



US012384833B2

(12) **United States Patent**  
Feng et al.

(10) **Patent No.:** US 12,384,833 B2  
(45) **Date of Patent:** Aug. 12, 2025

(54) **YEAST-BASED IMMUNOTHERAPY  
AGAINST CLOSTRIDIUM DIFFICILE  
INFECTION**

(71) Applicants: **Hanping Feng**, Ellicott City, MD (US); **James Eugene Galen**, Eldersburg, MD (US); **Kevin Chen**, Baltimore, MD (US); **Yixuan Zhu**, Ellicott City, MD (US)

(72) Inventors: **Hanping Feng**, Ellicott City, MD (US); **James Eugene Galen**, Eldersburg, MD (US); **Kevin Chen**, Baltimore, MD (US); **Yixuan Zhu**, Ellicott City, MD (US)

(73) Assignee: **UNIVERSITY OF MARYLAND,  
BALTIMORE**, Baltimore, MD (US)

(\*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 132 days.

(21) Appl. No.: **17/955,537**

(22) Filed: **Sep. 29, 2022**

(65) **Prior Publication Data**

US 2023/0340087 A1 Oct. 26, 2023

**Related U.S. Application Data**

(62) Division of application No. 15/768,331, filed as application No. PCT/US2016/056875 on Oct. 13, 2016, now Pat. No. 11,466,075.

(60) Provisional application No. 62/240,810, filed on Oct. 13, 2015.

(51) **Int. Cl.**

**C07K 16/12** (2006.01)  
**A61K 9/00** (2006.01)  
**A61K 36/064** (2006.01)  
**A61K 39/00** (2006.01)  
**A61K 45/06** (2006.01)  
**A61P 31/04** (2006.01)  
**CI2N 1/18** (2006.01)  
**G01N 33/569** (2006.01)

(52) **U.S. Cl.**

CPC ..... **C07K 16/1282** (2013.01); **A61K 9/0031** (2013.01); **A61K 9/0043** (2013.01); **A61K 9/0053** (2013.01); **A61K 36/064** (2013.01); **A61K 45/06** (2013.01); **A61P 31/04** (2018.01); **C12N 1/18** (2013.01); **G01N 33/56961** (2013.01); **A61K 2039/505** (2013.01); **C07K 2317/14** (2013.01); **C07K 2317/21** (2013.01); **C07K 2317/22** (2013.01); **C07K 2317/31** (2013.01); **C07K 2317/35** (2013.01); **C07K 2317/52** (2013.01); **C07K 2317/569** (2013.01); **C07K 2317/64** (2013.01); **C07K 2317/76** (2013.01); **C07K 2317/94** (2013.01)

(58) **Field of Classification Search**

None

See application file for complete search history.

(56) **References Cited**

**U.S. PATENT DOCUMENTS**

2002/0018749 A1	2/2002	Hudson et al.	
2003/0228665 A1*	12/2003	Gautvik .....	C12N 15/70
			435/254.2
2010/0272698 A1	10/2010	Stateva et al.	
2012/0237496 A1	9/2012	Birkenfeld et al.	
2013/0058962 A1	3/2013	Shoemaker et al.	
2014/0294826 A1*	10/2014	Shoemaker .....	C07K 16/1278
			424/134.1

**FOREIGN PATENT DOCUMENTS**

WO	WO-2015/100409 A2	7/2015
WO	WO-2016/127104 A2	8/2016

**OTHER PUBLICATIONS**

Yang et al. (JID vol. 210, pp. 964-972, Sep. 2014) (Year: 2014).<sup>\*</sup>  
Qamar et al., Infection and Immunity vol. 69, n.4, pp. 2762-2765, 2001. (Year: 2001).<sup>\*</sup>

Feng, Hanping, "Multispecific single domain antibody for treatment of Clostridium difficile infection," BIT's 5th international Congress of Antibodies, 2013, Hangzhou, China, 186.

Hamed et al., "Generation of a Uracil Auxotroph Strain of the Probiotic Yeast *Saccharomyces boulardii* as a Host for the Recombinant Protein Production," Avicenna J. Med. Biotech., 2013. 5(1):29-34.

Hudson et al., "Functional Heterologous Protein Expression by Genetically Engineered probiotic Yeast *Saccharomyces boulardii*," PLoS ONE, Nov. 12, 2014, 9(11):e112660, 1-12.

Kagaku to Seibutsu (Chemistry and Biology), 2012, vol. 50, No. 3, pp. 163-174.

Liu et al., "Expression of cellulase genes in *Saccharomyces cerevisiae* via d-integration subject to auxotrophic markers," Biotechnol Lett, 35:1303-1307 (2013).

Liu et al., "Single domain antibody-alkaline phosphatase fusion proteins for antigen detection—Analysis of affinity and thermal stability of single domain antibody," Journal of Immunological Methods, 2013, 393:1-7.

Palma et al., "Probiotic *Saccharomyces cerevisiae* strains as biotherapeutic tools: is there room for improvement?", Appl. Microbiol. Biotechnol., 2015, 99:6563-6570.

(Continued)

*Primary Examiner* — Jana A Hines

*Assistant Examiner* — Khatol S Shahnan Shah

(74) *Attorney, Agent, or Firm* — Foley & Lardner LLP

(57) **ABSTRACT**

Antibody-based binding agents derived from human and camelid immunoglobulins are described, as well as strains of yeast engineered to secrete the binding agents, and methods of treating and preventing *Clostridium difficile* infections using the engineered strains of yeast. These binding agents recognize and bind with specificity to *Clostridium difficile* toxin A and/or toxin B and in some cases exhibit toxin neutralizing activity. The binding agents include camelid V<sub>H</sub>H peptide monomers, linked groups of V<sub>H</sub>H peptide monomers, V<sub>H</sub>H peptide monomers joined to antibody Fc domains, and V<sub>H</sub>H peptide monomers joined to IgG antibodies.

**12 Claims, 21 Drawing Sheets**

**Specification includes a Sequence Listing.**

(56)

**References Cited**

**OTHER PUBLICATIONS**

Schmidt et al., "A Tetraspecific VHH-Based Neutralizing Antibody Modifies Disease Outcome in Three Animal Models of Clostridium difficile Infection," *Clin. Vaccine Immunol.*, Sep. 2016, 23(9):774-784.

Shi et al., "Engineering of Chromosomal Wax Ester Synthase Integrated *Saccharomyces cerevisiae* Mutants for Improved Biosynthesis of Fatty Acid Ethyl Esters," *Biotechnol. Bioeng.*, 2014, 111(9):1740-1747.

Spiess et al., "Alternative molecular formats and therapeutic applications for bispecific antibodies," *Molecular Immunology*, 2015, 67:95-106.

Vickers et al., "Dual gene expression cassette vectors with antibiotic selection markers for engineering in *Saccharomyces cerevisiae*," *Microbial Cell Factories*, 2013, 12(96):1-10.

Von Heijne et al., "Species-specific variation in signal peptide design, Implications for protein secretion in foreign hosts," *FEBS Letters*, Feb. 1989, 244(2):439-446.

Yang et al., "The protective effect of recombinant *Lactococcus lactis* oral vaccine on a *Clostridium difficile*-infected animal model," *BMC Gastroenterology*, 2013, 13:117, 13 pages.

\* cited by examiner

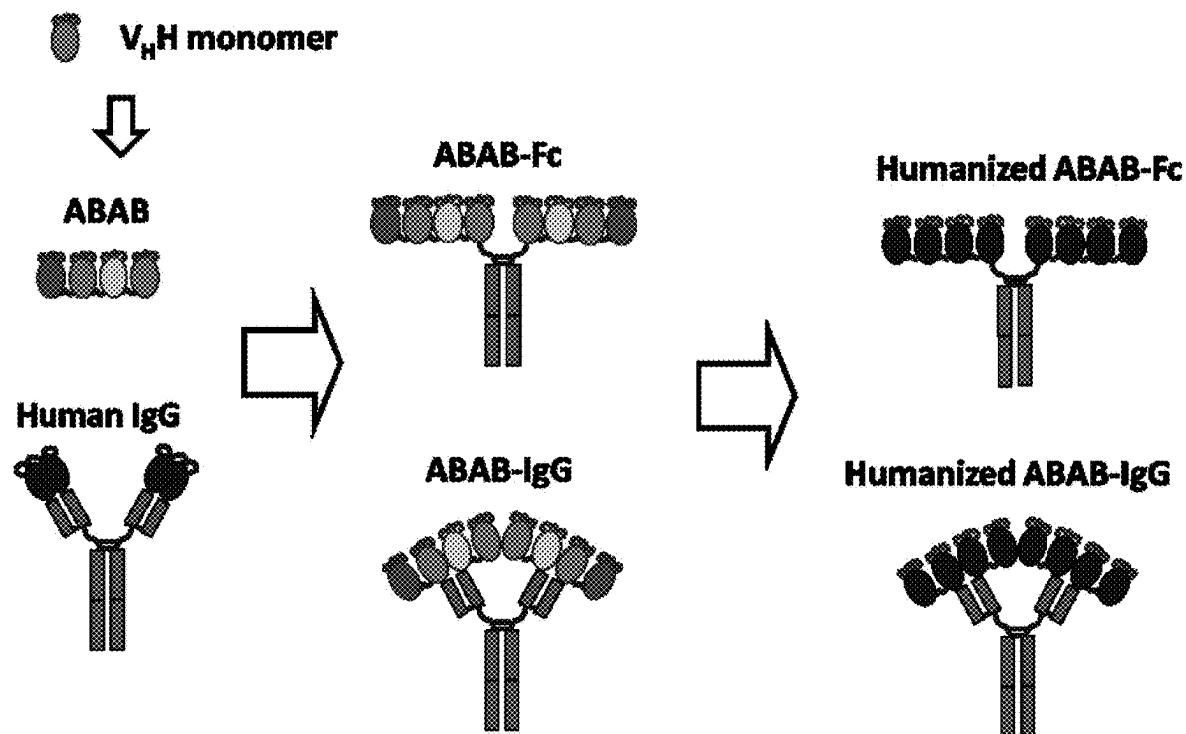


FIG. 1

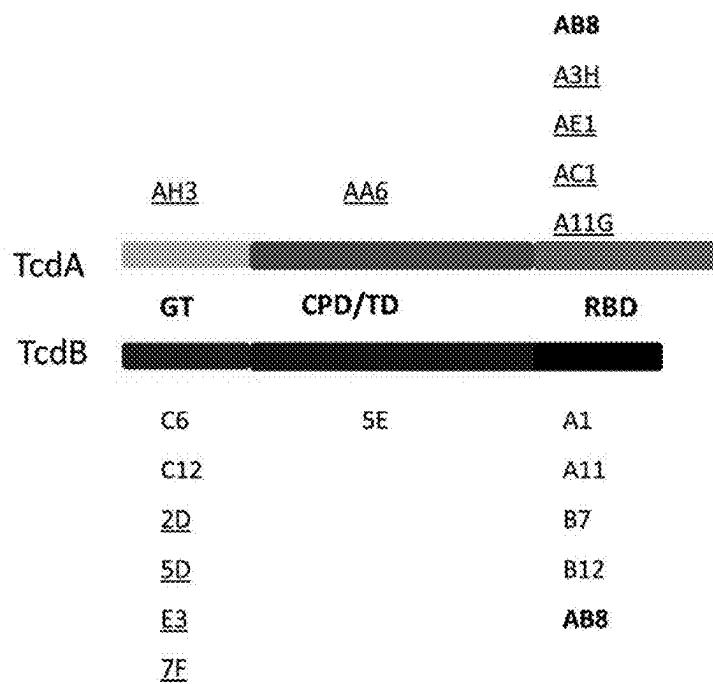


FIG. 2

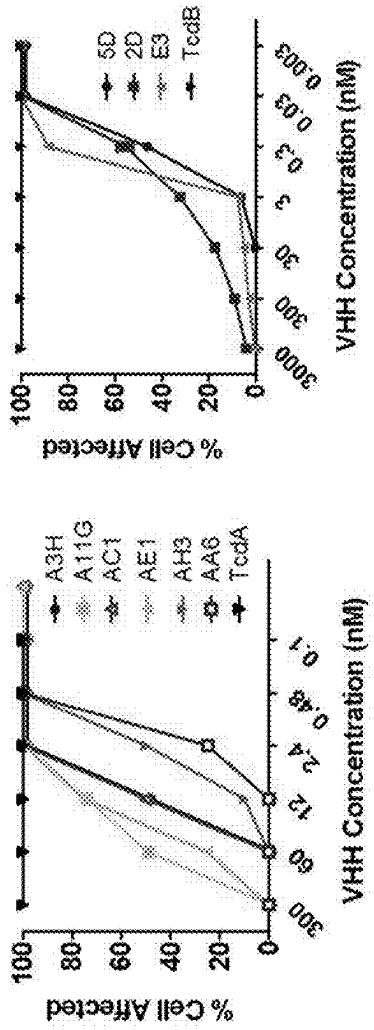


FIG. 3A

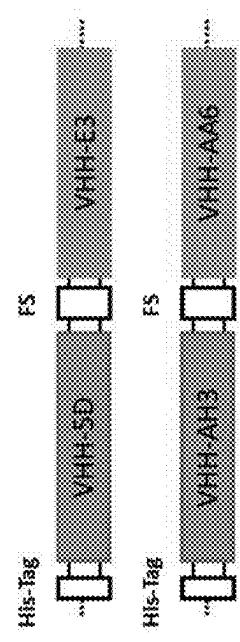


FIG. 3B

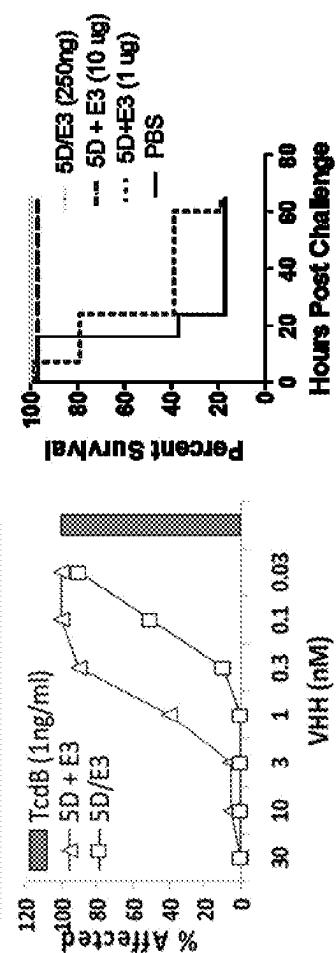


FIG. 3C

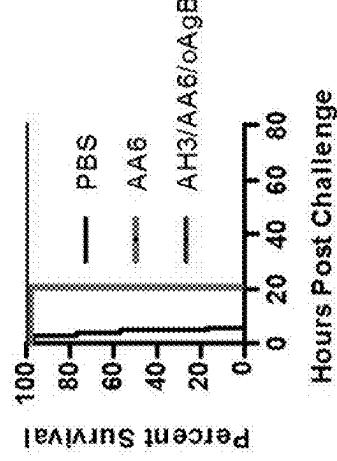


FIG. 3D

FIG. 3E

FIG. 3F

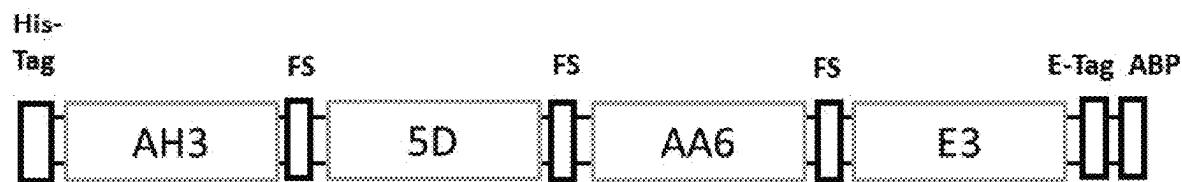


FIG. 4

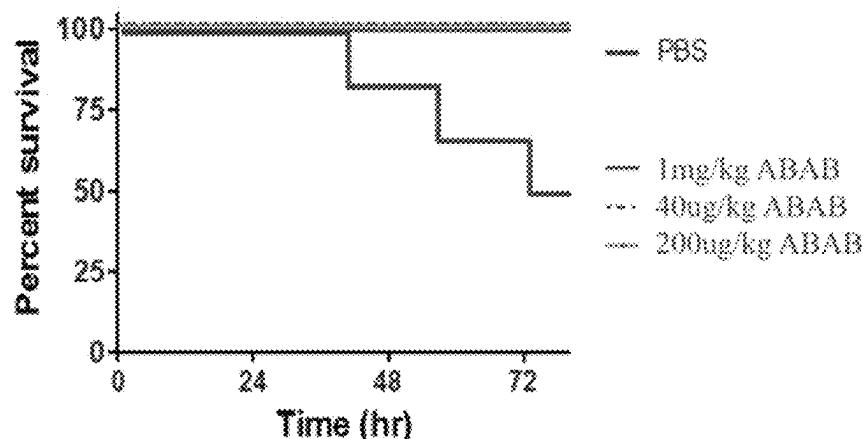


FIG. 5A

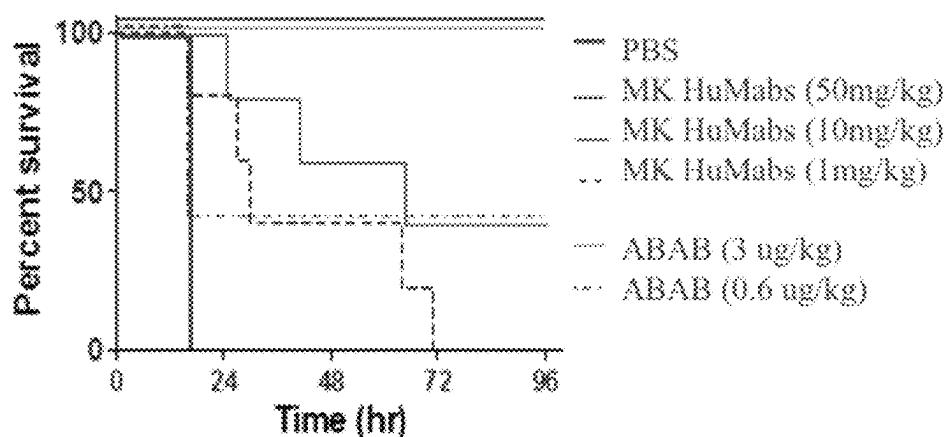


FIG. 5B

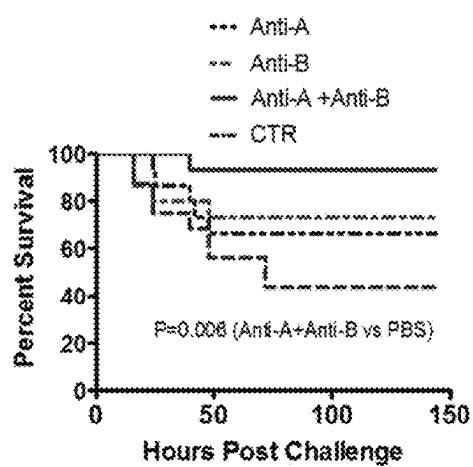


FIG. 6A

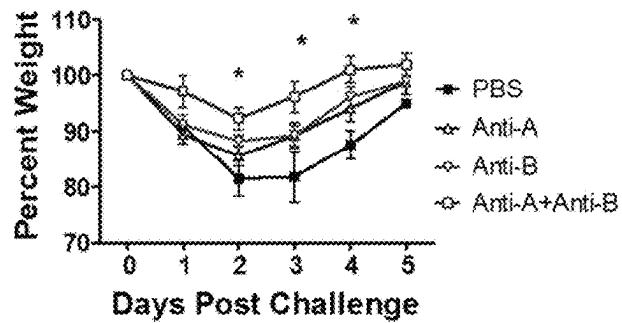


FIG. 6B

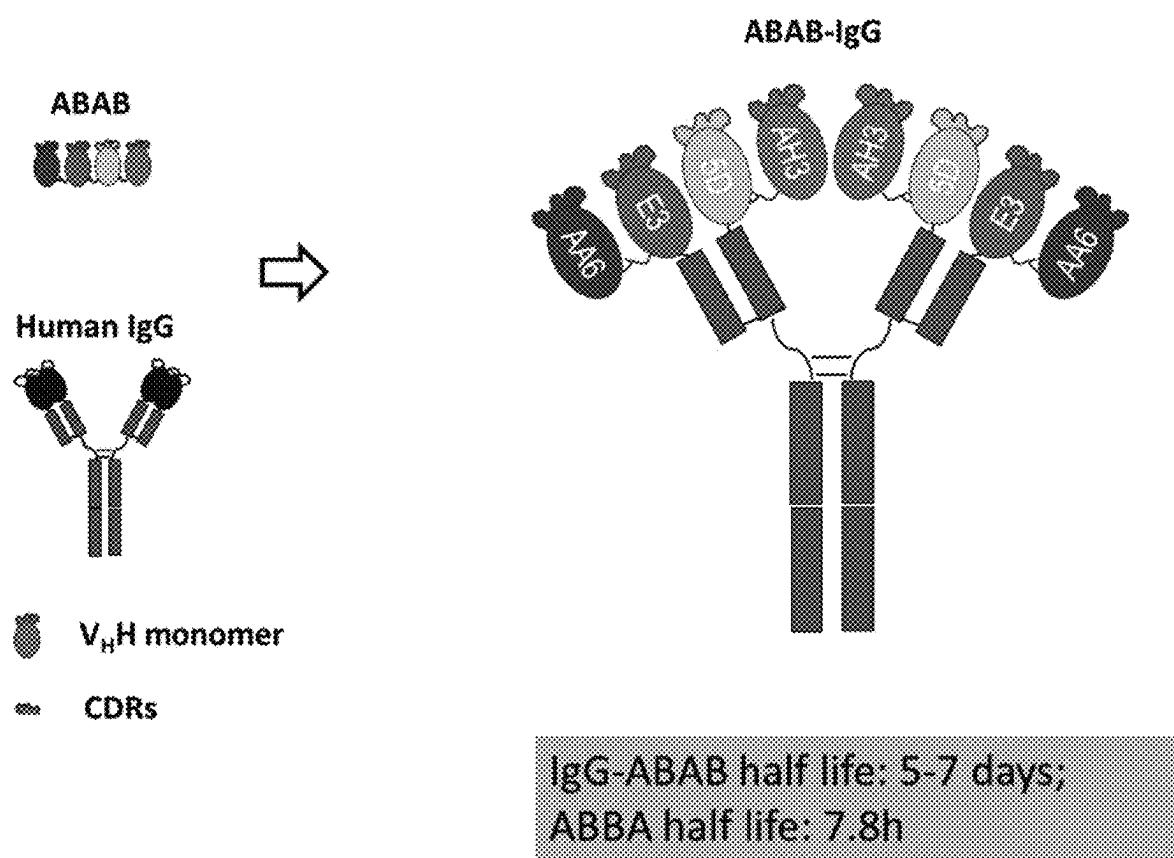
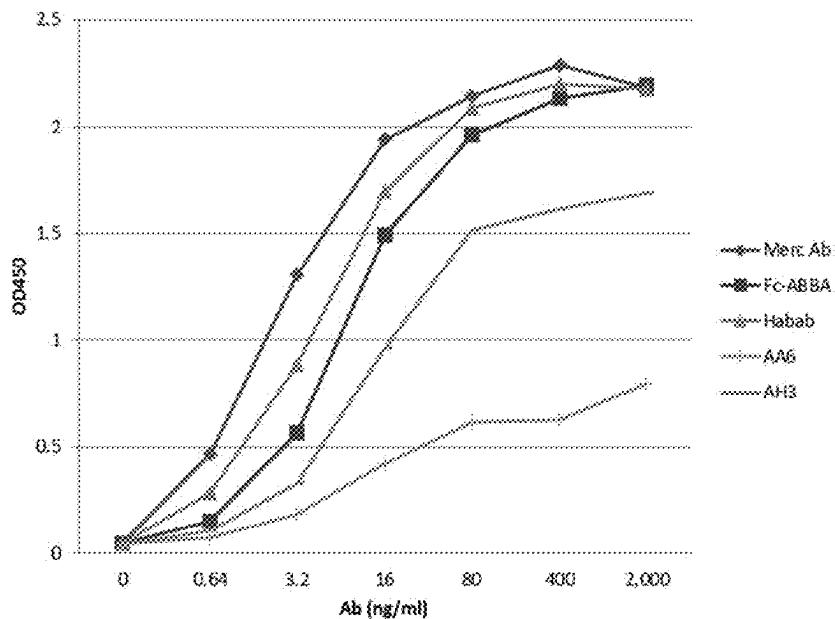
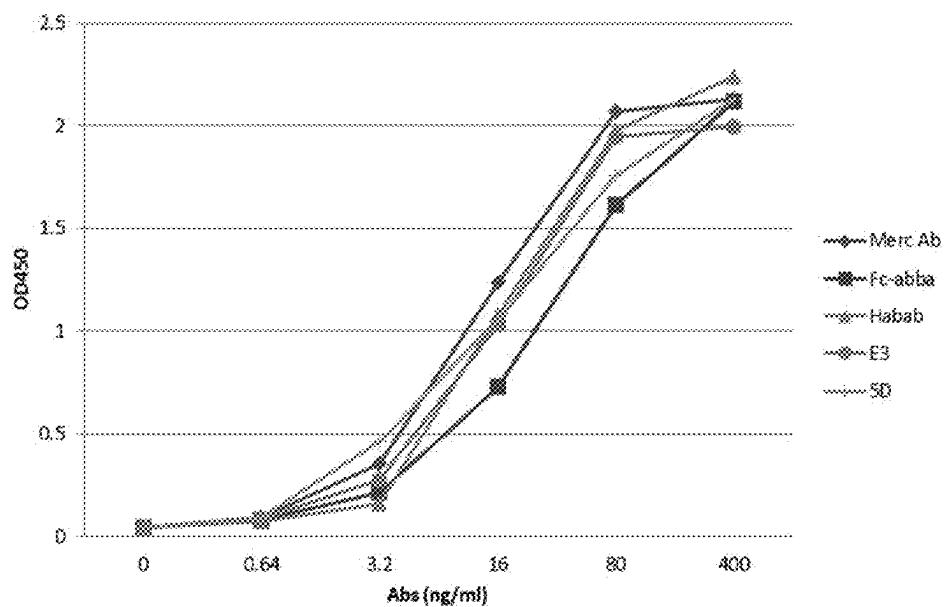
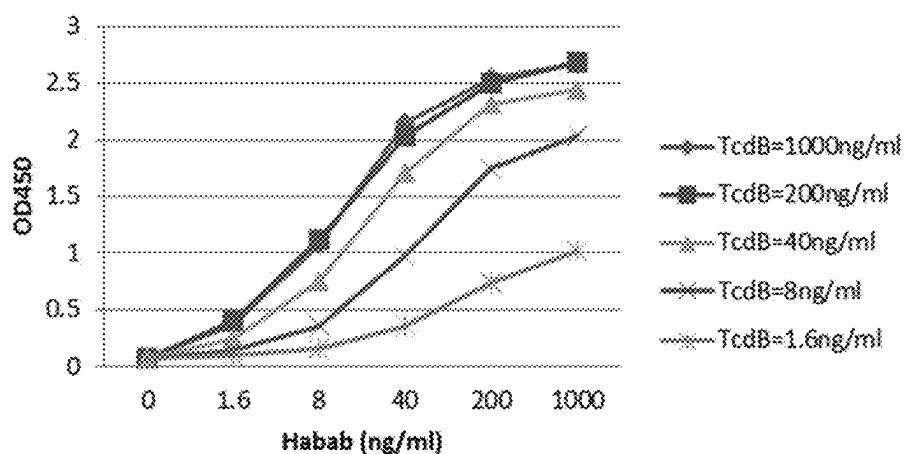
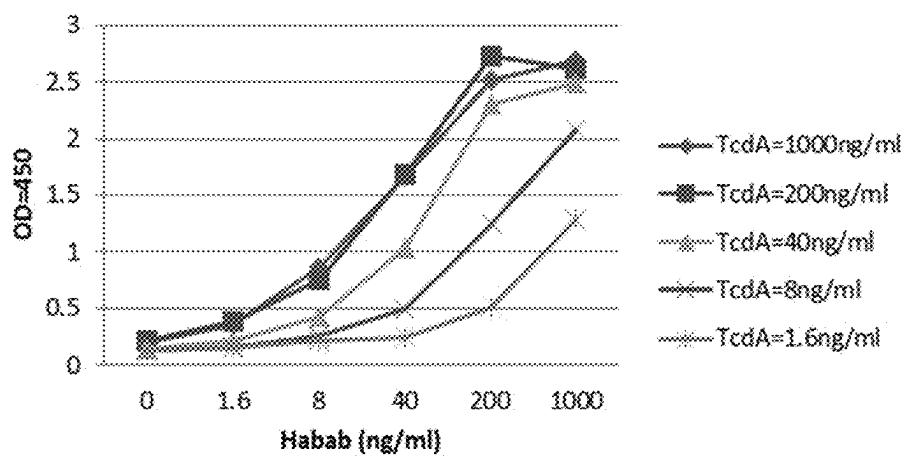


FIG. 7

**Habab binds to TxA****FIG. 8A****FIG. 8B**

**Coat TcdA to detect TcdB****FIG. 9A****Coat TcdB to detect TcdA****FIG. 9B**

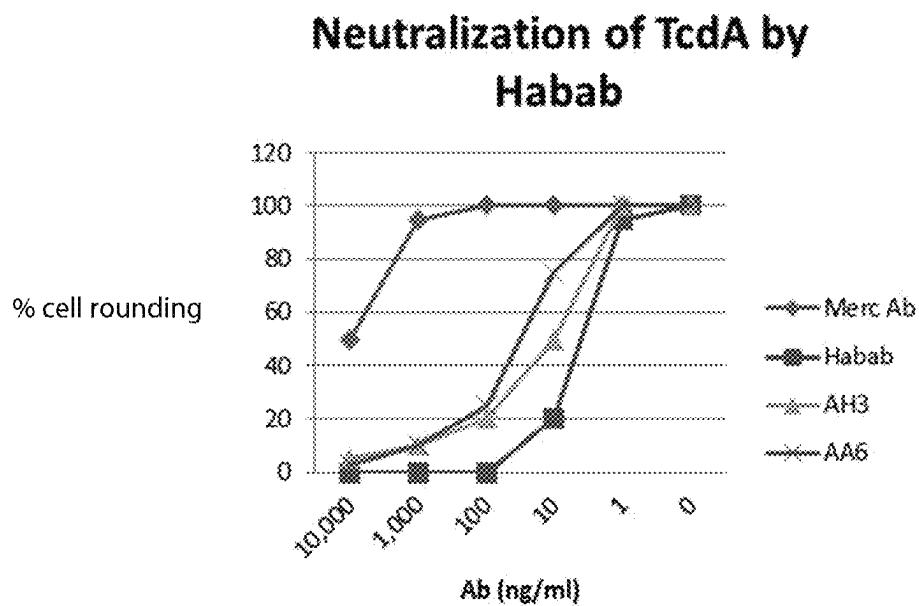


FIG. 10A

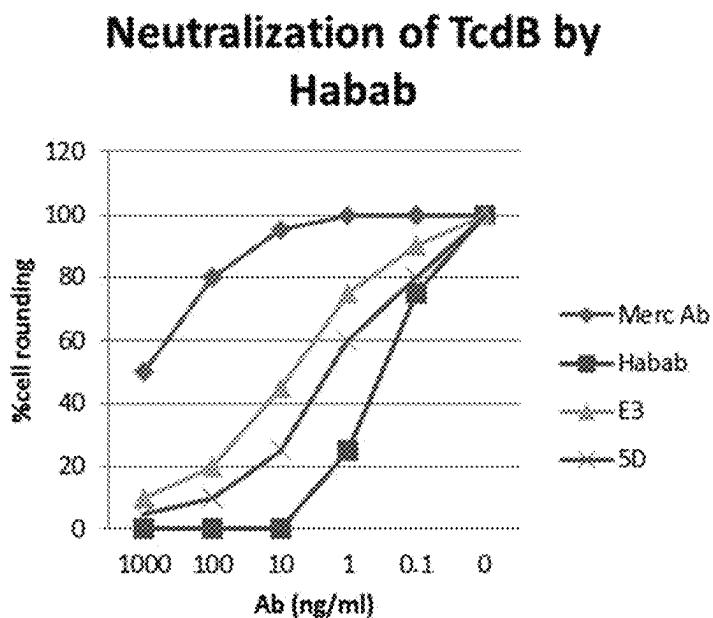


FIG. 10B

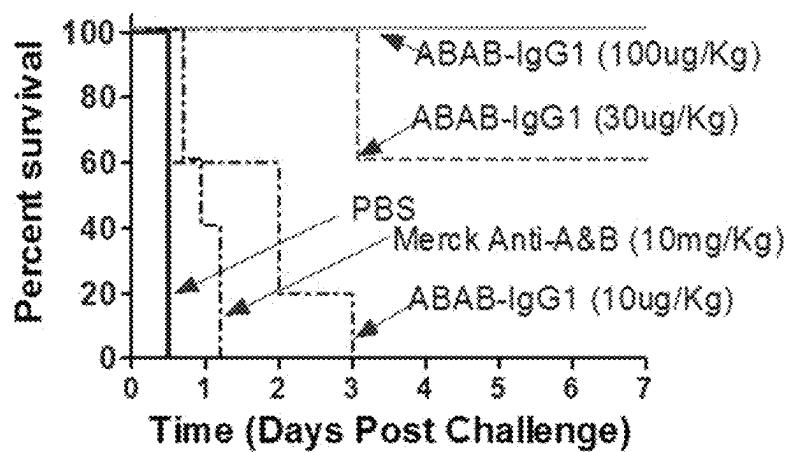


FIG. 11

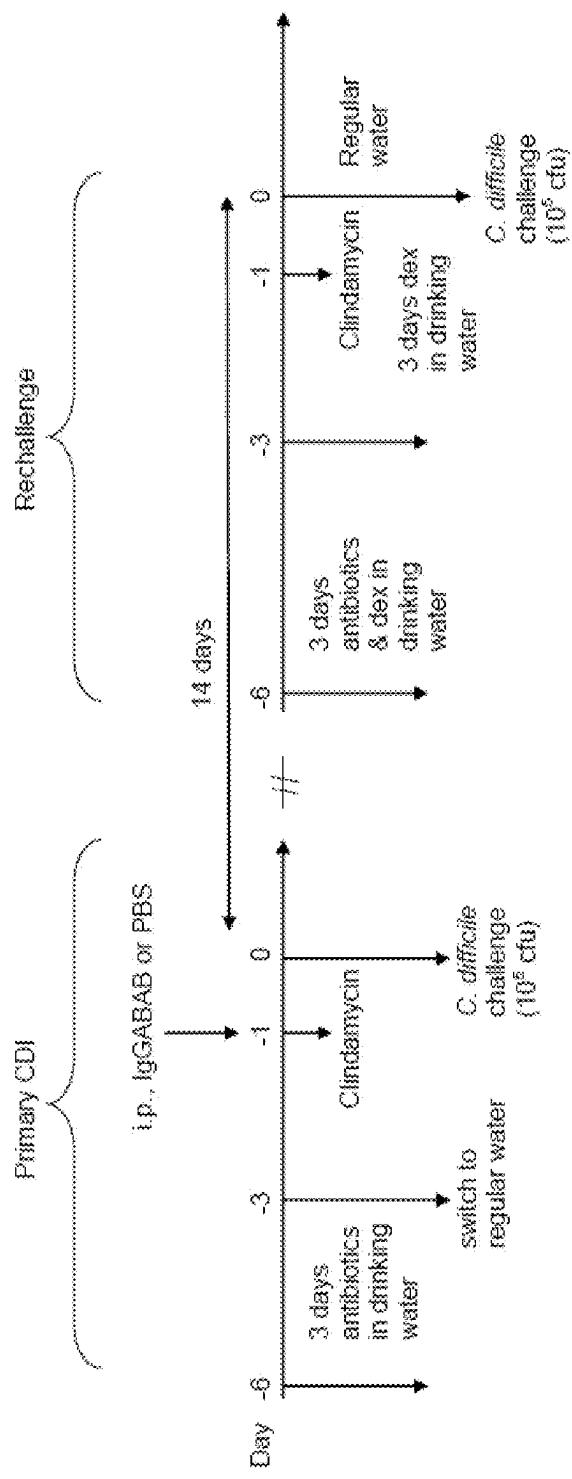


FIG. 12

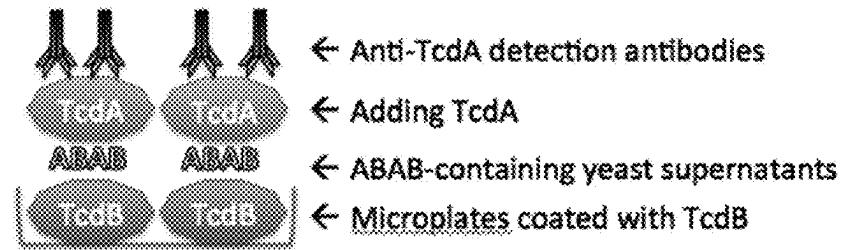


FIG. 13A

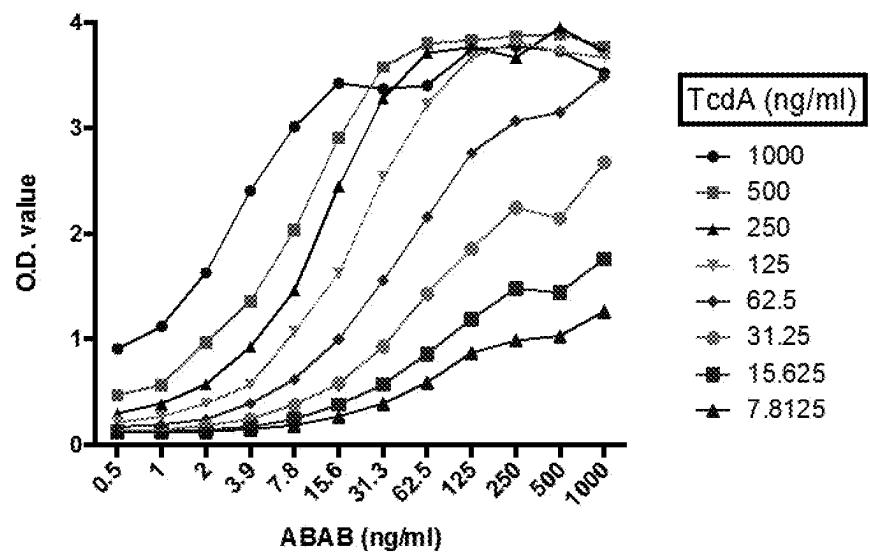


FIG. 13B

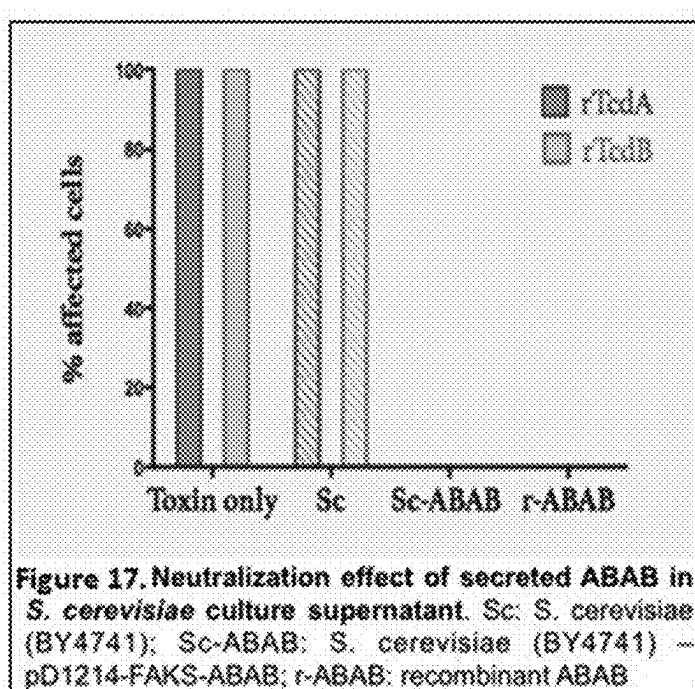


FIG. 14A

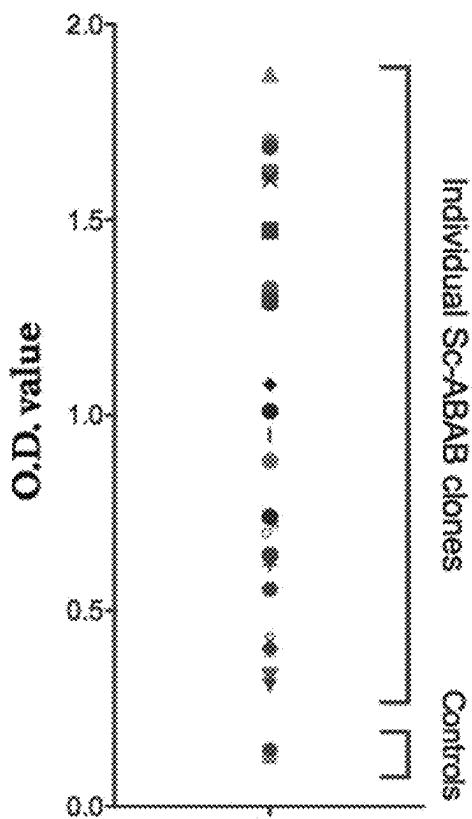


FIG. 14B

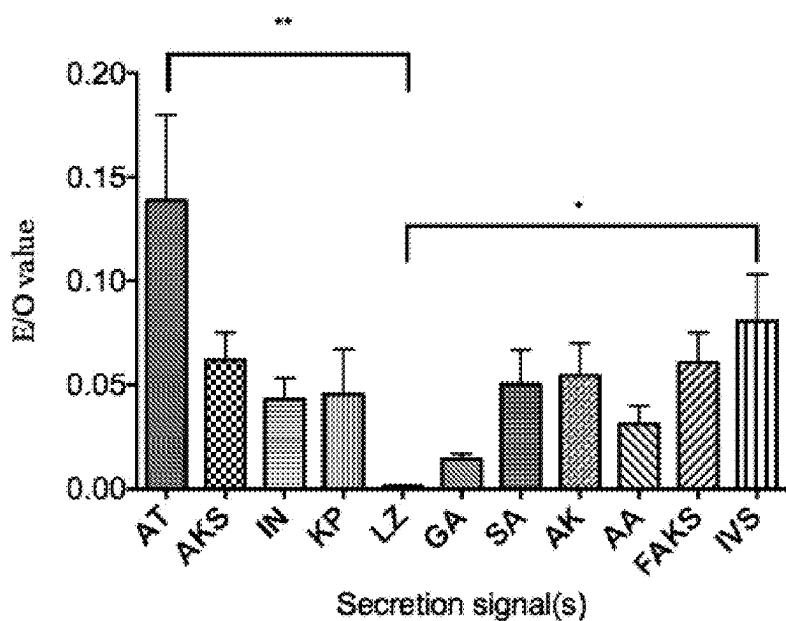


FIG. 15A

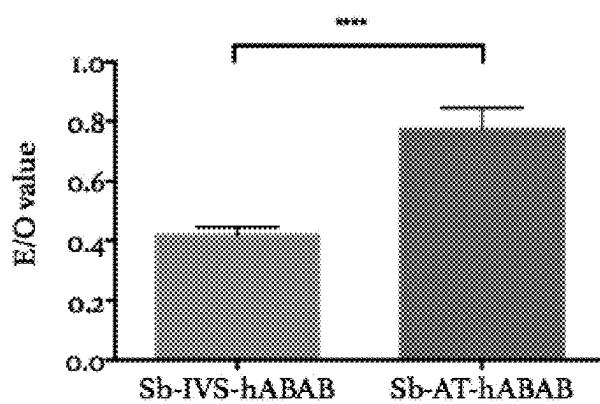
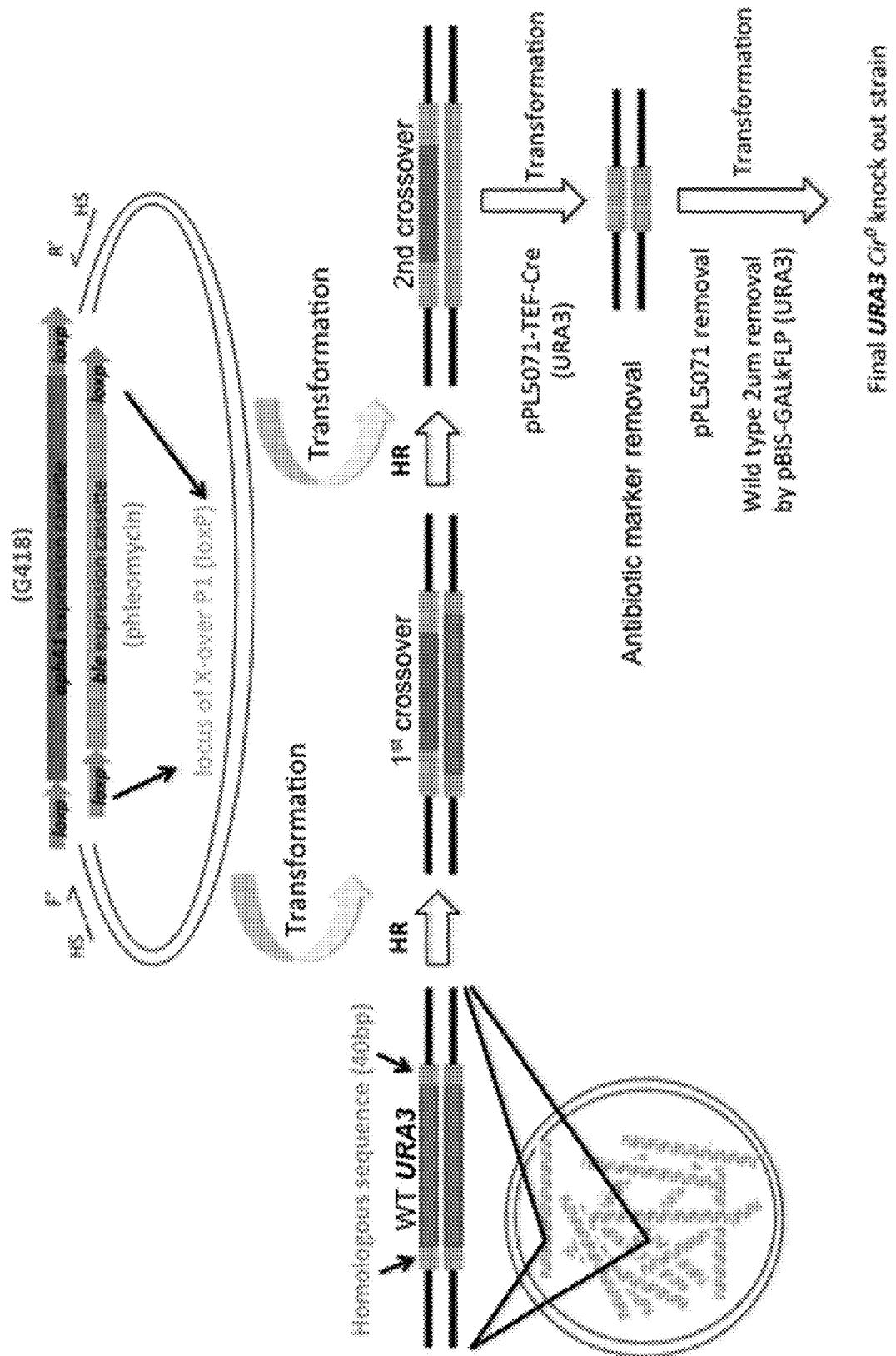


FIG. 15B



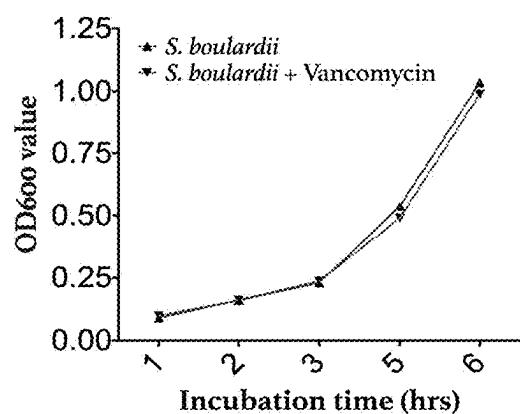


FIG. 17A

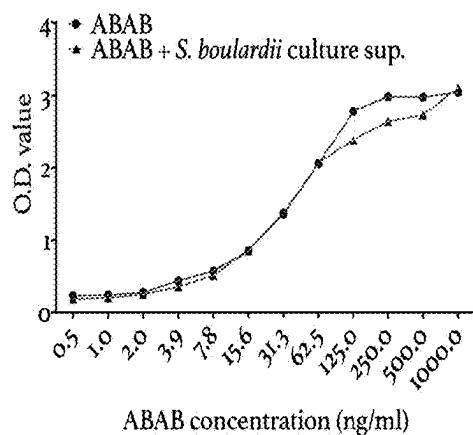


FIG. 17B

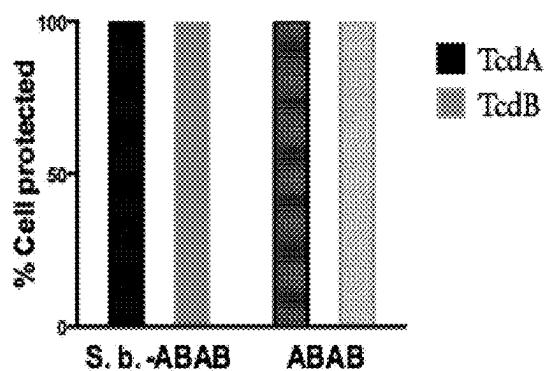


FIG. 17C

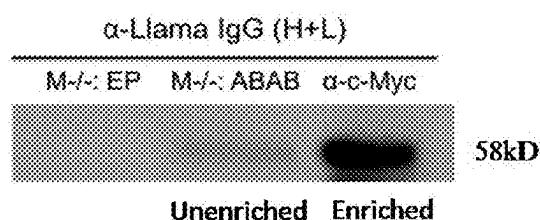


FIG. 17D

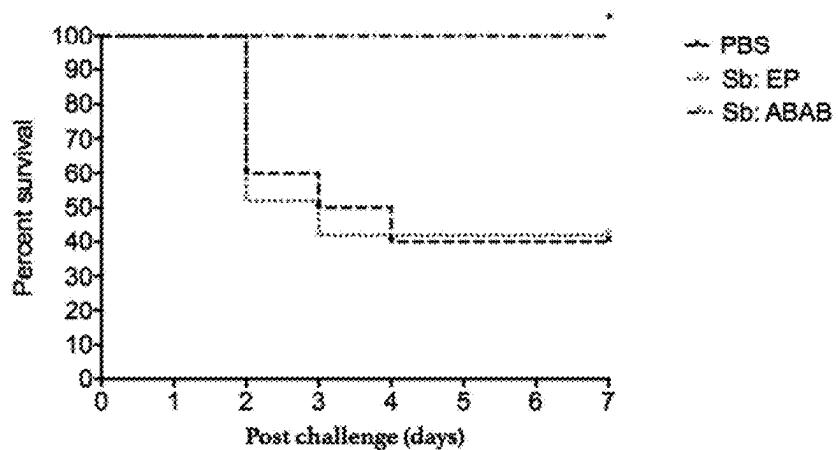


FIG. 18A

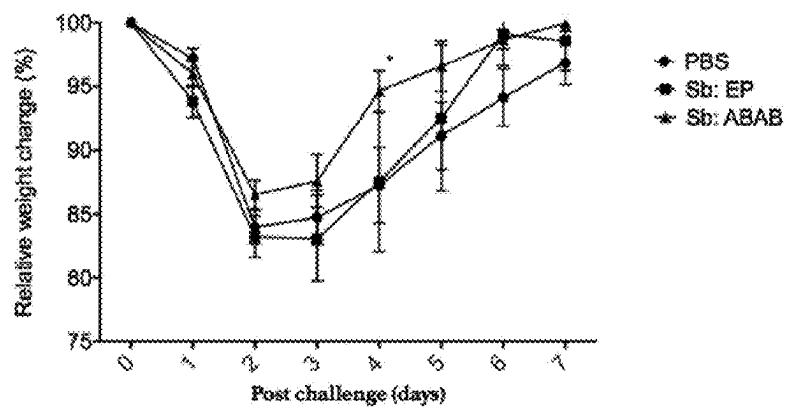


FIG. 18B

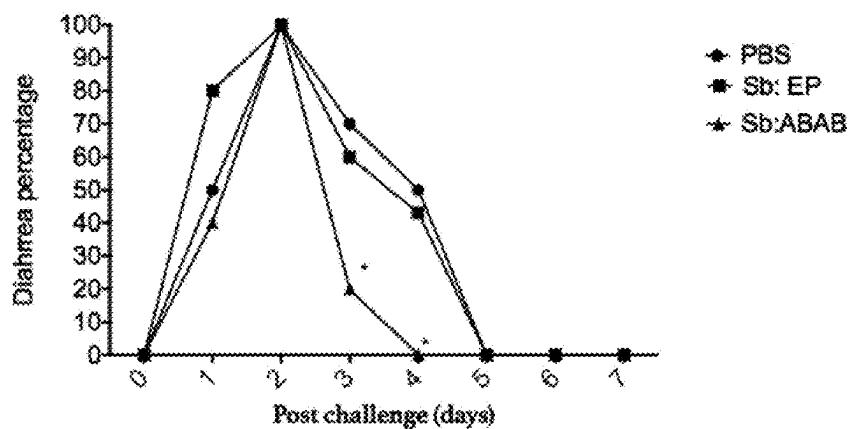


FIG. 18C

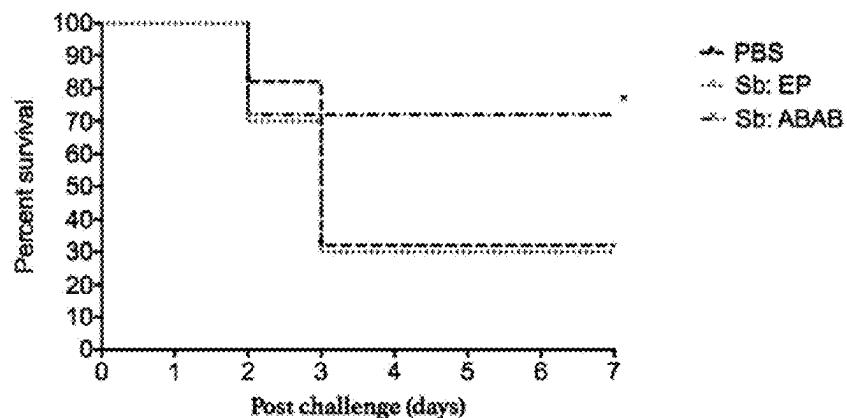


FIG. 19A

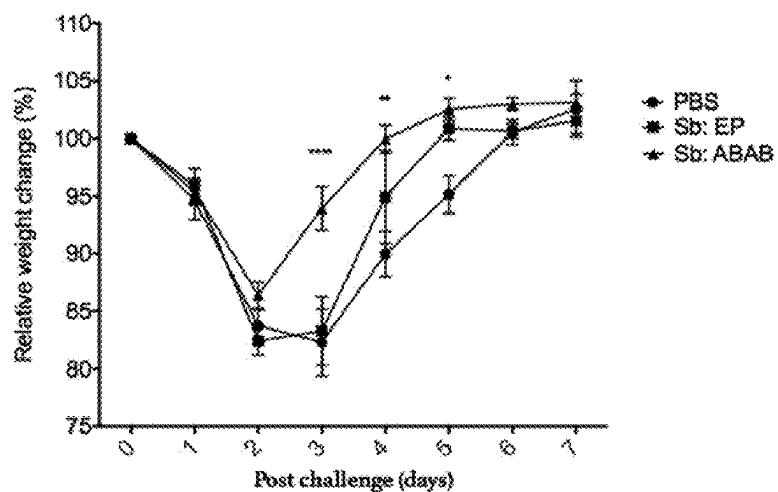


FIG. 19B

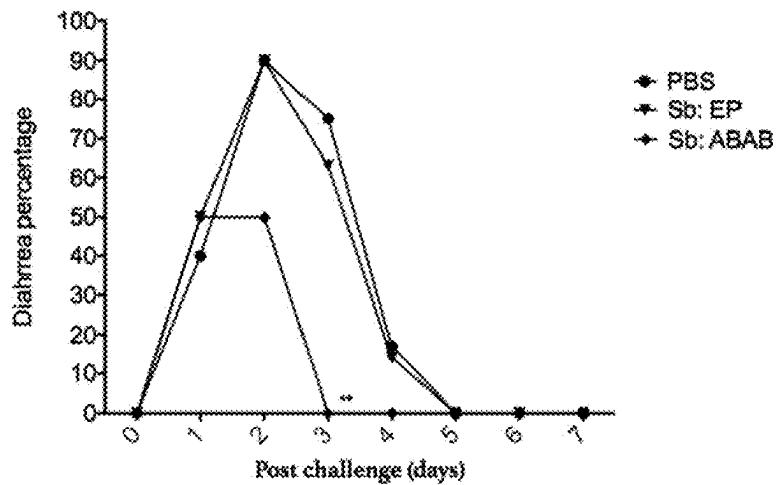


FIG. 19C

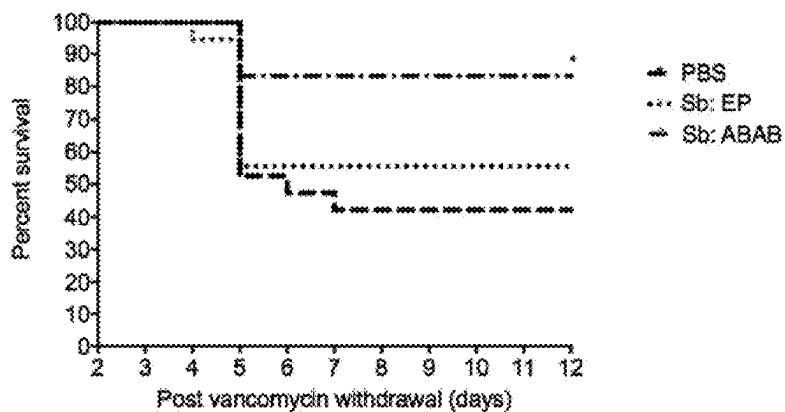


FIG. 20A

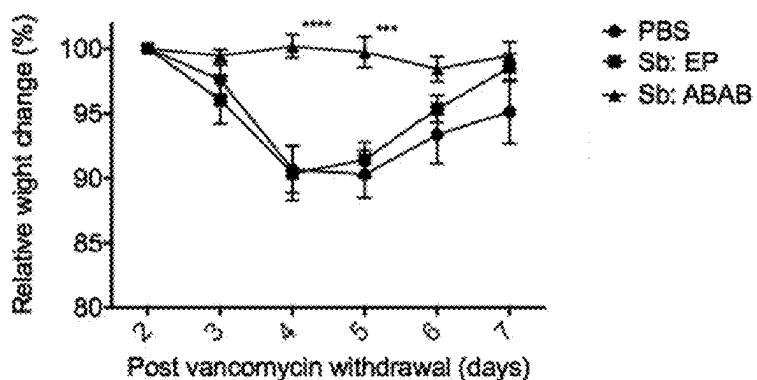


FIG. 20B

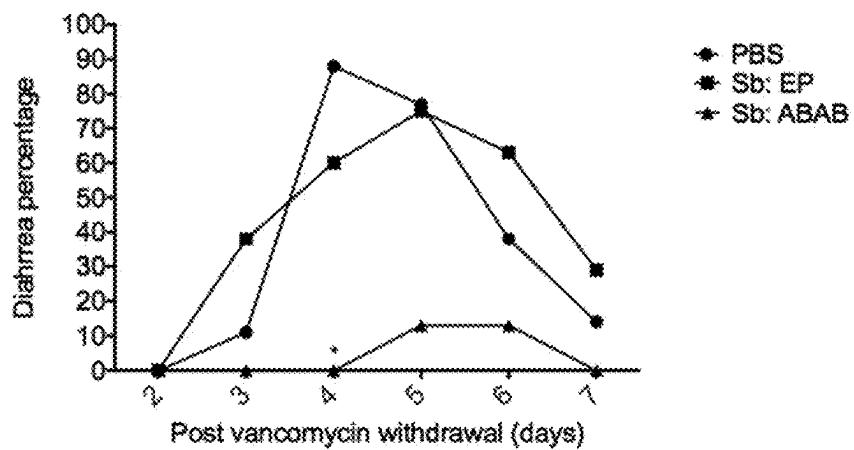


FIG. 20C

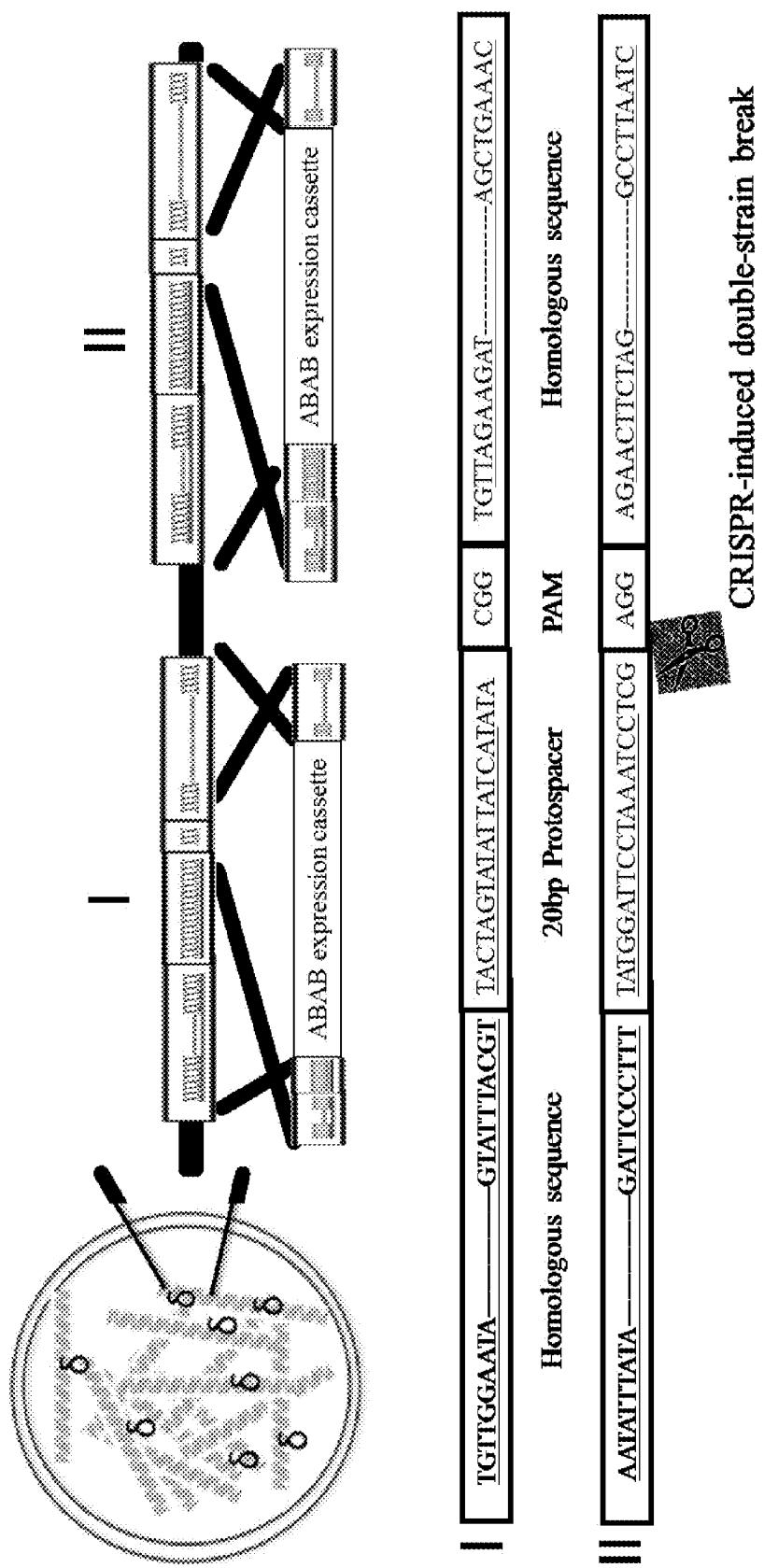


FIG. 21

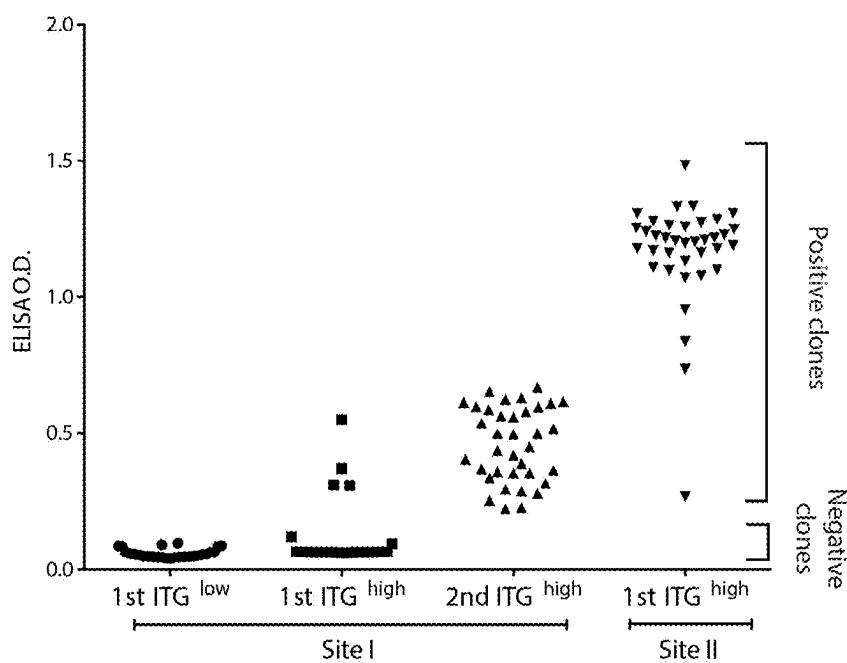


FIG. 22A

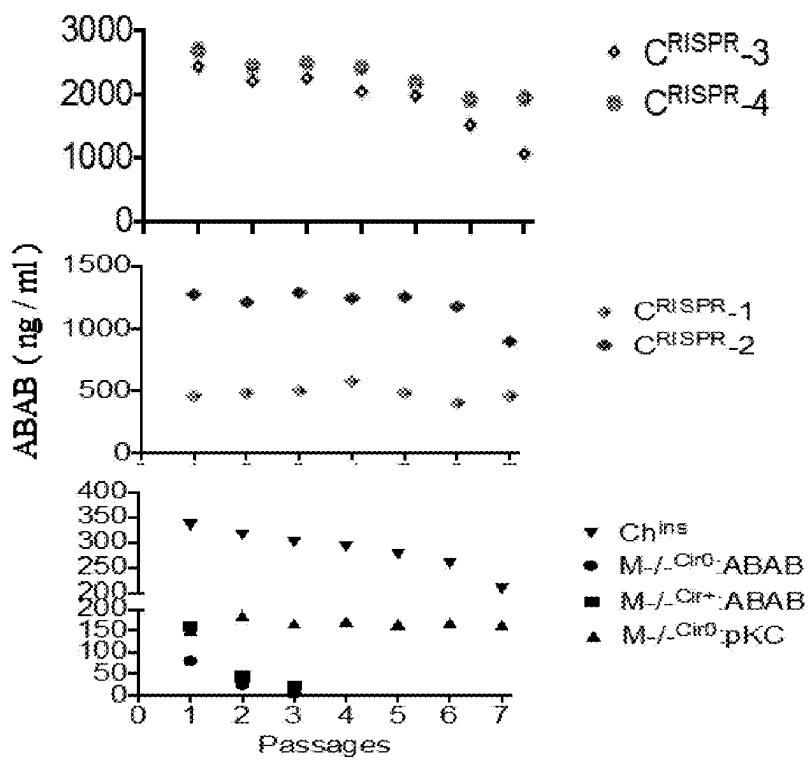


FIG. 22B

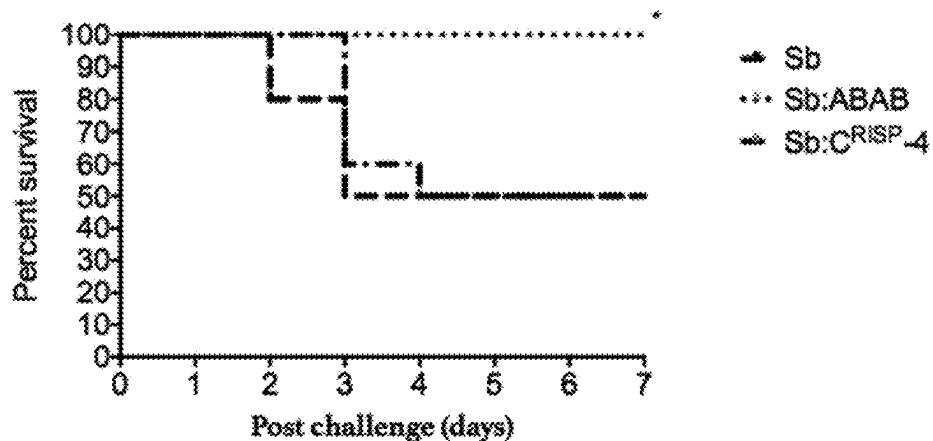


FIG. 23A

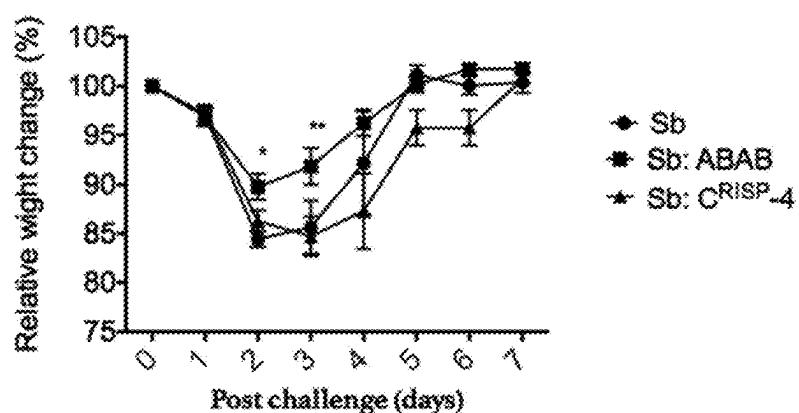


FIG. 23B

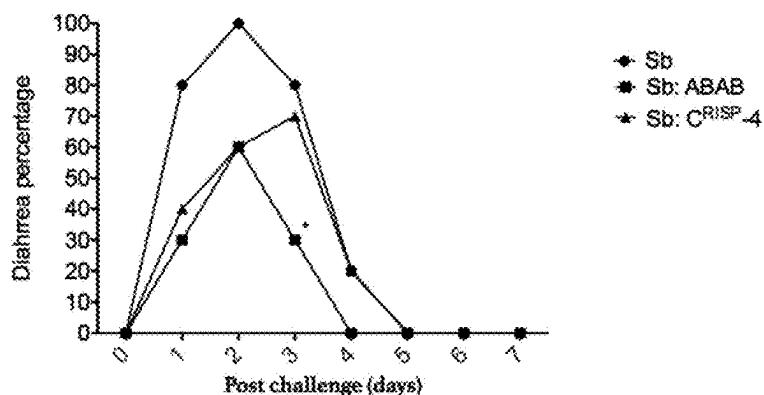


FIG. 23C

**1**

**YEAST-BASED IMMUNOTHERAPY  
AGAINST *CLOSTRIDIUM DIFFICILE*  
INFECTION**

**CROSS-REFERENCE TO RELATED  
APPLICATIONS**

This application is a divisional of U.S. Non-Provisional application Ser. No. 15/768,331, filed on Apr. 13, 2018, which is a national stage of International Patent Application No. PCT/US2016/056875, filed Oct. 13, 2016, which claims priority to U.S. Provisional Application No. 62/240,810, filed Oct. 13, 2015, the contents of each which are hereby incorporated by reference in their entirety.

**STATEMENT OF FEDERALLY SPONSORED  
RESEARCH AND DEVELOPMENT**

This invention was made with government support under Grant Numbers DK084509 and AI109776 awarded by the National Institutes of Health. The government has certain rights in the invention.

**SEQUENCE LISTING**

A sequence listing in electronic ST.26 XML format is filed with this application and incorporated herein by reference. The name of the XML file is “130507-0127\_ST26\_SL”; the file was created on Jan. 23, 2023; the size of the file is 378 KB.

**BACKGROUND**

The bacterium *Clostridium difficile* is the most common cause of nosocomial antibiotic-associated diarrhea as well as the etiologic agent of pseudomembranous colitis [1]. It is estimated that over 500,000 cases of *C. difficile*-associated disease (CDI) occur annually in the United States, with the annual mortality rate ranging from about 3-17%, depending on the strains. With the emergence of hypervirulent and antibiotic-resistant strains, the incidence of mortality in CDI patients is increasing rapidly [2].

CDI is mainly caused by the two *C. difficile* exotoxins TcdA and TcdB (as TcdA-TcdB-strains are avirulent) [21, 22]. The two toxins are structurally similar and exhibit a similar mode of action on host cells. Both toxins target host Rho GTPases, leading to their inactivation as well as cytoskeleton disorganization. The relative roles of the two toxins in the pathogenesis of CDI are not well understood, but it is clear that either toxin individually can cause CDI in animals [22,23].

The options for treating CDI patients are limited and the recurrence rate is high (20-35% of patients). Current standard treatment for CDI using antibiotics causes the disruption of microflora and results in a relapse rate approaching 35% [3,13]. While other interventions have been tried (e.g., probiotics, toxin-absorbing polymers, and toxoid vaccines), neither prevention nor treatment strategies have kept up with the increased incidence and severity of this infection. The risk of further episodes of CDI in recurrent patients can be more than 50% [14] and a subset of patients will have multiple recurrences. Recurrent CDI can be caused by the same strain or newly colonizing strains [15-18].

Newer immune-based therapies have been shown to be somewhat effective in clinical trials, including intravenous immunoglobulin (IVIG) against severe CDI [4-8] and human monoclonal antibodies against recurrent CDI [9].

**2**

Fidaxomicin, a narrow spectrum macrocyclic antibiotic, showed an effect similar to oral vancomycin on CDI but was significantly better at lowering the relapse rate [10]. Fecal transplantation is effective against refractory and recurrent CDI, but it is difficult to standardize and it is associated with risks [11,12].

CDI is a frustrating condition that is difficult to treat and may affect patients for months or even years, causing tremendous morbidity and mortality [19]. Accordingly, there is a need for new treatments for CDI, and means for preventing both primary and recurrent CDI in subjects at risk of developing CDI.

**BRIEF SUMMARY OF INVENTION**

Provided herein are antibody-based fusion protein binding agents that selectively bind *C. difficile* virulence factors TcdA and TcdB, and strains of the probiotic yeast *Saccharomyces* genetically engineered to express and secrete these *C. difficile* toxin binding agents. Both the yeast and the binding agents show utility in treating and preventing primary and recurrent CDI in a subject. Orally administered *Saccharomyces* secreting the binding agents in host intestines can relieve ongoing CDI and prevent recurrence.

The present invention is thus directed to *C. difficile* toxin binding agents, strains of *Saccharomyces* including, but not limited to, *Saccharomyces boulardii* engineered to produce the binding agents, methods of making the engineered strains of yeast, and methods of treating and preventing primary and recurrent CDI using the binding agents and the engineered strains of yeast, among other important features. Binding Agents

The binding agents of the present invention include simple V<sub>H</sub>H peptide monomers and linked groups of V<sub>H</sub>H peptide monomers (comprising 2, 3, 4, or more monomers), as well as more complex binding agents that comprise V<sub>H</sub>H peptide monomers joined to antibody Fc domains, as well as V<sub>H</sub>H peptide monomers joined to partial or full IgG antibodies.

In a first embodiment, the present invention is directed to binding agents comprising V<sub>H</sub>H peptide monomers and linked groups of V<sub>H</sub>H peptide monomers comprising two, three, four, or more monomers, each of which binds TcdA and/or TcdB, preferably with specificity. Thus, the invention encompasses V<sub>H</sub>H peptide binding agents comprising at least one V<sub>H</sub>H peptide monomer, wherein each V<sub>H</sub>H peptide monomer has binding specificity for an epitope of *C. difficile* toxin A (TcdA) or toxin B (TcdB). In certain aspects, these binding agents comprise two, three, four, or more linked V<sub>H</sub>H peptide monomers. The V<sub>H</sub>H peptide monomers include, but are not limited to, the V<sub>H</sub>H peptide monomers 5D (SEQ ID NO:1), E3 (SEQ ID NO:3), AA6 (SEQ ID NO:5), and AH3 (SEQ ID NO:7).

In aspects of this embodiment where two or more monomer are linked, the monomers may be linked by flexible peptide linkers, generally comprising between 10 and 20 amino acids. Suitable linkers include, but are not limited to, linker-1 (SEQ ID NO:9), linker-2 (SEQ ID NO:11), and linker-3 (SEQ ID NO:13).

In certain aspects of this embodiment, the binding agents bind to TcdA and/or TcdB with specificity. In certain aspects of this embodiment, the binding agents exhibit TcdA and/or TcdB neutralizing activity.

In a specific aspect of this embodiment, the binding agent comprises four linked V<sub>H</sub>H peptide monomers where two of the monomers have binding specificity for epitopes of TcdA and two of the monomers have binding specificity for

epitopes of TcdB. The epitopes of TcdA may be the same or different. The epitopes of TcdB may be the same or different.

In a specific aspect of this embodiment, the binding agent comprises the amino acid sequence set forth in SEQ ID NO:19 or a sequence variant thereof having at least 95% sequence identity thereto, and wherein the sequence variant retains TcdA and/or TcdB binding specificity, or the sequence variant retains toxin neutralizing activity, or both. In some instances, variant amino acids of the sequence variant are located in framework regions of the  $V_{H}H$  peptide monomers.

In a second embodiment, the invention is directed to binding agents comprising  $V_{H}H$  peptide monomers joined to IgG antibodies, where the binding agents bind TcdA and/or TcdB. In these IgG-based binding agents, the variable regions of the light and heavy chains of IgG antibodies are replaced by one, two, three, four or more of the  $V_{H}H$  peptide monomers.

In certain aspects of this embodiment, these binding agents comprise two, three, four, or more linked  $V_{H}H$  peptide monomers joined to the amino termini of IgG light and heavy chains in place of the variable regions. The  $V_{H}H$  peptide monomers include, but are not limited to, the  $V_{H}H$  peptide monomers 5D (SEQ ID NO:1), E3 (SEQ ID NO:3), AA6 (SEQ ID NO:5), and AH3 (SEQ ID NO:7).

In aspects of this embodiment where two or more monomer are linked, the monomers may be linked by flexible peptide linkers, generally comprising between 10 and 20 amino acids. Suitable linkers include, but are not limited to, linker-1 (SEQ ID NO:9), linker-2 (SEQ ID NO:11), and linker-3 (SEQ ID NO:13).

In a first sub-embodiment, the invention is directed to tetra-specific, octameric binding agents comprising an IgG antibody, two sets of linked first and second  $V_{H}H$  peptide monomers, and two sets of linked third and fourth  $V_{H}H$  peptide monomers, wherein the IgG antibody comprises two arms, each arm comprising a heavy chain lacking a variable region and a light chain lacking a variable region, and each chain having an amino terminus, wherein for each arm of the antibody, one set of linked first and second  $V_{H}H$  peptide monomers is joined to the amino terminus of the light chain, and one set of linked third and fourth  $V_{H}H$  peptide monomers is joined to the amino terminus of the heavy chain, and wherein the  $V_{H}H$  peptide monomers have binding specificity for an epitope of *Clostridium difficile* toxin A (TcdA) or toxin B (TcdB). This binding agent is termed "tetra-specific" as it recognizes four different toxin epitopes. It is termed "octameric" as it bears eight  $V_{H}H$  peptide monomers (two copies of the first monomer, two copies of the second monomer, two copies of the third monomer, and two copies of the fourth monomer).

