thr Gly Gly Cys Asp Ala Ser Ala Arg Asp Val
            180                 185                 190
```

```
Thr Val Thr Leu Pro Asp Tyr Pro Gly Ser Val Pro Ile Pro Leu Thr
        195                 200                 205

Val Tyr Cys Ala Lys Ser Gln Asn Leu Gly Tyr Tyr Leu Ser Gly Thr
    210                 215                 220

Thr Ala Asp Ala Gly Asn Ser Ile Phe Thr Asn Thr Ala Ser Phe Ser
225                 230                 235                 240

Pro Ala Gln Gly Val Gly Val Gln Leu Thr Arg Asn Gly Thr Ile Ile
                245                 250                 255

Pro Ala Asn Asn Thr Val Ser Leu Gly Ala Val Gly Thr Ser Ala Val
            260                 265                 270

Ser Leu Gly Leu Thr Ala Asn Tyr Ala Arg Thr Gly Gly Gln Val Thr
        275                 280                 285

Ala Gly Asn Val Gln Ser Ile Ile Gly Val Thr Phe Val Tyr Gln
    290                 295                 300
```

<210> 103
<211> 279
<212> PRT
<213> E. coli

<400> 103

```
Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro Ile Gly Gly Gly
1               5                   10                  15

Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Ala Val Asn Val Gly Gln
            20                  25                  30

Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys His Asn Asp Tyr
        35                  40                  45

Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg Gly Ser Ala Tyr
    50                  55                  60
```

```
Gly Gly Val Leu Ser Ser Phe Ser Gly Thr Val Lys Tyr Asn Gly Ser
65                  70                  75                  80

Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg Val Val Tyr Asn
                85                  90                  95

Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr Leu Thr Pro Val
            100                 105                 110

Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser Leu Ile Ala Val
        115                 120                 125

Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp Asp Phe Gln Phe
    130                 135                 140

Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val Pro Thr Gly Gly
145                 150                 155                 160

Cys Asp Ala Ser Ala Arg Asp Val Thr Val Thr Leu Pro Asp Tyr Pro
                165                 170                 175

Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala Lys Ser Gln Asn
            180                 185                 190

Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp Ala Gly Asn Ser Ile
        195                 200                 205

Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly Val Gly Val Gln
    210                 215                 220

Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn Thr Val Ser Leu
225                 230                 235                 240

Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu Thr Ala Asn Tyr
                245                 250                 255
```

```
Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val Gln Ser Ile Ile
            260                 265                 270


Gly Val Thr Phe Val Tyr Gln
        275


<210>  104
<211>  303
<212>  PRT
<213>  E. coli

<400>  104

Met Ile Val Met Lys Arg Val Ile Thr Leu Phe Ala Val Leu Leu Met
1               5                  10                  15


Gly Trp Ser Val Asn Ala Trp Ser Phe Ala Cys Lys Thr Ala Asn Gly
            20                  25                  30


Thr Ala Ile Pro Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn Leu
        35                  40                  45


Ala Pro Val Val Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser Thr
    50                  55                  60


Gln Ile Phe Cys His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr Val
65                  70                  75                  80


Thr Leu Gln Arg Gly Ser Ala Tyr Gly Gly Val Leu Ser Asn Phe Ser
                85                  90                  95


Gly Thr Val Lys Tyr Ser Gly Ser Ser Tyr Pro Phe Pro Thr Thr Ser
            100                 105                 110


Glu Thr Pro Arg Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp Pro
        115                 120                 125


Val Ala Leu Tyr Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala Ile
    130                 135                 140
```

```
Lys Ala Gly Ser Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn Asn
145                 150                 155                 160


Tyr Asn Ser Asp Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn Asn
                165                 170                 175


Asp Val Val Val Pro Thr Gly Gly Cys Asp Val Ser Ala Arg Asp Val
            180                 185                 190


Thr Val Thr Leu Pro Asp Tyr Pro Gly Ser Val Pro Ile Pro Leu Thr
        195                 200                 205


Val Tyr Cys Ala Lys Ser Gln Asn Leu Gly Tyr Tyr Leu Ser Gly Thr
    210                 215                 220


Thr Ala Asp Ala Gly Asn Ser Ile Phe Thr Asn Thr Ala Ser Phe Ser
225                 230                 235                 240


Pro Ala Gln Gly Val Gly Val Gln Leu Thr Arg Asn Gly Thr Ile Ile
                245                 250                 255


Pro Ala Asn Asn Thr Val Ser Leu Gly Ala Val Gly Thr Ser Ala Val
            260                 265                 270


Ser Leu Gly Leu Thr Ala Asn Tyr Ala Arg Thr Gly Gly Gln Val Thr
        275                 280                 285


Ala Gly Asn Val Gln Ser Ile Ile Gly Val Thr Phe Val Tyr Gln
    290                 295                 300


<210>  105
<211>  279
<212>  PRT
<213>  E. coli

<400>  105

Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro Ile Gly Gly Gly
1               5                   10                  15
```

```
Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val Val Asn Val Gly Gln
            20                  25                  30

Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys His Asn Asp Tyr
        35                  40                  45

Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg Gly Ser Ala Tyr
    50                  55                  60

Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val Lys Tyr Ser Gly Ser
65                  70                  75                  80

Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg Val Val Tyr Asn
                85                  90                  95

Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr Leu Thr Pro Val
            100                 105                 110

Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser Leu Ile Ala Val
        115                 120                 125

Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp Asp Phe Gln Phe
    130                 135                 140

Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val Pro Thr Gly Gly
145                 150                 155                 160

Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu Pro Asp Tyr Pro
                165                 170                 175

Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala Lys Ser Gln Asn
            180                 185                 190

Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp Ala Gly Asn Ser Ile
        195                 200                 205
```

```
Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly Val Gly Val Gln
    210                 215                 220

Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn Thr Val Ser Leu
225                 230                 235                 240

Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu Thr Ala Asn Tyr
                245                 250                 255

Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val Gln Ser Ile Ile
            260                 265                 270

Gly Val Thr Phe Val Tyr Gln
        275
```

<210> 106
<211> 300
<212> PRT
<213> E. coli

<400> 106

```
Met Lys Arg Val Ile Thr Leu Phe Ala Val Leu Leu Met Gly Trp Ser
1               5                   10                  15

Val Asn Ala Trp Ser Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile
            20                  25                  30

Pro Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Ala
        35                  40                  45

Val Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe
    50                  55                  60

Cys His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln
65                  70                  75                  80

Arg Gly Ala Ala Tyr Gly Gly Val Leu Ser Ser Phe Ser Gly Thr Val
                85                  90                  95
```

```
Lys Tyr Asn Gly Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro
            100                 105                 110

Arg Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu
        115                 120                 125

Tyr Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly
    130                 135                 140

Ser Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser
145                 150                 155                 160

Asp Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val
                165                 170                 175

Val Pro Thr Gly Gly Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr
            180                 185                 190

Leu Pro Asp Tyr Pro Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys
        195                 200                 205

Ala Lys Ser Gln Asn Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp
    210                 215                 220

Ala Gly Asn Ser Ile Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln
225                 230                 235                 240

Gly Val Gly Val Gln Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn
                245                 250                 255

Asn Thr Val Ser Leu Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly
            260                 265                 270

Leu Thr Ala Asn Tyr Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn
        275                 280                 285

Val Gln Ser Ile Ile Gly Val Thr Phe Val Tyr Gln
```

290 295 300

<210> 107
<211> 279
<212> PRT
<213> E. coli

<400> 107

Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro Ile Gly Gly Gly
1 5 10 15

Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Ala Val Asn Val Gly Gln
20 25 30

Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys His Asn Asp Tyr
35 40 45

Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg Gly Ala Ala Tyr
50 55 60

Gly Gly Val Leu Ser Ser Phe Ser Gly Thr Val Lys Tyr Asn Gly Ser
65 70 75 80

Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg Val Val Tyr Asn
85 90 95

Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr Leu Thr Pro Val
100 105 110

Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser Leu Ile Ala Val
115 120 125

Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp Asp Phe Gln Phe
130 135 140

Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val Pro Thr Gly Gly
145 150 155 160

```
Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu Pro Asp Tyr Pro
                165                 170                 175


Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala Lys Ser Gln Asn
            180                 185                 190


Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp Ala Gly Asn Ser Ile
        195                 200                 205


Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly Val Gly Val Gln
    210                 215                 220


Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn Thr Val Ser Leu
225                 230                 235                 240


Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu Thr Ala Asn Tyr
                245                 250                 255


Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val Gln Ser Ile Ile
            260                 265                 270


Gly Val Thr Phe Val Tyr Gln
        275


<210>  108
<211>  20
<212>  PRT
<213>  Artificial

<220>
<223>  Leader

<400>  108

Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1               5                   10                  15


Gly Ser Thr Gly
            20
```

<210> 109
<211> 163
<212> PRT
<213> H. pylori

<400> 109

Asp Ile Ile Lys Leu Leu Asn Glu Gln Val Asn Lys Glu Met Asn Ser
1 5 10 15

Ser Asn Leu Tyr Met Ser Met Ser Ser Trp Cys Tyr Thr His Ser Leu
20 25 30

Asp Gly Ala Gly Leu Phe Leu Phe Asp His Ala Ala Glu Glu Tyr Glu
35 40 45

His Ala Lys Lys Leu Ile Ile Phe Leu Asn Glu Asn Asn Val Pro Val
50 55 60

Gln Leu Thr Ser Ile Ser Ala Pro Glu His Lys Phe Glu Gly Leu Thr
65 70 75 80

Gln Ile Phe Gln Lys Ala Tyr Glu His Glu Gln His Ile Ser Glu Ser
85 90 95

Ile Asn Asn Ile Val Asp His Ala Ile Lys Ser Lys Asp His Ala Thr
100 105 110

Phe Asn Phe Leu Gln Trp Tyr Val Ala Glu Gln His Glu Glu Glu Val
115 120 125

Leu Phe Lys Asp Ile Leu Asp Lys Ile Glu Leu Ile Gly Asn Glu Asn
130 135 140

His Gly Leu Tyr Leu Ala Asp Gln Tyr Val Lys Gly Ile Ala Lys Ser
145 150 155 160

Arg Lys Ser

<210> 110
<211> 148
<212> PRT
<213> Artificial

<220>
<223> HBC

<400> 110

```
Met Asp Ile Asp Pro Tyr Lys Glu Phe Gly Ala Thr Val Glu Leu Leu
1               5                  10                  15


Ser Phe Leu Pro Ser Asp Phe Phe Pro Ser Val Arg Asp Leu Leu Asp
            20                  25                  30


Thr Ala Ser Ala Leu Tyr Arg Glu Ala Leu Glu Ser Pro Glu His Cys
        35                  40                  45


Ser Pro His His Thr Ala Leu Arg Gln Ala Ile Leu Cys Trp Gly Glu
    50                  55                  60


Leu Met Thr Leu Ala Thr Trp Val Gly Asn Asn Leu Glu Asp Ala Ser
65                  70                  75                  80


Arg Asp Leu Val Val Asn Tyr Val Asn Thr Asn Met Gly Leu Lys Ile
                85                  90                  95


Arg Gln Leu Leu Trp Phe His Ile Ser Cys Leu Thr Phe Gly Arg Glu
            100                 105                 110


Thr Val Leu Glu Tyr Leu Val Ser Phe Gly Val Trp Ile Arg Thr Pro
        115                 120                 125