In this sub-embodiment, the first, second, third and fourth  $V_{H}H$  peptide monomers each has binding specificity for a different epitope.

In certain aspects of this sub-embodiment, two of the  $V_{H}H$  peptide monomers have binding specificity for epitopes of TcdA and two of the  $V_{H}H$  peptide monomers have binding specificity for epitopes of TcdB.

In certain aspects of this sub-embodiment, the  $V_{H}H$  peptide monomers independently have binding specificity for an epitope in the glucosyltransferase domain, cysteine protease domain, translocation domain or receptor binding domain of TcdA or TcdB.

In a specific aspect of this sub-embodiment, the light (kappa) chain of the binding agent comprises the amino acid sequence set forth in SEQ ID NO:46 (AA6/E3 kappa) or a sequence variant having at least 95% sequence identity

thereto, and the heavy chain of the binding agent comprises the amino acid sequence set forth in SEQ ID NO:44 (AH3/5D heavy) or a sequence variant having at least 95% sequence identity thereto. As this binding agent is an IgG-based binding agent, it will be clear to the skilled artisan that two heavy chain polypeptides and two light chain polypeptides, having the noted amino acid sequences, will assemble through disulfide bonding to provide the complete binding agent. The sequence variants retain TcdA and/or TcdB binding specificity, or the sequence variants retain toxin-neutralizing activity, or both. The variant amino acids of the sequence variants may be located in framework regions of the  $V_{H}H$  peptide monomers.

In a second sub-embodiment, the invention is directed to bi-specific or tetra-specific, tetrameric binding agents comprising an IgG antibody and first, second, third and fourth  $V_{H}H$  peptide monomers, wherein the IgG antibody comprises two arms, each arm comprising a heavy chain lacking a variable region and a light chain lacking a variable region, and each chain having an amino terminus, wherein for a first arm of the antibody, the first  $V_{H}H$  peptide monomer is joined to the amino terminus of the light chain, and the second  $V_{H}H$  peptide monomer is joined to the amino terminus of the heavy chain, wherein for a second arm of the antibody, the third  $V_{H}H$  peptide monomer is joined to the amino terminus of the light chain, and the fourth  $V_{H}H$  peptide monomer is joined to the amino terminus of the heavy chain, and wherein the  $V_{H}H$  peptide monomers have binding specificity for an epitope of *Clostridium difficile* toxin A (TcdA) or toxin B (TcdB). When the binding agent is "tetra-specific", it recognizes four different toxin epitopes; when "bi-specific" it recognizes two different toxin epitopes. The binding agents are "tetrameric" as they bear four  $V_{H}H$  peptide monomers (when bi-specific, the first and third monomer have the same sequence and bind the same epitope, and the second and fourth monomers have the same sequence and bind the same epitope; when tetra-specific, each of the monomers has a different sequence and binds a different epitope).

When the binding agent is bi-specific, the first and second monomers have binding specificity for different epitopes, the first and third monomers have identical amino acid sequences, and the second and fourth monomers have identical amino acid sequences. One of the  $V_{H}H$  peptide monomers may have binding specificity for an epitope of TcdA and one of the  $V_{H}H$  peptide monomers may have binding specificity for an epitope of TcdB.

When the binding agent is tetra-specific, each of the  $V_{H}H$  peptide monomers has binding specificity for a different epitope. Two of the  $V_{H}H$  peptide monomers may have binding specificity for epitopes of TcdA and two of the  $V_{H}H$  peptide monomers may have binding specificity for epitopes of TcdB.

In certain aspects of this sub-embodiment, each of the  $V_{H}H$  peptide monomers has binding specificity for epitopes of TcdA.

In certain aspects of this sub-embodiment, each of the  $V_{H}H$  peptide monomers has binding specificity for epitopes of TcdB.

In certain aspects of this sub-embodiment, the  $V_{H}H$  peptide monomers independently have binding specificity for an epitope in the glucosyltransferase domain, cysteine protease domain, translocation domain or receptor binding domain of TcdA or TcdB.

In a specific aspect of this sub-embodiment, the light (kappa) chain of the binding agent comprises the amino acid sequence set forth in SEQ ID NO:40 (AA6 kappa) or a

sequence variant having at least 95% sequence identity thereto, and the heavy chain of the binding agent comprises the amino acid sequence set forth in SEQ ID NO:36 (AH3 heavy) or a sequence variant having at least 95% sequence identity thereto. As this binding agent is an IgG-based binding agent, it will be clear to the skilled artisan that two heavy chain polypeptides and two light chain polypeptides, having the noted amino acid sequences, will assemble through disulfide bonding to provide the complete binding agent. The sequence variants retain TcdA and/or TcdB binding specificity, or the sequence variants retain toxin neutralizing activity, or both. The variant amino acids of the sequence variant may be located in framework regions of the  $V_H$  peptide monomers.

In another specific aspect of this sub-embodiment, the light (kappa) chain of the binding agent comprises the amino acid sequence set forth in SEQ ID NO:42 (E3 kappa) or a sequence variant having at least 95% sequence identity thereto, and the heavy chain of the binding agent comprises the amino acid sequence set forth in SEQ ID NO:38 (5D heavy) or a sequence variant having at least 95% sequence identity thereto. As this binding agent is an IgG-based binding agent, it will be clear to the skilled artisan that two heavy chain polypeptides and two light chain polypeptides, having the noted amino acid sequences, will assemble through disulfide bonding to provide the complete binding agent. The sequence variants retain TcdA and/or TcdB binding specificity, or the sequence variants retain toxin neutralizing activity, or both. The variant amino acids of the sequence variants may be located in framework regions of the  $V_H$  peptide monomers.

In certain aspects of this embodiment and the sub-embodiments, the binding agents bind to TcdA and/or TcdB with specificity. In certain aspects of this embodiment, the binding agents exhibit TcdA and/or TcdB neutralizing activity.

In a third embodiment, the invention is directed to binding agents comprising  $V_H$  peptide monomers joined to antibody Fe domains, where the binding agents bind TcdA and/or TcdB. In these Fe domain-based binding agents, one, two, three, four or more of the  $V_H$  peptide monomers are joined to the hinge,  $C_{H2}$  and  $C_{H3}$  regions of each arm of Fc domain of an antibody heavy chain. Thus, the peptide monomers replace the Fab regions of an antibody.

In certain aspects of this embodiment, these binding agents comprise two, three, four, or more linked  $V_H$  peptide monomers joined to the amino termini of the arms of the Fc domains. The  $V_H$  peptide monomers include, but are not limited to, the  $V_H$  peptide monomers 5D (SEQ ID NO:1), E3 (SEQ ID NO:3), AA6 (SEQ ID NO:5) and AH3 (SEQ ID NO:7).

In aspects of this embodiment where two or more monomer are linked, the monomers may be linked by flexible peptide linkers, generally comprising between 10 and 20 amino acids. Suitable linkers include, but are not limited to, linker-1 (SEQ ID NO:9), linker-2 (SEQ ID NO:11), and linker-3 (SEQ ID NO:13).

In a first sub-embodiment, the invention is directed to tetra-specific, octameric binding agents comprising an antibody Fc domain and two sets of linked first, second, third and fourth  $V_H$  peptide monomers, wherein the antibody Fc domain comprises two arms, each arm comprising hinge,  $C_{H2}$  and  $C_{H3}$  regions of an antibody heavy chain, and each arm having an amino terminus, wherein for each arm of the Fc domain, one set of linked first, second, third and fourth  $V_H$  peptide monomers is joined to the amino terminus of the arm, and where the  $V_H$  peptide monomers have binding specificity for an epitope of *Clostridium difficile* toxin A (TcdA) or toxin B (TcdB). This binding agent is termed “tetra-specific” as it recognizes four different toxin epitopes. It is termed “octameric” as it bears eight  $V_H$  peptide monomers (two copies of the first monomer, two copies of the second monomer, two copies of the third monomer, and two copies of the fourth monomer).

In certain aspects of this sub-embodiment, the first, second, third and fourth  $V_H$  peptide monomers each has binding specificity for a different epitope.

In certain aspects of this sub-embodiment, two of the  $V_H$  peptide monomers have binding specificity for epitopes of TcdA and two of the  $V_H$  peptide monomers have binding specificity for epitopes of TcdB.

In certain aspects of this sub-embodiment, the  $V_H$  peptide monomers independently have binding specificity for an epitope in the glucosyltransferase domain, cysteine protease domain, translocation domain or receptor binding domain of TcdA or TcdB.

In a specific aspect of this sub-embodiment, the binding agent comprises the amino acid sequence set forth in SEQ ID NO:22 (ABAB-Fc) or a sequence variant having at least 95% sequence identity thereto, where the sequence variant retains TcdA and/or TcdB binding specificity, or the sequence variant retains toxin neutralizing activity, or both. As this binding agent is an Fc domain-based binding agent, it will be clear to the skilled artisan that two identical polypeptides, having the noted amino acid sequence, serve as the arms of the binding agent and that the arms will assemble through disulfide bonding to provide the complete binding agent. The variant amino acids of the sequence variant may be located in framework regions of the  $V_H$  peptide monomers.

In a second sub-embodiment, the invention is directed to bi-specific, tetrameric binding agents comprising an antibody Fc domain and two sets of linked first and second  $V_H$  peptide monomers, wherein the antibody Fc domain comprises two arms, each arm comprising hinge,  $C_{H2}$  and  $C_{H3}$  regions of an antibody heavy chain, and each arm having an amino terminus, wherein for each arm of the Fc domain, one set of linked first and second  $V_H$  peptide monomers is joined to the amino terminus of the arm, and where the  $V_H$  peptide monomers have binding specificity for an epitope of *Clostridium difficile* toxin A (TcdA) or toxin B (TcdB). This binding agent is termed “bi-specific” as it recognizes two different toxin epitopes. It is termed “tetrameric” as it bears four  $V_H$  peptide monomers (two copies of the first monomer, and two copies of the second monomer).

In certain aspects of this sub-embodiment, the first and second  $V_H$  peptide monomers have binding specificity for the same or different epitopes.

In certain aspects of this sub-embodiment, the  $V_H$  peptide monomers independently have binding specificity for an epitope in the glucosyltransferase domain, cysteine protease domain, translocation domain or receptor binding domain of TcdA or TcdB.

In a specific aspect of this sub-embodiment, the binding agent comprises the amino acid sequence set forth in SEQ ID NO:32 (AH3/5D-Fc) or a sequence variant having at least 95% sequence identity thereto, where the sequence variant retains TcdA and/or TcdB binding specificity, or the sequence variant retains toxin neutralizing activity, or both. As this binding agent is an Fc domain-based binding agent, it will be clear to the skilled artisan that two identical polypeptides, having the noted amino acid sequence, serve as the arms of the binding agent and that the arms will assemble through disulfide bonding to provide the complete

binding agent. The variant amino acids of the sequence variant may be located in framework regions of the  $V_H$ H peptide monomers.

In another specific aspect of this sub-embodiment, the binding agent comprises the amino acid sequence set forth in SEQ ID NO:34 (AA6/E3-Fc) or a sequence variant having at least 95% sequence identity thereto, where the sequence variant retains TcdA and/or TcdB binding specificity, or the sequence variant retains toxin neutralizing activity, or both. As this binding agent is an Fc domain-based binding agent, it will be clear to the skilled artisan that two identical polypeptides, having the noted amino acid sequence, serve as the arms of the binding agent and that the arms will assemble through disulfide bonding to provide the complete binding agent. The variant amino acids of the sequence variant may be located in framework regions of the  $V_H$ H peptide monomers.

In certain aspects of this embodiment and the sub-embodiments, the binding agents bind to TcdA and/or TcdB with specificity. In certain aspects of this embodiment, the binding agents exhibit TcdA and/or TcdB neutralizing activity.

The invention includes humanized variants of each the binding agents provided in the various embodiments and aspects defined herein. Likewise, the invention includes epitope binding fragments of each the binding agents provided in the various embodiments and aspects defined herein.

#### Polynucleotides, Expression Vectors, and Host Cells

The invention includes polynucleotides comprising nucleotide sequences encoding each the binding agents provided in the various embodiments and aspects defined herein, as well as complementary strands thereof. The invention also includes expression vectors (e.g., bacterial and yeast) comprising the polynucleotides, and host cells (e.g., bacterial, yeast, mammalian, insect) comprising the expression vectors. The invention further includes methods of producing the binding agents define herein, comprising culturing the host cells under conditions promoting expression of the binding agents encoded by the expression vectors, and recovering the binding agents from the cell cultures.

#### Engineered Strains of *S. boulardii*

In a fourth embodiment, the invention is directed to strains of *Saccharomyces* yeast, such as *S. cerevisiae* and *S. boulardii*, engineered to produce one or more of the binding agents defined herein. In preferred aspects, the engineered strains of *Saccharomyces* yeast secrete the binding agents.

The identity of the *Saccharomyces* yeast strain is only limited in that it can be engineered to produce, and preferably secrete, one or more of the binding agents of the invention. In preferred aspects of the invention, the strain of *Saccharomyces* yeast engineered to produce one or more of the binding agents is *S. cerevisiae* or *S. boulardii*. The invention thus encompasses an engineered strain of *S. cerevisiae* that produces one or more of the binding agents defined herein, as well as an engineered strain of *S. cerevisiae* that secretes one or more of the binding agents defined herein. The invention also encompasses an engineered strain of *S. boulardii* that produces one or more of the binding agents defined herein, as well as an engineered strain of *S. boulardii* that secretes one or more of the binding agents defined herein.

In an example of this embodiment, the invention is directed to engineered strains of *Saccharomyces* yeast that produce a binding agent comprising a  $V_H$ H peptide monomer or linked groups of  $V_H$ H peptide monomers comprising

two, three, four, or more monomers, each of which binds TcdA and/or TcdB, preferably with specificity. Thus, the invention encompasses engineered strains of *Saccharomyces* yeast that produces  $V_H$ H peptide binding agents comprising at least one  $V_H$ H peptide monomer, wherein each  $V_H$ H peptide monomer has binding specificity for an epitope of *C. difficile* toxin A (TcdA) or toxin B (TcdB). In certain aspects, these binding agents comprise two, three, four, or more linked  $V_H$ H peptide monomers. The  $V_H$ H peptide monomers include, but are not limited to, the  $V_H$ H peptide monomers 5D (SEQ ID NO:1), E3 (SEQ ID NO:3), AA6 (SEQ ID NO:5), and AH3 (SEQ ID NO:7).

In another example of this embodiment, the invention is directed to engineered strains of *Saccharomyces* yeast that produce binding agents comprising  $V_H$ H peptide monomers joined to IgG antibodies, where the binding agents bind TcdA and/or TcdB, as defined herein. In these IgG-based binding agents, the variable regions of the light and heavy chains of IgG antibodies are replaced by one, two, three, four or more of the  $V_H$ H peptide monomers.

In further example of this embodiment, the invention is directed to engineered strains of *Saccharomyces* yeast that produce binding agents comprising  $V_H$ H peptide monomers joined to antibody Fc domains, where the binding agents bind TcdA and/or TcdB, as defined herein. In these Fc domain-based binding agents, one, two, three, four or more of the  $V_H$ H peptide monomers are joined to the hinge,  $C_H2$  and  $C_H3$  regions of each arm of Fc domain of an antibody heavy chain. Thus, the peptide monomers replace the Fab regions of an antibody.

In yet another example of this embodiment, the invention is directed to an engineered strain of *Saccharomyces* yeast that produces a tetra-specific, tetrameric binding agent, wherein the binding agent comprises linked first, second, third and fourth  $V_H$ H peptide monomers, and wherein the  $V_H$ H peptide monomers independently have binding specificity for an epitope of *Clostridium difficile* toxin A (TcdA) or toxin B (TcdB). In certain aspects, the first, second, third and fourth  $V_H$ H peptide monomers each has binding specificity for a different epitope. In certain aspects, the two of the  $V_H$ H peptide monomers have binding specificity for epitopes of TcdA and two of the  $V_H$ H peptide monomers have binding specificity for epitopes of TcdB. In certain aspects, the  $V_H$ H peptide monomers independently have binding specificity for an epitope in the glucosyltransferase domain, cysteine protease domain, translocation domain or receptor binding domain of TcdA or TcdB.

In a preferred example of this embodiment, the invention is directed to an engineered strain of yeast, wherein the binding agent is ABAB, wherein the first and third monomers have binding specificity for epitopes of TcdA and the first and third monomers are  $V_H$ H peptide monomers AH3 (SEQ ID NO:7) and AA6 (SEQ ID NO:5), respectively, and wherein the second and forth monomers have binding specificity for epitopes of TcdB and the second and forth monomers are  $V_H$ H peptide monomers 5D (SEQ ID NO:1) and E3 (SEQ ID NO:3), respectively. In certain aspects, the ABAB binding agent comprises the amino acid sequence set forth in SEQ ID NO:19, or a sequence variant having at least 95% sequence identity thereto, wherein the sequence variant retains TcdA and/or TcdB binding specificity, or the sequence variant retains toxin neutralizing activity, or both. In certain aspects, the ABAB binding agent further comprises an N-terminal secretion signal selected from the AT secretion signal (MRFPSIFTAVLFAASSALA (SEQ ID NO:99)) and the IVS secretion signal (MLLQAFLFL-LAGFAAKISA (SEQ ID NO:103)).

In certain aspects, the ABAB binding agent is expressed from a plasmid within the yeast, wherein the ABAB binding agent comprises the amino acid sequence set forth in SEQ ID NO:107, or a sequence variant having at least 95% sequence identity thereto, and wherein the sequence variant retains TcdA and/or TcdB binding specificity, or the sequence variant retains toxin neutralizing activity, or both. The plasmid may be, but is not limited to, pCEV-URA3-TEF-AT-yABAB-cMyc (SEQ ID NO:88).

In certain aspects, the ABAB binding agent coding sequence is integrated into a chromosome of the strain of yeast, wherein the ABAB binding agent comprises the amino acid sequence set forth in SEQ ID NO:109, or a sequence variant having at least 95% sequence identity thereto, and wherein the sequence variant retains TcdA and/or TcdB binding specificity, or the sequence variant retains toxin neutralizing activity, or both.

Aspects of this embodiment include engineered strains of *Saccharomyces* yeast that produce a therapeutic protein having binding specificity for a unique epitope of *Clostridium difficile* toxin A (TcdA) or toxin B (TcdB), or both. Preferably, the engineered strain of *Saccharomyces* yeast is *S. cerevisiae* or *S. boulardii*. A therapeutic protein is any protein that can bring about an improvement or cure in a medical condition in a subject, or that can inhibit or prevent a medical condition from developing in a subject. Suitable therapeutic protein include, but are not limited to, proteins that (a) replace a protein that is deficient or abnormal; (b) augment an existing pathway; (c) provide a novel function or activity; (d) interfere with a molecule or organism; and (e) deliver other compounds or proteins, such as a radionuclide, cytotoxic drug, or effector proteins. Therapeutic proteins also include antibodies and antibody-based drugs, Fc fusion proteins, anticoagulants, blood factors, bone morphogenetic proteins, engineered protein scaffolds, enzymes, growth factors, hormones, interferons, interleukins, and thrombolytics. Therapeutic proteins further include bispecific monoclonal antibodies (mAbs) and multispecific fusion proteins, mAbs conjugated with small molecule drugs, and proteins with optimized pharmacokinetics.

#### Methods of Making Engineered Strains of *S. boulardii*

The invention is also directed to methods of making strains of *Saccharomyces* yeast engineered to produce one or more of the binding agents defined herein.

The invention thus encompasses a method of preparing a strain of *Saccharomyces* yeast engineered to produce one or more of the binding agents defined herein comprising (a) transforming a strain of *Saccharomyces* yeast with an expression vector encoding the binding agent, and (b) screening the yeast of (a) for production of the binding agent. In a certain aspect, the expression vector is plasmid pCEV-URA3-TEF-AT-yABAB-cMyc (SEQ ID NO:88).

The invention thus encompasses a method of preparing a strain of *Saccharomyces* yeast engineered to produce one or more of the binding agents defined herein comprising (a) chromosomally integrating a polynucleotide sequence encoding the binding agent into the genome of the strain of *Saccharomyces* yeast, and (b) screening the yeast of (a) for production of the binding agent. In certain aspects, the chromosomal integration is performed via:

- (a) amplifying a polynucleotide sequence encoding the ABAB binding agent from plasmid pCEV-G4-Km-TEF-AT-yABAB hAA6T83N-tagless (SEQ ID NO:90) using primers containing (i) nucleic acid sequence homologous to a selected yeast chromosomal integration site and (ii) nucleic acid sequence homologous to

regions 5' and 3' of ABAB binding agent coding sequence of the plasmid, to produce an integration cassette,

(b) transforming yeast with the integration cassette produced in (a) with pCRI-Sb-61 (SEQ ID NO:91) or pCRI-Sb-62 (SEQ ID NO:92) to induce a double stranded break within the corresponding yeast chromosomal delta sites under conditions promoting spontaneous integration of the integration cassette into the site of the double stranded break,

(c) screening the transformed yeast of (b) for production of the ABAB binding agent.

In certain aspects of these methods, the strain of *Saccharomyces* yeast engineered to produce the binding agents is an auxotrophic strain of *Saccharomyces* yeast, such as a ura3-strain of yeast. A ura3-strain of yeast can be utilized under ura3 selection.

In certain aspects of these methods, the strain of *Saccharomyces* yeast engineered to produce the binding agents is *S. cerevisiae* or *S. boulardii*.

In certain aspects of these methods, the screening is performed using an immunoassay, such as an ELISA. Pharmaceutical Formulations

The invention includes pharmaceutical formulations comprising one or more of the binding agents defined herein and a pharmaceutically acceptable carrier or diluent. The invention also includes pharmaceutical formulations comprising one or more of the engineered strains of *Saccharomyces* yeast defined herein and a pharmaceutically acceptable carrier or diluent. In certain aspects, the *Saccharomyces* yeast is *S. cerevisiae* or *S. boulardii*.

#### Methods of Treating and Preventing

In a sixth embodiment, the invention is directed to methods of treating or preventing a disease symptom induced by *C. difficile* in a subject comprising administering a therapeutically-effective amount of one or more binding agents and/or one or more engineered strains of *Saccharomyces* yeast as defined herein to a subject having *C. difficile* infection or a risk of developing *C. difficile* infection. In preferred aspects, the *Saccharomyces* yeast is *S. cerevisiae* or *S. boulardii*.

In certain aspects of this embodiment, the disease symptom induced by *C. difficile* is diarrhea.

In a seventh embodiment, the invention is directed to methods of neutralizing *C. difficile* toxin TcdA and/or TcdB in a subject infected by *C. difficile* comprising administering a therapeutically-effective amount of one or more binding agents and/or one or more engineered strains of *Saccharomyces* yeast as defined herein to a subject having *C. difficile* infection. In preferred aspects, the *Saccharomyces* yeast is *S. cerevisiae* or *S. boulardii*.

In an eighth embodiment, the invention is directed to methods of treating or preventing *C. difficile* infection in a subject comprising administering a therapeutically-effective amount of one or more of the binding agents and/or one or more engineered strains of *Saccharomyces* yeast as defined herein to a subject having *C. difficile* infection or a risk of developing *C. difficile* infection. In preferred aspects, the *Saccharomyces* yeast is *S. cerevisiae* or *S. boulardii*. In certain aspects of the eighth embodiment, the method further comprises administering a therapeutically-effective amount of an antibiotic to the subject.

In a ninth embodiment, the invention is directed to methods of maintaining normal bowel function in a subject having a *C. difficile* infection comprising administering a therapeutically-effective amount of one or more of the binding agents and/or one or more engineered strains of

## 11

*Saccharomyces* yeast as defined herein to a subject having *C. difficile* infection or a risk of developing *C. difficile* infection. In preferred aspects, the *Saccharomyces* yeast is *S. cerevisiae* or *S. boulardii*. In certain aspects of the ninth embodiment, the method further comprises administering a therapeutically-effective amount of an antibiotic to the subject.

In certain aspects of the methods, the binding agent is in a pharmaceutical formulation comprising the binding agent and a pharmaceutically acceptable carrier or diluent.

In certain aspects of the methods, the therapeutically-effective amount of the binding agent is between 10 ug/kg and 100 mg/kg of the agent per body weight of the subject.

In certain aspects of the methods, the agent is administered to the subject orally, parenterally or rectally.

In certain aspects of the methods, the engineered strain of *Saccharomyces* yeast is in a pharmaceutical formulation comprising the engineered strain and a pharmaceutically acceptable carrier or diluent. In preferred aspects, the *Saccharomyces* yeast is *S. cerevisiae* or *S. boulardii*.

In certain aspects of the methods, the therapeutically-effective amount of the engineered strain of *Saccharomyces* yeast is between 10 ug/kg and 100 mg/kg of the engineered strain per body weight of the subject. In preferred aspects, the *Saccharomyces* yeast is *S. cerevisiae* or *S. boulardii*.

In certain aspects of the methods, the engineered strain of *Saccharomyces* yeast is administered to the subject orally, nasally or rectally. In preferred aspects, the *Saccharomyces* yeast is *S. cerevisiae* or *S. boulardii*.

The foregoing has outlined rather broadly the features and technical advantages of the present invention in order that the detailed description of the invention that follows may be better understood. Additional features and advantages of the invention will be described herein, which form the subject of the claims of the invention. It should be appreciated by those skilled in the art that any conception and specific embodiment disclosed herein may be readily utilized as a basis for modifying or designing other structures for carrying out the same purposes of the present invention. It should also be realized by those skilled in the art that such equivalent constructions do not depart from the spirit and scope of the invention as set forth in the appended claims. The novel features which are believed to be characteristic of the invention, both as to its organization and method of operation, together with further objects and advantages will be better understood from the following description when considered in connection with the accompanying figures. It is to be expressly understood, however, that any description, figure, example, etc. is provided for the purpose of illustration and description only and is by no means intended to define the limits of the invention.

## BRIEF DESCRIPTION OF DRAWINGS

FIG. 1. Illustration of strategies for making binding agents of the invention.

FIG. 2. A diagram of *C. difficile* toxins TcdA and TcdB, showing the glucosyltransferase domains (GT), cysteine protease domains (CPD), translocation domains (TD) and receptor binding domains (RBD) of each toxin. V<sub>H</sub>Hs that recognize and bind the different toxin domains are shown. Those that are underlined are those that have toxin-neutralizing activity.

FIGS. 3A-3F. Monomeric or dimeric V<sub>H</sub>Hs possess potent neutralizing activity. V<sub>H</sub>Hs block cell rounding induced by TcdA (FIG. 3A) or TcdB (FIG. 3B) at nM concentrations. (FIG. 3C) Diagram of two heterodimers

## 12

against TcdA or TcdB. His<sub>(6)</sub> tag on N-terminus facilitates purification; a flexible spacer (FS) separate the two V<sub>H</sub>Hs. (FIG. 3D) Dimer 5D/E3 increases its neutralizing activity at least 10-fold over a simple mix of the two V<sub>H</sub>Hs. Heterodimers fully protected mice from lethal ip challenge with TcdB (FIG. 3E) or TcdA (FIG. 3F).

FIG. 4. Diagram of ABAB. His-tag and E-tag are epitope tags for purification and detection, respectively. FS: flexible linker; ABP: albumin binding peptide.

FIGS. 5A-5B. ABAB is highly potent in protecting mice from *C. difficile* spore (FIG. 5A) and toxin (FIG. 5B) challenge. MK HuMabs: a mixture of Merck anti-TcdA (actoxumab) and anti-TcdB (bezlotoxumab) human monoclonal antibodies that are undergoing clinical trials.

FIGS. 6A-6B. Anti-toxin sera against both toxins protect mice from CDI. Mice were i.p. injected with 50 ul alpaca anti-sera against TcdA ("Anti-A"), TcdB ("Anti-B"), TcdA+TcdB ("Anti-A+Anti-B") or with 100 ul presera or PBS ("CTR") for 4 hours before *C. difficile* spore (UK1 strain, 10<sup>6</sup> spores/mouse) inoculation. Mouse survival (FIG. 6A; Anti-A+Anti-B vs. PBS, p=0.006) and weight loss (FIG. 6B) are illustrated (\*, p<0.05 between Anti-A+Anti-B vs. control).

FIG. 7. The diagram of the ABAB and ABAB-IgG molecules.

FIGS. 8A-8B. ELISA analysis of binding of ABAB-IgG to TcdA (FIG. 8A) and TcdB (FIG. 8B) as compared with the binding of the individual VHVs to the respective toxins.

FIGS. 9A-9B. Sandwich ELISA analysis of simultaneous binding of the tetraspecific antibody IgG-ABAB to both TcdA and TcdB. FIG. 9A shows serially diluted ABAB-IgG added to ELISA plates coated with TcdA (TxA), followed by TcdB (TxB). FIG. 9B shows serially diluted ABAB-IgG added to ELISA plates coated with TcdB (TxB), followed by TcdA (TxA).

FIGS. 10A-10B. ABAB-IgG neutralizing activities against TcdA (FIG. 10A) and TcdB (FIG. 10B).

FIG. 11. Graph showing in vivo neutralizing activity of ABAB-IgG against *C. difficile* infection in mice versus Merck antibodies against TcdA and TcdB (actoxumab and bezlotoxumab).

FIG. 12. Design of studies on the effects of prophylactic ABAB-IgG against *C. difficile* infection.

FIGS. 13A-13B. Bi-specific sandwich ELISA. (FIG. 13A) A diagram of toxins and antibodies setup in ELISA. (FIG. 13B) O.D. reading of various TcdA concentrations; 125 ng/ml of TcdA was chosen for subsequent ELISA.

FIGS. 14A-14B. Activity of ABAB secreted by Sc-ABAB. (FIG. 14A) Neutralizing effect of secreted ABAB in *S. cerevisiae* culture supernatant. Sc: *S. cerevisiae* (BY4741); Sc-ABAB: *S. cerevisiae* (BY4741)-pD1214-FAKS-ABAB; r-ABAB: recombinant ABAB. ABAB in the supernatant of Sc-ABAB is able to fully protect cells from intoxication. ELISA O.D. readings of supernatants from individual Sc-ABAB clones (FIG. 14B).

FIGS. 15A-15B. ABAB secretion level with various secretion signals. (FIG. 15A) ABAB secretion measured by ELISA and normalized against cell density based on O.D. 600 in *S. cerevisiae*. Statistical significance was determined by Kruskal-Wallis test followed by Dunn's Multiple comparison test. \*p<0.05\*\*p<0.01 (FIG. 15B) ABAB secretion measured by ELISA and normalized against cell density based on O.D. 600 in *S. boulardii*. Statistical significance was determined by Mann Whitney test. \*\*\*p<0.0001.

FIG. 16. A diagram of targeted deletion of chromosomally encoded genes by homologous recombination in *S. boulardii*.

FIGS. 17A-17D. *S. boulardii* URA3Δ/Δ expressing ABAB. (FIG. 17A) Growth comparison in YPD containing vancomycin (1 mg/ml) versus without. (FIG. 17B) ABAB stability in *S. boulardii* culture supernatant after 2 hours of incubation determined by ELISA. (FIG. 17C) Neutralizing activity of ABAB from the culture supernatant of *S. boulardii* URA3Δ/Δ expressing ABAB. (FIG. 17D) ABAB detection in *S. boulardii* URA3Δ/Δ expressing ABAB culture supernatant by western blot. Enriched: ABAB contains c-Myc tag at the end of C-terminus and was further concentrated using α-c-Myc tag antibodies.

FIGS. 18A-18C. Protection of *S. boulardii* expressing ABAB in CDI prevention in mice. (FIG. 18A) Survival rate, (FIG. 18B) Weight loss, (FIG. 18C) Diarrhea incident, throughout the course of infection were recorded and presented. \*significance as determined by Fisher's exact test with two tailed and 95% confidence interval; p value is 0.0108 for FIG. 18A and regular two-way ANOVA (not repeated measures) followed by Dunnett's multiple comparison test was used for FIG. 18B and FIG. 18C, \*P≤0.05. "Sb:EP" is *S. boulardii* with the empty plasmid; "Sb: ABAB" is *S. boulardii* expressing ABAB.

FIGS. 19A-19C. Protection of *S. boulardii* expressing ABAB in treating CDI mice. (FIG. 19A) Survival rate, (FIG. 19B) Weight loss, (FIG. 19C) Diarrhea incident, throughout the course of infection were recorded and presented. \*significance as determined by Fisher's exact test with two tailed and 95% confidence interval; p value is 0.0256 for FIG. 19A; regular two-way ANOVA (not repeated measures) followed by Dunnett's multiple comparison test for FIG. 19B and FIG. 19C. \*P≤0.05\*\*P≤0.01\*\*\*\*P≤0.0001 for FIG. 19B and FIG. 19C. "Sb:EP" is *S. boulardii* with the empty plasmid; "Sb: ABAB" is *S. boulardii* expressing ABAB.

FIGS. 20A-20C. Protection of *S. boulardii* expressing ABAB in CDI recurrent mice. (FIG. 20A) Survival rate, (FIG. 20B) Weight loss, (FIG. 20C) Diarrhea incident, throughout the course of infection were recorded and presented. \*significance as determined by Fisher's exact test with two tailed and 95% confidence interval; p value is 0.017 for FIG. 20A; regular two-way ANOVA (not repeated measures) followed by Dunnett's multiple comparison test for FIG. 20B and FIG. 20C. \*P≤0.05\*\*\*P≤0.001\*\*\*\*P≤0.0001 for FIG. 20B and FIG. 20C. "Sb:EP" is *S. boulardii* with the empty plasmid; "Sb: ABAB" is *S. boulardii* expressing ABAB.

FIG. 21. A diagram of δ site-targeted chromosomal integration using CRISPR. Ty1-H3 (Genbank accession no. M18706) was used to blast against draft genome of MYA796 to obtain δ site sequences. Compiled sequences were used to identify common protospacer adjacent motif (PAM) sites and protospacers. Two PAM site sequences were chosen based on best coverage for multiple sites and common homologous sequences located upstream and downstream of the protospacer and PAM sites for simple integration of ABAB expression cassette. PAM site "I" is provided in SEQ ID NO:93; PAM site "II" is provided in SEQ ID NO:94. Homologous recombination sequences used in primers to generate ABAB expression cassette by PCR are underlined.

FIGS. 22A-22B. ABAB secretion of *S. boulardii* using CRISPR-based targeting δ site chromosomal integration. (FIG. 22A) ABAB secretion measured by ELISA. ITG: ABAB integration cassette. Low: CRISPR plasmid to ITG ratio at 2; High: CRISPR plasmid to ITG ratio at 0.25. (FIG. 22B) ABAB secretion amount comparison. M-/−*Cir*<sup>0</sup>:pKC, M-/−*Cir*<sup>+</sup>:ABAB, M-/−*Cir*<sup>0</sup>:ABAB are plasmid based.

Ch<sup>Ins</sup>: single site target chromosomal integration of ABAB cassette through conventional homologous recombination. CRISPR1-2: ABAB cassette integration at site I. C<sup>CRISPR3-4</sup>: ABAB cassette integration at site II.

FIGS. 23A-23C. Protection of *S. boulardii* expressing ABAB in treating CDI mice. (FIG. 23A) Survival rate, (FIG. 23B) Weight loss, (FIG. 23C) Diarrhea incident, throughout the course of infection were recorded and presented. \*significance as determined by Fisher's exact test with two tailed and 95% confidence interval; p value is 0.0325 for FIG. 23A); regular two-way ANOVA (not repeated measures) followed by Dunnnett's multiple comparison test for FIG. 23B and FIG. 23C. \*P≤0.05\*\*P≤0.01 for FIG. 23B and FIG. 23C.

## DETAILED DESCRIPTION OF THE INVENTION

### I. Definitions

Unless otherwise noted, technical terms are used according to conventional usage. Definitions of common terms in molecular biology may be found, for example, in Benjamin Lewin, Genes VII, published by Oxford University Press, 2000 (ISBN 019879276X); Kendrew et al. (eds.); The Encyclopedia of Molecular Biology, published by Blackwell Publishers, 1994 (ISBN 0632021829); and Robert A. Meyers (ed.), Molecular Biology and Biotechnology: a Comprehensive Desk Reference, published by Wiley, John & Sons, Inc., 1995 (ISBN 0471186341); and other similar technical references.

As used herein, "a" or "an" may mean one or more. As used herein when used in conjunction with the word "comprising," the words "a" or "an" may mean one or more than one. As used herein "another" may mean at least a second or more. Furthermore, unless otherwise required by context, singular terms include pluralities and plural terms include the singular.

As used herein, "about" refers to a numeric value, including, for example, whole numbers, fractions, and percentages, whether or not explicitly indicated. The term "about" generally refers to a range of numerical values (e.g., +/−5-10% of the recited value) that one of ordinary skill in the art would consider equivalent to the recited value (e.g., having the same function or result). In some instances, the term "about" may include numerical values that are rounded to the nearest significant figure.

### II. The Present Invention

*C. difficile*-associated disease (CDI) is mainly caused by two large exotoxins, namely toxin A (TcdA) and toxin B (TcdB), produced by the bacteria. These toxins are structurally similar, large, single-chain proteins (TcdA is about 300 kD; TcdB is about 270 kD) that exhibit similar modes of action on host cells. Both toxins target host Rho GTPases, leading to enzyme inactivation, followed by cytoskeleton disorganization and apoptosis. In intestinal epithelial cells, TcdA catalyzes glucosylation of the Rho GTPases, leading to reorganization of the actin cytoskeleton with accompanying morphological changes such as complete rounding of cells and destruction of the intestinal barrier function. The toxins can individually cause CDI in animals, and TcdA<sup>−</sup>TcdB<sup>−</sup> strains of the bacteria are avirulent.

Numerous independent studies have demonstrated that neutralizing antibodies against the toxins confer protection against CDI [24-33]. Because TcdA and TcdB are essential

## 15

virulence factors for *C. difficile*, neutralizing antibodies produced against both toxins protect against toxigenic *C. difficile* infection in animal models [30-33]. In humans, high serum levels of antitoxin antibodies are associated with reduced disease severity and incidence of relapse [9,25,29].

Therefore, a preventative rationale for systemically and orally administered antitoxin antibodies exists. However, monoclonal antibodies targeting a single epitope are typically low affinity, and use of such antibodies runs the risk of inducing mutations within the epitopes of the toxins thereby creating additional strains. Thus, neutralizing antitoxins targeting multiple, key, and conserved toxin epitopes are highly desirable.

The present invention builds on existing knowledge regarding anti-TcdA and anti-TcdB antibodies for the treatment and prevention of CDI, and the symptoms of CDI. Provided herein are antibody-based, fusion protein binding agents derived from human and camelid immunoglobulins, optionally expressed by the probiotic yeast *Saccharomyces* strain in a subject. These binding agents recognize and bind with specificity to *C. difficile* TcdA and/or TcdB. Some of these binding agents exhibit toxin-neutralizing activity. These yeast-based immunotherapeutic can be used to treat or prevent primary and recurrent CDI, as well as the symptoms of primary and recurrent CDI. In preferred aspects, the *Saccharomyces* yeast is *S. cerevisiae* or *S. boulardii*.

As discussed in detail below, camelid animals (dromedary camels, Bactrian camels, wild Bactrian camels, llamas, alpacas, vicunas, and guanacos) produce a class of functional immunoglobulins that lack light chains and are thus heavy chain-only antibodies (HCabs) [34] with binding properties equivalent to those achieved by conventional IgG [35]. The V<sub>H</sub> domain of HCabs, called V<sub>H</sub>H, is similar to the conventional human V<sub>H</sub> domain but has unique sequence and structural characteristics [36]. DNA encoding this domain can be readily cloned and expressed in microbes to yield soluble protein monomers that retain the antigen-binding properties of the parent HCab. These V<sub>H</sub>H peptide monomer binding agents are small (~15 kDa), easy to produce, and generally more stable than conventional antibody body fragments [37-39]. V<sub>H</sub>Hs have been explored to treat intestinal diseases since they are relatively resistant to proteases and can be further engineered to enhance such properties [40]. They can also be produced as fusion proteins with human antibodies, such as IgG, and fragments of human antibodies, such as Fc domains.

The present invention utilizes the advantageous characteristics of HCabs in the production of binding agents that can be used in the treatment and prevention of CDI. As disclosed herein, V<sub>H</sub>H peptide monomers were screened for TcdA and TcdB epitope recognition and binding. Those monomers that exhibited epitope binding and had toxin-neutralizing activity were linked to produce the binding agents of the invention. The binding agents include simple V<sub>H</sub>H peptide monomers and linked groups of V<sub>H</sub>H peptide monomers (comprising 2, 3, 4, or more monomers), as well as more complex binding agents that comprise V<sub>H</sub>H peptide monomers joined to antibody Fc domains, as well as V<sub>H</sub>H peptide monomers joined to IgG antibodies (see FIG. 1).

Further, *Saccharomyces boulardii*, a Generally Regarded as Safe (GRAS) organism by the FDA, is commonly available over-the-counter for use in promoting intestinal health and amelioration of gastrointestinal illness due to diarrheal diseases. This yeast strain has been studied in multiple randomized double-blinded placebo-controlled clinical trials for both safety and efficacy against intestinal diseases including CDI [42-46]. *S. boulardii* treatment significantly

## 16

reduced CDI recurrence [44-46], and those recurrent patients had significantly less *S. boulardii* in stools than non-recurring patients [43]. The immune modulatory effects of *S. boulardii* that provide protection against *C. difficile* toxin-induced inflammation have been described [47-49]. In addition, *S. boulardii* may help in maintaining normal microbiota [50]; a recent clinical trial (NCT01473368) found that *S. boulardii* treatment can prevent some antibiotic-induced microbiome changes and, in parallel, can reduce antibiotic-associated diarrhea.

*S. cerevisiae* (commonly known as “brewer’s yeast”), which is genetically related to *S. boulardii*, has been used successfully to express V<sub>H</sub>Hs with high yield [51]. *S. boulardii* is physiologically distinct from *S. cerevisiae*, although genome analysis has revealed that both genomes are remarkably similar at the nucleotide level [52,53]. Therefore, molecular genetic tools previously developed for use in *S. cerevisiae* are now being used with *S. boulardii* [54-56], making this probiotic a candidate for engineering as a therapeutic agent against CDI.

There are several additional metabolic characteristics which make *S. boulardii* ideal for use as an oral therapeutic agent. In contrast to *S. cerevisiae*, *S. boulardii* grows well at 37° C. and it is more resistant to acidic environmental conditions [57], making this strain particularly well suited for better surviving and persisting in the human intestinal tract after oral administration. In addition, an experimental murine oral colonization model with *Saccharomyces* is well characterized [58]; using this model, protection has been reported against oral challenge with enteric pathogens such as *Salmonella Typhimurium* [58,59] and *Enteritidis* [60] in conventional mice orally treated with *S. boulardii*, as well as protection against CDI challenge in pretreated gnotobiotic animals [58,61]. The probiotic *S. boulardii*, genetically engineered to secrete V<sub>H</sub>H binding agents capable of neutralizing both TcdA and TcdB of *C. difficile*, could significantly improve the therapeutic capacity of this probiotic to disrupt both ongoing and recurrent CDI.

In view of the exceptional characteristics of *S. boulardii*, strains of *S. boulardii* expressing the binding agents defined herein were produced and tested. As described in the Examples, these yeast-based immunotherapeutics can be used to treat or prevent primary and recurrent CDI, as well as the symptoms of primary and recurrent CDI.

V<sub>H</sub>H Monomers & V<sub>H</sub>H Heterodimers

As initially reported in WO 16/127104, the inventors established an efficient platform to screen V<sub>H</sub>H monomers against specific domains of both *C. difficile* toxins. Using highly immunogenic atoxic holotoxins for immunization, and bioactive chimeric toxins (with normal domain functions) for screening, panels of V<sub>H</sub>H monomers binding to different domains of TcdA or TcdB were prepared. A majority of these V<sub>H</sub>H monomers possessed potent neutralizing activity and their binding to specific domains of TcdA and TcdB was determined (FIG. 2).

Several of the V<sub>H</sub>H monomers bind to highly conserved TcdA/TcdB epitopes. For example, the E3 V<sub>H</sub>H monomer binds to the Rho GTPase binding site and blocks glycosylation; the AH3 V<sub>H</sub>H monomer binds to the GT domain of the toxin; the 7F V<sub>H</sub>H monomer binds to cysteine protease cleavage sites and blocks GT domain cleavage and release. Some V<sub>H</sub>H monomers have potent toxin neutralizing activity, capable of blocking toxin cytotoxic activity at nM concentrations (monomers underlined in FIG. 2; see also FIGS. 3A and 3B). Table 1 references amino and nucleic acid sequences in the Sequence Listing for some of these V<sub>H</sub>H peptide monomers, both wild-type and codon-opti-

mized versions. While both the optimized and non-optimized versions can be used in the production of the various binding agents of the present invention, the codon-optimized versions are preferred for expression in mammalian cells.

The present invention includes each of the  $V_{H}H$  peptide monomers referenced in Table 1 as well as sequence variants thereof having at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity over the entire length of the peptide sequence and retaining the toxin binding and/or neutralizing activity of the wild-type peptide. The present invention also includes polynucleotide sequences encoding each of the  $V_{H}H$  peptide monomers of Table 1 and the sequence variants thereof, as well as complementary strands thereof.

TABLE 1

Name	Codon Optimized?	Location of epitope	SEQ ID NO for Amino Acid Seq.	SEQ ID NO for Nucleic Acid Seq.
SD	Yes	TcdB glucosyltransferase domain	1	2
E3	Yes	TcdB glucosyltransferase domain	3	4
AA6	Yes	TcdA cysteine protease domain	5	6
AH3	Yes	TcdA glucosyltransferase domain	7	8
SD	No	TcdB glucosyltransferase domain	48	49
E3	No	TcdB glucosyltransferase domain	50	51
AA6	No	TcdA cysteine protease domain	52	53
AH3	No	TcdA glucosyltransferase domain	54	55

To enhance the binding activity of the peptide monomers,  $V_{H}H$  peptide homo- and hetero-dimer binding agents were created, where two  $V_{H}H$  peptide monomers are linked (FIG. 3C). Homodimer binding agents comprise two identical monomers that bind identical epitopes on two different toxins. Heterodimer binding agents comprise two different monomers that bind two distinct epitopes of the same toxin or distinct epitopes on two different toxins. The  $V_{H}H$  heterodimers were found to possess substantially enhanced neutralizing activities compared with equimolar mixtures of the individual  $V_{H}H$  peptide monomers comprising the heterodimers (FIG. 3D). Indeed, heterodimers 5D/E3 and AH3/AA6 were found to fully protect mice from lethal systemic TcdB or TcdA challenge respectively, whereas mixed 5D and E3, or AA6 alone were only partially protective (FIGS. 3E and 3F).

The  $V_{H}H$  monomers in the homo- and hetero-dimers are linked using a short, flexible linker of between 10 and 20 amino acids. Suitable linkers include those provided in Table 2. Table 2 also includes codon-optimized versions of the three linkers. While both the optimized and non-optimized versions can be used in the production of the various binding agents of the present invention, the codon-optimized versions are preferred for expression in mammalian cells.

TABLE 2

Name	Codon Optimized?	SEQ ID NO for Amino Acid Seq.	SEQ ID NO for Nucleic Acid Seq.
Linker-1	Yes	9	10
Linker-2	Yes	11	12
Linker-3	Yes	13	14
Linker-1	No	56	57
Linker-2	No	58	59
Linker-3	No	60	61

It will be understood by the skilled artisan that minor changes can be made to the sequence of the flexible linker without departing from the properties of the peptide. Sequence variants of the flexible linker having at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity over the entire length of the peptide sequence and retaining properties of the linker upon which they are based may thus be used.

The present invention includes  $V_{H}H$  peptide homodimer binding agents comprising pairs of any of the monomers listed in Table 1, linked by a flexible linker as defined above. The present invention also includes  $V_{H}H$  peptide heterodimer binding agents comprising any combination of two of the monomers listed in Table 1, linked by a flexible linker as defined above. Exemplary heterodimers are provided in Table 3.

TABLE 3

Name	SEQ ID NO for Amino Acid Seq.	SEQ ID NO for Nucleic Acid Seq.
AH3-5D	15	16
AA6-E3	17	18
5D-E3	62	63
AH3-AA6	64	65

The present invention also includes sequence variants of the  $V_{H}H$  peptide homo- and hetero-dimers having at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity over the entire length of the protein sequence and retaining the toxin binding and/or neutralizing activity of the wild-type protein. The present invention further includes polynucleotide sequences encoding each the  $V_{H}H$  peptide homo-heterodimers and the sequence variants thereof, as well as complementary strands thereof.

The invention also includes  $V_{H}H$  peptide homo- and hetero-trimer binding agents where three monomers are linked using the flexible linkers defined above in Table 2. Any combination of the monomers of Table 1 may be used, including trimers comprising three copies of the same monomer, trimers comprising two copies of one monomer and a single copy of another, and trimers comprising three different monomers. Sequence variants of the  $V_{H}H$  peptide homo- and hetero-trimers are included in the invention, having at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity over the entire length of the protein sequence and retaining the toxin binding and/or neutralizing activity of the wild-type protein. The present invention further includes polynucleotide sequences encoding each the  $V_{H}H$  peptide homo- and hetero-trimers and the sequence variants thereof, as well as complementary strands thereof.

ABAB

The success of the peptide monomers and heterodimers allowed the inventors to develop binding agents comprising four linked  $V_{H}H$  peptide monomers. This was a goal of the research as earlier work had shown that the most useful agents in the treatment and prevention of CDI would be single antibodies that can simultaneously neutralize both TcdA and TcdB as this would be necessary in order to convey full protection against most pathogenic *C. difficile* strains. By creating tetra-specific binding agents that recognize and bind two epitopes on each of the toxins, the binding

19

and neutralizing activity of the proteins might be strengthened. Therefore, four domain (tetra-specific)  $V_H$ H binding agents were generated.

The tetra-specific, tetrameric binding agents can be prepared from any combination of the monomers of Table 1, where the monomers are linked using the flexible linkers of Table 2. These binding agents include those having four copies of the same monomer, those having three copies of the same monomer, those having two copies of the same monomer, those having four unique monomers, and variations therein. Sequence variants of the tetramers are included in the invention, having at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity over the entire length of the protein sequence and retaining the toxin binding and/or neutralizing activity of the wild-type protein. The present invention further includes polynucleotide sequences encoding each tetramer and the sequence variants thereof, as well as complementary strands thereof.

ABBA is a particular binding agent of the invention that comprises four linked  $V_H$ H monomers, AH3-E3-E3-AA6. ABBA thus has two identical monomers (E3) and two additional different monomers (AH3 and AA6) (See Table 1).

ABAB is another particular binding agent of the invention that comprises four linked  $V_H$ H monomers, each of which has binding specificity for a different epitope of TcdA or TcdB. ABAB is thus a tetra-specific, tetrameric binding agent that consists of four distinct neutralizing  $V_H$ H monomers, two against TcdA and two against TcdB. This structural feature allows ABAB to bind simultaneously to two distinct neutralizing epitopes on each toxin. As described below, affinity/avidity and neutralizing activity of ABAB is more than 3-logs higher than human monoclonal antibodies (HuMabs) currently undergoing clinical trials for treatment of CDI.

ABAB binding agent was prepared by linking  $V_H$ H monomers AH3, 5D, AA6, and E3 (Table 1) using flexible linkers (Table 2). This binding agent targets conserved, non-overlapping epitopes and has excellent toxin neutralizing activity. In the design of ABAB (FIG. 4),  $V_H$ H peptide monomers AH3 and AA6 were separated by placing the 5D between them because AH3 and AA6 bind to GT and TD

20

respectively (FIG. 2), which are spatially distant to each other. This design allowed AH3 and AA6 to bind to TcdA simultaneously.

The complete amino acid sequence comprising ABAB is provided in SEQ ID NO:19; the nucleic acid sequence encoding the protein is provided in SEQ ID NO:20. The present invention thus includes the ABAB binding agent provided in SEQ ID NO: 19, as well as sequence variants of the ABAB binding agent having at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity over the entire length of the protein sequence and retaining the toxin binding and/or neutralizing activity of the wild-type protein. The sequence variants include variants wherein the variant is humanized and/or wherein the amino acids are optimized for production and secretion by yeast.

The present invention further includes polynucleotide sequences encoding the ABAB binding agent (e.g., SEQ ID NO:20) and the sequence variants thereof, as well as complementary strands thereof.

Modified versions of the ABAB binding agent encompassed by the invention includes those having one or more of (i) a His<sub>(6)</sub>-tag (HHHHHH; SEQ ID NO:66) at the amino terminus of the protein to aid in purification, (ii) an E-tag (GAPVYPDPLEPR; SEQ ID NO:67) at the carboxy terminus of the protein to aid in detection; (iii) an albumin-binding peptide (ABP) (DICLPRWGCLWD; SEQ ID NO:21) at the carboxyl end of the construct to increase serum half-life of the protein as  $V_H$ H monomers have a half-life of 2-3 hr and inclusion of ABP can increase the serum half-life to 10 hr (see FIG. 4); and a D7 tag (SSAPTKAKRRVVQREKT; SEQ ID NO:112) at the carboxy terminus of the protein. The invention includes versions of the ABAB binding agent having one, two, three or four of these tags and peptides. An exemplary modified ABAB binding agent that includes the His tag and the D7 tag comprises the amino acid sequence set forth in SEQ ID NO:113 (the coding sequence is set forth in SEQ ID NO:114).

When yeast strains are engineered to produce ABAB, the protein can be also modified to include a secretion signal at the amino terminus of the protein. The secretion signal may be, but is not limited to, one of the sequences shown in Table 4.

TABLE 4

Secretion sequences for protein secretion in yeast		
Secretion signal	Amino acid sequence	Abbr.
$\alpha$ -factor_full ( <i>S. cerevisiae</i> )	MRFPSIFTAVLFAASSALAAPVNTTTEDETAQIPAEAVIGYSD LEGDFDVALPLPSNSTNNGLLFINTTTIASIAKEEGVSLEKRE AEA (SEQ ID NO: 96)	FAKS
$\alpha$ -factor_T_kex_stc ( <i>S. cerevisiae</i> )	MRFPSIFTAVLFAASSALAAPVNTTTEDELEGDFDVALPFSA SIAAKEEGVSLEKREA (SEQ ID NO: 97)	AKS
$\alpha$ -factor_T_kex ( <i>S. cerevisiae</i> )	MRFPSIFTAVLFAASSALAAPVNTTTEDELEGDFDVALP FSASIAKEEGVSLEKR (SEQ ID NO: 98)	AK
$\alpha$ -factor_T ( <i>S. cerevisiae</i> )	MRFPSIFTA VLFAASSALA (SEQ ID NO: 99)	AT
Alpha-amylase ( <i>Aspergillus niger</i> )	MVAWSLFLYGLQVAAPALA (SEQ ID NO: 100)	A.A.
Glucoamylase ( <i>Aspergillus awamori</i> )	MSFRSLLALSGLVCSGLA (SEQ ID NO: 101)	GA
Inulinase ( <i>Kluyveromyces maxianus</i> )	MKLAYSLLLPLAGVSA (SEQ ID NO: 102)	IN

TABLE 4-continued

Secretion sequences for protein secretion in yeast		
Secretion signal	Amino acid sequence	Abbr.
Invertase ( <i>S. cerevisiae</i> )	MLLQAFLFLLAGFAAKISA (SEQ ID NO: 103)	IVS
Killer protein ( <i>S. cerevisiae</i> )	MTKPTQVLVRSVSILFFITLLHLVVA (SEQ ID NO: 104)	KP
Lysozyme ( <i>Gallus gallus</i> )	MLGKNDPMCLVLVLLGLTALLGICQG (SEQ ID NO: 105)	LZ
Serum albumin ( <i>Homo sapiens</i> )	MKWVTFISLLFLFSSAYS (SEQ ID NO: 106)	SA

Exemplary modified ABAB binding agents that include an amino-terminal secretion signal include AT-ABAB and IVS-ABAB.

An exemplary modified ABAB binding agent that is expressed from a plasmid in yeast or bacteria includes the ABAB binding agent set forth in SEQ ID NO: 107, which is encoded by the polynucleotide sequence set forth in SEQ ID NO: 108.

An exemplary modified ABAB binding agent that is expressed in yeast after chromosomal integration includes the ABAB binding agent set forth in SEQ ID NO:109, which is encoded by the polynucleotide sequence set forth in SEQ ID NO: 110.

Each of the binding agents of the invention binds to TcdA and/or TcdB with specificity. In certain aspects of the invention, the binding agents exhibit TcdA and/or TcdB neutralizing activity.

For the sake of clarity it can be noted that as used herein, “mono-specific”, “bi-specific”, “tri-specific”, “tetra-specific”, etc., mean the particular binding agent binds to 1, 2, 3, 4, etc., different epitopes, respectively. As used herein, “monomeric”, “dimeric”, “trimeric”, “tetrameric”, etc., mean that the particular binding agent has 1, 2, 3, 4, etc., separate V<sub>H</sub>H peptide monomers that bind to the epitopes, respectively. Thus, a mono-specific, dimeric binding agent would display two V<sub>H</sub>H peptide monomers that bind to the same epitope (e.g., a homodimer), and a bi-specific, dimeric binding agent would have two V<sub>H</sub>H peptide monomers that bind to two different epitopes (e.g., a heterodimer). A tetra-specific, octameric binding agent has eight V<sub>H</sub>H peptide monomers that recognize four different epitopes.

#### V<sub>H</sub>-Fc

It is well known that chimeric Fc-fusion proteins have the potential of increasing the half-life of a protein in vivo. This strategy has been applied in several FDA approved drugs, such as Etanercept. A proof-of-principle study has shown that single-chain antibodies can be correctly assembled and expressed by B cells of transgenic mice carrying a mini-Ig construct encoding a dromedary V<sub>H</sub>H and the Fc domain of human IgG. Also EG2-Fc, a chimeric anti-EGFR/EGFRvIII V<sub>H</sub>H, exhibited excellent tumor accumulation in vivo and has pharmacokinetic properties that could improve glioblastoma targeting.

The present invention includes binding agents comprising V<sub>H</sub>H peptide monomers joined to antibody Fc domains (V<sub>H</sub>H-Fc), where the binding agents bind TcdA and/or TcdB. In these Fc domain-based binding agents, one, two, three, four or more of the V<sub>H</sub>H peptide monomers are joined to the

hinge, C<sub>H</sub>2 and C<sub>H</sub>3 regions of the Fc domain of an antibody heavy chain. Thus, the peptide monomers replace the Fab regions of the antibody.

The V<sub>H</sub>H peptide monomers may be any of those provided in Table 1 above and include 5D (SEQ ID NO:1), E3 (SEQ ID NO:3), AA6 (SEQ ID NO:5) and AH3 (SEQ ID NO:7) V<sub>H</sub>H peptide monomers. Where two or more monomers are linked, the monomers may be linked by flexible peptide linkers, generally comprising between 10 and 20 amino acids. Suitable linkers include those linkers provided in Table 2, such as linker-1 (SEQ ID NO:9), linker-2 (SEQ ID NO:11), and linker-3 (SEQ ID NO:13).

While the V<sub>H</sub>H-Fc will typically be composed of two identical chains that self-assemble intracellularly after production, the invention also includes V<sub>H</sub>H-Fc binding agents comprising two different Fc chains. In such circumstances, the sequence of the V<sub>H</sub>H monomer(s) alone may differ between the two Fc chains, or the Fc chains themselves may differ in sequence, or both the V<sub>H</sub>H monomer(s) and the Fc chains may differ in sequence.

One type of V<sub>H</sub>H-Fc binding agent is an octameric binding agent comprising an antibody Fc domain and first, 40 second, third and fourth V<sub>H</sub>H peptide monomers, where the V<sub>H</sub>H peptide monomers have binding specificity for an epitope of TcdA or toxin B TcdB, where the first, second, third and fourth V<sub>H</sub>H peptide monomers are linked together and joined to amino termini of both antibody Fc domains, 45 and where the antibody Fc domain comprises the hinge, C<sub>H</sub>2 and C<sub>H</sub>3 regions of an antibody heavy chain. Because this binding agent has four V<sub>H</sub>H peptide monomers, it can be mono-specific (where all of the monomers bind the same epitope), bi-specific (where the monomers bind two different epitopes), tri-specific (where the monomers bind three different epitopes), or tetra-specific (where the monomers bind four different epitopes).

A specific example of a tetra-specific V<sub>H</sub>H-Fc binding agent is the ABAB-Fc binding agent, a tetra-specific, octameric binding agent comprising an antibody Fc domain and two sets of linked first, second, third and fourth V<sub>H</sub>H peptide monomers, wherein the antibody Fc domain comprises two arms, each arm comprising hinge, C<sub>H</sub>2 and C<sub>H</sub>3 regions of an antibody heavy chain, and each arm having an amino terminus, wherein for each arm of the Fc domain, one set of linked first, second, third and fourth V<sub>H</sub>H peptide monomers is joined to the amino terminus of the arm, and where the V<sub>H</sub>H peptide monomers have binding specificity for an epitope of TcdA or TcdB (see FIG. 1). This binding agent is termed “tetra-specific” as it recognizes four different toxin epitopes. It is termed “octameric” as it bears eight V<sub>H</sub>H peptide monomers (two copies of the first monomer, two

copies of the second monomer, two copies of the third monomer, and two copies of the fourth monomer). ABAB-Fc was found to exhibit specific binding and neutralizing activity.

The ABAB-Fc binding agent was prepared by generating an expression vector encoding the  $V_{H}H$  peptide monomers AH3/5D/AA6/E3 (linked in the noted order) joined to a human IgG1 Fc domain. The  $V_{H}H$  peptide monomers were separated by flexible linkers of Table 2. The nucleic acid sequence encoding each chain is provided in SEQ ID NO:23. The amino acid sequence of each chain is provided in SEQ ID NO:22. Upon self-assembly of pairs of the chains after expression, the tetra-specific, octameric binding agent resulted. The invention includes the ABAB-Fc binding agent of SEQ ID NO:22, modified versions of ABAB binding agents as defined above, and sequence variants thereof having at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity over the entire length of the protein sequence and retaining the toxin binding and/or neutralizing activity of the wild-type protein. The present invention further includes polynucleotide sequences encoding these sequence variants and complementary strands thereof.

Mono-specific  $V_{H}H$ -Fc binding agents (5D-Fc, 5D-Fc, E3-Fc, AA6-Fc) and bi-specific  $V_{H}H$ -Fc binding agents (e.g., AH3/5D-Fc and AA6/E3-Fc) were also made using this Fc-fusion system. With respect to mono-specific binding agents, single  $V_{H}H$  peptide monomers were joined to human IgG1 Fc domains. Upon expression and assembly, pairs of the chains resulted in mono-specific, dimeric binding agents (when the chains were identical) or bi-specific, dimeric binding agents (when the chains were different). With respect to bi-specific binding agents, two linked  $V_{H}H$  peptide monomers ( $V_{H}H$  homo- or hetero-dimers) were joined to human IgG1 Fc domains. Upon expression and assembly, pairs of the chains resulted in bi-specific, tetrameric binding agents (when the chains were identical) or tetra-specific, tetrameric binding agents (when the chains were different). Table 5 provides the sequences for some of these binding agents.

TABLE 5

Name	SEQ ID NO for Amino Acid Seq.	SEQ ID NO for Nucleic Acid Seq.
5D-Fc	24	25
E3-Fc	26	27
AA6-Fc	28	29
AH3-Fc	30	31
AH3-5D-Fc	32	33
AA6-E3-Fc	34	35

Specific pairings with one monomer include: 5D-Fc+5D-Fc; E3-Fc+E3-Fc; AA6-Fc+AA6-Fc; AH3-Fc+AH3-Fc; 5D-Fc+E3-Fc; 5D-Fc+AA6-Fc; 5D-Fc+AH3-Fc; E3-Fc+AA6-Fc; E3-Fc+AH3-Fc; and AA6-Fc+AH3-Fc. Specific pairings with two monomers include: AH3-5D-Fc+AH3-5D-Fc; AA6-E3-Fc+AA6-E3-Fc; and AH3-5D-Fc+AA6-E3-Fc.

Bi-specific, tetrameric  $V_{H}H$ -Fc binding agents were produced comprising an antibody Fc domain and two sets of linked first and second  $V_{H}H$  peptide monomers, wherein the antibody Fc domain comprises two arms, each arm comprising hinge,  $C_{H}2$  and  $C_{H}3$  regions of an antibody heavy chain, and each arm having an amino terminus, wherein for each arm of the Fc domain, one set of linked first and second

$V_{H}H$  peptide monomers is joined to the amino terminus of the arm, and where the  $V_{H}H$  peptide monomers have binding specificity for an epitope of TcdA or TcdB. This binding agent is termed “bi-specific” as it recognizes two different toxin epitopes. It is termed “tetrameric” as it bears four  $V_{H}H$  peptide monomers (two copies of the first monomer, and two copies of the second monomer). The first and second  $V_{H}H$  peptide monomers may have binding specificity for the same or different epitopes. The  $V_{H}H$  peptide monomers may independently have binding specificity for an epitope in the glucosyltransferase domain, cysteine protease domain, translocation domain or receptor binding domain of TcdA or TcdB.

A specific example of a bi-specific, tetrameric  $V_{H}H$ -Fc binding agent comprises the amino acid sequence set forth in SEQ ID NO:32 (AH3/5D-Fc). The invention also includes sequence variants thereof having at least 95% sequence identity, where the sequence variant retains toxin-neutralizing activity. The variant amino acids of the sequence variant may be located in framework regions of the  $V_{H}H$  peptide monomers.

A specific example of a bi-specific, tetrameric  $V_{H}H$ -Fc binding agent comprises the amino acid sequence set forth in SEQ ID NO:34 (AA6/E3-Fc). The invention also includes sequence variants thereof having at least 95% sequence identity, where the sequence variant retains toxin-neutralizing activity. The variant amino acids of the sequence variant may be located in framework regions of the  $V_{H}H$  peptide monomers.

The  $V_{H}H$ -Fc binding agents bind to TcdA and/or TcdB with specificity. In certain aspects of the invention, the binding agents exhibit TcdA and/or TcdB neutralizing activity.

V<sub>H</sub>H-IgG

The present invention also includes binding agents comprising  $V_{H}H$  peptide monomers joined to more of an antibody than the Fc domain alone.  $V_{H}H$ -IgG binding agents comprise one, two, three, four or more of the  $V_{H}H$  peptide monomers are joined to the light (kappa or lambda) and heavy chains of an IgG antibody lacking the variable regions of the antibody. Thus, the peptide monomers replace the variable regions of the antibody.

The  $V_{H}H$  peptide monomers may be any of those provided in Table 1 above and include 5D (SEQ ID NO:1), E3 (SEQ ID NO:3), AA6 (SEQ ID NO:5) and AH3 (SEQ ID NO:7)  $V_{H}H$  peptide monomers. Where two or more monomers are linked, the monomers may be linked by flexible peptide linkers, generally comprising between 10 and 20 amino acids. Suitable linkers include those linkers provided in Table 2, such as linker-1 (SEQ ID NO:9), linker-2 (SEQ ID NO:11), and linker-3 (SEQ ID NO:13).

$V_{H}H$ -IgG binding agents include octameric binding agents comprising an IgG antibody and first, second, third and fourth  $V_{H}H$  peptide monomers, wherein the  $V_{H}H$  peptide monomers have binding specificity for an epitope of TcdA or TcdB, wherein first and second  $V_{H}H$  peptide monomers are linked together and joined to amino termini of both light chains of the antibody, wherein the light chains lack the antibody variable regions, and wherein third and fourth  $V_{H}H$  peptide monomers are linked together and joined to amino termini of both heavy chains of the antibody, wherein the heavy chains lack the antibody variable regions. Because this binding agent has four  $V_{H}H$  peptide monomers, it can be mono-specific (where all of the monomers bind the same epitope), bi-specific (where the monomers bind two different epitopes), tri-specific (where the monomers bind

three different epitopes), or tetra-specific (where the monomers bind four different epitopes).

A specific example of a tetra-specific  $V_{H}H$ -IgG binding agent is the ABAB-IgG binding agent, a tetra-specific, octameric binding agent comprising an IgG antibody, two sets of linked first and second  $V_{H}H$  peptide monomers, and two sets of linked third and fourth  $V_{H}H$  peptide monomers, wherein the IgG antibody comprises two arms, each arm comprising a heavy chain lacking a variable region and a light chain lacking a variable region, and each chain having an amino terminus, wherein for each arm of the antibody, one set of linked first and second  $V_{H}H$  peptide monomers is joined to the amino terminus of the light chain, and one set of linked third and fourth  $V_{H}H$  peptide monomers is joined to the amino terminus of the heavy chain, and wherein the  $V_{H}H$  peptide monomers have binding specificity for an epitope of TcdA or TcdB (see FIG. 1). This binding agent is termed “tetra-specific” as it recognizes four different toxin epitopes. It is termed “octameric” as it bears eight  $V_{H}H$  peptide monomers (two copies of the first monomer, two copies of the second monomer, two copies of the third monomer, and two copies of the fourth monomer). In certain aspects, the first, second, third and fourth  $V_{H}H$  peptide monomers may each have binding specificity for a different epitope. In certain aspects, two of the  $V_{H}H$  peptide monomers may have binding specificity for epitopes of TcdA and two of the  $V_{H}H$  peptide monomers may have binding specificity for epitopes of TcdB. In certain aspects, the  $V_{H}H$  peptide monomers independently have binding specificity for an epitope in the glucosyltransferase domain, cysteine protease domain, translocation domain or receptor binding domain of TcdA or TcdB.

A specific example of a tetra-specific, octameric ABAB-IgG binding agent comprises a light (kappa) chain having the amino acid sequence set forth in SEQ ID NO:46 (AA6/E3 kappa) or a sequence variant having at least 95% sequence identity thereto, and a heavy chain having the amino acid sequence set forth in SEQ ID NO:44 (AH3/5D heavy) or a sequence variant having at least 95% sequence identity thereto. In this aspect, the sequence variants retain toxin-neutralizing activity. The variant amino acids of the sequence variant may be located in framework regions of the  $V_{H}H$  peptide monomers. This binding agent was produced by preparing two separate expression vectors, the first encoding the  $V_{H}H$  peptide monomers AH3/5D (linked in the noted order) joined to the human IgG1 antibody heavy chain lacking the variable region and the second encoding the  $V_{H}H$  peptide monomers AA6/E3 (linked in the noted order) joined to the human IgG1 antibody light (kappa) chain lacking the variable region. The nucleotide sequence encoding the AA6/E3-IgG1 light (kappa) chain is provided in SEQ ID NO:47. The nucleotide sequence encoding the AH3/5D-IgG1 heavy chain is provided in SEQ ID NO:45. The invention includes sequence variants of ABAB-IgG having at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity over the entire length of the protein sequence and retaining the toxin binding and/or neutralizing activity of the wild-type protein. The present invention further includes polynucleotide sequences encoding these sequence variants and complementary strands thereof.

Bi-specific or tetra-specific, tetrameric IgG binding agents are included in the invention. Such binding agents comprise an IgG antibody and first, second, third and fourth  $V_{H}H$  peptide monomers, wherein the IgG antibody comprises two arms, each arm comprising a heavy chain lacking a variable region and a light chain lacking a variable region, and each

chain having an amino terminus, wherein for a first arm of the antibody, the first  $V_{H}H$  peptide monomer is joined to the amino terminus of the light chain, and the second  $V_{H}H$  peptide monomer is joined to the amino terminus of the heavy chain, wherein for a second arm of the antibody, the third  $V_{H}H$  peptide monomer is joined to the amino terminus of the light chain, and the fourth  $V_{H}H$  peptide monomer is joined to the amino terminus of the heavy chain, and where the  $V_{H}H$  peptide monomers have binding specificity for an epitope of TcdA or TcdB. When the binding agent is “tetra-specific”, it recognizes four different toxin epitopes; when “bi-specific” it recognizes two different toxin epitopes. The binding agents “tetrameric” as they bear four  $V_{H}H$  peptide monomers (when bi-specific, the first and second monomer have the same sequence and bind the same epitope, and the third and fourth monomers have the same sequence and bind the same epitope; when tetra-specific, each of the monomers has a different sequence and binds a different epitope).

When the binding agent is bi-specific, the first and third monomers have binding specificity for different epitopes, the first and second monomers have identical amino acid sequences, and the third and fourth monomers have identical amino acid sequences. In certain aspects, one of the  $V_{H}H$  peptide monomers has binding specificity for an epitope of TcdA and one of the  $V_{H}H$  peptide monomers has binding specificity for an epitope of TcdB.

When the binding agent is tetra-specific, each of the  $V_{H}H$  peptide monomers has binding specificity for a different epitope. In certain aspects, two of the  $V_{H}H$  peptide monomers have binding specificity for epitopes of TcdA and two of the  $V_{H}H$  peptide monomers have binding specificity for epitopes of TcdB.

In certain aspects, each of the  $V_{H}H$  peptide monomers has binding specificity for epitopes of TcdA. In other aspects, each of the  $V_{H}H$  peptide monomers has binding specificity for epitopes of TcdB.

In certain aspects, the  $V_{H}H$  peptide monomers independently have binding specificity for an epitope in the glucosyltransferase domain, cysteine protease domain, translocation domain or receptor binding domain of TcdA or TcdB.

A specific example of a bi-specific, tetrameric IgG binding agent comprises a light (kappa) chain having the amino acid sequence set forth in SEQ ID NO:40 (AA6 kappa) and a heavy chain having the amino acid sequence set forth in SEQ ID NO:36 (AH3 heavy). The invention also includes sequence variants thereof having at least 95% sequence identity, where the sequence variant retains toxin neutralizing activity. The variant amino acids of the sequence variant may be located in framework regions of the  $V_{H}H$  peptide monomers.

Another specific example of a bi-specific, tetrameric IgG binding agent comprises a light (kappa) chain having the amino acid sequence set forth in SEQ ID NO:42 (E3 kappa) and a heavy chain having the amino acid sequence set forth in SEQ ID NO:38 (5D heavy). The invention also includes sequence variants thereof having at least 95% sequence identity, where the sequence variant retains toxin neutralizing activity. The variant amino acids of the sequence variant may be located in framework regions of the  $V_{H}H$  peptide monomers.

Table 6 provides the sequences used to generate bi- and tetra-specific  $V_{H}H$ -IgG binding agents. Other suitable pairings include (i) 5D-IgG1-heavy chain+AA6-light (kappa or lambda) chain, and (ii) AH3-IgG1-heavy chain+E3-light (kappa or lambda) chain.

TABLE 6

Name	SEQ ID NO for Amino Acid Seq.	SEQ ID NO for Nucleic Acid Seq.
AH3-IgG1 heavy chain	36	37
5D-IgG1 heavy chain	38	39
AA6-IgG1 light (kappa) chain	40	41
E3-IgG1 light (kappa) chain	42	43
AH3/5D-IgG1 heavy chain	44	45
AA6/E3-IgG light (kappa) chain	46	47

However, the present invention includes IgG1 heavy chains joined to any of AH3, 5D, AA6 and E3, and IgG1 light (kappa or lambda) chains joined to any of AH3, 5D, AA6 and E3. Further, all possible combinations of the heavy and light (kappa or lambda) chains are encompassed herein. Humanized Binding Agents

Due to their small size and the high degree of identity of their framework to the human  $V_H$  framework of family III,  $V_H$  peptide monomers are expected to exhibit low immunogenicity when administered to humans. While the systemic application of small monovalent  $V_H$  monomers seems to induce little, if any, neutralizing antibody responses, protein immunogenicity generally increases with size and complexity. Two major hurdles for repeated and/or long-term in vivo use of  $V_H$  monomers are their likely short half-life and potential immunogenicity. To increase the valence and circulating half-life,  $V_H$  monomers can be fused with human IgG and Fc domains as discussed herein. To address possible immunogenicity, the  $V_H$  monomers can be humanized as needed without compromising their expression level, affinity, solubility, and stability. These strategies should result in good expression, stability, and solubility of humanized  $V_H$  monomers ( $hV_H$  monomers), while retaining the antigen specificity and affinity of the loop donor  $V_H$ .

$hV_H$  monomers that gain highest identity to human  $V_H$  gene(s) and possess the highest binding/neutralizing activity are selected, after which they are transferred into the  $V_H$ -multimers (e.g., ABAB),  $V_H$ -Fc and  $V_H$ -IgG constructs to generate fully humanized binding agents, such as fully humanized ABAB, ABAB-IgG and ABAB-Fc binding agents. The protein sequences of these humanized binding agents can be essentially identical to that of a human antibody variant, despite the non-human origin of some of its CDR segments that are responsible for the ability of the antibody to bind to its target antigen. Therefore, this strategy decreases the chance for potential immunogenicity in vivo and thus increase their safety and half-life in vivo.

The binding agents of the present invention thus encompasses humanized versions of each of the binding agents defined herein, comprising  $hV_H$  peptide monomers.

#### Epitope Binding Fragments

The binding agents of the invention include epitope binding fragments of each of the  $V_H$ -Fc and  $V_H$ -IgG binding agents defined herein. Because the  $V_H$ -Fc and  $V_H$ -IgG binding agents are comparable in structure to human IgG antibodies, where the variable regions are replaced by the  $V_H$  monomers, terms for human antibody fragments are also applicable to the such binding agents. The fragments include, but are not limited to, Fab fragments,  $F(ab')_2$  fragments, single chain Fv (scFv) antibodies, and fragments produced by an Fab expression library, as well as bi-specific antibody and triple-specific antibodies.

The  $V_H$ -Fc and  $V_H$ -IgG binding agents of the invention include fully human, humanized, and chimeric binding agents. The binding agents may be monoclonal or polyclonal. Further, the binding agents may be recombinant binding agents.

The binding agents may be produced in any species of animal, though preferably from a mammal such as a human, simian, mouse, rat, rabbit, guinea pig, horse, cow, sheep, goat, pig, dog or cat. For example, the binding agents can be 10 human or humanized, or any binding agent preparation suitable for administration to a human.

#### Polynucleotide, Expression Vectors, Host Cells and Method of Making

The invention includes polynucleotides comprising 15 nucleotide sequences encoding each the binding agents provided herein, as well as complementary strands thereof.

The invention also includes expression vectors comprising the polynucleotides, and host cells comprising the expression vectors. Suitable expression vectors include, e.g., 20 pcDNA3.1 and pSec-His, as well as plasmids used to transform yeast cells into producers and secretors of the binding agents of the invention. Suitable host cells include, e.g., Chinese hamster ovary cells (CHO cells), human embryonic kidney cells 293 (HEK 293 cells), yeast cells, and insect cells.

The invention further includes methods of producing the binding agents defined herein, comprising culturing the host cells under conditions promoting expression of the binding agents encoded by the expression vectors, and recovering 30 the binding agents from the cell cultures.

#### Engineered Strains of Yeast

Each of the binding agents of the invention may also be produced by engineered strains of *Saccharomyces* yeast. Accordingly, the invention is also directed to strains of 35 *Saccharomyces* yeast, such as *S. cerevisiae* and *S. boulardii*, engineered to produce one or more of the binding agents defined herein including, but not limited to,  $V_H$  monomer binding agents (see Table 1),  $V_H$  homodimer binding agents,  $V_H$  heterodimer binding agents (see Table 3), 40 ABAB binding agents,  $V_H$ -Fc binding agents (see Table 5),  $V_H$ -IgG binding agents (see Table 6), and epitope bidding fragments thereof. In preferred aspects, the engineered strains of *Saccharomyces* yeast secrete the binding agents.

45 The identity of the *Saccharomyces* yeast strain is only limited in that it can be engineered to produce, and preferably secrete, one or more of the binding agents of the invention. In preferred aspects of the invention, the strain of *Saccharomyces* yeast engineered to produce one or more of the binding agents is *S. cerevisiae* or *S. boulardii*. The invention thus encompasses an engineered strain of *S. cerevisiae* that produces one or more of the binding agents defined herein, as well as an engineered strain of *S. cerevisiae* that secretes one or more of the binding agents defined 50 herein. The invention also encompasses an engineered strain of *S. boulardii* that produces one or more of the binding agents defined herein, as well as an engineered strain of *S. boulardii* that secretes one or more of the binding agents defined herein. Suitable stains of yeast also include *Schizosaccharomyces pombe*, *Saccharomyces paradoxus*, and *Saccharomyces unisporus*.

55 *S. boulardii* is an FDA-designated Generally Regarded as Safe (GRAS) organism and it is commonly available over-the-counter for use in promoting intestinal health and amelioration of gastrointestinal illness due to diarrheal diseases. This species of yeast has been studied in multiple randomized double-blinded placebo-controlled clinical trials for 60

both safety and efficacy against intestinal diseases including CDI [42-46]. A suitable strain of *S. boulardii* is the *S. boulardii* strain MYA796 (ATCC, Manassas, VA).

A particular example of the engineered strains of *Saccharomyces* yeast of the invention is an engineered strain of *Saccharomyces* yeast that produces a binding agent comprising a  $V_{H}H$  peptide monomer or linked groups of  $V_{H}H$  peptide monomers comprising two, three, four, or more monomers, each of which binds TcdA and/or TcdB, preferably with specificity. Thus, the invention encompasses engineered strains of *Saccharomyces* yeast that produces  $V_{H}H$  peptide binding agents comprising at least one  $V_{H}H$  peptide monomer, wherein each  $V_{H}H$  peptide monomer has binding specificity for an epitope of *C. difficile* toxin A (TcdA) or toxin B (TcdB). In certain aspects, these binding agents comprise two, three, four, or more linked  $V_{H}H$  peptide monomers. The  $V_{H}H$  peptide monomers include, but are not limited to, the  $V_{H}H$  peptide monomers 5D (SEQ ID NO:1), E3 (SEQ ID NO:3), AA6 (SEQ ID NO:5), and AH3 (SEQ ID NO:7).

Another particular example of the engineered strains of *Saccharomyces* yeast of the invention is an engineered strain of *Saccharomyces* yeast that produces binding agents comprising  $V_{H}H$  peptide monomers joined to IgG antibodies, where the binding agents bind TcdA and/or TcdB, as defined herein. In these IgG-based binding agents, the variable regions of the light and heavy chains of IgG antibodies are replaced by one, two, three, four or more of the  $V_{H}H$  peptide monomers.

A further particular example of the engineered strains of *Saccharomyces* yeast of the invention is an engineered strain of *Saccharomyces* yeast that produces binding agents comprising  $V_{H}H$  peptide monomers joined to antibody Fc domains, where the binding agents bind TcdA and/or TcdB, as defined herein. In these Fc domain-based binding agents, one, two, three, four or more of the  $V_{H}H$  peptide monomers are joined to the hinge,  $C_{H}2$  and  $C_{H}3$  regions of each arm of Fc domain of an antibody heavy chain. Thus, the peptide monomers replace the Fab regions of an antibody.

An additional particular example of the engineered strains of *Saccharomyces* yeast of the invention is an engineered strain of *Saccharomyces* yeast that produces a tetra-specific, tetrameric binding agent, wherein the binding agent comprises linked first, second, third and fourth  $V_{H}H$  peptide monomers, and wherein the  $V_{H}H$  peptide monomers independently have binding specificity for an epitope of *Clostridium difficile* toxin A (TcdA) or toxin B (TcdB). In certain aspects, the first, second, third and fourth  $V_{H}H$  peptide monomers each has binding specificity for a different epitope. In certain aspects, the two of the  $V_{H}H$  peptide monomers have binding specificity for epitopes of TcdA and two of the  $V_{H}H$  peptide monomers have binding specificity for epitopes of TcdB. In certain aspects, the  $V_{H}H$  peptide monomers independently have binding specificity for an epitope in the glucosyltransferase domain, cysteine protease domain, translocation domain or receptor binding domain of TcdA or TcdB. Suitable  $V_{H}H$  peptide monomers include the AH3 monomer (SEQ ID NO:7), the AA6 monomer (SEQ ID NO:5), the 5D monomer (SEQ ID NO:1), and the E3 monomer (SEQ ID NO:3). Other monomers include, but are not limited to, those provided in Table 1.

In a preferred example, the invention is directed to an engineered strain of yeast, wherein the binding agent is ABAB, wherein the first and third monomers have binding specificity for epitopes of TcdA and the first and third monomers are  $V_{H}H$  peptide monomers AH3 (SEQ ID NO:7) and AA6 (SEQ ID NO:5), respectively, and wherein the

second and forth monomers have binding specificity for epitopes of TcdB and the second and forth monomers are  $V_{H}H$  peptide monomers 5D (SEQ ID NO:1) and E3 (SEQ ID NO:3), respectively.

The ABAB binding agent may comprise the amino acid sequence set forth in SEQ ID NO:19, or a sequence variant having at least 95% sequence identity thereto, wherein the sequence variant retains TcdA and/or TcdB binding specificity, or the sequence variant retains toxin neutralizing activity, or both.

The ABAB binding agent may also comprise an N-terminal secretion signal selected from the secretion signals provided in Table 4. In preferred aspects, the N-terminal secretion signal is the AT secretion signal (MRFPSIFTAVLFAASSALA (SEQ ID NO:99)) or the IVS secretion signal (MLLQ AFLFLLAGFAAKISA (SEQ ID NO:103)).

The ABAB binding agent may be expressed from a plasmid within the yeast. The plasmid may be, but is not limited to, pCEV-URA3-TEF-AT-yABAB-cMyc (SEQ ID NO:88). The ABAB binding agent encoded by the plasmid may comprise the amino acid sequence set forth in SEQ ID NO:107, or a sequence variant having at least 95% sequence identity thereto, and wherein the sequence variant retains TcdA and/or TcdB binding specificity, or the sequence variant retains toxin neutralizing activity, or both.

The ABAB binding agent may also be expressed from coding sequence integrated into a chromosome of yeast. The ABAB binding agent expressed from coding sequence integrated into a yeast chromosome may comprise the amino acid sequence set forth in SEQ ID NO:109, or a sequence variant having at least 95% sequence identity thereto, and wherein the sequence variant retains TcdA and/or TcdB binding specificity, or the sequence variant retains toxin neutralizing activity, or both.

The invention is also directed to engineered strains of *Saccharomyces* yeast that produce a therapeutic protein having binding specificity for a unique epitope of *Clostridium difficile* toxin A (TcdA) or toxin B (TcdB), or both. Preferably, the engineered strain of *Saccharomyces* yeast is *S. cerevisiae* or *S. boulardii*. A therapeutic protein is any protein that can bring about an improvement or cure in a medical condition in a subject, or that can inhibit or prevent a medical condition from developing in a subject. Suitable therapeutic protein include, but are not limited to, proteins that (a) replace a protein that is deficient or abnormal; (b) augment an existing pathway; (c) provide a novel function or activity; (d) interfere with a molecule or organism; and (e) deliver other compounds or proteins, such as a radionuclide, cytotoxic drug, or effector proteins. Therapeutic proteins also include antibodies and antibody-based drugs, Fc fusion proteins, anticoagulants, blood factors, bone morphogenetic proteins, engineered protein scaffolds, enzymes, growth factors, hormones, interferons, interleukins, and thrombolytics. Therapeutic proteins further include bispecific monoclonal antibodies (mAbs) and multispecific fusion proteins, mAbs conjugated with small molecule drugs, and proteins with optimized pharmacokinetics.

#### 60 Methods of Making Engineered Yeast Strains

The invention is also directed to methods of engineering strains of *Saccharomyces* yeast to produce one or more of the binding agents defined herein. The means used to produce the engineered strains of yeast are not particularly limited and there are a number of well-established techniques available for engineering yeast to produce homologous and heterologous proteins that will be known to the

31

skilled artisan. In certain aspects of these methods, *S. cerevisiae* or *S. boulardii* is engineered to produce the binding agents.

As an example, *Saccharomyces* yeast may be engineered to produce one or more of the binding agents defined herein by (a) transforming a strain of *Saccharomyces* yeast with an expression vector encoding the binding agent, and (b) screening the resulting yeast for production of the binding agent. In a certain aspect, the expression vector is plasmid pCEV-URA3-TEF-AT-yABAB-cMyc (SEQ ID NO:88). While this plasmid encodes a particular ABAB binding agent, the coding region for this binding agent can be replaced by the coding region of any of the binding agents defined herein.

As a further example, *Saccharomyces* yeast may be engineered to produce one or more of the binding agents defined herein by (a) chromosomally integrating a polynucleotide sequence encoding the binding agent into the genome of the strain of *Saccharomyces* yeast, and (b) screening the yeast of (a) for production of the binding agent. In certain aspects, the chromosomal integration is performed using a CRISPR technique [85-88]. As an example, such a method may include the steps of: (a) amplifying a polynucleotide sequence encoding the ABAB binding agent from plasmid pCEV-G4-Km-TEF-AT-yABAB hAA6T83N-tagless (SEQ ID NO:90) using primers containing (i) nucleic acid sequence homologous to a selected yeast chromosomal integration site and (ii) nucleic acid sequence homologous to regions 5' and 3' of ABAB binding agent coding sequence of the plasmid, to produce an integration cassette, (b) transforming yeast with the integration cassette produced in (a) with pCRI-Sb-δ1 (SEQ ID NO:91) or pCRI-Sb-δ2 (SEQ ID NO:92) to induce a double stranded break within the corresponding yeast chromosomal delta sites under conditions promoting spontaneous integration of the integration cassette into the site of the double stranded break, (c) screening the transformed yeast of (b) for production of the ABAB binding agent.

While the plasmid pCEV-G4-Km-TEF-AT-yABAB hAA6T83N-tagless encodes a particular ABAB binding agent, the coding region for this binding agent can be replaced by the coding region of any of the binding agents defined herein.

Suitable means used to screen the yeast for production of the binding agents will be readily apparent to the skilled artisan and include, but are not limited to immunoassays, such as an ELISA or a western blot.

#### Methods of Treatment and Prevention

The binding agents and engineered strains of *Saccharomyces* yeast of the invention can be used in methods of treating or preventing a disease symptom induced by *C. difficile* in a subject. These methods generally comprise administering a therapeutically-effective amount of one or more binding agents and/or one or more engineered strains of *Saccharomyces* yeast as defined herein to a subject having *C. difficile* infection or a risk of developing *C. difficile* infection. In certain aspects of this embodiment, the disease symptom induced by *C. difficile* is diarrhea.

The binding agents and engineered strains of *Saccharomyces* yeast of the invention can also be used in methods of neutralizing *C. difficile* toxin TcdA and/or TcdB in a subject infected by *C. difficile*. These methods generally comprise administering a therapeutically-effective amount of one or more binding agents and/or one or more engineered strains of *Saccharomyces* yeast as defined herein to a subject having *C. difficile* infection.

32

The binding agents and engineered strains of *Saccharomyces* yeast of the invention can further be used in methods of treating *C. difficile* infection in a subject. These methods generally comprise administering a therapeutically-effective amount of one or more of the binding agents and/or one or more engineered strains of *Saccharomyces* yeast as defined herein to a subject having *C. difficile* infection. These same methods can be used to treat CDI, as defined herein.

The binding agents and engineered strains of *Saccharomyces* yeast of the invention can also be used in methods of maintaining normal bowel function in a subject having a *C. difficile* infection. These methods generally comprise administering a therapeutically-effective amount of one or more of the binding agents and/or one or more engineered strains of *Saccharomyces* yeast as defined herein to a subject having *C. difficile* infection or a risk of developing *C. difficile* infection.

The binding agents and engineered strains of *Saccharomyces* yeast can also be used in immunoprophylaxis in order to prevent immediate CDI threats. In addition, passive immunoprophylaxis can be used to prevent both immediate and longer-term CDI threats. Each approach has its own particular advantages and is suitable to target a particular high-risk population. These methods generally comprises administering a therapeutically-effective amount of one or more of the binding agent and/or one or more engineered strains of *Saccharomyces* yeast as defined herein to a subject a risk of developing *C. difficile* infection.

In preferred aspects of the methods of the invention, the *Saccharomyces* yeast is *S. cerevisiae* or *S. boulardii*.

Each of the methods of the invention may include administration of the one or more binding agents and/or the one or more engineered strains of *Saccharomyces* yeast in one or more pharmaceutical formulations comprising the binding agents and/or the engineered strains of *Saccharomyces* yeast and a pharmaceutically acceptable carrier or diluent. In preferred aspects, the *Saccharomyces* yeast is *S. cerevisiae* or *S. boulardii*.

As used herein, the terms "treat", "treating", and "treatment" have their ordinary and customary meanings, and include one or more of: blocking, ameliorating, or decreasing in severity and/or frequency a symptom of a *C. difficile* infection or a *C. difficile*-related disease (CDI) in a subject; and/or partly or fully inhibiting the biological activity and/or promoting the immunologic clearance of *C. difficile* TcdA and/or TcdB in a subject infected with *C. difficile*; and/or growth, division, spread, or proliferation of *C. difficile* cells or a *C. difficile* infection in a subject. Treatment means blocking, ameliorating, decreasing, or inhibiting by about 1% to about 100% versus a subject in which the methods of the present invention have not been practiced. Preferably, the blocking, ameliorating, decreasing, or inhibiting is about 100%, 99%, 98%, 97%, 96%, 95%, 90%, 80%, 70%, 60%, 50%, 40%, 30%, 20%, 10%, 5% or 1% versus a subject in which the methods of the present invention have not been practiced.

As used herein, the terms "prevent", "preventing" and "prevention" have their ordinary and customary meanings, and include one or more of: stopping, averting, avoiding, alleviating or blocking *C. difficile* from colonizing, developing or progressing in a subject; and/or partly or fully inhibiting the biological activity and/or toxic effects of TcdA and/or TcdB in a subject infected with *C. difficile*; and/or stopping, averting, avoiding, alleviating or blocking the growth, division, spread, or proliferation of bacterial cells or bacterial infection in a subject. Prevention means stopping by at least about 95% versus a subject to which the preven-

## 33

tion has not been administered. Preferably, the stopping is about 100%, about 99%, about 98%, about 97%, about 96% or about 95%. The results of the prevention may continue for a period of days (such as 1, 2, 3, 4, 5, 6 or 7 days), weeks (such as 1, 2, 3 or 4 weeks) or months (such as 1, 2, 3, 4, 5, 6 or more months).

The method of treating and preventing provided herein can be supplemented by also administering a therapeutically-effective amount of an antibiotic to the subject. Preferably, the antibiotic will have antibacterial activity against *C. difficile*.

## Pharmaceutical Formulations

While the binding agents and engineered strains of *Saccharomyces* yeast may be administered directly to a subject, the methods of the present invention are preferably based on the administration of a pharmaceutical formulation comprising one or more binding agents and/or one or more engineered strains of *Saccharomyces* yeast, and a pharmaceutically acceptable carrier or diluent. Thus, the invention includes pharmaceutical formulations comprising one or more of the binding agents and/or one or more engineered strains of *Saccharomyces* yeast defined herein and a pharmaceutically acceptable carrier or diluent.

Pharmaceutically acceptable carriers and diluents are commonly known and will vary depending on the particular binding agent or engineered strains of *Saccharomyces* yeast being administered and the mode of administration. Examples of generally used carriers and diluents include, without limitation: saline, buffered saline, dextrose, water-for-injection, glycerol, ethanol, and combinations thereof, stabilizing agents, solubilizing agents and surfactants, buffers and preservatives, tonicity agents, bulking agents, and lubricating agents. The formulations comprising binding agents and/or engineered strains of *Saccharomyces* yeast will typically have been prepared and cultured in the absence of any non-human components, such as animal serum (e.g., bovine serum albumin).

Pharmaceutical formulations comprising one or more binding agents and/or one or more engineered strains of *Saccharomyces* yeast may be administered to a subject using modes and techniques known to the skilled artisan. Characteristic of CDI disease may make it more amenable to treatment and prevention using colonic delivery of therapeutic agents, i.e., targeted delivery of binding agents to the lower GI tract, e.g., the large intestine or colon. Other modes of delivery include, but are not limited to, oral, nasal, anal, and via intravenous injection or aerosol administration. Other modes include, without limitation, intradermal, subcutaneous (s.c., s.q., sub-Q, Hypo), intramuscular (i.m.), intraperitoneal (i.p.), intra-arterial, intramedullary, intracardiac, intra-articular (joint), intrasynovial (joint fluid area), intracranial, intraspinal, and intrathecal (spinal fluids).

Depending on the means of administration, the dosage may be administered all at once, such as with an oral formulation in a capsule or liquid, or slowly over a period of time, such as with an intramuscular or intravenous administration.

The amount of binding agents, alone or in a pharmaceutical formulation, administered to a subject is an amount effective for the treatment or prevention of infection. Thus, therapeutically effective amounts are administered to subjects when the methods of the present invention are practiced. In general, between about 1 ug/kg and about 1000 mg/kg of the binding agent per body weight of the subject is administered. Suitable ranges also include between about 50 ug/kg and about 500 mg/kg, and between about 10 ug/kg and about 100 mg/kg. However, the amount of binding agent

## 34

administered to a subject will vary between wide limits, depending upon the location, source, extent and severity of the infection, the age and condition of the subject to be treated, the means of administration, etc. A physician will ultimately determine appropriate dosages to be used.

The amount of the engineered strains of *Saccharomyces* yeast, alone or in a pharmaceutical formulation, administered to a subject is an amount effective for the treatment or prevention of infection. Thus, therapeutically effective amounts are administered to subjects when the methods of the present invention are practiced. In general, between about 1 ug/kg and about 1000 mg/kg of the engineered strains of *Saccharomyces* yeast per body weight of the subject is administered. Suitable ranges also include between about 50 ug/kg and about 500 mg/kg, and between about 10 ug/kg and about 100 mg/kg. However, the amount of the engineered strains of *Saccharomyces* yeast administered to a subject will vary between wide limits, depending upon the location, source, extent and severity of the infection, the age and condition of the subject to be treated, the means of administration, etc. A physician will ultimately determine appropriate dosages to be used.

Administration frequencies of the binding agents, the engineered strains of *Saccharomyces* yeast, and pharmaceutical formulations comprising the binding agents and/or engineered strains of *Saccharomyces* yeast will vary depending on factors that include the location of the bacterial infection, the particulars of the infection to be treated or prevented, and the mode of administration. Each formulation may be independently administered 4, 3, 2 or once daily, every other day, every third day, every fourth day, every fifth day, every sixth day, once weekly, every eight days, every nine days, every ten days, bi-weekly, monthly and bi-monthly.

The duration of treatment or prevention will be based on location and severity of the infection being treated or the relative risk of contracting the infection, and will be best determined by the attending physician. However, continuation of treatment is contemplated to last for a number of days, weeks, or months.

In each embodiment and aspect of the invention, the subject is a human, a non-human primate, bird, horse, cow, goat, sheep, a companion animal, such as a dog, cat or rodent, or other mammal. The subjects to which the methods of the present invention can be applied include subjects having an underlying disease or condition that makes them more susceptible to *C. difficile* infections.

The invention also provides a kit comprising one or more containers filled with one or more of the binding agents, one or more of the engineered strains of *Saccharomyces* yeast, or one or more pharmaceutical formulations comprising binding agents and/or the engineered strains of *Saccharomyces* yeast. The kit may also include instructions for use. Associated with the kit may further be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration.

## III. Examples

*V<sub>H</sub>H* Monomer and Heterodimer Binding Agents

An efficient platform to screen single domain (monomeric), mono-specific *V<sub>H</sub>H* peptide monomers against specific domains of toxins TcdA and TcdB was established. Using highly immunogenic atoxic holotoxins for immunization, and bioactive chimeric toxins (with normal domain

35

functions) for screening, panels of  $V_H$  monomers binding to different domains of TcdA or TcdB were prepared. A majority of these  $V_H$  monomers possessed potent neutralizing activity and their binding to specific domains was determined (FIG. 2). The atoxic holotoxins have point mutations at their enzymatic glucosyltransferase domains as described previously [33]. The bioactive chimeric toxins were created by switching the functional domains between TcdA and TcdB, which was also described previously [33].

Several of the  $V_H$  monomers bind to highly conserved TcdA/TcdB epitopes. For example,  $V_H$  E3 binds to the Rho GTPase binding site and blocks glycosylation;  $V_H$  AH3 binds to the GT domain of the toxin;  $V_H$  7F binds to cysteine protease cleavage sites and blocks GT domain cleavage and release. Some  $V_H$  monomers have potent neutralizing activity capable of blocking toxin cytotoxic activity at nM concentrations (See Table 1; FIGS. 3A and 3B).

To enhance the binding activity, two domain (dimeric), bi-specific  $V_H$  heterodimers were created (Table 3; FIG. 3C), allowing a single protein to target two distinctive epitopes of the toxins. These bi-specific  $V_H$  heterodimers possessed substantially enhanced neutralizing activities compared with equimolar mixtures of the same two  $V_H$  monomers (FIG. 3D). Heterodimers 5D/E3 and AH3/AA6 were found to fully protect mice from lethal systemic TcdB or TcdA challenge respectively, whereas mixed 5D and E3, or AA6 alone were only partially protective (FIGS. 3E and 3F).

A tetra-valent, tri-specific  $V_H$  binding agent (ABA) was generated by genetically fusing  $V_H$ s with the highest neutralizing activities targeting conserved, non-overlapping epitopes (AH3/E3/E3/AA6) [41]. This rationally designed toxin binder achieved a substantially enhancing binding affinity and neutralizing activity over the individual monomers and potent therapeutic efficacy against fulminant CDI. ABA was able to broadly neutralize toxins from 11 different TcdA<sup>+</sup>TcdB<sup>+</sup> *C. difficile* clinical isolates but failed to neutralize TcdB derived from two TcdA<sup>-</sup>TcdB<sup>+</sup> strains. The amino acid sequence of ABA is set forth in SEQ ID NO:111.

The  $V_H$  monomers comprising the heterodimers were linked using a flexible linker selected from SEQ ID NOs: 9-13 (Table 2).

#### ABAB Binding Agent

Four domain (tetrameric), tetra-specific  $V_H$  binding agents were generated by linking  $V_H$  monomers AH3, 5D, E3, and AA6, namely ABBA (AH3/5D/E3/AA6) and ABAB (AH3/5D/AA6/E3). These tetra-specific, tetrameric binding agent targets conserved, non-overlapping epitopes and had excellent toxin neutralizing activity. In the design of ABAB (FIG. 4),  $V_H$  peptide monomers AH3 and AA6 were separated by placing the 5D monomers between them because AH3 and AA6 bind to GT and TD respectively (FIG. 2), which are spatially distant to each other. This design allowed AH3 and AA6 to bind to TcdA simultaneously.

In the construction of the ABAB binding agent, flexible linkers were placed between the  $V_H$  monomers (see FIG. 4). The complete nucleic acid sequence encoding ABAB is provided in SEQ ID NO:20; the amino acid sequence of the protein is provided in SEQ ID NO:19.

In certain variants, a His<sub>(6)</sub>-tag was provided at the amino terminus of the protein to aid in purification, an E-tag was provided at the carboxy terminus of the protein to aid in detection, and/or an albumin-binding peptide (ABP, DICL-

36

PRWGCLWD; SEQ ID NO:21) was placed at the carboxy end of the construct to increase serum half-life of the protein (See FIG. 4).

ABAB was found to exhibit substantial enhanced binding affinity (Table 7) and neutralizing activity (Table 8) over the individual monomers and ABA. In Table 8, Vero cells were exposed to 5 ng/ml of TcdA in the presence of serially diluted AA6, AH3, ABAB or Merck anti-TcdA HuMab [9]. The minimal doses of antibodies protecting cells from TcdA-induced cell rounding are shown.

TABLE 7

	$V_H$ <sub>S</sub>	$K_{on}$ (Ms <sup>-1</sup> )	$K_{off}$ (s <sup>-1</sup> )	$K_D$ (nM)
TcdA	AH3	$2.20 \times 10^4$	$7.10 \times 10^{-4}$	32.0
	AA6	$3.52 \times 10^4$	$6.92 \times 10^{-4}$	19.7
	ABAB	$6.96 \times 10^5$	$1.21 \times 10^{-6}$	0.002
TcdB	5D	$1.52 \times 10^6$	$9.94 \times 10^{-4}$	0.65
	E3	$2.95 \times 10^6$	$9.4 \times 10^{-5}$	0.03
	ABAB	$1.79 \times 10^6$	$3.57 \times 10^{-6}$	0.002

TABLE 8

	AA6	AH3	ABAB	Merck Anti-TcdA HuMab
	8 nM	8 nM	0.25 nM	>10 nM

ABAB was also found to compete with all four individual  $V_H$  peptide monomers in a competition ELISA and can simultaneously bind to both TcdA and TcdB as determined by sandwich ELISA. Furthermore, ABAB is broadly reactive, capable of neutralizing toxins from the 13 different *C. difficile* strains that represent most of the current epidemic strains (Table 9).

TABLE 9

Strains	Ribo-type	REA type	PFGE type	Toxins	Place/date of isolation	ABAB neutralization	
R20291	27	Bl	NAP1	TcdA/Tcd B	London/2006	Yes	
CD196	27	Bl	NAP1	TcdA/Tcd B	France/1985	Yes	
	630	12	R	TcdA/Tcd B	Zurich/1982	Yes	
	M120	78	BK	NAP7, 8, 9	TcdA/Tcd B	UK/2007	Yes
BI-9	1	J	NAP2	TcdA/Tcd B	Gerding Collection	Yes	
	Liv024	1	J	NAP2	TcdA/Tcd B	Liverpool/2009	Yes
	Liv022	106	DH	NAP11	TcdA/Tcd B	Liverpool/2009	Yes
TL178	2	G	NAP6	TcdA/Tcd B	Belfast/2009	Yes	
	TL176	14	Y	NAP4	TcdA/Tcd B	Cambridge, UK/2009	Yes
	TL174	15			TcdA/Tcd B	Cambridge, UK/2009	Yes
CD305	23				TcdA/Tcd B	London/2008	Yes
	CFS	17			TcdB	Belgium/1995/ human	Yes
	M68	17			TcdB	Dublin/2006/ human	Yes

Since ABAB shows high potency in binding to and neutralizing both toxins, its efficacy in treating fulminant

CDI was evaluated. A single injection with as low as 40 µg/kg of ABAB one-day post *C. difficile* spore challenge reversed fulminant CDI in mice. None of the ABAB-treated mice died whereas 50% of control mice became moribund by 3 days post-infection (FIG. 5A). ABAB is 4-log more potent in preventing mortality after systemic challenge with TcdA and TcdB than the Merck HuMabs (FIG. 5B) [9]. Thus, ABAB possesses extraordinary in vivo efficacy against *C. difficile* toxins and spore challenge.

Animal and human studies demonstrated that passively administered antitoxin antibodies provide protection against CDI. The initial studies here also showed that antitoxin polysera protected mice from primary CDI (FIGS. 6A and 6B) and recurrent/relapse CDI. These findings and results from FIGS. 5A and 5B supported the hypothesis and provided the rationale for development of a parenteral ABAB immunization strategy for preventing CDI. To achieve the goal of optimizing ABAB for systemic delivery, chimeric and humanized ABAB were generated as illustrated in FIG. 1, i.e., V<sub>H</sub>-Fc and V<sub>H</sub>-IgG binding agents as well as the humanized proteins hV<sub>H</sub>-Fc and hV<sub>H</sub>-IgG, after which leading proteins were evaluated for in vivo neutralizing activity and protection in animal models. Details regarding the preparation and testing of the additional binding agents are provided in the following paragraphs.

#### ABAB-Fc

ABAB-Fc binding agent was prepared by generating an expression vector encoding the V<sub>H</sub> peptide monomers AH3/5D/AA6/E3 (linked in the noted order) joined to a human IgG1 Fc domain. The V<sub>H</sub> peptide monomers were separated by flexible linkers of Table 2. The nucleic acid sequence encoding the protein is provided in SEQ ID NO:23. ABAB-Fc was expressed and purified from stable transfected HEK293 cell line culture supernatant using protein A beads under conditions permitting disulfide bond formation and bi-valent molecule production. The expression levels were about 20 mg/L of culture supernatant. ABAB-Fc is fully functional in binding and neutralizing both TcdA and TcdB (data not shown). The amino acid sequence of ABAB-Fc is provided in SEQ ID NO:22.

Mono-specific V<sub>H</sub>-Fc binding agents (AH3-Fc, 5D-Fc, E3-Fc, AA6-Fc) and bi-specific V<sub>H</sub>-Fc binding agents (AH3/5D-Fc) and AA6/E3-Fc) were also made using this Fc-fusion system. Table 5 above provides the sequences for these additional binding agents.

#### ABAB-IgG

As illustrated in FIG. 1, bi-specific V<sub>H</sub>-IgG (AH3/5D-IgG and E3/AA6-IgG) can be generated by fusing monomers with human IgG heavy and light (kappa or lambda) chains separately. Tetra-specific V<sub>H</sub>-IgG (ABAB-IgG) binding agents can be generated by fusing dimers with human IgG heavy and light chains separately. Co-transfecting the heavy and light chain constructs generates the AH3/5D-IgG, E3/AA6-IgG and ABAB-IgG chimeric proteins. The separation of two V<sub>H</sub>s into heavy and light chains likely improves the yield and stability of bi-specific and tetra-specific V<sub>H</sub> chimeric proteins. This allows determination of whether V<sub>H</sub>-human IgG chimeric antibody helps the stability and efficacy of ABAB in vivo. Similarly, further improvement of in vivo half-life of ABAB-IgG can also be tested in ABAB-IgG variants with enhanced binding affinity to FcRn receptor.

Bi-specific (AH3/5D-IgG1 and E3/AA6-IgG1) and tetra-specific (ABAB-IgG1) IgG1 binding agents were prepared by co-transfected expression vectors encoding the heavy

and light (kappa) chain of each binding agent. The V<sub>H</sub>H peptide monomers were separated by flexible linkers of Table 2.

Bi-specific, tetrameric V<sub>H</sub>-IgG1 binding agents were produced by preparing two separate expression vectors, the first encoding a V<sub>H</sub> peptide monomer joined to the human IgG1 antibody heavy chain (C<sub>H</sub>1-Hinge-C<sub>H</sub>2-C<sub>H</sub>3) lacking the heavy chain variable region and the second encoding a V<sub>H</sub> peptide monomer joined to the human IgG1 antibody light (kappa) chain (CK) lacking the light chain variable region. These binding agents are bi-specific and tetrameric in that each light chain of the resulting binding agent is linked to a first V<sub>H</sub> monomer and each heavy chain of the resulting binding agent is linked to a second V<sub>H</sub> monomer. Table 6 above provides the sequences for these additional binding agents. Suitable pairings include (i) AH3-IgG1-heavy chain+AA6-light (kappa or lambda) chain, (ii) 5D-IgG1-heavy chain+E3-light (kappa or lambda) chain, (iii) 5D-IgG1-heavy chain+AA6-light (kappa or lambda) chain, and (iv) AH3-IgG1-heavy chain+E3-light (kappa or lambda) chain.

Tetra-specific, octameric ABAB-IgG binding agents were prepared. These binding agents are tetra-specific and octameric in that each light (kappa or lambda) chain of the resulting binding agent is joined to two (a first and second) linked V<sub>H</sub> monomers and each heavy chain of the resulting binding agent is joined to a two (a third and fourth) linked V<sub>H</sub> monomer, where the first, second, third and fourth monomers binds to a different epitope.

A particular tetra-specific, octameric ABAB-IgG (FIG. 7) binding agent was produced by preparing two separate expression vectors, the first encoding the V<sub>H</sub> peptide monomers AH3/5D (linked in the noted order) joined to the human IgG1 antibody heavy chain (C<sub>H</sub>1-Hinge-C<sub>H</sub>2-C<sub>H</sub>3) lacking the heavy chain variable region and the second encoding the V<sub>H</sub> peptide monomers AA6/E3 (linked in the noted order) joined to the human IgG1 antibody light (kappa) chain (CK) lacking the light chain variable region. The nucleotide sequence encoding the AH3/5D-IgG1 heavy chain is provided in SEQ ID NO:45; the amino acid sequence is provided in SEQ ID NO:44. The nucleotide sequence encoding the AA6/E3-IgG1 kappa chain is provided in SEQ ID NO:47; the amino acid sequence is provided in SEQ ID NO:46.

The bi-specific (AH3/5D-IgG1 and E3/AA6-IgG1) and tetra-specific (ABAB-IgG1) IgG1 binding agents were expressed and purified from stable transfected HEK293 cell line culture supernatant using protein A beads under conditions permitting disulfide bond formation and bi-valent molecule production. SDS-PAGE shows more than 90% purity of the purified ABAB-IgG1 with total molecular weight (light and heavy chains together) around 218 KDa on non-reduced gel (data not shown). The molecular weight of heavy chain is 68 KDa and light chain is 41 KDa showed on reduced gel.

The binding of ABAB-IgG1 to TcdA and TcdB was determined. FIGS. 8A-8B illustrate the comparison of binding ABAB-IgG1 to both toxins with the individual components (AH3, AA6, E3, and 5D). FIG. 8A shows the results of experiments where plates were coated with 1 µg/ml TcdA (TxA). Serially diluted ABAB-IgG was added in concentrations of 0, 0.64, 3.2, 16, 80, 400 and 2,000 ng/ml. The plates were washed and Merck Ab (anti-TcdA), Fc-ABBA (ABAB-Fc), Habab (ABAB-IgG), and V<sub>H</sub> anti-TcdB monomers AA6 and AH3 were added in the indicated amounts (ng/ml). Appropriate labeled antibodies were used for detection. FIG. 8B shows the results of experiments where plates were

coated with 1 ug/ml TcdB (TxB). Serially diluted ABAB-IgG was added in concentrations of 0, 0.64, 3.2, 16, 80 and 400 ng/ml. The plates were washed and Merck Ab (Anti-TcdB), Fc-abba (ABAB-Fc), Habab (ABAB-IgG), and V<sub>H</sub>H anti-TcdB monomers E3 and 5D were added in the indicated amounts (ng/ml). Appropriate labeled antibodies were used for detection.

As expected, the tetra-specific antibody can bind to TcdA and TcdB simultaneously as determined by sandwich ELISA (FIGS. 9A-9B). In a first set of experiments, plates were coated with 1 ug/ml TcdA (TxA). Serially diluted ABAB-IgG (Habab) was added in concentrations of 0, 1.6, 8, 40, 200 and 1000 ng/ml. The plates were washed and the following amounts of TcdB were added: 1.6, 8, 40, 200, and 1000 ng/ml. Mouse anti-TxB antibodies (500×) and goat anti-mouse-IgG-HRP (3000×) antibodies were used for detection. The results provided in FIG. 9A show that TxB is detected by coating TxA, suggesting IgG-ABAB binds to TxA/B simultaneously. In a second set of experiments, plates were coated with 1 ug/ml TcdB (TxB). Serially diluted ABAB-IgG (Habab) was added in concentrations of 0, 1.6, 8, 40, 200 and 1000 ng/ml. The plates were washed and the following amounts of TcdA were added: 1.6, 8, 40, 200, and 1000 ng/ml. Mouse anti-TxA antibodies (500×) and goat anti-mouse-IgG-RP (3000×) antibodies were used for detection. The results provided in FIG. 9B show that TxA is detected by coating TxB, again suggesting IgG-ABAB binds to TxA/B simultaneously.

The neutralizing activities of ABAB-IgG1 against cytopathic effects of the toxins on cultured cells were also examined. TcdA (100 ng/ml, FIG. 10A) was mixed with serially diluted Merck anti-TcdA human monoclonal antibody, ABAB-IgG1 (Hababa), and V<sub>H</sub>H anti-TcdA monomers AA6 and AH3 before adding to Vero cell monolayers in 100 ul culture medium and incubated at 37° C. for 24 hours. The results provided in FIG. 10A show that ABAB-IgG1 is at least 1000-fold more potent than Merck antibodies in neutralizing TcdA. In similar experiments, TcdB (10 pg/ml, FIG. 10B) was mixed with serially diluted Merck anti-TcdB human monoclonal antibody, ABAB-IgG1 (Hababa), and V<sub>H</sub>H anti-TcdB monomers E3 and 5D before adding to Vero cell monolayers in 100 ul culture medium and incubated at 37° C. for 24 hours. The results provided in FIG. 10B show that ABAB-IgG1 is at least 1000-fold more potent than Merck antibodies in neutralizing TcdB.

The in vivo neutralizing activities of ABAB-IgG1 were studied in a mouse model of CDI, the results of which are shown in FIG. 11. Mice were challenged with lethal dose of a mixed TcdA and TcdB (25 ng each toxin per mouse) and 4 hour later, ABAB-IgG (10, 30 or 100 ug/kg), a mixture of Merck anti-toxin A and anti-toxin B antibodies (10 mg/kg) or PBS was administered to the mice. The results demonstrate that the neutralizing activity of ABAB-IgG was much greater than the Merck antibody, and at lower concentrations.

#### Animal Testing of ABAB-IgG

The ABAB-IgG binding agent was tested in both prophylactic treatment and re-challenge survival assays. FIG. 12 provides the experimental design of both studies. 6-8 week old female C57 mice were used, and the conditions included PBS: 10 ml/kg, i.p., n=14; ABAB-IgG: 200 ug/kg, i.p., n=10; ABAB-IgG: 1 mg/kg, i.p., n=10; ABAB-IgG: 5 mg/kg, i.p., n=10.

Table 10 provides a summary of the results seen with prophylactic treatment of mice against *C. difficile* spores (UK1, a 027/BI/NAP1 epidemic strain). ABAB-IgG or PBS was administered one day prior to administrating of *C.*

*difficile* spores. As can be seen, ABAB-IgG showed dose-related prophylactic protection against CDI, where 5 mg/kg showed complete protection on all the parameters examined and 200 ug/kg was found to be more potent than 200 ug/kg of bi-specific V<sub>H</sub>H fusion antibody ABA [41].

TABLE 10

	Diarrhea								Survival		
	Occurrence	Day 1		Day 2		Weight Change		Day 2	Day 3	Day 4	
		score	score	Overall	Day 2	Day 3	Day 4				
200 μg/kg		—	—	—	✓	—	✓	✓	—	✓	✓
1 mg/kg	✓	—	✓	—	✓	✓	✓	✓	—	✓	
5 mg/kg	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	

Table 11 provides a summary of the results seen with re-challenge of mice against *C. difficile* spores. ABAB-IgG or PBS was administered 15 days prior to administrating of *C. difficile* spores. As can be seen, one dose of ABAB-IgG showed some protection against the CDI caused by re-challenge of spores, but the protection was much less efficient compared to that during the primary challenge. This may be due to the drop of the antibody level with time and the generation of antibody in the PBS group following primary challenge.

TABLE 11

	Diarrhea								Survival		
	Occurrence	Day 1		Day 2		Weight Change		Day 2	Day 3	Day 4	
		score	score	Overall	Day 2	Day 3	Day 4				
200 μg/kg	✓	✓	—	—	—	—	—	—	—	—	—
1 mg/kg	✓	—	—	—	—	—	—	—	✓	—	
5 mg/kg	—	—	—	✓	—	—	—	—	—	—	

Intestinal delivery of IgG-ABAB was also tested for protection of mice from fulminant CDI. After a single IgG-ABAB injection into mouse ceca after a laparotomy, mice were completely protected against fulminant CDI of death outcomes whereas 50% of control mice succumbed (data not shown). Disease progress and severity were assessed daily using a clinical scoring system modified from a previous publication [62], which included four criteria (activity level, posture, coat, and diarrhea) each graded on a scale from 0 to 4 and added together to generate a score with a maximum value of 16. A normal mouse would score 0 and a mouse found dead was scored as 16. Mice with scores equal to or higher than 11 should be euthanized. Only one mouse in the IgG-ABAB treatment group developed transient diarrhea whereas mice injected with PBS developed severe CDI disease symptoms (data not shown). Thus, Ig-ABAB manually delivered by injection into mouse intestines showed potent therapeutic efficacy.

#### Expression, Purification and Evaluation of Binding Agents

A variety of selection criteria is used to select the binding agents generated in the experiments described in the approaches herein. First, each of the constructs defined herein can be used in transient transfections of 293T cells for making small-scale recombinant proteins by Protein A affinity chromatography. The production yield of each construct can be determined by quantitative ELISA. Second, binding activity of recombinant proteins can be screened using

41

ELISA and surface plasmon resonance (SPR) to select constructs that preserve their original binding activities against the toxins. Third, the proteins are evaluated for neutralizing activity in *in vitro* assays (FIG. 3).

Accumulating observations indicate that polyreactivity and/or autoreactivity of *in vivo* recombinant binding agents are potential issues related to their *in vivo* safety and half-life. The application of the selected ABAB binding agents as a systemic binding agent for preventing primary acute CDI likely requires that the chimeric and humanized ABAB proteins are limited in polyreactivity and/or autoreactivity. Progress in protein proteomics has made it possible to screen for polyreactivity and autoreactivity of recombinant antibodies *in vitro*, which is a great tool for surrogate therapeutic antibodies. Therefore, selected humanized binding agents with good yield, high binding affinity, and potent neutralizing activity can be further tested for potential polyreactivity and autoreactivity using the auto-antigen microarray test and ProtoArray protein microarrays (Invitrogen).

From the above *in vitro* assays, candidate ABAB-Fc and ABAB-IgG binding agents can be evaluated for their *in vivo* toxicity, serum half-life, and immunogenicity.

#### Generation of *S. cerevisiae* Secreting ABAB (Sc-ABAB)

Means for *in vivo* production and delivery of the binding agents to the gut of subjects having CDI or at risk of developing CDI were developed. Because *S. cerevisiae* is genetically similar to *S. boulardii* [52,53] and genetic tools are readily available for *S. cerevisiae*, *S. cerevisiae* was first used for ABAB secretion validation.

A novel bi-specific sandwich ELISA method was first developed to evaluate ABAB secretion. The setting utilizes purified TcdA and TcdB as binding antigens for ABAB bispecificity and α-TcdA antibodies for detection (FIG. 13A). For standardization, plates were coated with TcdB (1 ug/ml) into which was added serially diluted ABBA ((AH3-E3-E3-AA6)) standard. Serial diluted rTcdA (1 ug/ml to 7.8 ng/ml) was then added. The capture of TcdA was then measured by adding monoclonal antibody against TcdA followed by HRP conjugated secondary antibody. The results for the standard curves are shown in FIG. 13B. Based on these results, a standard curve derived using 125 ng/ml of rTcdA was chosen for determining secretion levels of ABAB in yeast culture supernatants and used for all subsequent ELISA.

A shuttle plasmid (pD1214-FAKS) containing origins of replication from both *E. coli* (pUC) and yeast (2 micron circle), as well as a yeast auxotrophic selection marker URA3 (conferring the ability to synthesize uracil), was obtained from DNA 2.0 (Newark, CA). The sequence encoding ABAB (SEQ ID NO:20), and His tag (SEQ ID NO:66) and D7 tag (SEQ ID NO:112) at the N-terminus and C-terminus of ABAB respectively, was inserted into this plasmid backbone in which transcription was controlled by the strong constitutive yeast translational elongation factor promoter ( $P_{TEF}$ ) and extracellular secretion provided by fusion to the alpha mating factor secretion signal leader sequence (FAKS). The sequence of the resulting plasmid (pD1214-FAKS-His-hABAB-D7) is provided in SEQ ID NO:68.

Plasmid pD1214-FAKS-His-hABAB-D7 was transformed into the *S. cerevisiae* strain BY4741 (MATa his3 Δ1 leu2 Δ0 Met15 Δ0 ura3 Δ0), an URA3 knockout S288C-derivative laboratory strain. Yeast transformants were then cultured in YNB medium containing dropout mix without uracil (6.8 g YNB, 20 g glucose, 2 g dropout mix in 1 L of sterile ddH<sub>2</sub>O) at 250 rpm at 30° C. overnight to reach O.D.

42

1 in a shaker. The cells were then centrifuged down and lysed by sonication in 1×SDS loading buffer. After sonication, total cell lysates were treated at 98° C. for 5 minutes before loading on a SDS gel. Same amount of yeast control 5 cell lysates were loaded in each well except the control cells were not viable in YNB medium without uracil and therefore were cultured in YNB complimented with uracil.

Culture supernatants from 25 yeast transformants as well as 3 yeast control colonies were centrifuged to spin down 10 cells, and the cell-free supernatants were then diluted with 2.5% milk in PBS containing 0.05% of tween 20 at 1:3 ratio and screened by the ELISA as described above after 24 hrs of incubation in a shaker at 250 rpm and 30° C. FIG. 14B shows that all the yeast transformants secreted ABAB in 15 culture supernatant compared to the culture supernatant from the yeast control colonies.

A cell-based neutralizing assay was used to assess the biological activity of secreted ABAB in culture supernatant. In this assay, sufficient amount of toxin A or toxin B to cause 20 100% cell rounding in 4 hours were added with PBS, cell-free culture supernatant from BY4741 control colony or BY4741-ABAB colony. Recombinant ABAB was used a positive control. The biological activity of secreted ABAB in 25 culture supernatant was determined by the level of neutralizing activity to prevent cell rounding. Full length ABAB secreted from *S. cerevisiae* indeed retains its neutralizing activity when compared with purified recombinant ABAB (FIG. 14A). These combined results imply the plausibility of ABAB secretion by *S. boulardii*.

In further experiments, it was demonstrated that oral 30 gavage of mice with Sc-ABAB at doses of 10<sup>10</sup> CFU had no adverse effects on mice, and mice shed live Sc-ABAB as determined by plating feces on Sabouraud CAF-Agar (data not shown). Isolates recovered from mice retained their 35 ability to produce functional ABAB using the assay described above.

#### ABAB Secretion Optimization

ABAB secretion level is imperatively linked to *in vivo* therapeutic efficacy. Therefore, the possibility of further 40 optimizing ABAB secretion by replacing the existing FAKS secretion signal with a number of commercially available secretion signals was explored. Secretion sequences facilitate co-translational or post-translational translocation of heterogeneous proteins into the endoplasmic reticulum and Golgi compartments prior to extracellular export. Although α-mating factor is a commonly used signal sequence for heterologous protein secretion that typically generates good yields of the secreted proteins in *S. cerevisiae* [69,70], studies have shown that other secretion sequences from other proteins such as inulinase or invertase could be more suitable for secreting certain heterologous proteins [71,72].

11 different commercially available secretion signals (Table 4; DNA 2.0, Newark, CA) were genetically fused with ABAB individually under the control of TEF promoter 55 in the same pD1214 plasmid backbone. Plasmids encoding ABAB with alternative secretion signals include the following plasmids where the FAKS secretion signal is replaced by the noted new secretion signals from Table 4 and where both the his-tag and D7-tag are removed:

- 60 Plasmid pD1214-AKS-hABAB (SEQ ID NO:70)
- Plasmid pD1214-AK-hABAB (SEQ ID NO:71)
- Plasmid pD1214-AT-hABAB (SEQ ID NO:72)
- Plasmid pD1214-AA-hABAB (SEQ ID NO:73)
- Plasmid pD1214-GA-hABAB (SEQ ID NO:74)
- Plasmid pD1214-IN-hABAB (SEQ ID NO:75)
- Plasmid pD1214-IVS-hABAB (SEQ ID NO:76)
- Plasmid pD1214-KP-hABAB (SEQ ID NO:77)

Plasmid pD1214-LZ-hABAB (SEQ ID NO:78)

Plasmid pD1214-SA-hABAB (SEQ ID NO:79)

In addition, both the his-tag and D7-tag in the original ABAB construct (pD1214-FAKS-His-hABAB-D7) were removed to produce plasmid pD1214-FAKS-hABAB (SEQ ID NO:69) and culture incubation temperature was raised to 37° C. to better accommodate *in vivo* and clinical testing relevant scenarios. All 11 plasmids were then transformed in BY4741 and 5 independent colonies from each selective plate were selected to generate culture supernatants. The amount of secreted ABAB was determined by the same ELISA as described above. In addition, E/O value was used to provide a fair comparison across all groups. E/O value is defined by ELISA O.D. value normalizes against culture O.D. value. Two of the best secretion signals for ABAB were found to be AT and IVS (Table 4; FIG. 15A).

Due to the unavailability of an auxotrophic mutant strain for *S. boulardii*, another 2 um-based plasmid carrying the *aphA1* gene encoding resistance to G418 (pCEV-G4-Km; SEQ ID NO:80; a gift from Lars Nielsen & Claudia Vickers (Addgene plasmid #46819)) was used instead of pD1214 plasmids to confirm ABAB secretion in *S. boulardii*. The best two secretion signals for *S. cerevisiae* (AT and IVS) were fused with ABAB genetically and inserted in the pCEV-G4-Km plasmid backbone to generate plasmids pCEV-G4-Km-TEF-AT-hABAB\*(SEQ ID NO:81) and pCEV-G4-Km-TEF-IVS-hABAB\*(SEQ ID NO:82). Both plasmids were used to transform *S. boulardii* (strain MYA796) and ABAB secretion with AT and IVS in *S. boulardii* was comparable with *S. cerevisiae* as determined by ELISA (FIG. 15B). A further construct, pCEV-G4-Km-TEF-AT-hABAB (SEQ ID NO:83), was prepared which differs from pCEV-G4-Km-TEF-AT-hABAB\* in that it contains a molecular cloning site between the AT and hABAB sequence.

ABAB secretion was then further optimized by yeast codon optimization (yABAB) at the nucleotide level in the construct having the AT secretion signal, producing plasmid pCEV-G4-Km-TEF-AT-yABAB (SEQ ID NO:84). A sequence containing 40 nucleotides between  $P_{TEF}$  and ABAB coding sequence was also found to be dispensable for ABAB secretion and removed resulting in plasmid pCEV-G4-Km-TEF-X40-AT-yABAB (SEQ ID NO:85). A further sequence containing two restriction cloning sites between AT and ABAB sequence was found to negatively impact ABAB secretion and therefore this sequence was also omitted (plasmid pCEV-G4-Km-TEF-AT- $R^S$ yABAB; SEQ ID NO:115) for subsequent study to maximize ABAB secretion.

Next, the amount of secretion of the individual monomers was measured and AA6 was found to be secreted the least. To improve AA6 secretion, and thus further optimize ABAB secretion, a panel of key amino acid residues was utilized. A T83N mutation was found to improve AA6 secretion. In addition, *S. boulardii* carrying the hAA6 sequence was found to secrete more AA6 than the one carrying the yeast optimized yAA6 sequence. Therefore, a comparison was undertaken between ABAB carrying the T83N mutation within AA6 (AT-yABAB T83N; plasmid pCEV-G4-Km-TEF-AT-yABAB AA6T83N; SEQ ID NO:116) and ABAB where the yAA6 sequence was replaced by the hAA6 T83N sequence (AT-yABAB hAA6 T83N; plasmid pCEV-G4-Km-TEF-AT-yABAB hAA6T83N, which has the sequence of SEQ ID NO:90 but lacks the coding sequence for c-Myc) to determine which sequence exhibited better secretion. It was found that there was no significance difference between these constructs and AT-yABAB hAA6 T83N was con-

cluded as the final sequence moving forward. The nucleotide sequence encoding AT-yABAB hAA6 T83N is provided in plasmid pCEV-G4-Km-TEF-AT-yABAB hAA6T83N-taggless (SEQ ID NO:90). The amino acid sequence of AT-yABAB hAA6 T83N is provided in SEQ ID NO: 117.

#### Generation of an Auxotrophic *S. boulardii* Strain

The expression plasmid encoding ABAB can be cloned into the *S. boulardii* strain. The *S. boulardii* strain can tolerate normal body temperature and acidic conditions better than *S. cerevisiae*, which can improve efficacy as a novel oral yeast-based therapeutic strategy. Two modifications to a wild-type *S. boulardii* strain can be made to preserve the *in vivo* stability of the expression plasmid conferred by the yeast URA3 metabolic selection marker: 1) a diploid auxotrophic mutant carrying a deletion in both chromosomal alleles of URA3 can be constructed, and 2) the endogenous 2 micron circle can be cured from *S. boulardii* to prevent unintended recombination from interfering with ABAB expression.

The most straightforward and efficient method for constructing auxotrophic mutants in wild-type *Saccharomyces* strains involves targeted deletion of chromosomally encoded genes by homologous recombination, which occurs at very high frequencies in *Saccharomyces*. Complete deletion of the targeted gene is preferred over selection of spontaneous mutations which can revert back to the wild type. Thus a gene deletion is preferred for the haploid state in *S. cerevisiae* which is typically induced from wild-type diploid via sporulation using a nutritionally poor growth medium and incubating at low temperature (30° C.). However, *S. boulardii* is sporulation deficient and recalcitrant to formation of haploid cells under normal sporulation conditions [64,65]. A two-step process for deletion of both chromosomal gene alleles (e.g. URA3) was used in which each deletion step can be selected for. The process is outlined schematically in FIG. 16.

All chromosomal deletions were carried out by lithium acetate-facilitated genetic transformation [73] of linear DNA deletion cassettes. Lithium acetate-based transformation originated from a *S. cerevisiae* protocol and was found to be compatible with *S. boulardii* although *S. boulardii* was found to be much harder to transform [55,56]. The difference is around 100 fold. Transformation efficiency in *S. cerevisiae* can be improved by adjusting glucose concentration and heat shock time [74]. Therefore various glucose concentrations and heat shock times were incorporated in *S. boulardii* transformation for optimization. The best condition tested for *S. boulardii* was 2% glucose in preculture and 20 minutes of heat shock time at 42° C. and these conditions were used for all transformation procedures in all studies.

Two deletion cassettes containing the genes *aphA1* and *ble*, which confer resistance to G418 and phleomycin in yeast respectively, were generated by PCR using pCEV-G4-Km (SEQ ID NO:80) and pCEV-G4-Ph (SEQ ID NO:86) (a gift from Lars Nielsen & Claudia Vickers (Addgene plasmid #46820)) as templates. Both deletion cassettes are flanked by two locus of X-over P1 (*loxP*) in the same direction, allowing for antibiotic resistance genes spin out using Cre recombinase. 40 base pairs of homologous sequences upstream of URA3 promoter ( $P_{URA3}$ ) and downstream of the stop codon of URA3 were incorporated in PCR primers to generate two final deletion cassettes for site-specific gene deletion in *S. boulardii* (see FIG. 16). The exact sequence and location of URA3 gene on chromosome V on *S. boulardii* was mapped using URA3 gene annotation from online-published sequence from *Saccharomyces* genome database (SGD). Selection for crossover 1 replacing the first

URA3 allele with aphA1 deletion cassette is selected for using resistance to G418 [66]; the second crossover replacing the second URA3 allele with ble deletion cassette is selected for using resistance to phleomycin [75] (FIG. 16). The replacement of both URA3 alleles with aphA1 and ble deletion cassettes was evidenced by resistance to both antibiotics (data not shown) as well as lack of growth on minimal synthetic medium plates lacking uracil (data not shown). Yeast phenotype was also confirmed by growth on Sabouraud plate with chloramphenicol (100 µg/ml) (data not shown). In addition, three sets of unique primers targeting the URA3, aphA1 or ble genes in the URA3 chromosomal region was designed and performed PCR using wild type (WT), URA3Δ::aphA1/URA3 (1<sup>st</sup> crossover) and URA3Δ::aphA1/Δ::ble (2<sup>nd</sup> crossover) genomic DNA as templates. Expected PCR product sizes targeting the URA3, aphA1 or ble genes in the URA3 chromosomal region are 766 bp, 1183 bp, and 662 bp respectively. DNA electrophoresis of PCR products from WT, 1<sup>st</sup> crossover and 2<sup>nd</sup> crossover clones using these three sets of unique primers confirmed the absence of URA3 alleles and integration of the aphA1 and ble deletion cassettes of the 2<sup>nd</sup> crossover strain.

The 2<sup>nd</sup> crossover strain was then transformed with pPL5071\_TEF1-Cre\_URA3 (pPL5071; SEQ ID NO:95) [76] to remove the aphA1 and ble deletion cassettes. Strain carries pPL5071 expresses Cre recombinase constitutively under P<sub>TEF</sub>. Cre recombinase then targets loxp sequences flanking the aphA1 and ble deletion cassettes; this causes the excision of the aphA1 and ble deletion cassettes, leaving only one loxp site in the URA3 chromosomal region. Strains that underwent successful excision of the aphA1 and ble deletion cassettes cannot grow in the presence of either G418 or phleomycin; yet retain the loss of both URA3 alleles, therefore can only grow on minimal synthetic medium plate in the presence of uracil and showed no growth on minimal synthetic medium plate without uracil supplement.

Removal of pPL5071 was achieved by growth in YPD and selecting for colonies later grown on minimal synthetic medium containing uracil and the pyrimidine analog 5-fluoro-orotic acid (5-FOA) [77]. Strains possessing pPL5071 carry the URA3 gene that can synthesize the toxic intermediate 5-fluorodeoxyuridine a potent inhibitor of thymidylate synthetase, which interrupts DNA synthesis and leads to cell death and allows selection of strains that have lost pPL5071. The absence of pPL5071 also was confirmed by pPL5071 specific primers by PCR and DNA electrophoresis of the PCR product.

The 2 um plasmid is a very stable 6.1 kb plasmid that is ubiquitous in *Saccharomyces* strains. This plasmid confers no selective advantage to the yeast host organism, and it is remarkably stable due to the presence of an efficient REP1-REP2-STB plasmid partitioning system [68]. *S. boulardii* strains used also contain this plasmid as confirmed via PCR. To remove the 2 um plasmid, pBIS-GALkFLP-URA3 (SEQ ID NO:87) [67] was used to cure 2 um plasmid, followed by removal with uracil and 5-FOA. Loss of the 2 um plasmid was confirmed by PCR using primers specific for the origin of replication.

The auxotrophic strain of *S. boulardii* that results from these manipulations is termed *S. boulardii* URA3 Δ/Δ. Auxotrophic *S. boulardii* Strain for In Situ Delivery of ABAB

For constructing the auxotrophic *S. boulardii* strain for in situ delivery of ABAB, the aphA1 cassette of the plasmid pCEV-G4-Km-TEF-X40-AT-yABAB (SEQ ID NO:85) was replaced by the URA3 cassette from pD plasmid to generate

the plasmid pCEV-URA3-TEF-AT-yABAB-cMyc (SEQ ID NO:88). This plasmid was then used to transform *S. boulardii* URA3 Δ/Δ. The resulting strain secretes fully functional ABAB when compared with purified ABAB in a cell toxicity assay (FIG. 17C). Western blotting showed the corresponding ABAB band from *S. boulardii* culture supernatant using α-Llama antibodies conjugated with HRP (FIG. 17D). C-terminus end of ABAB contains c-Myc tag and can be further pulled down by α-c-Myc antibodies (FIG. 17D).

For empty plasmid (EP) control, AT-yABAB sequence was later removed from pCEV-URA3-TEF-AT-yABAB-cMyc (SEQ ID NO:88) to generate pCEV-URA3-TEF-cMyc (SEQ ID NO:89). *S. boulardii* URA3 Δ/Δ strain transformed with this plasmid results a strain complemented with URA3 but does not secrete ABAB. *S. boulardii* URA3 Δ/Δ strain secreting ABAB also showed no growth inhibition when cultured in YPD containing vancomycin (1 mg/ml) (FIG. 17A). This suggests *S. boulardii* can be co-administered with vancomycin typically used to treat CDI patients and secretes ABAB to treat ongoing CDI. In addition, purified ABAB is stable in culture supernatant collected from *S. boulardii* at O.D. 10 over 2 hours period of time suggests secreted ABAB is likely to diffuse out from *S. boulardii* without being degraded.

#### 25 Safety Assessment of *S. boulardii* Delivered Orally to Antibiotic-Treated Mice

Prior to evaluating whether *S. boulardii* URA3 Δ/Δ expressing ABAB can protect mice in CDI models [20,33, 62,78], a safety assessment was performed to determine safe 30 doses of *S. boulardii* in antibiotic-treated mice. In this safety assessment mice were first supplied with an antibiotic cocktail in their daily drinking water for three days and then switched to regular water. One day before oral delivery of *S. boulardii*, mice were injected with clindamycin intraperitoneally. This completes the antibiotic treatment for the mice and *S. boulardii* was then orally delivered to the mice for safety assessment, which includes monitoring of daily 35 weight change and persistence of *S. boulardii* in their stool samples of these antibiotic-treated mice. Mice exhibited no signs of illness and steadily weight increase during 6 days of 40 monitoring when 10<sup>10</sup> cells of *S. boulardii* were delivered orally consistent with the idea of *S. boulardii* as a GRAS organism. For the subsequent CDI mouse studies, however, only 10<sup>9</sup> cells of *S. boulardii* were given due to the ease of 45 pellet resuspension and less variability of the dosing amount to the mice, which can occur with high viscosity present in resuspension. *S. boulardii* also shows limited colonization in these antibiotic-treated mice GI tracts; three days after the final gavage, no detectable *S. boulardii* were recovered from 50 Sabouraud plate (data not shown).

#### Protection of *S. boulardii* Expressing ABAB Against Primary CDI in Mice

Protection of *S. boulardii* expressing ABAB was evaluated using established primary mouse CDI models. *S. boulardii* expressing ABAB was delivered either as preventative or treatment against primary CDI in mice. In brief, primary CDI was established in mice by supplementing a mixture of antibiotic into their drinking water for three days, and then intraperitoneal injection of clindamycin 24 hours prior to *C. difficile* spore challenge. 10<sup>5</sup> *C. difficile* spores (UK1, a 027/B1/NAP1 epidemic strain) were gavaged in the mice to induce CDI. For preventative evaluation, mice started receiving an oral dose of *S. boulardii* the day after switching to regular drinking water, which continued every day for 7 55 days. For therapeutic evaluation, mice received an oral dose of *S. boulardii* at 6, 24, 48, and 72 hours after spore challenge. Controls included PBS and *S. boulardii* trans-

formed with an empty plasmid. In both methods, mice receiving *S. boulardii* expressing ABAB were significantly protected against CDI-induced deaths (FIGS. 18A and 19A; PBS: negative control; Sb:EP: *S. boulardii* transformed with an empty plasmid; Sb:BAB: *S. boulardii* secreting ABAB). CDI mice typically suffered weight loss with most weight drops around day 2 to day 3 due to diarrhea and gradually recovered. Weights of mice receiving *S. boulardii* expressing ABAB recovered significantly sooner (FIGS. 18B and 19B) and had significant reduced percentage of diarrhea incidents after day 2 post challenge (FIGS. 18C and 19C). Protection of *S. boulardii* Expressing ABAB Against Recurrence CDI in Mice

Protection of *S. boulardii* expressing ABAB was evaluated against recurrence CDI in mice. To induce recurrent CDI, mice were given three days of antibiotic cocktail in their daily drinking water. After three days of antibiotic water, mice were then switched back to drinking regular water. One day before oral delivery of  $10^5$  *C. difficile* spores (UK1, a 027/BI/NAP1 epidemic strain), mice were injected with clindamycin intraperitoneally. Six hours after spore challenge, regular water was changed to water containing 0.5 mg/ml of vancomycin for six days and regular water was switched back again for the rest of study. Mice typically develop signs of CDI after 4 days of vancomycin withdrawal without another *C. difficile* spore challenge. During the course of recurrence model, *S. boulardii* was orally delivered along with vancomycin water once every day for 12 days. This model is used to evaluate protection efficacy of *S. boulardii* expressing ABAB for preventing CDI recurrence in mice. Survival rate, weight loss and diarrhea incident of these mice were monitored on a daily basis. Controls included PBS and *S. boulardii* transformed with an empty plasmid. Mice receiving *S. boulardii* expressing ABAB were significantly protected against recurrence-induced CDI deaths (FIG. 20A; PBS: negative control; Sb:EP: *S. boulardii* transformed with an empty plasmid; Sb:BAB: *S. boulardii* secreting ABAB). Similar to primary CDI mice, recurrent CDI mice also typically suffered weight loss with most weight drops around day 4 to day 5 after vancomycin water withdrawal. Mice receiving *S. boulardii* expressing ABAB were significantly protected from weight loss (FIG. 20B) and had significant reduced percentage of diarrhea recurrence incidents (FIG. 20C).

#### Stability Optimization of ABAB Cassette Through Chromosomal Integration

Genome editing using a CRISPR-Cas9 based system has been recently demonstrated both in *S. cerevisiae* and *S. boulardii* [79-81]. In addition, large fragment deletion can be achieved by targeting two guide sequences simultaneously [82]. Foreign genes are typically more steadily maintained when integrated into chromosomes versus introduced via plasmids when there is no selection pressure. However, chromosomal integration often requires multiple rounds of integration to achieve high copies. A protocol reported in a recent publication overcame this hurdle through targeting multiple copies of common sequences such as  $\delta$  sites in *S. cerevisiae* genome through CRISPR-induced double strand breaks and achieved concurrent integration of large fragments in these sites [83]. DNA double strand break can be repaired either by non-homologous end joining or homologous recombination; however, when endogenous homologous sequences are present, host preferentially uses homologous sequences to repair DNA double strand break by homologous recombination [83].

$\delta$  sites are long terminal repeats (LTRs) belong to the Ty element I and II and are the most abundant LTRs in *S. cerevisiae*. There are five types of Ty elements (1-5) represented by the class II transposon (retrotransposon) that is

more commonly found in *S. cerevisiae*. It is estimated that there are about 51 retrotransposons (Ty1-5) and 251  $\delta$  sites across *S. cerevisiae* genomes [84]. Such  $\delta$  sites are appealing target sequences for ABAB expression cassette integration into *S. boulardii* chromosomes. However, much less is known about  $\delta$  sites in *S. boulardii*. Therefore, Ty1-H3 (Genbank accession no. M18706) [84] was first used as a probe to survey Ty1-2 elements in *S. boulardii* strain MYA796 (ATCC, Manassas, VA) (draft genome obtained from NCBI) to identify possible Ty1-2 elements and their  $\delta$  sites in the *S. boulardii* genome. Surprisingly, no full Ty1-2 elements were found in MYA796. A total of 57  $\delta$  sites were found; this includes 44 full  $\delta$  sites and 12 partial sites as well as a partial Ty element containing 1 full  $\delta$  site identified across all 16 chromosomes (Table 12).

TABLE 12

	Number of $\delta$ sites and their distribution on MYA796 chromosomes			
	Full $\delta$ site	Full Ty1, 2 elements	Partial $\delta$ site $60 < X < 200$ bp	Partial Ty1, 2 element with full $\delta$ site
Ch I	0	0	1	0
Ch II	0	0	0	0
Ch III	1	0	0	0
Ch IV	5	0	1	1
Ch V	2	0	1	0
Ch VI	2	0	0	0
Ch VII	8	0	1	0
Ch VIII	2	0	0	0
Ch IX	3	0	0	0
Ch X	3	0	1	0
Ch XI	0	0	0	0
Ch XII	8	0	1	0
Ch XIII	2	0	1	0
Ch XIV	1	0	0	0
Ch XV	2	0	4	0
Ch XVI	5	0	1	0
Total	44	0	12	1

Due to *S. boulardii* diploid state; there are about 114  $\delta$  sites across the *S. boulardii* genome. To allow simple multiple  $\delta$  site targeting by CRISPR, all 57  $\delta$  site sequences were compiled for multiple sequence alignment using MUSCLE to identify protospacer adjacent motif (PAM) sites that present in high numbers among the 57  $\delta$  sequences. Two PAM sites were chosen based on the highest number of  $\delta$  sequences having uniformity in protospacers as the upstream and downstream sequences. The sequences of these PAM sites are illustrated in FIG. 21 and the specific sequences are as follows:

PAM Site I (SEQ ID NO: 93)  
TGTTGGAATAAAATCAACTATCATCTACTAACTAGTATTACGTTACTA  
GTA  
TATATTATCATATAACGGTGTAGAAGATGACGCAATGATGAGAAATAG  
TCATCTAAATTACTGGAGCTGAAAC  
TC  
AT  
PAM Site II (SEQ ID NO: 94)  
AATATTATAGAATTGTGTAGAATTGCAGATTCCCTTTATGGATTCTA  
AATCCTCGAGGAGAACTTAGTATACACATACCTAATATTATAGCCT  
TAATC  
T

In both Pam Site I and Pam Site II, the sequences underscored by a dashed line correspond to the upstream homologous sequences; the sequences underscored by a single line correspond to the 20 bp protospacers; the sequences underscored by a double line correspond to the PAM sequences; the sequences underscored by a wavy line correspond to the downstream homologous sequences.

These two PAM sites, accompanied by their common upstream and downstream homologous sequences within the  $\delta$  sites, allow simple chromosomal integration of ABAB expression cassettes into *S. boulardii* genomes. ABAB integration cassettes containing homologous recombination sequences were generated by PCR using primers containing the upstream homologous sequences with the last three nucleotides removed at the 3' end and the downstream homologous sequences with the first two nucleotides removed at the 5' end and the corresponding annealing sequences needed for PCR using plasmid pCEV-G4-Km-TEF-AT-yABAB hAA6T83N-tagless as template (SEQ ID NO:90).

PCR products of the ABAB integration cassette with CRISPR plasmids contain the corresponding guide sequence (pCRI-Sb-δ1 (SEQ ID NO:91) and pCRI-Sb-δ2 (SEQ ID NO:92)) were then cotransformed with *S. boulardii* for ABAB integrations into chromosomes independently and sequentially to target PAM site I and PAM site II. The ratio of PCR product to CRISPR plasmid was found to be important for generating successful integration clones (FIG. 22A; ITG<sup>low</sup> versus ITG<sup>high</sup>). In addition, a repeat transformation of the highest ABAB secretion clone from ITG<sup>high</sup> group with the same integration cassette and CRISPR plasmid did not further improve the overall ABAB secretion of independent clones (FIG. 22A; 2<sup>nd</sup> ITG<sup>high</sup>). ABAB secretion of the highest ABAB secretion clone (C<sup>CRISPR</sup>-2) from ITG<sup>high</sup> group was then further improved by cotransforming the second set of ABAB integration cassette containing the homologous recombination sequences and its corresponding guide sequence in CRISPR plasmid targeting site II (FIG. 22A). Two highest ABAB secretion clones, C<sup>CRISPR</sup>-3 and C<sup>CRISPR</sup>-4 were selected. ABAB secretion amount and stability over time of these four representative clones are shown in FIG. 22B. A preliminary mouse CDI study was performed. However, C<sup>CRISPR</sup>-4 was found to be not better than previously M-/-:ABAB clone that showed protection in a number of mouse CDI models (FIG. 23).

While the invention has been described with reference to certain particular embodiments thereof, those skilled in the art will appreciate that various modifications may be made without departing from the spirit and scope of the invention. The scope of the appended claims is not to be limited to the specific embodiments described.

## REFERENCES

All patents and publications mentioned in this specification are indicative of the level of skill of those skilled in the art to which the invention pertains. Each cited patent and publication is incorporated herein by reference in its entirety. All of the following references have been cited in this application:

1. Cloud, J. & Kelly, C. P. Update on *Clostridium difficile* associated disease. *Curr Opin Gastroenterol* 23, 4-9 (2007).
2. Kelly, C. P. & LaMont, J. T. *Clostridium difficile*—more difficult than ever. *N Engl J Med* 359, 1932-1940 (2008).

3. Barbut, F., et al. Epidemiology of recurrences or reinfections of *Clostridium difficile*-associated diarrhea. *J Clin Microbiol* 38, 2386-2388 (2000).
4. Hassoun, A. & Ibrahim, F. Use of intravenous immunoglobulin for the treatment of severe *Clostridium difficile* colitis. *Am J Geriatr Pharmacother* 5, 48-51 (2007).
5. Shahani, L. & Koirala, J. Intravenous immunoglobulin in treatment of *Clostridium difficile* colitis. *BMJ Case Rep* 2012(2012).
6. Saito, T., et al. Evidence of intravenous immunoglobulin as a critical supportive therapy against *Clostridium difficile* toxin-mediated lethality in mice. *J Antimicrob Chemother* 66, 1096-1099 (2011).
7. Abougergi, M. S. & Kwon, J. H. Intravenous immunoglobulin for the treatment of *Clostridium difficile* infection: a review. *Dig Dis Sci* 56, 19-26 (2011).
8. Sokol, H., Maury, E., Seksik, P., Cosnes, J. & Beaugerie, L. Single immunoglobulin infusion can reverse hemodynamic failure associated with severe *Clostridium difficile* colitis. *Am J Gastroenterol* 104, 2649-2650 (2009).
9. Lowy, I., et al. Treatment with monoclonal antibodies against *Clostridium difficile* toxins. *N Engl J Med* 362, 197-205 (2010).
10. Louie, T. J., et al. Fidaxomicin versus vancomycin for *Clostridium difficile* infection. *The New England journal of medicine* 364, 422-431 (2011).
11. Rao, K. & Young, V. B. Fecal Microbiota Transplantation for the Management of *Clostridium difficile* Infection. *Infectious disease clinics of North America* 29, 109-122 (2015).
12. Vyas, D., Aekka, A. & Vyas, A. Fecal transplant policy and legislation. *World journal of gastroenterology: WJG* 21, 6-11 (2015).
13. Tonna, I. & Welsby, P. D. Pathogenesis and treatment of *Clostridium difficile* infection. *Postgrad Med J* 81, 367-369 (2005).
14. McFarland, L. V., Elmer, G. W. & Surawicz, C. M. Breaking the cycle: treatment strategies for 163 cases of recurrent *Clostridium difficile* disease. *The American journal of gastroenterology* 97, 1769-1775 (2002).
15. O'Neill, G. L., Beaman, M. H. & Riley, T. V. Relapse versus reinfection with *Clostridium difficile*. *Epidemiol Infect* 107, 627-635 (1991).
16. Wilcox, M. H., Fawley, W. N., Settle, C. D. & Davidson, A. Recurrence of symptoms in *Clostridium difficile* infection—relapse or reinfection? *J Hosp Infect* 38, 93-100 (1998).
17. Johnson, S., Adelmann, A., Clabots, C. R., Peterson, L. R. & Gerding, D. N. Recurrences of *Clostridium difficile* diarrhea not caused by the original infecting organism. *J Infect Dis* 159, 340-343 (1989).
18. Tang-Feldman, Y., Mayo, S., Silva Jr, J., Jr. & Cohen, S. H. Molecular analysis of *Clostridium difficile* strains isolated from 18 cases of recurrent *Clostridium difficile*-associated diarrhea. *J Clin Microbiol* 41, 3413-3414 (2003).
19. Johnson, S. Recurrent *Clostridium difficile* infection: causality and therapeutic approaches. *Int J Antimicrob Agents* 33 Suppl 1, S33-36 (2009).
20. Sun, X., et al. Mouse relapse model of *Clostridium difficile* infection. *Infection and immunity* 79, 2856-2864 (2011).
21. Seal, D., et al. Treatment of relapsing *Clostridium difficile* diarrhoea by administration of a non-toxigenic strain. *Eur J Clin Microbiol* 6, 51-53 (1987).
22. Kuehne, S. A., et al. The role of toxin A and toxin B in *Clostridium difficile* infection. *Nature* (2010).

23. Lyras, D., et al. Toxin B is essential for virulence of *Clostridium difficile*. *Nature* 458, 1176-1179 (2009).
24. Libby, J. M. & Wilkins, T. D. Production of antitoxins to two toxins of *Clostridium difficile* and immunological comparison of the toxins by cross-neutralization studies. *Infection and immunity* 35, 374-376 (1982).
25. Katchar, K., et al. Association between IgG2 and IgG3 subclass responses to toxin A and recurrent *Clostridium difficile*-associated disease. *Clinical gastroenterology and hepatology: the official clinical practice journal of the American Gastroenterological Association* 5, 707-713 (2007).
26. Ghose, C., et al. Transcutaneous immunization with *Clostridium difficile* toxoid A induces systemic and mucosal immune responses and toxin A-neutralizing antibodies in mice. *Infection and immunity* 75, 2826-2832 (2007).
27. Sougioultzis, S., et al. *Clostridium difficile* toxoid vaccine in recurrent *C. difficile*-associated diarrhea. *Gastroenterology* 128, 764-770 (2005).
28. Kyne, L., Warny, M., Qamar, A. & Kelly, C. P. Association between antibody response to toxin A and protection against recurrent *Clostridium difficile* diarrhoea. *Lancet* 357, 189-193 (2001).
29. Leav, B. A., et al. Serum anti-toxin B antibody correlates with protection from recurrent *Clostridium difficile* infection (CDI). *Vaccine* 28, 965-969 (2010).
30. Fernie, D. S., Thomson, R. O., Batty, I. & Walker, P. D. Active and passive immunization to protect against antibiotic associated caecitis in hamsters. *Dev Biol Stand* 53, 325-332 (1983).
31. Kim, P. H., Iaconis, J. P. & Rolfe, R. D. Immunization of adult hamsters against *Clostridium difficile*-associated ileocectitis and transfer of protection to infant hamsters. *Infection and immunity* 55, 2984-2992 (1987).
32. Lyerly, D. M., Bostwick, E. F., Binion, S. B. & Wilkins, T. D. Passive immunization of hamsters against disease caused by *Clostridium difficile* by use of bovine immunoglobulin G concentrate. *Infection and immunity* 59, 2215-2218 (1991).
33. Wang, H., et al. A chimeric toxin vaccine protects against primary and recurrent *Clostridium difficile* infection. *Infection and immunity* 80, 2678-2688 (2012).
34. Hamers-Casterman, C., et al. Naturally occurring antibodies devoid of light chains. *Nature* 363, 446-448 (1993).
35. van der Linden, R., et al. Induction of immune responses and molecular cloning of the heavy chain antibody repertoire of *Lama glama*. *J Immunol Methods* 240, 185-195 (2000).
36. Harmsen, M. M. & De Haard, H. J. Properties, production, and applications of camelid single-domain antibody fragments. *Appl Microbiol Biotechnol* 77, 13-22 (2007).
37. Arbab Ghahroudi, M., Desmyter, A., Wyns, L., Hamers, R. & Muyldermans, S. Selection and identification of single domain antibody fragments from camel heavy-chain antibodies. *FEBS Lett* 414, 521-526 (1997).
38. van der Linden, R. H., et al. Comparison of physical chemical properties of llama V<sub>H</sub>H antibody fragments and mouse monoclonal antibodies. *Biochim Biophys Acta* 1431, 37-46 (1999).
39. Dumoulin, M., et al. Single-domain antibody fragments with high conformational stability. *Protein Sci* 11, 500-515 (2002).

40. Hussack, G., Hirama, T., Ding, W., Mackenzie, R. & Tanha, J. Engineered single-domain antibodies with high protease resistance and thermal stability. *PloS one* 6, e28218 (2011).
41. Yang, Z., et al. A novel multivalent, single-domain antibody targeting TcdA and TcdB prevents fulminant *Clostridium difficile* infection in mice. *The Journal of infectious diseases* 210, 964-972 (2014).
42. McFarland, L. V. Systematic review and meta-analysis of *Saccharomyces boulardii* in adult patients. *World J Gastroenterol.* 16, 2202-2222 (2010).
43. Elmer, G. W., McFarland, L. V., Surawicz, C. M., Danko, L. & Greenberg, R. N. Behaviour of *Saccharomyces boulardii* in recurrent *Clostridium difficile* disease patients. *Aliment. Pharmacol. Ther.* 13, 1663-1668 (1999).
44. Surawicz, C. M., et al. The search for a better treatment for recurrent *Clostridium difficile* disease: use of high-dose vancomycin combined with *Saccharomyces boulardii*. *Clin. Infect. Dis.* 31, 1012-1017 (2000).
45. McFarland, L. V., et al. A randomized placebo-controlled trial of *Saccharomyces boulardii* in combination with standard antibiotics for *Clostridium difficile* disease. *Jama* 271, 1913-1918 (1994).
46. Czerucka, D., Piche, T. & Rampal, P. Review article: yeast as probiotics—*Saccharomyces boulardii*. *Alimentary pharmacology & therapeutics* 26, 767-778 (2007).
47. Sougioultzis, S., et al. *Saccharomyces boulardii* produces a soluble anti-inflammatory factor that inhibits NF-κappaB-mediated IL-8 gene expression. *Biochemical and biophysical research communications* 343, 69-76 (2006).
48. Qamar, A., et al. *Saccharomyces boulardii* stimulates intestinal immunoglobulin A immune response to *Clostridium difficile* toxin A in mice. *Infection and immunity* 69, 2762-2765 (2001).
49. Chen, X., et al. *Saccharomyces boulardii* inhibits ERK1/2 mitogen-activated protein kinase activation both in vitro and in vivo and protects against *Clostridium difficile* toxin A-induced enteritis. *The Journal of biological chemistry* 281, 24449-24454 (2006).
50. Barc, M. C., et al. Molecular analysis of the digestive microbiota in a gnotobiotic mouse model during antibiotic treatment: Influence of *Saccharomyces boulardii*. *Anaerobe* 14, 229-233 (2008).
51. Gorlani, A., de Haard, H. & Verrips, T. Expression of VHJs in *Saccharomyces cerevisiae*. *Methods in molecular biology* 911, 277-286 (2012).
52. Khatri, I., et al. Gleaning evolutionary insights from the genome sequence of a probiotic yeast *Saccharomyces boulardii*. *Gut Pathog.* 5, 30 (2013).
53. Batista, T. M., Marques, E. T., Jr., Franco, G. R. & Douradinha, B. Draft Genome Sequence of the Probiotic Yeast *Saccharomyces cerevisiae* var. *boulardii* Strain ATCC MYA-796. *Genome announcements* 2, 578-579 (2014).
54. Hamed, H., et al. Generation of a uracil auxotroph strain of the probiotic yeast *Saccharomyces boulardii* as a host for the recombinant protein production. *Avicenna journal of medical biotechnology* 5, 29-34 (2013).
55. Hudson, L. E., et al. Functional heterologous protein expression by genetically engineered probiotic yeast *Saccharomyces boulardii*. *PLoS One* 9, e112660 (2014).
56. Douradinha, B., et al. Novel insights in genetic transformation of the probiotic yeast *Saccharomyces boulardii*. *Bioengineered* 5, 21-29 (2014).