Pro Ala Tyr Arg Pro Pro Asn Ala Pro Ile Leu Ser Thr Leu Pro Glu
    130                 135                 140


Thr Thr Val Val
145
```

<210> 111
<211> 133
<212> PRT
<213> Artificial

<220>
<223> Qbeta

<400> 111

```
Met Ala Lys Leu Glu Thr Val Thr Leu Gly Asn Ile Gly Lys Asp Gly
1               5                   10                  15


Lys Gln Thr Leu Val Leu Asn Pro Arg Gly Val Asn Pro Thr Asn Gly
            20                  25                  30


Val Ala Ser Leu Ser Gln Ala Gly Ala Val Pro Ala Leu Glu Lys Arg
        35                  40                  45


Val Thr Val Ser Val Ser Gln Pro Ser Arg Asn Arg Lys Asn Tyr Lys
    50                  55                  60


Val Gln Val Lys Ile Gln Asn Pro Thr Ala Cys Thr Ala Asn Gly Ser
65                  70                  75                  80


Cys Asp Pro Ser Val Thr Arg Gln Ala Tyr Ala Asp Val Thr Phe Ser
                85                  90                  95


Phe Thr Gln Tyr Ser Thr Asp Glu Glu Arg Ala Phe Val Arg Thr Glu
            100                 105                 110


Leu Ala Ala Leu Leu Ala Ser Pro Leu Leu Ile Asp Ala Ile Asp Gln
        115                 120                 125


Leu Asn Pro Ala Tyr
    130
```

<210> 112
<211> 9
<212> PRT

<213> Artificial

<220>
<223> Linker

<400> 112

Gly Ser Ser Gly Ser Gly Ser Gly Ser
1   5

<210> 113
<211> 5
<212> PRT
<213> Artificial

<220>
<223> Linker

<400> 113

Gly Gly Ser Gly Ser
1   5

<210> 114
<211> 14
<212> PRT
<213> Artificial

<220>
<223> Linker

<400> 114

Ser Gly Ser His His His His His His His His Gly Gly Ser
1   5   10

<210> 115
<211> 13
<212> PRT
<213> Artificial

<220>
<223> Linker

<400> 115

Ala Lys Phe Val Ala Ala Trp Thr Leu Lys Ala Ala Ala

1 5 10

<210> 116
<211> 16
<212> PRT
<213> Artificial

<220>
<223> Linker

<400> 116

Gly Gly Gly Gly Ser Leu Val Pro Arg Gly Ser Gly Gly Gly Gly Ser
1 5 10 15

<210> 117
<211> 16
<212> PRT
<213> Artificial

<220>
<223> Linker

<400> 117

Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Ala
1 5 10 15

<210> 118
<211> 17
<212> PRT
<213> Artificial

<220>
<223> Linker

<400> 118

Ser Gly Ser Phe Val Ala Ala Trp Thr Leu Lys Ala Ala Ala Gly Gly
1 5 10 15

Ser

<210> 119

<211> 12
<212> PRT
<213> Artificial

<220>
<223> Linker

<400> 119

Ser Gly Ser Gly Ser Gly Gly Gly Gly Gly Gly Ser
1   5   10

<210> 120
<211> 307
<212> PRT
<213> Artificial

<220>
<223> FimH_DNKQ_DG citopl pET22b+

<400> 120

Met Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro Ile Gly Gly
1   5   10   15

Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val Val Asn Val Gly
            20   25   30

Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys His Asn Asp
        35   40   45

Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg Gly Ser Ala
    50   55   60

Tyr Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val Lys Tyr Ser Gly
65   70   75   80

Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg Val Val Tyr
                85   90   95

Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr Leu Thr Pro
            100   105   110

```
Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser Leu Ile Ala
        115                 120                 125

Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp Asp Phe Gln
    130                 135                 140

Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val Pro Thr Gly
145                 150                 155                 160

Gly Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu Pro Asp Tyr
                165                 170                 175

Pro Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala Lys Ser Gln
            180                 185                 190

Asn Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp Ala Gly Asn Ser
        195                 200                 205

Ile Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly Val Gly Val
    210                 215                 220

Gln Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn Thr Val Ser
225                 230                 235                 240

Leu Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu Thr Ala Asn
                245                 250                 255

Tyr Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val Gln Ser Ile
            260                 265                 270

Ile Gly Val Thr Phe Val Tyr Gln Asp Asn Lys Gln Ala Asp Val Thr
        275                 280                 285

Ile Thr Val Asn Gly Lys Val Val Ala Lys Gly Ser Gly His His His
    290                 295                 300

His His His
```

305

<210> 121
<211> 308
<212> PRT
<213> Artificial

<220>
<223> FimH_PGDGN_DG citopl pET22b+

<400> 121

```
Met Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro Ile Gly Gly
1               5                   10                  15


Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val Val Asn Val Gly
            20                  25                  30


Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys His Asn Asp
        35                  40                  45


Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg Gly Ser Ala
    50                  55                  60


Tyr Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val Lys Tyr Ser Gly
65                  70                  75                  80


Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg Val Val Tyr
                85                  90                  95


Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr Leu Thr Pro
            100                 105                 110


Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser Leu Ile Ala
        115                 120                 125


Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp Asp Phe Gln
    130                 135                 140


Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val Pro Thr Gly
```

```
145                 150                 155                 160

Gly Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu Pro Asp Tyr
                165                 170                 175

Pro Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala Lys Ser Gln
            180                 185                 190

Asn Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp Ala Gly Asn Ser
        195                 200                 205

Ile Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly Val Gly Val
    210                 215                 220

Gln Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn Thr Val Ser
225                 230                 235                 240

Leu Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu Thr Ala Asn
                245                 250                 255

Tyr Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val Gln Ser Ile
            260                 265                 270

Ile Gly Val Thr Phe Val Tyr Gln Pro Gly Asp Gly Asn Ala Asp Val
        275                 280                 285

Thr Ile Thr Val Asn Gly Lys Val Val Ala Lys Gly Ser Gly His His
    290                 295                 300

His His His His
305
```

<210> 122
<211> 306
<212> PRT
<213> Artificial

<220>
<223> FimH_DGG_PGDGN_DG citopl pET22b+

<400> 122

```
Met Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro Ile Gly Gly
1               5                  10                  15

Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val Val Asn Val Gly
            20                  25                  30

Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys His Asn Asp
        35                  40                  45

Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg Gly Ser Ala
    50                  55                  60

Tyr Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val Lys Tyr Ser Gly
65                  70                  75                  80

Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg Val Val Tyr
                85                  90                  95

Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr Leu Thr Pro
            100                 105                 110

Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser Leu Ile Ala
        115                 120                 125

Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp Asp Phe Gln
    130                 135                 140

Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val Pro Thr Cys
145                 150                 155                 160

Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu Pro Asp Tyr Pro Gly
                165                 170                 175

Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala Lys Ser Gln Asn Leu
            180                 185                 190
```

```
Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp Ala Gly Asn Ser Ile Phe
        195                 200                 205


Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly Val Gly Val Gln Leu
    210                 215                 220


Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn Thr Val Ser Leu Gly
225                 230                 235                 240


Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu Thr Ala Asn Tyr Ala
                245                 250                 255


Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val Gln Ser Ile Ile Gly
            260                 265                 270


Val Thr Phe Val Tyr Gln Pro Gly Asp Gly Asn Ala Asp Val Thr Ile
        275                 280                 285


Thr Val Asn Gly Lys Val Val Ala Lys Gly Ser Gly His His His His
    290                 295                 300


His His
305


<210>  123
<211>  299
<212>  PRT
<213>  Artificial

<220>
<223>  FimH_PGDGN_DG citopl PET24 Tagless

<400>  123

Met Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro Ile Gly Gly
1               5                   10                  15


Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val Val Asn Val Gly
            20                  25                  30
```

```
Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys His Asn Asp
        35                  40                  45

Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg Gly Ser Ala
    50                  55                  60

Tyr Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val Lys Tyr Ser Gly
65                  70                  75                  80

Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg Val Val Tyr
                85                  90                  95

Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr Leu Thr Pro
            100                 105                 110

Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser Leu Ile Ala
        115                 120                 125

Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp Asp Phe Gln
    130                 135                 140

Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val Pro Thr Gly
145                 150                 155                 160

Gly Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu Pro Asp Tyr
                165                 170                 175

Pro Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala Lys Ser Gln
            180                 185                 190

Asn Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp Ala Gly Asn Ser
        195                 200                 205

Ile Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly Val Gly Val
    210                 215                 220

Gln Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn Thr Val Ser
```

225 230 235 240

Leu Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu Thr Ala Asn
245 250 255

Tyr Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val Gln Ser Ile
260 265 270

Ile Gly Val Thr Phe Val Tyr Gln Pro Gly Asp Gly Asn Ala Asp Val
275 280 285

Thr Ile Thr Val Asn Gly Lys Val Val Ala Lys
290 295

<210> 124
<211> 298
<212> PRT
<213> Artificial

<220>
<223> FimH_PGDGN_DG citopl tagless no Met

<400> 124

Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro Ile Gly Gly Gly
1 5 10 15

Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val Val Asn Val Gly Gln
20 25 30

Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys His Asn Asp Tyr
35 40 45

Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg Gly Ser Ala Tyr
50 55 60

Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val Lys Tyr Ser Gly Ser
65 70 75 80

Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg Val Val Tyr Asn

```
                85                  90                  95

Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr Leu Thr Pro Val
            100                 105                 110

Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser Leu Ile Ala Val
        115                 120                 125

Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp Asp Phe Gln Phe
    130                 135                 140

Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val Pro Thr Gly Gly
145                 150                 155                 160

Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu Pro Asp Tyr Pro
                165                 170                 175

Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala Lys Ser Gln Asn
            180                 185                 190

Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp Ala Gly Asn Ser Ile
        195                 200                 205

Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly Val Gly Val Gln
    210                 215                 220

Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn Thr Val Ser Leu
225                 230                 235                 240

Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu Thr Ala Asn Tyr
                245                 250                 255

Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val Gln Ser Ile Ile
            260                 265                 270

Gly Val Thr Phe Val Tyr Gln Pro Gly Asp Gly Asn Ala Asp Val Thr
        275                 280                 285
```

Ile Thr Val Asn Gly Lys Val Val Ala Lys
290 295

<210> 125
<211> 188
<212> PRT
<213> Artificial

<220>
<223> VH of mAb 926

<400> 125

Gln Val Gln Leu Gln Gln Ser Gly Ala Glu Leu Ala Thr Pro Gly Ala
1 5 10 15

Ser Val Lys Met Ser Cys Lys Ala Ser Gly Tyr Thr Ser Thr Asn Tyr
20 25 30

Trp Ile His Trp Val Lys Gln Arg Pro Gly Gln Gly Leu Glu Trp Ile
35 40 45

Gly Tyr Ile Asn Pro Thr Ser Gly Tyr Thr Glu Tyr Asn Gln Asn Phe
50 55 60

Lys Asp Lys Ala Thr Leu Thr Ala Asp Lys Ser Ser Ser Thr Ala Tyr
65 70 75 80

Met Gln Leu Thr Ser Leu Thr Ser Glu Asp Ser Ala Val Tyr Tyr Cys
85 90 95

Ala Arg Gly Val Ile Arg Asp Phe Trp Gly Gln Gly Thr Thr Leu Thr
100 105 110

Val Ser Ser Ala Lys Thr Thr Ala Pro Ser Val Tyr Pro Leu Ala Pro
115 120 125

Val Cys Gly Asp Thr Thr Gly Ser Ser Val Thr Leu Gly Cys Leu Val
130 135 140