57. Fietto, J. L., et al. Molecular and physiological comparisons between *Saccharomyces cerevisiae* and *Saccharomyces boulardii*. *Canadian journal of microbiology* 50, 615-621 (2004).
58. Martins, F. S., et al. Screening of yeasts as probiotic based on capacities to colonize the gastrointestinal tract and to protect against enteropathogen challenge in mice. *The Journal of general and applied microbiology* 51, 83-92 (2005).
59. Martins, F. S., Veloso, L. C., Arantes, R. M. & Nicoli, J. R. Effects of yeast probiotic formulation on viability, revival and protection against infection with *Salmonella enterica* ssp. *enterica* serovar *Typhimurium* in mice. *Lett. Appl. Microbiol.* 49, 738-744 (2009).
60. Wu, D., Teng, D., Wang, X., Dai, C. & Wang, J. *Saccharomyces boulardii* prevention of the hepatic injury induced by *Salmonella Enteritidis* infection. *Can. J. Microbiol.* 60, 681-686 (2014).
61. Elmer, G. W. & Corthier, G. Modulation of *Clostridium difficile* induced mortality as a function of the dose and the viability of the *Saccharomyces boulardii* used as a preventative agent in gnotobiotic mice. *Can. J. Microbiol.* 37, 315-317 (1991).
62. Chen, X., et al. A mouse model of *Clostridium difficile*-associated disease. *Gastroenterology* 135, 1984-1992 (2008).
63. Pawlowski, S. W., et al. Murine model of *Clostridium difficile* infection with aged gnotobiotic C57BL/6 mice and a BI/NAP1 strain. *The Journal of infectious diseases* 202, 1708-1712 (2010).
64. McFarland, L. V. *Saccharomyces boulardii* is not *Saccharomyces cerevisiae*. *Clin. Infect. Dis.* 22, 200-201 (1996).
65. Edwards-Ingram, L., et al. Genotypic and physiological characterization of *Saccharomyces boulardii*, the probiotic strain of *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* 73, 2458-2467 (2007).
66. Panchal, C. J., Whitney, G. K. & Stewart, G. G. Susceptibility of *Saccharomyces* spp. and *Schwanniomyces* spp. to the aminoglycoside antibiotic G418. *Appl. Environ. Microbiol.* 47, 1164-1166 (1984).
67. Tsalik, E. L. & Gartenberg, M. R. Curing *Saccharomyces cerevisiae* of the 2 micron plasmid by targeted DNA damage. *Yeast* 14, 847-852 (1998).
68. Chan, K. M., Liu, Y. T., Ma, C. H., Jayaram, M. & Sau, S. The 2 micron plasmid of *Saccharomyces cerevisiae*: a miniaturized selfish genome with optimized functional competence. *Plasmid* 70, 2-17 (2013).
69. Brake, A. J., et al. Alpha-factor-directed synthesis and secretion of mature foreign proteins in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America* 81, 4642-4646 (1984).
70. Porro, D., Sauer, M., Branduardi, P. & Mattanovich, D. Recombinant protein production in yeasts. *Molecular biotechnology* 31, 245-259 (2005).
71. Chung, B. H., Nam, S. W., Kim, B. M. & Park, Y. H. Highly efficient secretion of heterologous proteins from *Saccharomyces cerevisiae* using inulinase signal peptides. *Biotechnology and bioengineering* 49, 473-479 (1996).
72. Hahn, M. S. & Chung, B. H. Secretory expression of human growth hormone in *Saccharomyces cerevisiae* using three different leader sequences. *Biotechnology and Bioprocess Engineering* 6, 306-309 (2001).
73. Gietz, R. D. & Schiestl, R. H. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.* 2, 31-4 (2007).
74. Konishi, T. & Harata, M. Improvement of the transformation efficiency of *Saccharomyces cerevisiae* by altering carbon sources in pre-culture. *Biosci. Biotechnol. Biochem.* 78, 1090-1093 (2014).
75. Gatignol, A., Baron, M. & Tiraby, G. Phleomycin resistance encoded by the ble gene from transposon Tn 5 as a dominant selectable marker in *Saccharomyces cerevisiae*. *MGG Mol. Gen. Genet.* 207, 342-348 (1987).
76. Macdonald, C. & Piper, R. C. Puromycin- and methotrexate-resistance cassettes and optimized Cre-recombinase expression plasmids for use in yeast. *Yeast* 32, 423-438 (2015).
77. Boeke, J. D., LaCroute, F. & Fink, G. R. A positive selection for mutants lacking orotidine-5'-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. *Mol Gen Genet* 197, 345-346 (1984).
78. Yang, Z. et al. Intravenous adenovirus expressing a multi-specific, single-domain antibody neutralizing TcdA and TcdB protects mice from *Clostridium difficile* infection. *Pathog. Dis.* 74, ftw078 (2016).
79. DiCarlo, J. E. et al. Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res.* 41, 4336-43 (2013).
80. Bao, Z. et al. Homology-Integrated CRISPR-Cas (Hi-CRISPR) System for One-Step Multigene Disruption in *Saccharomyces cerevisiae*. *ACS Synth. Biol.* 4, 585-594 (2015).
81. Liu, J.-J. et al. Metabolic engineering of a probiotic *Saccharomyces boulardii*. *Appl. Environ. Microbiol.* AEM.00057-16 (2016). doi:10.1128/AEM.00057-16
82. Hao, H. et al. Large fragment deletion using a CRISPR/Cas9 system in *Saccharomyces cerevisiae*. (2016). doi: 10.1016/j.ab.2016.07.008.This
83. Shi, S., Liang, Y., Zhang, M. M., Ang, E. L. & Zhao, H. A highly efficient single-step, markerless strategy for multi-copy chromosomal integration of large biochemical pathways in *Saccharomyces cerevisiae*. *Metab. Eng.* 33, 19-27 (2016).
84. Kim, J. M. et al. Transposable Elements and Genome Organization: A Comprehensive Survey of Retrotransposons Revealed by the Complete *Saccharomyces cerevisiae*. *Genome Research.* 8(5): 464-478 (1998).
85. Bao Z, Xiao H, Liang J, Zhang L, Xiong X, Sun N, Si T, Zhao H. Homology-Integrated CRISPR-Cas (Hi-CRISPR) System for One-Step Multigene Disruption in *Saccharomyces cerevisiae*. *ACS Synth Biol.* 4(5):585-94 (2015).
86. DiCarlo J E, Norville J E, Mali P, Rios X, Aach J, Church G M. Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res* 41:4336-43 (2013).
87. Liu J-J, Kong I I, Zhang G-C, Jayakody L N, Kim H, Xia P-F, Kwak S, Sung B H, Sohn J-H, Walukiewicz H E, Rao C V, Jin Y-S. Metabolic engineering of a probiotic *Saccharomyces boulardii*. *Appl Environ Microbiol.* 82(8): 2280-7 (2016).
88. Shi S, Liang Y, Zhang M M, Ang E L, Zhao H. A highly efficient single-step, markerless strategy for multi-copy chromosomal integration of large biochemical pathways in *Saccharomyces cerevisiae*. *Metab Eng* 33:19-27 (2016).

## SEQUENCE LISTING

Sequence total quantity: 117

SEQ ID NO: 1 moltype = AA length = 127  
 FEATURE Location/Qualifiers  
 REGION 1..127  
 note = Codon-optimized VHH peptide monomer 5D  
 source 1..127  
 mol\_type = protein  
 organism = synthetic construct

SEQUENCE: 1  
 QVOLVESGGG LVQPGGSLRL SCEASGFTLD YYGIGWFRQP PGKEREAVSY ISASARTILY 60  
 ADSVKGRFTI SRDNNAKNAVY LQMNSLKRED TAVYYCARRR FSASSVNRWL ADDYDVWGRG 120  
 TQAVASS 127

SEQ ID NO: 2 moltype = DNA length = 381  
 FEATURE Location/Qualifiers  
 misc\_feature 1..381  
 note = Codon-optimized VHH peptide monomer 5D  
 source 1..381  
 mol\_type = other DNA  
 organism = synthetic construct

SEQUENCE: 2  
 caagtgcaac tgggttgaatc tgggggaggc ttggtaacaac ctgggggatc cctgagactc 60  
 tcttcgcagg cttccggatt cacccatggc tactatggca tcggctggc ccggcaggcc 120  
 ccaggaaagg agccggaggc cggttcatac attatggcca gtgccccgac catactgtac 180  
 gcgactctg tgaaggagacg ctttaccatc tcttagggaca atgccaaaaa tgctgtgtac 240  
 ctgc当地atga acagcctcaa gcgggaggat accgcagtgt actactgcgc gagacggcgc 300  
 ttctccgtt ctatcggtga tagatggctg gccgacgact acgcgtgtg gggacggggc 360  
 acacagggtgg ctgtctcgag c 381

SEQ ID NO: 3 moltype = AA length = 111  
 FEATURE Location/Qualifiers  
 REGION 1..111  
 note = Codon-optimized VHH peptide monomer E3  
 source 1..111  
 mol\_type = protein  
 organism = synthetic construct

SEQUENCE: 3  
 QVQLVESGGG LVQTGGSLRL SCASSGSIAG FETVTWSRQA PGKSLQWVAS MTKTNNEIYS 60  
 DSVKGRFIIS RDNAKNTVYL QMNSLKPEDT GVYFCKGPTEL RGQQGIQTVS S 111

SEQ ID NO: 4 moltype = DNA length = 333  
 FEATURE Location/Qualifiers  
 misc\_feature 1..333  
 note = Codon-optimized VHH peptide monomer E3  
 source 1..333  
 mol\_type = other DNA  
 organism = synthetic construct

SEQUENCE: 4  
 caagttcagc tggtcgaatc cggggggcga ctgggtccaga cagggggctc cctgaggctc 60  
 tctctgtcat cttecggaag catcgccgc ttcgagaccg tgacctggc tcggcaggct 120  
 cccggaaagt ctctcgatg ggctgcttc atgactaaga ctaacaaacga gatctactct 180  
 gactcgtga aaggccgtt catcatttct agagataacg ctaaaaaacac agtgttatctg 240  
 caagatgaata gtctcaaacc tgaagacaca ggcgtgtatt tctgtaaaggc tcctgagctg 300  
 agggggccagg gcatccagg aacagtctcg agt 333

SEQ ID NO: 5 moltype = AA length = 121  
 FEATURE Location/Qualifiers  
 REGION 1..121  
 note = Codon-optimized VHH peptide monomer AA6  
 source 1..121  
 mol\_type = protein  
 organism = synthetic construct

SEQUENCE: 5  
 QLQLVETGGG LVQPGGSLRL SCAASGFTFS DYVMTWVRQA PGKGPWEIAT INTDGSTMRD 60  
 DSTKGRTFIS RDNAKNTLYL QMTSLKPEDT ALYYCARGRV ISASAIRGAV RGPQTQVTVS 120  
 S 121

SEQ ID NO: 6 moltype = DNA length = 363  
 FEATURE Location/Qualifiers  
 misc\_feature 1..363  
 note = Codon-optimized VHH peptide monomer AA6  
 source 1..363  
 mol\_type = other DNA  
 organism = synthetic construct

SEQUENCE: 6  
 caactgcagc tggtagagac agggggcggc tttagttcagc ctgggggtc tctcagactg 60  
 tcatgcgtc cctctggctt taccttcagt gactacgtga tgacatgggt ccggcaagct 120  
 ccaggaaagg ggcctgtactt attaatacag atggcagcac aatgcggac 180

-continued

---

```

gactccacaa agggggcggtt caccatttc agagacaacg ccaagaatac tctgtacctt 240
cagatgacca gtctgaaacc cgaggacact gctctgtact atttgcaag aggccgggtg 300
atctctgtttt ccgcatacag aggcgcagta aggggccctg gaacacaagt aactgtctcg 360
          agc                                     363

SEQ ID NO: 7      moltype = AA  length = 126
FEATURE           Location/Qualifiers
REGION            1..126
note = Codon-optimized VHH peptide monomer AH3
source             1..126
mol_type = protein
organism = synthetic construct

SEQUENCE: 7
QVQLVETGGL VQPQGSRLS CAASGFTLDY SSIGWFRQAP GKEREJVSCI SSSGDSTKYA 60
DSVKGRFTTS RDNAKNTVYL QMNSLPDDT AVYYCAAFRA TMCGVFPLSP YGKDDWGKGT 126
LTVSS

SEQ ID NO: 8      moltype = DNA  length = 378
FEATURE           Location/Qualifiers
misc_feature       1..378
note = Codon-optimized VHH peptide monomer AH3
source             1..378
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 8
caggtacagc tggtggagac gggggggctg gtacaacccg gcgggtcaact gagggtttcc 60
tgtggccat ctgggttac actggattt tcgtccatag ggtgggttcg gcaggctcc 120
ggcaaaagacg gtgaggggggt ctcatgtatt agtagtagt gtgatagcac aaagtaccc 180
gattccgtaa agggccgggt tacaaacctcc agggataatg ctaagaaacac cgtatatctc 240
cagatgaact ctctgaagcc cgacgatacg gcgcgtatata actgtgcggc ttccaggcgg 300
actatgtcgcg cgcgtgttccc tctgagccct tacggcaagg acgactgggg caaggggacc 360
ctgggtgaccg ttcgtcgat 378

SEQ ID NO: 9      moltype = AA  length = 15
FEATURE           Location/Qualifiers
REGION            1..15
note = Codon-optimized flexible linker 1
source             1..15
mol_type = protein
organism = synthetic construct

SEQUENCE: 9
GGGGSGGGGS GGGGS                                         15

SEQ ID NO: 10     moltype = DNA  length = 45
FEATURE           Location/Qualifiers
misc_feature       1..45
note = Codon-optimized flexible linker 1
source             1..45
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 10
ggcggtggag ggtctgggtgg gggaggctca ggggggtggag gcagc                                         45

SEQ ID NO: 11      moltype = AA  length = 16
FEATURE           Location/Qualifiers
REGION            1..16
note = Codon-optimized flexible linker 2
source             1..16
mol_type = protein
organism = synthetic construct

SEQUENCE: 11
GGGSGGGGGG GSAGGGS                                         16

SEQ ID NO: 12      moltype = DNA  length = 48
FEATURE           Location/Qualifiers
misc_feature       1..48
note = Codon-optimized flexible linker 2
source             1..48
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 12
ggtggcggaa gcggagggggg cagcgggggt gggagcggtg ggggcagc                                         48

SEQ ID NO: 13      moltype = AA  length = 15
FEATURE           Location/Qualifiers
REGION            1..15
note = Codon-optimized flexible linker 3
source             1..15
mol_type = protein

```

-continued

---

organism = synthetic construct

SEQUENCE: 13  
GGGGSGGGGS GGGGS 15

SEQ ID NO: 14 moltype = DNA length = 45  
FEATURE Location/Qualifiers  
misc\_feature 1..45  
note = Codon-optimized flexible linker 3  
source 1..45  
mol\_type = other DNA  
organism = synthetic construct

SEQUENCE: 14  
gggggaggcg gttcaggccg tggggatct ggccgggggtg gatcc 45

SEQ ID NO: 15 moltype = AA length = 269  
FEATURE Location/Qualifiers  
REGION 1..269  
note = AH3-5D heterodimer  
source 1..269  
mol\_type = protein  
organism = synthetic construct

SEQUENCE: 15  
QVQLVETGGG LVQPGGSLRL SCAASGFTLD YSSIGWFRQA PGKEREGRVSC ISSSGDSTKY 60  
ADSVKGRFTT SRDNNAKNTVY LQMNSLPDD TAVYYCAAFR ATMCGVFPLS PYGKDDWGKG 120  
TLTVTSSGGG GSGGGGGGGG GSQVQLVESG GGLVQPGGSL RLSCEASGFT LDYYGIGWFR 180  
QPPGKEREAV SYISASARTI LYADSVKGRF TISRDNAKNA VYLQMNLSLR EDTAVYYCAR 240  
RPPSASSVNR WLADDYDVWG RGTQVAVSS 269

SEQ ID NO: 16 moltype = DNA length = 807  
FEATURE Location/Qualifiers  
misc\_feature 1..807  
note = AH3-5D heterodimer  
source 1..807  
mol\_type = other DNA  
organism = synthetic construct

SEQUENCE: 16  
caggtacacg tggtgagac gggggggggg ctggtacaac caggcgggtc actgaggctt 60  
tcctgtgcgc catctgggtt cacatggat tattcgttca tagggtgtt tcggcaggct 120  
cctggcaag agcgtgaggg ggtctcatgt attatgtat gtggatgtat cacaatgtac 180  
gccgattccg taaaggccg gtttacaacc tccaggata atgtaaaggaa caccgtatata 240  
ctccagatga actctctgaa gcccgcacat acggccgtat attactgtgc ggcttcagg 300  
ggcactatgt cccttgcgcg ccttacggca aggacgactg gggcaagggg 360  
acccttggta cggtatcttc agggcggttgg ggggttggggctc aggggggttgg 420  
ggcagccagg tgcaactggg tgaatctggg ggaggcttgg tacaacctgg gggatccctg 480  
agactctttt cggaggccctc cggattcacc ttggactact atggcatcgg ctggttccgc 540  
cagccccccac ggaaggagcg ggaggccgtt tcatacattta gtgcgcgtgc cgggaccata 600  
ctgtacgcag actctgtgaa gggacgcgtt accatctcta gggacaatgc caaaaatgtct 660  
gtgtacctgc aaatgaacag cctcaagcg gaggataccg cagtgtacta ctgcgcgaga 720  
cggcgcttcccg cctgttcttag cgtgaataga tggctggccg acgactacga cgtgtgggaa 780  
cggggcacac aggtggctgt ctcgagc 807

SEQ ID NO: 17 moltype = AA length = 247  
FEATURE Location/Qualifiers  
REGION 1..247  
note = AA6-E3 heterodimer  
source 1..247  
mol\_type = protein  
organism = synthetic construct

SEQUENCE: 17  
QLQLVETGGG LVQPGGSLRL SCAASGFTFS DYVMTWVRQA PGKGPEWIAT INTDGSTMRD 60  
DSTKGRFTIS RDNAKNTLYL QMTSLKPDT ALYYCARGRV ISASAIRGAV RGPGTQVTVS 120  
SGGGGSGGGG SGGGGSQVQL VESGGGLVQT GGSLRLSCAS SGSIAGFETV TWSRQAPGKS 180  
LQWVASMTKT NNEIYSDSVK GRFIISRDNA KNTVYLMNLS LRPEDTGVYF CKGPELRGQG 240  
IQTVSS 247

SEQ ID NO: 18 moltype = DNA length = 741  
FEATURE Location/Qualifiers  
misc\_feature 1..741  
note = AA6-E3 heterodimer  
source 1..741  
mol\_type = other DNA  
organism = synthetic construct

SEQUENCE: 18  
caactgcacg tggtagagac agggggccgc ttagttcagc ctggagggtc ttcgcactg 60  
tcatgcgcgtc cctctgggtt tacattcgtt gactacgttga tgacatgggt ccggcaagct 120  
ccagggaaagg ggcctgtatgt gatcgctact attaatacag atggcagcac aatgcggac 180  
gactccacaa agggggcggtt caccatttcc agagacaacg ccaagaatac tctgtacctt 240  
cagatgacca gtctgaaacc cgaggacact gctctgtact attgtgcaag agggccgggtg 300  
atctctgttcccg cctgtatcg agggccctcg gaacacaggtt aaccgtttca 360

-continued

tccgggggag	cgcggttcagg	cggtggggga	tctggcgggg	gtggatccca	agttcagctg	420
gtcgaatccg	ggggcgact	ggtccagaca	gggggctcc	tgaggcttc	ctgtcatct	480
tccggaaagca	tcgccccgtt	cgagaccgtg	acctggctc	gccaggctcc	cggaaagtct	540
ctgcagtggg	tcgcttccat	gactaagact	aacaacgaga	tctactctga	ctcagtgaaa	600
ggccgcgttca	tcatttctag	agataacgct	aaaacacag	tgtatctgca	gatgaatagt	660
ctcaaaccgt	aagacacagg	cgtgttatttc	tgttaagggtc	ctgagctgag	ggccagggc	720
atccaggtaa	cagtctcgag	t				741

SEQ ID NO: 19	moltype = AA	length = 532				
FEATURE	Location/Qualifiers					
REGION	1..532					
	note = ABAB binding agent					
source	1..532					
	mol_type = protein					
	organism = synthetic construct					
SEQUENCE: 19						
QVQLVETGGG	LVQPGGSLRL	SCAASGFTLD	YSSIGWFRQA	PGKEREVGSC	ISSSGDSTKY	60
ADSVVKGRFTT	SRDNNAKNTVY	LQMNSLKPDD	TAVYYCAAFR	ATMCGVFPLS	PYGKDDWGKG	120
TLVTVSSGGG	GSGGGGSGGG	GSQVOLVESP	GLGVQPGGSL	RLSCEASGFT	LDYYGIGWFR	180
QPPGKEREAV	SYISASARTI	LYADSVKGRF	TISRDNAKNA	VYLQMNSLKR	EDTAVYYCAR	240
RPFPSASSVNR	WLADDYDVWG	RGTQAVVSSG	GGSGGGSGGG	SGGGSQLQLV	ETGGGLVQPG	300
GSLRLSCAAS	GFTFSDYVMT	WVRQAPGKGP	EWIATINTDG	STMRDDSTKG	RFTISRDNNAK	360
NTLYLQMTSL	KPEDTALYYC	ARGVVISASA	IRGAVERGPGT	QTVTVSSGGG	SGGGGGGGG	420
SQVQLVESGG	GLVQTGGSLR	LSCASSGSIA	GFETVTWSRQ	APGKSLQWA	SMTKTNNEIY	480
SDSVVKGRFII	SRDNNAKNTVY	LQMNSLKPED	TGVYFCKGPE	LRGQGIQVTV	SS	532

SEQ ID NO: 20	moltype = DNA	length = 1596				
FEATURE	Location/Qualifiers					
misc_feature	1..1596					
	note = ABAB binding agent					
source	1..1596					
	mol_type = other DNA					
	organism = synthetic construct					
SEQUENCE: 20						
caggtagacg	tgggtggagac	gggggggaggg	ctggtacaaac	caggcgggtc	actgaggctt	60
tccctgtcgc	catctgggtt	cacactggat	tattcgtcca	tagggtgttt	tcggcaggct	120
cctggcaaaag	agcgtgaggg	ggtctcatgt	attagtagta	gtggtgatag	cacaaagtac	180
gcccgttccg	taaaggcccg	gtttacaacc	tccagggtata	atgctaagaa	caccgtatata	240
ctccagatga	actctctgaa	gcccgcacat	acggccgtat	attactgtgc	ggctttcagg	300
gcccgtatgt	ggggcgtgtt	ccctctgago	ccttacggca	aggacgactg	ggcaagggg	360
accctgttga	ccgtatccctc	aggcggtgga	gggtctgtt	ggggaggctc	aggggggttgg	420
ggcagccagg	tgcaacttgtt	tgaatctggg	ggaggcttgc	tacaacctgg	gggatccctg	480
agactctctt	gcccggccctc	cgaggatcacc	ttggactact	atggcatcg	ctggttccgc	540
cagccccccag	ggaaggagcg	ggaggccgtt	tcatacattta	gtgcaagtgc	ccggaccata	600
ctgtacgcag	actctgtgaa	gggacgtttt	accatctcta	gggacaatgc	caaaaatgt	660
gtgtacactgc	aatgaaacag	cctcaagcgg	gaggataccat	cgatgtacta	ctgcgcgaga	720
cggcgcttct	ccgcttcttag	cgtaataga	tggctggccc	acgactacga	cgtgtgggg	780
cggggcacac	aggtggctgt	gtcttccgtt	ggcggaaacg	gagggggcag	cggggggttgg	840
agcggtgggg	gcagccaaact	gcagctggta	gagacaggggg	gcccgttagt	tcagctgttgg	900
gggtctctca	gactgtccatg	cgctgcctt	ggctttactt	tcagtacta	cgtgtatgaca	960
tgggtccccc	aaagctccagg	gaagggcct	gagtggatgc	ctacttaa	tacagatggc	1020
agcacaatgc	ggggacactc	cacaaggggg	cggttccacca	tttccagaga	caacgc当地	1080
aataactctgt	accttcagat	gaccagtctg	aaaccgcagg	acactgtct	gtactattgt	1140
gcaagaggcc	gggtgtatctc	tgcttccgt	atcagaggcg	cagtaagggg	cccttgaaaca	1200
caggttaaccg	tttcatccgg	gggggggggt	tcaggcgggt	ggggatctgg	cggggggttgg	1260
tcccaactgtc	actctgtcga	atccggggc	qgactgttc	agacaggggg	ctccctgtgg	1320
ctctctgtg	catctccgg	aagcatcgcc	ggtttcgaga	ccgtgcaccc	gtctcgccag	1380
gtctccggga	agtctctgca	gtgggtcgct	tccatgacta	agactaaacaa	cgagatctac	1440
tctgactcag	tgaaaggccg	cttcatcatt	tctagagata	acgtaaaaaa	cacagtgtt	1500
ctgcagatga	atagtccaa	acctgaaac	acaggcgtgt	atttctgtaa	gggtcctgtag	1560
ctgagggggc	agggcatcca	ggtaacagtc	tcgagt			1596

SEQ ID NO: 21	moltype = AA	length = 12
FEATURE	Location/Qualifiers	
REGION	1..12	
	note = Albumin-binding peptide	
source	1..12	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 21		
DICLPRWGCL WD		12

SEQ ID NO: 22	moltype = AA	length = 761
FEATURE	Location/Qualifiers	
REGION	1..761	
	note = ABAB-Fc binding agent	
source	1..761	
	mol_type = protein	
	organism = synthetic construct	

-continued

SEQUENCE: 22  
 QVQLVETGGG LVQPGGSSLRL SCAASGFTLD YSSIGWFRQ PGKEREVGSC ISSSGDSTKY 60  
 ADSVKGRFTT SRDNNAKNTVY LQMNLSLPDD TAVYYCAAFL ATMCGVFPLS PYGKDDWGKG 120  
 TLTVTSSGGG GSGGGGSGGG GSQVQLVESG GGLVQPGGSL RLSCEASGFT LDYYGIGWFR 180  
 QPPGKEREAV SYISASARTI TISRDNNAKNTVY LQMNLSLPDD TAVYYCAAFL ATMCGVFPLS PYGKDDWGKG 240  
 RRFSASSVNR WLADDYDVWG RGTQAVVSSG GGSGGGSGGG SGGSQSQLQV ETGGGLVQPG 300  
 GSRLRSLCAAS GFTFSDYVMT WVRQAPGKGP EWIAITINTDG STMRDDSTKG RFTISRDNAK 360  
 NTLYLQMITS KPEDTALYYC ARGRVISASA IRGAVERPGT QVTVSSGGGG SGGGGGGGGG 420  
 SQVQLVESGG GLVQTCGSSLR LSCASSGTSIA GFETVTVWSRQ APGKSQIOWVA SMTKTNEIY 480  
 SDSVKGRFLII SRDNNAKNTVY LQMNLSLPDD TGTVYFCKGP ETDATVYYCAR 540  
 PPCPAPELII GPSVFLFPPI PKDTLMISRT PEVTCVVVDV SHEDPEVKFN WYVDGVEVHN 600  
 AKTKPREEQY NSTYRVVSVL TVLHQDWLNG KEYKCKVSNK ALPAPIEKTI SKAKGQPREP 660  
 QVYTLPPSRE EMTKNQVSLL CLVKGFYPSD IAVEWESNQ PENNYKTPP VLDSDGSFFL 720  
 YSKLTVDKSR WQQGNVFSCS VMHEALHNHY TQKSLSLSPG K 761

SEQ ID NO: 23                    moltype = DNA length = 2286  
 FEATURE                        Location/Qualifiers  
 misc\_feature                1..2286  
 note = ABAB-Fc binding agent  
 source                        1..2286  
 mol\_type = other DNA  
 organism = synthetic construct  
 SEQUENCE: 23  
 caggtagacgc tggtggagac ggggggaggg ctggtacaac caggcggtc actgaggctt 60  
 tctctgtccgc catctgggtt cacatggat tattcgtcca tagggtgggt tcggcaggct 120  
 cctggcaaaag acgcgtgggg ggtctcatgt attagtagt gtggtgatag cacaaggatc 180  
 gcccattccg taaaaggccg gtttacaacc tccaggata atgctaagaa caccgttat 240  
 ctccagatga actctctgaa gccegacat acggccgtat attactgtgc ggcttcagg 300  
 gggactatgt gggcgctgtt ccctctgtggc ctttacggca aggacgactg gggcaagggg 360  
 accctggta cggatccctc aggcgggtgg gggctctggg gggggggctc aggggggttgg 420  
 ggcagccagg tgcaacttgtt tgaaatcccggg gggggcttgg tacaacctgg gggatccctg 480  
 agactcttctt gggaggccctc cggattcacc ttgactact atggcatccg ctgggtccgc 540  
 cagccccccagg ggaaggagcg ggaggccgtt tcatacatatta gtggcagtgc ccggaccata 600  
 ctgtacgcgactctgtggaa gggacatccatc accatcttca gggacaatgc caaaaatgtct 660  
 gtgtacatgc aaatgaacac ccttaaaggcg gaggataccg cagtgtacta ctgcgcgaga 720  
 cggcgcttctt cggcttcttag cgttaataga tggctggccg acgactacga cgtgtgggaa 780  
 cggggcacac aggtggctgt gtcttcgggtt ggcggaaacg gggggggcag cgggggttggg 840  
 agccgggtgggg gcaagccaaact cggatcttgcg gggccgttgg gggccgttgg 900  
 ggggtcttcata cactgtcatg cggctcttgcg ggtttaacctc tggacta cgtgtatgaca 960  
 tgggtccggc aagctccagg gaaggggctt gggatggatcg ctactattaa tacagatggc 1020  
 agcacaatgc gggacgactc cacaaggggg cggttccacca tttccagaga caacgccaag 1080  
 aataactctgtt accttcatgcg gaccatgtgg aaaccccgagg acactgtctt gtactatgt 1140  
 gcaagaggccc ggggtatctc tgcttcgggtt atcagaggcg cagaaggggg ccctggaaaca 1200  
 caagtaaccg tttcatccgg gggaggccgtt tcaggccgtt gggggatctgg cgggggttgg 1260  
 tcccaagttc agctggtcga atccggggcgc ggactggtcc agacaggggg ctccctgagg 1320  
 ctctccgtgc catctccggc aaggatccgcg ggcttcggatc cggtaatgcgtt gtctccgcag 1380  
 gctcccccggaa agtcttcgcg gtgggtcgat tccatgacta agactaaacaa cgagatctac 1440  
 tcttgacttcgat tgaaaggcccg cttcatccgtt tcttagatata acgctaaaaa cacagtgtat 1500  
 ctgcagatga atagttctca accttgcgatc acaggccgtt atttctgtaa ggggtcttgag 1560  
 ctggggggccggc aaggccatccgcg ggttacatccgcg tccggccgtt ccggacaaaac tcacacatgc 1620  
 ceaccgtgcc cggccatgcg actctggggg ggaccgtcgat tttccctttt ccccccaaaa 1680  
 cccaaaggaca cccatcatgtat cttccggacc cttcgaggatca catcgctgtt ggtggacgtt 1740  
 agccacacggc accctcgaggat caagtccaaat tggatcgatgg acggccgtt ggtgcataat 1800  
 gccaagacaaacccgcgggg gggccgttgg aacccggccgtt accctgttggt cagccgtctc 1860  
 acccttcgtgc accaggactg gcttgcgttgg aaggatgttca agtgcacatgc ctccaaacaaa 1920  
 gcccctcccg ccccatcgat gaaaaccatc tccaaaggccca aaggccggcc ccggacacca 1980  
 cagggttaca ccctggccc atccggggag gagatgacca agaaccaggat cagccgttgc 2040  
 tgccctggatca aaggccgttca tcccgccgcg atccggccgtt ggtggggaggg caatggccag 2100  
 ccggagaaacttccca caccatccgcg tggccgttggact ccggccgttgc ctccatccgtt 2160  
 tatagcaacgc tcaccgttgcg acaaggccgcg gggacgttccctt ctcatgttgc 2220  
 gtgtatgcatg aggtcttcgcg acaaccatc acggccgttgc ggttcccttgcgttgc 2280  
 aaatgaacac 2286

SEQ ID NO: 24                    moltype = AA length = 356  
 FEATURE                        Location/Qualifiers  
 REGION                        1..356  
 note = 5D-Fc binding agent  
 source                        1..356  
 mol\_type = protein  
 organism = synthetic construct  
 SEQUENCE: 24  
 QVQLVESGGG LVQPGGSSLRL SCEASGFTLD YYGIGWFRQ PGKEREAVSY ISASARTILY 60  
 ADSVKGRFTI SRDNNAKNAVY LQMNLSLKRED TAVYYCARRR FSASSVNRWL ADDYDVWGRG 120  
 TQVAVSSGSD KTHTCPPCPA PELLPGPSVLFPPKPKDML MISRTPEVTC VVVDVSHEDP 180  
 EVKFNWYWDG VEVHNAKTKP REEQYNSTYR VVSVLTVLHQ DWLNGKEYKC KVSNKALPAP 240  
 IEKTISKAKG QPREPQVYTL PPSREEMTKN QVSLTCLVKG FYPYPSDIATEWV ESNGOOPENNY 300  
 KTTTPVLDSD GSFFFLYSKLT VDKSRWQOQN VFSCSVMHEA LHNHYTQKSL SLSPGK 356

SEQ ID NO: 25                    moltype = DNA length = 1071

-continued

---

FEATURE	Location/Qualifiers
misc_feature	1..1071 note = 5D-Fc binding agent
source	1..1071 mol_type = other DNA organism = synthetic construct
SEQUENCE: 25	cagggtgcaac tgggttgaatc tggggggaggc ttgggtacaac ctggggggatc cctgagactc 60 tcttgcagg ctcggcgatt cacccgttgc tactatggca tcggctgggt ccggccggcc 120 ccaggaaagg aegggggaggc cggttcataad attagtgcaca gtggccggac catactgtac 180 gcagactctg tgaaggggacg ctttaccatc tcttagggaca atgccaaaaa tgctgtgtac 240 ctgcaaatgaa acagcctcaa gccccgggat accgcgtgt actactgcgc gagacggcgc 300 tttcccggtt ctatcggtttagatggctg gcccggact acggcgtgt gggacggggc 360 acacagggtgg ctgttgcgtgg cggatccgc aaaaactcaca catggccacc gtgcggcaga 420 cctgaactcc tggggggggacc gtcgtcttc ctcttccccca caaaacccaa ggacacctc 480 atgatctccc ggacccctga ggtcacatgc gtgggtggg acgtgagcca cgaagaccc 540 gagggtcaatgtt ctaatcggttgc cgtggacggc gtgggggttgc ataatggca gacaaaggcg 600 cggggggggc agtacaacagc cacgttccgt gtggcgttgcg teectcacgt cctgcaccc 660 gactggctgttgc atggcaagggat gtaaaatgttgc aagggttccca acaaaggccct cccagcccc 720 atcgagaaaa ccatctccaa agccaaagggg cagccccggg aaccacagggt gtacacctg 780 ccccccatccc gggggggatgg gaccaaaaggc cagggtcggcc tgacccctg ggtcaaaaggc 840 ttctatccatc cgcacatcgc cgtgggtgg gagacatgttgc ggcacccggaa gaacaactac 900 aagaccacgc ctcccggttgc gggacttgcgat ggttcccttc tctctatag caagctcacc 960 gtggacaaga gcaagggtggca gcaggggaaat gtcttctcat gtccgtgtat gcatgaggct 1020 ctgcacaacc actacacgcgca gaagggccctc tccctgttgc cgggttaatg a 1071
SEQ ID NO: 26	moltype = AA length = 340
FEATURE	Location/Qualifiers
REGION	1..340 note = E3-Fc binding agent
source	1..340 mol_type = protein organism = synthetic construct
SEQUENCE: 26	QVLVESGGG LVQTGGSLRL SCASSGSIAG FETVTWSRQA PGKSLQVVAS MTKTNNIEIYS 60 DSVKGRFIIS RDNAKNTVYL QMNSLKPEDT GYVFCKGPTEL RGQQIQVTVS SGSDKTHTCP 120 PCPAPELLGG PSVFLFPPKP KDTLMISRTP EVTCVVVDVS HEDPEVKFNW YVDGVVEVHNA 180 KTKPREEQYN STYRVVSVLTL VLVGDWLNGK EYKCKVSNKA LPAPIEKTTIS KAKGQPREPQ 240 VYTLPPSRREE MTKNQVSLTC LVKGFYPSDI AVEWESNGQP ENNYKTTPPV LDSDGSFLFLY 300 SKLTVDKSRW QQGNVFSCSV MHEALHNHYT QKSLSLSPKG 340
SEQ ID NO: 27	moltype = DNA length = 1023
FEATURE	Location/Qualifiers
misc_feature	1..1023 note = E3-Fc binding agent
source	1..1023 mol_type = other DNA organism = synthetic construct
SEQUENCE: 27	caagttcaggc tgggtcgaatc cggggggcggg ctgggtccagg cagggggctc cctgaggctc 60 tcttgtcgtt cttccggaaatc cttccggaaatc cttccggaaatc cttccggaaatc cttccggaaatc 120 ccccggaaatc ctctgcgtgtgg cttccgttcc atgactaaga ctaacaaacgc gatctactct 180 gactcgtgttgc aaggccgtttgc catatccatc agagataacgc ctaaaaaacac acgttatctg 240 caagatgttgc gtctcaaaatc tgaagacaca ggcgtgttattc tctgttaaggg tccttgaggctg 300 agggggccagg gcatccaggatc aacatgttgc acgggatccgc aacaaactca cacatggcca 360 ccctgtccccc caccgttgcactt cctggggggatc cccgtcgttgc tctcttcccccc 420 aaggacaccc tcatgtatctc cgggacccctt gaggttccatc ggcgtgttgc ggcgtgttgc 480 caccgttgcactt cttccgttcc atcgttgcactt cttccgttcc atcgttgcactt cttccgttcc 540 aagacaaatc cccgttgcactt cttccgttcc atcgttgcactt cttccgttcc atcgttgcactt cttccgttcc 600 gtctgttgcactt cttccgttcc atcgttgcactt cttccgttcc atcgttgcactt cttccgttcc atcgttgcactt cttccgttcc 660 cttcccgccccc cccgttgcactt cttccgttcc atcgttgcactt cttccgttcc atcgttgcactt cttccgttcc atcgttgcactt cttccgttcc 720 gttgtacaccctt cttccgttcc atcgttgcactt cttccgttcc atcgttgcactt cttccgttcc atcgttgcactt cttccgttcc atcgttgcactt cttccgttcc 780 cttgttgcactt cttccgttcc atcgttgcactt cttccgttcc atcgttgcactt cttccgttcc atcgttgcactt cttccgttcc atcgttgcactt cttccgttcc 840 gagaacaaatc cccgttgcactt cttccgttcc atcgttgcactt cttccgttcc atcgttgcactt cttccgttcc atcgttgcactt cttccgttcc 900 agcaaggttgcactt cttccgttcc atcgttgcactt cttccgttcc atcgttgcactt cttccgttcc atcgttgcactt cttccgttcc atcgttgcactt cttccgttcc 960 atgcgttgcactt cttccgttcc atcgttgcactt cttccgttcc atcgttgcactt cttccgttcc atcgttgcactt cttccgttcc atcgttgcactt cttccgttcc 1020 tga 1023
SEQ ID NO: 28	moltype = AA length = 350
FEATURE	Location/Qualifiers
REGION	1..350 note = AA6-Fc binding agent
source	1..350 mol_type = protein organism = synthetic construct
SEQUENCE: 28	QLQLVETGGG LVQPGGSLRL SCAASGFTFS DYVMTWVRQA PGKGPWEIAT INTDGSTMRD 60 DSTKGRFTIS RDNAKNTLYL QMNTSLKPEDT ALYYCARGRV ISASAIRGAV RGPGTQVTVS 120 SGSDKTHTCP PCPAPELLGG PSVFLFPPKP KDTLMISRTP EVTCVVVDVS HEDPEVKFNW 180

---

## US 12,384,833 B2

67

68

-continued

YVDGVEVHNA KTKPREEQYN STYR VVS VLT VLHQDWLNGK EYKCKVSNKA LPAPIEKTI S 240  
 KAKGQPREGQ VYTLPSSREE MTKNQVSLTC LVKGFYPSDI AVEWESENQGP ENNYKTTPV 300  
 LSDGSFFLY SKLTVDKSRW QQGNVFSCSV MHEALHNHYT QKSLSLSPGK 350

SEQ ID NO: 29 moltype = DNA length = 1053  
 FEATURE Location/Qualifiers  
 misc\_feature 1..1053  
 note = AA6-Fc binding agent  
 source 1..1053  
 mol\_type = other DNA  
 organism = synthetic construct

SEQUENCE: 29  
 caactgcagc tggttagagac aggggccgc ttatgttcagg ctggagggtc tctcagactg 60  
 tcatgcgtc cctctggc ttacccgt gactacgtg tgacatgggt ccgcgaact 120  
 ccaggaaagg ggcctgagtg gatecgctact attaatacac atggcagcac aatgcggac 180  
 gactccacaa aggggccgtt caccattcc agagacaacg ccaagaatac tctgtacctt 240  
 caagatgacca gctctaaacc cgagacact gctctgtact attgtcaag aggccgggtg 300  
 atctctgtt cccgtatcag aggccgacta agggccgtt gaacacaagt aactgtctcg 360  
 acgcggatcg aaaaaactca cacatggcca ccgtgccccag cacctgacta cctgggggga 420  
 ccgtcgtctt cccttccc cccaaaccc aaggacacc tcatgtatcc ccggaccct 480  
 gaggtcacat ggctgggtt ggacgtgago caccggacact tgagggtcaa gttcaactgg 540  
 tacgtggacg cgctggaggt gcataatgc aagacaacg cgcggggaga gcagtacaac 600  
 agcacgtacc gtgtggtcag cgtctcacc gtcctgaccaggactggtt gaatggcaag 660  
 gagtacaagt gcaaggctc caacaaagcc ctcccaccc ccatcgagaa aaccatctcc 720  
 aaaggccaaag ggccaggcccc agaaccacac gtgtacaccc tgccccccatcc ccggggaggag 780  
 atgaccaaga accaggctc cctgtatcc ctggtcaaaag gtttctatcc cagcgacatc 840  
 gccgtggagt gggagacaa tggccagccg gagaacaact acaagaccac gcctccctgt 900  
 ctggactccg acggctctt cttectctat agcaagctc ccgtggacaa gaggcagggtt 960  
 cagcaggaaa acgtttctc atgtccgtt atgcattggg ctctgcacaa ccactacacg 1020  
 cagaagaccc tttccctgtt ccgggtaaa tga 1053

SEQ ID NO: 30 moltype = AA length = 356  
 FEATURE Location/Qualifiers  
 REGION 1..356  
 note = AH3-Fc binding agent  
 source 1..356  
 mol\_type = protein  
 organism = synthetic construct

SEQUENCE: 30  
 QVQLVETGGG LVQPGGSLRL SCAASGFTLD YSSIGWFRQA PGKEREGRVSC ISSSGDSTKY 60  
 ADSVVKGRFTT SRDNNAKNTVY LQMNSLKPDD TAVYYCAAFR ATMCGVFPLS PYGKDDWGKG 120  
 TLTVTSSGSD KTHTCPCPVY PELLPPGPSVF LFPPPKPDTL MISRTPEVTC VVVDVSHEDP 180  
 EVKFNWVYDG VEVHNAKTPK REEQYNSTYR VVSVLTVLHQ DWLNGKEYKC KVSNKALPAP 240  
 IEKTISAKAG QPREPOVYTL PPSREEMTKN QVSLTCLVKG FYPSDIAVIEW ESNGOPENNY 300  
 KTTPPVLDSD GSFFFLYSKLT VDKSRWQQGN VFSCSVMHEA LHNHYTQKSL SLSPGK 356

SEQ ID NO: 31 moltype = DNA length = 1071  
 FEATURE Location/Qualifiers  
 misc\_feature 1..1071  
 note = AH3-Fc binding agent  
 source 1..1071  
 mol\_type = other DNA  
 organism = synthetic construct

SEQUENCE: 31  
 caggtagc tggtggagac gggggggggg ctggtagacaa caggccgggtc actggggctt 60  
 tctctgtccg catctgggtt cacactggat tatttcgtcc taggggtgtt tcggcaggct 120  
 cctggcaag agcgtggggg ggtctcatgt attagtagta gtgggtatag cacaaggatc 180  
 gccgatccg taaaggccg gtttacaacc tccagggtata atgctaaggaa caccgttat 240  
 ctccagatga actctctgaa gcccacgtt acggccgtat attachgtgc ggcttcagg 300  
 gcgactatgt ggccgtgtt ccctctgago ctttacggca aggacactg gggcaagggg 360  
 accctggta cctgtctcgag cggatccgac aaaactcaca catggccacc gtggccagca 420  
 ctctgaactcc tggggggacc gtcaacttcc ctttccccca caaaacccaa ggacaccctc 480  
 atgatctccc ggacccctga ggtcacatgc gtgggtgtt ggccgttgc cgaagaccc 540  
 gaggtcaagt tcaactggta ctggacccggtt gggagggttc ataatgcacaa gacaaggccg 600  
 cggggaggac agtacaacag cacgtaccgt gtggtcagcg tcttcaccgt cctgcaccag 660  
 gactggctga atggcaagga gtacaagtgc aaggcttccaa acaaaggccct cccagccccc 720  
 atcgaaaaaa ccattcccaaa agccaaaggag cagcccccaag aaccacaggt gtacaccctg 780  
 ccccccattccccc gggaggagat gccaaggaaac cagggtcagcc tgacccgtt ggtcaaaaggc 840  
 ttctatccca ggcacatcgc ctggaggtttt gggaccaatg ggcaggccggaa gacaactac 900  
 aagaccacgc ctccctgtt ggactccgac ggcttccctt tcttcatacg caagtcacc 960  
 gtggacaaga gcaagggtggca gcaggggaaac gtcttctcat gtcctgtat gcatgaggct 1020  
 ctgcacaacc actacacgcgca gaagacccctc tccctgtctc cgggtaaat a 1071

SEQ ID NO: 32 moltype = AA length = 498  
 FEATURE Location/Qualifiers  
 REGION 1..498  
 note = AH3-5D-Fc binding agent  
 source 1..498  
 mol\_type = protein

-continued

organism = synthetic construct

SEQUENCE: 32

```
QVQLVETGGG LVQPGGSLRL SCAASGFTLD YSSIGWFRQA PGKEREGRVSC ISSSGDSTKY 60
ADSVKGRFTT SRDNNAKNTVY LQMNSLPKDD TAVYYCAAFR ATMCGVFPLS PYGKDDWGKG 120
TLTVSSGGG GSAGGGGGGG GSQVQLVESG GGLVQPGSSL RLSCEASGFT LDYYGIGWFR 180
QPPGKEREAV SYISASARTI LYADSVKGRF TISRDNAKNA VYLQMNLSKR EDTAVYYCAR 240
RFRSASSVNR WLADDYDVWG RGTQAVVSSG SDKHTCPCP PAPELLGGPS VFLFPKPKD 300
TLMISRTPEV TCVVVDSHE DPEVKFNWYV DGVEVHNNAKT KPREEQYNST YRVVSLTVL 360
HQDWLNKEY KCKVSNKALP APIKTISKA KGOPREPQVY TLPPSRREMT KNQVSLTCLV 420
KGFYPSDIAV EWESNGOPEN NYKTTTPVLD SDGSFFLYSK LTVDKSRWQQ GNVFSCVMH 480
EALHNHYTQK SLSLSPKG 498
```

SEQ ID NO: 33 moltype = DNA length = 1497

FEATURE Location/Qualifiers

misc\_feature 1..1497 note = AH3-5D-Fc binding agent

source 1..1497 mol\_type = other DNA

organism = synthetic construct

SEQUENCE: 33

```
caggtacagc tggtgagac gggggggggg ctggtacaac caggcggtc actgaggctt 60
tcctgtccg catctgggtt cacactggat tattcgtcca taggggttt tcggcaggct 120
cctggcaag agcgtgaggg ggtctcatgt attagtagta gtgggtatag cacaaggatc 180
ggcgattccg taaagggccg gtttacaacc tccagggata atgctaagaa caccgtatat 240
ctccagatga actctctgaa gcccacatg acggccgttat attactgtgc ggcttcagg 300
gcgactatgt gggcggttt ccctatgago ccttacggca aggacgactg gggcaagggg 360
accctgttga cccgtatctc agggcggtt gggctctgtt gggggggctc aggggggttgg 420
ggcagccagg tgcaactggg tgaatctggg ggaggcttgg tacaacctgg gggatccctg 480
agactcttt gggggccctc cggatttacc ttggactat atggcatccg ctggttccgc 540
cagccccccg ggaaggaggcg ggaggccgtt tcatacattha gtggcagtgc cccgaccata 600
ctgtacgcaactctgtgaa gggacgtttt accatctcta gggacaatgc caaaaatgt 660
gtgtactctgc aaatgaacag cctcaagcg gaggataccg cagtgtacta ctgcgcgaga 720
cggcgcttctt cccgttcttag cgtgaataga tggctggccg acgactacga cgtgtgggaa 780
cggggcacac aggtggctgtt ctgcagccacatggggccaa ctcacacatg cccacccgtgc 840
ccagcacccat aactcttggg gggacgtca gtcttccctc tccccccaaa accccaaaggac 900
accctcatgaa tctccggac ccctgagggtc acatgcgtt tggtggacgt gagccacgaa 960
gaccctgagg tcaagtcaa ctggtaatgtt gacggcggtt aggtgcataa tgccaagaca 1020
aaaggccggg aggacgacta caacacgacg taccgtgttgc tcaagcgtctt caccgtctg 1080
caccaggact ggctgaatgg caaggatgtc aagtgcgaa tcttccaaacaa agccctccca 1140
gccccatcg agaaaaccat ctccaaagcc aaaggccggc cccgagaacc acagggttac 1200
accctgcccc catccggga ggagatgacc aagaaccagg tcagcgtac ctgcctgtc 1260
aaaggcttctt atcccgacgtt catcgccgtt gagggtggaga gcaatggca gccggggaaac 1320
aactacaaga ccacgcctcc cgtgtggac tccgacggctt ctttctctt ctatagcaag 1380
cttaccgttgg acaagagcgtt gttggcggcgtt tcttcatgttc cgtgtatgtt 1440
gaggctgttca acaaccacta cacgcagaag agcctctccc tggctccggg taaatgaa 1497
```

SEQ ID NO: 34 moltype = AA length = 476

FEATURE Location/Qualifiers

REGION 1..476 note = AA6-E3-Fc binding agent

source 1..476 mol\_type = protein

organism = synthetic construct

SEQUENCE: 34

```
QLQLVETGGG LVQPGGSLRL SCAASGFTFS DYVMTWVRQA PGKGPWEIAT INTDGSTM RD 60
DSTKGRFTIS RDNAKNTLYL QMTSLKPEDT ALYCARGRV ISASAIRGAV RGPGTQVTVS 120
SGGGGSGGGG SGGGGSQVQL VESGGGLVQT GGSLRLSCAS SGSIAGFETV TWSRQAPGKS 180
LQWVASMTKT NNEIYSDSVL GRFIISRSNA KNTVYLOMNS LKPEDTGYYF CKGPRLRGQG 240
IQVTVSSGSD KTHTCPKCPA PELLGGPSV LFPPPKPKDTL MISRTPEVTC VVVDVSHEPD 300
EVKFNWVTDG VEVHNNAKTKP REEQYNSTYR VVSVLTVLHQ DWLNGKEYKC KVSNKALPAP 360
IEKTISKAKG QPREPQVYTL PPSREEMTKN QVSLTCLVKG FYPDSIAVEW ESNGQPENNY 420
KTPPPVLDSD GSFFFLYSKLT VDKSRWQQGN VFSCSVMHEA LHNHYTQKSL SLSPKG 476
```

SEQ ID NO: 35 moltype = DNA length = 1431

FEATURE Location/Qualifiers

misc\_feature 1..1431 note = AA6-E3-Fc binding agent

source 1..1431 mol\_type = other DNA

organism = synthetic construct

SEQUENCE: 35

```
caactgcgcg tggtaagac gggggggggc ttatgttgcg ctggagggtc tctcagactg 60
tcatgcgtc cctctggctt tacattcgtt gactacgtga tgacatgggt cccccaagct 120
ccaggaaagg ggcgttgcgtt gatcgtactt attaatcagc atggcagcac aatgcggac 180
gactccacaa aggggcggtt caccatttcc agagacaacg ccaagaatac tctgtacctt 240
cagatgacca gtctgaaacc cgaggacact gctctgtactt atgtgcaag aggccgggtt 300
atctctgtttt ccgttatcgtt agggcgacta agggggccctt gaacacaggt aaccgttca 360
tccggggggag ggggttgcgtt cgggtggggg tctggggggg gtggatccca agttcagctg 420
gtcgaatccg gggggcgactt ggtccagaca gggggctccc tgaggcttc cttgtgcattt 480
```

-continued

```
tccggaaagca tcgcccgtt cgagaccgtg acctggctc gccaggctcc cgggaagtct 540
ctgcagtggg tcgcttccat gactaagact aacaacgaga tctactctga ctcaagtaaa 600
ggccgcctca tcatttctag agataacgct aaaaacacag tgcatactgca gatgaatagt 660
ctcaaacctg aagacacagg cgtgtatttc tgtaagggtc ctgagctgag gggccaggc 720
atccaggtaa cagtcctcgag cggatccgac aaaactcaca catgcccccc gtgcccagca 780
cctgaactcc tggggggacc gtcaatcttc ctcttccccca caaaacccaa ggacacccctc 840
atgatctcc gggacccctga ggtcacatgc gtgggtggg acgtgagcca cgaagaccc 900
gaggtaactg tcaactggta cgtggacggc gtggagggtc ataatgccaa gacaaagccg 960
cgggaggagc agtacaacag caacgtacgt gtggcagcc tcctcacccgt cctgcaccc 1020
gactggctga atggcaaggg gtacaagtgc aagggtctcca acaaaggccct cccagcccc 1080
atcgagaaaa caatctccaa agccaaagggg cagccccccgg aaccacaggt gtacacctg 1140
cccccatccc gggaggagat gaccaagaac caggtcagcc tgacactgct ggtcaaaaggc 1200
ttctatccca gogacatcgc cgtggagggtg gagagcaatg ggcagccggaa gaacaactac 1260
aagaccaccc ttcccgatgc ggactccgatgg ggttctcat tcctatctatgc caagtcacc 1320
gtggacaaga gcaagggtggca gcaggggaaatc gtcttctcat gtcctgtat gcatgaggct 1380
ctgcacaacc actacacgca gaagacccctc tccctgtctc cgggttaatg a 1431
```

```
SEQ ID NO: 36 moltype = AA length = 457
FEATURE Location/Qualifiers
REGION 1..457
note = AH3-IgG1-heavy chain
source 1..457
mol_type = protein
organism = synthetic construct
SEQUENCE: 36
QVQLVETGGG LVQPGGSLRL SCAASGFTLD YSSIGWFRQA PGKEREVSC ISSSGDSTKY 60
ADSVVKGRFTT SRDNNAKNTVY LQMNLSLPDD TAVYYCAFR ATMCGVFPLS PYGKDDWGKG 120
TLTVTSSAST KGPSVFPLAP SSXSTSGGTA ALGCLVKDYF PEPVTWSNS GALTSGVHTF 180
PAVLQSSGLY SLSSVVTVPSS SSLGTQTYIC NVNHKPSNTK VDKRVEPKSC DKTHTCPSCP 240
APELLGGPSV FLFPPPKPKDT LMISRTPEVT CVVVDVSHED PEVKFNWYVD GVEVHNATK 300
PRREEQYNSTY RVSVSLTVLH QDWLNGKEYK CKVSNKALPA PIEKTISKAK GQPREPQVYT 360
LPPSREEMTK NQVSLTCLVK GFYPSDIAVE WESNGQPENN YTTPPVLDSDGSFFLYSKL 420
TVDKSRWQQG NVFSCSVMHE ALHNHYTQKS LSLSPKG 457
```

```
SEQ ID NO: 37 moltype = DNA length = 1374
FEATURE Location/Qualifiers
misc_feature 1..1374
note = AH3-IgG1-heavy chain
source 1..1374
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 37
caggtagacggc tggtggagac gggggggggg ctggtagacac caggccggc actgaggctt 60
tccctgtcgc catctgggtt cacactggat tattcgtccaa taggggtggg tcggcaggct 120
cctggcaaaag agcgtggggg ggtctcatgt attagtagta gtgggtatag cacaaggatc 180
ggccgatccg taaaggccgc gtttacaacc tccagggtata atgctaaggaa caccgtatata 240
ctccagatgc actctgttgc gcccgaatc acggccgtat attachtgc ggccttcagg 300
ggactatgtt ggggggtttt ccctctggc ctttacggca aggacactg gggcaagggg 360
accctgggtgaa cccgtctcgag tgccgtcgacc aaggggccat cggcttccc gctagcaccc 420
tccctccaaatc gacacttgcgg gggcacageg ggcctgggctt gctgggtcaatggactatcc 480
cccgaaactggc tgacggcttc gttggaaactca gggccctgtt ccacggccgtt gcaacacttc 540
ccggctgtcc tacagtcctc aggacttac cccttcgtca gctgtggatc cgtgcctcc 600
agcagcttgc gcacccagac ctatctgc aacgtgaatc acaagcccaag caacaccaag 660
gtggacaaga gagggtggccaa caaatctgtt gacaaaactt acacatgccc accgtgccca 720
gcacctgttgc tccctgggggg accgtcgttgc ttccctttcc ccccaaaacc caaggacacc 780
ctcatgtatcc cccggacccc tgaggatcata tgctgtgggg tggacgttgcg ccacggaaatc 840
cctggaggatca agttcaactg gtacgtggac ggcgtggagg tgcataatgc caagacaaatg 900
cccgccggagg agcgttacaaatc cagccgtatc cgtgtggatc gctgtccatc cgtcttcac 960
caggacttgc tgaatggccaa ggaggatcata tgcacgggtt ccaacaaaggc cctcccaaggcc 1020
cccatcgaga aaacatctc ccaaaatggcc gggcagccccca gagaacccca ggtgtacacc 1080
ctggccccat cccggggggaa gatgaccaatc aaccggatca gctgtccatc cctgtcaaa 1140
ggcttctatcc cccggggggaaatc cggccgtgggg tggggagggca atggggccgggg ggagaacaaatc 1200
tacaagacccca gcccggccgtt gctggacttgc gacgggtccat tcccttctca tagcaagtc 1260
accgtggacca agagcgggttgc gacggggggaaatc cgtgtccatc gatgtcatgatgg 1320
gctctgtcaca accactacac gcaaggatc cttccctgtt cccggggtaatc 1374
```

```
SEQ ID NO: 38 moltype = AA length = 457
FEATURE Location/Qualifiers
REGION 1..457
note = 5D-IgG1-heavy chain
source 1..457
mol_type = protein
organism = synthetic construct
SEQUENCE: 38
QVQLVESGGG LVQPGGSLRL SCEASGFTLD YYGIGWFRQP PGKEREAVSY ISASARTILY 60
ADSVVKGRFTI SRDNNAKNAVY LQMNLSLKRED TAVYYCARRR FSASSVNRL ADDYDVWGRG 120
TQAVVSSAST KGPSVFPLAP SSXSTSGGTA ALGCLVKDYF PEPVTWSNS GALTSGVHTF 180
PAVLQSSGLY SLSSVVTVPSS SSLGTQTYIC NVNHKPSNTK VDKRVEPKSC DKTHTCPSCP 240
APELLGGPSV FLFPPPKPKDT LMISRTPEVT CVVVDVSHED PEVKFNWYVD GVEVHNATK 300
```

-continued

---

PREEQYNSTY	RVVSVLTVLH	QDWLNGKEYK	CKVSNKALPA	PIEKTISKAK	GQPREPQVYT	360
LPPSREEMTK	NQVSLTCLVK	GFYPSDIAVE	WESNGQPENN	YKTPPVLDs	DGSFFLYSKL	420
TVDKSRWQQG	NVFSCSVMHE	ALHNHYTQKS	LSLSPGK			457

---

SEQ ID NO: 39	moltype = DNA length = 1374	
FEATURE	Location/Qualifiers	
misc_feature	1..1374	
	note = 5D-IgG1-heavy chain	
source	1..1374	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 39		
caaggtaaac tgggtgaatc	tgggtacaac ctgggggatc cctgagactc	60
tcttgcgagg ctcggatt	cacctggac tactatggca tggctgggtt ccgcaggccc	120
ccagggaaagg agcgggaggc	cgtttccatac attagtgcctt gtcggccggac catactgtac	180
gcagactctg tgaaggagc	ctttaccatc tctaggagaca atgccaaaa tgctgtgtac	240
ctgcaaatga acagcttcaa	gcggaggat acccgactgt actactgcgc gagacggcgc	300
ttctccgctt ctatcgtaa	tagatggctg gcccggact acgacgtgtg gggacgggac	360
acacagggtgg ctgttcggag	cgctcgacc aaggggccat cggcttccc gctagcacccc	420
tcctccaaga gcacactctgg	gggcacagcg gcccgggtt gcctggtaa ggactacttc	480
cccgAACCTG tgacggcttc	gtggaaactca ggccggctgtt ccageggcgt gcacaccctt	540
ccggctgtcc tacatgttcc	aggacttcatc tccctcagca ggtgtgtgac cgtgcctcc	600
agcagcttgg gcacccagac	ctatcatctgc aacgtgaatc acaagccccag caacaccaag	660
gtggacaaga gagttgagcc	caaattctgt gacaaaactc acacatgcc accgtgcccc	720
gcacctaaca ttctgggggg	accgctcgttcc ccccaaaaacc caaggacacc	780
ctcatgtatcc cccggacccc	tgaggatcaca tggctgggtt tggacgttgac ccacgaagac	840
cctgaggatc agttcaactg	tgatctggac ggcgtggagg tgcataatgc caagacaaag	900
ccggggggagg agcagtacaa	cagcacgtac cgtgtggtaa ggtgtctcac cgctctgcac	960
caggacttgtc tgaatggcaa	ggagttacaag tgcaagggtctt ccaacaaagc cctcccaagcc	1020
cccatcgaga aaccatctc	caaagccaaa gggcagcccc gagaaccaca ggtgtacacc	1080
ctgccccccat cccggggagg	gttgcggatcc aaccagggtca gcctggactt cctggtaaaa	1140
ggtttctatc ccagcgacat	cgccgtggag tgggagggca atgggagggc ggagaacaaac	1200
tacaagacca cgcctccgt	gctggactcc gacggctctt tttctctta tagcaagctc	1260
accgtggaca agagcgatgt	gcacggggggg aacgttctt catgtccgt gatgtatgg	1320
gtctcgacca accactacac	gcagaagagc ctctccctgt cccgggtaa atga	1374
SEQ ID NO: 40	moltype = AA length = 228	
FEATURE	Location/Qualifiers	
REGION	1..228	
	note = AA6-IgG1-kappa chain	
source	1..228	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 40		
QLQLVETGGG LVOPGGSLRL	SCAASGFTFS DYVMTWVRQA PGKGPEWIAT INTDGSTMRD	60
DSTKGRPTIS RDNAKNTLYL	QMITSLKPEDT ALYYCARGRV ISASAIRGAV RGPGTQVTVS	120
SRTVAAPSVF IPPPSDEQLK	SGTASVYVCLL NNFPREAVK QWKVDNALQS GNSQESVTEQ	180
DSKDSTSYLS STLTLKADY	EKHKVYACEV THQGLSSPVT KSFNRGEC	228
SEQ ID NO: 41	moltype = DNA length = 687	
FEATURE	Location/Qualifiers	
misc_feature	1..687	
	note = AA6-IgG1-kappa chain	
source	1..687	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 41		
caactgcacg tggtagagac	agggggccgc ttagttcagc ctgggggtt tctcagactg	60
tcatgcgtc cctctggctt	tacattcagt gactacgtg tgacatgggt ccgcacact	120
ccagggaaagg ggcctgagtg	gatecgtact attaatacag atggcagcac aatgcggac	180
gactccacaa agggggcggtt	caccatttcc agagacaacg ccaagaatac tctgtaccc	240
caagatgcaca gtctgaaacc	cgaggacact gctctgtactt attgtcaag aggccgggt	300
atctctgtt ccgtatcgt	aggcgcgtt agggggccctg gaacacaatg aactgtctcg	360
agccgtatccgg tggctgcacc	atcttccctc catctgtatc gcaatggaaa	420
tctggaaatg cctctgttgt	gtgcgtgtctt aataactctt atccctggaaa ggccaaatgt	480
cagtggaaagg tggataacgc	cctccaaatcg ggttaactccc aggagatgtt cacagacgc	540
gacagcaagg acagcaccta	cagccatcgac agcacccctgatcgacaa agcagactac	600
gagaaacaca aagtctacgc	ctgcgaatgtc acccatcagg gcctgatc gcccgtcaca	660
aagagcttca acaggggaga	gtgttga	687
SEQ ID NO: 42	moltype = AA length = 218	
FEATURE	Location/Qualifiers	
REGION	1..218	
	note = E3-IgG1-kappa chain	
source	1..218	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 42		
QVQLVESGGG LVQTGGSLRL	SCASSGSIAG FETVTWSRQA PGKSLQWVAS MTKTNNEIYS	60

-continued

---

DSVKGRFIIS RDNAKNTVYL QMNSLKPEDT GYVFKGPTEL RGQGIQVTVS SRTVAAPSVF 120  
 IFPPSDEQLK SGTASVCLL NNFYPREAKV QWKVDNALQGS GNSQESVTEQ DSKDSTYSL 180  
 STLTLSKADY EKHKVYACEV THQGLSSPVT KSFNRGEC 218

SEQ ID NO: 43 moltype = DNA length = 657  
 FEATURE Location/Qualifiers  
 misc\_feature 1..657  
 note = E3-IgG1-kappa chain  
 source 1..657  
 mol\_type = other DNA  
 organism = synthetic construct

SEQUENCE: 43  
 caagtccagc tggtegaatc cggggcgga ctgggtccaga cagggggctc cctgaggctc 60  
 tcctgtgcat ttccggaaag catggccggc ttccgagaccg tgacactggc tcggcaggct 120  
 cccggaaagt ctctgcagtg ggtcgcttc atgactaaga ctaacaacga gatctactct 180  
 gactcagtga aaggccgtt catcattct agagataacg ctaaaaacac agtgtatctg 240  
 caqatgaata gtctcaaac tgaacacaca ggcgtgtatt tctgtaaaggg tcctgagctg 300  
 agggggccagg gcattccagg aacagtctcg aecgcgtacgg tggctgcacc atctgtttc 360  
 atcttccccg catctgtatgc gcaagtgtaaa tctgtgaactg cctctgttgt gtgcctgctg 420  
 aataactctt atcccagaga ggc当地aaagta cagtggaaagg tggataacgc cctccaatcg 480  
 ggtaactccc aggagagtgt cacagagcag gacagcaagg acagcaccta cagccctcagc 540  
 agcaccctga cgctgagcaa agcagactac gagaacacaca aagtctacgc ctgcgaagtc 600  
 acccatcagg gcctgagctc gcccgtcaca aagagcttca acaggggaga gtgttga 657

SEQ ID NO: 44 moltype = AA length = 599  
 FEATURE Location/Qualifiers  
 REGION 1..599  
 note = AH3-5D-IgG1 heavy chain  
 source 1..599  
 mol\_type = protein  
 organism = synthetic construct

SEQUENCE: 44  
 QVOLVETGGG LVQPGSSLRL SCAASGFTLD YSSIGWFRQA PGKEREGVSC ISSSGDSTKY 60  
 ADSVKGRFTT RDNAKNTVY LQMNSLKPDD TAVYYCAA FR ATMCGVFPLS PYGKDDWGKG 120  
 TLTVTSSGGG GSGGGGSSGG GSQVQLVESP GGLVQPGSSL RLSCEASGFT LDYYGIGWFR 180  
 QPPGKEREAV SYISASARTI LYADSVKGRF TISRDNAKNA VYLMQMSLKR EDTAVYYCAR 240  
 RRPASSVNR WLADDYDVWG RGTQAVVSSA STKGPSVPFL APSSSKTSSG TAALGCLVKD 300  
 YFPEPVTVSW NSGALTSGVH TFPAVLQSSG LYSLSSVVTV PSSSLGTQTY ICNVNHPKSN 360  
 TKVDKRVEPK SCDKTHTCPP CAPELLGGP SVPLFPPKPK DTLMISRTP ETCVVVDVSH 420  
 EDPEVKPNWY VDGVEVHNAK TKPREEQYNS TYRVVSVLTV LHQDWLNGKE YKCKVSNKAL 480  
 PAPIEKTIK AKGQPREPQV YTLPSSREEM TKNQVSLTCL VKGFYPSDIA VEWESNGQPE 540  
 NYKTTPPVLDSDGSFFLYS KLTVDKSRWQ QGNVFSCCSVME HEALHNHYTQ KSLSLSPKG 599

SEQ ID NO: 45 moltype = DNA length = 1800  
 FEATURE Location/Qualifiers  
 misc\_feature 1..1800  
 note = AH3-5D-IgG1 heavy chain  
 source 1..1800  
 mol\_type = other DNA  
 organism = synthetic construct

SEQUENCE: 45  
 caaggtaacgc tggtgagac gggggggggg ctggtacaac caggcggtc actgaggctt 60  
 tcctgtgcgc catctgggtt cacactggat tattcgtccca taggggtgtt tcggcaggct 120  
 cctggcaaaag aecgtggggg ggtctcatgt attagtagta gtgggtatag cacaatgtac 180  
 gccgattccg taaaggccg gtttacaacc tccaggatgtatgtaaaggaa caccgtatat 240  
 ctccaggatgc actctgtaa gcccgcacat acggccgtat attactgtgc ggcttcagg 300  
 ggcactatgt gcccgcgtt ccctctgago ccttacggca aggacgactg gggcaagggg 360  
 accctgggtga ccgtatccctc aggcgggtgg gggctctggggggggctc aggggggtgg 420  
 ggcagccagg tgcactgtt tgaatctgggg ggaggcttggggatccctgg gggatccctg 480  
 agactctttt gggaggccctc ccgttaccac tggactact atggcatccg ctggttccgc 540  
 cagccccccag ggaaggagcg ggaggccgtt tcatacatta gtgccagtgc cccgaccata 600  
 ctgtacgcg actctgtaa gggacgtttt accatcttca gggacaatgc caaaaatgtct 660  
 gtgtacactgc aaatgaacacg cctcaagccg gaggatcccg cagtgtacta ctgcgcgaga 720  
 cggcgctttt ccgttccatcg cgttaataga tggctggcccg acgactacga cgtgtgggg 780  
 cggggccacac aggtggctgt ctgcagccgg tgcaccaagg gcccacatgtt cttccgcata 840  
 gcaccctctt ccaagacac ctctggggggc acagccggccc tgggctgcgtt ggtcaaggac 900  
 tacttcccccg aacctgtgcg ggtctcggtt aactcaggccg ccctgaccag cggcggtc 960  
 accttcccccg ctgttcataca gtcctcaggat ctctactccctc ctgcgcgtt ggtgaccgtg 1020  
 ccctccacca gtttggggcac ccacgacttcc atctgcacac tgaatcacaa gcccacac 1080  
 accaagggtgg acaagagagt tgagccaaa tcttgcataa aaactcacac atgcccaccc 1140  
 tgcccagcac ctgaactccctt gggggggaccc tcaacttccctt tttttttttt aaaacccaaag 1200  
 gacaccctca tgatctcccg gaccccttagt gtcacatgcg tgggtgggca cgtgaggccac 1260  
 gaagaccctg aggtcaagttt caactgtgtac gtggacggccg tggagggtggca taatgcac 1320  
 acaaaggccg gggaggagca gatcaacacg acgttacccgtt tggctacgtt ccttacccgtc 1380  
 ctgcaccagg actggctgaa tggcaaggag tacaagtgcg aggtctccaa caaaggccctc 1440  
 ccagccccca tcgagaaaaac catctccaaa gccaaaggccg accacaggtg 1500  
 tacaccctgc cccatccccg ggaggagatg accaagaacc aggtcagctt gacctgcctg 1560  
 gtcaaaaggct tctatccccg cgacatcgcc gtggagtgggg agagcaatgg gcagccggag 1620  
 aacaactaca agaccacgccc tccctgtctg gactccgcac gtccttctt cctctatagc 1680

-continued

---

```
aagctcaccg tggacaagag caggtggcag caggggaacg tcttctcatg ctccgtatg 1740
catgaggctc tgcacaacca ctacacgcag aagacgcctc ccctgtcccc gggtaatga 1800
```

```
SEQ ID NO: 46      moltype = AA length = 354
FEATURE          Location/Qualifiers
REGION           1..354
note = AA6-E3-IgG1 light chain
source            1..354
mol_type = protein
organism = synthetic construct

SEQUENCE: 46
QLQLVETGGG LVQPGGSLRL SCAASGFTFS DYVMTWVRQA PGKGPEWIAT INTDGSTMRD 60
DSTKGRFTIS RDNAKNTLYL QMITSLKPEDT ALYYCARGRV ISASAIRGAV RGPGTQVTVS 120
SGGGGSGGGG SGCGGSQVQL VESGGGLVQT GGSLRLSCAS SGSIAGFETV TWSRQAPGKS 180
LQWVASMTKT NNEIYSDSVK GRFIISRDN A KNTVYLOMNS LKPEDITGVYF CKGPRLRGQG 240
IQVTVSSRTV AAPSVFIFPP SDEQLKSGTA SVVCLLNFFY PREAKVQWKV DNALQSGNSQ 300
ESVTEQDSKD STYSLSSTLT LSKADYEKHK VYACEVTHQG LSSPVTKSFN RGEC 354
```

```
SEQ ID NO: 47      moltype = DNA length = 1065
FEATURE          Location/Qualifiers
misc_feature     1..1065
note = AA6-E3-IgG1 light chain
source            1..1065
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 47
caactgcagc tggtagagac agggggcgcc tacatttcgt gctggagggtc tctcagactg 60
tcatgcgctg cttctggctt tacattcgt gactacgtg tgacatgggt ccgccaagct 120
ccaggaaagg ggcctgagtg gatcgctact attaatacag atggcagcac aatgcgggac 180
gactccacaa agggggcggtt caccatttcg agagacaaacg ccaagaatac tctgtacactt 240
cagatgcacca gtctgaaacc cggggcagact gctctgtact atttgtcaag agggccgggtg 300
atctctgctt cccgtatcag agggggcctg gaacacaggt aaccgttca 360
tccggggggg ggggttcagg cggtgggggta tctggggggg gtggatcca agttcagctg 420
gtcgaatccg gggggcggact ggccacacaa gggggctccc tgaggcttc ctgtgcacat 480
tccggaaagca tggccggctt cgagacgtg acctgggttc gccagggtccc cggggaaatct 540
ctgcagtggg tcgcttccat gactaagact aacaacgaga tctacttgc ctcagtggaaa 600
ggccgcttca tcatttctag agataacgcg aaaaacacag tggatctgca gatgaatagt 660
ctcaaaccctt aagacacagg cgtgtatcc tggatgggtc ctgagctgag gggccaggcc 720
atccaggtaa cagtctcgag ccgtacggtg gctgcacat cttgttccat cttccggcca 780
tctgtatgac agttgaaatc tggacttgc tctgttggtt gcctgtgaa taacttctat 840
cccagagagg ccaaagtaca gtggaaagggtg gataacgccc tccaaatcggg taactccag 900
gagagtgtca cagacggaga cagacacatca gctcagcag cacccgtacg 960
ctgagcaaaag cagactacga gaaacacaaa gtcatacgcc gctgaaatcc ccatcaggcc 1020
ctgagctcgc cggtcacaa aggttcacac agggggagat gttga 1065
```

```
SEQ ID NO: 48      moltype = AA length = 127
FEATURE          Location/Qualifiers
REGION           1..127
note = VHH peptide monomer 5D
source            1..127
mol_type = protein
organism = synthetic construct

SEQUENCE: 48
QVOLVESGGG LVQPGGSLRL SCEASGFTLD YYGIGWFROP PGKEREAVSY ISASARTILY 60
ADSVKGRFTI SRDNNAKNAVY LQMNSLKRED TAVYYCARRR FSASSVNRWL ADDYDVWGRG 120
TQVAVASS 127
```

```
SEQ ID NO: 49      moltype = DNA length = 381
FEATURE          Location/Qualifiers
misc_feature     1..381
note = VHH peptide monomer 5D
source            1..381
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 49
caggtgcagc tcgtggagtc aggtggaggc ttgggtgcagc ctggggggtc tctgagactc 60
tccctgtgaag cctctggatt cacttggat tattatgtta taggctgttt ccgcgcggcc 120
ccaggaaagg agcgcgaggc ggtctcatat attagtgcacat gtcggccgtac gatattgttat 180
gcagatccg tgaaggcccg atttaccatc tccagagacaa atgccaagaa cgcgggttat 240
ctacaaaatgaa acagcctgaa acgtgaggac acggctgtctt attactgtgc gagggggcga 300
ttctccgcgt ctatgtttaa tagatggctt gccgacgtact atgacgtctg ggggtcggggg 360
acccaggtcg cgggtgcctc a 381
```

```
SEQ ID NO: 50      moltype = AA length = 111
FEATURE          Location/Qualifiers
REGION           1..111
note = VHH peptide monomer E3
source            1..111
mol_type = protein
```

-continued

organism = synthetic construct

SEQUENCE: 50  
 QVQLVESGGG LVQTGGSLRL SCASSGSIAG FETVTWSRQA PGKSLQWVAS MTKTNNEIYS 60  
 DSVKGRFIIS RDNAKNTVYL QMNSLKPEDT GVFCKGPTEL RGQGIQVTVS S 111

SEQ ID NO: 51 moltype = DNA length = 333  
 FEATURE Location/Qualifiers  
 misc\_feature 1..333  
 note = VHH peptide monomer E3  
 source 1..333  
 mol\_type = other DNA  
 organism = synthetic construct

SEQUENCE: 51  
 cagggtgcaggc tcgtggagtc gggcgaggc ttgggtcaga ctggggggtc tctgagactc 60  
 tccctgtcat cctctggaaag tatecgccgt ttccgaaacc tgacctggc ccggccaggct 120  
 cctggaaaagt cgctccagtg ggtegcacatcg atgactaaaa ctaataacga gatctatca 180  
 gactccgtga aggcccatt catcatctcc agagacaacg ccaagaatac ggtgtatcta 240  
 caaatgaaca gcctgaaacc tgaggacaca ggegtctatt ttgtttaagg tcctgaggttg 300  
 aggccccagg ggatccaggc caccgtctcc tcg 333

SEQ ID NO: 52 moltype = AA length = 121  
 FEATURE Location/Qualifiers  
 REGION 1..121  
 note = VHH peptide monomer AA6  
 source 1..121  
 mol\_type = protein  
 organism = synthetic construct

SEQUENCE: 52  
 QLQLVETGGG LVQPGGSLRL SCAASGFTPS DYVMTWVRQA PGKGPEWIAT INTDGSTMRD 60  
 DSTKGRFTIS RDNAKNTLYL QMNTSLKPEDT ALYYCARGRV ISASAIRGAV RGPGBTQVTVS 120  
 S 121

SEQ ID NO: 53 moltype = DNA length = 363  
 FEATURE Location/Qualifiers  
 misc\_feature 1..363  
 note = VHH peptide monomer AA6  
 source 1..363  
 mol\_type = other DNA  
 organism = synthetic construct

SEQUENCE: 53  
 cagttgcaggc tcgtggagac agggggaggc ttgggtcagc ctgggggggtc tctgagactc 60  
 tccctgtcaggc cctctggatt cacgttcactg gactacgtca tgacccgtt ccggccaggct 120  
 ccaggaaaagg ggcccgaatg gatcgcact attaatacccg acggtagcac gatgcgttat 180  
 gactccacaa aaggcccatt caccatctcc agagacaacg ccaagaacac actgtatctg 240  
 caaatgacca gcctgaaacc ggaggacacg gccctgtatt actgtgcgag aggccgcgtg 300  
 atctccgcct ccgcgataag agggggccggg ggaccaggc caccgtctcc 360  
 tca 363

SEQ ID NO: 54 moltype = AA length = 126  
 FEATURE Location/Qualifiers  
 REGION 1..126  
 note = VHH peptide monomer AH3  
 source 1..126  
 mol\_type = protein  
 organism = synthetic construct

SEQUENCE: 54  
 QVQLVETGGL VQPQGGSLRLS CAASGFTLDY SSIGWFRQAP GKEREGVSCI SSSGDSTKYA 60  
 DSVKGRFTTS RDNAKNTVYL QMNSLKPDDT AVYYCAAFRA TMCGVFPPLSP YGKDDWGKGT 120  
 LTVSS 126

SEQ ID NO: 55 moltype = DNA length = 378  
 FEATURE Location/Qualifiers  
 misc\_feature 1..378  
 note = VHH peptide monomer AH3  
 source 1..378  
 mol\_type = other DNA  
 organism = synthetic construct

SEQUENCE: 55  
 caggtgcaggc tcgtggagac ggggggctt gtgcagcctt gggggctct gagactctcc 60  
 tggcaggcctt ctggatttac tttggattat tcgtccatag gttgggtccg ccaggcccca 120  
 gggaaaggaggc gtgagggggt ctcatgtatt agtagtagtg gtgatgcac aaagtatgca 180  
 gactccgtga aggcccatt caccatctcc agagacaacg ccaagaacac ggtgtatctg 240  
 caaatgaaca gcctgaaacc tgacgacaca gccgtttatt actgtgcagc ttttagggcg 300  
 actatgtcgc ggtgttccc ccttagcccc tacggcaagg acgactgggg caaagggacc 360  
 ctggtcaccg tctcttca 378

SEQ ID NO: 56 moltype = AA length = 15  
 FEATURE Location/Qualifiers  
 REGION 1..15

-continued

---

```

source          note = Flexible linker 1
                1..15
                mol_type = protein
                organism = synthetic construct
SEQUENCE: 56
GGGGSGGGGS GGGGS                                         15

SEQ ID NO: 57      moltype = DNA length = 45
FEATURE          Location/Qualifiers
misc_feature     1..45
note = Flexible linker 1
source           1..45
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 57
ggcggtggc gctctggtgg cggcggtcc ggtggcggtg gcagc          45

SEQ ID NO: 58      moltype = AA length = 15
FEATURE          Location/Qualifiers
REGION           1..15
note = Flexible linker 2
source           1..15
mol_type = protein
organism = synthetic construct

SEQUENCE: 58
GGGGSGGGGS GGGGS                                         15

SEQ ID NO: 59      moltype = DNA length = 45
FEATURE          Location/Qualifiers
misc_feature     1..45
note = Flexible linker 2
source           1..45
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 59
ggtgaggaggc gttcaggccgg aggtggctct ggcggtggc gttcc          45

SEQ ID NO: 60      moltype = AA length = 15
FEATURE          Location/Qualifiers
REGION           1..15
note = Flexible linker 3
source           1..15
mol_type = protein
organism = synthetic construct

SEQUENCE: 60
GGGGSGGGGS GGGGS                                         15

SEQ ID NO: 61      moltype = DNA length = 45
FEATURE          Location/Qualifiers
misc_feature     1..45
note = Flexible linker 3
source           1..45
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 61
ggcggtgtc gctctggtgg cggcggtcc ggtggcggtg gcagc          45

SEQ ID NO: 62      moltype = AA length = 259
FEATURE          Location/Qualifiers
REGION           1..259
note = 5D-E3 heterodimer
source           1..259
mol_type = protein
organism = synthetic construct

SEQUENCE: 62
QVOLVESGGG LVQPGGSLRL SCEASGFTLD YYGIGWFRQP PGKEREAVSY ISASARTILY 60
ADSVKGRTFI SRDNNAKNAY LQMETNSLKR EDTAVYYCAR RRFSASSVNR WLADDYDVWG 120
RGTQVAWSGG GGGSGGGGSG GGGSQVQLVE SGGLVQTGG SLRLSCASSG SIAGFETVTW 180
SRQAPGKSLQ WVASMBTTKT NNEIYSDSVK GRPIISRDNA KNTVYLQMET NSLPEDTGV 240
YFCKGPRLRG QGIQVTVSS                                         259

SEQ ID NO: 63      moltype = DNA length = 759
FEATURE          Location/Qualifiers
misc_feature     1..759
note = 5D-E3 heterodimer
source           1..759
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 63

```

---

-continued

---

```
caggtgcagc tcgtggagtc aggtggaggc ttgggtgcagc ctggggggtc tctgagactc 60
tctgtgaag cctctggatt cacttggat tattatggta taggctggtt ccggccagccc 120
ccagggaaagg agcgcgaggc ggtctcatat attagtgcac gtgcccgtac gatattgtat 180
gcagattccg tgaaggggcg atttaccatc tccagagaca atgccaagaa cgccgtgtat 240
ctacaaatga acagcttgaa acgtggggac acggctgtctt attactgtgc gagggggcg 300
tttcccgctg ctatgttta tagatggctt ggccacgtac atgacgtctg ggggggggg 360
acccaggctc cggtgtccctc agggcggtt ggctctggg gcccgggttc cggtggcggt 420
ggcagccagg tgcagctgtt ggagtgggc ggaggcttgg tgcaacttgg ggggtctctg 480
agactctctt gtgcacttc tggaaatgtt gcccgggttcc aaaccgtgac ctgggtccgc 540
caggctctg gaaagtgcgtt ccagtgggtc gcatcgatgtt ctaaaactaa taacgagatc 600
tattcagact ccgtgaaggg ccgttcatc atcccaagaa acaacgcocaa gaatacggtg 660
tatctacaaa tgaacagcct gaaaccttgg gacacaggcg tctatttttgg taaaggctt 720
gagttgaggg gccagggtt ccagggttcc accgttggc 759
```

```
SEQ ID NO: 64 moltype = AA length = 272
FEATURE Location/Qualifiers
REGION 1..272
note = AH3-AA6 heterodimer
source 1..272
mol_type = protein
organism = synthetic construct
SEQUENCE: 64
QVQLVETGGL VQPQGSSLRLS CAASGFTLDY SSIGWFRQAP GKEREGVSCI SSSGDSTKYA 60
DSVKGRFTTS RDNAKNTVYL QMETNSLKPQ DTAVYYCAAF RATMETCGVF PLSPYKGDDW 120
GKGTLVTVSS GGGGSGGGGS GGGGSQLQLV ETGGLVQPG GSLRLSCAAS GFTFSDYVME 180
TTWVRQAPGK GPEWIATINT DGSTMETRDD STKGRFTISR DAKNTLYLQ METTSLKPED 240
TALYYCAGR VRGPQTVTSS 272
```

```
SEQ ID NO: 65 moltype = DNA length = 786
FEATURE Location/Qualifiers
misc_feature 1..786
note = AH3-AA6 heterodimer
source 1..786
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 65
caggtgcagc tcgtggagac gggggggctt gtcaggttgc ggggggtctt gagactctcc 60
tgtgcagctt ctggattcac ttggattttc tcgtccatag gtcgggttcgg ccaggcccca 120
ggggaggagc tgagggggtt ctcatgtttagt agtagatgtt gtgatgttgc aagatgtca 180
gactccgtga agggccgatt cacccatcc agagacaac ccaagaaacac ggtgtatctg 240
caaataatgttgc gcttggggcc cttttttttt actgtgttgc ttttagggcg 300
actatgttgc ggtgttcccc cttttttttt tacggcaatgg acggatgggg caaaaggacc 360
ctgggtcaccg ttcctcttggg tctgggtggc tctgggtggc ggggttccgg tggcggtggc 420
agccaggatgc agctcggttgc gacagggggg ggtttttttt acggatgggg gttttttttt 480
ctctccgtgttgc cttttttttt attttttttt agtggacttgc tttttttttt ggtttttttt 540
gttcccggttgc aagggtttttt atggatgttgc actttttttt cttttttttt cttttttttt 600
gatgacttgc aaaaaggccg attttttttt tccagagaca acggccaaacacactgttat 660
ctgttttttttccgttgc aacggggggc acggatggggt attactgttgc gagggccgc 720
gttcccggttgc cttttttttt aagggtttttt gttttttttt cggggggccca ggtttttttt 780
tccctttttttt 786
```

```
SEQ ID NO: 66 moltype = AA length = 6
FEATURE Location/Qualifiers
REGION 1..6
note = six-histidine tag
source 1..6
mol_type = protein
organism = synthetic construct
SEQUENCE: 66
HHHHHHH 6
```

```
SEQ ID NO: 67 moltype = AA length = 13
FEATURE Location/Qualifiers
REGION 1..13
note = E-tag for protein purification
source 1..13
mol_type = protein
organism = synthetic construct
SEQUENCE: 67
GAPVPYPDPL EPR 13
```

```
SEQ ID NO: 68 moltype = DNA length = 6985
FEATURE Location/Qualifiers
misc_feature 1..6985
note = Plasmid pD1214-FAKS-His-hABAB-D7
source 1..6985
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 68
```

-continued

---

tcagaattgg ttaattgggtt gtaacactga cccttatttttgc taaatacatt	60
caaataatgtt tccgctcatg agacaataac cctgataaat gttcaataaa tattgaaaaaa	120
ggaagaatat gagtttcaaa catttccgtg tcgcctttaat tccctttttt gcggcattt	180
gccttcctgt ttttgcac ccagaacacg tggtaaaagt aaaagatgtt gaagatcagt	240
tgggtgcacg aatgggttac atcacaatgg atcacaacag cgtaaaatgtt ctggagatgt	300
ttcgccccca gaaacgtttt ccaatgttgc gcaattttaa agttctgttgc tggtggcg	360
tattatcccg tatttgcgc gggcaagago aactcggtcg cccgcatacac tatttgcaga	420
atgacttggt tgtagtactca ccagtccacag aaaagcatct tacggatggc atgacagtaa	480
gagaattatg cagtgcgtc ataaccatga gtgataacac tgccggccaaat ttacttcgtt	540
caacgatcg gaggaccgaag gagetaacccg ctttttgca caacatgggg gatcatgtt	600
ctcgcccttgc tctgtggaa ccggactgttgc atgaaggccat accaaacgac gagegtgtt	660
ccacgatgcc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	720
ctttagtcc tccggcaacaa ttaatagatg ggatgggggc ggataaaaggat gcaggacac	780
ttctgcgtcgc gggcccttcgg gctgggttgc ttatttgcgtt taaatccggc gcccgtgtt	840
gtgggttcggc cggatcatcgc gcaatgttgc gggccatgtt taagcccttcc cgtagtgc	900
ttatcttacac gacggggagt cggcaacta tggatgttgc aaatagacag atcgtgttgc	960
taggtgcctt actgtttaa catttttttttgc ttttttttttgc ttttttttttgc	1020
gcccgcgtcg ttccactgttgc ctttttttttgc ttttttttttgc ttttttttttgc	1080
tccttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	1140
ggtttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	1200
agcgcagata ccaaataactt ttcttcgttgc gttagccgttgc tttagccccc accttcaagaa	1260
ctctgttagca cccgttccat accttgcgttgc gctaatttttttgc tttagccgttgc	1320
ttggcgataag ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	1380
gccccccggc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	1440
cggacttgcgttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	1500
ggggggggggc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	1560
agggggggggc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	1620
tcgattttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	1680
cttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	1740
ccctgatttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	1800
tgccgggttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	1860
cccgccatcg cccgttccat ttttttttttgc ttttttttttgc ttttttttttgc	1920
cttcctcttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	1980
ctctgttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	2040
tggaaattttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	2100
ataaaacggc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	2160
tttacttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	2220
ccatcttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	2280
acttccacttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	2340
tttgagggttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	2400
tttgcatttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	2460
aaaagggggggc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	2520
ggggggggggc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	2580
acacttggatttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	2640
gtctcatgttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	2700
tttaccaacttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	2760
cccgacgatc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	2820
cctctgttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	2880
ggggggggggc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	2940
gaatcttgggg gggggggggc ttttttttttgc ttttttttttgc ttttttttttgc	3000
ggatccaccc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	3060
tttgcatttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	3120
ggacgcgttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	3180
ctcaagcggttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	3240
gtgaatagatgttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	3300
ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	3360
cagcttggtag ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	3420
gttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	3480
aaggggggcc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	3540
acaaaggggcc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	3600
accagtctgttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	3660
gttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	3720
ggggggggggc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	3780
tccggggggcc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	3840
agcatcgccg ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	3900
tgggtcggtt ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	3960
tttcatcttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	4020
cctcgaaatgttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	4080
gttaacatgttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	4140
taataaagggttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	4200
tcccgcttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	4260
tttataatgttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	4320
tacagacgttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	4380
cgttgcgttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	4440
catcatttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	4500
taatcttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	4560
gcataatgttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	4620
caaaaaccatc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	4680
ctactctatcc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc	4740

-continued

---

```

acttgtgtgc ttcatggat gttegtacca ccaagaaattt actggaggtt gttgaagcat 4800
taggtccaa aatttgtta ctaaaaaacat atgtggatat ctgactgtat tttccatgg 4860
aggcacagt taagcccta aaggcattat ccgccaagta caatttttt ctcttcgaag 4920
atagaaaatt tgctgacatt ggtataacag tcaaattgca gtactctgct ggtgtataca 4980
gaatagcaga atggccagac attacaatg cacacgggtt ggtggggcca ggttattttt 5040
gcgggttgaa gcaggccgca gaagaagtaa caaaggaaac tagggccct ttgtatgttag 5100
cagaattgtc atgcaaggc tccctatcta ctggagaata tactaagggt actgttgaca 5160
tttgcgaaag cgacaaagat ttgttatcg gctttattgc tcaaagagac atgggtggaa 5220
gagatgaaagg ttacgatgg ttgtatgtc caccgggtt ggggtttagat gacaaggag 5280
atgcattggg tcaacatg agaaccatgg atgtatgtt ctctacaggat tctgacattt 5340
ttattgtggg aagaggacta ttgttcaaggd gaagggtatc taaggtagag ggtgaacgtt 5400
acagaaaaggc aggctggaa gcataattgtt gaagatgcgg ccagcaaaac taaaaactg 5460
tattataatgt aaatgtatc atactaaatc cacaattatgtt agcttcaattt taattatc 5520
agtatttacc acgttacatc tccaattatca aaggaaatgtt tagcttgcgaa ggatgagact 5580
aatccaaatggc aggagtggca gcatatagaa cagcttacagg gtatgtgtt agggaaagcata 5640
cgataccccc catggatgg gataatatca caggaggatc tagactactt ttcattctac 5700
ataaatagac gcataatgtt acgcattttt gcataaacac gcactatgcc gtttcttca 5760
tgttatatat tatatacgca acacgcgtt ataggtgcgtt cgtgaacagt gagctgtatg 5820
tgcgcagtc ggcgttcattt ttcggaaaggctt ctcgttttcg gaaacgtt gaaatgttca 5880
ttccgaaggctt cttttttttt agaaaatgtt ggaacttcgtt agcgcttttgggaaaacaaa 5940
gcgcctctgaa gtcgcactt caaaaaaccccaaaaacgcggcc gggactgttac gggactctaa 6000
aatatttgcgaa ataccgttcc cacaacattt gctcaaaatgtt atctcttgc tatataatctc 6060
tgcgttatatat ccttccatc cacccttcgcg tcccttgcg tcccttgcg tcccttgcg 6120
ctcgaccatc acatttttt tttttttttt tagtattactt ctttagacaa aaaaatttgcg 6180
gtaaagaactt ttcatagatgtt gatgttggaaa caatacgaaa atgttacat ttccttatacg 6240
tagtatataatg agacaaaatgtt gaaagaaatccgg ttcataattt ttcgttccatc gaaatgttca 6300
caacgcgttccatc atcccttgcg cacaatgtt ggcggatccatc catcggtatc gaaatataatc 6360
ggggatgcctt ttatgttgcgaa aaaatgcacc cgcggatccatc cttggatccatc gtttttttgcg 6420
gaagtggatgtt cgggtttttt ttatgttgcgaa gaaaatgttac accaaatgtt ctttttttgcg 6480
acottaaacggg accttacatgtt cacaatgtt gtcgttacatc gtttttttgcg gggccacaaag 6540
gagaaaaaaa gtaatgttacatc atcccttgcg ttttttttttggggatccatc gggccatccatc 6600
gttgcgttccatc atcccttgcg ctttttttttgcg gggccatccatc gggccatccatc 6660
cttgcgttccatc atcccttgcg ctttttttttgcg gggccatccatc gggccatccatc 6720
tttttttttttgcg gggccatccatc gtttttttttgcg gggccatccatc gggccatccatc 6780
catttttttttgcg gggccatccatc gtttttttttgcg gggccatccatc gggccatccatc 6840
tttgcgttccatc atcccttgcg ctttttttttgcg gggccatccatc gggccatccatc 6900
aaatgttgcgc gggccatccatc gtttttttttgcg gggccatccatc gggccatccatc 6960
catgagacaa taaccctgtat attgg 6985

```

```

SEQ ID NO: 69 moltype = DNA length = 6916
FEATURE Location/Qualifiers
misc_feature 1..6916
note = Plasmid pD1214-FAKS-hABAB
source 1..6916
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 69
tcagaattttt ttaattttttt gtaacactga ccccttattttt ttttttttcaatacatt 60
caaataatgtt tccgttcatg agacaataac cctgataat gttcaatataa tattggaaaaaa 120
ggaaagatattt gatgttccatc cattttccgtt tcggccctttat tcccttttgcggcatttt 180
gccttcctgtt ttttgcgttccatc cccggatccatc gggccatccatc gggccatccatc 240
tgggtgttccatc atcccttgcg ttttttttttgcg gggccatccatc gggccatccatc 300
tttcgttccatc agaacttgcg ttttttttttgcg gggccatccatc gggccatccatc 360
tattatcccgat ttttttttttgcg gggccatccatc gggccatccatc gggccatccatc 420
atgacttttttgcg ttttttttttgcg gggccatccatc gggccatccatc gggccatccatc 480
gagaatttttgcg ttttttttttgcg gggccatccatc gggccatccatc gggccatccatc 540
caacgcgttccatc atcccttgcg ctttttttttgcg gggccatccatc gggccatccatc 600
cttcgttccatc atcccttgcg gggccatccatc gggccatccatc gggccatccatc 660
ccacgttccatc atcccttgcg gggccatccatc gggccatccatc gggccatccatc 720
cttcgttccatc atcccttgcg gggccatccatc gggccatccatc gggccatccatc 780
tttcgttccatc atcccttgcg gggccatccatc gggccatccatc gggccatccatc 840
gttgcgttccatc atcccttgcg gggccatccatc gggccatccatc gggccatccatc 900
tttatcttccatc atcccttgcg gggccatccatc gggccatccatc gggccatccatc 960
taggtgttccatc atcccttgcg gggccatccatc gggccatccatc gggccatccatc 1020
ggcgccgttccatc atcccttgcg gggccatccatc gggccatccatc gggccatccatc 1080
tttcgttccatc atcccttgcg gggccatccatc gggccatccatc gggccatccatc 1140
gttgcgttccatc atcccttgcg gggccatccatc gggccatccatc gggccatccatc 1200
agcgttccatc atcccttgcg gggccatccatc gggccatccatc gggccatccatc 1260
cttcgttccatc atcccttgcg gggccatccatc gggccatccatc gggccatccatc 1320
tttcgttccatc atcccttgcg gggccatccatc gggccatccatc gggccatccatc 1380
ggcgccgttccatc atcccttgcg gggccatccatc gggccatccatc gggccatccatc 1440
cgaaacttgcg ttttttttttgcg gggccatccatc gggccatccatc gggccatccatc 1500
ggcgccgttccatc atcccttgcg gggccatccatc gggccatccatc gggccatccatc 1560
aggggggaaac gcttgcgttccatc atcccttgcg gggccatccatc gggccatccatc 1620
tgcgttccatc atcccttgcg gggccatccatc gggccatccatc gggccatccatc 1680
cttcgttccatc atcccttgcg gggccatccatc gggccatccatc gggccatccatc 1740
ccctgttccatc atcccttgcg gggccatccatc gggccatccatc gggccatccatc 1800
tgcgttccatc atcccttgcg gggccatccatc gggccatccatc gggccatccatc 1860
cccgccatccatc atcccttgcg gggccatccatc gggccatccatc gggccatccatc 1920

```

-continued

---

cttcctctag	ggtgtcgta	attaccgt	ctaaaggttt	ggaaaagaaaa	aaagtgaccg	1980
cctcggttct	ttttctcg	cgaaaaagg	aataaaaat	tttatcacgt	ttcttttct	2040
tgaaaat	tttttgat	tttttct	ttcgatgacc	ccccattgat	attaagtt	2100
ataaaacgg	ttcaatttct	caagttcg	tttcat	ttgttctat	tacaacttt	2160
tttacttctt	gotcattaga	aagaagat	agcaatctaa	tctaagttt	aatgagattc	2220
ccatctattt	tcaccgtgt	tttgcgt	gcttctctg	cattggctgc	ccctgtttaac	2280
actaccact	aagacgagac	tgctcaaatt	ccagctga	cagttatcgg	ttactctgac	2340
cttgagggtt	attcgacgt	cgctgtttt	ccttctct	actccactaa	caacgggtt	2400
ttgttccat	acaccat	cgcttccat	gtgtcgaa	aagagggtt	ctctctcgag	2460
aaaagagagg	ccgaagat	cgaggat	ctgggggaga	cggggggg	gtgttacaa	2520
ccaggcgggt	cactgaggct	ttctgtgc	gcatctgggt	tcacactgga	ttatcgctc	2580
atagggtgt	ttcggcaggc	tcctggcaaa	gagcgtgagg	gggtctcatg	tatttagtagt	2640
agtgggtata	gcacaaat	cgccgat	gtaaaggggc	gggttacaac	ctccagggt	2700
aatgctaaga	acacggata	tctccat	aactctgt	agccccacg	taaggccgt	2760
tattactgt	cggttccat	ggcactat	tgccgcgt	ccctctgag	cccttacggc	2820
aggagcact	ggggcaaggg	gaccctgg	accgtatct	caggcgg	agggtctgg	2880
gggggggggt	cagggggtgg	aggcaggc	gtgcaact	ttgaatctgg	gggaggctt	2940
gtacaacctg	ggggatccc	gagactct	tgcgggccc	ceggattcac	cttggactac	3000
tatggcatcg	gtgttccg	ccagccccca	gggaaaggagc	gggaggcc	ttcatacat	3060
agtggccat	ccgggaccat	actgtacg	gactctgt	agggacg	taccatct	3120
aggggacaat	ccaaat	tgtgtac	ctgtgt	gctcaacg	ggaggat	3180
gcagtgtact	actggcgg	acggcgtt	tcgcgtt	gctgtatag	atggctggc	3240
gacgactacg	acgtgtgg	acggggcaca	cagggtgt	tgtcttccg	ttgcgaagc	3300
ggagggggca	gggggggtgg	gaggggtgg	ggcagccac	tgca	agagacaggg	3360
ggogggttag	ttcggcttgg	agggtctc	agactgt	gctgtc	tggcttacc	3420
ttca	gtact	atgggtcc	caagcttgc	ggaaaggggc	tgagtggatc	3480
gtctactt	atacatgt	cgacaaat	cgggac	ccacaaagg	gggg	3540
atttccag	acaacgca	gaatact	taccc	tgc	gatgtt	3600
gacactgtc	tgtactatt	tgcaagagg	cggtgtat	ctgttcc	tatcagagg	3660
gcagtaagg	gccttgg	acaggttac	gttcat	ggggaggccc	ttcaggccgt	3720
ggggggatct	gggggggtgg	cattaaat	cggttgc	aatccgggg	cggtgttgc	3780
cgacaggg	gtccctgt	gctctct	gcatctt	gaagcat	cggttccag	3840
accgtgac	gttctcg	ggctccccc	aagtct	atgcgttgc	ttccatgact	3900
aagactaaca	acagatct	ctctgact	gtgaaaggc	gttcat	ttctagagat	3960
aacgctaa	acacagt	tgcgtat	aatatct	aactct	gacaggcgt	4020
tatttctgt	agggtct	gctgagg	cgaggcat	aggtta	acgttgcgt	4080
tgaatcatgt	aatttagt	gtcac	ttttttt	ccat	ttttttt	4140
accggaaa	aaggaggt	acaacgtt	gtctgtt	ttttttt	ttttttt	4200
atgttagt	taagaaac	ttttttt	ttttttt	ttttttt	ttttttt	4260
gtgtacat	taacattat	atgtaaa	tttgc	ttttttt	ttttttt	4320
gtcttaattt	ggggccctc	acctgcac	aaaaag	tttgc	ttttttt	4380
ttttttttt	ttttttt	tttgc	ttttttt	ttttttt	ttttttt	4440
aacagaagg	agaacgaa	aaggagca	gactt	ttttttt	ttttttt	4500
gtgtgaaga	aacatgaa	ttccca	tttgc	ttttttt	ttttttt	4560
gcaggaaac	aagataat	atgtgaa	ctacata	ttttttt	ttttttt	4620
ctatgtt	gtgtcgca	ctttaat	tcatgc	ttttttt	ttttttt	4680
tttcattt	tttgc	tttgc	ttttttt	ttttttt	ttttttt	4740
aaattttttt	actaaaaa	catgtt	ttttttt	ttttttt	ttttttt	4800
ttaaggcc	aaaggcatt	tcgc	ttttttt	ttttttt	ttttttt	4860
tttgcgtat	ttgtat	gtcaat	ttttttt	ttttttt	ttttttt	4920
aatggcaga	cattacga	gcacacgg	ttttttt	ttttttt	ttttttt	4980
agcaggccg	agaagaag	aaaaaaac	ctagagg	tttgc	ttttttt	5040
catgcagg	ctccctat	actgg	ttttttt	ttttttt	ttttttt	5100
gegacaaa	ttttgtt	ggctt	cttca	ttttttt	ttttttt	5160
gttacgtt	gttggat	acacccgg	ttttttt	ttttttt	ttttttt	5220
gtcaacag	tagaacc	gtatgtt	ttttttt	ttttttt	ttttttt	5280
gaagaggact	atttgc	ggaagg	ttttttt	ttttttt	ttttttt	5340
caggctgg	agcatat	agaagat	ttttttt	ttttttt	ttttttt	5400
taaatgc	tatactaa	tcacaaat	gtatgtt	ttttttt	ttttttt	5460
ccacgtat	atccaat	aaaggaaat	atagcat	ttttttt	ttttttt	5520
gaggagtgc	agcatat	acagttaa	ggtgt	ttttttt	ttttttt	5580
gcatgaaat	ttttttt	ttttttt	ttttttt	ttttttt	ttttttt	5640
cgcatata	ta	atgtat	ttttttt	ttttttt	ttttttt	5700
atatacaggc	aaacacgc	ttatgttgc	ttttttt	ttttttt	ttttttt	5760
cgcggttgc	tttgc	ttttttt	ttttttt	ttttttt	ttttttt	5820
ttctatttct	tagaaatgt	aggaaactt	ttttttt	ttttttt	ttttttt	5880
agtgcactt	tcaaaaa	aaaaacgc	ttttttt	ttttttt	ttttttt	5940
aataccgtt	ccacaaat	tgctt	ttttttt	ttttttt	ttttttt	6000
ttccctatata	acttccat	ccac	ttttttt	ttttttt	ttttttt	6060
taatattttt	atgtttat	ctgtt	ttttttt	ttttttt	ttttttt	6120
attcata	tgatgc	aaaat	ttttttt	ttttttt	ttttttt	6180
gagacaaaat	agaagaa	tttgc	ttttttt	ttttttt	ttttttt	6240
cactttctgt	tcacaaat	tgcg	ttttttt	ttttttt	ttttttt	6300
tttatcttgc	aaaaatgc	ccgc	ttttttt	ttttttt	ttttttt	6360
tcaggcgtt	tttatgttgc	agaaaat	ttttttt	ttttttt	ttttttt	6420
gacctacat	gcaaaaa	atcaagag	tgcat	ttttttt	ttttttt	6480
agtaatctt	gatgtt	ttttttt	ttttttt	ttttttt	ttttttt	6540
aaaagaag	taatgtt	ttttttt	ttttttt	ttttttt	ttttttt	6600
aaaaatgc	tcagat	ttttttt	ttttttt	ttttttt	ttttttt	6660

---

-continued

```
tacaaaaatg aagcacagat tcttcgttgg taaaatagcg cttdcgcttgcattttgt 6720  
tctgtaaaaa tgcaactcag attttttttt tgaaaaatta gcgctctcgccgttcatttt 6780  
tgttctacaa aatagaagcac agatgctcg ttccagggtggc acttttcggg gaaatgtcg 6840  
ccgaaccctt atttttttat ttttctaaat acattcaaat atgtatccgc tcatgagaca 6900  
ataaccctga tattgg 6916
```

-continued

tcaaaaggatc	ttcttgcgat	ccttttttc	tgcgctgtat	ctgtgtcttg	caaacaaaaa	3960
aaccaccgc	accagcggt	gtttgttgc	cgatcaaga	gctacaact	cttttccga	4020
agttactgt	cttcagcaga	gcccacatac	caaatactgt	tctttactgt	tagccgtagt	4080
tagccccaca	cttcagaac	tctgttagac	cgcctacata	cctcgctctg	ctaatacgat	4140
taccagtgc	tgctgcggat	ggcgtataat	cgtgttttac	cgggtggac	tcaagacgtat	4200
agttaccgg	taaggcgca	cggtcggggt	gaacgggggg	ttcgtgcaca	cagcccgact	4260
tggaggeca	gacctacacc	gaactcgagat	acccatacgg	tgagatgtat	gaaacgcgca	4320
cgttccccc	aggggagaa	ggccgacagg	atccggtaag	cggcgggtc	ggaacaggag	4380
agcgcacgg	ggaggttcca	ggggggaaac	cctgttatct	ttatagatct	tgccgggttc	4440
gccacactcg	acttgagcgt	cgattttgt	gatgtcgctc	agggggcg	agectatgg	4500
aaaacgcgg	caacggggcc	tttttacggt	tcctggcc	ttgtggcct	tttgtccaca	4560
tgttcttc	tcgcgttacc	ctgttactg	ttgtataccg	tattacccgg	ttttagtggag	4620
ctgtatccgc	tcgggggtcg	cgacgtatgg	cttcggaaat	tttactactc	tttttacttc	4680
ttccagattt	tctcgactc	cgcgcatacg	cgttacactt	caaacaaccc	aagcacagca	4740
tactaaat	cccttcttcc	ttccctcttgc	gtgtgttac	ttaccggat	taaaagggtt	4800
gaaaagaaaa	aagtgcacgc	ctcggtttt	tttcttcgtc	gaaaaaggaa	ataaaaattt	4860
ttatcaacgt	tctttttttt	gaaaattttt	ttttttgtt	ttttttctt	tcgtatggct	4920
cccatgtata	ttaaagttaa	taaacggact	tcaatttctc	aagtttcagt	ttcatttttc	4980
ttgttcttat	acaatctttt	ttacttcttg	ctcatttagaa	agaaaggata	gcaatctaat	5040
ctaaatgtta	aatgagattt	ttcttcataat	ttactgtgt	ttatctgcga	gcataccccc	5100
cattaaatgc	tcctatgtca	actacaacag	aatgtatatt	aaaggggat	ttcgatgttt	5160
ctgttttgc	attttccggcc	agcattgtgt	ctaaagaaga	aggggtatct	ctcgagaaaa	5220
gagagggtca	agccatcgac	gtacagctgg	ttggagacggg	ggggagggtg	gtacaaaccag	5280
gccccgtact	gaggctttcc	ttgtccgcat	ctgggttcac	actggattat	tcgttccat	5340
tggtgtttcg	cgagggtctt	ggccaaaggg	gtgagggggt	ctatgtatt	agttagtgg	5400
gtgtatagca	aaagtacgac	gatttgcgtt	agggccgggt	tacaacctcc	agggataatg	5460
ctaagaacac	cgatataatc	cagatgaact	ctctgaagec	cgacatatacg	gcgcgtatatt	5520
actgttgcgg	tttcaggggc	actatgtgc	ttcggttccc	tctggccct	tafcggcaagg	5580
acgactgggg	caaggggggcc	cttgcgtaccc	tatcttcagg	cttgcggggg	ttgtgtgggg	5640
gagggtcagg	gggtgggggc	agccagggtc	aactgtgtt	atctggggga	ggttgtggat	5700
aacctgggg	atcccttgaga	ctcttcttgc	aggccctccg	attacccttg	gactactatg	5760
geatcggtt	gttccggccag	ccccccaggga	aggagggggg	ggccgttca	tatcttgcgt	5820
ccatgtccgc	gaccatactc	tafcgtacat	ctgtgttgg	acgttattacc	atcttcgtgg	5880
acaatgtccaa	aatatgtgt	taccttgcata	tgaacacgtt	caagccggag	gatacccgag	5940
tgtaactact	cgcgagacgg	cgcttccatcc	cttcttagct	gaatatgttg	cttgcgcgaat	6000
actacgtacgt	gtggggggcc	ggccacacagg	ttgtgtgtc	ttccgggtgc	ggaacgggg	6060
ggggcagccg	gggtgggggc	gggtggggcc	gccaatgtca	ctgtgttagat	acaggggggcc	6120
gtttatgttc	gtctgggggg	tctctcagac	tgtcatgtgc	tgccctgtgc	tttaccttca	6180
gtgactacgt	gtgcgtatgg	gttccggccaa	ctccaggggaa	gggggtctgt	ttggatcgta	6240
cttataatac	agatggcaga	acgtggccgg	acgttccac	aaaggggccg	ttccaccattt	6300
ccagggacaaa	cgccaaatgt	actctgtacc	ttcagatgtac	cagtgttaaa	ccccggagaca	6360
ctgtctgtat	ctattgtgc	agaggccggg	tgatctctgc	ttccgttatac	agaggcgacg	6420
taaggggccc	tggaaacacag	gtaaaccgtt	catcgggggg	aggccgttca	ggccgtgggg	6480
gatctgggg	gggtgtggatcc	caagggttgc	ttgtgttgcata	ccccggccga	ttgtgtccca	6540
caggggggcc	cttgcgggtcc	tcctgtgtatc	cttccggaaag	catccggccg	ttccggacgg	6600
tgacctgttc	tcggcaggat	ccccggaaat	ctctgtgtgt	ggtcgttcc	atgactaaaga	6660
ctaacaacga	gatctactt	gactcgtatg	aaggccgtt	catcttttct	agagataaa	6720
ctaaaaacaa	agtgtatctg	cgatgtatgt	gtctcaacc	tgaagacaca	gggtgttatt	6780
tgtgtatgtt	tcttcgtatc	aggggcccaat	qcataccat	aaactcttc	act	6833

```
SEQ ID NO: 71          moltype = DNA    length = 6821
FEATURE                Location/Qualifiers
misc_feature           ..6821
note = Plasmid pD1214-AK-hABAB
source                 1..6821
mol_type = other DNA
organism = synthetic construct
```

SEQUENCE: 71 Organism = synthetic construct

ggtaaatca	tgtaaattagt	tatgtcacgc	ttacattcac	gccctcccc	cacatccgt	60
ctaaccgaaa	aggaaggagt	tagacaacct	gaagtcttagg	tcccttatta	ttttttata	120
gttattgttag	tattaagaac	gttattata	tttcaaaattt	ttctttttt	tctgtacaga	180
cgcgttacg	catgttaaacat	tatactgaaa	accttgcttg	agaaggttt	gggcgcgtc	240
aaggcttaa	tttgcggccc	ctcacccgtca	cgcaaaaagc	tttcaattt	aatttcattat	300
tttttttta	ttctttttt	tgatttgcgt	ttctttgaaa	ttttttgtat	tcggtaatct	360
ccgaacacaaa	ggaagaacga	aggaaggagc	acagacttag	atttgtatata	atacgcatat	420
gttagtggta	agaaacatca	attttcccg	tatttttaac	ccaaactgcac	agaacaaaaaa	480
ccacggaaa	acgaagataaa	atatgtccg	aaatgtacata	taaggaaacgt	gtgtctactt	540
atcttagtcc	tggtgtggcc	aaatgtat	atatcatgca	cgaaaagcaa	acaaacttgt	600
gtgcttcatt	ggatgttctgt	accaccaagg	aattactgg	gttagttgaa	gcatttaggtc	660
ccaaaatttg	tttactaaaa	acacatgtgg	atatcttgc	tgattttcc	atggaggggca	720
cagttaaggc	gctaaaggca	ttatccggca	agttacaattt	tttactctt	gaagatagata	780
aatttgcgt	cattgttaat	acagtcaat	tgcaactgtc	tgccgggtgt	tacagataat	840
cagaatggcc	agacatcaag	aatgcacacg	gtgtgtgtgg	cccaaggttt	gttagcgtt	900
tgaaggcaggc	ggcagaagaa	gtaacaaagg	aaccttagagg	cctttgtat	tttagcagaat	960
tgtcatgc	gggtccctta	tctactggag	aatataactaa	gggtactgtt	gacattgcga	1020
aaacgcacaaa	agattttgtt	atcggttta	ttgtctaaag	agacatgggt	ggagagatgt	1080
aagggttacg	ttgggttgtt	atgcacacccg	gtgtgtgttt	agatgacaaag	ggatgtcgt	1140
tgggtcaaca	gtatagaaccc	gtggatgtat	ttgtcttac	aggatctgc	attttatttt	1200
tttgaagagg	actatttgc	aaggaaagg	atgtcaaggt	agagggtgtt	cgttacagaaa	1260

-continued

---

aagcaggctg	ggaagcatat	tttggagaat	ggggccagca	aaactaaaaa	actgtattat	1320
aagtaaatgc	atgtatacta	aactcacaaa	tttagagcttc	aatttaataat	tatcgattat	1380
taccacacgt	atgtatccaat	atcaaaggaa	atgatagcat	tgaaggatga	gactaatcca	1440
attgaggagt	ggcagcatat	agaacagcta	aagggtatgt	ctgaaggaaag	catacgatac	1500
cccgcatgga	atgggataat	atcaggaggag	gtactagact	accccccattc	ctcataataat	1560
agacgcataat	aagtacgcat	ttaaagcataa	acacgcacta	tgcgggttctt	ctcatgtata	1620
tatataataca	ggcaaacacgc	agatataggt	gccccgtgaa	cagtgcgtg	tatgtgcgca	1680
gctcgcttg	cattttcgga	agcgcgtt	ttcggaaacg	cttggaaagtt	cctattccga	1740
agtccattt	ctctagaaag	tataggaat	tcagagccgt	tttggaaaacc	aaaagcgctc	1800
tgaagtgcga	cttccaaaaa	accaaaaaac	cacccggactg	atacgagct	ctaaatattat	1860
ggeaaataccg	cttccaaaaa	cattgtctaa	aagtatctct	ttgtctatata	tctctgtgt	1920
atatccctat	ataacccattc	cateccaccc	tegtccttg	aacttgcac	taaactcgac	1980
ctctacattt	ttttagttat	ttctctatgt	tactctttat	acaaaaaaat	tgtgtaaag	2040
actattcata	gagtgtatcg	aaaacaatac	gaaaatgtaa	acattttccat	tacgtagtat	2100
atagagacaa	aatagaagaa	accgttcata	attttctgac	caatggaaagaa	tcatcaacgc	2160
tatcacttc	tggtcacaat	gtatgcgaa	tccacatcg	tatagaatat	aatcgggat	2220
gcctttatct	tgaaaaaaatg	cacccggcago	ttcgcgtatg	atcgatataac	gccccggatg	2280
gagtcaggct	ttttttatgg	aagagaaaaat	agacacccaa	gtagccttct	tctaaacett	2340
acggacccat	agtgcacaaa	gttataaqua	gactgcattt	tagagccac	aaaggagaaaa	2400
aaaagtaatc	taagatgctt	ttgttagaaaa	atagcgtct	cgggatgc	ttttgttagaa	2460
caaaaaaaagaa	gtatagatcc	tttgggtgg	aaatagcgct	ctcgcggttc	atttctgttc	2520
tgtaaaaaatg	cagtcgtat	tctttgtt	aaaaattatgc	gtctcggt	tgcattttt	2580
ttttacaaaa	atagaacaca	gatttttgt	ttgttaaaaata	gccccccgc	tttgcattt	2640
tgttctgtta	aatgcagct	cagattctt	ttttggaaaa	ttagcgtct	cgccgtgc	2700
ttttgttcta	aaaaatgaag	cacagatgt	tcgttgcagg	ggcacccccc	ggggaaatgt	2760
gccccggaaacc	ccttattttgt	tatttttcta	aatacatccaa	aatatgtatc	cgccatgtag	2820
acaataaacc	tgatattgtt	caccaatgtt	taattttgtt	taacactgac	cccttatttt	2880
ttatttttct	aaatacattc	aaatatgtat	ccgctcatga	gacaataacc	ctgataaaatg	2940
cttcaataat	attgaaaaaaag	gaagaatatgt	agttatcaac	atttccgtt	cgcccttatt	3000
cccttttttgc	ccgcatttttgc	ccttcttgc	tttgcctacc	cagaaacgc	gggtaaagta	3060
aaagatgtcg	aagatcgat	gggtgcacga	ttgggttaca	tgcgcactgaa	tctcaacgc	3120
ggttaatgtcc	ttgagatgtt	tcgcggcga	gaacgggttc	caatgtatgc	cacttttaaa	3180
gttctgtat	gtggcgcgg	attatccgt	ttgacgcgg	ggcaagagca	actcggtcgc	3240
cgccatcaact	atttcgtat	tgacttgcgtt	ggtactcact	cagtccacaga	aaagcatctt	3300
acggatggca	tgacagtaaq	aaatgtatgc	agtgctgc	taaccatgag	tgataaact	3360
gcccccaact	tactctgac	aacgatcgga	ggacccgaagg	agctaaccgc	tttttgcac	3420
aacatggggg	atcatgttaac	tcgccttgc	cggtggaaac	cgaggatgt	tgaaggccata	3480
ccaaacgcacg	acgcgtacac	cacgtgc	gtacgcgtat	caacaacgtt	gccccaaact	3540
ttaacttgcg	aactacttac	tctatgttgc	cgccatcaat	taatagactg	gatggaggcg	3600
gataaaatgt	caggaccact	tctgcgtcg	ggcccccgg	ctggctgtt	tattgtctgt	3660
aaatccggag	ccgggtgagcg	ttgttctcg	ggtatcatcg	cagccgttgg	gccagatgg	3720
aaagcccttc	gtatcgatgt	tatccatccac	acggggagatc	aggccactat	ggatgtacgc	3780
aatagacaga	tcgcgtatgt	agggtgcata	ctgttgcata	atgggtact	catgacccaaa	3840
atcccttaac	gtgagttacg	cgccgtcg	tccactgtac	gtcagacccc	gtaaaaaaaga	3900
tcaaaggatc	ttcttgatgt	ccttttttgc	tgcgctaat	ctgtgtcttgc	caaacaaaaa	3960
aaccacccgt	accaggdtgt	tttttttgc	cggtatcaaga	gttaccaact	ctttttccga	4020
aggttaatcc	tttcgcaga	gcgcagata	caatatactgt	tcttctatgt	tagccgtat	4080
tagccacca	tttcaagaa	tctgtatgc	cgccatcata	cctcgctct	ctaattctgt	4140
taccagtggc	tgctgcgt	ggcgataat	cgtgttcttac	cggtttggac	tcaagacat	4200
agttacccgg	taaggcgcac	cggttgcgg	gaacgggggg	ttcgtgcaca	cagcccaat	4260
tggagcgaac	gacccatcacc	gaactgtat	acccatccac	tgatgtatgc	gaaaggccca	4320
cgttcccgaa	aggggaaaaag	ggccacgggt	atccgttaag	eggcagggtc	ggaacaggag	4380
agcgccacg	ggagcttcca	ggggggaaac	cctggatctt	ttatagtct	gtcggtt	4440
ccacacccat	atttgcgt	cgatgtttgt	gtatgtcg	agggggggcc	agccatattg	4500
aaaacgcgcac	caacgcggcc	tttttacgtt	tcctggcctt	ttgtgtggct	tttgcataca	4560
tgttcttcc	tgcgttattcc	cctgttctgt	ttggataacc	tattttcc	tttggatgt	4620
ctgtatccgc	tcgggttgcgt	cgaggatata	tttctacttcc	ttttttactc	4680	
ttccagattt	tccgcgtact	cgccgcata	cgatccactt	aaaaacaccc	aagcacagca	4740
tactaaatcc	cccccttttgc	ttccctatgt	gtgtcgat	tttccatccat	ttaaagggtt	4800
gaaaagaaaa	aaatgtacccgc	ctcggttctt	ttttttcg	ttttttcg	ttttttatgt	4860
ttatcacgtt	tcttttttgc	gaaaattttt	ttttttatgt	ttttttatgt	ttcgatgtat	4920
cccaattata	ttaaagttaa	taaaaggact	tcaatttctc	aaatgtatgt	tttcattttt	4980
ttttttctt	acaacttctt	ttacttcttgc	tcattttatgt	aaaaaaggat	gtcaatctat	5040
ctaaatgtttaa	aatgtatgtt	ccttcaattt	ttactgtgt	tttattgtca	gtatccctcg	5100
cattatgtgc	tccagtcac	actacaacag	aaatgtatgt	agaagggtat	ttcgatgtt	5160
ctgttttgc	attttgcgc	agcattgtgt	tcttgcata	ttttttatgt	ttttttatgt	5220
gaatgcgtt	acagctgtgt	gaggggttttt	ttttttatgt	ttttttatgt	ttttttatgt	5280
ggttttctgt	tcgcgcata	gggttccac	tggatatttc	gtccatagg	ttgtttcg	5340
aggcttctgg	caaagacgt	gggggggtt	catgtttagat	ttgtgtgtt	gtatgtat	5400
agtaacgcgt	ttccgtatgt	ggccgggtt	caacccatcc	ggataatgt	aagaacaccc	5460
tataatcttca	gtatgtatgt	ctgatgtatgt	acgatcgc	ctgtatattac	tgtgcgtt	5520
tcaggggcact	tatgtgcggc	gtgttccctt	ttggccat	ttttttatgt	ttttttatgt	5580
agggggaccct	gggtgacgt	tcctcaggcg	ttgggggtt	ttgtgggggg	ttgtgggggg	5640
gtggaggccat	ccagggtgc	ctgggttgc	ttgggggtt	ttgtggat	ttgtggat	5700
ccctggatct	tccttgcgt	ggcccccgtt	ttttttatgt	ttttttatgt	ttttttatgt	5760
tccggccagg	cccaggaa	gagccgggg	ccgtttcata	tattttatgt	ttttttatgt	5820
ccataactgt	cgccatgtt	gtgaagggg	ttttttatgt	ttttttatgt	ttttttatgt	5880
atgtgtgt	cttgcataat	aacgcctca	ttccgtatgt	ttttttatgt	ttttttatgt	5940
cgagacggcg	tttcccgct	tctagcgt	ttttttatgt	ttttttatgt	ttttttatgt	6000

-continued

---

```

ggggacgggg cacacagggt gctgtgttcc cgggtgggg aagcggggg ggcaacgggg 6060
gtgggagcg tgggggcagc caactgcagc tggtagagac agggggcgcc ttatgttac 6120
ctggagggtc tctcaagactc tcatgcgtc cctctggct taccttcagt gactacgtga 6180
tgacatgggt ccgccaagct ccaggaaagg ggcctgagtg gatecgtaact attaataacag 6240
atggcagcac aatgcggggac gactccacaa agggggcggtt caccattttc agagacaacg 6300
ccaagaatac tctgtaccc catgtacca gctgtgaaacc cgaggacact gctctgtact 6360
atttgtcaag aggccgggtg atctctgtt ccgttatcag aggcgcagta agggggcctg 6420
gaacacaggta aaccgtttca tccggggag gcggttcagg cggtggggga tctggcgccc 6480
gtggatccca agttcagctg gtcaatccg ggggcggact ggtccagaca gggggctccc 6540
tgaggcttc ctgtgtacat tccggaaagca tccggcggtt cgaggacgtt acctgggttc 6600
gcaggctcc cgggaaagtt ctgcgtgttccat gactaaagct aacaacgaga 6660
tctactctga ctcaagtaaa ggccgcttca tcatttctag agataacgct aaaaacacag 6720
tgtatctca gatgaatagt ctcaaacctg aqacacaggc cggttatttc tgtaagggtc 6780
ctgagcttag gggccaggc atccaggtaa cagtctcgag t 6821

```

```

SEQ ID NO: 72 moltype = DNA length = 6707
FEATURE Location/Qualifiers
misc_feature 1..6707
note = Plasmid pD1214-AT-hABAB
source 1..6707
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 72

```

```

ggtaaatca tgtaattatgt tatgtcacgc ttacattcac gcccctcccc cacatccgt 60
cttaaccgaaa aggaaggagt tagacaacctt gaagtctagg tccctattta ttttttata 120
gttatgttag tattaaggaaac gttttttttt ttccaaattt ttctttttt tctgtacaga 180
cgcgtgtacg catgtacat tataactgaaa accttgcgtt agaaggttt gggacgcctg 240
aaggctttaa ttgcggccc ctcacctgtca cgcaaaaagg ttttcaattt aatttcatcat 300
ttttttttt ttctttttt tgatgttccgtt ttcttttataa ttttttttgc tccgttattt 360
ccgaacagaa ggaagaacgaa aggaaggaggc acagacttagt atgggtatattt atacgcatt 420
gtatgttgc agaaacatgt aatttgcggcc tatttcttaac ccaactgcac agaacaaaaa 480
ccacggcggaa acgaagataa atcatgtcga aagctacata taagggacgt gctgtactc 540
atccctgtcc ttgtgtcgcc aagcttattt atatcatgtca cgaaaaggca acaaacttgt 600
gtgttccattt ggtatgttccgtt accaccaagg aattactgtca gttatgttgc gcatgttgc 660
ccaaatttt tttactaaaaa acacatgtgg atatcttgc tgatttttcc atggaggcga 720
cagtttaagcc gctaaaggca ttatccggca agtacaattt ttacttcttca gaagatagaa 780
aatttgcgtca cattgtgtat acagtcaat tgcgttactc tgcgggtgtt tacagaatag 840
cagaatggcc agacatttgcg aatgcacacgg gtgtgggtggg cccaggattt gtttgcgtt 900
tgaaggcggc ggcagaagaa gtaacaaggaa aaccttaggg ctttttgcgtt ttagcgaat 960
tgcgtatgtca gggcttccat tctacttgcgtt aatatactaa gggtaactgtt gacattgcga 1020
aaagcgcacaa agatttgtt atccgttta ttgtcaaaag agacatgggtt ggaagagatg 1080
aagggttacgg tttgtgttgcgtt atgacacccgg gtgtgggtttt agatgacaaag ggagatgcat 1140
tgggtcaaca attagataacc gttgtgtatgc tttgtcttac aggtatgtcatttatttgc 1200
tttggaaagg actatttgcgtt aagggaaagggtt atgctaagggtt aggggttgcgtt ctttttgcgtt 1260
aagcaggctg ggaagcatat ttgagaagat gcccggcggca aaactaaaaa actgttattt 1320
aaatgtatgtc atgtatacta aacttccat ttagacttcc aattttatttata tttttttttt 1380
tacccacgtt atgatccat atcaaaagggtt atgatagcat tgaaggatgtt gactaatcca 1440
atttggggatgtt ggcacgtt atgacacccgg tttttttttt tttttttttt tttttttttt 1500
cccgcatggaa atgggataat atcacagggtt gtacttagactt accttttccatc ctatataat 1560
agacacgttcaat aatgtatgtca ttaacgttcaat acacgcactt tgcgttccat tttttttttt 1620
tatataatataca ggcacacggc agatgttgcgtt ggcacgttcaat tttttttttt 1680
gttcgcgttgc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1740
agtttcttattt ctcttagaaatg tttttttttt tttttttttt tttttttttt tttttttttt 1800
tgaatgtcgca ctttccaaaaa accaaaaaaaggc caccggacttgc taacgttcaat tttttttttt 1860
gcgaataccgg ctttccaaaaa cattgttcaat aatgtatgtca tttttttttt tttttttttt 1920
atatccctat ataaaccttccat tttttttttt tttttttttt tttttttttt tttttttttt 1980
ctctacattt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2040
actatttccat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2100
atagagacaa attagaaagaa accgttccat tttttttttt tttttttttt tttttttttt 2160
tatcaacttccat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2220
gcctttatct tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2280
gagtccggct tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2340
acggacccatc agtgcacaaa gtttatcaaga gactgcatttca tagacgttccat tttttttttt 2400
aaaatgtatgtca tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2460
caaaaaaaaggaa gttatgttccat tttttttttt tttttttttt tttttttttt tttttttttt 2520
tgtttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2580
ttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2640
tgtttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2700
ttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2760
gcgcggaaacc ctttccat tttttttttt tttttttttt tttttttttt tttttttttt 2820
acaataaccc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2880
ttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2940
cttcaataat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3000
cccttttttgc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3060
aaagatgttcaat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3120
ggtaatgttccat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3180
gtttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3240
cgcataactt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3300
acggatggca tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3360

```

-continued

---

gggcccaact tacttctgac aacgatcgga ggaccgaaagg agctaaccgc tttttgcac 3420  
 aacatgggg atcatgtaac tcgccttgat cggtggaaac cggagctgaa tgaagccata 3480  
 ccaaacgacg agcgtgacac cacgatgct gtacgatgg caacaacgtt gcgcaacta 3540  
 ttaactggcg aactacttac tetagttcc cgccaacaat taatagactg gatggaggcg 3600  
 gataaaagtgc caggaccact tctgcgtcg gcccttcgg ctggctggg tattgtctgat 3660  
 aaatccggag ccggtgagcg tggttctcg ggtatcatcg cagcgtggg gccagatgg 3720  
 aagccctccc gtatcgtagt tatctacagc acggggagtc aggcaactat ggatgaaacga 3780  
 aatagacaga tcgctgagat aggtgctca ctgattaagec attggtaact catgacaaa 3840  
 atcccttaac gtggatgtac cgccgtcg tccactgagc gtcagaccc gtagaaaaaga 3900  
 tcaaaaggatc ttcttgat ctttttttc tgcgtgtaat ctgtgtcg taaacaaaaaa 3960  
 aaccaccgc accagccgtg gtttttgc cggtatcaaga gtcataactt cttttccga 4020  
 aggttaactgg cttcagcaga ggcagatad caaatactgt tcttctagtg tagccgtag 4080  
 tagccctacca cttaaaagca tctgtacagc cgccatatac ctcgtctcg ctaatccgt 4140  
 taccatggc tgctgtcgat ggcataatg ctgtgtttac cgggtggac tcaagacat 4200  
 agttaccggta taaggcgcag cggccgggtt gaaacggggggg ttctgtgcaca cagcccgat 4260  
 tggagcgaac gacctacacc gaactgagat acctacagcg tgtagctatga gaaagccca 4320  
 cgcttcccgaa aggggagaaacg gccggacaggt atccggtaaaatg cggcagggtc ggaacaggag 4380  
 agegcacgag gggatctca gggggaaacg cttggatatct ttatagtctt gtcgggttc 4440  
 gccacccctcg acttgcgtg cgtatgtt gatgtctcg aggggggccg agccatatgg 4500  
 aaaacgcgcg caacgcggcc tttttacggt tcttgcctt ttgtggcct tttgtcaca 4560  
 ttttcttcc tgcgttatec cctgttctg tggataacccg tattacccg tttgagtgag 4620  
 ctgtatccgc tccgggtcgat gcaggatatac ttccaaatgt ttctactcc tttttactc 4680  
 ttccaggat ttcggactc cgccatcg cgttacactt caaaacaccc aagcacagca 4740  
 tactaaattt ccccttttc ttctcttagg gtgtcgtaa ttacccgtac taaaggttt 4800  
 gaaaagaaaa aagtgaccgc ctgcgtttttt ttttttcgtc gaaaaggca ataaaaattt 4860  
 ttatcacgtt tttttttttt gaaaattttt ttttttgatt tttttctt tcgtgaccc 4920  
 cccattgata ttaaattttttaaaacggact tcaattttca aagtttcaatg ttcattttc 4980  
 ttgttctatt acaactttt ttacttcttg ctcatagaa agaaagcata gcaatcta 5040  
 ctaaagttaa aatgagatttt cttcaattt ttactgtgtt ttatttcgca gcatctccg 5100  
 cattaggtat gcaaggatcag ctgggggaga cggggggggg gttgttacaa ccaggccgggt 5160  
 cactggggct tccctgtcc gcatctgggtt tccactggta ttatctgtcc atagggtgg 5220  
 ttccggcggc tccctggcaaa gagcggtgagg gggtctcatg tatttagtgc atgggtgata 5280  
 gcacaaaatgat cgccgatcc gtaaaaggcc ggtttacaccc ctccagggtat aatgtcta 5340  
 acaccgtat tccctggatg aactctgtca agccccgtca tacggccgttattactgt 5400  
 cggcttccatggcactatg tccctgtgt tccctctgat ggttgcgttcc aaggacact 5460  
 ggggcaaggcc gaccctgggtt accgtatctt caggccgtgg agggtctggt gggggagggt 5520  
 caggggggtgg aggcagccag gtgcacttggg ttgaatctgg gggaggctg gtacaactg 5580  
 ggggatccctt gggatctctt tccggggatcc cccggatttcc ctgggactac tatggcatcg 5640  
 gctgggttccg ccaggccca gggaggagcc gggaggccgtt ttcatcattt atgtccatgt 5700  
 cccggaccat actgtacgc gactctgtca agggacgtt taccatctt agggacacat 5760  
 caaaaatgc tggatcttgc caaatgaaaca gctcaacgc ggaggatacc gcaatgtact 5820  
 actgcggcg acggccgttcc tccgttccat gctgtatggc atggctggcc gacgactacg 5880  
 acgtgtgggg acggcccaaa cagggtgtc tggatctccg tggcggaaac gggggggca 5940  
 ggggggggtgg ggggggtgg ggcacggccaa tggatctggt agagacaggg gggggcttag 6000  
 ttcagccctgg aggggtcttc agactgtcat gctgtccctt tggcttacc ttcaatgt 6060  
 acgtgtatgac atgggtccgc caaggtccca gggggggcc tggatctggc gctactattt 6120  
 atacatggc cggccatgtt cggggactt ccacaaagggg ggggttacc atttccagag 6180  
 acaacgcacaat gatactctt tccatcaga tggatcttgc tggatcttgc gacactgtc 6240  
 tggatcttgc tggatcttgc tggatcttgc tggatcttgc tggatcttgc gacactgtc 6300  
 gcccgttgcac acggatcaac gtttcatcc gggggggggg ttcaggccgtt gggggatctg 6360  
 ggggggggtgg atcccaatgtt cggatcttgc aatccgggggggggggggggggggggg 6420  
 gtcctctgtt gatcttgc tggatcttgc gacactgtc tggatcttgc tggatcttgc 6480  
 ggtctcgccaa ggttcccgat gatcttgc tggatcttgc tggatcttgc tggatcttgc 6540  
 acggatcttgc ttcgtactca gtttgcgtt ttcatcattt aatgttgc ttttgcgtt 6600  
 acacagtgttca ttcgtactca aatgttgc ttttgcgtt ttcatcattt aatgttgc 6660  
 aggggtctggt gtttgcgtt ttcatcattt aatgttgc ttttgcgtt ttcatcattt aatgttgc 6707

SEQ ID NO: 73 moltype = DNA length = 6710  
 FEATURE Location/Qualifiers  
 misc\_feature 1..6710  
 note = Plasmid pD1214-AA-hABAB  
 source 1..6710  
 mol\_type = other DNA  
 organism = synthetic construct

SEQUENCE: 73  
 ggtaaatca tgtaattatgt tatgtcacgc ttacattcac gcccctcccc cacatccgt 60  
 cttaaccgaaa aggaaggagt tagacaacat gaagtctagg tccctattta tttttttata 120  
 gttatgttag tattaagaac gttatattt tttcaattttt tctgtacaga 180  
 cgcgtgtacg catgtacat tataactgaa accttgcgtt agaagggttt gggacgcctcg 240  
 aaggctttaa ttgcggccc ctcacctgc caaaaaaggcc ttttcaattt aatttcatcat 300  
 tttttttttt ttcctttttt tgatttgcgtt ttctttgaa ttttttgtat tcggtaatct 360  
 ccgaacacaa ggaaggacgc aggaaggaggc acagacttgc atgggttat ataccgtat 420  
 gtagtgttgc agaaacatgc aatttgcgtt tattcttacac ccaactgcac agaacaaaaa 480  
 ccacggccaa acggatataa atcatgtca aagcttacata taaggaaatgt gctgtactc 540  
 atccctgttcc tggatcttgc aatgttgc ttttgcgtt ttcatcattt aatgttgc 600  
 gttatgttagt gatcttgc ttcatcattt aatgttgc ttttgcgtt ttcatcattt aatgttgc 660  
 cttttttttt ttcctttttt tgatttgcgtt ttctttgaa ttttttgtat tcggtaatct 720  
 cttttttttt ttcctttttt tgatttgcgtt ttctttgaa ttttttgtat tcggtaatct 780  
 aatgttgc ttttgcgtt ttcatcattt aatgttgc ttttgcgtt ttcatcattt aatgttgc 840

-continued

---

cagaatgggc agacattacg aatgcacaa	gtgtggtggg cccaggattt gtttagcggtt	900
tgaagcaggc ggcagaagaa gtaacaagg aaccttagagg	ccttttgatg ttagcagaat	960
tgtcatgca gggctcccta tctactggag aataactaa	gggtactgtt gacattgcga	1020
aaagcgacaa agatttgtt atcggttta ttgctcaag	agacatgggt ggaagagatg	1080
aagggtacga ttgggttatt atgacaccgg gtgtgggtt	atgatgacaag ggagatgcat	1140
tgggtcaaca gtatagaacc gtggatgat ttgtctct	aggatgtgc attattattg	1200
ttggaagagg actatttgc aagggaaaggg atgctaagg	agaggggtgaa cgttacagaa	1260
aaggcaggctg ggaagcatat ttgagaagat	gcccgcacgaa aaactaaaaa actgtattat	1320
aagtaatatgc atgtatacta aactcacaaa tttagacgttc	aatttaatata tatcagtat	1380
tacccacgcg atgatccat atcaaaggag atgatagcat	tgaaggatg gactaatcca	1440
attggaggatg ggcacatata aacacgcgta aagggtatgt	ctgagggaaag catacgat	1500
cccgcatgga atgggataat atcacaggag gtactagact	acctttcatc ctacataat	1560
agacgcataat aagtacgcataa acacgcacta	tgccgttctt ctcatgtata	1620
tatataataca ggcacacgc agatataagggt	gacgtgtt gctgtcgca	1680
gtctcgcttg cattttcgga agcgcgtt ttccggaaacg	catttggaaaccc ctttggaaaccc	1740
agttcttatt ctctagaaag tataggaact tcagacgcgt	aaaaggcgctc	1800
tgaagtgcga ctttcaaaa accaaaaacc caccggactg	taacgagctt ctaaaatatt	1860
gcaaataccg ctccccaaa cattgtcaaa aagtatctc	ttgttatata tctctgtgt	1920
atatccctat ataacctacc catccaccc ttgcgttcc	aacttcgcata taaactcgac	1980
ctctacattt tttatgttta tctctagtat tactctttag	acaaaaaaaat ttttggatgt	2040
actatttcata gagtgaatcg aaaacaaatc gaaaatgtaa	acattttctc tacgtat	2100
atagagacaa aatagaagaa accgttccata attttgcac	tttttgcacat gatgttgc	2160
tatcacttcc ttgtccaaaat gatgtcgcaat ccacatcg	tatagaatata aatcgggat	2220
gcctttatct tgaaaaatg caccggcago ttgcgttaga	atcagtaaac gcggaaatgt	2280
gagtccggct tttttatgg aagagaaaaa agacacccaaa	gtacgcgttct tctaaaccc	2340
acgggactac agtgcacaaa gttatcaaga gactgcattt	tagagegcac aaaggagaaa	2400
aaaaggtaaatc taatgtgtt ttgtggaaa atagcgctt	cgggatgtat ttttgtgaa	2460
caaaaaaaaaaa gtatagattt ttttgtggta aaatagcgct	ctcgctgttgc attttgttgc	2520
tgtaaaaatg cagctcagat ttcttttttgg aaaaatttgc	gtctcgctgt tgcatttttgc	2580
tttttacaaa atgaacgcata gattttccgt	ttgttacaaaat ggcgtttccgc gttgcattttgc	2640
ttgttctgtttaaa attagtgcgtt caggatctt	ttttggaaaat ttagcgctt cgcgttgc	2700
tttttgttctt caaaaatgaag cacagatgt	ttgttccgttgc ggcacttttgc gggaaatgt	2760
gccccggaaacc ctattttgtt tatttttctaa aatacattca	aatatgttacgc cgctcatgt	2820
acaaataacc ttttttttttgc cagaattttgtt	taatttgttgc taacactgc ccctattttgt	2880
ttattttttca aataatgtat ccgcgtatgc gacaatacc	tttttttttgc tgcataat	2940
cttcaataat atgaaaaatg gaagaatatg agtatttcaac	tttttttttgc tgcataat	3000
cccttttttgc cggcatttttgc ctttctgtt ttttgttgc	tttttttttgc tgcataat	3060
aaagatgttca aagatcgatgtt ggggtgcacga	tttttttttgc tgcataat	3120
ggtaagatcc ttggatgttttgc cggcatttttgc	tttttttttgc tgcataat	3180
tttttttttttttgc ttttttttttgc ttttttttttgc	tttttttttgc tgcataat	3240
cgcatacact atttcagaa tgactttgtt ggtacttccac	tttttttttgc tgcataat	3300
acggatggca tggacgttgc agaaattatgc agtgcgttccac	tttttttttgc tgcataat	3360
ggggcccaact tactttgtac aacgtatgtt ggggggggg	tttttttttgc tgcataat	3420
aacatggggg atcatgttac tcgcgttgcgtt ctttttttttgc	tttttttttgc tgcataat	3480
ccaaacgacg acgctgttgcgtt ctttttttttgc tgcataat	tttttttttgc tgcataat	3540
tttaacttgcgc aactacttac ttttttttttgc tgcataat	tttttttttgc tgcataat	3600
gataaaatggc cggggatccat ttttttttttgc tgcataat	tttttttttgc tgcataat	3660
aaatccggat ctttttttttgc tgcataat	tttttttttgc tgcataat	3720
aaggccctccc gtatcgatgtt ttttttttttgc tgcataat	tttttttttgc tgcataat	3780
aatagacaga ttttttttttgc tgcataat	tttttttttgc tgcataat	3840
atcccttaac ttttttttttgc tgcataat	tttttttttgc tgcataat	3900
tcaaggatc ttttttttttgc tgcataat	tttttttttgc tgcataat	3960
aaccaccgtt accagggttgc ttttttttttgc tgcataat	tttttttttgc tgcataat	4020
aggtaacttgc ttccgtcaga ggcgcataat ttttttttttgc tgcataat	tttttttttgc tgcataat	4080
tagcccccaatc ttttttttttgc tgcataat	tttttttttgc tgcataat	4140
taccaggatgc ttttttttttgc tgcataat	tttttttttgc tgcataat	4200
agttaccggta aaggcgcag ctttttttttgc tgcataat	tttttttttgc tgcataat	4260
tggagcgaaac gacccatcacc ttttttttttgc tgcataat	tttttttttgc tgcataat	4320
cgccgttccgc aggggagaaatggc ctttttttttgc tgcataat	tttttttttgc tgcataat	4380
agccgcacgcg ggatgttcaat gggggaaacgc ttttttttttgc tgcataat	tttttttttgc tgcataat	4440
gccacccctgc acttgcgttgc ttttttttttgc tgcataat	tttttttttgc tgcataat	4500
aaaaacgcgcg caacgcggcc ttttttttttgc tgcataat	tttttttttgc tgcataat	4560
tgttctttttc ttttttttttgc tgcataat	tttttttttgc tgcataat	4620
ctgtatccgc ttttttttttgc tgcataat	tttttttttgc tgcataat	4680
ttccagattt ttttttttttgc tgcataat	tttttttttgc tgcataat	4740
tactaaatcccttcc ttttttttttgc tgcataat	tttttttttgc tgcataat	4800
aaaaaaaatggc aacttgcgttgc ttttttttttgc tgcataat	tttttttttgc tgcataat	4860
ttatcagttt ttttttttttgc tgcataat	tttttttttgc tgcataat	4920
cccaattgtata ttttttttttgc tgcataat	tttttttttgc tgcataat	4980
ttgttcttattt ttttttttttgc tgcataat	tttttttttgc tgcataat	5040
ctaagtttaa attggttgc ttttttttttgc tgcataat	tttttttttgc tgcataat	5100
ctgcatttgcg aatgcgttgc ttttttttttgc tgcataat	tttttttttgc tgcataat	5160
ggtcacttgcg gtttccgttgc ttttttttttgc tgcataat	tttttttttgc tgcataat	5220
ggtttcggca ggcgttccgc ttttttttttgc tgcataat	tttttttttgc tgcataat	5280
atagcacaatc ttttttttttgc tgcataat	tttttttttgc tgcataat	5340
agaacaccgtt atatctccat ttttttttttgc tgcataat	tttttttttgc tgcataat	5400
gtgcggctt caggccatcacttgc ttttttttttgc tgcataat	tttttttttgc tgcataat	5460
actggggccaa gggggccatcacttgc ttttttttttgc tgcataat	tttttttttgc tgcataat	5520
gtcagggggg tggggccatcacttgc ttttttttttgc tgcataat	tttttttttgc tgcataat	5580

-continued

---

ctgggggatc	cctgagactc	tcttgcgagg	cctccggatt	cacccggac	tactatggca	5640
tcggctgtt	ccggcagccc	ccagggagg	agcgggggc	cgtttcatac	attatggcca	5700
gtccccggac	catactgtac	gcagactctg	tgaagggagc	cttaccatc	tctagggaca	5760
atgccaaaaa	tgctgtgtac	ctgcaaatga	acagcctaa	gccccggat	accgcagtgt	5820
actactgcgc	gagacggcgc	ttctcccggt	ctagcgtgaa	tagatggctg	ggcgcacact	5880
acgacgtgtg	gggacggggc	acacagggtg	ctgtgtcttc	cgggtgggga	agcggggggg	5940
gcagcggggg	ttggagcgggt	ggggcagcc	aactgcacgt	ggtagagaca	ggggggggct	6000
tagttcagcc	ttggagggtct	ctcagactgt	catgcgtgc	ctctggctt	acccctca	6060
actacgtgt	gacatgggtc	cgccaaagtc	cagggaaagg	gcctgagttg	atcgctacta	6120
ttaatacaca	ttggcagcaca	atggcggacg	actccacaaa	ggggcgggtc	accatttcca	6180
gagacaacgc	caagaatact	ctgtacctc	agatgacca	tctgaaaccc	gaggacactg	6240
ctctgtacta	ttgtgaaga	ggccgggtga	tctctgttcc	cgctatcaga	ggcgcagtaa	6300
ggggccctgg	aacacggta	accgttccat	ccgggggagg	cggttccaggc	ggtgggggat	6360
ctggcggggg	ttggatccaa	gttcaatgg	tcaatccgg	gggggggggt	gtccagacag	6420
ggggctccct	gaggctctcc	tgtcatctt	ccggaaagcat	cgccgggttc	gagaccgtga	6480
cctggtgtcg	ccaggctccc	ggggagtctc	tgcagtgggt	cgcttccatg	actaagacta	6540
acaacgcagat	ctactgtac	tcgtgaaag	gccgccttc	cattttcaga	gataacgc	6600
aaaacacacat	gtatctgcag	atgaatagt	tcaaacctga	agacacaggc	gtgtatttct	6660
gttaagggtcc	ttggatgggg	ggccggcc	tccaggtaac	agtctcgagt		6710

---

SEQ_ID NO: 74	moltype = DNA	length = 6704				
FEATURE	Location/Qualifiers					
misc_feature	1..6704					
	note = Plasmid pD1214-GA-hABAB					
source	1..6704					
	mol_type = other DNA					
	organism = synthetic construct					
SEQUENCE: 74						
ggtaaatca	tgttaattgt	tatgtcacgc	ttacattcac	gccctcccc	cacatccgct	60
ctaaccgaaa	aggaaggagt	tagacaac	taaagtctagg	tcccttattt	ttttttata	120
gttatgttag	tattaagac	gttattata	tttcaaattt	ttctttttt	tctgtacaga	180
cgcgtgtacg	catgtacat	tatactgaaa	accttgcgtt	agaagggttt	gggacgc	240
aaggctttaa	tttgccggcc	ctcacctca	cgcaaaaac	tttcaatc	aattcatcat	300
ttttttttt	ttctttttt	tgatctgg	ttctttgtt	ttttttgtat	tcggtaatct	360
ccgaaacagaa	ggaagaa	ccggagg	acagacttag	atggatata	atacgcata	420
gtatgttga	agaaacat	aattggcc	tattctta	ccaaactgcac	agaacaaaaa	480
ccagcagaa	acgaagata	atcatgtc	gaagctacata	taaggaacgt	gctgtactc	540
atctctgtcc	ttgtgtgc	aaatcttta	atatcatgc	cgaaaagccaa	acaaacttgt	600
gtgttctt	ggatgtcg	accaccaagg	attactgg	gttagttgaa	gcattaggtc	660
ccaaaattt	tttactaaa	acatctgtgg	atatcttgac	tgatttttcc	atggagg	720
cagttaa	gc	ttatccgca	agtacaattt	tttacttcc	gaagatagaa	780
aatttgc	tttgc	atcgttta	acatgttac	tgccgggtt	tacaaat	840
cagaatgggc	agacat	aatgcac	gtgtgggtt	cccggtt	gttagcg	900
tgaagcggc	ggcagaa	gtacaaagg	acccttaggg	ccttttgat	tttagc	960
tgtcatgc	gggccttca	tctactgg	aatatactaa	gggtactgtt	gacattgc	1020
aaagcgacaa	agatgtt	atcggttta	ttgtcaaa	agacatgggt	ggaagagat	1080
agggttacg	ttgggtt	atgcaccc	gtgtgggtt	atagcaag	ggagatgc	1140
tgggtcaaca	gtataga	ttgtatgt	ttgtcttc	aggatctgc	atttattt	1200
ttggaa	aggagg	actatttgc	aaggaaagg	atgtcaagg	agagggtt	1260
aacgcagg	gttgc	atgcacat	tttgcacat	gtccgcac	aaactt	1320
aatgtat	actat	aacttacaa	tttaggtt	tttaggtt	ttatgtt	1380
tacccacgc	atgtat	atcaaggaa	atgtat	tgaaggat	gactaatcc	1440
attgaggagt	ggcag	atgtat	agggtat	ctgaagg	catacgata	1500
cccgcatgg	atggata	atcacagg	gtactagat	acctttatc	ctatataat	1560
agacgcata	aatgtac	ttaa	acacgcacta	tgccgtt	ttatgtt	1620
tatataata	ggcaacac	atgtat	gttgcgtt	ttgtat	ttatgtt	1680
gtcg	cggt	catttc	agcgtcg	ttcggaaac	tttgcgtt	1740
atgttctt	ttct	atgtat	ttgtat	tttgcgtt	tttgcgtt	1800
tgaatgc	cc	tttgcgtt	atgtat	tttgcgtt	tttgcgtt	1860
gcgtat	cc	tttgcgtt	atgtat	tttgcgtt	tttgcgtt	1920
atatccat	ataac	atccac	ttgtat	tttgcgtt	tttgcgtt	1980
cttacat	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	2040
actatttata	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	2100
atagagacaa	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	2160
tatcactt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	2220
gccttata	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	2280
gagt	cggt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	2340
acggac	atgt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	2400
aaaatgtat	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	2460
caaaaat	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	2520
tgtaaaaat	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	2580
ttttacaaa	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	2640
tgttctgt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	2700
ttttgttca	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	2760
gcgcggaa	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	2820
acaataaccc	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	2880
ttatccat	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	2940
cttcaataat	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	3000
cccttttttgc	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	3060