```
Lys Gly Tyr Phe Pro Glu Pro Val Thr Leu Thr Trp Asn Ser Gly Ser
145                 150                 155                 160


Leu Ser Ser Gly Val His Thr Phe Pro Ala Val Leu Gln Ser Asp Leu
                165                 170                 175


Tyr Thr Leu Ser Ser Ser Val Thr Val Thr Ser Ser
            180                 185


<210>  126
<211>  112
<212>  PRT
<213>  Artificial

<220>
<223>  VL (kappa) of mAb 926

<400>  126

Asp Val Leu Met Thr Gln Thr Pro Leu Ser Leu Pro Val Ser Leu Gly
1               5                   10                  15


Asp Gln Ala Ser Ile Ser Cys Arg Ser Ser Gln Asn Ile Val His Asn
            20                  25                  30


Asn Gly Asn Thr Tyr Leu Glu Trp Tyr Leu Gln Ser Pro Gly Gln Ser
        35                  40                  45


Pro Lys Leu Leu Ile Tyr Lys Val Ser Asn Arg Phe Ser Gly Val Pro
    50                  55                  60


Asp Arg Phe Ser Gly Ser Gly Ser Gly Thr Asp Phe Thr Leu Lys Ile
65                  70                  75                  80


Ser Arg Val Glu Ala Glu Asp Leu Gly Val Tyr Tyr Cys Phe Gln Gly
                85                  90                  95


Ser His Val Pro Phe Thr Phe Gly Ser Gly Thr Lys Leu Glu Ile Lys
            100                 105                 110
```

<210> 127
<211> 191
<212> PRT
<213> Artificial

<220>
<223> VH of mAb 475

<400> 127

```
Gln Val Gln Leu Gln Gln Ser Gly Ala Glu Leu Val Arg Pro Gly Ser
1               5                   10                  15


Ser Val Lys Ile Ser Cys Lys Ala Ser Gly Tyr Ala Phe Ser Ser Tyr
            20                  25                  30


Trp Met Asn Trp Val Lys Gln Arg Pro Gly Gln Gly Leu Glu Trp Ile
        35                  40                  45


Gly Gln Ile Tyr Pro Arg Asp Gly Asp Thr Asn Tyr Asn Gly Lys Phe
    50                  55                  60


Met Asp Lys Val Thr Leu Thr Ala Asp Lys Ser Ser Asn Thr Ala Tyr
65                  70                  75                  80


Met Gln Leu Ser Ser Leu Thr Ser Glu Asp Ser Ala Val Tyr Phe Cys
                85                  90                  95


Glu Val Gly Arg Gly Phe Tyr Gly Met Asp Tyr Trp Gly Gln Gly Thr
            100                 105                 110


Ser Val Thr Val Ser Ser Ala Lys Thr Thr Ala Pro Ser Val Tyr Pro
        115                 120                 125


Leu Ala Pro Val Cys Gly Asp Thr Thr Gly Ser Ser Val Thr Leu Gly
    130                 135                 140


Cys Leu Val Lys Gly Tyr Phe Pro Glu Pro Val Thr Leu Thr Trp Asn
145                 150                 155                 160
```

Ser Gly Ser Leu Ser Ser Gly Val His Thr Phe Pro Ala Val Leu Gln
                165   170   175

Ser Asp Leu Tyr Thr Leu Ser Ser Ser Val Thr Val Thr Ser Ser
            180   185   190

<210> 128
<211> 106
<212> PRT
<213> Artificial

<220>
<223> VL (kappa) of mAb 475

<400> 128

Asp Ile Val Met Thr Gln Ser Pro Lys Phe Met Ser Thr Ser Val Gly
1   5   10   15

Asp Arg Val Ser Val Thr Cys Lys Ala Ser Gln Asn Val Ser Asn Val
            20   25   30

Ala Trp Tyr Gln Gln Lys Pro Gly Gln Ser Pro Lys Ala Met Ile Tyr
        35   40   45

Ser Ala Ser Tyr Arg Tyr Ser Gly Val Pro Gly Arg Phe Thr Gly Ser
    50   55   60

Gly Ser Gly Thr Asp Phe Thr Leu Thr Ile Asn Asn Val Gln Ser Glu
65   70   75   80

Asp Leu Ala Thr Tyr Phe Cys Gln Gln Asn Ser Ser Phe Pro Phe Thr
                85   90   95

Phe Gly Gly Gly Thr Lys Leu Glu Ile Lys
            100   105

<210> 129
<211> 487
<212> PRT

<213> Artificial

<220>
<223> D_FimHDG_Fer_GSG4

<400> 129

```
Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1               5                  10                  15


Gly Ser Thr Gly Asp Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile
            20                  25                  30


Pro Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val
        35                  40                  45


Val Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe
    50                  55                  60


Cys His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln
65                  70                  75                  80


Arg Gly Ser Ala Tyr Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val
                85                  90                  95


Lys Tyr Ser Gly Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro
            100                 105                 110


Arg Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu
        115                 120                 125


Tyr Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly
    130                 135                 140


Ser Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser
145                 150                 155                 160


Asp Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val
                165                 170                 175
```

```
Val Pro Thr Gly Gly Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr
            180                 185                 190

Leu Pro Asp Tyr Pro Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys
        195                 200                 205

Ala Lys Ser Gln Asn Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp
    210                 215                 220

Ala Gly Asn Ser Ile Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln
225                 230                 235                 240

Gly Val Gly Val Gln Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn
                245                 250                 255

Asn Thr Val Ser Leu Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly
            260                 265                 270

Leu Thr Ala Asn Tyr Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn
        275                 280                 285

Val Gln Ser Ile Ile Gly Val Thr Phe Val Tyr Gln Pro Gly Asp Gly
    290                 295                 300

Asn Ala Asp Val Thr Ile Thr Val Asn Gly Lys Val Val Ala Lys Gly
305                 310                 315                 320

Ser Gly Gly Gly Gly Asp Ile Ile Lys Leu Leu Asn Glu Gln Val Asn
                325                 330                 335

Lys Glu Met Asn Ser Ser Asn Leu Tyr Met Ser Met Ser Ser Trp Cys
            340                 345                 350

Tyr Thr His Ser Leu Asp Gly Ala Gly Leu Phe Leu Phe Asp His Ala
        355                 360                 365

Ala Glu Glu Tyr Glu His Ala Lys Lys Leu Ile Ile Phe Leu Asn Glu
```

```
    370                 375                 380

Asn Asn Val Pro Val Gln Leu Thr Ser Ile Ser Ala Pro Glu His Lys
385                 390                 395                 400

Phe Glu Gly Leu Thr Gln Ile Phe Gln Lys Ala Tyr Glu His Glu Gln
                405                 410                 415

His Ile Ser Glu Ser Ile Asn Asn Ile Val Asp His Ala Ile Lys Ser
            420                 425                 430

Lys Asp His Ala Thr Phe Asn Phe Leu Gln Trp Tyr Val Ala Glu Gln
        435                 440                 445

His Glu Glu Glu Val Leu Phe Lys Asp Ile Leu Asp Lys Ile Glu Leu
    450                 455                 460

Ile Gly Asn Glu Asn His Gly Leu Tyr Leu Ala Asp Gln Tyr Val Lys
465                 470                 475                 480

Gly Ile Ala Lys Ser Arg Lys
                485
```

<210> 130
<211> 492
<212> PRT
<213> Artificial

<220>
<223> D_FimHDG_Fer0.5_deglyc_tagless

<400> 130

```
Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1               5                   10                  15

Gly Ser Thr Gly Asp Phe Ala Cys Lys Thr Ala Gln Gly Thr Ala Ile
            20                  25                  30

Pro Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val
```

```
        35                  40                  45

Val Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe
    50                  55                  60

Cys His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln
65                  70                  75                  80

Arg Gly Ser Ala Tyr Gly Gly Val Leu Ser Gln Phe Ser Gly Thr Val
                85                  90                  95

Lys Tyr Ser Gly Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro
            100                 105                 110

Arg Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu
        115                 120                 125

Tyr Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly
    130                 135                 140

Ser Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser
145                 150                 155                 160

Asp Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val
                165                 170                 175

Val Pro Thr Gly Gly Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr
            180                 185                 190

Leu Pro Asp Tyr Pro Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys
        195                 200                 205

Ala Lys Ser Gln Asn Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp
    210                 215                 220

Ala Gly Asn Ser Ile Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln
225                 230                 235                 240
```

```
Gly Val Gly Val Gln Leu Thr Arg Gln Gly Thr Ile Ile Pro Ala Gln
                245                 250                 255

Asn Thr Val Ser Leu Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly
            260                 265                 270

Leu Thr Ala Asn Tyr Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn
        275                 280                 285

Val Gln Ser Ile Ile Gly Val Thr Phe Val Tyr Gln Pro Gly Asp Gly
    290                 295                 300

Asn Ala Asp Val Thr Ile Thr Val Asn Gly Lys Val Val Ala Lys Gly
305                 310                 315                 320

Ser Gly Gly Gly Gly Gly Gly Met Leu Lys Pro Glu Met Ile Glu Lys
                325                 330                 335

Leu Asn Glu Gln Met Asn Leu Glu Leu Tyr Ser Ser Leu Leu Tyr Gln
            340                 345                 350

Gln Met Ser Ala Trp Cys Ser Tyr His Gly Phe Glu Gly Ala Ala Ala
        355                 360                 365

Phe Leu Arg Arg His Ala Gln Glu Glu Met Thr His Met Gln Arg Leu
    370                 375                 380

Phe Asp Tyr Leu Thr Asp Thr Gly Asn Leu Pro Arg Ile Asp Thr Ile
385                 390                 395                 400

Pro Ser Pro Phe Ala Glu Tyr Ser Ser Leu Asp Glu Leu Phe Gln Glu
                405                 410                 415

Thr Tyr Lys His Glu Gln Leu Ile Thr Gln Lys Ile Asn Glu Leu Ala
            420                 425                 430

His Ala Ala Met Thr Asn Gln Asp Tyr Pro Thr Phe Asn Phe Leu Gln
```

```
        435                 440                 445

Trp Tyr Val Ala Glu Gln His Glu Glu Glu Lys Leu Phe Lys Ser Ile
    450                 455                 460

Ile Asp Lys Leu Ser Leu Ala Gly Lys Ser Gly Glu Gly Leu Tyr Phe
465                 470                 475                 480

Ile Asp Lys Glu Leu Ser Thr Leu Asp Thr Gln Asn
                485                 490
```

<210> 131
<211> 318
<212> PRT
<213> Artificial

<220>
<223> FimHDG_deglyc_tagless

<400> 131

```
Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1               5                   10                  15

Gly Ser Thr Gly Phe Ala Cys Lys Thr Ala Gln Gly Thr Ala Ile Pro
            20                  25                  30

Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val Val
        35                  40                  45

Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys
    50                  55                  60

His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg
65                  70                  75                  80

Gly Ser Ala Tyr Gly Gly Val Leu Ser Gln Phe Ser Gly Thr Val Lys
                85                  90                  95

Tyr Ser Gly Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg
```

```
            100                 105                 110

Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr
        115                 120                 125

Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser
    130                 135                 140

Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp
145                 150                 155                 160

Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val
                165                 170                 175

Pro Thr Gly Gly Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu
            180                 185                 190

Pro Asp Tyr Pro Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala
        195                 200                 205