-continued

aagatgctc	aagatcagg	gggtgcacga	gtgggttaca	tcaactgg	tctcaacagc	3120
ggtaagatcc	ttgagatgtt	tcgccccaa	gaacgtttc	caatgatgag	cactttaaa	3180
gttctgtat	gtggcgccgt	attatcccg	attgacgcgg	ggcaagagca	actccgtccg	3240
cgcataact	attctcagaa	tgacttggt	gagtactca	cagtcacaga	aaagcatct	3300
acggatgcca	tgacagtaaq	agaattatqc	aqgtctcc	taaccatq	tgataactact	3360
ggggcaact	tacttctgac	aacgatcgga	ggaccgaagg	agctaaccgc	tttttgcac	3420
aacatgggg	atcatgtaac	tcgccttgt	cgtgggaac	cgaggactgaa	tgaagccata	3480
ccaaacagc	agcgtgcac	cacgtgccc	tgacgatgg	caacaacgt	gcccggacta	3540
ttaactggcc	aactttaac	tctagcttc	cgccacaaca	taataactg	gtggggccg	3600
gataaaatgg	caggaccact	tctgogcteg	gcccttccgg	ctggctgggt	tattgtctgat	3660
aatccggag	ccggtgagcg	tggttctcgc	ggatatcatq	cagcgctggg	gccagatgg	3720
aaggccctcc	gtatecgtagt	tatctacaq	acggggggatc	agggcaactat	ggatgaacgca	3780
aatagacaga	tcgctgatag	agggtgcctc	ctgattaago	atgttaact	catgaccaaa	3840
atcccttaac	gtgagttacq	cgccggctgt	tccactggc	gtcagacccc	gtagaaaaaa	3900
tcaaaggatc	ttcttgatag	ccttttttc	tgcgctqat	ctgctgctt	caaacaaaaa	3960
aaccacccgt	accagcggt	gttttttgtc	cggatcaaga	gtaccaact	ctttttccga	4020
aggttaactg	cttcagcaga	cgccagatac	caaaatctgt	ttttctatgt	tagcctgt	4080
tagccccca	cttcagaac	tctgtacgg	ccgcatacata	cctcgctctg	ctaatctgt	4140
taccagtggc	tgctgcccgt	ggcgtataq	cgtgtcttac	cggttggac	tcaagacgt	4200
agttaccgg	taaggccgag	cggtcggt	gaacgggggg	ttctgcaca	cagccggact	4260
tggagcgca	gacactacac	gaactgt	actacacgg	tgatctatg	gaaaggccca	4320
cgcttcccg	agggagaaag	cgccgacagg	atccggtaaq	cgccggggtc	gaaacaggag	4380
agcgcacgag	ggaggttca	ggggggaaac	cctggatatc	ttatagttct	gtcggtttc	4440
ccacccctcg	acttgagcgt	cgattttgc	gatgtctcgtc	agggggccgg	agcctatgg	4500
aaaacggcc	caacgccc	tttttacgg	ttctgggtt	ttctgggtt	tttgcacca	4560
tgtttttcc	tgctgtatcc	cctgtatcc	tggtataacc	tattaccgc	tttgatgt	4620
ctgataccgc	tcggggctgt	gcaggatata	cttc当地at	tttctactc	ttttttactc	4680
ttccagattt	tctcggaact	cgcgcatcgc	cgtaccactt	caaaacaccc	aagcacagca	4740
tactaaat	ccctctttt	ttccctttag	gtgtctttaa	ttaccctgtac	taaaaggttt	4800
aaaaaaatgg	aagtgaccgc	ctcggtttt	tttttgcgtc	gaaaaggca	ataaaaaattt	4860
ttatcagtt	tctttttttt	gaaaattttt	ttttttgtt	ttttttctt	ttcatgtact	4920
ccatttata	tttaaagttaa	taaaacggact	tcaatttctc	aagtttca	tttcattttc	4980
ttttttttat	acaatctttt	ttacttctt	ctcatttagaa	agaaagatca	gcaatctaat	5040
ctaagttaa	aatgttcttt	agatcactt	tagctctgc	cggttgggtt	tgtatgttt	5100
ttggcaatca	ggtagacgt	gtggagacgg	gggggggggt	tgatcaacca	ggccgggttc	5160
tgagggtttc	ctgtgcgc	tctgggttca	cactggatta	ttcgtccata	gggttggttc	5220
ggcaggctcc	tggcaagag	cgtgaggggg	tctcatgtat	tagtagta	ggtgatgac	5280
caaaatgtac	cgatcccg	aaaggccgt	ttacaacc	caggataat	gctaagaaca	5340
ccgttatctat	ccatgtacac	tctctga	ccgcacata	ggccgtatata	tagtgcgg	5400
cttcagggc	gactatgtc	ggcgtgttcc	ctctgagccc	ttacggcaag	gacgactggg	5460
gcaaggggac	cctggtacc	gtatcttc	gcccgtgggg	gtctgggtgg	ggaggctcag	5520
gggggtgggg	caggccagg	caactgggtt	aatctgggg	aggctgttgc	caactctgggg	5580
gatccctgtat	actcttttc	ggggctccg	gattcaccc	ggactactat	ggcatcggt	5640
ggttccgc	gccccccagg	aaaggccgtt	aggccgttcc	ataccatgt	ggcgttgccc	5700
ggaccatact	gtacgcagac	tctgtgaagg	gacgcttac	catctctagg	gacaatgca	5760
aaaatgtct	gtatcttc	atagaacggc	tcaacgggg	ggataccgc	gtgtactact	5820
ggggagac	gggttctcc	gttttacgg	tgatatagt	gtggccgac	gactacgac	5880
tgtggggac	gggcacacag	gtgggtgtt	cttcctgtt	cggaaggccg	ggggggccgg	5940
gggggtgggg	cggtggggcc	agccaaactgc	agctgtttaga	gacagggggc	ggcttagtt	6000
agctctgggg	gtctcttc	ctgtcatcg	ctggctctgg	ctttaacc	atgtactacg	6060
tgtatgatct	ggtccggccaa	gtccaggga	agggggctga	ttggatcgct	actataata	6120
catgtggcc	caacaatgg	gacgacttc	aaaggccgg	tttcacccat	ttccagagaca	6180
acgccaagaa	tactctgtac	cttcagatga	ccagtcgtaa	acccggaggac	actgtctgt	6240
actattgtc	aagaggccgg	gtgatctctg	cttcctgtat	caagggcgc	gtaaaggggcc	6300
ctggaaacaca	ggtaaccgc	tcatccgggg	gaggccgtt	aggccgttgg	ggatctcgcc	6360
gggggtggat	ccaaatgtc	ctggtca	ccggggccgg	actgttccag	acagggggg	6420
ccctggatct	cttcctgtca	tcttcggaa	gatctccgg	cttcgagacc	gtgacccgtt	6480
ctcgccaggc	ccccggga	tctctgtc	gggtcgcttc	catgactaa	actacaacag	6540
agatctactc	tgactctgt	aaaggccgc	tcatcatcc	tagatataac	gctaaaaaca	6600
catgtatct	gcatgtatct	agtcataa	ctgaagacac	aggcgctgtat	ttctgtatgg	6660
qtcctqact	qaggqqccaa	ggccatccaaq	taacatgttc	qagt		6704

```
SEQ ID NO: 75          moltype = DNA    length = 6698
FEATURE                  Location/Qualifiers
misc_feature            1..6698
note = Plasmid pD1214-1N-hABAB
source                  1..6698
mol_type = other DNA
organism = synthetic construct
```

SEQUENCE:	75	Organism = Synechococcus elongatus				
ggttaaatca	tgtatattagt	tatgtcacgc	ttacatttac	gcccctcccc	cacatccgt	60
ctaaccggaa	aggaaggagt	tagacaacct	gaagttttag	tcccttattt	tttttttttt	120
gttatgttag	tattaagaac	gttattttttt	tttcaattt	tttttttttt	tctgttacaga	180
ccgcgttacg	catgttacat	tatactgtaaa	accttttttt	agaaggtttt	ggggacgtcg	240
aaggctttaa	tttgcggccc	ctcacctgtca	cgaaaaaaagc	ttttcaattt	aatttcatcat	300
ttttttttta	ttcttttttt	tgatttcggt	ttttttgaaa	tttttttttt	tcggtaatct	360
ccgaaacagac	ggaqaacgca	aggaaggagc	acagacttag	atggatatat	atacgcatat	420
gtatgttga	agaaacatgt	atttggccat	tatttttttt	ccaactgcac	agaacaaaaaa	480
ccacggaaac	acgaaatgtaa	atcatgtgtca	aaactttttt	taaggaaacgt	gtctgtactc	540

-continued

---

atcctagtc	tgttgtgc	aagctattta	atatcatgca	cggaaaagcaa	acaaacttgt	600
gtgttattt	ggatgtcg	accaccaagg	aattactgga	gttagttgaa	gcatttagtc	660
ccaaaattt	tttactaaa	acacatgtgg	atatcttgac	tgatttttcc	atggaggggca	720
cagtttaagcc	gctaaaggca	ttatccgcca	agtacaattt	tttacttttc	gaagatagaa	780
aatttgcg	cattggtaat	acatgcataat	tgcaacttgc	tgccgggtgt	tacagaatag	840
cagaatgggc	agacattacg	aatgcacacg	gtgtgggtgg	cccaggattt	gttagcggtt	900
tgaagcggc	ggcagaagaa	gtaacaaagg	aaccttagagg	ccttttgatg	ttagcagaat	960
tgtcatgca	gggctcccta	tctactggag	aatataactaa	gggtactgtt	gacattgcga	1020
aaagcgacaa	agattttgtt	atcggttta	ttgctcaaaag	agacatgggt	ggaagagatg	1080
aagggtacg	ttgggttatt	atgacacccg	gtgtgggttt	atagacaag	ggagatgcatt	1140
tgggtcaaca	gtatagaacc	tggtatgtg	ttgtctctac	aggatctgc	atttatttt	1200
tttggaaagagg	actatttgc	aaggaaaggg	atgctaagg	agagggtgaa	cgttacagaa	1260
aagcaggcgt	ggaagcatat	tttggaaagat	gcccggccaa	aaactaaaaa	actgttattat	1320
aagtaaatgc	atgtatata	aactcaaaaa	tttaggtttt	aattttatata	tatcgttatt	1380
tacccacgc	atgtatcaat	atcaaggaa	atgatagcat	tgaaggatga	gactaatcca	1440
attgaggagt	ggcagccat	agaacagct	aagggtatg	ctgaaggaaag	catacgat	1500
cccgcatgga	atgggataat	atcacaggag	gtactagat	accttttcatt	ctacataat	1560
agacgcataat	aagtacgc	tttggatata	acacgcacta	tgccgttctt	ctcatgtata	1620
tatataatata	ggcaacacgc	atgataggt	gccccgttga	cagtgttgc	tatgtgcgca	1680
gctcgcttg	cattttcgga	agcgtcggtt	ttcgaaacg	ctttgttgc	cctattccga	1740
agttccattt	ttcttagaaag	tataggaaat	tcagagcgct	tttgaaaacc	aaaagcgctc	1800
tgaagtcgca	ctttccaaaa	accaaaaaac	caccggact	taacgagct	ctaaatatt	1860
gcaataatcc	tttcccaaaa	cattgttca	aaatgttctt	tttgcataat	tctctgtgt	1920
atatccctat	ataacacc	catccaccc	tegttcttgc	aacttgcatt	taaactcgac	1980
ctctacattt	tttatgttta	ttcttgcatt	tactctttgc	acaaaaaaat	tgttagtaaa	2040
actatttata	gagtgcatt	aaaacaatac	gaaaatgtt	acattttccat	tacgttagat	2100
atagagacaa	aatagaagaa	accgttcaat	tttttgcatt	caatggaaat	tcatcaacgc	2160
tatcacttc	tttttttttt	tttgcataaa	gtatgcgca	tccacatgg	tatagaatat	2220
gcctttatct	tgaaaaaaat	caccggcgg	tttgcgtatg	atcgttacaa	gccccggatg	2280
gagttaggt	tttttttttt	aagagaaaaat	agacacccaa	gtatgttgc	tatgttgc	2340
acggacccat	agtgcataaa	tttacaaaga	gactgttca	tagagcgca	aaaggagaaa	2400
aaaagtaatc	taagatgtt	tttttttttt	atagcgctt	cgggatgtat	ttttgttagaa	2460
caaaaaaaagaa	gtatagattt	tttgcgttgc	aaatagcgct	ctcgctgtgc	attttgttgc	2520
tgtaaaaatgt	cagcttgcatt	tttttttttt	aaaataggat	gttctcgctg	tgcatatttt	2580
tttttttttt	atagagcaca	gatttttgc	tttttttttt	tttttttttt	tttttttttt	2640
tgttctgtt	aaatgcagct	cagatttttt	tttttttttt	tttgcgttct	cgccgttgc	2700
tttttttttt	caaaatgtaa	cacagatgt	tttttttttt	tttttttttt	tttttttttt	2760
ggccggaaacc	tttttttttt	tattttttca	aatacattca	aatatgttac	cgcttcatgag	2820
acaataatccc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	2880
tttatttttt	aaatataattt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	2940
cttcaataat	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	3000
cccttttttgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	3060
aaagatgttgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	3120
ggtttagatcc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	3180
gttctgttat	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	3240
cgccatcaact	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	3300
acggatggca	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	3360
ggccgcaact	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	3420
aatatgggggg	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	3480
ccaaacgcac	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	3540
ttaacttgcg	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	3600
gataaaatgt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	3660
aaatccggag	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	3720
aaaggcccttc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	3780
aatagacaga	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	3840
atcccttaac	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	3900
tcaaaggatc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	3960
accacccgtt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	4020
aggttaacttgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	4080
tagcccaacca	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	4140
taccagtggc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	4200
agtttacccg	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	4260
tggagcgaac	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	4320
cgccatccgg	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	4380
agcgacacg	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	4440
gccacacctg	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	4500
aaaacccgg	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	4560
ttatcagttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	4620
ccatttggata	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	4680
tttgcattttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	4740
tactaaat	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	4800
aaaaagaaaa	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	4860
ttatcagttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	4920
ccatttggata	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	4980
tttgcattttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	5040
ctaagtttaa	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	5100
tgcaggatc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	5160
tttcctgtgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	5220
ctcctggca	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	5280

-continued

---

```

acggccgattc cgtaaagggc cggtttaca cctccaggga taatgctaag aacaccgtat 5340
atctccagat gaactctcg aagcccgacg atacggccgt attacttgt gcccgttca 5400
gggcgactat gtgcggcgtg ttccctctga gccttacgg caaggacac tggggcaagg 5460
ggaccctgtt gaccgtatcc tcaggcggtg gagggtctgg tgaaaaagggc tcaggggttg 5520
gagggcoca ggtcaactg gttaaatcg gggggggctt ggtacaactt gggggatccc 5580
tgagacttc ttgcgaggcc ctggactca ctatggactt ggcgttcc 5640
gcggcccccc agggaggag cgggaggccg tttcatacat tagtgcctgt gccggacca 5700
tactgtacgc agactctgtg aaggacgtt ttaccatctc tagggacaat gccaaaatg 5760
ctgtgtactt gcaaatacg acgcgttac gggaggatca cgcgttac tactgcgcga 5820
gacggcgtt tccgttctt acgcgttac gatggcttgc cgcgttac gacgttggg 5880
gacggggccac acagggttgtt gtgtttccg gtggcggaa eggaggggcc acgggggttg 5940
ggagcgttgg gggcagccaa ctgcgttgg tagagacagg gggcggctt gttcagcctg 6000
gagggttctt cagactgtca tgcgttctt ctggcttac ctgcgttac tacgttgc 6060
catgggttcc ccaacttca gggggggcc ctgggttgat cgcttactt aatacagatg 6120
gcacgttccat gggggcgttcc tccacaaagg ggccgttccat cattttccaga gacaacccca 6180
agaataactct gtacatttcg atgaccatgc tggaaaccgc ggacactgtt ctgtactatt 6240
gtgcagagg cccgggttgc tctgttccg ctatcgatgg cgcgttggg ggccttggaa 6300
cacaggttcc acgttccatcc gggggggccg gttcaggggg tggggatctt ggggggggtt 6360
gttccaaatg ttagtgcgtc gatccgggg gggactgtt ccacacagggg ggctccctga 6420
ggctcttctt tgcattttcc ggaagcatcg cccgttccg gaccgttcc gtttctcgcc 6480
aggctccccc gaaatgttctg cagtgggtcg ctccatgtac taagactaac aacgagatct 6540
actctgactc agtggaaaggc cgcgttccat tttcttagata taacgtttaa aacacagtgt 6600
atctgcagat gaatgttcc aacatggaa acacaggctt gtatccgtt aagggttctt 6660
agctgagggg ccacggccatc caggtaacag tctcgagt 6698

```

SEQ ID NO: 76                    moltype = DNA length = 6707  
 FEATURE                        Location/Qualifiers  
 misc\_feature                1..6707  
                               note = Plasmid pD1214-IVS-hABAB  
 source                        1..6707  
                               mol\_type = other DNA  
                               organism = synthetic construct

SEQUENCE: 76

```

gtttaaatca tgtaatagt tatgtcacgc ttacattcac gcctccccc cacatccgt 60
ctaaccgaaa aggaaggagt tagacaacct gaagtcttag tcccttattta ttttttata 120
gttatgttag tattaagaac gttattttata ttccaaattt ttctttttt tctgtacaga 180
cgcgtgtacg catgttacat tatactgaaa accttgcgtt agaagggttt gggacgttgc 240
aaggctttaa ttgcggccccc ctcaacttgcg cgcggggggcc ttttcaattt aatttcatcat 300
ttttttttt ttctttttt ttatgttgcgtt ttcttttgcata tttttttttgcata 360
ccgaacacaa ggaagaaacga aggaaggagc acagacttag attggatata atacggcatat 420
gtagtgttgc agaaacatgc aattggcccg tatttcttac ccaactgttgc acgaaacaaaa 480
ccacggggcc acggaaatccat atcgatgtca aagcttacata taagggactgt gctgttactc 540
atccctatgtcc ttgttgcgtcc aagcttacata atcgatgtca cggaaacggaa acggaaatgtt 600
gtgttgcattt ggtatgttgcgtt accaccaagg aattactgtt gtttgcgtt gcatgggttgc 660
ccaaaatttgc ttactttttt acacatgtgg atatcttgcg ttttttttcc atgggggca 720
caigtttgcgca gctaaaggcc ttatccggcc agtacaattt ttacttcttgc gaaatgttgc 780
aatttgcgttgc aatttgcgttgc aatgttgcgtt tgcgttgcgtt tacagaatag 840
cagaatggcc acacatttgc aatgttgcgtt tgcgttgcgtt cccagggtt gtttgcgttgc 900
tgaaggcggc ggcggaaaggaa gtaacaaagg aaccttagagg ctttttgcgtt ttagcgttgc 960
tgcgttgcgtt gggccatccat tctatgtggg aatataactaa ggttactgtt gacatttgcg 1020
aaagcggccaa agattttttgcgtt atcggttttgcgtt ttttttttttgcgtt ggttactgtt 1080
aagggttgcgtt ttgggttgcgtt atcgaccccg gtttgcgttgcgtt agatgttgcgtt ggttactgtt 1140
tgggttgcgtt gtttgcgttgcgtt gtttgcgttgcgtt ttttttttttgcgtt ggttactgtt 1200
ttggaaaggacttgcgtt aagggttgcgtt gtttgcgttgcgtt gtttgcgttgcgtt 1260
aaggcggccaa ggttactgttgcgtt gtttgcgttgcgtt gtttgcgttgcgtt 1320
aagtttttttttgcgtt gtttgcgttgcgtt gtttgcgttgcgtt gtttgcgttgcgtt 1380
tacccacgcgtt atgttgcgtt atcaaaaggaa atgttgcgtt gtttgcgttgcgtt gtttgcgttgcgtt 1440
attgggggttgcgtt gtttgcgttgcgtt gtttgcgttgcgtt gtttgcgttgcgtt 1500
ccggccatccat ttttttttttgcgtt gtttgcgttgcgtt gtttgcgttgcgtt 1560
agacggccatccat aatgttgcgtt gtttgcgttgcgtt gtttgcgttgcgtt 1620
tatatataca ggcacacccat agatgttgcgtt gtttgcgttgcgtt gtttgcgttgcgtt 1680
gttcgttgcgtt gtttgcgttgcgtt gtttgcgttgcgtt gtttgcgttgcgtt 1740
agtttgcgttgcgtt gtttgcgttgcgtt gtttgcgttgcgtt gtttgcgttgcgtt 1800
tgaatgttgcgtt gtttgcgttgcgtt gtttgcgttgcgtt gtttgcgttgcgtt 1860
gcgttgcgtt gtttgcgttgcgtt gtttgcgttgcgtt gtttgcgttgcgtt 1920
atatccatcatat aatgttgcgtt gtttgcgttgcgtt gtttgcgttgcgtt 1980
cttctatcatat ttttttttttgcgtt gtttgcgttgcgtt gtttgcgttgcgtt 2040
actatttgcgtt gtttgcgttgcgtt gtttgcgttgcgtt gtttgcgttgcgtt 2100
atagagacaa aatgttgcgtt gtttgcgttgcgtt gtttgcgttgcgtt 2160
tatcacttc ttgttgcgtt gtttgcgttgcgtt gtttgcgttgcgtt 2220
gcgttgcgtt gtttgcgttgcgtt gtttgcgttgcgtt gtttgcgttgcgtt 2280
gatgttgcgtt gtttgcgttgcgtt gtttgcgttgcgtt gtttgcgttgcgtt 2340
acggccatccat aatgttgcgtt gtttgcgttgcgtt gtttgcgttgcgtt 2400
aaaatgttgcgtt gtttgcgttgcgtt gtttgcgttgcgtt gtttgcgttgcgtt 2460
caaaatgttgcgtt gtttgcgttgcgtt gtttgcgttgcgtt gtttgcgttgcgtt 2520
tgtaatgttgcgtt gtttgcgttgcgtt gtttgcgttgcgtt gtttgcgttgcgtt 2580
tttttttttgcgtt gtttgcgttgcgtt gtttgcgttgcgtt gtttgcgttgcgtt 2640
tttttttttgcgtt gtttgcgttgcgtt gtttgcgttgcgtt gtttgcgttgcgtt 2700
tttttttttgcgtt gtttgcgttgcgtt gtttgcgttgcgtt gtttgcgttgcgtt 2760

```

-continued

---

```

gcgccggaaacc cctattttgtt tatttttcta aatacattca aatatgtatc cgctcatgag 2820
acaataaacc tgatatttgtt cagaatttgtt taatttgtt taacactgac ccctatttgt 2880
ttatTTTCTT aaatacattc aaatatgtat ccgctcatga gacaataacc ctgataaatg 2940
cttcaataat attgaaaaag gaagaatatg agtattcaac atttccgtgt cgcccttatt 3000
ccctttttt cggcattttt ccttcctgtt ttgcctcacc cagaaacgtt ggtaaaagta 3060
aaagatgtg aagatcagg tggtgacgca gttgggttaca tggaaactgga tctcaacagc 3120
ggttaagatcc tttagatgtt tcggccgaa gaacgtttt caatgtatc cacttttaaa 3180
gttctgtat gtggcgcggt attatcccgtt attgacgcccc ggcaagagca actcggtcgc 3240
cgatataact attcctcaggaa tgacttgtt gagaacttcac cagtacacaa aaagcatctt 3300
acggatggca tgacgatggaa agaaatgtt agtgcgtgcacaa taaacctgag tgataaacat 3360
geggccaact tacttctgac aacatcgga gggccggaaagg agctaaccgc tttttgcac 3420
aacatggggg atcatgttac acgttcgtat cgttgggaaac cggagatgaa tgaagccata 3480
ccaaacacgacg aacgtgcacac cactgtatc gtacgtatc ggacaaacgtt ggcggaaacta 3540
ttaactggcg aactacttac ttagcttc cggcaacaaat taatagactg gatggggcg 3600
gataaaatgtt caggaccact tctgcgtcg gcccctccgg ctggctgtt tatttgcgtat 3660
aaatccggag cccgtgagcg tggttctcg ggtatcatcg cagcgttggg gccagatgtt 3720
aaggccctttt ctatgtatc acgttcgtatc acggggagatc aggcaactat ggtgaacga 3780
aatagacaga ttcgtatcaggat aggtgcctca ctgattaage attgttaact catgaccaaa 3840
atcccttaac gtgaggtaatc cggcgctgtt tccactgtac gtcagacccc gtaaaaaaga 3900
tcaaaggatc ttcttgcgtat ccttttttc tgcgcgtat ctgcgtgtt caaacaaaaa 3960
aaccacccgtt accagcggtt gtttttgtt cggatcaaga gctaccaact cttttccga 4020
aggttaacttgg cttcagcaga ggcagatcaca caataactgtt ttttctatgtt tagccgtatg 4080
tagccccacca cttcaagaac tctgtatc cgcctatcata ctcgcgtctg ctaatctgt 4140
taccagtggc tgctgcgtat ggcgataatc cgtgttcttac cgggttggac tcaagacat 4200
agttacccggta aaggccggcag cgggtgggtt gaaacgggggg ttcgtgcaca cagcccaat 4260
tggagcgaac gacccatcacc gaaactgtatc acctacagcg tgactatgaa gaaacggcc 4320
cggttccggaa aaggagaaatgg gggccggatgtt atccggtaatc cggcagggtt ggaacaggag 4380
agcgcaegag ggagatcataa gggggaaacgtt cctggatctt ttatgtctt gtcgggtttc 4440
gccacacccgtt atttgcgtat cgtttttgtt gatgcgttgc tggggggccgg agccatgtt 4500
aaaacacggcc caacggggcc tttttacgtt ttcgtggctt ttgtgtcataa 4560
tggttttttcc tgcgttattcc cttgtatctg tggataacccg ttttgcgtt tttgagtgtt 4620
ctgtatccgc tccgggtcgatc gcaggatataat cttccaaatgtt ttttctatctt 4680
ttccagatattt ttcgttactt cgcgtatcg cgttaccatc caaaacaccc aagcacacgt 4740
tactaaattttccctt ttccttcttgcgtt gttgtcgatc ttacccgtatc taaagggttt 4800
gaaaagaaaaa aagtgtatcgc ttcgtttttt ttttttgcgtt gaaaaggcca ataaaaatgtt 4860
ttatcagttt tttttttttt gaaaatgtt ttttttgcgtt ttcgtatcc 4920
ccattgtataa ttaaactgtt tttttttttt ttttttgcgtt ttcgtatcc 4980
ttgttcttattt acaactttttt ttttttgcgtt ttttttgcgtt ttttttgcgtt 5040
ctaagttttt aatgttactt cttccatc ttttttgcgtt ttttttgcgtt ttttttgcgtt 5100
tctcccaatc gaaatgttactt ttttttgcgtt ttttttgcgtt ttttttgcgtt 5160
caactgtatcgc ttttttgcgtt ttttttgcgtt ttttttgcgtt ttttttgcgtt 5220
ttccggcggcc ttttttgcgtt ttttttgcgtt ttttttgcgtt ttttttgcgtt 5280
gcacaaatgtt ttttttgcgtt ttttttgcgtt ttttttgcgtt ttttttgcgtt 5340
acaccgtata ttttttgcgtt ttttttgcgtt ttttttgcgtt ttttttgcgtt 5400
cggttttccatc gtttttgcgtt ttttttgcgtt ttttttgcgtt ttttttgcgtt 5460
ggggcggccatc atctgtatcgc gacttgcgtt ttttttgcgtt ttttttgcgtt 5520
caggggggccatc ttttttgcgtt ttttttgcgtt ttttttgcgtt ttttttgcgtt 5580
caggggggccatc ttttttgcgtt ttttttgcgtt ttttttgcgtt ttttttgcgtt 5640
gttgggttccg ctttttgcgtt ttttttgcgtt ttttttgcgtt ttttttgcgtt 5700
cccccggccatc atctgtatcgc gacttgcgtt ttttttgcgtt ttttttgcgtt 5760
caaaaaaaaatgtt ttttttgcgtt ttttttgcgtt ttttttgcgtt ttttttgcgtt 5820
acttgcgtatcgc ttttttgcgtt ttttttgcgtt ttttttgcgtt ttttttgcgtt 5880
acgtgttgggg acggggccatc ttttttgcgtt ttttttgcgtt ttttttgcgtt 5940
ggggggggggccatc ttttttgcgtt ttttttgcgtt ttttttgcgtt ttttttgcgtt 6000
tttccatc ttttttgcgtt ttttttgcgtt ttttttgcgtt ttttttgcgtt 6060
acgtatgtatcgc ttttttgcgtt ttttttgcgtt ttttttgcgtt ttttttgcgtt 6120
atataatgtt ttttttgcgtt ttttttgcgtt ttttttgcgtt ttttttgcgtt 6180
acaaacccgtt ttttttgcgtt ttttttgcgtt ttttttgcgtt ttttttgcgtt 6240
tgcgtatgtt ttttttgcgtt ttttttgcgtt ttttttgcgtt ttttttgcgtt 6300
ggccctggaaac acaggatcataa ttttttgcgtt ttttttgcgtt ttttttgcgtt 6360
ggggggggccatc ttttttgcgtt ttttttgcgtt ttttttgcgtt ttttttgcgtt 6420
gttgggttccg ttttttgcgtt ttttttgcgtt ttttttgcgtt ttttttgcgtt 6480
gggttccatc ttttttgcgtt ttttttgcgtt ttttttgcgtt ttttttgcgtt 6540
acggatcataa ttttttgcgtt ttttttgcgtt ttttttgcgtt ttttttgcgtt 6600
acacagtgtt ttttttgcgtt ttttttgcgtt ttttttgcgtt ttttttgcgtt 6660
agggttccatc ttttttgcgtt ttttttgcgtt ttttttgcgtt ttttttgcgtt 6707

```

SEQ ID NO: 77 moltype = DNA length = 6728  
 FEATURE Location/Qualifiers  
 misc\_feature 1..6728  
 note = Plasmid pD1214-KP-hABAB  
 source 1..6728  
 mol\_type = other DNA  
 organism = synthetic construct

SEQUENCE: 77  
 ggtaaaatca tgtaattatgt tatgtcacgtt tacatttac gcccctcccc cacatccgt 60  
 ctaaccgaaa aggaaggagt tagacaacctt gaaatgtatc tttttttata 120  
 gttatgtttag tattaaatgtt tttcaatgtt ttctttttt tctgtacaga 180  
 cggcgatcgtacatcatatactgtaaa accttgcgtt agaagggtttt gggacgtcgt 240

-continued

---

aaggctttaa tttggggccc ctcacctgca cgcaaaaagc ttttcaattc aattcatcat	300
ttttttta ttcttttt tgatttcgt ttctttgaa ttttttgat tcggtaatct	360
cgcgaacagaa ggaagAACGA aggaaggago acagacttag attggatat atacgcat	420
gtagtgtga agaaacatga aattgcccag tattcttaac ccaactgcac agaacaaaaa	480
ccagcaggaa acgaagataa atcatgtca aagotacata taaggaacgt gctgtactc	540
atcttagtcc ttgtgtcgcc aagcttta atatcatgca cgaaagcaa acaaacttgt	600
gtgcttcatt ggatgtcgt accacaagg aattactgga gtagttgaa gcattaggtc	660
ccaaatgtt tttactaaaa acacatgtgg atatcttgac tgattttcc atggggcca	720
cagttaaaggc gctaaaggc ttatccgca agtacaattt ttactcttc gaagatagaa	780
aatttgcgtga cattgtgtat acagtcaatg tgcaagtactc tgccgggtgtc tacagaatag	840
cagaatggcc agacattacg aatgcacagc gtgtgggtgg cccaggattt gtagcggtt	900
tgaaggcggc ggcagaagaa gtaacaagg aaccttagagg cttttgtg ttagcagaat	960
tgtcatgcaaa gggctccctt tctatcgtag aatatactata gggtaactgtt gacattgcga	1020
aaagcgcacaa agatttgtt atcggtttt tgtagctaaag agacatgggt ggaagagatg	1080
aaggttacgaa ttgggtgtt atgacacccg gtgtgggtt agatgacaag ggagatgcat	1140
tgggtcaaca gtatagaacc gtggatgtat ttgtctctc aggtatgtac attattattt	1200
ttggaaaggc actatgtca auggggaaaggg atgctaagggt agagggtgaa cgttacagaa	1260
aagcaggctat tgagaagat gggccacagc aaactaaaaa actgtattat	1320
aagtaaatgc atgtataacta aactcacaat tttagagttc aatttaataat tatcgttat	1380
tacccacgcg atgatccat atcaaaggaa atgatagcat tgaaggatga gactaatcca	1440
attggaggat ggcacgatcat agaaacagct aagggttagt ctgaaggaaag catacgat	1500
cccgcatggaa atgggataat atcacaggag gtactagat acctttccatc ctacataat	1560
agacgcacat aagtacgcat ttaaagcataa acacgcacta tgccgttcc ctcatgtata	1620
tatataataca ggcaacacgc agatataagggt ggcacgtgaa cagttagtgc tatgtcgca	1680
gtctcggtt catggcgaa agcgcgtt ttccggaaacg ctttgaaggctt cctattccga	1740
agtttcttatt ctcttagaaatg tataaggactt ctagagcgctt ttggaaaccaaa aaaaagcgctc	1800
tgaagtcgca cttccacaaa aaaaaaaacg caccggactt taacgacgcta ctaaaatatt	1860
ggeaatacccg cttccacaaa cattgctca aagtatctt ttgctatata tctctgtgt	1920
atatccctat ataacctacc catccaccc ttgctccctt aacttgcate taaactcgac	1980
cttotacattt ttatgttta ttcttagat tactcttagt aaaaaaaaaat tgtagtaaga	2040
actattcata gagtgaatcg aaaaacatac gaaaatgtaa acattttccata tacgttagat	2100
atagagacaa aataagaagaa accgttcata attttctgac caatgaagaa tcataacgc	2160
tatcacttcc ttgttcacaaa gtatgcgca tccacatcgat tataagatata aatcggggat	2220
gcctttatcg taaaaaaatg caccggcggc ttccgttagt atcagtaaaac gccccggaaatg	2280
gagttaggcg tttttttatgg aagaaaaat agacacccaa gtagccttct tctaacccat	2340
acggacccatc agtgcacaaa gttatcaaga gactgcatttta tagagggcacaa aaggagaaaa	2400
aaaagtaatc taagatgtt ttgttggaaa atagcgctct cgggatgtat tttttagaa	2460
aaaaaaagaa gtataatgtt ttgttggtaa aaatagcgctt ctcgcgttgc attttctgtt	2520
tgtaaaaatgc cagtcgtatc ttcttttttggg aaaaatttgcgat ttccgtcgat tgcattttt	2580
ttttacaaaa atgaacacaa gattttcgat ttgttggaaaat ggcctttccg gttgat	2640
tgttctgtaa aatgcagatc cagattctt ttgttggaaaat ttagcgctct cgcgttgcatt	2700
ttttgttctt caaaatgttgcgat cagatgtatc ttccgtcgat ggcacttttcc gggggaaatgt	2760
gcggcggacc cttatgttgc ttttttctta aataatgtatc cgcgtat	2820
acaaataaccc tgatatttttgc cagaattttgtt taatgttgggtaa acatgtac cccttatttgc	2880
ttattnnttctt aaatacattc aaataatgtatc ccgcgtatgc gacaataacc ctgataat	2940
cttcaataat atgaaaaatg gaagatataatc agtatttcacat tttccgtgtt cgccttattt	3000
cccttttttgc cggcatttttgc ctttctgtt tttgttccac cagaaacgcgtt ggttggaaatgt	3060
aaagatgtgtc aagatcgatgg ggggttccgcgat ggttgggttaca tgcgacttgcgat ttttttgcac	3120
ggtaagatcc ttgagatgtt tccggccggaa gaaatgttgc ttttttgcac	3180
tttctgtatcg ttggcgccgtt attatccgtt attgacgcggc ggcacagacaa actcggttgc	3240
cgcatacact attcgcgaa tgacttgcgtt ggttggggatc cagtcacaga aagcatctt	3300
acggatggca tgacatgttgc gaaatgtatc agtgcgttgc ttttttgcac	3360
gcggccaaact ttttttgcac aacatgttgc ggcacggaaagg agctaaccgc ttttttgcac	3420
aacatggggg atcatgttgc tccgttgcgtt ctttttgcac	3480
ccaaacgcacgc acgcgttgcac ctcgttgcgtt gtttttgcac	3540
ttaacttgcg acactttatc ttttttgcac ctttttgcac	3600
gataaaatgtt ctttttgcac ctttttgcac ttttttgcac	3660
aaatccggatc cccggatggcg ttttttgcac ttttttgcac	3720
aaggccctccgtt ttttttgcac ttttttgcac	3780
atagacatccatc ttttttgcac ttttttgcac	3840
atcccttaac gtttttgcac ttttttgcac	3900
tcaaggatccatc ttttttgcac ttttttgcac	3960
aaccacccgtt accagccgtt ttttttgcac ttttttgcac	4020
aggtaacttgc ttccgttgc ggcacatgttgc ttttttgcac	4080
tagccccacca ttttttgcac ttttttgcac	4140
taccatgttgc ttttttgcac ttttttgcac	4200
agtttccggatc ttttttgcac ttttttgcac	4260
tggagcgacac gtttttgcac ttttttgcac	4320
cgtttccggatc ttttttgcac ttttttgcac	4380
agcgacacgc gggatgttgc gggggaaacgc ttttttgcac ttttttgcac	4440
gccacccctgc attttgcac ttttttgcac ttttttgcac	4500
aaaaacgcacgc caacccggcc ttttttgcac ttttttgcac	4560
tgttctttcc ttttttgcac ttttttgcac	4620
ctgtatccgc ttttttgcac ttttttgcac	4680
tttccatccgc ttttttgcac ttttttgcac	4740
tactaaatccgc ttttttgcac ttttttgcac	4800
aaaaagaaaa aagtgcaccc ttttttgcac ttttttgcac	4860
ttatcagatcc ttttttgcac ttttttgcac	4920
cccaatttgcata ttttttgcac ttttttgcac	4980

-continued

---

```

ttgttctatt acaacttttt ttacttcttg ctcatcattaa agaaagcata gcaatcta 5040
ctaagtttaa aatgacaaaa ccaacccaag tcttggtag atctgtatcc attctttct 5100
ttatcacttt attgcattctg gttgttgcata tgccaggataca gctgggttag acggggggag 5160
ggctggataca accaggcggt tcactgaggo ttccctgtgc cgcatctggg ttcacactgg 5220
attatcgctc catagggtgg ttccggcgg ctctggcaaa agagcgtagg ggggtctcat 5280
gtatttagtag tagtggtagat agcacaagg acggcgatcc cgtaaaggcc cggtttacaa 5340
cctccaggaa taatgtcaag aacaccgtat atctccagat gaactctctg aagcccgacg 5400
atacggccgt atattactgt ggggttca gggcactat gtggggctgt ttcccctctga 5460
gccttacgg caaggacgac tggggcaagg ggacccctggt gaccgtatcc tcaggcggtg 5520
gagggttgg tgggggagggc tcagggggttgg gggcagocca ggtcaactg gttgaatctg 5580
ggggagggtt ggtacaacctt gggggatccc tgagactctc ttgcggggcc tccggatcca 5640
ccttggacta ctatggcata ggctgggtcc gccagcccc agggaggag cggggggccg 5700
tttcacatcat tagtggcactg gcccggacca tactgtaccc agactctgtg aaggggacgt 5760
ttaccatctc tagggcaaat gccaaaatgt ctgtgtaccc gcaaatgaaag agcctcaagc 5820
ggggaggatc cgcgtgtac tactgcgcga gacggcgctt ctccgttctc agcgtgaata 5880
gatggctggc cgacgactac gacgtgtggg gacggggcac acaggtgggt gtgtcttccg 5940
gtggcggaaag cggggggggc acgggggggtg ggacgggtgg gggcagccaa ctgcagctgg 6000
tagagacagg ggggggttca gttcagctg ggggggtctc cagactgtca tgccgtgtc 6060
ctggcttac cttcgttac tacgtgtatca ctgggttcc ccaagtcaca gggggggcc 6120
ctgagtggtt ctgtactatt aatacagatg gcagcacaat gggggacgac tccacaaaagg 6180
ggcggttccac cattttccaga gacaacgcga agaaatactt ctacccatcg atgaccatc 6240
tgaaccccgaa ggacactgtc ctgtactat gtgcaagagg cgggggtgatc tctgttccg 6300
ctatcagagg cgcgttacgg gggccctggaa cacaggttac cgtttcatcc gggggggccg 6360
gttcaggcggt tgggggatct ggggggggtt gatccaaagt tcaagctggc gaatccgggg 6420
gcggactgtt ccagacaggg ggctccctga ggctctccgt tgcacatcc ggaaggatcg 6480
ccggcttccac gaccgttacg tggtctcgac aggttcccgga gaatgtctcg cagtgggtcg 6540
cttccatgtac taagactaac aacagatgtt actctgtactc agtggaaaggc cgcttcata 6600
tttctagaga taacgttaaa aacacagtgt atctgcagat gaatagtctc aaacctgaag 6660
acacaggcgt gtatttctgt aagggttccgt agctgagggg ccagggcata caggttacag 6720
tctcgagt

```

```

SEQ ID NO: 78 moltype = DNA length = 6728
FEATURE Location/Qualifiers
misc_feature 1..6728
note = Plasmid pD1214-LZ-hABAB
source 1..6728
mol_type = other DNA
organism = synthetic construct

```

```

SEQUENCE: 78
ggtaaatca tgtaatttagt tatgtcacgt ttacattcac gcccctcccc cacatccgt 60
ctaaccgaaa aggaaggagt tagacaacctt gaagtctagg tccctattta ttttttata 120
gttatgttag tattaagaac gttttttttt ttccaaattt ttcttttttt tctgtacaga 180
cgctgttacg catgttacat tataactggaa accttgcgtt gaaagggttt gggacgtcg 240
aaggcttaa ttgcggcccc ctcacactca cgccaaaacg ttttcaattt aatccatcat 300
ttttttttt ttcttttttt tgatttcgtt ttctttgtaa tttttttgtat tcggtaatct 360
ccgaacacaaa ggaagaaacggg aggaaaggggc acagacttgcg atgggtatata atacccat 420
gtatgttgcg agaaacatgt aatttcccgat tattcttaaattt ccaatgtcaca agaaacaaaaa 480
ccacggggaa acgaagataa atcatgtcga aagtttccatata taagggaaatgt gtcgtactc 540
atccctgtcc ttgttgcgtcc aagtttccatata atatcatgtca cgaaaacgaa acaaacttgt 600
gtgttccattt ggatgttgcgtt accacaaggg aatatactggc gttagttggaa gcattagggtc 660
ccaaattttgg tttactaaaaa aacatgttgcg atatcttgcg tttttttcc atggggggca 720
cgtttaagcc gttttttttt ttatccggca agtacaattt tttacttcc tcggatggat 780
aatttgcgtt catttttttgcg aacgttccatata tgcgttccgc tcggatggat 840
cagaatggcc agacatttcg aatgttccatata tgcgttccgc tcggatggat 900
tgaaggccggc ggcggaaagaa gtaacaaaagg aaccttaggg ctttttttttgcg ttttttttttgcg 960
tgcgtatccaa gggcccttca ttttttttttgcg ttttttttttgcg ttttttttttgcg 1020
aaagcgacaa agattttttttgcg ttttttttttgcg ttttttttttgcg ttttttttttgcg 1080
aagggttccggc ttgggttggat atggatccggc gtgtgggtttt agatgttgcg gggatgtcg 1140
tgggttccggc gtatgttgcgtt ttttttttttgcg ttttttttttgcg ttttttttttgcg 1200
tttggatgggg actatccatata aagggggggggg atgttgcgtt aagggttgcg ctttttttttgcg 1260
aaggccggccg ttttttttttgcg ttttttttttgcg ttttttttttgcg ttttttttttgcg 1320
aaggccggccg ttttttttttgcg ttttttttttgcg ttttttttttgcg ttttttttttgcg 1380
tacccacgtt atgttgcgtt atccaaagggg atgttgcgtt ttttttttttgcg ttttttttttgcg 1440
atggggatgtt ggcggccatata aaggccggccg atgttgcgtt ttttttttttgcg ttttttttttgcg 1500
cccgcatggc atggggatata atccacgggg gtttttttttgcg ttttttttttgcg ttttttttttgcg 1560
aaggccggccg ttttttttttgcg ttttttttttgcg ttttttttttgcg ttttttttttgcg 1620
tatatataca gggccacacggc agatgttgcgtt gggccacgttgcgtt ttttttttttgcg 1680
gtctcggtt ctttttttttgcg ttttttttttgcg ttttttttttgcg ttttttttttgcg 1740
agtttccattt ttatccatata aaggccggccg atgttgcgtt ttttttttttgcg ttttttttttgcg 1800
tgaaggccggc ttttttttttgcg ttttttttttgcg ttttttttttgcg ttttttttttgcg 1860
gcaatccatata aaggccggccg ttttttttttgcg ttttttttttgcg ttttttttttgcg 1920
atccatccatata aaggccggccg ttttttttttgcg ttttttttttgcg ttttttttttgcg 1980
ctcttacattt ttatccatata aaggccggccg ttttttttttgcg ttttttttttgcg 2040
acttccatata aaggccggccg ttttttttttgcg ttttttttttgcg ttttttttttgcg 2100
atagagacaa aatagaagaa accgttccatata ttttttttttgcg ttttttttttgcg 2160
tatccacttc ttgttccatata aaggccggccg ttttttttttgcg ttttttttttgcg 2220
gcctttatct tgaaaaatgtt cccggccggc ttttttttttgcg ttttttttttgcg 2280
gaggccggccg ttttttttttgcg ttttttttttgcg ttttttttttgcg ttttttttttgcg 2340
acggccatata aaggccggccg ttttttttttgcg ttttttttttgcg ttttttttttgcg 2400

```

-continued

---

aaaagaatac	taagatgctt	tgttagaaaa	atagcgctc	cgggatgcat	ttttgttagaa	2460
caaaaaaagaa	gtatagattc	tttgttgta	aatagcgct	ctcgcggtgc	atttctgttc	2520
tgtaaaaatg	cagctcagat	tctttgttg	aaaaattagc	gctctcgctg	tgcattttg	2580
tttacaaaaa	atgaagcaca	gattttcg	tggtaaaata	gcccggccgc	tttgcatttc	2640
tgttctgtaa	aatgcagct	cagatctt	gtttgaaaaa	ttagcgctc	cgcggtgc	2700
ttttgttcta	caaaaatgaag	cacagatgt	tcgttccagg	ggcactttc	ggggaaatgt	2760
gcccggaaacc	cttattttgtt	tatTTTCTA	aatacattca	aatatgtatc	cgctcatgag	2820
acaataaacc	tgatattgg	cagaatttgg	taattgg	taacactgac	cccttattgt	2880
ttatTTTCTC	aaatacattc	aaatatgtat	ccgctcatg	gacaataacc	ctgataatg	2940
cttcaataat	attggaaaag	gaagaatata	agtattcaac	atttccgtgt	cgcccttatt	3000
cccttttttg	cgccatTTTG	ccttctgtt	tttgc	caccaggt	ggtaaaggt	3060
aaagatgtc	agagtcgtt	gggtgcacga	gtgggttaca	tgcgacttga	tctcaacagc	3120
gtaagatcc	tttagatgtt	tcggccgg	gaacgtttt	caatgtgag	cactttaaa	3180
gttctgtat	gtggccgg	attatccgt	attgacgccc	ggcaagagca	actcggtc	3240
cgcatacact	attctcagaa	tgacttgg	gagta	tcac	aaagcatctt	3300
acggatggc	tgacagtaag	agaattatgc	agtgtgc	caacatgag	tgataacact	3360
gcccccaact	tacttctgc	aacatgcgg	ggaccgaa	agctaaccgc	tttttgcac	3420
aacatgggg	atcatgtaac	tcgcgttgc	cggtggaa	cgagatgt	tgaagccata	3480
ccaaacgac	agcgtgacac	cacatgc	gtacgcgt	caacaacgt	qcgcaacta	3540
ttaactggcg	aactacttac	tctagttcc	cgcaacaat	taatagactg	gatggaggcg	3600
gataaaagt	ttgggttgc	cggtgg	gtc	ccat	gtgaaaaga	3900
aaatccggag	ccgggtgc	ttgttctgc	ggat	tcat	cg	3660
aaggcccttcc	gtatgtat	tatccat	acttccgt	ggat	gtgtgg	3720
aatagacaga	tcgctgat	tttttttttt	acggggatc	aggcaactat	ggatgaa	3780
atccccat	gtggatgt	cgccgtgt	tttttttttt	tttttttttt	tttttttttt	3840
tcaaaggat	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	3900
aaccaccgc	accagggt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	3960
aggttaactgg	tttcagcaga	tttttttttt	tttttttttt	tttttttttt	tttttttttt	4020
tagccccacca	tttcaagaac	tttttttttt	tttttttttt	tttttttttt	tttttttttt	4080
taccatggc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	4140
agtttccgg	taaggcgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	4200
tgtagcgac	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	4260
cgatcccg	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	4320
tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	4380
agcgcac	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	4440
ggatgttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	4500
ggccatcttgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	4560
aaaacgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	4620
tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	4680
ttccatgat	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	4740
tactaaattt	ccccctttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	4800
gaaaagaaaa	aaatgttgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	4860
tttatcagtt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	4920
ccatttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	4980
tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	5040
ctaaatgttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	5100
tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	5160
tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	5220
tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	5280
tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	5340
tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	5400
tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	5460
tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	5520
tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	5580
tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	5640
tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	5700
tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	5760
tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	5820
tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	5880
tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	5940
tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	6000
tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	6060
tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	6120
tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	6180
tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	6240
tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	6300
tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	6360
tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	6420
tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	6480
tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	6540
tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	6600
tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	6660
tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	6720
tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	6728

---

SEQ ID NO: 79

moltype = DNA length = 6704

FEATURE

Location/Qualifiers

misc\_feature

1..6704

source

note = Plasmid pD1214-SA-hABAB

1..6704

-continued

---

mol\_type = other DNA  
 organism = synthetic construct

**SEQUENCE: 79**  
 ggtaaatca tgtaattagt tatgtcacgc ttacattca cccctcccc cacatccgt 60  
 ctaaccgaaa aggaaggagt tagacaacctt gaagtcagg tccctattttt tttttttata 120  
 gttatgttag tattaagaac gttatttata ttccaaattt ttctttttt tctgtacaga 180  
 cgcggtacat catgtacat tatactgaaa accttgcgtt agaagggttt gggacgtcg 240  
 aaggctttaa ttgcggccc ctcacctgcg cgcggggaa aatccatcat 300  
 ttctttttt ttcctttttt tgatgtcggt ttttttgaa ttcttttgat tcggtaatct 360  
 ccgaacagaga gagaaggacga aggaaggcgc acagacttag atttgtat atacgcattat 420  
 gtatgttga agaaacatga aatttgcgg tttttttttt tttttttttt tttttttttt 480  
 ccaggcggaa acgaagataa atcatgtcgaa aagttttttt tttttttttt tttttttttt 540  
 atccatgtcc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 600  
 gtatgttgcggtt accaccacggg aatccatgtcgaa gttatgttgcggtt 660  
 ccaaaatttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 720  
 cagtttgcgca gttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 780  
 aattttgttgcgca tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 840  
 cagaatggcgc agacattacg aatccatgtcgaa gttatgttgcggtt 900  
 tgaagcggc gggcggggaa gttttttttt tttttttttt tttttttttt tttttttttt 960  
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1020  
 aaaaacccggc aatccatgtcgaa gttatgttgcggtt 1080  
 aagggttgcgca tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1140  
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1200  
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1260  
 aagggttgcgca tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1320  
 aatccatgtcgaa gttatgttgcggtt 1380  
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1440  
 attttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1500  
 cccgcgttgcgca tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1560  
 aatccatgtcgaa gttatgttgcggtt 1620  
 aatccatgtcgaa gttatgttgcggtt 1680  
 gttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1740  
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1800  
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1860  
 gttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 1920  
 aatccatgtcgaa gttatgttgcggtt 1980  
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2040  
 aatccatgtcgaa gttatgttgcggtt 2100  
 aatccatgtcgaa gttatgttgcggtt 2160  
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2220  
 gttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2280  
 aatccatgtcgaa gttatgttgcggtt 2340  
 aatccatgtcgaa gttatgttgcggtt 2400  
 aatccatgtcgaa gttatgttgcggtt 2460  
 aatccatgtcgaa gttatgttgcggtt 2520  
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2580  
 aatccatgtcgaa gttatgttgcggtt 2640  
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2700  
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2760  
 aatccatgtcgaa gttatgttgcggtt 2820  
 aatccatgtcgaa gttatgttgcggtt 2880  
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2940  
 aatccatgtcgaa gttatgttgcggtt 3000  
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 3060  
 aatccatgtcgaa gttatgttgcggtt 3120  
 aatccatgtcgaa gttatgttgcggtt 3180  
 aatccatgtcgaa gttatgttgcggtt 3240  
 aatccatgtcgaa gttatgttgcggtt 3300  
 aatccatgtcgaa gttatgttgcggtt 3360  
 aatccatgtcgaa gttatgttgcggtt 3420  
 aatccatgtcgaa gttatgttgcggtt 3480  
 aatccatgtcgaa gttatgttgcggtt 3540  
 aatccatgtcgaa gttatgttgcggtt 3600  
 aatccatgtcgaa gttatgttgcggtt 3660  
 aatccatgtcgaa gttatgttgcggtt 3720  
 aatccatgtcgaa gttatgttgcggtt 3780  
 aatccatgtcgaa gttatgttgcggtt 3840  
 aatccatgtcgaa gttatgttgcggtt 3900  
 aatccatgtcgaa gttatgttgcggtt 3960  
 aatccatgtcgaa gttatgttgcggtt 4020  
 aatccatgtcgaa gttatgttgcggtt 4080  
 aatccatgtcgaa gttatgttgcggtt 4140  
 aatccatgtcgaa gttatgttgcggtt 4200  
 aatccatgtcgaa gttatgttgcggtt 4260  
 aatccatgtcgaa gttatgttgcggtt 4320  
 aatccatgtcgaa gttatgttgcggtt 4380  
 aatccatgtcgaa gttatgttgcggtt 4440  
 aatccatgtcgaa gttatgttgcggtt 4500  
 aatccatgtcgaa gttatgttgcggtt 4560

-continued

---

tgttcttcc	tgcggttatcc	cctgattctg	tggtataaccg	tattaccggc	tttgagttag	4620
ctgtatccgc	tcggggctgt	gcaggatata	cttcaaaatg	tttctactcc	tttttactc	4680
ttccagattt	tctcgactc	cgcgcacgc	cgtaccactt	aaaaacaccc	aagcacagca	4740
tactaaattt	ccccctttc	ttccctctagg	gtgtcgtaa	ttaccggtag	taaaggttt	4800
aaaaagaaaa	aagtgaccgc	ctcggtttc	tttcttcgtc	aaaaaaggca	ataaaaattt	4860
ttatcagttt	tctttttctt	gaaaatttt	ttttttgtt	ttttttctt	tcgatgacct	4920
ccattgtata	ttaaagttaa	taaacggact	tcaatttctc	aagtttca	ttcattttc	4980
tttgttctatt	acaactttt	ttacttcttgc	tcattagaa	agaaagcata	gcaatcta	5040
ctaagttaa	aatgaaatgg	gttacttttta	tctcaactttt	tttccctgttc	tccagcgctt	5100
actctatgc	ggtagatgt	ttggagacgg	ggggagggtt	ggtagaca	accgggggtc	5160
tgaggcttcc	ctgtgcgc	tctgggttca	cactggattt	ttcgtccata	gggtgggtt	5220
ggcagggcttcc	ttggcaagag	cgtgggggg	tctcatgtat	tagtagtagt	ggtgatagca	5280
caaagtacgc	cgattcgta	aaggcccggt	ttacaaccc	caggataat	gctaagaaca	5340
cgtatatact	ccatgtttt	ttctcgat	ccgacgata	ggccgtat	tactgtgcgg	5400
cttcaggcgc	gactatgtgc	ggcggttcc	ctctgagccc	ttacggcaag	gacgacttgg	5460
gcaagggggac	cttgggtgacc	gtatccctcg	gggggtgggg	gtctgggtt	ggaggcttc	5520
gggggtgggg	caggcagggtt	caactgggtt	aatctggggg	aggcttggta	caacctgggg	5580
gatcccttgat	actcttgc	gaggecccg	gattcaccc	ggactactat	ggcatcggt	5640
ggttccgc	ccccccagg	aaaggccgg	aggccgtttc	atataatttt	gccagtggcc	5700
ggaccataact	gtacgcagac	tctgtgaagg	gacgcttac	catctttagg	gacaatgcca	5760
aaaatgtgt	gtacgttgc	atgaacagcc	tcaaggccgg	ggataccgc	gtgtactact	5820
gcgcgagac	gcgttctcc	gcttctagec	tgaatagat	gtctggccac	gactacgac	5880
tgtggggac	gggcacacag	gtggctgt	cttccgggtt	cgaaagccgg	ggggccagcg	5940
gggggtgggg	cggtgggggc	agccaactgc	agctggtaga	gacagggggc	ggctttagtt	6000
agccgtggagg	gtctcttgc	ctgtcatgt	cttgccttc	agtactac	6060	
tgatgatc	ggtccggccaa	gctccaggaa	agggggctgt	gttgcgttgc	actattaa	6120
catgttgc	cacaatgcgg	gacacttca	caaaaggccgg	tttcacattt	tccagagaca	6180
acgccaagaa	tactctgtac	cttcagatga	ccagttgtaa	ccccgaggac	actgtctgt	6240
actattgtgc	aaggccgg	gtatcttgc	cttccgttat	cagaggccca	gtaaggggcc	6300
ctggaacaca	ggttaaccgtt	tcatccgggg	ggggcggtt	aggccgtgg	ggatctggcg	6360
gggggtgggg	ccaaatgttgc	ctgggtcaat	ccggggccgg	acttgcgttgc	acagggggct	6420
cctcgaggct	ctccctgtgc	tcttcggaa	gcatcgccgg	cttcgagacc	gtgacctgtt	6480
ctgcggccgg	tcccgggaaag	tctctgcgt	gggtcggttc	catgactaa	actaacaacg	6540
agatctactt	tgactctgt	aaaggccgc	tcatcatttt	tagagataac	gctaaaaaaca	6600
catgttatct	gcatgttgc	agtttca	ctgaagacac	aggccgttat	ttctgttaagg	6660
gtcctgtact	ggggccgg	ggcatccagg	taacagtctc	gagt		6704

---

SEQ ID NO: 80	moltype = DNA	length = 6574				
FEATURE	Location/Qualifiers					
misc_feature	1..6574					
source	note = Plasmid pCEV-G4-Km					
	1..6574					
	mol_type = other DNA					
	organism = synthetic construct					
SEQUENCE: 80						
ggagcgcacct	catgtatatac	ctgagaaa	aacctgaccc	acagggaaaga	gttactcaag	60
aataagaattt	tgcgttttaa	aacctaagag	tcactttaaa	atttgcatac	acttattttt	120
tttataactt	atthaataat	aaaaatcata	aatcataaga	aattcgotta	ttttagaagtgc	180
tcaacaacgt	atcttaccaac	gatttgcac	ttttccatct	tttcgttaat	ttctggcaag	240
gtagacaacgc	cgacaaaccc	gatttgcac	tttgcacca	ttctggccaa	gaatttgc	300
ttaaagatgt	agatcttatac	gtctgtatcc	ttgttaatcc	tcgtactact	ttttttgtt	360
aaaatttttttt	aaactttttt	ttttttgtt	tattttttgt	tcttagaaaa	gacaagtgc	420
gcttgcgttgc	tcttgcgtt	tttattttttt	acaatagct	caatgcaga	atagattcga	480
acatttgc	gtatttgcac	atatctgtc	tattttatct	ttttttttt	cagtctact	540
atattttgtt	ttttcttct	tttccatctt	tcgttgcac	ataagttaca	ggggatctcg	600
aaagatgtt	gggattttcc	tttcaaaagac	qacttttgc	catcaattt	ttccctgttgc	660
cctctgaaaa	ttatccagca	gaagcaat	taaaagatga	acctcaga	aacacgcagg	720
ggcccgaaat	tgttcttgc	agaagatgt	ggtcataaaa	agttttatcc	ctggaaaaaaa	780
atattttgtt	tgcctttct	gagaattttt	tcgaatttgc	ttgtgc	ccac tgcgttgc	840
tctcgatgtt	gtggccat	ttccacatgt	ttgttgcgtt	aaaaaaaaaa	agatttctca	900
ccccgcacaa	atccctgc	acccttcatt	tccacggggc	cgccgc	cacaatagct	960
tcaaaatgtt	tctacttctt	ttttacttac	ccagattttc	tcggactcc	cgccatcgcc	1020
taccacttca	aaacacccaa	gcacacata	ctaaatttcc	cctcttttctt	cctcttaggg	1080
gtcggttattt	accgtacta	aagggtttgg	aaagaaaaaa	gagaccgcct	cgttttttt	1140
tcttcgttgc	aaaaggcaat	aaaaattttt	atcacgttcc	tttttttttgc	aaatttttt	1200
ttttgtttttt	tttcttttcc	gatgaccttcc	cattgtat	taagttata	aacggcttcc	1260
atattttcttca	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	1320
catttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	1380
ctcaactatgtt	ggcccgccgg	tcgacatgtt	acagaatgtt	atttccgaag	aagacctcg	1440
gtaagcttgg	taccggggct	agctaaagatc	cgctctaa	gaaaaggaa	gagtttagaca	1500
acctgaatgt	taggttccat	tttattttttt	tatagttat	tttagtattaa	gaacgttatt	1560
tatattttca	attttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	1620
gaaaaccccttgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	1680
ccaaacgcgc	ggggagggcg	tttttgcgtt	tttttttttt	tttttttttt	tttttttttt	1740
ctcgctgcgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	1800
acgggttattcc	acagaatc	gggataacgc	aggaaagaac	atgtgagcaa	aaggccagca	1860
aaaggccagg	aaccgtaaaa	aggccgcgtt	gttggcggtt	ttccataggg	tccgcccccc	1920
tgacgagcat	cacaaaaatc	gacgttca	tcagagggttgc	cgaaacccga	caggactata	1980
aagataccatcg	gggtttcccc	cttggaaatcg	cctctgtgc	tctctgttgc	cgaccctgccc	2040

-continued

---

gcttacccgg	tacctgtccg	cctttctcc	ttcgggaa	gtggcgctt	ctcatagctc	2100
acgcgttag	tatctcagtg	cggtgttagt	cggtcgctc	aagctgggt	gtgtgcacga	2160
acccccccgt	cagccccacc	gctgcgcct	atccggtaac	tatcgctt	agtccaaccc	2220
ggtaagacac	gacttatcgc	caactggcago	agccacttgt	aacaggattt	gcagagcgag	2280
gtatgttagc	gggtgtacag	agtccgttgc	gtggggccct	aactacggct	acactagaag	2340
gacagtattt	ggtagatcgc	ctctgttcgaa	ggcaggattt	ttccggaaaaa	gatgtttgtt	2400
ctcttgcatt	ggccaaacaaa	ccacccgtgg	tagcgttgtt	ttttttgtt	gcaaggcagca	2460
gattacgcgc	agaaaaaaag	gatctcaaga	agatccctt	atctttctt	cggggctgt	2520
cgtcgttgg	aacgaaaact	cacgtaagg	gattttggtc	atgagattat	caaaaaggat	2580
cttcacatcg	atccccctaa	ataaaaaaat	aagtttttaaa	taatcttataa	gtatataatga	2640
gtaaacttgg	tctgcacgat	aaacaatgtt	aatcgttgc	gcacatctt	cagcqatctg	2700
tctatccat	tcatccatag	ttgcgtact	cccgttgt	tagataacta	cgatcggga	2760
gggcttacca	tctggccca	gtgcgtcaat	gataccgcga	gaccacgtt	caccggctcc	2820
agattatca	gcaataaaacc	agccagccgg	aaggggccag	cgccaggat	gtctgcacac	2880
tttacccgc	tccatccat	cttataattt	ttggccggaa	gctagat	gtatgtcgcc	2940
agttaatagt	ttgcgcacg	ttgttgcatt	tgctacagg	atcgttgtt	cacgctgc	3000
gtttggat	gttcgttgc	gtccgggtt	ccaaacgtca	aggcgat	catgtatccc	3060
catgttgc	aaaaaaagcg	ttatgttcc	cgggtttcc	atcgttgt	gaagtaagt	3120
ggccgcgtg	ttatcaact	ttgttattgg	agcactgtat	aattcttta	ctgtatgtcc	3180
atccgtaa	tgctttctg	tgactgttgc	gtactcaacc	aagtcttctt	gagaataatgt	3240
atgcggcga	ccgggttgc	cttgcggcc	gtcaatcgg	gataatacc	cgccacatag	3300
cagaacttta	aaagtgtca	tcattttggaa	acgttcttgc	ggggcaaaac	tctcaaggat	3360
cttacccgt	tttgatgtca	ttgtcgat	accactgt	gcacccaa	gatcttcag	3420
atctttact	ttcaccagcg	tttctgggt	agcaaaaaca	ggaaggcaaa	atgcggcaaa	3480
aaaggaaata	aggccgacac	ggaaaatgtt	aatactca	cttctcc	ttcaatattta	3540
ttgaaggatt	tatcagggtt	attgtctcat	gacggat	atatttgaat	gtatgttagaa	3600
aaataaaca	atagggggtt	tcgtccat	ttcccgaaaa	gtgcaatct	aacgaatgt	3660
ctgtgctca	ttttgttagaa	caaaaatgc	acgcgagag	gtaat	caaacaaga	3720
atctgatgt	catatataca	gaacagaaa	gcaacgcga	agcgat	taccaacgaa	3780
gaatctgtc	ttcacattttt	taaaaacaaa	atgcacgc	agacgcgtaa	ttttccaaac	3840
aaagaatctg	agctgcattt	ttacacaa	gaaatgc	gcccgc	tatattacca	3900
acaagaata	tatacttctt	ttttgttca	aaaaatgc	tcggagag	gtatttttgc	3960
taacaaagca	tcttagatta	cttttttgc	tttgcgttgc	tcctataat	cagtcgttgc	4020
ataactttt	gactgttgc	ttccgttgc	tttagaaga	gtaat	tttttttttgc	4080
cttccat	aaaaaaagcc	gactcactt	cccgcggtt	ctgattact	gcaagacgt	4140
gggtgcattt	tttcaagata	aaggcatccc	cgatttat	ctatccgat	gtggatgtcg	4200
catactttgt	gaacagaaag	tgatagcgtt	gatgattt	cattgttgc	aaaattatga	4260
acggtttctt	ctatgttgc	tctatataact	acgtat	tttttttca	tttttgcatt	4320
gtttcgatt	cactctat	atagtttca	tcataat	tttgcattaa	gagaatact	4380
agagataaac	aaaaaaatgt	tagaggtca	gtttagatgc	aatttca	agcgaaggat	4440
ggatgggtag	gttatata	tttttttttgc	gatatacgac	agatataat	agcaaaagaga	4500
caatgtttgt	ggaaaggcgta	tttgcattttt	tttagtgcgt	cgatcgtc	ccgtgcgtt	4560
ttgggtttt	gaaaggcgt	tttgcattttt	cttcagagc	tttttttgc	tctgttgc	4620
ctatactttc	tagaaatag	gaaacttcg	atagaaat	tttttgcattt	tttttttgc	4680
gcgcgttccg	aaatgcacg	cgacgtgc	acatacg	cacttgttac	gtgcacact	4740
atctgttgc	ttgcgttgc	atataatata	atgaga	aaaggat	tttttttttgc	4800
ctttaatcg	taattatgt	cgatcatttt	ttgttgcgtt	tttttttttgc	tttttttttgc	4860
gtgtatattt	ccccatccat	gggggttgc	gtatgttgc	tttttttgc	tttttttttgc	4920
gttctatat	tcgtccactt	tcaattttgt	ttgttgc	tttttttgc	tttttttttgc	4980
tttgcattttt	atcatgttgc	acaacccat	atataactt	tttttttgc	tttttttttgc	5040
tttgcattttt	tttttttgc	tttttttgc	tttttttgc	tttttttgc	tttttttttgc	5100
gacatggagg	ccccatccat	tttttttgc	tttttttgc	tttttttgc	tttttttttgc	5160
tgtgtatgtc	ccccatccat	tttttttgc	tttttttgc	tttttttgc	tttttttttgc	5220
cattttgtc	ccccatccat	tttttttgc	tttttttgc	tttttttgc	tttttttttgc	5280
caggaaacac	ccccatccat	tttttttgc	tttttttgc	tttttttgc	tttttttttgc	5340
aatataaaatg	tttttttgc	tttttttgc	tttttttgc	tttttttgc	tttttttttgc	5400
tttgcattttt	tttttttgc	tttttttgc	tttttttgc	tttttttgc	tttttttttgc	5460
ctcacgttgc	tttttttgc	tttttttgc	tttttttgc	tttttttgc	tttttttttgc	5520
gggtctcg	tttttttgc	tttttttgc	tttttttgc	tttttttgc	tttttttttgc	5580
atgcgttgc	tttttttgc	tttttttgc	tttttttgc	tttttttgc	tttttttttgc	5640
agatggatcg	tttttttgc	tttttttgc	tttttttgc	tttttttgc	tttttttttgc	5700
ccgtacttcc	tttttttgc	tttttttgc	tttttttgc	tttttttgc	tttttttttgc	5760
aggtttaga	tttttttgc	tttttttgc	tttttttgc	tttttttgc	tttttttttgc	5820
ttggccgtt	tttttttgc	tttttttgc	tttttttgc	tttttttgc	tttttttttgc	5880
gttctgttgc	tttttttgc	tttttttgc	tttttttgc	tttttttgc	tttttttttgc	5940
acgagcgat	tttttttgc	tttttttgc	tttttttgc	tttttttgc	tttttttttgc	6000
tcccacccgg	tttttttgc	tttttttgc	tttttttgc	tttttttgc	tttttttttgc	6060
agggggaaattt	tttttttgc	tttttttgc	tttttttgc	tttttttgc	tttttttttgc	6120
atcttgcatt	tttttttgc	tttttttgc	tttttttgc	tttttttgc	tttttttttgc	6180
tttcaaaaata	tttttttgc	tttttttgc	tttttttgc	tttttttgc	tttttttttgc	6240
atgtgtttttt	tttttttgc	tttttttgc	tttttttgc	tttttttgc	tttttttttgc	6300
tgtatgtttt	tttttttgc	tttttttgc	tttttttgc	tttttttgc	tttttttttgc	6360
tttttttgc	tttttttgc	tttttttgc	tttttttgc	tttttttgc	tttttttttgc	6420
atgcgttgc	tttttttgc	tttttttgc	tttttttgc	tttttttgc	tttttttttgc	6480
ccatccatcc	tttttttgc	tttttttgc	tttttttgc	tttttttgc	tttttttttgc	6540
tatacgtat	tttttttgc	tttttttgc	tttttttgc	tttttttgc	tttttttttgc	6574

---

-continued

---

**misc\_feature** 1..8211  
 note = Plasmid pCEV-G4-Km-TEF-AT-hABAB\*  
**source** 1..8211  
 mol\_type = other DNA  
 organism = synthetic construct  
**SEQUENCE: 81**  
 gccttaggatc agatttcctt caattttac tgctgttta ttgcgcagcat cctccgcatt 60  
 agctggggcc cacgtgtatgc aggttacatgt ggtggagacgg gggggagggc tggtacaacc 120  
 aggccgggtca ctggaggctt cctgtggccg atctgggttc acactggatc attcgtccat 180  
 agggtgttgc cggcaggctc ctggcaagaaga gcggtgggggt gtctcatgtt tttagtag 240  
 tggtgtatgc acaaagtacg ccgtatcccgtt aaaggccggg tttacaacctt ccaggatataa 300  
 tgctaaagaac accgtatatac tccagatgaa ctctctgaag cccgaecata cggccgtata 360  
 ttactgtcgcc gttttcaggg cgtatgttg cgccgtgttc cttctgaggcc ttacggc 420  
 ggacgactgg ggcaaggggg ccctgtgtac ctgtatccta ggcgggtgggg ggtctggg 480  
 gggagggtca ggggggtgggg gcaaggccgtt gcaactgtt gaaatctggggg gaggctttgg 540  
 acaacactggg ggatccctga gactctttg cgaggcctcc ggattcacct tggactacta 600  
 tggcatacgcc tgggttcggc agccccccgg gaaggagcggg gaggccgtt catacatatag 660  
 tgcagtcggcc cggaccatac tgtaacggaga tgatgtgtggaa ggacgatcca ccatcttgc 720  
 ggacaatggcc aaaaatgtcg tgtaacttca aatgaacagc ctcgaaggccg aggataccgc 780  
 agtgtactac tgcgegagac ggccgttctc cgcttcttgcgtt gtaatagat ggctggccga 840  
 cgactcagac gtgtggggac gggggacacaa ggtggctgtg tttccgggtg gcgaaaggccg 900  
 agggggcage ggggggtgggg ggggtggggg cagccaaactg cagttggtag agacaggggg 960  
 cggttcaatgtt cagccgtggc ggttccatcgt gctgttccatgtt gctttaccc 1020  
 cagtgtactac gtgtatgc ggggtccggca agtccagggg aaggggccgtt agtggatccgc 1080  
 tactattaaat acagatggca gcacaatcgccg ggacgacttcc acaaaaggccg ggtttccat 1140  
 ttccagagac aacggccaaa atactctgtt ctttccatgtt accgttctgaa aaccggaggaa 1200  
 cactgcttcgt tactattgtt caagggccg ggttcatccatgtt gttccatcgtt tcagaggccgc 1260  
 agtaaggggc cctggaaacac aggttacccgtt ttcatccgggg gggggccgtt cagggccgttgg 1320  
 gggatctggc ggggggtggat cccaaatgtca gctggatccggaa tccggggggccg gactgttccaa 1380  
 gagagggggc ttccctggggcc tttccatgtt atttccggca agcatccggc ggttccggagac 1440  
 cgtgacatggcc tttccatgtt ccggggggcc gtttccatgtt cccatgttccatgtt 1500  
 gactaacaac gagatctact ctgactcgtt gaaaggccgc ttccatccatgtt cttagatgtt 1560  
 cgtataaaaac acagtgtatc tgccatgtt gtttccatgtt cctgttccatgtt tcagggccgtt 1620  
 ttctgttgc ggttccatgtt tgccatgtt gggccatgtt gtaacatgtt caggtgtccgc 1680  
 cgttccatgtt gtttccatgtt ccggggggcc gtttccatgtt accgttccatgtt aataccgtt 1740  
 gtttccatgtt ttccatgtt ttccatgtt atttccatgtt taagaacgtt atttccatgtt 1800  
 caaattttc ttccatgtt gtttccatgtt gtacagacgc gtttccatgtt gtaacatgtt actgttccatgtt 1860  
 ttggcttcgttca ggggtttgggg acgttccatgtt atccagccgtt gtttccatgtt actgttccatgtt 1920  
 gggggggcggcc ggggtggggcc gtttccatgtt tttccatgtt cctgttccatgtt tgactccgtt 1980  
 cgttccatgtt tttccatgtt cccatgtt cttccatgtt gtttccatgtt gtttccatgtt 2040  
 tccacagaat cccatgtt ccggggccggcc gtttccatgtt aacatgtt gtttccatgtt 2100  
 aggaacgttca aaaaaggccgc gtttccatgtt tttccatgtt ggctccggccccc ccctgttccatgtt 2160  
 catcacaaaa acatgttccatgtt gtttccatgtt aacatgtt gtttccatgtt 2220  
 cccatgttccatgtt ccggggccggcc gtttccatgtt cttccatgtt gtttccatgtt 2280  
 ggataccatgtt ccggggccggcc gtttccatgtt aacatgtt gtttccatgtt 2340  
 aggttccatgtt gtttccatgtt gtttccatgtt cttccatgtt gtttccatgtt 2400  
 gtttccatgtt gtttccatgtt cttccatgtt aacatgtt gtttccatgtt 2460  
 cccatgttccatgtt ccggggccggcc gtttccatgtt aacatgtt gtttccatgtt 2520  
 gggggccggcc gtttccatgtt aacatgtt gtttccatgtt 2580  
 ttgttccatgtt gtttccatgtt gtttccatgtt aacatgtt gtttccatgtt 2640  
 tccggccaaacccatgtt gtttccatgtt aacatgtt gtttccatgtt 2700  
 cccatgttccatgtt ccggggccggcc gtttccatgtt aacatgtt gtttccatgtt 2760  
 tggacccatgtt ccggggccggcc gtttccatgtt aacatgtt gtttccatgtt 2820  
 tagatccatgtt ccggggccggcc gtttccatgtt aacatgtt gtttccatgtt 2880  
 ttgttccatgtt ccggggccggcc gtttccatgtt aacatgtt gtttccatgtt 2940  
 cccatgttccatgtt ccggggccggcc gtttccatgtt aacatgtt gtttccatgtt 3000  
 cccatgttccatgtt ccggggccggcc gtttccatgtt aacatgtt gtttccatgtt 3060  
 tccatgttccatgtt ccggggccggcc gtttccatgtt aacatgtt gtttccatgtt 3120  
 gtttccatgtt aacatgtt gtttccatgtt aacatgtt gtttccatgtt 3180  
 agtggccatgtt ccggggccggcc gtttccatgtt aacatgtt gtttccatgtt 3240  
 atggccatgtt ccggggccggcc gtttccatgtt aacatgtt gtttccatgtt 3300  
 tccatgttccatgtt ccggggccggcc gtttccatgtt aacatgtt gtttccatgtt 3360  
 gtttccatgtt aacatgtt gtttccatgtt aacatgtt gtttccatgtt 3420  
 agtggccatgtt ccggggccggcc gtttccatgtt aacatgtt gtttccatgtt 3480  
 ccggccatgtt ccggggccggcc gtttccatgtt aacatgtt gtttccatgtt 3540  
 tccatgttccatgtt ccggggccggcc gtttccatgtt aacatgtt gtttccatgtt 3600  
 tccatgttccatgtt ccggggccggcc gtttccatgtt aacatgtt gtttccatgtt 3660  
 actttccatgtt ccggggccggcc gtttccatgtt aacatgtt gtttccatgtt 3720  
 ataaaggccgc aacccatgtt ccggggccggcc gtttccatgtt aacatgtt gtttccatgtt 3780  
 atttccatgtt ccggggccggcc gtttccatgtt aacatgtt gtttccatgtt 3840  
 caaataggggg ttccatgtt ccggggccggcc gtttccatgtt aacatgtt gtttccatgtt 3900  
 tccatgttccatgtt ccggggccggcc gtttccatgtt aacatgtt gtttccatgtt 3960  
 ctggccatgtt ccggggccggcc gtttccatgtt aacatgtt gtttccatgtt 4020  
 tgggttccatgtt ccggggccggcc gtttccatgtt aacatgtt gtttccatgtt 4080  
 ctggccatgtt ccggggccggcc gtttccatgtt aacatgtt gtttccatgtt 4140  
 atccatgttccatgtt ccggggccggcc gtttccatgtt aacatgtt gtttccatgtt 4200  
 gcatcttccatgtt ccggggccggcc gtttccatgtt aacatgtt gtttccatgtt 4260  
 ttggccatgtt ccggggccggcc gtttccatgtt aacatgtt gtttccatgtt 4320  
 ataaaaaaaggccgc aacccatgtt ccggggccggcc gtttccatgtt aacatgtt gtttccatgtt 4380

-continued

---

tttttcaag ataaaggcat ccccgattt attctatacc gatgtggatt gcgcatactt 4440  
 tgtaacaga aagtgatagc gttgatgatt ctccattgtt cagaaaattt tgaacggttt 4500  
 ctcttatTTt gtctctatTTt actacgtata ggaaatgttT acatTTTcgtt attgtttcg 4560  
 attcactcta tgaatagttc ttactacaat tttttgttT aaagagtaat actagagata 4620  
 aacataaaaa atgttaggtt caggttttgc tgcagggtttc aggagggaaa ggtggatggg 4680  
 taggttatTTt agggatatag cacaggttttgc tttttttttt tttttttttt 4740  
 tttttttttttttt ggaaatgtt gttgtttttt tttttttttt tttttttttt 4800  
 tttttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 4860  
 tttttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 4920  
 cgaaaatgtc acgcgatgtc cgcacataca gtcactgtt ctcgtcgtttt cttatcgtc 4980  
 gtgttgcgtt tttttttttt tttttttttt tttttttttt tttttttttt 5040  
 gtcgtactttt atgcgtctt tttttttttt tttttttttt tttttttttt 5100  
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 5160  
 atgttgcgtc tttttttttt tttttttttt tttttttttt tttttttttt 5220  
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 5280  
 aggtcttagag atctgtttt cttttttttt tttttttttt tttttttttt 5340  
 agggcccaagttt tttttttttt tttttttttt tttttttttt tttttttttt 5400  
 gtcgtccgtt tttttttttt tttttttttt tttttttttt tttttttttt 5460  
 atggcccac tttttttttt tttttttttt tttttttttt tttttttttt 5520  
 acgtccccctt cttttttttt tttttttttt tttttttttt tttttttttt 5580  
 aagggttaaggcc tttttttttt tttttttttt tttttttttt tttttttttt 5640  
 gatacgttc tttttttttt tttttttttt tttttttttt tttttttttt 5700  
 ttccaggccc tttttttttt tttttttttt tttttttttt tttttttttt 5760  
 cgataatgtc tttttttttt tttttttttt tttttttttt tttttttttt 5820  
 aggtttgtttt tttttttttt tttttttttt tttttttttt tttttttttt 5880  
 caacatgcg tttttttttt tttttttttt tttttttttt tttttttttt 5940  
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 6000  
 agaagaatat tttttttttt tttttttttt tttttttttt tttttttttt 6060  
 gtttgcatttc tttttttttt tttttttttt tttttttttt tttttttttt 6120  
 tcaggcccaat tttttttttt tttttttttt tttttttttt tttttttttt 6180  
 taatggcc tttttttttt tttttttttt tttttttttt tttttttttt 6240  
 ggttccgtt tttttttttt tttttttttt tttttttttt tttttttttt 6300  
 attataggtt tttttttttt tttttttttt tttttttttt tttttttttt 6360  
 catccatgtt tttttttttt tttttttttt tttttttttt tttttttttt 6420  
 atatgttattt tttttttttt tttttttttt tttttttttt tttttttttt 6480  
 ttttttttta tttttttttt tttttttttt tttttttttt tttttttttt 6540  
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 6600  
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 6660  
 aatcgtatgtt tttttttttt tttttttttt tttttttttt tttttttttt 6720  
 gtttgcgttt tttttttttt tttttttttt tttttttttt tttttttttt 6780  
 agtttattttt tttttttttt tttttttttt tttttttttt tttttttttt 6840  
 ctggaccatcc tttttttttt tttttttttt tttttttttt tttttttttt 6900  
 ctttttaatggcc tttttttttt tttttttttt tttttttttt tttttttttt 6960  
 cataagaatat tttttttttt tttttttttt tttttttttt tttttttttt 7020  
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 7080  
 accaaacccctt tttttttttt tttttttttt tttttttttt tttttttttt 7140  
 taatccatgtt tttttttttt tttttttttt tttttttttt tttttttttt 7200  
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 7260  
 atagctgcca tttttttttt tttttttttt tttttttttt tttttttttt 7320  
 ttatcttttt tttttttttt tttttttttt tttttttttt tttttttttt 7380  
 gtttgcataa tttttttttt tttttttttt tttttttttt tttttttttt 7440  
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 7500  
 aagatgaacc tttttttttt tttttttttt tttttttttt tttttttttt 7560  
 cataaaaaatgtt tttttttttt tttttttttt tttttttttt tttttttttt 7620  
 aatttgcgtt tttttttttt tttttttttt tttttttttt tttttttttt 7680  
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 7740  
 acggggcccg tttttttttt tttttttttt tttttttttt tttttttttt 7800  
 gtttgcgtt tttttttttt tttttttttt tttttttttt tttttttttt 7860  
 aatattttttt tttttttttt tttttttttt tttttttttt tttttttttt 7920  
 gaaaaaaatgtt tttttttttt tttttttttt tttttttttt tttttttttt 7980  
 acgttttttt tttttttttt tttttttttt tttttttttt tttttttttt 8040  
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 8100  
 ctattacaac tttttttttt tttttttttt tttttttttt tttttttttt 8160  
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 8211

SEQ ID NO: 82 moltype = DNA length = 8202  
 FEATURE Location/Qualifiers  
 misc\_feature 1..8202  
 note = Plasmid pCEV-G4-Km-TEF-IVS-hABAB\*  
 source 1..8202  
 mol\_type = other DNA  
 organism = synthetic construct

SEQUENCE: 82  
 catgttactt caaggctttctt ttttttttttt tttttttttt tttttttttt tttttttttt 60  
 cggccacgtt atgcagggtt cttttttttt tttttttttt tttttttttt tttttttttt 120  
 gtcactgtttt cttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 180  
 gtttccggat gtttccggat gtttccggat gtttccggat gtttccggat gtttccggat 240  
 tagcacaaatg tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 300  
 gaacacccgtt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 360

-continued

---

tgccggcttgc	aggggcacta	tgtgcccgt	gttcccctctg	agcccttacg	gcaaggacga	420
ctggggcaag	gggaccctgg	tgaccgtato	ctcaggcggt	ggggggctcg	gtgggggagg	480
ctcaggggggt	ggagggcagcc	aggtgcaact	ggttgaatct	ggggggagct	tggtacaacc	540
tgggggatcc	ctgagactc	cttgcggag	ctccggattc	accttggact	actatggcat	600
cggctgggtc	cgccaccccc	caggaaaggga	gcggggaggcc	gtttcataca	ttagtgccag	660
tgccggacc	atactgtacg	cagactctgt	gaagggacgc	tttaccatct	ctagggacaa	720
tgccaaaat	gtgtgttacc	tgcaaatgaa	cagaoctaag	cgggaggata	ccgcagtgta	780
ctactgacgc	agacggcgct	tctccgcttc	tagcgtgaat	agatggctgg	ccgacgacta	840
cgcacgtgtgg	ggacggggcc	cacagggtgc	tgtgtcttc	gggtggggaa	gcggaggggg	900
cagcgggggt	ggggcgggtg	ggggcggcc	actgcagctg	gttagagacag	ggggggggct	960
agttcagoc	gggggggtctc	tgcactgtc	atgcgtgc	tctggcttca	ccttcagtga	1020
ctacgtatg	acatgggtcc	gccaagetc	agggaaagggg	cctgagtgg	tcgctactat	1080
taatacatag	ggcgcacaa	tgcggggcga	ctccacaaag	gggcgggtca	ccatitccag	1140
agacaacgc	aagaataactc	tgtacccctca	gtgaggacgt	ctgaaacccg	aggacactgc	1200
tctgtactat	tgtcaagag	gcgggggtat	ctctgcttc	gtatcagag	gcgcagtaag	1260
ggggccctgga	acacaggtaa	ccggttcatc	cgggggggge	gttgcaggcg	gtgggggatc	1320
tggcgggggt	ggatcccaag	ttcagctgtt	cgaaatccggg	ggcggactgg	tccagacagg	1380
gggctcgttgc	aggctcgttgc	gtgcacatc	cggaagccte	gcccggcttgc	agaccgtgac	1440
cttgcgtctgc	caggctcccg	ggaaatctct	qcaatgggtc	gttccatgat	ctaagactaa	1500
caacgagatc	tactctgact	cagtgaaagg	ccgcatttc	atttctagag	ataacgctaa	1560
aaacacagt	tatctgcaga	tgaatagtc	caaaccgtaa	gacacaggcg	tgtatttctg	1620
taagggttcc	gagctgagggg	gccaggccat	ccaggtaaca	gtctcgat	aagggtaccgc	1680
ggctagatca	gatccgctc	aaccggaaag	gaaggaggta	gacaacctga	agtctaggtc	1740
cctattttt	tttttatag	tatgttagt	ttaagaacgt	tattttatatt	tcaaattttt	1800
ctttttttc	tgtacagacg	cgtgtacgc	tgttaacatt	tactgaaaac	cttgcgttgc	1860
agggttttgg	gacgctcgaa	gatccacgt	cattaatgaa	toggccaaacg	cgccggggaa	1920
ggcgggttgc	gttattggcg	cttccctcg	tcttcgttca	ctgcacgcgt	gcgcgtcg	1980
gttcgggtgc	ggcgacgggt	atcagctac	tcaaaggccg	taatacgggt	atccacagaa	2040
tcagggata	acgcagggaaa	gaacatgtg	gcaaaaggcc	agcaaaaggc	caggAACCGT	2100
aaaaggccgc	cggtgtgtgc	gttttccat	aggctccgc	ccccctgacga	gcataccaaa	2160
aatcgacgc	caagtgcag	tgccggaaac	ccgacaggac	tataaagata	ccaggcg	2220
ccccctggaa	gttccctcg	gcgcctctc	gttccgacc	tgccgttac	cggtacactg	2280
tcogcccttc	tcccttcggg	aagctggcg	cttctctata	gctcacgcgt	taggtatctc	2340
agttcggtgt	aggctgtgt	cttcaagct	ggetgtgtgc	acgaacccccc	cggtcagccc	2400
gaccgcgtcg	cottatccgg	taactatctg	cttgcgttca	accccgta	acacgactta	2460
tcgcccactgg	cagcagccac	tggtaacagg	attagcagag	cgaggatgt	aggcggtgt	2520
acagagtct	tgaagtgg	gcctaaactac	ggctacacta	gaaggacat	attttggat	2580
tgcgcttcgc	tgaaggccat	tacccctcg	aaaagagttt	gtatgccttgc	atccggcaaa	2640
caaaccaccc	ctggtagccg	tggttttt	gttgcgttgc	agcagattac	gcgcagaaaa	2700
aaaggatctc	aagaagatcc	tttgatctt	tctacgggtt	ctgacgtca	gtggaaacgaa	2760
aactcacgtt	aagggttttt	ggtcatgaga	ttatcaaaaa	ggatcttac	ctagatcctt	2820
ttaaattaaa	aatagat	taatcaatc	taaagtatata	atgagtaaac	tttgtctgc	2880
agttaccaat	gottaatcag	tgaggacact	atctcagcg	tctgttctt	tcgttcatcc	2940
atagtgcct	gactccccgt	cgtttagata	actacgat	gggagggtt	accatctggc	3000
cccagtgc	caatgatacc	gctgagaccca	cgctcaccgg	ctccagattt	atcagcaata	3060
aaccgcgc	ccggggccgg	cgagcgcaga	agtgtgtctc	caacttatac	cgccctccatc	3120
cagtctata	attttgtccg	ggaagatc	gtaatgttgc	cgccgtttaa	tagttgcgc	3180
acagtgtgt	ccattgtctac	aggatcg	gtgtcgtcg	cgtcgttgg	tatggcttgc	3240
ttcagctccg	gttcccaacg	atcaaggcg	gttacatgt	ccccccatgtt	gtgcaaaaaa	3300
ggcggttgc	cttccgttgc	tccgcgttgc	gtcagaatgt	agttggccgc	agtgttatac	3360
ctcatgttgc	ttggcgtatc	gtataattct	cttactgtc	tgcccatccgt	aaatgtctt	3420
tctgtgtact	tggtgtactc	aaccaaatc	tcttcgaaat	agtgtatgc	gcgcaggagt	3480
tgctcttcgc	ccggcgtcaat	acgggataat	accgcgcac	atagcagaac	ttttaaaatgt	3540
ctcatatgt	aaaaacatcc	tccggggcc	aaactctca	ggatcttacc	gtctgttgc	3600
tccagttcgc	tgttaacccac	tggtgcaccc	aactgtat	cagcatctt	tactttcacc	3660
acgcgtttgc	gtgtgtgttgc	aaacaggaa	caaaaatcc	caaaaaagggg	aataaggccg	3720
acacggaaat	gttgaatact	catacttcc	cttttcaat	attattgt	cattttatcg	3780
ggttttatgtc	tcatgacgg	atatactt	gaatgttgc	agaaaaataa	acaaatagg	3840
gttccgcgc	cattttcccg	aaaatgtcc	cctgaacgc	gcatgttgc	ttcattttgt	3900
agaacacaaa	tgcaacgcga	gagcgttaat	tttcaaca	agaatctg	gtgttgcatt	3960
taacaaacac	aatatgcacg	cgaaagcgct	attttacaa	cgaagaatct	gtgttgcatt	4020
tttgtaaac	aaaaatgcac	cgccggagcg	ctaattttc	aaacaaaagaa	tctgacgtgc	4080
atttttacag	aacagaaat	caacgcgaga	gctgttgc	atccaaacaa	aatctatact	4140
tcttttttgt	tctacaaaaa	tgcattccga	gagcgttgc	tttctaaacaa	agcatcttgc	4200
attacttttt	ttctctttgc	tgctgttgc	atatgttgc	cttgcataact	ttttgcact	4260
taggtccgtt	aaaggtagaa	gaaggctact	tttgtgttgc	ttttctctc	cataaaaaaa	4320
gcctgactcc	acttcccg	ttaatgttgc	actagcgac	ctgcgggtgc	attttttca	4380
gataaaggcc	tcccgat	tatttatac	cgatgttgc	tgccataact	tttgtgaacag	4440
aaagtatag	cggtgtatgc	tcttcattgc	tcagaaaat	atgaacgg	tcttcatttgc	4500
tgtctctata	tactacgtat	aggaaatgtt	tacattttcg	tattttttc	gatttact	4560
atgaatatgt	cttactacaa	ttttttgtc	taaagatgt	tactagagat	aaacataaaa	4620
aatgtatgg	tcgatgttgc	atgcacgtt	aaaggcgac	agggtggatgg	gtatgttata	4680
tagggatata	gcacagagat	atatacgaaa	gagatactt	tgtagatgt	tttgtggaa	4740
ggttattcgca	atattttgt	agtcgttgc	agtcgttgc	gtttttgttgc	ttttgttgc	4800
gcgtcttcgc	agcgcttttgc	gtttcaaaa	gcgcgttgc	gttcttatac	tttcttagata	4860
ataggaactt	cggaatagga	acttcaaa	gttccgaa	acgagcgc	ccgaaaatgc	4920
aacgcgagct	gcccacatac	agctcaact	tcacgtcg	ccttatatctg	cgtgttgc	4980
gttatatat	atacatgaga	agaacggc	agtgctgttgc	tatgtttaaa	tgctgtactt	5040
atgcgttca	tttatgttgc	atgaaaggta	gtctgttgc	ttatccat	ttatccat	5100

-continued

---

ccatgcgggg tatacgatgc ttcccttcaactacccttt agctgttcta tatgtctcca	5160
ctccctcaatt ggattagtct catcccaa tgctatcatt tcctttgata ttggatcatg	5220
gtagacaacc ctaataataa cttecgataaa tgtatgttat acgaagttat taggtctaga	5280
gatctgtta gttgcctcg tccccggcgg gtcacccggc cagcgacatg gaggcccaga	5340
atacccttc tgacagtctt gacgtcgca gtcaggggc atgatgtgac tgcggcccg	5400
acattttagcc catacatccc catgtataat cattgcatacatactt gatggccgca	5460
cggcgcgaag caaaaattac ggctctcg gtcagacctg cgagcaggaa aacgtcccc	5520
tcacagacgc gttgaattgt cccacacgcg cgccctgtaa gagaataataa aaaggtagg	5580
atttgccact gaggttctt ttcatataat ttccttttaa aatctgtca ggatcacgtt	5640
ctcacatcac atccaaacat aaacaaccat gggtaaggaa aagactcaacg ttgcggggcc	5700
gegattaaat tccaacatgg atgctgattt atatgggtat aaatgggtc gcgataatgt	5760
cgggcaatca ggtgcgacaa tctatcgatt gtatggaa gccgatgcg cagagtgtt	5820
tctgaaacat ggcaaaatggc gctgtcccaat tgatgttata gatgagatgg tcagactaaa	5880
ctggcgacat gaaattatgc ctctggcgc catcaaggat ttatccgta ctctgtatg	5940
tgcatggta tcaccactg cgatccccgg caaaacacgaa ttccaggat tagaaata	6000
tcctgattca ggtgaaataa ttgttgatgc gctggcagttt tcctgccc ggttgcattc	6060
gattccctgtt tgtaattgtc cttaaacag cgatcgcgta ttgcgttcg ctcaaggcga	6120
atcacgaatg aataaacgggt tggtgatgc gatgtattt gatgacgac gtaatggctg	6180
gcctgtggaa caaqtgtggaa aaaaaatgc taagctttt ccattctca cggattcagt	6240
cgtcactcat ggtgatttct cacttgataa ccttattttt gacgaggggaa aattaatagg	6300
ttgtattgtat gtggacggag tcggaaatcg agaccgatatac caggatctt ccattctatg	6360
gaactgcctc ggtgatcttc ctccatccat acagaaacgg cttttcaaa aatatggat	6420
tgataatcgt gatataataa aatttcgtt gtcattgtat tcgcgtatgtt ttttataatc	6480
agtaactgaca ataaaaagat tcttggttt aagaacttgtt catttgtata gtttttttat	6540
attgttagtt ttctatataatcatttgc agcgtgtttt atatttttt tcgcctcgac	6600
atcatctgcc cagatcgaa gttaaatgtcg cagaaggataa tattatgtatgcg	6660
tgaatgtcg tccgtatctgatgcgatccat cgtatcaac gcgcgcattcc agtgcgaaa	6720
acagagcttc gagaaccctt aatataactt cgtataatgtt atgcataatgc aagttatgt	6780
gtgatatcgatccacttggc acggacccca tgcgtatccctt gagaagacaa cctgacccatc	6840
aggaaaagagt tactcaagaa taagaatttt cgttttaaaa ctaaaggatc actttttttttt	6900
ttgtatcacatcattttttt ttaacttat ttaataataa aataatataaa tcataaagaaa	6960
ttcgcttatt tagaagtgaa aacaacgtat ctaccacgcgaa ttggacccctt ttccatcttt	7020
tcgtaaattt ctggcaaggtt agacaaggccg acaaccccttgc ttggagacctt gaccaaacc	7080
ctggcgaagaattgtt aagatgtatgc atctttatgcg atcttcgtt ccgtatccatc	7140
gatactagg ttttgatataa aattttttttt atcttttttgc ttgtgttta ttcttggttc	7200
ttagaaaaaaa caagttggc ttgttgatgc ttgtatgtttt attatttttt aatacgatgc	7260
aatgaagaat agatcgaaat ttgttgcattt atggcattat atcgatcttca tttatactttt	7320
tttttttcaatcgtttagatgatgtt ttccttcattt tcatttttttcaatcgtttagatgatgtt	7380
aagtttacagg gatgtcgaa atatgtgttggg gatgttttttcaatcgtttagatgatgtt	7440
tctaattttt ctttggcgcgaaatccatcgtttagatgatgtt	7500
ctcagaagaaacacgcgatccacttgcgatccatcgtttagatgatgtt	7560
tttattcccttccatcgtttagatgatgtt	7620
gctgccttcatcgtttagatgatgtt	7680
aaaaaaaaaaaaatccatcgtttagatgatgtt	7740
gccgcacacaatccatcgtttagatgatgtt	7800
ggactccgcgatccatcgtttagatgatgtt	7860
tcttcttccatcgtttagatgatgtt	7920
gaccgcctcgatccatcgtttagatgatgtt	7980
tttcttgcataatccatcgtttagatgatgtt	8040
agtaataaaatccatcgtttagatgatgtt	8100
ctttttttatccatcgtttagatgatgtt	8160
caaggatccatcgtttagatgatgtt	8202

---

SEQ ID NO: 83 moltype = DNA length = 8190  
 FEATURE Location/Qualifiers  
 misc\_feature 1..8190  
 note = Plasmid pCEV-G4-Km-TEF-AT-hABAB  
 source 1..8190  
 mol\_type = other DNA  
 organism = synthetic construct

SEQUENCE: 83  
 catgagattt ctccaaat ttaatcgatgc gcatctccg cattagctat 60  
 gcaggatcgatcg tgggtgggaa cggggggggg gctggtacaa ccggcgccgtt cactggggat 120  
 ttatcgatgc gcatctccgatgc gtcacatcgatgc ttatcgatgc atagggtgtt ttggcgaggc 180  
 tcctggcaaa gagcgatggggatgggtatcgatgc tattatgtatgc agtgggtatgc gcaaaatgtatgc 240  
 cgccgattcc gtaaaaggccgatccatcgtttagatgatgtt gatgtgtatgc acacccgtata 300  
 tcctccatcgatgc aactctcgatgc agccggccgcgatccatcgtttagatgatgtt 360  
 gggcgactatcgatgc tggggggggatgggtatcgatgc acacccgtata 420  
 gaccctgtgttccatcgtttagatgatgtt gggggggggatgggtatcgatgc acacccgtata 480  
 aggccggccatcgatgc tggcgatctggatgggtatcgatgc acacccgtata 540  
 gagactctctcgatgc tggcgaggccatcgatgc tggcgaggccatcgatgc acacccgtata 600  
 ccacccgtata 660  
 actgttacatcgatgc gactctcgatgc tggcgaggccatcgatgc acacccgtata 720  
 tggcgatctggatgggtatcgatgc acacccgtata 780  
 acggcgatccatcgatgc tggcgatctggatgggtatcgatgc acacccgtata 840  
 acggggccatcgatgc tggcgatctggatgggtatcgatgc acacccgtata 900  
 gagcgatccatcgatgc tggcgaggccatcgatgc acacccgtata 960  
 aggggtcttcatcgatgc tggcgatctggatgggtatcgatgc acacccgtata 1020  
 atgggttccatcgatgc tggcgatctggatgggtatcgatgc acacccgtata 1080

-continued

---

cagcacaatg	cgggacgact	ccacaaaagg	gcgggttacc	atttccagag	acaacgcca	1140	
gaatactctg	taccttcaga	tgaccagtct	gaaacccgag	gacactgctc	tgtactattg	1200	
tgcagaggc	cggggtatct	ctgcttccgc	tatcagaggc	gcagtaaggg	gccctggaa	1260	
acaggtaacc	gtttcatccg	ggggggcg	ttcaggcggt	gggggatctg	gcgggggtgg	1320	
atccccagg	cagctgtcg	aatccccggg	cgactgtcg	cacacagggg	gtccctcgag	1380	
gctctccgt	gcatcttccg	gaagcatcg	cggttccgag	accgtgact	ggttctcgcca	1440	
gggtccccc	aagtctctgc	agtgggtcg	tccatgact	aagactaaca	acgagatcta	1500	
ctctgactca	gtgaaaaggcc	gcttcatcg	ttcttagagat	aacgctaaaa	acacagtgt	1560	
tctgcagatg	aatagtctca	aacctgaaga	cacaggcg	tatctctgt	agggttctgt	1620	
gctgaggggc	cagggcattcc	aggtaacat	ctcaggtta	ggtagccgg	ctagctaaga	1680	
tccgcttaa	ccggaaaagg	aggagttga	caacctgaag	tctaggccc	tatttat	1740	
tttatagtt	tgtagtattt	aagaacgtt	tttatattt	aaattttct	ttttttctg	1800	
taacagacgc	tgtagccat	taacattta	ctgaaaacct	tgcttgagaa	gggtttggga	1860	
cgtcgaaga	tccagtcg	taatgaatc	ggccaaacg	cgggggaggg	cggttgcgt	1920	
attgggcgt	tttccgcctt	ctcgctact	gactcg	gctcggtc	tcgggtcg	1980	
cgagcgtat	cagtcactc	aaaggcggt	atacgggt	ccacagaatc	aggggataac	2040	
gcaggaaaga	acatgtgac	aaaaggccca	ggaaaccgtaa	aaaggcccg	2100		
ttgctggcgt	tttccatag	getccccc	cctgacgag	atcacaaaa	tgcacgctca	2160	
agtcaaggt	ggcggaaaccc	gacagacta	taaagatacc	aggcg	tttccctcgaa	2220	
tcctctgtc	gctctctgt	tccgacc	ccgcttacc	gatactgtc	cgcccttctc	2280	
ccttcggaa	gggtggcg	ttctccat	tcacgtgt	ggtatctcg	ttcgggtgt	2340	
gtcggtcg	ccaaagttgg	ctgtgtcg	gac	aaacccccc	ttcageccga	2400	
ttagccgt	atatcgct	tgatgtcc	ccggtaa	gac	acttatac	gccactgg	2460
gcagccactg	gtaacaggat	tagcagage	aggtatgt	g	gggtgtctac	agagttctt	2520
aagtggtgc	taaactacgg	ctacacta	aggacagt	tttgtatctg	cgctctgt	2580	
aaggccgtt	ccttcggaa	aagagtgt	agetcctgt	ccggcaaa	aaaccaccgt	2640	
ggttagccgt	gtttttgt	ttcgaag	ccatggc	cgaaaaaaa	aggatctca	2700	
gaagatcc	tgatctt	tacgggt	gacgct	ggaacgaaa	ctcacgtt	2760	
gggattttgg	tcatgagatt	atcaaaaagg	atcttca	ccctt	ttt	aaattaaaa	2820
tgaatgtt	atcaactcta	aagtatata	gagtaaaactt	ggtctgac	ttaccaat	2880	
ttaatcg	aggacccat	ctcagcgt	tgttattt	tttcat	ccat	agttgc	2940
ctcccccgt	tgtagataac	tacat	ccat	ggggcttac	catctggccc	3000	
atgataccgc	gagacccac	ctcacc	ccat	ggattt	ca	ccagccag	3060
gaaaggccgc	agcgcgaa	tggttcc	acttatac	cc	cttcat	cc	3120
tgttgcgg	aagtaggt	tgatgtc	ccat	tttgc	cc	cggttgc	3180
attgctac	gatcgctgt	gtc	ccgtt	ggta	tttgc	cc	3240
tcccaacgt	caaggcg	tacat	gtt	ccat	gtt	cc	3300
ttcggtc	cgatcg	tgatgt	ccat	gtt	ccat	act	3360
gcagcactg	aaatttct	tact	gtat	ccat	gtt	tttgc	3420
gagtagact	ccaa	gtt	ccat	gtt	ccat	cc	3480
gcgtcaatac	gggataatac	cg	ccgc	ccat	gtt	ccat	3540
aaacgtt	cgggggc	actt	ccgc	ccat	gtt	ccat	3600
taacccact	gtgcac	ctgat	ccat	gtt	ccat	cc	3660
tgaacaaa	caggaggc	aaat	ccgc	ccat	gtt	ccat	3720
tgaatactca	tacttct	tttca	at	ccat	gtt	ccat	3780
atgagccgt	acatattt	atgtt	ccat	gtt	ccat	cc	3840
tttcccgaa	aagtggcacc	tgaacat	ccat	gtt	ccat	cc	3900
caacgcgaga	gogctattt	ttcaacaaa	aat	ctgt	ccat	cc	3960
atgcaacgc	aaagcgctat	tttacca	ac	gaaat	ctgt	ccat	4020
aaatgcaac	cgagacgcgt	aattttt	ccat	gtt	ccat	cc	4080
cagaaatgc	acgcgag	gtt	ccat	gtt	ccat	cc	4140
tacaaaaatg	catcccgaga	ccgt	ccat	gtt	ccat	cc	4200
ctctttgt	cgctctata	tgca	ccat	gtt	ccat	cc	4260
ggttagaaga	aggctactt	gggt	ccat	gtt	ccat	cc	4320
ttcccggtt	tactgttac	tgca	ccat	gtt	ccat	cc	4380
cccgattata	ttcttac	atgtt	ccat	gtt	ccat	cc	4440
tttgcattt	ttcatt	ccat	gtt	ccat	cc	4500	
ctacgtat	gaaatgtt	cattt	ccat	gtt	ccat	cc	4560
tactacaatt	ttttgttca	aaag	ccat	gtt	ccat	cc	4620
gagtttagat	gcaagtt	ccat	gtt	ccat	cc	4680	
acagagat	atagcaa	gata	ccat	gtt	ccat	cc	4740
atttttagat	ctcgat	ccat	gtt	ccat	cc	4800	
cgctttgtt	ttcatt	ccat	gtt	ccat	cc	4860	
gaataggaa	ttccaa	ccat	gtt	ccat	cc	4920	
gcacat	actgtt	ccat	gtt	ccat	cc	4980	
acatgagaag	aacggat	tg	ccat	gtt	ccat	cc	5040
atgttagat	gaaaggat	tgat	ccat	gtt	ccat	cc	5100
tcgtat	cattt	ccat	gtt	ccat	cc	5160	
attagtc	tcctt	ccat	gtt	ccat	cc	5220	
taatataact	tgtat	ccat	gtt	ccat	cc	5280	
ttgcctgtc	cccgccgg	caccc	ccat	gtt	ccat	cc	5340
acagtctt	cg	ccat	gtt	ccat	cc	5400	
tacat	ccat	gtt	ccat	cc	5460		
aaaatttac	ccctc	ccat	gtt	ccat	cc	5520	
tgaattgtt	ccac	ccat	gtt	ccat	cc	5580	
ggttctt	tcatata	cc	ccat	gtt	ccat	cc	5640
cggacataa	acaaccat	gta	ccat	gtt	ccat	cc	5700
caacat	ccat	gtt	ccat	cc	5760		
tgcgacaat	tatc	gtt	ccat	gtt	ccat	cc	5820

-continued

caaaggtagc	gttgc当地atg	atgttacaga	tgagatggtc	agactaaact	ggctgacgga	5880
atttatgcct	cttc当地accat	tcaagcattt	tatccgtact	cctgtatgt	catggttact	5940
caccactgcg	atccccggca	aaacagcattt	ccaggtatta	gaagaatatac	ctgattcagg	6000
tgaaaatattt	gttgc当地g	tggc当地gtt	cctgc当地gg	ttgc当地tca	ttc当地gtt	6060
taatttgcct	ttaacagcg	atcgcttgc	tcgtc当地cg	caggcg	cacgaaatgaa	6120
taacggggtt	gttgc当地g	gtgatgttgc	tgacgagcgt	aatggc当地	ctgtt当地	6180
agtctggaaa	gaaatgcata	agctttgc	atttgc当地	gattc当地	tcactc当地	6240
tgatttgc	cttgat	ttat	ttttgaa	ttat	atagg	6300
tgga	cgac	acc	ggat	cc	at	6360
tgat	ttt	catt	aaa	gg	ttt	6420
tatgat	tttgc	atgt	cgat	at	tttgc	6480
aaa	aaat	tttgc	gatgtt	tat	tttgc	6540
ctat	tttgc	aatatgtt	cg	tttgc	gcac	6600
atgc	gaat	ttatgtt	tttgc	at	tttgc	6660
gtat	gc当地	ttatgtt	atcg	at	gtgt	6720
gtat	act	acta	cc	tg	ggaa	6780
gaac	ccct	tata	atcg	tat	cgat	6840
ccact	ggag	ccat	ctat	ga	acac	6900
ctca	aga	attt	tttgc	taa	gtac	6960
at	tttgc	tttgc	tttgc	tttgc	tttgc	7020
gaag	gttca	caac	gtat	acc	tttt	7080
ggca	aggtag	aaac	ggcc	tttgc	tttgc	7140
tgta	tttt	at	tttgc	tttgc	tttgc	7200
ttgt	at	tttgc	tttgc	tttgc	tttgc	7260
agtg	gatgtt	tttgc	tttgc	tttgc	tttgc	7320
attc	gaat	tgt	tttgc	tttgc	tttgc	7380
tct	atgt	tttgc	tttgc	tttgc	tttgc	7440
at	tc	tttgc	tttgc	tttgc	tttgc	7500
cg	cagg	tttgc	tttgc	tttgc	tttgc	7560
aaaa	aaat	tttgc	tttgc	tttgc	tttgc	7620
tgcat	tttgc	tttgc	tttgc	tttgc	tttgc	7680
ttoct	atccc	gcaca	at	tttgc	tttgc	7740
at	gc	aaat	tttgc	tttgc	tttgc	7800
tcgc	cttca	aat	tttgc	tttgc	tttgc	7860
tagg	gttgc	tttgc	tttgc	tttgc	tttgc	7920
tctt	tttgc	tttgc	tttgc	tttgc	tttgc	7980
ttt	tttgc	tttgc	tttgc	tttgc	tttgc	8040
gtctt	caatt	tctca	tttgc	tttgc	tttgc	8100
cttgc	catt	aga	tttgc	tttgc	tttgc	8160
at	acgact	at	tttgc	tttgc	tttgc	8190

SEQ ID NO: 84            moltype = DNA length = 8190  
 FEATURE                Location/Qualifiers  
 misc\_feature          1..8190  
 note = Plasmid pCEV-G4-Km-TEF-AT-yABAB  
 source                1..8190  
 mol\_type = other DNA  
 organism = synthetic construct

SEQUENCE: 84  
 ggaggc当地 ctatgc当地 ac ctggaaacg aacctgaccc acaggaaaga gttactcaag 60  
 aataaggaaat ttctgtttaa aacctaaaggat tcactttaaa atttgatatac actttttttt 120  
 ttataactt atttaataat aaaaatcata aatcataaga aattcgotta ttttagaagg 180  
 tcaacaacgt atcttaccaac gatttgaccc tttccatct ttctgtaat ttctggcaag 240  
 gttagacaacg cgacaacccctt gatttgagac ttgacccatcc ttctggcgaa gaatttgtaa 300  
 ttaagagctc agattttac gcgc当地atcc ttgtatatac tcgactatag tttttgcatt 360  
 aaaatattttt aacttttttgc tttttgtt tttttgttgc tttttgttgc tttttgttgc 420  
 gcttgc当地 ttttgatgtt ttattttatcc acaatagctg ctaatgaaaga atagattcga 480  
 acatttgc当地 gatattggcat atatcgctc tattttatcc ttttttttgc tttttgttgc 540  
 atatttgttgc ttttgc当地 ttttgc当地 ttttgc当地 ttttgc当地 ttttgc当地 600  
 aaagatgttgc gggattttgc cttggaaatcc gacttttgc catcttgc当地 ttctttgttgc 660  
 cctctgaaaat ttatccagca gaagcaaatg taaaatgtatcc acctcagaag aacacgagg 720  
 gggccgaaat ttgttgc当地 agaaggatgttgc ggttgc当地aaaat gtttttttgc ctggaaaat 780  
 aattttgc当地 ttgttgc当地 ttttgc当地 ttttgc当地 ttttgc当地 ttttgc当地 840  
 tcttgatgttgc gttggatgttgc ttccatccatgt ttgttgc当地 ttttgc当地 ttttgc当地 900  
 ccccgcaacaat ttctctgc当地 acccccttgc当地 tccacggggcc cggccgcaaca caccatacg 960  
 tcaaaatgtt ttacttgc当地 ttttgc当地 ccagatccatgc tggactccg cgc当地ccgc当地 1020  
 taccacttca aacacccatgc gcacagcata ctaaatttcc cctcttcc tctcttgc当地 1080  
 gtcgttattt accccgtacta aacggatgttgc aaaaaaaaaa gagacccgctt cgttccatgtt 1140  
 tcttc当地gc当地 aaaaaggat ttttgc当地 ttttgc当地 ttttgc当地 ttttgc当地 1200  
 ttttgc当地 ttttgc当地 ttttgc当地 ttttgc当地 ttttgc当地 ttttgc当地 1260  
 aatttctca aatgttgc当地 ttttgc当地 ttttgc当地 ttttgc当地 ttttgc当地 1320  
 cattttgc当地 ttttgc当地 ttttgc当地 ttttgc当地 ttttgc当地 ttttgc当地 1380  
 ctcactatag gggccggcc tccatgc当地 atttcccttca attttgc当地 ttttgc当地 1440  
 cgc当地ccgc当地 tccgatccatgc ttttgc当地 ttttgc当地 ttttgc当地 ttttgc当地 1500  
 tcaaccaggat ggttagttgc gattatccatgc ttttgc当地 ttttgc当地 ttttgc当地 1560  
 ttcaataggat ttttgc当地 ttttgc当地 ttttgc当地 ttttgc当地 ttttgc当地 1620  
 cagtctgttgc gactcaacta aatatgttgc当地 ctccgttaag ggttagatca ctacttcaag 1680  
 agataacgc当地 aaaaatacag tctacttgc当地 aatgttgc当地 ttttgc当地 ttttgc当地 1740  
 agtcttattac ttttgc当地 ttttgc当地 ttttgc当地 ttttgc当地 ttttgc当地 1800

-continued

---

cggtaaaagat	gactggggta	aaggtaacttt	agttactgtc	tcatccggtg	gtgggtggc	1860
cgggtgggtt	ggtagtgggt	gtgggtggtc	tcaagttcaa	ttagtagaaat	ccgggtgggt	1920
tttagtccaa	ctgggtggta	gtttaagatt	atcctgcgaa	gcaagtgggt	ttacattaga	1980
ttattacggt	atcggttgtt	ttagacaacc	acctggtaaa	gaaagagaag	ctgtcttta	2040
tatcccggt	agtgcagaag	ctatattgtt	cgcagatttc	gtttagggta	gattcacaaat	2100
ttcaagagac	aatgcgaaga	acgtgtttt	tttgcataatg	aacttttta	agagagaaga	2160
caccgcagg	tattactgt	ccagaagaag	attttctgt	tcttcagtc	acagatgggt	2220
agcagacgat	tatgtatgtt	ggggtagagg	tacacaagtc	gcccgtaaat	ctgggtgggt	2280
ttccgggtgt	ggtagtgggt	gtgggtctgg	tgggtggttca	caatttgcatt	tagtagaaac	2340
tgggtgggtt	ttgggttcaac	cagggtggtc	cttggatattt	agttgtgtc	cattctgggtt	2400
tactttctt	gattacgtt	tgatcatggt	cajacaacgt	ccaggtaaaag	gtccctgaatg	2460
gatcgctaca	attaatacc	acggttccac	aatgagagat	gacagtacca	agggttagatt	2520
caactattca	agagacaaac	ctaagaacac	atgttactta	caaatttgcatt	cttttgcataac	2580
agaagacac	gatttatatt	actgtggccag	aggtagatgt	atatccgcac	gtgtatccat	2640
aggtgcagta	agagggtctgt	gtactcaatg	taatgttctt	tcagggtggc	gccccgtgtt	2700
cggcgggggt	tctgggggt	gtgggtcaca	agtccaaattt	gttagatctg	gtgggtgggtt	2760
agttcaaaat	gggtgttcat	tgagattatc	ctgggttcc	agtgggttca	ttgcagggtt	2820
cggaaactgtt	acatgttca	gacaatgtt	aggttaatctt	tttgcataatg	tcgcctcaat	2880
gaccaagact	aacaacgaaa	tctattctga	ttcagttaaat	ggtagatcca	ttatttcaag	2940
agataatgt	aaaaaacacc	tttatttgc	aatgaactca	tttgcatttgc	aaagataactgg	3000
tgtttacttc	tgcaagggtt	ctgatattaa	aggtaatgtt	attcaatgtt	cagtttcttc	3060
ataagggtacc	cggggctatgt	aaatccgtt	ctaaccgtt	aaaggaaat	tagacaacat	3120
gaagtctagg	tccctattta	tttttttata	gtttagttag	tatattttata	gttattttata	3180
tttcaattt	ttctttttt	tctgtacaga	cgcggttac	catgttacat	tataactgaaa	3240
acottgttg	agaagggtttt	ggggacgtcg	aaatccgcac	tgcattttat	aatccgcacaa	3300
cggcgggggg	gaggggggtt	ggcttattttt	cgcttccgtt	cttcccttgc	cactgtactcg	3360
ctgcgtctgg	tcgttccgtt	ggggcggatcg	gtatcgttcc	acttacggc	ggtaatccgg	3420
ttatccacag	aatcagggg	taacgcagga	aaagacatgt	gagcaaaagg	ccagcaaaag	3480
gccaggaacc	gtaaaaaggg	cgcggttgc	gggtttttcc	ataggctcc	ccccctgtac	3540
gacgatcaca	aaaatccgac	ctcaagtcac	gggtggccag	accggacagg	actataaaaga	3600
taccaggcgt	ttccccctgg	aaatccctc	gtgcgtcttc	ctgttccgac	cctgcccgtt	3660
acccggatacc	tgtccgctt	tctcccttc	ggaaagcgtgg	cgcttctca	tagtcaacgc	3720
tgttaggtatc	tcaatgttgc	gttagtgcgtt	cgcttcaatgc	ttgggtgtt	gcacgaaatcc	3780
cccggttgc	cgaccgttgc	cgccgttgc	ggtaactatc	gttcttgc	caacccggta	3840
agacacgact	tatgcacat	ggcgcgcac	acttgcataa	ggatttgcac	agcggaggat	3900
gtaggcgggt	ctacagatgt	tttgcataact	acggcttacac	tagaaggaca	3960	
gtatttggta	tctgcgtct	gcttgcgtt	tttgcataact	ggaaaatgtt	tttgcgtct	4020
tgtatccggca	aacaaaccac	cgcttgcgtt	gggtgttttt	tttgcataat	ggcagcagat	4080
acgcgcagaa	aaaaggatc	tcaagaatgt	cctttgtatc	tttgcatacg	gtctgcacgt	4140
cagtggaaac	aaaacttcac	ttaagggtt	tttgcataat	tttgcatacg	gtctgcacgt	4200
acctagatcc	ttttaaat	aaaatgtt	tttgcataat	tttgcataat	tttgcataat	4260
acttggcttc	acagtatcc	atgttcaat	atgttcaat	atgttcaat	atgttcaat	4320
tttcgttcat	ccatgttgc	ctgacttccc	gttgcgttgc	taatgttcaat	acggggggc	4380
ttaccatctg	gccccgttgc	tgcataat	ccgcgttgc	cacgcgttcc	ggcttccat	4440
ttatcagcaa	taaaccagcc	agccggaaag	gccgagcgc	gaagttgtt	tgcataat	4500
tccgccttca	tccatgtat	taatgttgc	cgggaaatgt	gatgttgc	tgcataat	4560
aatagtttgc	gcaacgttgc	tccatgtat	cgaggcatcg	tttgcgttgc	ctcgctgtt	4620
ggtatgggtt	cattcgttgc	cggtttccaa	cgatcaatgc	tttgcgttgc	atccccatgt	4680
tttgttgc	aaagggtttag	ctcccttcgtt	cctccgtatc	tttgcataat	taatgttgc	4740
gcagtgttat	cacttgcgtt	tatgcgttgc	ctgcataat	cttcttactgt	catgcacat	4800
gtaaatgtt	tttgcgttgc	tttgcgttgc	tttgcataat	tttgcataat	tttgcataat	4860
ccggcggccg	gttgcgttgc	ccccgggtca	atacgggata	tttgcataat	tttgcataat	4920
actttttaa	tttgcgttgc	tttgcataat	tttgcataat	tttgcataat	tttgcataat	4980
ccgctgttgc	gtatgttgc	gtatgttgc	acttgcgttgc	ccaaatgttgc	tttgcataat	5040
tttacttgc	ccatgttgc	tttgcataat	tttgcataat	tttgcataat	tttgcataat	5100
ggaataatgg	cgacacggaa	atgttgcata	cttcatatct	tttgcataat	tttgcataat	5160
agcatttatc	agggttattt	tctcatgttgc	ggatacatat	tttgcataat	tttgcataat	5220
aaacaaatag	gggttccggc	cacatttccc	cgaaaatgtc	cacctgttgc	aagcatctgt	5280
gcttcattat	tttgcataat	tttgcataat	tttgcataat	tttgcataat	tttgcataat	5340
gagctgttgc	ttagatgttgc	tttgcataat	tttgcataat	tttgcataat	tttgcataat	5400
ctgtgttca	ttttgttata	acaaaaatgc	aacgcgttgc	tttgcataat	tttgcataat	5460
aatctgttgc	gttgcgttgc	tttgcataat	tttgcataat	tttgcataat	tttgcataat	5520
agaaatctta	tttgcgttgc	tttgcataat	tttgcataat	tttgcataat	tttgcataat	5580
aaagcatctt	agatattttt	tttgcgttgc	tttgcataat	tttgcataat	tttgcataat	5640
ctttttgcac	tttgcgttgc	tttgcataat	tttgcataat	tttgcataat	tttgcataat	5700
tccatataaaa	aaaggcttgc	tttgcataat	tttgcataat	tttgcataat	tttgcataat	5760
gcattttttc	aaagatggaa	cattttccat	tttgcataat	tttgcataat	tttgcataat	5820
ctttgttgcac	ggggatggaa	tttgcataat	tttgcataat	tttgcataat	tttgcataat	5880
tttcttctat	tttgcgttgc	tttgcataat	tttgcataat	tttgcataat	tttgcataat	5940
tcgatttact	ctatgttgc	tttgcataat	tttgcataat	tttgcataat	tttgcataat	6000
ataaaacataa	aaaatgttgc	tttgcataat	tttgcataat	tttgcataat	tttgcataat	6060
gggttaggtt	tatagggtt	tttgcataat	tttgcataat	tttgcataat	tttgcataat	6120
gtttgttggaa	ggggatggaa	tttgcataat	tttgcataat	tttgcataat	tttgcataat	6180
tttttggaaa	gttgcgttgc	tttgcataat	tttgcataat	tttgcataat	tttgcataat	6240
actttcttgc	aaataggaaac	tttgcataat	tttgcataat	tttgcataat	tttgcataat	6300
ttccggaaat	gtcaacgcgc	tttgcataat	tttgcataat	tttgcataat	tttgcataat	6360
tgcgtgttgc	ctgtatatat	tttgcataat	tttgcataat	tttgcataat	tttgcataat	6420
aatgcgttgc	tatattgttgc	tttgcataat	tttgcataat	tttgcataat	tttgcataat	6480
tattatccca	ttccatgttgc	tttgcataat	tttgcataat	tttgcataat	tttgcataat	6540

-continued

---

tatcatgtgc	cactcctcaa	ttggattagt	ctcataccttc	aatgctatca	tttccttga	6600
tattggatca	tggttagacaa	ccctaataat	aacttcgtat	aatgtatgt	atacgaaagt	6660
attaggctca	gagatctgtt	tagcttgct	cgtccccccc	gggtcacccg	gccagcgaca	6720
tggaggecca	gaataaccctc	cttgacagtc	ttgacgtgcg	cagtcaggg	gcatgatgt	6780
actgtgcgcc	gtacattnag	cccatacato	cccatgtata	atcatttgc	tccatacatt	6840
ttgatggccg	cacggcgcga	agcaaaaatt	acggctcte	gtgcagacc	tgcgagcagg	6900
gaaacgtcc	cctcacagac	gctgttaatt	gtccccacgc	cgcgcctc	tagagaataa	6960
taaaaggtta	ggatttgcca	ctgagggtct	tcttcataat	acttccttt	aaaatcttc	7020
taggatcacag	ttctcatac	acatccgaa	ataaacaaacc	atgggtaaagg	aaaagactca	7080
cgtttcgagg	ccgcgatcaa	attcacaa	ggatgtgtat	ttatatgggt	ataaatgggc	7140
tcgcgataat	gtcgccgcaat	cagggtgcac	aatctatcg	ttgtatggga	agcccgatgc	7200
gccagagttg	ttctgtaaac	atggcaaaagg	tagcgttgc	aatgtatgt	cagatgagat	7260
ggtcagacta	aactggctc	cggaaattat	gctcttcgc	accatcaagc	attttatccg	7320
tactctgtat	gtgcgttgt	tactcataac	tgccatcccc	ggcaaaacag	cattccagg	7380
attagaagaa	tatctgtatt	cagggtaaaa	tattttgtat	gcccgtgc	tgttctgcg	7440
ccgggttgc	tcgattctg	tttgttaatt	tccttttaac	agcgatcg	tatccgtct	7500
cgctcaggcg	caatcacgaa	tgaataacgg	tttgggttgc	gcccgtgtt	ttgtatggca	7560
gctaatggc	ttgcgttgtt	aacaatgtc	gaaagaaatg	cataagctt	tgccatttctc	7620
accggattca	gtcgtcactc	atggtgattt	ctcacttgc	aaccttattt	ttgacgagg	7680
gaaaattaata	ggttgtattt	atgttggac	agtcgaaatc	gcagaccgat	accaggatct	7740
tgccatctca	tggaactg	tcgggtggat	tttccttc	ttacagaaac	ggcttttca	7800
aaaatatgtt	attgataatc	ctgatgtat	taaatttgc	tttgcgttgc	tgctcgatga	7860
gtttttctaa	tctgactgt	caataaaaaa	attttgttgc	tcaagaaattt	gtcattttgc	7920
tagttttttt	atattgtat	tggttctattt	taatcaaattt	ttagcgtat	ttatattttt	7980
tttcgcctcg	acatcatctg	cccagatgc	aagtttgc	cgccagaaatg	aatatcatgc	8040
gtcaatctgt	tgtgtatgt	ggtcgtatca	ctgctgtcg	ttcgatatac	acgcgcct	8100
ccagtgctg	aaacgagctc	tcgagaaacc	ttatataac	ttcgatataat	gtatgtata	8160
cgaagttatt	aggtgatatac	agatccacta				8190

---

SEQ ID NO: 85                    moltype = DNA length = 8202  
 FEATURE                         Location/Qualifiers  
 misc\_feature                 1..8202  
 note = Plasmid pCEV-G4-Km-TEF-X40-AT-yABAB  
 source                         1..8202  
 mol\_type = other DNA  
 organism = synthetic construct

SEQUENCE: 85  
 ggagcgcacct catgtatac ctgagaaago aacctgaccc acaggaaaga gttactcaag 60  
 aataagaatt ttcgtttaa aacctaaagag tcactttaaa atttgcatac actttttttt 120  
 ttataactt attaataat aaaaatcata aatcataaga aattcgotta ttttagaagt 180  
 tcaacaacgt atctaccaac gatttgaccc tttccatctt ttctgtataat ttctggcaag 240  
 gttagcaacgc cgacaaacctt gatttgagac ttgaccaaaac ctctggcgaa gaattttgtaa 300  
 ttaagagctc agatcttac gtcgtcactc ttgtatccca tgcataactg tttttgtt 360  
 aaataaaaaa aaactttttt tttttgtttt tatttttttgc tctttagaaaa gacaatgt 420  
 gcttgggttgc ttttgcataat ttattttttt acaatagctg caaatagaaga atagattcga 480  
 acattgtgaa gtatttgcat atatctgttca tattttatact tttttttttt cagtttctgt 540  
 atattttttgc ttttcttctt ttcatttttca tcaatgttca ataagttaca ggggatctcg 600  
 aaagatgtgtt gggatttttcc ttttgcataat ttgttgcac gactttttgc catcaattt ttccatttttttgc 660  
 cctctgaaaa ttatccagca gaagcaatgtg taaaatgtca acctcagaag aacacgcagg 720  
 ggcccaaaat ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 780  
 aatttttgcgtt ttttttttttgc gagaatttttgc tcaatttttttgc ttttttttttgc ttttttttttgc 840  
 tcttgatgttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 900  
 ccccgacaaa atttctgtat acccccttatttcc ttttttttttgc ttttttttttgc ttttttttttgc 960  
 tcaaaatgtt ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 1020  
 taccacttca aaacacccaa gcacacgtata ctaaattttcc ttttttttttgc ttttttttttgc 1080  
 gtctgttattt accccgtacta aagggttgc aaaaatggatccgc ttttttttttgc ttttttttttgc 1140  
 ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 1200  
 ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 1260  
 aattttctcaaa ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 1320  
 catttagaaag aaagcatgtc aatctatctt aatgtttttat tacaaggatc catgaggattt 1380  
 ctttcaattt ttactgtctgt ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 1440  
 ttgtgttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 1500  
 gcatcagggtt ttatctgtat ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 1560  
 gaaaagagaag gtgttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 1620  
 gttttaggtt gttttaggtt gttttaggtt gttttaggtt gttttaggtt gttttaggtt 1680  
 aactcattaa agccagatgtc cacagcgttcc ttttttttttgc ttttttttttgc ttttttttttgc 1740  
 tgccgtgtat ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 1800  
 actgtctcat cccgggttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 1860  
 gtttcaatttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 1920  
 tgccgtgtat ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 1980  
 gttttaggtt gttttaggtt gttttaggtt gttttaggtt gttttaggtt gttttaggtt 2040  
 gttttaggtt gttttaggtt gttttaggtt gttttaggtt gttttaggtt gttttaggtt 2100  
 caaatgtactt ctttgcataat ttgttgcataat ttgttgcataat ttgttgcataat ttgttgcataat 2160  
 ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 2220  
 caagtcgcgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 2280  
 gttttaggtt gttttaggtt gttttaggtt gttttaggtt gttttaggtt gttttaggtt 2340  
 agttttaggtt gttttaggtt gttttaggtt gttttaggtt gttttaggtt gttttaggtt 2400  
 caagctccat ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 2460  
 agatgtatgtc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 2520

-continued

---

tacttacaaa	tgacctcttt	gaaaccagaa	gacaccgcat	tatattactg	tgccagagg	2580	
agagtcatat	ccggcagtgc	tatcagagg	gcagtaagag	gtcctggtag	tcaagttaca	2640	
gtctcttcag	gtggcgccgg	tagtggcg	ggcggttctg	gcccgtgg	ttcacaagtc	2700	
caatttgttag	aatctgg	tggtttagtt	caaactgg	gttcattgag	attatctgc	2760	
gcttccatgc	gttccatgc	agggttcgaa	actgttacat	ggtaaagaca	agctccaggt	2820	
aaatctttgc	aatgggtcgc	ctcaatgacc	aagactaaca	acgaaatcta	ttctgattca	2880	
gttaagggtt	gattcattat	ttcaagagat	aatgctaaaa	acaccgttta	tttgcaatg	2940	
aactcattga	agccagaaga	tactgggtt	tacttctgca	agggtcctg	attaagaggt	3000	
caaggattat	aagtaaact	ttcttcgtgc	gacatggaa	agaagg	ttccaaagaa	3060	
gacctcgagt	aagcttggta	cecgccgt	ctaagatccg	ctctaa	aaaggaaagga	3120	
gttagacaa	cttgcgtct	ggcccttatt	tat	tttttttt	tagttatgt	3180	
acgttattta	tat	ttttttttt	tttctgtaca	gacgcgtgt	cgcgttac	3240	
attatactga	aaacccgt	tgagaagg	tttggacg	cgaagatcca	gtcgttac	3300	
tgaatccgg	aacgcgg	gagagg	tttgcgtt	ggcgcttctc	cggttcc	3360	
ctcaactgact	cgctcg	ggcgttccg	ctgcgg	cggtatc	tcactcaag	3420	
gccccgtat	ggttatcc	agaatcagg	gataac	gaaagaa	gtgagcaaa	3480	
ggcccaaa	aggccagg	ccgtt	ggccgttgc	ttgg	ccatagg	3540	
cgcgttcc	acgacatca	caaaaatc	cgctca	agagg	ttgg	aaaccgcaca	3600
ggactataa	gataccagg	gttcc	gaaag	ctcc	tcgtcg	3660	
acccgttcc	ttacccgata	cctgtcc	tttcc	cggtt	tttct	3720	
catagtcac	gtgttgg	tctcgtt	gtgttgg	ttcg	cttccaa	3780	
gtgcacgac	ccccgttca	gccc	gtgc	ttttt	ccatagg	3840	
tccaa	ccgg	tttcc	tttgc	tttgc	tttgc	3900	
agagcgagg	atgttgg	tgctc	atgttgg	tttgc	tttgc	3960	
actagaagg	cagtat	tat	tttgc	tttgc	tttgc	4020	
gttggat	cttgat	ccaa	accc	gggg	ttttt	4080	
aagcagcaga	ttac	aaaa	tttgc	tttgc	tttgc	4140	
gggtctgac	ctc	aaaa	tttgc	tttgc	tttgc	4200	
aaaaggatct	tcac	tttt	tttgc	tttgc	tttgc	4260	
atataatgt	aaacttgg	tgac	tttgc	tttgc	tttgc	4320	
gcccgttgc	tatttgc	atccat	tttgc	tttgc	tttgc	4380	
atacggagg	gttatt	tttgc	tttgc	tttgc	tttgc	4440	
ccggcttcc	at	tttgc	tttgc	tttgc	tttgc	4500	
cctgcgtt	at	tttgc	tttgc	tttgc	tttgc	4560	
agttcgcc	ttaa	tttgc	tttgc	tttgc	tttgc	4620	
cgttcgtt	tttgc	tttgc	tttgc	tttgc	tttgc	4680	
tgatcccca	tttgc	tttgc	tttgc	tttgc	tttgc	4740	
agtaatgtt	ccgc	tttgc	tttgc	tttgc	tttgc	4800	
gtcatgc	ccgt	tttgc	tttgc	tttgc	tttgc	4860	
gaatagtgt	tgcc	tttgc	tttgc	tttgc	tttgc	4920	
ccacatgca	gaactt	tttgc	tttgc	tttgc	tttgc	4980	
tcaggatct	tacc	tttgc	tttgc	tttgc	tttgc	5040	
tcttcagcat	ttttt	tttgc	tttgc	tttgc	tttgc	5100	
gcccgttttt	agg	tttgc	tttgc	tttgc	tttgc	5160	
caatattatt	gaagcattt	tcagg	tttgc	tttgc	tttgc	5220	
atttgc	aaa	tttgc	tttgc	tttgc	tttgc	5280	
cgaagcatct	gtgtt	tttgc	tttgc	tttgc	tttgc	5340	
aacaaagat	tttgc	tttgc	tttgc	tttgc	tttgc	5400	
ccaaacgtt	atctgt	tttgc	tttgc	tttgc	tttgc	5460	
tttcaacaa	aaatctgt	ctgc	tttgc	tttgc	tttgc	5520	
tttcaacaa	aaatctgt	tttgc	tttgc	tttgc	tttgc	5580	
tat	tttgc	tttgc	tttgc	tttgc	tttgc	5640	
gtctcttgat	aa	tttgc	tttgc	tttgc	tttgc	5700	
cttattttt	cttccat	aaa	tttgc	tttgc	tttgc	5760	
gttgcgtt	tttgc	tttgc	tttgc	tttgc	tttgc	5820	
tttgcgtt	tttgc	tttgc	tttgc	tttgc	tttgc	5880	
aattatgt	tttgc	tttgc	tttgc	tttgc	tttgc	5940	
tttgcgtt	tttgc	tttgc	tttgc	tttgc	tttgc	6000	
gttaatgtat	aaat	tttgc	tttgc	tttgc	tttgc	6060	
cggaaagg	tttgc	tttgc	tttgc	tttgc	tttgc	6120	
ctttgcgtt	tttgc	tttgc	tttgc	tttgc	tttgc	6180	
gttgcgtt	tttgc	tttgc	tttgc	tttgc	tttgc	6240	
tgaatgtt	tttgc	tttgc	tttgc	tttgc	tttgc	6300	
gaaaacgt	tttgc	tttgc	tttgc	tttgc	tttgc	6360	
cgcac	tttgc	tttgc	tttgc	tttgc	tttgc	6420	
tttgcgtt	tttgc	tttgc	tttgc	tttgc	tttgc	6480	
tttgcgtt	tttgc	tttgc	tttgc	tttgc	tttgc	6540	
ctttagctt	tttgc	tttgc	tttgc	tttgc	tttgc	6600	
catttc	tttgc	tttgc	tttgc	tttgc	tttgc	6660	
ctatacgtt	tttgc	tttgc	tttgc	tttgc	tttgc	6720	
cgccca	tttgc	tttgc	tttgc	tttgc	tttgc	6780	
ggccatgt	tttgc	tttgc	tttgc	tttgc	tttgc	6840	
catccatata	tttgc	tttgc	tttgc	tttgc	tttgc	6900	
cctgcgtt	tttgc	tttgc	tttgc	tttgc	tttgc	6960	
tgttagagaa	tataaa	tttgc	tttgc	tttgc	tttgc	7020	
ttaaaatctt	gttgcgtt	tttgc	tttgc	tttgc	tttgc	7080	
gaaaaagact	cacgtt	tttgc	tttgc	tttgc	tttgc	7140	
gtataaatgg	gttgcgtt	tttgc	tttgc	tttgc	tttgc	7200	
gaagcccgat	gttgcgtt	tttgc	tttgc	tttgc	tttgc	7260	

-continued

---

tacagatgag atggtcagac taaaactggct gacggaattt atgcctcttc cgaccatcaa	7320
gcattttatc cgtactcctg atgatgcattt gttactcacc actgcgcattcc ccggcaaaac	7380
agcattccag gtattagaag aataccttga ttccaggtaa aatattgttg atgcgttgtc	7440
agtgttccgtg cgcccggtgc attcgattcc ttgtttgtat tgccctttt acagcgatcg	7500
cgttattttgtt ctcgcgttccg cgccatccag aatgaataaacc gggtttgggtt atgcgaggta	7560
ttttgtatgac gagcgtaatg gttggcgtt gtaacaaggaa tgaaaaaaa tgcataaagct	7620
tttgccatcc tcacccggattt cagtcgttcc tcatgttgat ttctacttataaatactt	7680
ttttgacgag gggaaaattaa taggttgat ttgtgtggaa cgagtcggaa tcgcagaccg	7740
ataccaggat ttgcgttcc tatggaaactg ctccgggtggat ttttctccctt cattacagaa	7800
acgggtttt caaaataatcgt gtatgtatcc ttctgtatcc aataaaatttcg agtttcattt	7860
gtatgtccatgat gatgttttctt aatcgtactt gacaataaaa agatgtttt tttcaagaaac	7920
tttgatgttcc tatagttttt ttatattgtt gttgttctat ttaatcaaa ttgttagcgtg	7980
atttatattt ttttgcgttcc cgacatccatc tgcccaatccatc gcaaggtaag tgccagaaaa	8040
gtaatataatcgt ggcgttccatc tgatgtatcc ttgtgtccatc tactgtgtc gattcgatcc	8100
taacgcgcgc accaggatcc ttccggaaac cttcaatataaacttccgtatc	8160
atgtatgtcta tacgaaggatcc tttaggtgtata tcagatccatc ta	8202

---

SEQ ID NO: 86            moltype = DNA length = 6151  
 FEATURE                Location/Qualifiers

misc\_feature            1..6151  
 note = Plasmid pCEV-G4-Ph  
 source                1..6151  
 mol\_type = other DNA  
 organism = synthetic construct

SEQUENCE: 86

ggagcgcacccatgcataac ctggaaaaacg aacctgaccc acaggaaaga gttactcaag	60
aataagaattt ttcgttttaa aacctaagag tcactttaa atttgatatac acttattttt	120
tttataactt atttaataat aaaaatcata aatcataaga aattcgotta tttagaagtgg	180
tcacacaacgt atctaccaac gatttgaccc ttccatctt ttccgttaat ttctggcaag	240
gtagacaacgc cgacacaatcc gattggagat ttgaccaaac ctctggggaa gaattgttaa	300
ttaagaggttc acatccatc ttgcgttccatcc ttgtatccca tcgatactatg tttttgtt	360
aaaaataaaaaa aaactttttt tttttgtttt tatttttttgc ttcttagaaaaa gacaaggatgt	420
gcttggggat ttttgcgtt ttatattttt acaatagctg caaatgaaaga atagattcga	480
acattgtgaa gttatgtccatc atatgcgttcc tattttatact tttttttt cagttctgt	540
atattttgtt ttttccctt ttttccatcc ttgtgtccca ataaggatca ggggatctcg	600
aaagatgttcc gggattttc ttggaaaaacg gacttttgc catctaaattt ttccctgttg	660
cctctgaaaa ttatccagca gaagcaatgg taaaagatga acctcagaag aacacgcagg	720
ggcccggaaat ttttgcgttcc gagaatgtgg ggttcataaaaaa agttttttcc ttggaaaaaaa	780
aattttgtgtt tgccttccgtt gagaattttt tcgaatttgc ttgtgtccac tcgtatcgat	840
tctggaaatgtt gttggccatcc ttccaccatc ttgttcccttcc aaaaaaaaaa agatttctca	900
ccccgcacaa attcctgcatt acccccttatttcc ttccacggggc cggccgcacca caccatacg	960
tcacaaatgtt ttttcccttcc ttttacttcc ttccatccatc tcggactccg cgcacatcccg	1020
taccacttca aaacacccaa gcacacgttccatc ttccatccatc ttccatccatc	1080
gtcggttaatcc accccgtacta aagggttggaa aaagaaaaaaa gagacccctt cgtttctt	1140
tcttcgttgcgaaaaggatcaat aaaaatttttt atcacgttcc tttttttgc aaatttttt	1200
ttttgtttttt ttttcccttcc ttgtgtccatcc ttgtgtatccatc aacgggtttcc	1260
aattttcttca attttccatc tttttttttgc ttgttattttt aactttttttt actttttgtt	1320
catttagaaag aaagcatatcc aatctaaatcc aatgtttttatc tacaaggatcc tcgtatccatc	1380
ctcaactatccatc ggcccgccgc tcgacatggaa acagaatgttccatc ttccatccatc	1440
gtaaatgttcc taccgtccgcgtt agctaaatccatc cgtctcaatccatc gaaaggaaag gatgttgcata	1500
acctgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc	1560
tatattttca atttttccatc tttttttgcatccatc cttttttttccatc ttgttgcgttccatc	1620
aaaaaccccttgc ttggggatccatc cttttttttccatc ttgttgcgttccatc ttgttgcgttccatc	1680
ccaaacgcgcg gggggaggatccatc ttgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc	1740
ctcgcgttgcgc tccgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc	1800
acgggttatccatc acagaatccatc ggatccatc ttgttgcgttccatc ttgttgcgttccatc	1860
aaaggccagg aaccgtaaaaa aggccgcgttccatc ttgttgcgttccatc ttgttgcgttccatc	1920
tgacgacatccatc ttgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc	1980
agatattccatc ttgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc	2040
gttttccatc ttgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc	2100
acgctgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc	2160
accccccgttccatc ttgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc	2220
ggtaagacac gacttccatc ttgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc	2280
gtatgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc	2340
gacagtatttccatc ttgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc	2400
ctcttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc	2460
ggggcttccatc ttgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc	2520
cgccatccatc ttgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc	2580
cttcacccatc ttgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc	2640
gttaaacttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc	2700
tcttatttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc	2760
ggggcttccatc ttgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc	2820
agatattccatc ttgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc	2880
tttattccatc ttgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc	2940
agttatattccatc ttgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc	3000
gttttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc	3060
catgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc	3120
ggccgcgttccatc ttgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc	3180
atccgtatccatc ttgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc ttgttgcgttccatc	3240

-continued

tatgcggcga	ccgagggtct	cttgcggcgc	gtcaatacgg	gataataccg	cggcacatag	3300
cagaacttta	aaagtgcctca	tcattggaaa	acgttcccg	ggcgaaaaac	tctcaaggat	3360
cttaccgtcc	tttagatgcc	gttcgtatgt	accacctgt	gcaccacta	gatcttcacg	3420
atcttttact	ttcacccagg	ttcttgggt	agcaaaaaac	ggaaaggaaa	atgcgcgaaa	3480
aaaggaaata	aggcgacac	ggaaatgtt	aataactata	cttccctt	ttaaatatta	3540
ttgaagcatt	tatcagggtt	attgtctcat	gagcggtata	atatttgaat	gtatttagaa	3600
aaataaaacaa	atagggggtc	cgcgcacatt	ccccggaaaa	gtgcacccgt	aacgaagtc	3660
ctgtgcctca	ttttgttagaa	caaaaatgc	acgcggagac	gctaatttt	caaaacaaga	3720
atctgagctg	catttttaca	gaacacaatg	gcaacgcgaa	agcgcttattt	taccaacgaa	3780
gaatctgtgc	ttcatttttg	taaaacaaaa	atgcacacg	agagcgctaa	ttttcaaac	3840
aaagaatctg	agctgcattt	ttacagaaaca	gaaatgcac	gcgagagcgc	tattttacca	3900
acaagaatgc	tatactttt	ttttgtteta	caaaaatgc	tcccgagago	gctattttt	3960
taacaaacga	tcttagatgt	ctttttttt	cttttgtcg	ctctataatg	cagctcttgg	4020
ataacttttt	gcactgttag	tccgttaagg	ttagaagaag	gctacttttg	tgtcttattt	4080
ctttccatca	aaaaaaacgt	gactccactt	ccccgggtt	ctgtactta	gcaagactgc	4140
gggtgcattt	tttcaagata	aaaggcatcc	cgattattat	ctataccgt	gtggatttgc	4200
catactttgt	gaacagaatgg	tgatagcgtt	gatgatctt	cattgtcg	aaaaatatgaa	4260
acgggttctt	ctattttgc	tctatatact	acgtatagga	aatgtttaca	ttttcgtatt	4320
gttttgcatt	cactctatga	atagttctta	ctacaatttt	ttttgcattaa	gagaataact	4380
agagataaaatc	ataaaaaaatg	tagagggtca	gtttagatgc	aagttcaagg	agcgaaagggt	4440
ggatgggttag	gttataatgg	gatataacgac	agagatataat	accaaagaga	tacttttgaa	4500
caatgtttgt	ggaagcggtt	ttcgcataat	tttagtagct	cgttacagtc	cggtgcgtt	4560
ttttgtttttt	gaaagtgcgt	cttcagagcg	cttttgcgtt	tcaaaagcgc	tctgaaagg	4620
ctatactttc	tagaaatgat	gaacttccga	ataggaactt	caaaagcgtt	ccggaaacgaa	4680
gegcttcgg	aaatgcac	cgatgtccgc	acatacagct	cactttca	gtgcacacta	4740
tatctgcgtg	ttgcctgtat	atatatatac	atgagaagaa	cggtatagtg	cgtgtttatg	4800
cttaaatgcg	tacttataatg	cgtctatatt	tgtaggtatg	aaggtagtct	agtacccctt	4860
gttgatattat	ccccatccat	gggggggtatc	gtatgttcc	ttcgcacta	ccctttagct	4920
gtttctatatg	ctgcacccct	tcaatgttg	tagtcttcc	cttcataatgt	atccatccat	4980
tttgatattgg	atcatggtag	acaaccccta	atataactt	gtataatgt	tgctatacga	5040
agttttaggg	tctagatgtc	tggttagatgt	gctctgc	cgccgggtt	ccggccacg	5100
gacatgggg	cccagaatatt	cctcttgc	agtttgc	tgcgcagtc	aggggcat	5160
ttgtgcgttc	ggccgtatcat	tttagccata	catccccat	tataatccat	tgatccatca	5220
cattttgtatg	ggcgacggc	gccaaggaaa	aattacggct	cctcgctca	gacctgcgag	5280
caggggaaacg	ctcccccac	agacgcgtt	aatttgc	acgcggcgc	cctgttagaga	5340
aatataaaaacg	gtttaggat	ggccactgagg	ttttttttt	ataacttcc	tttttttttt	5400
tttgcgttagat	acatgttca	catcacatcc	gaacataaaac	aaccatgggt	atgaccgacc	5460
aaggcgcacgc	caacctgcac	tcacggat	tcgatccac	cgccgcctt	tatgaaagg	5520
tgggtcttcgg	aatcggtttc	cgggacggcc	gtcggtat	cctccagegc	ggggatctca	5580
tgtgtggagtt	cttcgcctcc	ccccgggtcg	atccccctcg	gagttttgtt	agetcgtgc	5640
tttaggtgtca	cgacccgtccg	gaggttccatc	ggcagtgca	atccgtccg	atccggaaa	5700
ccagcagcgg	ctatecgcgc	atccatggcc	ccgaactgca	ggagtgggga	ggcagcatgg	5760
ccgcgttgg	cgacccggac	ggggacgc	tgcgccgtat	acagaacgaa	ttgtttgcag	5820
geatctcgtat	atcagtgatct	actaaataaa	gattttgtt	ttaaagaat	tgtoattttt	5880
atagttttt	tatattgtatg	ttgttattat	ttaatccat	tttgcgtgt	tttatataat	5940
ttttcgcctc	gacatcatct	gcccagatgc	gaagttaaatg	gcccggaaa	taatatcatg	6000
cgtaatcgt	atgtgtatgc	tggtcgatct	actgtgtcg	attcgatact	aacccgcgc	6060
ttccatgtgc	aaaacggat	ctcgaaacc	cttaataataa	cttcgtataa	tgtatgtat	6120
acdaaaattat	taatgtat	caatccatc	a			6151

```
SEQ ID NO: 87 moltype = DNA length = 6861
FEATURE Location/Qualifiers
misc_feature 1..6861
note = Plasmid pbIS-GALkFLP-URA
source 1..6861
mol_type = other DNA
organism = synthetic construct
```

-continued

---

aaatttagagc ttcaatttaa ttatatacg tattacccta tgcgggtgta aataccgcac	1320
agatgcgtaa ggaaaaataa ccgcatcagg aaattgtaaa cgtaattat ttgttaaaat	1380
tccggttaaa ttttggtaa atcagctat ttttaaccat ataggccaa atccggaaaa	1440
tcccttataa atcaaagaa tagaccgaga taggggttag tgggttcca gtttggaca	1500
agagtccact attaaagaac gtggacttca acgtcaaaagg gcgaaaaacc gtctatcagg	1560
gcgatggccc actacgtaa cccatccccct aatcaaggat tttggggtcg aggtgcgt	1620
aaggactaaa tcggaccctt aaaggggccccc cccgatttag agtttgcgg gggaaaggccg	1680
cgaacgtggc gagaaaggaa gggaaaggaa cggaaaggccg gggcgttgg ggcgtggca	1740
gtgttgcggc caccgtgcg gtaaccacca caccgcggc gcttaatgcg ccgtatcagg	1800
gcccgtgcgc cccatgcgc ttcagggtgc gcaactgtt ggaactggc tcgggtcggg	1860
cctcttcgcg attacggccag ctggcggaaag ggggatgtgc tcaaggccg ttaaagggtgg	1920
taacgccagg gtttccag tcaacgtt gtaaaacgc gggcagttag cggcgttaat	1980
acgactact ataggccgaa ttgggttaccg gggcccccctt cgagggtgac ggtatcgt	2040
agttgtat cgaatggggc cagccgggg gatccaaaaat tcatcgctt gctgatataat	2100
taccccgaaaa ataaaggctaa aaaaactatc gcatatccat cctatgtt ttaatttgat	2160
tcgttcat tttttttttt gggccagggt tactgttcaat ttttccctt cataaccata	2220
aaagctgtaa ttgttgcatttttttggggatgtgc ggcggggcc acatctgcgt	2280
ttcaggaacg cggccggcc gacggggcc gacggggcc gagggttctt tcggggggct	2340
gtcacccgcg cggccggctt taatccgtat tcaatataat caatggccat ttaaaggctat	2400
tactgaaatg tccaaagaga aggtttttt aggctaatg aatggggctc tttacatttc	2460
cacaacatataat aagtttgcatttccat gatgttatgat atgttatat ggtgttatg	2520
ccatgttataat tgatttttttgcgttccat gttccatccaa aaaaaaaatg aaaaatttttgcgt	2580
gggtcgacatg ccacaatttgcgttccat gttccatccaa aaaaaacccca cctaaagggtc ttgttgcgtca	2640
gtttgtggaa aggtttggaa gacccgttccat gttccatccaa aaaaaatg aaaaatttttgcgt	2700
aaaccttataat ttgttgcgttccat gttccatccaa cggaaacacca atcaagagag ccacattcat	2760
gagctataat actatccat gcaatttcgttccat gttccatccaa aatcactccaa	2820
gtttaaataat aagacgcggaa aagcaatccat ttttttttttgcgttccat gttccatccaa aatgttgcgt	2880
tgotttggggaa ttacaattttaat ttttttttttgcgttccat gttccatccaa aatgttgcgt	2940
tattgttaatg agtttgcatttccat gttccatccaa aatgttgcgttccat gttccatccaa aatgttgcgt	3000
ccacagtaaa aaaaatgttccat gttccatccaa aatgttgcgttccat gttccatccaa aatgttgcgt	3060
tgagaaaaatccat ttttttttttgcgttccat gttccatccaa aaaaaaaatg aaaaatttttgcgt	3120
ccatttccttccat ttttttttttgcgttccat gttccatccaa aaaaaaaatg aaaaatttttgcgt	3180
tgatccggaaat ttttttttttgcgttccat gttccatccaa aaaaaaaatg aaaaatttttgcgt	3240
agtgcacagatccat gttccatccaa aaaaaaaatg aaaaatttttgcgttccat gttccatccaa aaaaaaaatg aaaaatttttgcgt	3300
gatcgatccat ttttttttttgcgttccat gttccatccaa aaaaaaaatg aaaaatttttgcgt	3360
agtttttttttgcgttccat gttccatccaa aaaaaaaatg aaaaatttttgcgt	3420
cttagtccat ttttttttttgcgttccat gttccatccaa aaaaaaaatg aaaaatttttgcgt	3480
aaaaaaatgcg cccaaatctccat ttttttttttgcgttccat gttccatccaa aaaaaaaatg aaaaatttttgcgt	3540
gggcctaaacg gggccggccat gttccatccaa aaaaaaaatg aaaaatttttgcgt	3600
ggccaggccaa acgttataatccat gttccatccaa aaaaaaaatg aaaaatttttgcgt	3660
ttctcggtatccat ttttttttttgcgttccat gttccatccaa aaaaaaaatg aaaaatttttgcgt	3720
taatccatccat gggggatgtgc gttccatccaa aaaaaaaatg aaaaatttttgcgt	3780
acgatccatccat gttccatccaa aaaaaaaatg aaaaatttttgcgt	3840
cataaaataga cccatataatccat gttccatccaa aaaaaaaatg aaaaatttttgcgt	3900
atgttatataat atatacaggccat gttccatccaa aaaaaaaatg aaaaatttttgcgt	3960
gttgcgtccat ttttttttttgcgttccat gttccatccaa aaaaaaaatg aaaaatttttgcgt	4020
atttccatccat ttttttttttgcgttccat gttccatccaa aaaaaaaatg aaaaatttttgcgt	4080
ttgttcccttccat ttttttttttgcgttccat gttccatccaa aaaaaaaatg aaaaatttttgcgt	4140
tgtgtgaaat ttttttttttgcgttccat gttccatccaa aaaaaaaatg aaaaatttttgcgt	4200
taaaggccgttccat ttttttttttgcgttccat gttccatccaa aaaaaaaatg aaaaatttttgcgt	4260
cggtttccat gttccatccaa aaaaaaaatg aaaaatttttgcgt	4320
gagaggccgttccat ttttttttttgcgttccat gttccatccaa aaaaaaaatg aaaaatttttgcgt	4380
ggtcgttccat ttttttttttgcgttccat gttccatccaa aaaaaaaatg aaaaatttttgcgt	4440
agaatccatccat gttccatccaa aaaaaaaatg aaaaatttttgcgt	4500
ccgtaaaatccat gttccatccaa aaaaaaaatg aaaaatttttgcgt	4560
aaaaaaatccat gttccatccaa aaaaaaaatg aaaaatttttgcgt	4620
gtttcccttccat gttccatccaa aaaaaaaatg aaaaatttttgcgt	4680
ctctgtccat ttttttttttgcgttccat gttccatccaa aaaaaaaatg aaaaatttttgcgt	4740
tctcgttccat gttccatccaa aaaaaaaatg aaaaatttttgcgt	4800
ggccggccat ttttttttttgcgttccat gttccatccaa aaaaaaaatg aaaaatttttgcgt	4860
cttccatccat ttttttttttgcgttccat gttccatccaa aaaaaaaatg aaaaatttttgcgt	4920
tgtctacatccat ttttttttttgcgttccat gttccatccaa aaaaaaaatg aaaaatttttgcgt	4980
tatctgtccat ttttttttttgcgttccat gttccatccaa aaaaaaaatg aaaaatttttgcgt	5040
caaaacaaatccat ttttttttttgcgttccat gttccatccaa aaaaaaaatg aaaaatttttgcgt	5100
aaaaaaatccat ttttttttttgcgttccat gttccatccaa aaaaaaaatg aaaaatttttgcgt	5160
cgaaaaactccat ttttttttttgcgttccat gttccatccaa aaaaaaaatg aaaaatttttgcgt	5220
ccctttccat ttttttttttgcgttccat gttccatccaa aaaaaaaatg aaaaatttttgcgt	5280
tgacagttccat ttttttttttgcgttccat gttccatccaa aaaaaaaatg aaaaatttttgcgt	5340
atccatccat ttttttttttgcgttccat gttccatccaa aaaaaaaatg aaaaatttttgcgt	5400
tggcccccgttccat ttttttttttgcgttccat gttccatccaa aaaaaaaatg aaaaatttttgcgt	5460
aataaaacccat ttttttttttgcgttccat gttccatccaa aaaaaaaatg aaaaatttttgcgt	5520
catccatccat ttttttttttgcgttccat gttccatccaa aaaaaaaatg aaaaatttttgcgt	5580
gcccggccat ttttttttttgcgttccat gttccatccaa aaaaaaaatg aaaaatttttgcgt	5640
tttccatccat ttttttttttgcgttccat gttccatccaa aaaaaaaatg aaaaatttttgcgt	5700
aaaaggccgttccat ttttttttttgcgttccat gttccatccaa aaaaaaaatg aaaaatttttgcgt	5760
atcaactccat ttttttttttgcgttccat gttccatccaa aaaaaaaatg aaaaatttttgcgt	5820
cttttctgttccat ttttttttttgcgttccat gttccatccaa aaaaaaaatg aaaaatttttgcgt	5880
gagttgtccat ttttttttttgcgttccat gttccatccaa aaaaaaaatg aaaaatttttgcgt	5940
agtgcgtccat ttttttttttgcgttccat gttccatccaa aaaaaaaatg aaaaatttttgcgt	6000

-continued

gagatccagt	tcgatgtaac	ccactcggtc	acccaactga	tcttcagcat	cttttacttt	6060
caccacggtt	tctgggttgag	caaaaaacagg	aaggccaaaat	gcgcgcaaaaaa	agggaataaag	6120
ggcgacacgg	aaatgttgaa	tactcaatct	ttccctttt	caatattttt	gaaggcattta	6180
tcagggttat	tgtctcatga	gcccgtatcat	attttgtatgt	atttagaaaaaa	ataacaataat	6240
agggggttccg	cgcacatttc	ccccggaaaatg	gccacccctggg	tcctttcat	cacgtgtat	6300
aaaaataat	ataatttaaa	ttttttataaa	taaatatataa	attttaaaaaa	agaagaaatgg	6360
aaaaagaattt	aaagaaaaaa	tagtttttgt	tttcggaaaga	tgtaaaaagac	tctagggggg	6420
tcggccaaacaa	atactaccctt	ttatcttgc	cttcctgtc	tcaggatttttt	atgcggaaat	6480
gtttcatctt	gtctgtgtag	aagaccacac	acggaaaatcc	tgtgtattttta	cattttactt	6540
atcggttaatc	gaatgtatata	ctattttttttt	tcgtttttttt	gtctataaaaa	tatatatgtt	6600
aaatgcgtt	ttttgttgaaaa	tttttttttttt	cttttttttttt	tttttttttttt	tttttttttttt	6660
actcttctac	tttttttttttt	tactttctaa	aatccaaata	caaaacataaa	aaataaaataa	6720
acacagagta	aatttcccaa	ttatccatc	ataaaaaagat	acggaggcgcg	tgttaagttt	6780
aggcaacggc	tccgttccat	gaaaacatttt	ttatcatgac	attaaccttat	aaaaataggcc	6840
qatccacqaq	qccttttcqtc	c				6861

```
SEQ ID NO: 88          moltype = DNA    length = 6857
FEATURE                  Location/Qualifiers
misc_feature             1..6857
note = Plasmid pCEV-URA3-TEF-AT-yABAB-cMyc
source                   1..6857
mol_type = other DNA
organism = synthetic construct
```