Lys Ser Gln Asn Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp Ala
    210                 215                 220

Gly Asn Ser Ile Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly
225                 230                 235                 240

Val Gly Val Gln Leu Thr Arg Gln Gly Thr Ile Ile Pro Ala Gln Asn
                245                 250                 255

Thr Val Ser Leu Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu
            260                 265                 270

Thr Ala Asn Tyr Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val
        275                 280                 285

Gln Ser Ile Ile Gly Val Thr Phe Val Tyr Gln Pro Gly Asp Gly Asn
    290                 295                 300
```

Ala Asp Val Thr Ile Thr Val Asn Gly Lys Val Val Ala Lys
305 310 315

<210> 132
<211> 319
<212> PRT
<213> Artificial

<220>
<223> D_FimHDG_deglyc_tagless

<400> 132

Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1 5 10 15

Gly Ser Thr Gly Asp Phe Ala Cys Lys Thr Ala Gln Gly Thr Ala Ile
20 25 30

Pro Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val
35 40 45

Val Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe
50 55 60

Cys His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln
65 70 75 80

Arg Gly Ser Ala Tyr Gly Gly Val Leu Ser Gln Phe Ser Gly Thr Val
85 90 95

Lys Tyr Ser Gly Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro
100 105 110

Arg Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu
115 120 125

Tyr Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly
130 135 140

```
Ser Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser
145                 150                 155                 160


Asp Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val
                165                 170                 175


Val Pro Thr Gly Gly Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr
            180                 185                 190


Leu Pro Asp Tyr Pro Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys
        195                 200                 205


Ala Lys Ser Gln Asn Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp
    210                 215                 220


Ala Gly Asn Ser Ile Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln
225                 230                 235                 240


Gly Val Gly Val Gln Leu Thr Arg Gln Gly Thr Ile Ile Pro Ala Gln
                245                 250                 255


Asn Thr Val Ser Leu Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly
            260                 265                 270


Leu Thr Ala Asn Tyr Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn
        275                 280                 285


Val Gln Ser Ile Ile Gly Val Thr Phe Val Tyr Gln Pro Gly Asp Gly
    290                 295                 300


Asn Ala Asp Val Thr Ile Thr Val Asn Gly Lys Val Val Ala Lys
305                 310                 315


<210>  133
<211>  318
<212>  PRT
<213>  Artificial
```

<220>
<223> FimHDG_N7Q_tagless

<400> 133

```
Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1               5                  10                  15


Gly Ser Thr Gly Phe Ala Cys Lys Thr Ala Gln Gly Thr Ala Ile Pro
            20                  25                  30


Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val Val
        35                  40                  45


Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys
    50                  55                  60


His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg
65                  70                  75                  80


Gly Ser Ala Tyr Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val Lys
                85                  90                  95


Tyr Ser Gly Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg
            100                 105                 110


Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr
        115                 120                 125


Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser
    130                 135                 140


Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp
145                 150                 155                 160


Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val
                165                 170                 175


Pro Thr Gly Gly Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu
```

```
            180                 185                 190

Pro Asp Tyr Pro Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala
        195                 200                 205

Lys Ser Gln Asn Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp Ala
    210                 215                 220

Gly Asn Ser Ile Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly
225                 230                 235                 240

Val Gly Val Gln Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn
                245                 250                 255

Thr Val Ser Leu Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu
            260                 265                 270

Thr Ala Asn Tyr Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val
        275                 280                 285

Gln Ser Ile Ile Gly Val Thr Phe Val Tyr Gln Pro Gly Asp Gly Asn
    290                 295                 300

Ala Asp Val Thr Ile Thr Val Asn Gly Lys Val Val Ala Lys
305                 310                 315


<210>  134
<211>  493
<212>  PRT
<213>  Artificial

<220>
<223>  D_FimHDG_Fer_deglyc_tagless

<400>  134

Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1               5                   10                  15

Gly Ser Thr Gly Asp Phe Ala Cys Lys Thr Ala Gln Gly Thr Ala Ile
```

```
            20                  25                  30

Pro Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val
        35                  40                  45

Val Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe
    50                  55                  60

Cys His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln
65                  70                  75                  80

Arg Gly Ser Ala Tyr Gly Gly Val Leu Ser Gln Phe Ser Gly Thr Val
                85                  90                  95

Lys Tyr Ser Gly Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro
            100                 105                 110

Arg Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu
        115                 120                 125

Tyr Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly
    130                 135                 140

Ser Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser
145                 150                 155                 160

Asp Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val
                165                 170                 175

Val Pro Thr Gly Gly Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr
            180                 185                 190

Leu Pro Asp Tyr Pro Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys
        195                 200                 205

Ala Lys Ser Gln Asn Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp
    210                 215                 220
```

```
Ala Gly Asn Ser Ile Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln
225                 230                 235                 240


Gly Val Gly Val Gln Leu Thr Arg Gln Gly Thr Ile Ile Pro Ala Gln
                245                 250                 255


Asn Thr Val Ser Leu Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly
            260                 265                 270


Leu Thr Ala Asn Tyr Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn
        275                 280                 285


Val Gln Ser Ile Ile Gly Val Thr Phe Val Tyr Gln Pro Gly Asp Gly
    290                 295                 300


Asn Ala Asp Val Thr Ile Thr Val Asn Gly Lys Val Val Ala Lys Ser
305                 310                 315                 320


Gly Ser Gly Ser Gly Gly Gly Gly Gly Gly Ser Asp Ile Ile Lys Leu
                325                 330                 335


Leu Asn Glu Gln Val Asn Lys Glu Met Gln Ser Ser Asn Leu Tyr Met
            340                 345                 350


Ser Met Ser Ser Trp Cys Tyr Thr His Ser Leu Asp Gly Ala Gly Leu
        355                 360                 365


Phe Leu Phe Asp His Ala Ala Glu Glu Tyr Glu His Ala Lys Lys Leu
    370                 375                 380


Ile Ile Phe Leu Asn Glu Asn Asn Val Pro Val Gln Leu Thr Ser Ile
385                 390                 395                 400


Ser Ala Pro Glu His Lys Phe Glu Gly Leu Thr Gln Ile Phe Gln Lys
                405                 410                 415


Ala Tyr Glu His Glu Gln His Ile Ser Glu Ser Ile Asn Asn Ile Val
```

```
            420                 425                 430

Asp His Ala Ile Lys Ser Lys Asp His Ala Thr Phe Asn Phe Leu Gln
        435                 440                 445

Trp Tyr Val Ala Glu Gln His Glu Glu Glu Val Leu Phe Lys Asp Ile
    450                 455                 460

Leu Asp Lys Ile Glu Leu Ile Gly Asn Glu Asn His Gly Leu Tyr Leu
465                 470                 475                 480

Ala Asp Gln Tyr Val Lys Gly Ile Ala Lys Ser Arg Lys
                485                 490
```

<210> 135
<211> 493
<212> PRT
<213> Artificial

<220>
<223> FimH_PGDGN_DG_GGS4-Ferritn j96

<400> 135

```
Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1               5                   10                  15

Gly Ser Thr Gly Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro
            20                  25                  30

Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val Val
        35                  40                  45

Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys
    50                  55                  60

His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg
65                  70                  75                  80

Gly Ser Ala Tyr Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val Lys
```

```
                85                  90                  95

Tyr Ser Gly Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg
            100                 105                 110

Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr
        115                 120                 125

Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser
    130                 135                 140

Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp
145                 150                 155                 160

Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val
                165                 170                 175

Pro Thr Gly Gly Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu
            180                 185                 190

Pro Asp Tyr Pro Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala
        195                 200                 205

Lys Ser Gln Asn Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp Ala
    210                 215                 220

Gly Asn Ser Ile Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly
225                 230                 235                 240

Val Gly Val Gln Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn
                245                 250                 255

Thr Val Ser Leu Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu
            260                 265                 270

Thr Ala Asn Tyr Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val
        275                 280                 285
```

```
Gln Ser Ile Ile Gly Val Thr Phe Val Tyr Gln Pro Gly Asp Gly Asn
    290                 295                 300

Ala Asp Val Thr Ile Thr Val Asn Gly Lys Val Val Ala Lys Gly Gly
305                 310                 315                 320

Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Asp Ile Ile Lys Leu Leu
                325                 330                 335

Asn Glu Gln Val Asn Lys Glu Met Asn Ser Ser Asn Leu Tyr Met Ser
            340                 345                 350

Met Ser Ser Trp Cys Tyr Thr His Ser Leu Asp Gly Ala Gly Leu Phe
        355                 360                 365

Leu Phe Asp His Ala Ala Glu Glu Tyr Glu His Ala Lys Lys Leu Ile
    370                 375                 380

Ile Phe Leu Asn Glu Asn Asn Val Pro Val Gln Leu Thr Ser Ile Ser
385                 390                 395                 400

Ala Pro Glu His Lys Phe Glu Gly Leu Thr Gln Ile Phe Gln Lys Ala
                405                 410                 415

Tyr Glu His Glu Gln His Ile Ser Glu Ser Ile Asn Asn Ile Val Asp
            420                 425                 430

His Ala Ile Lys Ser Lys Asp His Ala Thr Phe Asn Phe Leu Gln Trp
        435                 440                 445

Tyr Val Ala Glu Gln His Glu Glu Glu Val Leu Phe Lys Asp Ile Leu
    450                 455                 460

Asp Lys Ile Glu Leu Ile Gly Asn Glu Asn His Gly Leu Tyr Leu Ala
465                 470                 475                 480

Asp Gln Tyr Val Lys Gly Ile Ala Lys Ser Arg Lys Ser
```

485 490

<210> 136
<211> 318
<212> PRT
<213> Artificial

<220>
<223> FimH_PGDGN_DG j96

<400> 136

Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1 5 10 15

Gly Ser Thr Gly Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro
20 25 30

Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val Val
35 40 45

Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys
50 55 60

His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg
65 70 75 80

Gly Ser Ala Tyr Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val Lys
85 90 95

Tyr Ser Gly Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg
100 105 110

Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr
115 120 125

Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser
130 135 140

Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp

```
145                 150                 155                 160

Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val
                165                 170                 175

Pro Thr Gly Gly Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu
            180                 185                 190

Pro Asp Tyr Pro Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala
        195                 200                 205

Lys Ser Gln Asn Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp Ala
    210                 215                 220

Gly Asn Ser Ile Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly
225                 230                 235                 240

Val Gly Val Gln Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn
                245                 250                 255

Thr Val Ser Leu Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu
            260                 265                 270

Thr Ala Asn Tyr Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val
        275                 280                 285

Gln Ser Ile Ile Gly Val Thr Phe Val Tyr Gln Pro Gly Asp Gly Asn
    290                 295                 300

Ala Asp Val Thr Ile Thr Val Asn Gly Lys Val Val Ala Lys
305                 310                 315


<210>  137
<211>  594
<212>  PRT
<213>  Artificial

<220>
<223>  FIMHDG_PADRE-encapsuline536
```

<400> 137

```
Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1               5                  10                  15

Gly Ser Thr Gly Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro
            20                  25                  30

Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Ala Val
        35                  40                  45

Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys
    50                  55                  60

His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg
65                  70                  75                  80

Gly Ser Ala Tyr Gly Gly Val Leu Ser Ser Phe Ser Gly Thr Val Lys
                85                  90                  95

Tyr Asn Gly Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg
            100                 105                 110

Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr
        115                 120                 125

Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser
    130                 135                 140

Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp
145                 150                 155                 160

Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val
                165                 170                 175

Pro Thr Gly Gly Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu
            180                 185                 190
```

```
Pro Asp Tyr Pro Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala
        195                 200                 205

Lys Ser Gln Asn Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp Ala
    210                 215                 220

Gly Asn Ser Ile Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly
225                 230                 235                 240

Val Gly Val Gln Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn
                245                 250                 255

Thr Val Ser Leu Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu
            260                 265                 270

Thr Ala Asn Tyr Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val
        275                 280                 285

Gln Ser Ile Ile Gly Val Thr Phe Val Tyr Gln Pro Gly Asp Gly Asn
    290                 295                 300

Ala Asp Val Thr Ile Thr Val Asn Gly Lys Val Val Ala Lys Phe Val
305                 310                 315                 320

Ala Ala Trp Thr Leu Lys Ala Ala Ala Met Glu Phe Leu Lys Arg Ser
                325                 330                 335

Phe Ala Pro Leu Thr Glu Lys Gln Trp Gln Glu Ile Asp Asn Arg Ala
            340                 345                 350

Arg Glu Ile Phe Lys Thr Gln Leu Tyr Gly Arg Lys Phe Val Asp Val
        355                 360                 365

Glu Gly Pro Tyr Gly Trp Glu Tyr Ala Ala His Pro Leu Gly Glu Val
    370                 375                 380

Glu Val Leu Ser Asp Glu Asn Glu Val Val Lys Trp Gly Leu Arg Lys
```

```
385                 390                 395                 400

Ser Leu Pro Leu Ile Glu Leu Arg Ala Thr Phe Thr Leu Asp Leu Trp
                405                 410                 415

Glu Leu Asp Asn Leu Glu Arg Gly Lys Pro Asn Val Asp Leu Ser Ser
            420                 425                 430

Leu Glu Glu Thr Val Arg Lys Val Ala Glu Phe Glu Asp Glu Val Ile
        435                 440                 445

Phe Arg Gly Cys Glu Lys Ser Gly Val Lys Gly Leu Leu Ser Phe Glu
    450                 455                 460

Glu Arg Lys Ile Glu Cys Gly Ser Thr Pro Lys Asp Leu Leu Glu Ala
465                 470                 475                 480

Ile Val Arg Ala Leu Ser Ile Phe Ser Lys Asp Gly Ile Glu Gly Pro
                485                 490                 495

Tyr Thr Leu Val Ile Asn Thr Asp Arg Trp Ile Asn Phe Leu Lys Glu
            500                 505                 510

Glu Ala Gly His Tyr Pro Leu Glu Lys Arg Val Glu Glu Cys Leu Arg
        515                 520                 525

Gly Gly Lys Ile Ile Thr Thr Pro Arg Ile Glu Asp Ala Leu Val Val
    530                 535                 540

Ser Glu Arg Gly Gly Asp Phe Lys Leu Ile Leu Gly Gln Asp Leu Ser
545                 550                 555                 560

Ile Gly Tyr Glu Asp Arg Glu Lys Asp Ala Val Arg Leu Phe Ile Thr
                565                 570                 575

Glu Thr Phe Thr Phe Gln Val Val Asn Pro Glu Ala Leu Ile Leu Leu
            580                 585                 590
```

Lys Phe

<210> 138
<211> 494
<212> PRT
<213> Artificial

<220>
<223> FIMH_NOD-FERRITN_J96

<400> 138

```
Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1               5                  10                  15


Gly Ser Thr Gly Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro
            20                  25                  30


Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val Val
        35                  40                  45


Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys
    50                  55                  60


His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg
65                  70                  75                  80


Gly Ser Ala Tyr Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val Lys
                85                  90                  95


Tyr Ser Gly Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg
            100                 105                 110


Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr
        115                 120                 125


Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser
    130                 135                 140
```

```
Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp
145                 150                 155                 160

Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val
                165                 170                 175

Pro Thr Gly Gly Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu
            180                 185                 190

Pro Asp Tyr Pro Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala
        195                 200                 205

Lys Ser Gln Asn Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp Ala
    210                 215                 220

Gly Asn Ser Ile Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly
225                 230                 235                 240

Val Gly Val Gln Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn
                245                 250                 255

Thr Val Ser Leu Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu
            260                 265                 270

Thr Ala Asn Tyr Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val
        275                 280                 285

Gln Ser Ile Ile Gly Val Thr Phe Val Tyr Gln Pro Gly Asp Gly Asn
    290                 295                 300

Ala Asp Val Thr Ile Thr Val Asn Gly Lys Val Val Ala Lys Ser Gly
305                 310                 315                 320

Ser His His His Gly Ser Gly Gly Gly Gly Gly Ser Asp Ile Ile Lys
                325                 330                 335

Leu Leu Asn Glu Gln Val Asn Lys Glu Met Asn Ser Ser Asn Leu Tyr
```

```
            340                 345                 350

Met Ser Met Ser Ser Trp Cys Tyr Thr His Ser Leu Asp Gly Ala Gly
        355                 360                 365

Leu Phe Leu Phe Asp His Ala Ala Glu Glu Tyr Glu His Ala Lys Lys
    370                 375                 380

Leu Ile Ile Phe Leu Asn Glu Asn Asn Val Pro Val Gln Leu Thr Ser
385                 390                 395                 400

Ile Ser Ala Pro Glu His Lys Phe Glu Gly Leu Thr Gln Ile Phe Gln
                405                 410                 415

Lys Ala Tyr Glu His Glu Gln His Ile Ser Glu Ser Ile Asn Asn Ile
            420                 425                 430

Val Asp His Ala Ile Lys Ser Lys Asp His Ala Thr Phe Asn Phe Leu
        435                 440                 445

Gln Trp Tyr Val Ala Glu Gln His Glu Glu Glu Val Leu Phe Lys Asp
    450                 455                 460

Ile Leu Asp Lys Ile Glu Leu Ile Gly Asn Glu Asn His Gly Leu Tyr
465                 470                 475                 480

Leu Ala Asp Gln Tyr Val Lys Gly Ile Ala Lys Ser Arg Lys
                485                 490
```

<210> 139
<211> 497
<212> PRT
<213> Artificial

<220>
<223> FIMH_DG_NOALFA-ferritin_J96

<400> 139

```
Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
```

```
1               5                   10                  15

Gly Ser Thr Gly Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro
            20                  25                  30

Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val Val
        35                  40                  45

Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys
    50                  55                  60

His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg
65                  70                  75                  80

Gly Ser Ala Tyr Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val Lys
                85                  90                  95

Tyr Ser Gly Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg
            100                 105                 110

Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr
        115                 120                 125

Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser
    130                 135                 140

Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp
145                 150                 155                 160

Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val
                165                 170                 175

Pro Thr Gly Gly Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu
            180                 185                 190

Pro Asp Tyr Pro Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala
        195                 200                 205
```

```
Lys Ser Gln Asn Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp Ala
    210                 215                 220

Gly Asn Ser Ile Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly
225                 230                 235                 240

Val Gly Val Gln Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn
                245                 250                 255

Thr Val Ser Leu Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu
            260                 265                 270

Thr Ala Asn Tyr Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val
        275                 280                 285

Gln Ser Ile Ile Gly Val Thr Phe Val Tyr Gln Pro Gly Asp Gly Asn
    290                 295                 300

Ala Asp Val Thr Ile Thr Val Asn Gly Lys Val Val Ala Lys Gly Gly
305                 310                 315                 320

Gly Gly Ser Leu Val Pro Arg Gly Ser Gly Gly Gly Gly Ser Asp Ile
                325                 330                 335

Ile Lys Leu Leu Asn Glu Gln Val Asn Lys Glu Met Asn Ser Ser Asn
            340                 345                 350

Leu Tyr Met Ser Met Ser Ser Trp Cys Tyr Thr His Ser Leu Asp Gly
        355                 360                 365

Ala Gly Leu Phe Leu Phe Asp His Ala Ala Glu Glu Tyr Glu His Ala
    370                 375                 380

Lys Lys Leu Ile Ile Phe Leu Asn Glu Asn Asn Val Pro Val Gln Leu
385                 390                 395                 400

Thr Ser Ile Ser Ala Pro Glu His Lys Phe Glu Gly Leu Thr Gln Ile
```

```
               405                 410                 415

Phe Gln Lys Ala Tyr Glu His Glu Gln His Ile Ser Glu Ser Ile Asn
            420                 425                 430

Asn Ile Val Asp His Ala Ile Lys Ser Lys Asp His Ala Thr Phe Asn
        435                 440                 445

Phe Leu Gln Trp Tyr Val Ala Glu Gln His Glu Glu Glu Val Leu Phe
    450                 455                 460

Lys Asp Ile Leu Asp Lys Ile Glu Leu Ile Gly Asn Glu Asn His Gly
465                 470                 475                 480

Leu Tyr Leu Ala Asp Gln Tyr Val Lys Gly Ile Ala Lys Ser Arg Lys
               485                 490                 495

Ser
```

<210> 140
<211> 495
<212> PRT
<213> Artificial

<220>
<223> FIMH_NOD_S_HIS_FERRITN_J96

<400> 140

```
Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1               5                   10                  15

Gly Ser Thr Gly Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro
            20                  25                  30

Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val Val
        35                  40                  45

Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys
```

```
    50                  55                  60

His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg
65                  70                  75                  80

Gly Ser Ala Tyr Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val Lys
                85                  90                  95

Tyr Ser Gly Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg
            100                 105                 110

Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr
        115                 120                 125

Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser
    130                 135                 140

Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp
145                 150                 155                 160

Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val
                165                 170                 175

Pro Thr Gly Gly Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu
            180                 185                 190

Pro Asp Tyr Pro Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala
        195                 200                 205

Lys Ser Gln Asn Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp Ala
    210                 215                 220

Gly Asn Ser Ile Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly
225                 230                 235                 240

Val Gly Val Gln Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn
                245                 250                 255
```

```
Thr Val Ser Leu Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu
            260                 265                 270

Thr Ala Asn Tyr Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val
        275                 280                 285

Gln Ser Ile Ile Gly Val Thr Phe Val Tyr Gln Pro Gly Asp Gly Asn
    290                 295                 300

Ala Asp Val Thr Ile Thr Val Asn Gly Lys Val Val Ala Lys Ser Gly
305                 310                 315                 320

Ser His His His His His His His His His Gly Gly Ser Asp Ile Ile Lys
                325                 330                 335

Leu Leu Asn Glu Gln Val Asn Lys Glu Met Asn Ser Ser Asn Leu Tyr
            340                 345                 350

Met Ser Met Ser Ser Trp Cys Tyr Thr His Ser Leu Asp Gly Ala Gly
        355                 360                 365

Leu Phe Leu Phe Asp His Ala Ala Glu Glu Tyr Glu His Ala Lys Lys
    370                 375                 380

Leu Ile Ile Phe Leu Asn Glu Asn Asn Val Pro Val Gln Leu Thr Ser
385                 390                 395                 400

Ile Ser Ala Pro Glu His Lys Phe Glu Gly Leu Thr Gln Ile Phe Gln
                405                 410                 415

Lys Ala Tyr Glu His Glu Gln His Ile Ser Glu Ser Ile Asn Asn Ile
            420                 425                 430

Val Asp His Ala Ile Lys Ser Lys Asp His Ala Thr Phe Asn Phe Leu
        435                 440                 445

Gln Trp Tyr Val Ala Glu Gln His Glu Glu Glu Val Leu Phe Lys Asp
```

```
    450                 455                 460

Ile Leu Asp Lys Ile Glu Leu Ile Gly Asn Glu Asn His Gly Leu Tyr
465                 470                 475                 480

Leu Ala Asp Gln Tyr Val Lys Gly Ile Ala Lys Ser Arg Lys Ser
                485                 490                 495
```

<210> 141
<211> 498
<212> PRT
<213> Artificial

<220>
<223> FIMH_DG_PGDGN-PADRE-Ferritin536

<400> 141

```
Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1               5                   10                  15

Gly Ser Thr Gly Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro
            20                  25                  30

Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Ala Val
        35                  40                  45

Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys
    50                  55                  60

His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg
65                  70                  75                  80

Gly Ser Ala Tyr Gly Gly Val Leu Ser Ser Phe Ser Gly Thr Val Lys
                85                  90                  95

Tyr Asn Gly Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg
            100                 105                 110

Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr
```

```
        115                 120                 125

Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser
    130                 135                 140

Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp
145                 150                 155                 160

Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val
                165                 170                 175

Pro Thr Gly Gly Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu
            180                 185                 190

Pro Asp Tyr Pro Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala
        195                 200                 205

Lys Ser Gln Asn Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp Ala
    210                 215                 220

Gly Asn Ser Ile Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly
225                 230                 235                 240

Val Gly Val Gln Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn
                245                 250                 255

Thr Val Ser Leu Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu
            260                 265                 270

Thr Ala Asn Tyr Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val
        275                 280                 285

Gln Ser Ile Ile Gly Val Thr Phe Val Tyr Gln Pro Gly Asp Gly Asn
    290                 295                 300

Ala Asp Val Thr Ile Thr Val Asn Gly Lys Val Val Ala Lys Ser Gly
305                 310                 315                 320
```

```
Ser Phe Val Ala Ala Trp Thr Leu Lys Ala Ala Ala Gly Gly Ser Asp
                325                 330                 335

Ile Ile Lys Leu Leu Asn Glu Gln Val Asn Lys Glu Met Asn Ser Ser
            340                 345                 350

Asn Leu Tyr Met Ser Met Ser Ser Trp Cys Tyr Thr His Ser Leu Asp
        355                 360                 365

Gly Ala Gly Leu Phe Leu Phe Asp His Ala Ala Glu Glu Tyr Glu His
    370                 375                 380

Ala Lys Lys Leu Ile Ile Phe Leu Asn Glu Asn Asn Val Pro Val Gln
385                 390                 395                 400

Leu Thr Ser Ile Ser Ala Pro Glu His Lys Phe Glu Gly Leu Thr Gln
                405                 410                 415

Ile Phe Gln Lys Ala Tyr Glu His Glu Gln His Ile Ser Glu Ser Ile
            420                 425                 430

Asn Asn Ile Val Asp His Ala Ile Lys Ser Lys Asp His Ala Thr Phe
        435                 440                 445

Asn Phe Leu Gln Trp Tyr Val Ala Glu Gln His Glu Glu Glu Val Leu
    450                 455                 460

Phe Lys Asp Ile Leu Asp Lys Ile Glu Leu Ile Gly Asn Glu Asn His
465                 470                 475                 480

Gly Leu Tyr Leu Ala Asp Gln Tyr Val Lys Gly Ile Ala Lys Ser Arg
                485                 490                 495

Lys Ser
```

<210> 142

<211> 498
<212> PRT
<213> Artificial

<220>
<223> fimh-DG_ferritina-linkerAlpha

<400> 142

Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1 5 10 15

Gly Ser Thr Gly Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro
20 25 30

Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val Val
35 40 45

Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys
50 55 60

His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg
65 70 75 80

Gly Ser Ala Tyr Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val Lys
85 90 95

Tyr Ser Gly Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg
100 105 110

Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr
115 120 125

Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser
130 135 140

Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp
145 150 155 160

Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val

```
                165                 170                 175

Pro Thr Gly Gly Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu
            180                 185                 190

Pro Asp Tyr Pro Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala
        195                 200                 205

Lys Ser Gln Asn Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp Ala
    210                 215                 220

Gly Asn Ser Ile Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly
225                 230                 235                 240

Val Gly Val Gln Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn
                245                 250                 255

Thr Val Ser Leu Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu
            260                 265                 270

Thr Ala Asn Tyr Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val
        275                 280                 285

Gln Ser Ile Ile Gly Val Thr Phe Val Tyr Gln Pro Gly Asp Gly Asn
    290                 295                 300

Ala Asp Val Thr Ile Thr Val Asn Gly Lys Val Val Ala Lys Ala Glu
305                 310                 315                 320

Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Ala Asp
                325                 330                 335

Ile Ile Lys Leu Leu Asn Glu Gln Val Asn Lys Glu Met Asn Ser Ser
            340                 345                 350

Asn Leu Tyr Met Ser Met Ser Ser Trp Cys Tyr Thr His Ser Leu Asp
        355                 360                 365
```

```
Gly Ala Gly Leu Phe Leu Phe Asp His Ala Ala Glu Glu Tyr Glu His
    370                 375                 380

Ala Lys Lys Leu Ile Ile Phe Leu Asn Glu Asn Asn Val Pro Val Gln
385                 390                 395                 400

Leu Thr Ser Ile Ser Ala Pro Glu His Lys Phe Glu Gly Leu Thr Gln
                405                 410                 415

Ile Phe Gln Lys Ala Tyr Glu His Glu Gln His Ile Ser Glu Ser Ile
            420                 425                 430

Asn Asn Ile Val Asp His Ala Ile Lys Ser Lys Asp His Ala Thr Phe
        435                 440                 445

Asn Phe Leu Gln Trp Tyr Val Ala Glu Gln His Glu Glu Glu Val Leu
    450                 455                 460

Phe Lys Asp Ile Leu Asp Lys Ile Glu Leu Ile Gly Asn Glu Asn His
465                 470                 475                 480

Gly Leu Tyr Leu Ala Asp Gln Tyr Val Lys Gly Ile Ala Lys Ser Arg
                485                 490                 495

Lys Ser
```

<210> 143
<211> 492
<212> PRT
<213> Artificial

<220>
<223> FIMH-PADRE-Ferritin536_noSpaces

<400> 143

```
Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1               5                   10                  15
```

```
Gly Ser Thr Gly Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro
            20                  25                  30

Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Ala Val
        35                  40                  45

Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys
    50                  55                  60

His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg
65                  70                  75                  80

Gly Ser Ala Tyr Gly Gly Val Leu Ser Ser Phe Ser Gly Thr Val Lys
                85                  90                  95

Tyr Asn Gly Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg
            100                 105                 110

Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr
        115                 120                 125

Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser
    130                 135                 140

Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp
145                 150                 155                 160

Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val
                165                 170                 175

Pro Thr Gly Gly Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu
            180                 185                 190

Pro Asp Tyr Pro Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala
        195                 200                 205

Lys Ser Gln Asn Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp Ala
```

```
    210                 215                 220

Gly Asn Ser Ile Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly
225                 230                 235                 240

Val Gly Val Gln Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn
                245                 250                 255

Thr Val Ser Leu Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu
            260                 265                 270

Thr Ala Asn Tyr Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val
        275                 280                 285

Gln Ser Ile Ile Gly Val Thr Phe Val Tyr Gln Pro Gly Asp Gly Asn
    290                 295                 300

Ala Asp Val Thr Ile Thr Val Asn Gly Lys Val Val Ala Lys Phe Val
305                 310                 315                 320

Ala Ala Trp Thr Leu Lys Ala Ala Ala Asp Ile Ile Lys Leu Leu Asn
                325                 330                 335

Glu Gln Val Asn Lys Glu Met Asn Ser Ser Asn Leu Tyr Met Ser Met
            340                 345                 350

Ser Ser Trp Cys Tyr Thr His Ser Leu Asp Gly Ala Gly Leu Phe Leu
        355                 360                 365

Phe Asp His Ala Ala Glu Glu Tyr Glu His Ala Lys Lys Leu Ile Ile
    370                 375                 380

Phe Leu Asn Glu Asn Asn Val Pro Val Gln Leu Thr Ser Ile Ser Ala
385                 390                 395                 400

Pro Glu His Lys Phe Glu Gly Leu Thr Gln Ile Phe Gln Lys Ala Tyr
                405                 410                 415
```

```
Glu His Glu Gln His Ile Ser Glu Ser Ile Asn Asn Ile Val Asp His
            420                 425                 430

Ala Ile Lys Ser Lys Asp His Ala Thr Phe Asn Phe Leu Gln Trp Tyr
        435                 440                 445

Val Ala Glu Gln His Glu Glu Glu Val Leu Phe Lys Asp Ile Leu Asp
    450                 455                 460

Lys Ile Glu Leu Ile Gly Asn Glu Asn His Gly Leu Tyr Leu Ala Asp
465                 470                 475                 480

Gln Tyr Val Lys Gly Ile Ala Lys Ser Arg Lys Ser
                485                 490
```

<210> 144
<211> 493
<212> PRT
<213> Artificial

<220>
<223> FIMH_DG_GSG4-ferritin_J96

<400> 144

```
Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1               5                   10                  15

Gly Ser Thr Gly Asp Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile
            20                  25                  30

Pro Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val
        35                  40                  45

Val Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe
    50                  55                  60

Cys His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln
65                  70                  75                  80
```

```
Arg Gly Ser Ala Tyr Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val
                85                  90                  95

Lys Tyr Ser Gly Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro
            100                 105                 110

Arg Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu
        115                 120                 125

Tyr Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly
    130                 135                 140

Ser Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser
145                 150                 155                 160

Asp Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val
                165                 170                 175

Val Pro Thr Gly Gly Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr
            180                 185                 190

Leu Pro Asp Tyr Pro Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys
        195                 200                 205

Ala Lys Ser Gln Asn Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp
    210                 215                 220

Ala Gly Asn Ser Ile Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln
225                 230                 235                 240

Gly Val Gly Val Gln Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn
                245                 250                 255

Asn Thr Val Ser Leu Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly
            260                 265                 270

Leu Thr Ala Asn Tyr Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn
```

```
        275                 280                 285

Val Gln Ser Ile Ile Gly Val Thr Phe Val Tyr Gln Pro Gly Asp Gly
    290                 295                 300

Asn Ala Asp Val Thr Ile Thr Val Asn Gly Lys Val Val Ala Lys Ser
305                 310                 315                 320

Gly Ser Gly Ser Gly Gly Gly Gly Gly Gly Ser Asp Ile Ile Lys Leu
                325                 330                 335

Leu Asn Glu Gln Val Asn Lys Glu Met Asn Ser Ser Asn Leu Tyr Met
            340                 345                 350

Ser Met Ser Ser Trp Cys Tyr Thr His Ser Leu Asp Gly Ala Gly Leu
        355                 360                 365

Phe Leu Phe Asp His Ala Ala Glu Glu Tyr Glu His Ala Lys Lys Leu
    370                 375                 380

Ile Ile Phe Leu Asn Glu Asn Asn Val Pro Val Gln Leu Thr Ser Ile
385                 390                 395                 400

Ser Ala Pro Glu His Lys Phe Glu Gly Leu Thr Gln Ile Phe Gln Lys
                405                 410                 415

Ala Tyr Glu His Glu Gln His Ile Ser Glu Ser Ile Asn Asn Ile Val
            420                 425                 430

Asp His Ala Ile Lys Ser Lys Asp His Ala Thr Phe Asn Phe Leu Gln
        435                 440                 445

Trp Tyr Val Ala Glu Gln His Glu Glu Glu Val Leu Phe Lys Asp Ile
    450                 455                 460

Leu Asp Lys Ile Glu Leu Ile Gly Asn Glu Asn His Gly Leu Tyr Leu
465                 470                 475                 480
```

```
Ala Asp Gln Tyr Val Lys Gly Ile Ala Lys Ser Arg Lys
                485                 490


<210>  145
<211>  498
<212>  PRT
<213>  Artificial

<220>
<223>  FIMH_DG_PGDGN-SGS_PADRE-Ferritinj96

<400>  145

Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1               5                   10                  15


Gly Ser Thr Gly Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro
            20                  25                  30


Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val Val
        35                  40                  45


Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys
    50                  55                  60


His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg
65                  70                  75                  80


Gly Ser Ala Tyr Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val Lys
                85                  90                  95


Tyr Ser Gly Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg
            100                 105                 110


Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr
        115                 120                 125


Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser
    130                 135                 140
```

```
Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp
145                 150                 155                 160

Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val
                165                 170                 175

Pro Thr Gly Gly Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu
            180                 185                 190

Pro Asp Tyr Pro Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala
        195                 200                 205

Lys Ser Gln Asn Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp Ala
    210                 215                 220

Gly Asn Ser Ile Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly
225                 230                 235                 240

Val Gly Val Gln Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn
                245                 250                 255

Thr Val Ser Leu Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu
            260                 265                 270

Thr Ala Asn Tyr Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val
        275                 280                 285

Gln Ser Ile Ile Gly Val Thr Phe Val Tyr Gln Pro Gly Asp Gly Asn
    290                 295                 300

Ala Asp Val Thr Ile Thr Val Asn Gly Lys Val Val Ala Lys Ser Gly
305                 310                 315                 320

Ser Phe Val Ala Ala Trp Thr Leu Lys Ala Ala Ala Gly Gly Ser Asp
                325                 330                 335

Ile Ile Lys Leu Leu Asn Glu Gln Val Asn Lys Glu Met Asn Ser Ser
```

```
            340                 345                 350

Asn Leu Tyr Met Ser Met Ser Ser Trp Cys Tyr Thr His Ser Leu Asp
        355                 360                 365

Gly Ala Gly Leu Phe Leu Phe Asp His Ala Ala Glu Glu Tyr Glu His
    370                 375                 380

Ala Lys Lys Leu Ile Ile Phe Leu Asn Glu Asn Asn Val Pro Val Gln
385                 390                 395                 400

Leu Thr Ser Ile Ser Ala Pro Glu His Lys Phe Glu Gly Leu Thr Gln
                405                 410                 415

Ile Phe Gln Lys Ala Tyr Glu His Glu Gln His Ile Ser Glu Ser Ile
            420                 425                 430

Asn Asn Ile Val Asp His Ala Ile Lys Ser Lys Asp His Ala Thr Phe
        435                 440                 445

Asn Phe Leu Gln Trp Tyr Val Ala Glu Gln His Glu Glu Glu Val Leu
    450                 455                 460

Phe Lys Asp Ile Leu Asp Lys Ile Glu Leu Ile Gly Asn Glu Asn His
465                 470                 475                 480

Gly Leu Tyr Leu Ala Asp Gln Tyr Val Lys Gly Ile Ala Lys Ser Arg
                485                 490                 495

Lys Ser
```

<210> 146
<211> 497
<212> PRT
<213> Artificial

<220>
<223> fimh-ferritina-linkerNONa

<400> 146

```
Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1               5                  10                  15


Gly Ser Thr Gly Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro
            20                  25                  30


Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val Val
        35                  40                  45


Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys
    50                  55                  60


His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg
65                  70                  75                  80


Gly Ser Ala Tyr Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val Lys
                85                  90                  95


Tyr Ser Gly Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg
            100                 105                 110


Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr
        115                 120                 125


Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser
    130                 135                 140


Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp
145                 150                 155                 160


Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val
                165                 170                 175


Pro Thr Gly Gly Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu
            180                 185                 190
```

```
Pro Asp Tyr Pro Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala
        195                 200                 205

Lys Ser Gln Asn Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp Ala
    210                 215                 220

Gly Asn Ser Ile Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly
225                 230                 235                 240

Val Gly Val Gln Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn
                245                 250                 255

Thr Val Ser Leu Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu
            260                 265                 270

Thr Ala Asn Tyr Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val
        275                 280                 285

Gln Ser Ile Ile Gly Val Thr Phe Val Tyr Gln Pro Gly Asp Gly Asn
    290                 295                 300

Ala Asp Val Thr Ile Thr Val Asn Gly Lys Val Val Ala Lys Gly Gly
305                 310                 315                 320

Gly Gly Ser Leu Val Pro Arg Gly Ser Gly Gly Gly Gly Ser Asp Ile
                325                 330                 335

Ile Lys Leu Leu Asn Glu Gln Val Asn Lys Glu Met Asn Ser Ser Asn
            340                 345                 350

Leu Tyr Met Ser Met Ser Ser Trp Cys Tyr Thr His Ser Leu Asp Gly
        355                 360                 365

Ala Gly Leu Phe Leu Phe Asp His Ala Ala Glu Glu Tyr Glu His Ala
    370                 375                 380

Lys Lys Leu Ile Ile Phe Leu Asn Glu Asn Asn Val Pro Val Gln Leu
```

385 390 395 400

Thr Ser Ile Ser Ala Pro Glu His Lys Phe Glu Gly Leu Thr Gln Ile
405 410 415

Phe Gln Lys Ala Tyr Glu His Glu Gln His Ile Ser Glu Ser Ile Asn
420 425 430

Asn Ile Val Asp His Ala Ile Lys Ser Lys Asp His Ala Thr Phe Asn
435 440 445

Phe Leu Gln Trp Tyr Val Ala Glu Gln His Glu Glu Glu Val Leu Phe
450 455 460

Lys Asp Ile Leu Asp Lys Ile Glu Leu Ile Gly Asn Glu Asn His Gly
465 470 475 480

Leu Tyr Leu Ala Asp Gln Tyr Val Lys Gly Ile Ala Lys Ser Arg Lys
485 490 495

Ser

<210> 147
<211> 492
<212> PRT
<213> Artificial

<220>
<223> FIMH-Ferritinj96_noSpaces

<400> 147

Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1 5 10 15

Gly Ser Thr Gly Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile Pro
20 25 30

Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Val Val

```
        35                  40                  45

Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe Cys
    50                  55                  60

His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln Arg
65                  70                  75                  80

Gly Ser Ala Tyr Gly Gly Val Leu Ser Asn Phe Ser Gly Thr Val Lys
                85                  90                  95

Tyr Ser Gly Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro Arg
            100                 105                 110

Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu Tyr
        115                 120                 125

Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly Ser
    130                 135                 140

Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser Asp
145                 150                 155                 160

Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val Val
                165                 170                 175

Pro Thr Gly Gly Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr Leu
            180                 185                 190

Pro Asp Tyr Pro Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys Ala
        195                 200                 205

Lys Ser Gln Asn Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp Ala
    210                 215                 220

Gly Asn Ser Ile Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln Gly
225                 230                 235                 240
```

```
Val Gly Val Gln Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn Asn
                245                 250                 255

Thr Val Ser Leu Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly Leu
            260                 265                 270

Thr Ala Asn Tyr Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn Val
        275                 280                 285

Gln Ser Ile Ile Gly Val Thr Phe Val Tyr Gln Pro Gly Asp Gly Asn
    290                 295                 300

Ala Asp Val Thr Ile Thr Val Asn Gly Lys Val Val Ala Lys Phe Val
305                 310                 315                 320

Ala Ala Trp Thr Leu Lys Ala Ala Ala Asp Ile Ile Lys Leu Leu Asn
                325                 330                 335

Glu Gln Val Asn Lys Glu Met Asn Ser Ser Asn Leu Tyr Met Ser Met
            340                 345                 350

Ser Ser Trp Cys Tyr Thr His Ser Leu Asp Gly Ala Gly Leu Phe Leu
        355                 360                 365

Phe Asp His Ala Ala Glu Glu Tyr Glu His Ala Lys Lys Leu Ile Ile
    370                 375                 380

Phe Leu Asn Glu Asn Asn Val Pro Val Gln Leu Thr Ser Ile Ser Ala
385                 390                 395                 400

Pro Glu His Lys Phe Glu Gly Leu Thr Gln Ile Phe Gln Lys Ala Tyr
                405                 410                 415

Glu His Glu Gln His Ile Ser Glu Ser Ile Asn Asn Ile Val Asp His
            420                 425                 430

Ala Ile Lys Ser Lys Asp His Ala Thr Phe Asn Phe Leu Gln Trp Tyr
```

```
        435                 440                 445

Val Ala Glu Gln His Glu Glu Glu Val Leu Phe Lys Asp Ile Leu Asp
    450                 455                 460

Lys Ile Glu Leu Ile Gly Asn Glu Asn His Gly Leu Tyr Leu Ala Asp
465                 470                 475                 480

Gln Tyr Val Lys Gly Ile Ala Lys Ser Arg Lys Ser
                485                 490
```

<210> 148
<211> 493
<212> PRT
<213> Artificial

<220>
<223> FIMH_DG_PGDGN-GGS4-Ferritin536

<400> 148

```
Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1               5                   10                  15

Gly Ser Thr Gly Asp Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile
            20                  25                  30

Pro Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Ala
        35                  40                  45

Val Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe
    50                  55                  60

Cys His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln
65                  70                  75                  80

Arg Gly Ser Ala Tyr Gly Gly Val Leu Ser Ser Phe Ser Gly Thr Val
                85                  90                  95

Lys Tyr Asn Gly Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro
```

```
            100                 105                 110

Arg Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu
        115                 120                 125

Tyr Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly
    130                 135                 140

Ser Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser
145                 150                 155                 160

Asp Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val
                165                 170                 175

Val Pro Thr Gly Gly Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr
            180                 185                 190

Leu Pro Asp Tyr Pro Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys
        195                 200                 205

Ala Lys Ser Gln Asn Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp
    210                 215                 220

Ala Gly Asn Ser Ile Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln
225                 230                 235                 240

Gly Val Gly Val Gln Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn
                245                 250                 255

Asn Thr Val Ser Leu Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly
            260                 265                 270

Leu Thr Ala Asn Tyr Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn
        275                 280                 285

Val Gln Ser Ile Ile Gly Val Thr Phe Val Tyr Gln Pro Gly Asp Gly
    290                 295                 300
```

```
Asn Ala Asp Val Thr Ile Thr Val Asn Gly Lys Val Val Ala Lys Gly
305                 310                 315                 320

Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Asp Ile Ile Lys Leu
                325                 330                 335

Leu Asn Glu Gln Val Asn Lys Glu Met Asn Ser Ser Asn Leu Tyr Met
            340                 345                 350

Ser Met Ser Ser Trp Cys Tyr Thr His Ser Leu Asp Gly Ala Gly Leu
        355                 360                 365

Phe Leu Phe Asp His Ala Ala Glu Glu Tyr Glu His Ala Lys Lys Leu
    370                 375                 380

Ile Ile Phe Leu Asn Glu Asn Asn Val Pro Val Gln Leu Thr Ser Ile
385                 390                 395                 400

Ser Ala Pro Glu His Lys Phe Glu Gly Leu Thr Gln Ile Phe Gln Lys
                405                 410                 415

Ala Tyr Glu His Glu Gln His Ile Ser Glu Ser Ile Asn Asn Ile Val
            420                 425                 430

Asp His Ala Ile Lys Ser Lys Asp His Ala Thr Phe Asn Phe Leu Gln
        435                 440                 445

Trp Tyr Val Ala Glu Gln His Glu Glu Glu Val Leu Phe Lys Asp Ile
    450                 455                 460

Leu Asp Lys Ile Glu Leu Ile Gly Asn Glu Asn His Gly Leu Tyr Leu
465                 470                 475                 480

Ala Asp Gln Tyr Val Lys Gly Ile Ala Lys Ser Arg Lys
                485                 490
```

<210> 149

<211> 164
<212> PRT
<213> Artificial

<220>
<223> 1EUM_0_5 - stabilized E. coli ferritin

<400> 149

```
Leu Lys Pro Glu Met Ile Glu Lys Leu Asn Glu Gln Met Asn Leu Glu
1               5                  10                  15


Leu Tyr Ser Ser Leu Leu Tyr Gln Gln Met Ser Ala Trp Cys Ser Tyr
            20                  25                  30


His Gly Phe Glu Gly Ala Ala Ala Phe Leu Arg Arg His Ala Gln Glu
        35                  40                  45


Glu Met Thr His Met Gln Arg Leu Phe Asp Tyr Leu Thr Asp Thr Gly
    50                  55                  60


Asn Leu Pro Arg Ile Asp Thr Ile Pro Ser Pro Phe Ala Glu Tyr Ser
65                  70                  75                  80


Ser Leu Asp Glu Leu Phe Gln Glu Thr Tyr Lys His Glu Gln Leu Ile
                85                  90                  95


Thr Gln Lys Ile Asn Glu Leu Ala His Ala Ala Met Thr Asn Gln Asp
            100                 105                 110


Tyr Pro Thr Phe Asn Phe Leu Gln Trp Tyr Val Ala Glu Gln His Glu
        115                 120                 125


Glu Glu Lys Leu Phe Lys Ser Ile Ile Asp Lys Leu Ser Leu Ala Gly
    130                 135                 140


Lys Ser Gly Glu Gly Leu Tyr Phe Ile Asp Lys Glu Leu Ser Thr Leu
145                 150                 155                 160


Asp Thr Gln Asn
```

<210> 150
<211> 164
<212> PRT
<213> Artificial

<220>
<223> 1EUM_2 - stabilized E. coli ferritin

<400> 150

```
Leu Lys Pro Glu Met Ile Glu Lys Leu Asn Glu Gln Met Asn Leu Glu
1               5                   10                  15


Leu Tyr Ser Ser Leu Leu Tyr Gln Gln Met Ser Ala Trp Cys Ser Tyr
            20                  25                  30


His Gly Phe Glu Gly Ala Ala Ala Phe Leu Arg Arg His Ala Gln Glu
        35                  40                  45


Glu Met Thr His Met Gln Arg Leu Phe Asp Tyr Leu Thr Asp Thr Gly
    50                  55                  60


Asn Leu Pro Arg Ile Asn Thr Ile Pro Ser Pro Phe Ala Glu Tyr Ser
65                  70                  75                  80


Ser Leu Asp Glu Leu Phe Gln Glu Thr Tyr Lys His Glu Gln Leu Ile
                85                  90                  95


Thr Gln Lys Ile Asn Glu Leu Ala His Ala Ala Met Thr Asn Gln Asp
            100                 105                 110


Tyr Pro Thr Phe Asn Phe Leu Gln Trp Tyr Val Ala Glu Gln His Glu
        115                 120                 125


Glu Glu Lys Leu Phe Lys Ser Ile Ile Asp Lys Leu Ser Leu Ala Gly
    130                 135                 140


Lys Ser Gly Glu Gly Leu Tyr Phe Ile Asp Lys Glu Leu Ser Thr Leu
```

145 150 155 160

Asp Thr Gln Asn

<210> 151
<211> 164
<212> PRT
<213> Artificial

<220>
<223> 1EUM_2_5 - stabilized E. coli ferritin

<400> 151

Leu Lys Pro Glu Met Ile Glu Lys Leu Asn Glu Gln Met Asn Leu Glu
1 5 10 15

Leu Tyr Ser Ser Leu Leu Tyr Gln Gln Met Ser Ala Trp Cys Ser Tyr
20 25 30

His Gly Phe Glu Gly Ala Ala Ala Phe Leu Arg Arg His Ala Gln Glu
35 40 45

Glu Met Thr His Met Gln Arg Leu Phe Asp Tyr Leu Thr Asp Thr Gly
50 55 60

Asn Leu Pro Arg Ile Asn Thr Val Pro Ser Pro Phe Ala Glu Tyr Ser
65 70 75 80

Ser Leu Asp Glu Leu Phe Gln Glu Thr Tyr Lys His Glu Gln Leu Ile
85 90 95

Thr Gln Lys Ile Asn Glu Leu Ala His Ala Ala Met Thr Asn Gln Asp
100 105 110

Tyr Pro Thr Phe Asn Phe Leu Gln Trp Tyr Val Ala Glu Gln His Glu
115 120 125

Glu Glu Lys Leu Phe Lys Ser Ile Ile Asp Lys Leu Ser Leu Ala Gly

130 135 140

Lys Ser Gly Glu Gly Leu Tyr Phe Ile Asp Lys Glu Leu Ser Thr Leu
145 150 155 160

Asp Thr Gln Asn

<210> 152
<211> 164
<212> PRT
<213> Artificial

<220>
<223> 1EUM_6- stabilized E. coli ferritin

<400> 152

Leu Lys Pro Glu Met Ile Glu Lys Leu Asn Glu Gln Met Asn Leu Glu
1 5 10 15

Leu Tyr Ser Ser Leu Leu Tyr Gln Gln Met Ser Ala Trp Cys Ser Tyr
20 25 30

His Gly Phe Glu Gly Ala Ala Ala Phe Leu Arg Arg His Ala Gln Glu
35 40 45

Glu Met Thr His Met Gln Arg Leu Phe Asp Tyr Leu Thr Asp Thr Gly
50 55 60

Asn Leu Pro Arg Ile Asn Thr Val Glu Ser Pro Phe Ala Glu Tyr Ser
65 70 75 80

Ser Leu Asp Glu Leu Phe Gln Glu Thr Tyr Lys His Glu Gln Leu Ile
85 90 95

Thr Gln Lys Ile Asn Glu Leu Ala His Ala Ala Met Thr Asn Gln Asp
100 105 110

Tyr Pro Thr Phe Asn Phe Leu Gln Trp Tyr Val Ser Glu Gln His Glu

115 120 125

Glu Glu Lys Leu Phe Lys Ser Ile Ile Asp Lys Leu Ser Leu Ala Gly
130 135 140

Lys Ser Gly Glu Gly Leu Tyr Phe Ile Asp Lys Glu Leu Ser Thr Leu
145 150 155 160

Asp Thr Gln Asn

<210> 153
<211> 498
<212> PRT
<213> Artificial

<220>
<223> Fusion of FimH_DG to E.Coli Ferritin stabilized 0.5

<400> 153

Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1 5 10 15

Gly Ser Thr Gly Asp Phe Ala Cys Lys Thr Ala Asn Gly Thr Ala Ile
20 25 30

Pro Ile Gly Gly Gly Ser Ala Asn Val Tyr Val Asn Leu Ala Pro Ala
35 40 45

Val Asn Val Gly Gln Asn Leu Val Val Asp Leu Ser Thr Gln Ile Phe
50 55 60

Cys His Asn Asp Tyr Pro Glu Thr Ile Thr Asp Tyr Val Thr Leu Gln
65 70 75 80

Arg Gly Ser Ala Tyr Gly Gly Val Leu Ser Ser Phe Ser Gly Thr Val
85 90 95

Lys Tyr Asn Gly Ser Ser Tyr Pro Phe Pro Thr Thr Ser Glu Thr Pro

```
            100                 105                 110

Arg Val Val Tyr Asn Ser Arg Thr Asp Lys Pro Trp Pro Val Ala Leu
        115                 120                 125

Tyr Leu Thr Pro Val Ser Ser Ala Gly Gly Val Ala Ile Lys Ala Gly
    130                 135                 140

Ser Leu Ile Ala Val Leu Ile Leu Arg Gln Thr Asn Asn Tyr Asn Ser
145                 150                 155                 160

Asp Asp Phe Gln Phe Val Trp Asn Ile Tyr Ala Asn Asn Asp Val Val
                165                 170                 175

Val Pro Thr Gly Gly Cys Asp Val Ser Ala Arg Asp Val Thr Val Thr
            180                 185                 190

Leu Pro Asp Tyr Pro Gly Ser Val Pro Ile Pro Leu Thr Val Tyr Cys
        195                 200                 205

Ala Lys Ser Gln Asn Leu Gly Tyr Tyr Leu Ser Gly Thr Thr Ala Asp
    210                 215                 220

Ala Gly Asn Ser Ile Phe Thr Asn Thr Ala Ser Phe Ser Pro Ala Gln
225                 230                 235                 240

Gly Val Gly Val Gln Leu Thr Arg Asn Gly Thr Ile Ile Pro Ala Asn
                245                 250                 255

Asn Thr Val Ser Leu Gly Ala Val Gly Thr Ser Ala Val Ser Leu Gly
            260                 265                 270

Leu Thr Ala Asn Tyr Ala Arg Thr Gly Gly Gln Val Thr Ala Gly Asn
        275                 280                 285

Val Gln Ser Ile Ile Gly Val Thr Phe Val Tyr Gln Pro Gly Asp Gly
    290                 295                 300
```

```
Asn Ala Asp Val Thr Ile Thr Val Asn Gly Lys Val Val Ala Lys Ser
305                 310                 315                 320

Gly Ser His His His His His His His His Gly Gly Ser Met Leu Lys
                325                 330                 335

Pro Glu Met Ile Glu Lys Leu Asn Glu Gln Met Asn Leu Glu Leu Tyr
            340                 345                 350

Ser Ser Leu Leu Tyr Gln Gln Met Ser Ala Trp Cys Ser Tyr His Gly
        355                 360                 365

Phe Glu Gly Ala Ala Ala Phe Leu Arg Arg His Ala Gln Glu Glu Met
    370                 375                 380

Thr His Met Gln Arg Leu Phe Asp Tyr Leu Thr Asp Thr Gly Asn Leu
385                 390                 395                 400

Pro Arg Ile Asp Thr Ile Pro Ser Pro Phe Ala Glu Tyr Ser Ser Leu
                405                 410                 415

Asp Glu Leu Phe Gln Glu Thr Tyr Lys His Glu Gln Leu Ile Thr Gln
            420                 425                 430

Lys Ile Asn Glu Leu Ala His Ala Ala Met Thr Asn Gln Asp Tyr Pro
        435                 440                 445

Thr Phe Asn Phe Leu Gln Trp Tyr Val Ala Glu Gln His Glu Glu Glu
    450                 455                 460

Lys Leu Phe Lys Ser Ile Ile Asp Lys Leu Ser Leu Ala Gly Lys Ser
465                 470                 475                 480

Gly Glu Gly Leu Tyr Phe Ile Asp Lys Glu Leu Ser Thr Leu Asp Thr
                485                 490                 495

Gln Asn
```