SEQUENCE: 88  
 agcttttcaa ttcaattcat cattttttt ttattttttt ttttgatttc ggtttcttgg 60  
 aaatttttt gatccggtaa tctccgaaaca gaaggaaaga cgaaggaagg agcacagact 120  
 tagatggta tatatacgcg tatgtatgt tgaagaaca taaaatggcc ctagattttt 180  
 aacccaactg cacagaacaa aaaccagcg gaaacagaaga taaatcatgt cgaagactac 240  
 atataaggaa cgtgctgcg ctcacccatg tcctgtgtc gccaagctat ttaatatcat 300  
 gcacaaaaact caaacaaact gtgtgtgtc attggatgtt cgtatccacca aggaattact 360  
 ggatgtatgt gaagcattag gtcccaaaat ttgtttaata aaaaacatcg tggatatttt 420  
 gactgtatgtt tccatgggg gcacagttaa gcccgtttaaaag gcattatccg ccaagtacaa 480  
 ttttttactt ttcgaagata gaaaatttgc tgacatgtt aatacagtc aatttgcgtt 540  
 ctctgggggt gtatacagaa tagcagaatg ggcagacatt aacgtatgc acggctgtgt 600  
 gggccccaggat attgttagcg ttgttgaagca ggcggcggaa gaagtaacaa aggaaacttag 660  
 aggccttttt atgttagcag aattgtcatg ccaaggctcc ctatctactg gagaatataatc 720  
 taagggtact gttgacatcg cgaaaaggcg caaagatttt gtatccggc ttatgtctca 780  
 aagagacatg ggttggaaagat atgaaggatc cgttgggtt atttgcacat ccggctgtgg 840  
 ttttagatgc aaggggatg cattgggttca acagtataga accgttggatg atgttgttcc 900  
 tacaggatct gacattatata ttgttggaaag aggactattt gcaagggaa gggatgtctaa 960  
 ggttagagggt gaacgttaca gaaaagcagg ctgggaagca tatttggaaa gatggggccca 1020  
 gcaaaaaactaa aaaactgtat tataatgtaaa tgcatgtata ctaaaactac aaattagagc 1080  
 ttcataattaa ttatcatgtt tatttcccg ggcgcacaca ctagatcttcc aaaaatgtttc 1140  
 tactccctt ttactcttcc agattttcc gggactcccg catccgcgtt ccacttcaaa 1200  
 acacccaagc acagcataact aaattcccc tctttttcc tctagggtgt cgttaattac 1260  
 ccgtacttaaa ggtttggaaa agaaaaaaaga gaccggctcg ttttttttcc ttcgtcgaaa 1320  
 aaggcaataaa aaattttttt cactgtttt ttttctggaaa attttttttt tttgtttttt 1380  
 tctcttgcg tgaccctccca ttgatattttt agtataaaaaa cggcttccaa ttttcaatgt 1440  
 ttcagttca tttttcttgt tctattacaa ctttttttac ttcttgctca ttggaaagaa 1500  
 agcatatcgaa tctaattctaa gtttttaat caaggatcca ttagattttcc ticaattttt 1560  
 actgtgtttt tatttgcgcg atccctccgc ttagatgtc aagtcaattt ggttggaaacc 1620  
 gggtgggtgtt tagttcaacc aggttgtatgt tttagatattt ctgtgtctgc atcagggtttt 1680  
 acattggattt attcttcaat aggttgggtc agacaagctc ctggtaaaga aagagaagggt 1740  
 gtttcttgcg tatttcggatc ttgtgtactca actaaatatg ctgtactccgt taagggttgc 1800  
 ttcaactactt caagagatata cgcgtaaaaaa acgttctact tgcataatggaa ctatccaa 1860  
 ccagatgaca cagcagtcg ttaacttgcg gcttttagag ccacatgtg cgggtgttcc 1920  
 ccatttgtctc ttacggtaa agatgactgg ggttaaaggta ctttagttac tgcatttcattcc 1980  
 ggttgggtgtt gttccgggtgg ttgttggatgt ggttgggtgtt gtttcaatgt tcaatttagt 2040  
 gaatccgggtt gtgggttagt ccaacccgtt ggttagttaa gattatctg cgaacaaatgt 2100  
 gggttttacat tagattatca cggatccgtt tggtttagac aaccacccgtt taaggaaagaa 2160  
 gaagctgtctt ttatatttc cgtttagtgcg agaactatata tgcacccgtt ttctgttaag 2220  
 ggttagattca caatttcaag agacaatggc aagaacccgtt tttatgtca aatgaactct 2280  
 ttgaagagac aagacaccgc atttttttac tgcacccgtt ggttggggta gaagatttt tgcatttc 2340  
 gtcaacagat ggttgcgcg cgtttagatgt gtttggggta gggatccacca agtcccccgtt 2400  
 agttctgtgtt gtgggtccgg tggttggatgt ggttgggtgtt ctgggtgtgg ttccacaattt 2460  
 caatttagtagt aaacttgggg tgggttggatgtt caacccagggtt gttcccttgcg attaagtgtt 2520  
 gtcgtatctt gtttttactt ctctgtatcc gttatgcattt ggttgcacca agtccccgtt 2580  
 aaaggctctt aatggatcgc tcaattatca accgcacccgtt ccacatgtag agatgacatgt 2640  
 accaagggtt gatttcaatctt tcaagatgtt aacgcataaga acacattgtt cttccaaatgt 2700  
 acctcttgcg aaccagaaga caccgcatttattactgtt ccagaggtagt agtcatatcc 2760  
 gccagtgctca tcagagggtc agtaaagggtt cctggtaactt aagttacagt ctcttcagggt 2820  
 gggccgggggtt gtggggggcc gggccgttgcg ggttgggtgtt cacaaggctca atttggatgtt 2880  
 tctgggtgtt gttttagttca aacttgggtgtt cttatgttgcgat ttccttcgttcc 2940  
 tccatccgtcgtt gtttccaaatc ttgttacatgtt ctaacccgtt ccacccgtt atctttccaa 3000  
 tgggtcgctt caatgaccaaa gactaacaac gaaatctattt ctgatccgtt taagggttaga 3060  
 ttcatttattt caagagatata tgcataatggt accgttattt tgcataatggaa ctatccgtt 3120  
 ccagaagatca ctgggtgttcc tctccgcg ggttgcgttca taagggttagt aggttacatgtt 3180  
 gtaacatgtt ctccgtcgtt cttccgttccatggacacaa aagtgttgcgtt ccgaacaaaga cctccgtt 3240  
 gtttgggttcc gccggccgttcc aacgcacccgtt ctaacccgtt ccacccgtt atctttccaa 3300

-continued

```
SEQ ID NO: 89          moltype = DNA  length = 5199
FEATURE                Location/Qualifiers
misc_feature           1..5199
note = Plasmid pCEV-URA3-TEF-cMyc
source                 1..5199
                      mol_type = other DNA
                      organism = synthetic construct
```

SEQUENCE:	89	Organism - Synthetic Construct	
agcttttcaa	ttcaattcat	catttttttt ttatttttt ttttgatttc ggtttcttgc	60
aaatttttt	gattcgtaa	tctccgaaca gaaggaaga cgaaggaaagg agcacagact	120
tagatggta	tatatacgc	tatgtatgt tgaaagaaaa tggaaatggc cagtattttc	180
aacccaaact	cacagaacaa	aaaccaggac gaaacgaaaga taaaatcatgt cggaaactac	240
atataaggaa	cgtgtgccta	ctcatcttag tcctgttgcg gcacagatgt ttaatatcat	300
gcacgaaaag	caaacaact	tgtgtgttcc attggatgtt cgtaccacca aggaattact	360
ggagttagt	gaagcattag	gtccccaaat ttgtttacta aaaacacatg tggatcttgc	420
gactgattt	tccatgggg	gcacagttaa ggccgctaaag gccattatccg ccaagttacaa	480
ttttttactt	ttcgaagata	gaaaatttgc tgacatttgt aatacagtca aatttgcgttca	540
cttcgcgggt	gtatacagaa	tagcagaatg ggcagacattt acgaatgcac acgggtgtgg	600

-continued

ggcccaggat	attgttagcg	gttgaagca	ggggcagaaa	gaagtaacaa	aggaacctag	660
aggcctttg	atgttagcg	aattgtcatg	caagggctc	ctatctactg	gagaataatac	720
taagggtact	gttgcatttgc	cgaaaagcga	caaagatttt	gttgcggct	ttatgtctca	780
aagagacatg	ggtggaaagag	atgaaggta	cgattggtt	attatgcac	ccgggtgtgg	840
tttagatgac	aaggagatg	cattgggtca	acagataga	accgtggatg	atgttgctc	900
tacaggatct	gacattatta	tttgtggaa	aggactattt	gcaaaggaa	gggatgtctaa	960
ggtagagggt	gaacgttaca	gaaaaggcgg	ctggaaagca	tatttgagaa	gatgcggcca	1020
gcaaaactaa	aaaactgtat	tataagtaaa	tgcattgtata	ctaaactcac	aaattagagc	1080
ttaattttaa	ttatgtatcg	tatttcccc	gccgcacaca	ccatagotc	aaaatgttc	1140
tactccttt	ttactcttc	agatttttc	ggactcccg	catecgccgt	ccacttcaaa	1200
acacccaa	acacatata	aaatttccc	tctttcttc	tctagggtgt	cgttaattac	1260
ccgtactaa	ggtttggaa	agaaaaaaaga	gaccgcctcg	tttcttttc	ttcgtcgaaa	1320
aaggcaataa	aaatttttat	cacgtttc	tttcttgcataa	ttttttttt	ttgattttt	1380
tctcttttgc	tgaccccca	ttgtatatta	agtaataaa	cgggtttcaa	ttttcaagt	1440
ttcagttca	tttttttgc	tctattacaa	cttttttac	ttcttgcota	tttagaaagaa	1500
agcatagcaa	tctaattctaa	gtttaatta	caaggatctc	gacatggaa	agaagttgtat	1560
ttccgaaagaa	gacccgtgt	aagcttggta	ccgcggctag	ctaagatccg	ctctaacccga	1620
aaaggaaagaa	gttagacaac	ctgaatgtca	ggtccctatt	tagttatgtt	1680	
agtattaa	acgttattt	tattttaaa	ttttttttt	tttctgtaca	gacgcgtgt	1740
cgcattgtac	attatactga	aaacccgtct	tgagaagggt	ttgggacgct	cgaagatcca	1800
gctgcattaa	tgaatccggcc	aacgcgcggg	gagagggcggt	ttgcgtatgt	ggcgctttc	1860
cgcttctcg	ctcaactgt	cgcgcgtc	cgctgttgc	ctgcgttgc	cggtatcagc	1920
tcactcaaa	goggtaatac	ggttattc	agaatccagg	gataacccgc	gaaagaacat	1980
gtgagcaaaa	ggccagcaaa	aggccaggaa	ccgtaaaaaa	ccgcgttgc	ttggcggttt	2040
ccataggotc	ccccccctgc	acgaggatca	caaaaatcga	cgctcaatgc	agaggtggcg	2100
aaacccgaca	ggacttataaa	gataccaggc	gttcccccc	ggaagetccc	tcgtgcgtc	2160
tctctttccg	accctccgc	ttaccggata	ctcttccgc	tttcttccctt	cgggaaacgcgt	2220
ggcgcttct	catagtcac	gctgttagta	tetcagttc	gtgttaggtc	ttcgtctcaa	2280
gtcggttgt	gtgcaccaac	ccccccgtca	gcccgcaccc	tgccgcctat	ccggtaacta	2340
tctgttgc	tccacccccc	taagacacga	cttattgcoc	ctggcagcag	ccactggtaa	2400
caggatttgc	agagcgagg	atgtaggcgg	tgtctacag	tttttgcata	ggtggccctaa	2460
ctacggctac	actagaaggaa	cagtattttg	tatctgcgtc	ctgtgtca	cagttacatt	2520
cggaaaaaa	gttggtagt	tttgatccgg	caaacaacc	accgcgtgt	gccccgttt	2580
ttttgttgc	aaggcaggc	ttacccgcgt	aaaaaaaggaa	tctcaaga	atcccttgc	2640
ctttctacg	gggtctgc	cctcgttgc	cgaaaaactca	cggttgcata	ttttggcata	2700
gagattatca	aaaaggatct	tcacccat	ctttttaaat	taaaaatgaa	gtttaataatc	2760
aatctaaagt	atatatgtat	aaacttggtc	tgacagttac	caatgtttaa	tcagtggaggc	2820
acccatctca	gogatctgtc	tatttgcgt	atccatagtt	gcctgactcc	ccgtcgctgt	2880
gataactacg	ataccggagg	gttccattac	ttggcccgat	gtgcgtatga	taccgcgaga	2940
ccccacgtca	ccggccccc	atttacago	aataaacc	ccagccggaa	gggcccggcg	3000
cagaagtgt	cctgcacactt	tatccgcctc	catccagtct	attaattgtt	gccgggaa	3060
tagagtaat	atgtccgc	ttatagtttgc	ggccaaacgt	tttgccat	ctacaggcat	3120
cgtgggtc	cgctcgctgt	tttgcgttgc	tttgcattc	cccggttccc	aacgatcaag	3180
ccggatgttca	tgatccccca	ttgtgtgc	aaaacccgtt	agctccctcg	gtccctccgt	3240
cgttgcaga	agtaagtgg	ccgcgtgtt	atcaactcat	tttgcgttgc	caactgcataa	3300
ttctctact	gtcatgtccat	ccgcgtatgt	tttttgcgt	actgggtgt	actcaacccaa	3360
gtcatttgc	gaaatgttgc	tgccgcgg	gagttgtct	tgccggcg	caataccggaa	3420
taataccgc	ccacatagca	gaactttaaa	agtgctcatc	attggaaaac	gttcttccgg	3480
cgcaaaactc	tcaaggatct	taccgttgc	gagatccat	tcgtgttac	ccactcg	3540
acccaactcg	tcttcgcac	cttttactt	caccgcgt	tttgcgttgc	caaaaacagg	3600
aaggcaaaa	ggccaaaaaa	agggaataa	ggcgacacgg	aatatgttga	tactcataact	3660
cttcctttt	caatattatt	gaacatttgc	tcagggttat	tgtctatgc	gcccgtatcat	3720
atttgaatgt	atttagaaaa	ataaaacaaat	aggggttcc	cgcacatttc	cccgaaaagt	3780
cccacccgt	ccgacatctt	gtgttcatt	tttgcgtatgc	aaaatgcac	gcgagacgc	3840
taattttca	aacaaagaaat	ctgagatgc	tttttgcata	acagaaatgc	aacgcgaaag	3900
cgtatatttt	ccaaacaa	atctgttgc	cattttgttgc	aaacaaaat	gcaacccgc	3960
agcgctaatt	tttcaacaa	agaatctgc	ctgcatttttgc	acagaacaga	aatgcacgc	4020
gagagcgtca	tttttaccaac	aaagaaatct	tacttcttttgc	aaaatgcac	4080	
ccgagagcgc	tatttttctca	acaaacgcac	ttagattact	tttttgcgt	tttgcgt	4140
ctataatgc	gtcttgcatt	aaacttttgc	actgttagtgc	cggttgcgtt	agaagaaggc	4200
tactttggt	tctatatttgc	cttccat	aaaagcgttgc	ctccacttcc	ccgcgttact	4260
gattactgc	gaagctgc	gttcatttttgc	taaagataaa	ggcattcccg	attatatttgc	4320
ataccatgt	ggatgttgc	tacttgcgt	acagaaatgt	atgcgttgc	tgattctca	4380
tttgtc	tttgtc	tttgttgc	ttttgttgc	tatataactac	gtatggaaa	4440
tgtttacatt	ttcgttgc	tttgcatttgc	ctctatgc	tttttttgc	actatatttttgc	4500
tgtctaaaga	gtataactat	agataaaat	aaaaatgttgc	gagggtcgat	tttagatgc	4560
gttcaacccgt	ccaaaggatgt	tttttttttttgc	tttttttttttgc	tttttttttttgc	tttttttttttgc	4620
caaaagagata	cttttgc	atgtttgttgc	aaagccgttgc	tttttttttttgc	tttttttttttgc	4680
tttacatgt	ccgcgttgc	tttttttttttgc	tttttttttttgc	tttttttttttgc	tttttttttttgc	4740
aaaacccgtc	tgaatgttgc	tttttttttttgc	tttttttttttgc	tttttttttttgc	tttttttttttgc	4800
aaggcgttcc	ggaaacccgc	gtttccgttgc	tttttttttttgc	tttttttttttgc	tttttttttttgc	4860
ctgttgc	ccgcacccata	tttttttttttgc	tttttttttttgc	tttttttttttgc	tttttttttttgc	4920
ccatgttgc	tttttttttttgc	tttttttttttgc	tttttttttttgc	tttttttttttgc	tttttttttttgc	4980
ggtagtgc	tacccctgt	tttttttttttgc	tttttttttttgc	tttttttttttgc	tttttttttttgc	5040
caccaactacc	tttttttttttgc	tttttttttttgc	tttttttttttgc	tttttttttttgc	tttttttttttgc	5100
tcaatgtat	tttttttttttgc	tttttttttttgc	tttttttttttgc	tttttttttttgc	tttttttttttgc	5160
ataatgtat	tttttttttttgc	tttttttttttgc	tttttttttttgc	tttttttttttgc	tttttttttttgc	5199

-continued

FEATURE	Location/Qualifiers
misc_feature	1..7208 note = Plasmid pCEV-G4-Km-TEF-AT-γABAB hAA6T83N-tagless
source	1..7208 mol_type = other DNA organism = synthetic construct
SEQUENCE: 90	
	cacacaccat agcttcaaaa tggttctact cttttttac tcttccagat tttctcgac 60 tcgcgcacat ggcgttacac ttccaaacac ccaagcacag catactaatt ttccccttt 120 tttccctata ggggtcggtt aattttccgtt actaaagggtt tgaaaaaggaa aaaagagacc 180 gcgtcggttc tttttttcg tcggaaaaagg caataaaaaat ttttatcagc ttctttttc 240 ttgaaaattt tttttttgta tttttttc tttcgatgac ctccccattgtat tttttttttt 300 aataaaacggt ttcaatttca tcaaggtttca gtttcattttt tttttttttt ttacaacttt 360 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 420 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 480 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 540 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 600 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 660 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 720 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 780 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 840 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 900 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 960 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 1020 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 1080 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 1140 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 1200 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 1260 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 1320 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 1380 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 1440 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 1500 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 1560 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 1620 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 1680 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 1740 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 1800 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 1860 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 1920 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 1980 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 2040 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 2100 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 2160 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 2220 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 2280 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 2340 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 2400 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 2460 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 2520 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 2580 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 2640 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 2700 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 2760 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 2820 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 2880 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 2940 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 3000 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 3060 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 3120 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 3180 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 3240 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 3300 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 3360 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 3420 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 3480 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 3540 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 3600 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 3660 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 3720 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 3780 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 3840 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 3900 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 3960 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 4020 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 4080 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 4140 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 4200 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 4260 ttttaacttct ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 4320

-continued

---

gcaacgcgag	agcgctaatt	tttcaaaca	agaatcttag	ctgcattttt	acagaacaga	4380
aatgcacgc	gaaaggccta	tttaccac	gaagaatctg	tgttcattt	ttgtaaaca	4440
aaaatgcaac	gcgagagcgc	taattttca	aacaaagaat	ctgagctgc	ttttcacaga	4500
acagaaatgc	aacgcgagag	cgtatttta	ccaacaaga	atctatactt	ctttttgtt	4560
ctacaaaaat	goatcccgag	agcgttat	ttctaaaca	gcatcttaga	ttacttttt	4620
tctcccttgt	gogctctata	atgcagtct	ttgataactt	tttgactgt	aggccgtta	4680
aggttagaag	aaggctactt	ttgtgtctat	tttctctcc	ataaaaaaag	cctgactcca	4740
cttcccgct	ttactgatta	ctagcgaago	tgccgggtca	tttttcaag	ataaaggccat	4800
cccccattat	attctatacc	gatgtggatt	gcccatactt	tgtgaacaga	aagtgtatgc	4860
gtttagatgt	cttcatttgtt	cagaaaatta	tgaacccgtt	tttctatttt	gtctctat	4920
actacgtata	gaaaaatgtt	acatctctgt	attgttttc	attcactcta	tgaataggte	4980
ttactacaat	tttttgtct	aaagagtaat	actagagata	aacataaaaa	atgttagaggt	5040
cagatttttaga	tgcaagttca	aggagcggaa	ggtggatgg	taggttat	agggatata	5100
cacagagata	tataggaa	agataactt	gagaatgtt	tgtggaaacg	gtattcgc	5160
tatTTTtagta	gtctgttaca	gtccggcgt	tttttgtt	tttggaaatgt	cgtcttcaga	5220
gcccgtttgg	ttttcaaaag	cgtctgtt	ttcttatact	ttcttagagaa	taggaactt	5280
ggaataggaa	cttcaaaagc	tttccggaa	cgagcgtt	cgaaaaatgc	acgcggactg	5340
cgcacatata	gtctactgtt	cacgtcgac	ctatactgt	gtgttgcctt	tatataatata	5400
tacatgaaag	gaaacggcata	gtcgctgtt	atgtttaat	gctgttat	atgcgtctat	5460
ttatgttaga	tgaaaggtag	tctgttac	cctgttat	tatcccattt	catgcgggg	5520
atcgtagtct	tccttcagca	ctaccctta	gtcttctat	atgcgtccac	tcctcaattt	5580
gatttagtct	atcccttaat	gtcttctat	cctttgtat	tggatccatgg	tagacaacc	5640
ttaataataac	ttcgatataat	gtatgttata	cgaatgtt	aggcttagag	atctgtttag	5700
cttgctctgt	ccccgcgggg	tcaccggc	agegacatgg	aggcccagaa	taccctcctt	5760
gacagtcgtt	acgtgcgcag	ctcaggggca	tgatgttact	gtcgccccgt	catttagccc	5820
atacatcccc	atgtataatc	atttgcatt	atacattttt	atggccgcac	ggcgcgaagc	5880
aaaatttcgt	gtcttcgtt	gcagacctt	gaggcaggaa	acgccttcc	cacagacccg	5940
ttgaattgtc	cccaacgcgc	gccccgttag	agaaatataa	aaggtaggaa	tttgcactg	6000
aggttcttct	ttcatatact	tcctttttaaa	atctgttact	gatacgttc	tcacatcaca	6060
tccogaacata	aaaaacccat	ggtaaaggaa	agactcaact	ttcgaggccg	cgattaaat	6120
ccacatcgtt	tgctgttata	tatgggtata	aatgggtctt	cgataatgc	gggcaatctag	6180
gtgcgacaaat	ctatcgatt	tatgggaacg	ccgtatgc	agagttgtt	ctgaaacat	6240
gcaaaaggtag	cttgcacat	gatgttac	atgagatgt	cagactaaac	ttgctgcacgg	6300
aatttgc	tcttcggacc	atcaacatt	ttatccgtt	tccgtatgt	gcatggttac	6360
tcaccactgc	gtccccggc	aaaaacgt	tccaggat	agaagaaat	cctgtatt	6420
gtgaaaatata	tggtgttgc	ctggcaatgt	tccgtcgcc	gttgcattt	attctgtt	6480
gtaattgtcc	ttttaaacagc	gatcgctat	ttcgtctcg	tcaggcgc	tcacaaat	6540
ataaacgggtt	gggttgcgt	agtgttattt	atgacgcgg	taatggotgg	cctgttgaac	6600
aagtctggaa	agaaatgtat	aaatgttgc	cattcttacc	ggatgttgc	gtcactat	6660
gtgatttttc	acttgcataac	tttatttttgc	acgaggggaa	attatagg	tgtattgt	6720
ttggacgagt	cggaatcgca	gaccgatacc	aggatcttgc	catcctatgg	aactgcctt	6780
gtgagttttc	tccttcata	cagaaacccg	ttttcaaaa	atatgttatt	gataatct	6840
atatgtat	tttgcgtt	catttgcatt	tcgtatgtt	tttctat	gtactgacaa	6900
taaaaagatt	ttgttttca	agaacttgc	atttgcatt	ttttttata	ttgtatgt	6960
tctatTTTaa	tcaaatagtt	gcgtat	tatTTTTT	cgccctcgaca	tcatct	7020
agatgcgtt	ttaatgcgtt	agaaatgtt	atcatgcgtt	aactgtatgt	aatgtgtt	7080
cgctatactg	ctgtcgat	gataacta	ccgcacat	gtgtcgaaa	cgagctct	7140
agaaccccta	ataataactt	gtataatgtt	tgctat	acttatttgc	tgatatcaga	7200
tccactag						7208

---

SEQ ID NO: 91                  moltype = DNA    length = 10379  
 FEATURE                         Location/Qualifiers  
 misc\_feature                 1..10379  
 note = Plasmid pCRI-Sb-delta1  
 source                        1..10379  
 mol\_type = other DNA  
 organism = synthetic construct

SEQUENCE: 91

tcgcgcgaac	ccaaactgcac	agaacaaaaa	cctgcaggaa	acgaagataa	atcatgtca	60
aagctatata	taaggAACgt	gctgtactt	atcttagtcc	tgttgcgtcc	aagctat	120
atatcatgc	cgaaaaacaa	acaaaactt	gtgttcat	ggatgttct	accaccaagg	180
aattactgta	gttagttgtt	gcattagtt	ccaaaatttt	tttactaaaa	acacatgtgg	240
atatcttgc	tgatgttcc	atggaggggc	cagttacgc	gtataaggca	ttatccgcca	300
agtacaat	tttacttctt	agacagaa	aatttgc	cattgttatt	acagtcaat	360
tgcaatgc	tcgcgggtgt	tacagaaat	cagaatgggg	agacatttac	aatgcacacg	420
gtgtgtgtgg	ccccaggatt	gttagcggtt	tgaagcggc	ggcagaagaa	gtaaacaagg	480
aacctagg	ctttttgtat	tttagcagaa	ttgtcat	gggccttca	tctactggg	540
aatataactt	gggttactgtt	gacattgc	agagcgc	aaattttgg	atcggttta	600
ttgtctca	agacatgggt	ggaaagatgt	aaagggttca	tttgcgttatt	atgcacccg	660
gtgtgtgtgg	agatgacaag	ggagacgc	tgggtcaaca	gtatagaacc	gtggatgt	720
tcgtttctac	aggatctgac	attattattt	tttgcgttatt	acttatttgc	aaggaaagg	780
atgtctat	agagggtgaa	cgttacagaa	aaagcaggct	ggaagcatat	ttgagaagat	840
cgccgcac	aaactaaaa	actgttatt	aagtaatgc	atgtataact	aactcacaaa	900
tttagatgtt	aatttataat	tatctgtt	taccctgtc	tcgatgtt	aaaaagataa	960
tgtatgtt	tgctttactt	catatttata	cagaaacttgc	atgtttttt	tcgatgtat	1020
acaagggtat	tatctgtac	tttgcgtt	aactctat	tttgcgtt	cctcttggc	1080
tagcggtaa	ggtgcgtt	tttccaccc	ctacaatgtt	ctgttcaaaa	gatTTTgtt	1140
aaacgcgtt	gaagtgtaa	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	1200
aaagataat	gtatcggtt	tagatgtat	ctgttttgc	tttgcgtt	tttgcgtt	1260

-continued

---

tattatata tagttttaga gctatgtgt tttgaatggg cccaaaacctt tttttgtttt	1320
ttatgtctgc atagttca aatgtttcta ctccccctt actcttcag attttctcg 1380	
actccgcgca tcgcccgtacc acttcaaaac acccaagcac agcataactaa atttccctc	1440
tttcttcctc taggggtgtcg ttaatttaccg gtactaaagg tttggaaaag aaaaaagagg	1500
cgcgcctcggtt tctttttttt cggtggaaaaa ggcaataaaa atttttatca cgtttcttt	1560
tcttgaaaat tttttttttt tctttcgatg acctcccccatt gatatttaag 1620	
ttataaaaaacg gtctcaatt tctcaagtt cagtttcat tttcttgttc tattacaat 1680	
ttttttactt ctgttcatt agaaaagaaag catagcaatc taatctaagt tttcttagatg	1740
gattataaag atgacgtatc caaaccttc aaaaagaaga gaaaggcgtca taagaatatac	1800
tcaataggcg tagatgtatcg cacaatagc gtcggatggg cggtgtatc tggatgat	1860
aagggttcgtt ctaaaaggtt caaggttcgtg gggaaaaatcag accggcacag tatcaaaaaa	1920
aatctttagt gggctttttt atttgacagt ggagagacag cggaaagcgc tctgtctcaa	1980
cggacacgtc gtggaaaggta tacacgtcgq aagaatcgta tttttatc acaggagat	2040
tttccaaatg agatggcgaa agttagatg agttttttt atcgtatcg agagtctttt	2100
tttgtggaaag aagacaagaa gcatgtatcg cttccatattt ttggaaatata agtagatgaa	2160
gttgcatttac atgagaata tccaaatcattt tttccatgtc gaaaaaaaaattt ggttagatct	2220
actttaaaag cggatgttcg cttaaatctat ttggccattt cgcattatgt taagttcgt	2280
ggtcattttt tggatgggg agatttat cctgttatc gtgtgttggg caaaacttattt	2340
atccaggatgg tttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	2400
gttagatgtca aaggcgattt ttctgcacga tttagttaat caagacgatt agaaaatctc	2460
attgtcgacg tccccgggtt gaaaggaaaat ggcttattttt ggtttttttt tgcttgc	2520
tttgggttttgc cccccccattttt tttttttttt tttttttttt tttttttttt tttttttttt	2580
ctttccaaatg atacttcgtc ttggatgtttaat gataattttt tttttttttt tttttttttt	2640
tatgctgatt ttggggggc agtcaagaat ttatcgatg ctatttttt tttttttttt	2700
ctaaaggtttaa atacttcaatc aacttcaaggct ccccttacgtt cttttttttt tttttttttt	2760
gatgttccatc atcaagactt gactttttttaa aaggttttttgg gttttttttt tttttttttt	2820
aagtataaag aatctttttt tttttttttt tttttttttt tttttttttt tttttttttt	2880
ggagcttagcc aagaagaattt ttttttttt tttttttttt tttttttttt tttttttttt	2940
actggaggat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	3000
gacaacggctt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	3060
caagaagactt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	3120
tttcgaattt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	3180
acttcggaaatgt ctggaaac aattttttttt tttttttttt tttttttttt tttttttttt	3240
gtttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	3300
aaagtacttac caaaatccat tttttttttt tttttttttt tttttttttt tttttttttt	3360
aaagtcaat atgtttttttt tttttttttt tttttttttt tttttttttt tttttttttt	3420
aaaggccattt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	3480
gaaggtttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	3540
agatttatgt ttccatgtt tttttttttt tttttttttt tttttttttt tttttttttt	3600
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	3660
tttgaagata gggagatgt tttttttttt tttttttttt tttttttttt tttttttttt	3720
aagggtatgtt aacatgtttt tttttttttt tttttttttt tttttttttt tttttttttt	3780
tttgcattatgt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	3840
gatgggtttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	3900
gaagacatttca aaaaagcaca agtgcgttgc caaggcgatc gttttttttt tttttttttt	3960
aattttatgtt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	4020
gaatgggtttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	4080
aaaaatcaga caactttttt tttttttttt tttttttttt tttttttttt tttttttttt	4140
gaaggtatgtt aagaattttttt tttttttttt tttttttttt tttttttttt tttttttttt	4200
tttgcattatgt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	4260
caaggatatttca tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	4320
tttccttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	4380
aaatccggata acgtttttttt tttttttttt tttttttttt tttttttttt tttttttttt	4440
tttccattttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	4500
ggagggtttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	4560
caaatctttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	4620
aatgtttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	4680
tttccggatatttca tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	4740
gatgtttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	4800
tcgggtttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	4860
gagcaagaaa taggtttttttt tttttttttt tttttttttt tttttttttt tttttttttt	4920
tccaaatggat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	4980
aatggggaaaatcg tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	5040
gtttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	5100
tccaaaggatgt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	5160
tgggatccat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	5220
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	5280
atccacaatgt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	5340
ggatataaaggat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	5400
tttagaaaacgtgtgtt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	5460
ctggcttcgtt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	5520
aagggttgc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	5580
tttagatgatgtt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	5640
aattttatgtt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	5700
gcagaaaataatgtt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	5760
tatttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	5820
acttcattccat cttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	5880
ggaggtgtttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	5940
ttatgtttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	6000

-continued

---

tcgaagttc	acgataaaaa	gataatattc	tacttttgc	tcccaccgcg	tttgctagca	6060
cgagtgaaca	ccatccctcg	cctgtgagtt	gtacccattc	ctctaaactg	tagacatgg	6120
agcttcagca	gtgttcgtta	tgtacggcat	cctccaacaa	acagtcgggt	atagtttgc	6180
ctgtccctct	gaatcgagtc	cctcgatatt	tctcatacta	gttctagaga	tctgccaatt	6240
gaacataaca	ttgttagttac	ataactagt	aatatggttt	ggcacacatt	aaaagtataa	6300
aaactatctg	aattacgaat	tacatataat	ggtcataaaa	atcaatcaat	catctgtgt	6360
tttatatgtc	tcttatctaa	gtataagaat	atccatagt	aatattcaact	tacgttacct	6420
ttaaccttgt	aatcattgtc	aacaggatat	gttaacgacc	cacattgata	aacgctagta	6480
tttcttttc	ctcttcatt	tgggccgtt	tctctatact	ccccatata	ctgtttctt	6540
tcgttccgtat	tggttagtgc	ttggggctc	gtggccaca	tgttacgttg	ttgggtctgc	6600
gggtggaaac	gaaactctgg	gagtcgtat	cattcaaaa	acgatagcaa	6660	
gttaaaaataa	ggctagtcg	ttatcaactt	gaaaaagtgg	caccgtgt	gtgtttttt	6720
tccatatacca	acttcaatt	taatttttt	tttttaattt	tcattttt	gcgatataca	6780
aagaccctgc	aggcatgca	gcttggctg	atcatggta	tagtgtt	ctgtgtgaaa	6840
ttgttatccg	ctcacaatc	cacacaatc	acgagccgg	agcataaagt	gtaaaagctg	6900
gggtgcctaa	tgagttagtgc	aactcacatt	aatttgcgtt	cgctcaactgc	ccgcttcca	6960
gtcgccaaac	ctgtcgcc	agctgcattt	atgaatcgcc	caacgcccgg	ggagaggcgg	7020
tttgcgtatt	ggcgcgtt	ccgettctc	gtcaactgac	tegtgttgt	cggtcggt	7080
gctcgccgca	gggttatacg	ctcaatcaaa	ggcggtaata	cggatatacc	cagaatcagg	7140
ggataaacga	ggaaaagaaca	tgtgagcaaa	aggccagcaa	aaggccagga	accgtaaaaa	7200
ggccgctgt	ctggcggtt	tccataggt	ccggcccccc	gacgacatc	acaaaaatcg	7260
acgtctcaatg	cagagggtgc	gaaaccggac	aggactataa	agataccagg	cgttttcccc	7320
tggaaatcc	tcctgtcgct	tccctgttc	gaccctgcgg	tttaccggat	acctgttcgc	7380
cttctccct	tggggaaagcg	tttgcgtt	tcatagctc	cgctgttagt	atctcagg	7440
ggtgttagtgc	gttcgttcca	agcttgggt	tgtgcacgaa	ccccccgttc	agccccaccc	7500
ctgcgcctta	tccggtaact	atctgcgttgc	gtccaaacccc	gttagacacg	acttattcgcc	7560
actggcagca	gcaactggta	acaggattat	caagagccgg	tatgttggcg	gtgtacaga	7620
gttcttgaag	tggtggctta	actacggta	cactagaaga	acagtattt	gtatctgcgc	7680
tctgtcaag	ccagttaccc	tcggaaaaag	agttgttagc	tcttgatccg	gcaaaacaac	7740
cacccgttgt	acggcggtt	ttttgttgc	caaggcagca	attacgogca	gaaaaaaaaagg	7800
atetcgaa	gatccattgc	tcccttca	gggttctgc	gtctgttgc	acgaaaactc	7860
acgttaagg	attttgtca	tgagattatc	aaaaaggatc	tccacataga	tccttttaaa	7920
ttaaaatgt	atgttttaat	caatctaaag	tatatatgt	taaaacttgt	ctgacagatta	7980
ccaatgetta	atcaatgttgc	cacctatcc	acgatctgt	ctattttgtt	catccatagt	8040
tgcctgtact	cccgctgtgt	agataactac	gatacgggg	ggcttccat	ctggccccag	8100
tgcgtcaatg	ataccggcgc	tcccacgtc	acccggctca	gattttatca	caataaaaa	8160
gccaggccgga	aggggccgagc	gcagaagttg	tccgtcaact	ttatccgc	ccatccagtc	8220
tattatgtt	tgccgggaat	cttagatgt	tagttcgca	gttaatagt	tgccgaacgt	8280
tgttgccatt	gtcacggca	tcgtgtgtc	acgctcggt	tttggatgg	cttcattcag	8340
ctccggttcc	caacgatcaa	ggcggat	atgtatccc	atgttgtca	aaaaaggcgt	8400
tagetccctc	ggtcctccga	tcgttgtc	aagtaagtgg	gcccagtgt	tatcaactat	8460
gtgttatggca	gcaactgcata	atcttcttc	tgtcatgcca	tccgtgttgc	atctttctgt	8520
gactgggtgag	tactcaacca	agtcatctgc	agaatagtgt	atgcggcgc	cgagttgtc	8580
ttggccggc	tcaataccgg	ataataccgc	gccacatagc	agaactttaa	aagtgtcat	8640
cattggaaa	cggttctcgg	ggcggaaact	ctcaaggatc	ttaccgttgt	tgagatccag	8700
tccgtatgt	ccactctcg	cacccactg	atcttcacgc	tcttttactt	tcaccagcgt	8760
ttctgggtga	ccaaaaacacg	gaaggcaaaa	tgcgcggaaa	aaggggataa	gggcgcacacg	8820
gaaaatgttga	atactcatac	tcccttctt	tcaatattat	tgaaggat	atcagggtt	8880
ttgtctcatg	agcggataca	tatttgaatg	tatttagaaa	aataaaacaaa	taggggttcc	8940
ggccgatattt	ccccggaaaat	tgccacccgt	acgaagcatc	tgtgtttccat	tttggatgg	9000
aaaatgtca	ccgcggagcg	ctaattttt	aaacaaagaa	tctgtatgc	atttttacag	9060
aacagaaatg	caacgcgaaa	cgcttattt	accaacga	aatctgtgt	tcatttttgt	9120
aaaacaaaaa	tgcaacgcga	gagcgctat	tttcaaaaca	aagaatctg	gctgcattt	9180
taaaaaatc	aatatcgac	cgagagcgt	attttacca	caaagaatct	atacttctt	9240
tttgcgttca	aaaaatgtat	cccgagagc	cttattttt	aaacaagatc	cttagattac	9300
ttttttttc	ttttgtcgcc	tctataatgc	agttcttgc	taacttttt	cactgttgt	9360
ccgttaagg	taagaagaagg	ctactttgtt	gtctatttt	tcttcataa	aaaaaggcgt	9420
actccacttc	cccgctttac	tgtagtact	cgaaatgtcg	gggtgcattt	ttcaagataa	9480
aggcattcccc	gattatatac	tataccgtat	tggattgcgc	atacttttgt	aacagaaatgt	9540
gtatgcgtt	atgttcttc	atggtagtca	aaatattgaa	cggttcc	tattttgtt	9600
ctatataacta	cgtataggaa	atgttacat	tttgcgtat	ttttcgatc	actctatgaa	9660
tagttcttac	tacaatttt	ttgttcttaa	agtaatacta	gagataaaaca	taaaaaatgt	9720
agagggtcgag	tttagatgc	agttaatgc	ggcggaaatgt	gttgggttgc	ttatataagg	9780
atatacgaca	gagatataca	gtccaaatgc	acggatgttgc	aatgttgcgt	gaagcggtat	9840
tcgcaatatt	tttagtagtgc	tttacagtcc	gggtgcgtttt	ttgggttttt	aaagtgcgtc	9900
ttcagagcgc	ttttgttttt	caaaagcgt	ctgaagtcc	tatactttt	agagaatagg	9960
aacttcggaa	taggaatcc	aaagggttcc	cgaaaaacag	cgcttccgaa	aatgcacacg	10020
gagctcgcc	catacgtc	acttgttcc	tcgcacccat	atctgtgt	tgccgtata	10080
tatataataca	tgagaagaac	ggcatagtgc	gtgtttatgc	ttaaatgcgt	acttataatgc	10140
gtcttattat	gttaggtgaa	aggtgtatgc	gtacccctgc	tgatattatc	ccatccatgt	10200
cggggatgt	atgttccat	ttcaatgtca	tttgcgttgt	ttctatatgc	tgccactcc	10260
caattggatt	agtctcatcc	ttcaatgtca	tttgcgttgt	ttctatatgc	tgccactcc	10320
aaccattatt	atcatgacat	taacctataa	aaataggcgt	atcacggggc	ccttcgtc	10379

---

SEQ ID NO: 92

moltype = DNA length = 10379

FEATURE

Location/Qualifiers

misc\_feature

1..10379

note = Plasmid pCRI-Sb-delta2

source

1..10379

-continued

mol\_type = other DNA  
 organism = synthetic construct

SEQUENCE: 92

```

tcgcgcgaac ccaactgcac agaacaaaaa cctgcaggaa acgaagataa atcatgtcga 60
aagctacata taaggaacgt gctgtacto atcctagtctc tgggtgtcc aagctattna 120
atatacatgca cgaaaagcaa acaaattgtt gtgtttcatc ggatgttgc accaccaagg 180
aattactgga gttagttgaa gcattagtc ccaaaatttg ttactaaaa acacatgtgg 240
atatcttgac tgattttcc atgggggca cagttagcc gctaaaggca ttatccgcca 300
agtacaattt ttactcttc gaagacagaa aatttgcgtc cattggtaat acagtcaaat 360
tgcgtactc tgcgggtgtc tacaatagc cagaatgggc agacatcaag aatgcacacg 420
gtgtgtggc cccagggtt gttacgggtt tgaagcaggc ggcagaaagg gtaacaagg 480
aaccttagagg cttttgcgtt ttagcagaat tgcgtatgc gggctccct tctactggag 540
aatataactaa gggatgtt gacatgcga agagcgcacaa agattttgtt atcggttta 600
ttgtcaaaag agacatgggtt ggaagagatg aaggttgcga ttgggtgatt atgacacccg 660
gtgtgtgggtt agatgacaag ggacgcgt tgggtcaaca gtatagaacc gtggatgtg 720
tgcgttctac agatgtcgtc attattattt tggaaagagg actatttgc aagggaaagg 780
atgcgtaaagg agagggtgtt cgttacagaa aacgcaggctt ggaagcatat ttgagaagg 840
ggggccagca aactaaaaaa actgttattt aagtaaatgc atgtatata aactcacaaa 900
tttagagcttc aatttaattt ttagcttatc taccctgtc tgcgttta gaaaagataa 960
tgcgtatgtt tgcgttctact catatttata cagaaacttg atgtttttt tcgagttat 1020
acaagggtgt tacatgtacg tttgaagttt aactcttagat tttgtgtgc cctcttggc 1080
tagcggtaaa ggtgcgttcc tttcacacc ctacaatgtt tgggtcaaaa gattttgttc 1140
aaacgcgtgtc gaagttttttt ggtgtcgca tgggtcggtt tggaaactt ctccgcgtt 1200
aaagataaat gatcggttt tagagctatg ctgttttgc tggcccggaa actatggatt 1260
ctttaatctt cgggttttgc gctatgtgtt tttgaatggt cccaaaaactt tttttgttt 1320
ttatgtctgc atagcttcaat tttgttctt ctccttccatg attttctcg 1380
actccgcgcga tggccgttcc acttcaaaaac acccaacgc acgtactaa atttccctc 1440
tttcttctc taggggtgtcg ttaatttaccg gtactaaagg tttggaaaag aaaaagagg 1500
ccgcgcgtt ttttttttgc cgttcaaaa ggcataaaaaa atttttatca ctgcgtt 1560
tcttggaaat ttttttttgc gatgttttttgc ttttgcgtt acctccattt gatatttttgc 1620
ttaataaaacg gtcttcattt tctcaagttt ctttttttttgc ttttttttttgc ttttttttttgc 1680
ttttttactt ctgtcttccattt agaaaaaaaat catagcaatc taatctaattt ttttttttttgc 1740
gattataaaag atgacgtatca caaaccccttca aaaaagaaga gaaagggtcgta taagaaatac 1800
tcaataggct tagatgtatccg cacaatagcgtt gtcggatggg cgggtgtatc tgcgtatgtt 1860
aagggttccgtt ctttttttttgc ctttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 1920
aatctttagt gggcttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 1980
cgacacgttccgtt ctttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 2040
tttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 2100
tttgggtggaaat gacatgtatccg ctttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 2160
gtttgttccattt atgagaaatata tccaaacttccatc ttttttttttgc ttttttttttgc ttttttttttgc 2220
acttataaaat gggatttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 2280
gggttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 2340
atcccgatgttccattt atgatgtatccg ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 2400
gttagatgttccattt atgatgtatccg ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 2460
attgtctccattt atgagaaatata tccaaacttccatc ttttttttttgc ttttttttttgc ttttttttttgc 2520
tttgggtttccattt atgagaaatata tccaaacttccatc ttttttttttgc ttttttttttgc ttttttttttgc 2580
ctttccaaatgttccattt atgatgtatccg ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 2640
tatgtctccattt atgatgtatccg ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 2700
ctaaagatgttccattt atgatgtatccg ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 2760
gatgtatgttccattt atgatgtatccg ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 2820
aagtataaaat gaaatgtatccg ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 2880
ggagctatccgtt ctttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 2940
actgagaaatata ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 3000
gacaacccgtt ccattttccatc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 3060
caagaagactt ccattttccatc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 3120
tttcgtatgttccattt atgatgtatccg ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 3180
actcggttccattt atgatgtatccg ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 3240
gtttcgatgttccattt atgatgtatccg ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 3300
aaagtacttccattt atgatgtatccg ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 3360
aagggttccattt atgatgtatccg ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 3420
aaagccatttccattt atgatgtatccg ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 3480
gaagattttccattt atgatgtatccg ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 3540
agatgtatccgtt ctttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 3600
tttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 3660
tttgggtttccattt atgatgtatccg ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 3720
aagggttccattt atgatgtatccg ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 3780
tttgggtttccattt atgatgtatccg ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 3840
gatgttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 3900
gaagacatgttccattt atgatgtatccg ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 3960
aatttagtcttccattt atgatgtatccg ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 4020
gaatttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 4080
gaaatcgttccattt atgatgtatccg ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 4140
gaagggttccattt atgatgtatccg ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 4200
tttgcgttccattt atgatgtatccg ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 4260
caagaatgttccattt atgatgtatccg ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 4320
tttgcgttccattt atgatgtatccg ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 4380
aaatcggttccattt atgatgtatccg ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 4440
cttctaaatccgtt ctttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 4500
ggaggtttccattt atgatgtatccg ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 4560

```

-continued

-continued

---

tttttttctc ctttgtgcgc tctataatgc agtctcttgc taactttttg cactgttaggt 9360  
 ccgttaagg tagaagaagg ctactttgtt gtctatttc tcttcataa aaaaagcctg 9420  
 actccacttc ccgcgtttac tgattactag cgaagctgcg ggtgcattt ttcaagataa 9480  
 aggcatcccc gattatattc tataccgtat tggattgcgc atactttgtt aacagaaagt 9540  
 gatagcgctg atgatccctc attgttgatc aaattatgaa cggtttctc tattttgtt 9600  
 ctatataacta cgtataggaa ttgcattaca tttcgatcc actctatgaa 9660  
 tagttcttac tacaattttt ttgtcttaag agtaataacta gagataaaaca taaaaaatgt 9720  
 agaggtcgag tttagatgca agttaaaggaa ggcgaagggtt gatgggttagg ttatataagg 9780  
 atatagcaca gagatataat gcaaaaggat accttttgcg aatgtttgtt gaagcggtat 9840  
 tcgcaatatt tttagatgctt gttacagtcc ggtgcgtttt tgggtttttt aaagtgcgtc 9900  
 ttcagagcgc ttttggttt caaaaggctt ctgaaaggatc tatactttt agagaatagg 9960  
 aacttcggaa taggaacttc aaaggcttc cgaaaacacgag cgcttccgaa aatgcacgc 10020  
 gagctgcgc catacacttc actgttccacg tggcacttacat atctgcgttgc tgcctgtata 10080  
 tatataatcata tggagaaggac ggcataatgcg ggttggatgt tttttatgcg acttataatgc 10140  
 gtttattttt gttagatgaa agtaggtctt gtagcttcgtt tgatattatc ccattccatg 10200  
 cggggtatcg tttttttttt tcagactac ctttttagtgc ttctatatgc tgccactcct 10260  
 caattggatt agtctcatcc ttcaatgctt tcaatttcgtt tgatatttgcg tcatctaaga 10320  
 aaccattattt atcatgacat taacctataa aaataggcgat atcacggggc ctttcgtc 10379

SEQ ID NO: 93                moltype = DNA length = 126  
 FEATURE                        Location/Qualifiers  
 source                        1..126  
 mol\_type = other DNA  
 organism = *Saccharomyces boulardii*  
 SEQUENCE: 93  
 tggtaataa aaaatcaact atcatctact aacttagtatt tacgttacta gtatattatc 60  
 atatacggtg tttagaaggatc acgcaatga tgagaaatag tcatctaaat tagtggaaagc 120  
 tggaaac 126

SEQ ID NO: 94                moltype = DNA length = 105  
 FEATURE                        Location/Qualifiers  
 source                        1..105  
 mol\_type = other DNA  
 organism = *Saccharomyces boulardii*  
 SEQUENCE: 94  
 aatattttata gaattgtgtttaa gaattgcaga ttcccttttta tggattccctt aatccctcgag 60  
 gagaacttct agtatactca catacctaat attatagcct taatc 105

SEQ ID NO: 95                moltype = DNA length = 6919  
 FEATURE                        Location/Qualifiers  
 misc\_feature                1..6919  
 note = Plasmid pPL5071\_TEF1-Cre\_URA3  
 misc\_feature                2382  
 note = n is a, c, g, or t  
 source                        1..6919  
 mol\_type = other DNA  
 organism = synthetic construct  
 SEQUENCE: 95  
 agaaaccatt attatcatga cattaaccta taaaatagg cgtatcacga ggccctttcg 60  
 ttcgcgcgt ttcggtgatc acggcgaaaa cctctgacatc atgcagctcc cggagacgg 120  
 cacagtttgcgatc ttgttgcggcg atgcggggatc gagacaaaggcc cgtcaggggc cgtcaggccc 180  
 tggggggggggcgttgcggatc ttgttgcggatc gagacaaaggcc cgtcaggccc 240  
 gcaccatacc acagtttgcgatc ttgttgcggatc gagacaaaggcc cgtcaggccc 300  
 tggggggggggcgttgcggatc ttgttgcggatc gagacaaaggcc cgtcaggccc 360  
 ggagcacaga ttgttgcggatc ttgttgcggatc gagacaaaggcc cgtcaggccc 420  
 cccagatttc ttaaccaac ttgttgcggatc ttgttgcggatc gagacaaaggcc cgtcaggccc 480  
 gtcggaaatgcg atatataatgcg aacgtgtgcg tactccatctt tactctgttgcg 540  
 atttaatataatgcg atatataatgcg aacgtgtgcg tactccatctt tactctgttgcg 600  
 caaggaaatgcg ttgttgcggatc ttgttgcggatc gagacaaaggcc cgtcaggccc 660  
 tggggggggggcgttgcggatc ttgttgcggatc gagacaaaggcc cgtcaggccc 720  
 cgcgttgcgatc ttgttgcggatc ttgttgcggatc gagacaaaggcc cgtcaggccc 780  
 caaaatgcg ttgttgcggatc ttgttgcggatc gagacaaaggcc cgtcaggccc 840  
 acacgggtgtc ttgttgcggatc ttgttgcggatc gagacaaaggcc cgtcaggccc 900  
 aaaggaaatgcg ttgttgcggatc ttgttgcggatc gagacaaaggcc cgtcaggccc 960  
 tggggggggggcgttgcggatc ttgttgcggatc gagacaaaggcc cgtcaggccc 1020  
 ctttatttgcg atatataatgcg aacgtgtgcg tactccatctt tactctgttgcg 1080  
 acccggtgtc ttgttgcggatc ttgttgcggatc gagacaaaggcc cgtcaggccc 1140  
 tgatgttgcg ttgttgcggatc ttgttgcggatc gagacaaaggcc cgtcaggccc 1200  
 aaggatgcgt aaggatgcgt ttgttgcggatc ttgttgcggatc gagacaaaggcc cgtcaggccc 1260  
 aagatgcgtc ttgttgcggatc ttgttgcggatc gagacaaaggcc cgtcaggccc 1320  
 acaaatttgcg ttgttgcggatc ttgttgcggatc gagacaaaggcc cgtcaggccc 1380  
 acagatgcgt aaggatgcgt ttgttgcggatc ttgttgcggatc gagacaaaggcc cgtcaggccc 1440  
 atttgcgttgcg ttgttgcggatc ttgttgcggatc gagacaaaggcc cgtcaggccc 1500  
 aatcccttgcg ttgttgcggatc ttgttgcggatc gagacaaaggcc cgtcaggccc 1560  
 caagatgcgt aaggatgcgt ttgttgcggatc ttgttgcggatc gagacaaaggcc cgtcaggccc 1620  
 gggcgttgcg ttgttgcggatc ttgttgcggatc gagacaaaggcc cgtcaggccc 1680  
 taaaaggacta aatcggaaacc ttgttgcggatc ttgttgcggatc gagacaaaggcc cgtcaggccc 1740  
 ggcgttgcg ttgttgcggatc ttgttgcggatc gagacaaaggcc cgtcaggccc 1800  
 aagatgtttgcg ttgttgcggatc ttgttgcggatc gagacaaaggcc cgtcaggccc 1860

-continued

---

gggcgcgtcg	cgcattcgc	cattcaggct	gcccactgt	tgggaaggc	gatecggtcg	1920
ggccttcgc	ctattacgc	cc agtggcgaa	agggggatg	gtcgaagc	gattaagg	1980
ggtaacgc	cca ggtttccc	agtca cgc	ttgtaaaac	acggcc	agcgcgc	2040
atac gact	ca tataggcg	aattgggt	cggccgaa	ttaaaggc	cgagcgt	2100
aaaaccc	tttcaagc	tttca	atgttacat	cgtaac	cg tctgtac	2160
aaaaaaag	aaat	tttca	ataaaataa	gttctt	ataacataa	2220
taaatagg	ctttagt	ctttagt	actc	tttccgtt	tgactcgagg	2280
gggggagg	gtgaatgt	gctgtac	actaattaca	tgactc	tcgacaatt	2340
caaac	tacc	caagagtc	ccaaactc	acatca	angcaaaacc	2400
tcttctcg	ttgcattc	cacaaat	tcagcact	gcatactt	gacaggaa	2460
gcaacgg	ata ttggat	atcagg	ctatcg	gtttaa	gacaatgac	2520
gctggc	act	aatattat	gttca	agcacta	at tttccgt	2580
tccttctc	taatgt	tttca	ataattat	caatca	cgcgtaat	2640
getaatgc	ccat	tttccgt	tttca	atccat	tttccgt	2700
tatagttc	at	aggcgc	tttccgt	tttca	agggttac	2760
aaactcc	ccgc	ccata	tctcg	cttgc	ccagacc	2820
ccaggat	cttgc	ccat	tttccgt	cgccgt	tttccgt	2880
aaatcc	ccagg	ccat	tttccgt	caatcg	tttccgt	2940
ttttctg	ccgg	ccaa	tttccgt	tttccgt	tttccgt	3000
tcatecg	ccat	tttccgt	tttccgt	tttccgt	tttccgt	3060
taacc	ccat	tttccgt	tttccgt	tttccgt	tttccgt	3120
tatctt	aaac	tttccgt	tttccgt	tttccgt	tttccgt	3180
ccagaat	ccat	tttccgt	tttccgt	tttccgt	tttccgt	3240
cctgg	ccat	tttccgt	tttccgt	tttccgt	tttccgt	3300
cgttt	ccat	tttccgt	tttccgt	tttccgt	tttccgt	3360
gcag	ccat	tttccgt	tttccgt	tttccgt	tttccgt	3420
ctgg	ccat	tttccgt	tttccgt	tttccgt	tttccgt	3480
gaacc	ccat	tttccgt	tttccgt	tttccgt	tttccgt	3540
gcattt	ccat	tttccgt	tttccgt	tttccgt	tttccgt	3600
togaac	ccat	tttccgt	tttccgt	tttccgt	tttccgt	3660
gtaaat	ccat	tttccgt	tttccgt	tttccgt	tttccgt	3720
gccc	ccat	tttccgt	tttccgt	tttccgt	tttccgt	3780
tttt	ccat	tttccgt	tttccgt	tttccgt	tttccgt	3840
aaaaat	ccat	tttccgt	tttccgt	tttccgt	tttccgt	3900
ttaact	ccat	tttccgt	tttccgt	tttccgt	tttccgt	3960
agagaa	ccat	tttccgt	tttccgt	tttccgt	tttccgt	4020
cgg	ccat	tttccgt	tttccgt	tttccgt	tttccgt	4080
gagg	ccat	tttccgt	tttccgt	tttccgt	tttccgt	4140
agtcc	ccat	tttccgt	tttccgt	tttccgt	tttccgt	4200
tttcc	ccat	tttccgt	tttccgt	tttccgt	tttccgt	4260
tgtgt	ccat	tttccgt	tttccgt	tttccgt	tttccgt	4320
taaag	ccat	tttccgt	tttccgt	tttccgt	tttccgt	4380
cgctt	ccat	tttccgt	tttccgt	tttccgt	tttccgt	4440
gagagg	ccat	tttccgt	tttccgt	tttccgt	tttccgt	4500
ggtcgt	ccat	tttccgt	tttccgt	tttccgt	tttccgt	4560
agaat	ccat	tttccgt	tttccgt	tttccgt	tttccgt	4620
cgt	ccat	tttccgt	tttccgt	tttccgt	tttccgt	4680
caaaa	ccat	tttccgt	tttccgt	tttccgt	tttccgt	4740
gttcccc	ccat	tttccgt	tttccgt	tttccgt	tttccgt	4800
cctgt	ccat	tttccgt	tttccgt	tttccgt	tttccgt	4860
tetc	ccat	tttccgt	tttccgt	tttccgt	tttccgt	4920
gccc	ccat	tttccgt	tttccgt	tttccgt	tttccgt	4980
cttat	ccat	tttccgt	tttccgt	tttccgt	tttccgt	5040
tgt	ccat	tttccgt	tttccgt	tttccgt	tttccgt	5100
tatct	ccat	tttccgt	tttccgt	tttccgt	tttccgt	5160
caaa	ccat	tttccgt	tttccgt	tttccgt	tttccgt	5220
aaaaa	ccat	tttccgt	tttccgt	tttccgt	tttccgt	5280
cgaaa	ccat	tttccgt	tttccgt	tttccgt	tttccgt	5340
ccttt	ccat	tttccgt	tttccgt	tttccgt	tttccgt	5400
tgacag	ccat	tttccgt	tttccgt	tttccgt	tttccgt	5460
atccat	ccat	tttccgt	tttccgt	tttccgt	tttccgt	5520
tttcc	ccat	tttccgt	tttccgt	tttccgt	tttccgt	5580
aaaaa	ccat	tttccgt	tttccgt	tttccgt	tttccgt	5640
catcc	ccat	tttccgt	tttccgt	tttccgt	tttccgt	5700
gccc	ccat	tttccgt	tttccgt	tttccgt	tttccgt	5760
tttcc	ccat	tttccgt	tttccgt	tttccgt	tttccgt	5820
aaaaa	ccat	tttccgt	tttccgt	tttccgt	tttccgt	5880
atca	ccat	tttccgt	tttccgt	tttccgt	tttccgt	5940
ctttt	ccat	tttccgt	tttccgt	tttccgt	tttccgt	6000
gagg	ccat	tttccgt	tttccgt	tttccgt	tttccgt	6060
agtg	ccat	tttccgt	tttccgt	tttccgt	tttccgt	6120
cacc	ccat	tttccgt	tttccgt	tttccgt	tttccgt	6180
cacc	ccat	tttccgt	tttccgt	tttccgt	tttccgt	6240
ggcg	ccat	tttccgt	tttccgt	tttccgt	tttccgt	6300
tcagg	ccat	tttccgt	tttccgt	tttccgt	tttccgt	6360
aggg	ccat	tttccgt	tttccgt	tttccgt	tttccgt	6420
aaa	ccat	tttccgt	tttccgt	tttccgt	tttccgt	6480
aaa	ccat	tttccgt	tttccgt	tttccgt	tttccgt	6540
tcgcca	ccat	tttccgt	tttccgt	tttccgt	tttccgt	6600

---

-continued

---

gtttcatctt gtctgttag aagaccac acgaaaatcc tgtgattta cattttactt 6660  
 atcgtaatc gaatgtatat ctatataatc tgctttctt gtctaataaa tatatatgtt 6720  
 aagtacgccc ttgttgaaa tttttaaac ctttgttat tttttttct tcattccgtt 6780  
 actcttctac cttctttatt tacttctaa aatccaaata caaaacataa aaataaataa 6840  
 acacagaga aattcccaa ttattccatc attaaaagat acgaggcgcg tgtaagttac 6900  
 aggcaagcga tccgtccta 6919

SEQ ID NO: 96 moltype = AA length = 89  
 FEATURE Location/Qualifiers  
 source 1..89  
 mol\_type = protein  
 organism = *Saccharomyces cerevisiae*

SEQUENCE: 96 MRPPSIFTAV LFAASSALAA PVNTTTEDET AQIPAEAVIG YSDLEGDFDV AVLPPSNSTN 60  
 NGLLFINTTI ASIAKEEGV SLEKREAEA 89

SEQ ID NO: 97 moltype = AA length = 61  
 FEATURE Location/Qualifiers  
 source 1..61  
 mol\_type = protein  
 organism = *Saccharomyces cerevisiae*

SEQUENCE: 97 MRPPSIFTAV LFAASSALAA PVNTTTEDEL EGDFDVAVLP FSASIAAKEE GVSLEKREAE 60  
 A 61

SEQ ID NO: 98 moltype = AA length = 57  
 FEATURE Location/Qualifiers  
 source 1..57  
 mol\_type = protein  
 organism = *Saccharomyces cerevisiae*

SEQUENCE: 98 MRPPSIFTAV LFAASSALAA PVNTTTEDEL EGDFDVAVLP FSASIAAKEE GVSLEKR 57

SEQ ID NO: 99 moltype = AA length = 19  
 FEATURE Location/Qualifiers  
 source 1..19  
 mol\_type = protein  
 organism = *Saccharomyces cerevisiae*

SEQUENCE: 99 MRPPSIFTAV LFAASSALA 19

SEQ ID NO: 100 moltype = AA length = 20  
 FEATURE Location/Qualifiers  
 source 1..20  
 mol\_type = protein  
 organism = *Aspergillus niger*

SEQUENCE: 100 MVAWSLFLY GLQVAAPALA 20

SEQ ID NO: 101 moltype = AA length = 18  
 FEATURE Location/Qualifiers  
 source 1..18  
 mol\_type = protein  
 organism = *Aspergillus awamorii*

SEQUENCE: 101 MSFRSLLALS GLVCSGLA 18

SEQ ID NO: 102 moltype = AA length = 16  
 FEATURE Location/Qualifiers  
 source 1..16  
 mol\_type = protein  
 organism = *Kluyveromyces marxianus*

SEQUENCE: 102 MKLAYSLLLP LAGVSA 16

SEQ ID NO: 103 moltype = AA length = 19  
 FEATURE Location/Qualifiers  
 source 1..19  
 mol\_type = protein  
 organism = *Saccharomyces cerevisiae*

SEQUENCE: 103 MLLQAFLFLL AGFAAKISA 19

SEQ ID NO: 104 moltype = AA length = 26  
 FEATURE Location/Qualifiers  
 source 1..26  
 mol\_type = protein  
 organism = *Saccharomyces cerevisiae*

SEQUENCE: 104

-continued

MTKPTQVLVR SVSILFFITL LHLVVA	26
SEQ ID NO: 105 moltype = AA length = 26 FEATURE Location/Qualifiers source 1..26 mol_type = protein organism = Gallus gallus	
SEQUENCE: 105 MLGKNPMPMCL VLVLLGLTAL LGICQG	26
SEQ ID NO: 106 moltype = AA length = 18 FEATURE Location/Qualifiers source 1..18 mol_type = protein organism = Homo sapiens	
SEQUENCE: 106 MKWVTIFISLL PLFSSAYS	18
SEQ ID NO: 107 moltype = AA length = 566 FEATURE Location/Qualifiers REGION 1..566 note = ABAB binding protein expressed from plasmid source 1..566 mol_type = protein organism = synthetic construct	
SEQUENCE: 107 MRPSPSIPTAV LFAASSALAM QVQLVETGGG LVOPGGSLRL SCAASGFTLD YSSIGWFRQA 60 PGKEREGVSC ISSSGDSTKY ADSVKGRFTT SRDNNAKNTVY LQMNSLKPPD TAVYYCAAFR 120 ATMCGVPLS PYGKDDWGKG TLTVTSSGGG GSGGGGSGGG GSQVQLVESG GGLVOPGGSL 180 RLSCEASGFT LDYYGIGWFR QPPGKEREAV SYISASARTI LYADSVKGRF TISRDNNAKNA 240 VYLQMNLSLR EDTAVYYCAR RRFSAASSVNR WLADDYDVWG RGTQVAVSSG GGSGGGGGGG 300 SGGGSQLQLV ETGGGLVQPG GSLRLSCAAS GFTFSDYVMT WVRQAPGKGP EWIATINTDG 360 STMRRDDSTKG RTTISRDNAK NTLYLQMTSL KPDTALYYC ARGRVISASA IRGAVRGPGT 420 QVTVSSGGGG SGGGGSGGGG SQVLTQVESGG GLVQTGGSLR LSCASSGSIA GFETVTWSRQ 480 APGKSLQWVA SMTKTNNIEY SDSVKGRFII SRDNNAKNTVY LQMNSLKPED TGVYFCKGPE 540 LRGQQGIQVTV SSVDMEQKLI SEEDLE	566
SEQ ID NO: 108 moltype = DNA length = 1701 FEATURE Location/Qualifiers misc_feature 1..1701 note = Polynucleotide sequence encoding ABAB binding protein expressed from plasmid source 1..1701 mol_type = other DNA organism = synthetic construct	
SEQUENCE: 108 atgagatttc ctccaatttt tactgttgtt ttattcgcag catcctccgc attagctatg 60 caagtacaat tggttgaaac cgggtgttgtt ttgttcaac cagggtgttag ttttagattta 120 tcttgcgtg catcagggtt tacattggat tattcttcaa taggttgttt cagacaagct 180 cctcgtaaag aaagagaagg tggttcttgcatatccagt ctgggtgactc aactaaat 240 gtctgactccg ttaagggttag attcaactact tcaagagata acgctaaaaa tacagtctac 300 ttgcaatgtg actcattaaa gcccattgtac acacgacttcttatttttgcgctttttaga 360 gccaccatgt gcgggttattt cccattgtct ccttacggta aagatgactg gggtaaaggt 420 actttagtta ctgttccatc cgggtgttgtt gggtccgggt gtgtgttgtt tggttgttgtt 480 ggttctcaag ttcaatgtt agaatccgg tgggttgtttag ttcaacccgg tggttgttta 540 agattatcct gogaqaacag tggtttata tttagattttt acggatcatgg ttgggtttaga 600 caaccacctg gtaaaagaaag agaagctgtc tcttatattt ccgtatgtc aagaactata 660 ttgtacgcg attctgttaaa gggtagatcc acaatttcaaa gagacaatgc caagaacgct 720 gtttatttgc aaatgaactc ttgttgcggaa gaagacaccg cagtttata ctgtgcggaa 780 agaagattttt ctgttccatc agtcaacaga tgggttagcag acgattatga ttgttgttgtt 840 agaggtcac ac aagtccgtt aagttcttgtt ggtggccgg gtttgttgtt tggttgttgtt 900 tctgggttgtt gttcacaattt gcaattttttt gaaacttgggtt gtttgttgtt tcaaccagg 960 ggttcccttgc gattaagggt tgctgtatc ggtttttactt tctctgtatca cgtttatgaca 1020 tgggtcagac aagctccagg ttaaagggtt gaaatggatcc ctacaattaa taccggccgt 1080 tccacaatgtg gagatgacag tacaagggtt agatttcaactt ttcaagagata taacgctaa 1140 aacacattgtt atttacaaat gaccttcttgc aaaccagaag acaccgcatt atattactgt 1200 gccagagggtt gtaatgttccatc cgccagggtt atcagagggtt cagtaagagg tcctgttact 1260 caagtttacatc tctcttcagg tggggccggcc agtggccggcc ggggttctgg cgggttgttgtt 1320 tcacaaggatcc aattttgttcaatc tttttttttt ggtttttttt aacttgggtt ttcatgtt 1380 ttatcttgcg cttccatgtgg ttccatgtca ggtttttttt gttttttttt ctgttacatc gtcaagacaa 1440 gctccaggatcc aatcttgcg atgggttgcgc tcaatgttccatc agactaacaa cgaaatctat 1500 tctgttgcg ttaagggtttagt atttattttt tcaagagata atgtctaaaaa caccgtttat 1560 ttgcaatgtg actcattgtttt gccaaggatc actgttgtttt atctctgtcaaa ggggttgtt 1620 ttaaagggttca aaggatgttca agtaacaggatc tttttttttt acatgttccatc gtcaagacaa 1680 tcccaaggatcc ac tttttttttt tttttttttt tttttttttt tttttttttt aacttggaa 1701	1701
SEQ ID NO: 109 moltype = AA length = 557 FEATURE Location/Qualifiers REGION 1..557	

-continued

note = ABAB binding protein expressed from chromosomal  
integration  
source 1..557  
mol\_type = protein  
organism = synthetic construct

SEQUENCE: 109  
MRFP SIFTAV LFAASSALAM QVQLVETGGG LVQPGGSLRL SCAASGFTLD YSSIGWFRQA 60  
PGKEREGVSC ISSSGDSTKY ADSVKGRFTT SRDNAKNTVY LQMNSLKPDD TAVYYCAAFR 120  
ATMCGVPLS PYGKDWDKGK TLTVTSSGGG GSAGGGGGGG GSQVQLVESG GGLVQPGGSL 180  
RLSCEASGFT DLYYIGWFR QPPGKEREAV SYISASARTI LYADSVKGRF TISRDNAKNA 240  
VYLQMNLSLRK EDTAVYYCAR RRFSAASSVNR WLADDYDVWG RGTQVAVSSG GGSGGGGGGG 300  
SGGGSQLQLV ETGGGLVQPG GSLRLSCAAS GFTFSDYVMT WVRQAPGKGP EWIATINTDG 360  
STMRDSTTKG RFTISRDNAF NTLYQMNLSL KPEDTALYYC ARGRVISASA IRGA VRGPGT 420  
QVTVSSGGGG SGGGGGGGG SQVQLVESG GLVQTGGSLR LSCASSGSIA GFETVWSRQ 480  
APGKSLQWVA SMTKTNEIY SDSVKGRFII SRDNAKNTVY LQMNSLKPED TGVYFCKGPE 540  
LRGQGIQVTV SSVDAAS 557

SEQ ID NO: 110 moltype = DNA length = 1674  
FEATURE Location/Qualifiers  
misc\_feature 1..1674  
note = Polynucleotide sequence encoding ABAB binding  
protein expressed from chromosomal integration  
source 1..1674  
mol\_type = other DNA  
organism = synthetic construct

SEQUENCE: 110  
atgagatttc cttcaatttt tactgctgtt ttattcgcacatccctccgc attagctatg 60  
caagtacaat tgggtgaaac cgggtgggtt ttagttcaac cagggtggtag ttttaggatta 120  
tcttgtgtcatcagggtt tacattgtt tattcttcaac taggttgggtt cagacaagct 180  
cctggtaaagg aaagagaagg tggttcttgatataccgtt ctgggtgactc aactaaatat 240  
gctgactccg ttaagggttag attcaactact tcaagagata acggctaaaa tacagtctac 300  
tttcaaatgaa actcaattaa gccagatgac acagcagtctt attactgtgc cgcttttaga 360  
gccaccatgtt gggtgttattt cccattgtt ctttacggtaa agatgactg gggtaaagg 420  
acttttagttt ctgttcttc cgggtgggtt gggttgggtt gttgtggtag tgggtgggtt 480  
gggtctcaag ttcaatttagt agaatccgggtt ggttgggtttag tccaaccttgg tggtagtta 540  
agatttatctt gogaagcaag tgggtttaca tttagattttt acggatctggg ttgggttaga 600  
caaccacatgtt gtaaaaggaaag agaagctgtc tcttattttt ccgtactgtc aagaactata 660  
tttgtacogat atctgttaaa gggtagatcc acatatttcaaa gagacaatgc caagaacgct 720  
gtttatttttcaaaatgttcaagaga gaagacaccg cagtttata ctgttccac 780  
agaagattttt ctgttcttc agtcaacaga tggtagcag acgattatgt tgggtgggtt 840  
agaggttacac aagtccgtt aagtctgtt ggtgggttccg gtgggtggtag tgggtgggtt 900  
tcctgggtgtt gttcacatgtt cgcgttccgtt gggacccgggg gaggcttagt tcagctccgg 960  
gggtccctgtt gactctgtt tgcacccctt ggatccaccc tcaatgtactc cgtgtgacc 1020  
tgggtccccc aagctccagg gaaggggccgtt ggtggatccg caactattaa tactgtatggg 1080  
agcacaatgtc cgcacgactc cacaaggcgcggtt cttccacca tctccagaga caacgcacaa 1140  
aacactctgtt atctgttcaatgtt gaaacgtctt gaaacccggggg acactgtctt gtatctgt 1200  
gcaagaggccc ggggtgtatctt tgcctccgtt atcagaggccg cagtcacagg ccctggaaacc 1260  
caggttccggc tttcgacggg tggggccggg tgggtggccgg ggggttccgg ggggttgggt 1320  
tcacaagtcc aattgtttaga atctgggtgtt ggttttagttt aaactgggtt ttcattttaga 1380  
ttatctctgtt ctccatgtt tttccatgtt gtttccatgtt ctgttacatgttcaagacaa 1440  
gtctccatgtt aatcttctgtt atgggtcgcc tcaatgtacttca gactaaacaa cggaaatctt 1500  
tctgttccatgtt ttaagggttag atttttttcaagagata atgctaaaaa caccgtttat 1560  
tttcaatgttcaacttctgtt gtttccatgtt gtttccatgtt ctgttacatgttcaagacaa 1620  
tttcaatgttcaacttctgtt gtttccatgtt gtttccatgtt ctgttacatgttcaagacaa 1674

SEQ ID NO: 111 moltype = AA length = 549  
FEATURE Location/Qualifiers  
REGION 1..549  
note = ABA binding agent  
source 1..549  
mol\_type = protein  
organism = synthetic construct

SEQUENCE: 111  
QVQLVETGGL VQPQGSLRLS CAASGFTLDY SSIGWFRQAP GKEREGVSCI SSSGDSTKYA 60  
DSVKGRFTTS RDNAKNTVYL QMNSLKPDT AVYYCAAFRA TMCGVFLSP YGKDDWGKGT 120  
LVITVSSEPKT PKPQPTSGGG GSAGGGGGGG GSQVQLVESG GGLVQPGGSL RLSCASSGS 180  
AGFETVWSR QAPGKSLQWV ASMTKTNEIY YSDSVKGRFI ISRDNAKNTVY LQMNSLKP 240  
DTGVYFCKGP ELRGQGQIQT VSSEPKTPKP QTSIAAGGGG SGAGGGGGGG SLQAMAASQ 300  
VQLVESGGL VQTGGSLRLS CASSGSIAFG ETVTWSRQAP GKSLOVWASM TKTNEIYSD 360  
SVKGRFIIISR DNAKNTVYLQ MNSLKPDTG VYFCKGPRLR QGQIQTVSS GGGGGGGGG 420  
GGGGSWAAQL QLVETGGGLV QPGGSLRLSC AASGFTFSYD VMTWVRQAPG KGPEWIATIN 480  
TDGSTMRDDSTKGRFTISRD NAKNTLYLQM TSLKPDTAL YYCARGRVIS ASAIRGAVRG 540  
PGTQVTVSS 549

SEQ ID NO: 112 moltype = AA length = 15  
FEATURE Location/Qualifiers  
REGION 1..15  
note = D7 tag  
source 1..15

-continued

---

mol\_type = protein  
 organism = synthetic construct

**SEQUENCE: 112**  
 APTKAKRRVV QREKT

15

SEQ ID NO: 113      moltype = AA length = 553  
 FEATURE                Location/Qualifiers  
 REGION                1..553  
 note = His-hABAB-D7 binding agent  
 source                1..553  
 mol\_type = protein  
 organism = synthetic construct

**SEQUENCE: 113**  
 HHHHHHQVQL VETGGGLVQP GGSLRLSCAA SGFTLDYSSI GWFRQAPGKE REGVSCISSS 60  
 GDESTKYADSV KGRFTTSRDN AKNTVYLQMN SLKPDDTAVY YCAAFRATMC GVFPPLSPYKGK 120  
 DDWGKGTLVT VSSGGGGSGG GGSGGGGSQV QLVESSGGLV QPGGSLRLSC EASGFTLDYY 180  
 GIGWFROPPG KEREAVSYI ASARTYAD SVKGRFTISR DNAKNAYVLQ MNSLKREDTA 240  
 VYYCARRRFS ASSVNRWLAQ DYDVWGRGTQ VAVSSGGGGG GSAGGGGGGG SQLQVETGG 300  
 GLVQPGGSLR LSCKAASGFTF SDYVMTWVRQ APGKGPEWIA TINTDGSTM R DDSTKGRFTI 360  
 SRDANKNTLY LQMITSKPED TALYYCARGR VISASAIRGA VRGPGTQVTV SSGGGGGGG 420  
 GSGGGGSQVQ LVESSGGLVQ TGGSRLSCA SSGSIAGFET VTWSRQAPGK SLQNVASMTK 480  
 TNNEIYSDSV KGRFIISRDN AKNTVYLQMN SLKPEDTGVY FCKGPRLRGQ GIQTVSSAP 540  
 TKAKRRVVQR EKT

553

SEQ ID NO: 114      moltype = DNA length = 1665  
 FEATURE                Location/Qualifiers  
 misc\_feature          1..1665  
 note = Polynucleotide sequence encoding His-hABAB-D7  
 binding agent  
 source                1..1665  
 mol\_type = other DNA  
 organism = synthetic construct

**SEQUENCE: 114**  
 catcatcatc atcatcatca ggtacagctg gtggagacgg gggggggg ggtacaacca 60  
 ggccgggtcac tgaggcttc ctgtgcccga tctgggtca cactggatta ttctgtccata 120  
 ggggtggttc ggcaggctcc tggcaaaagag cgtgaggggg tctcatgtat tagtagtagt 180  
 ggtgatagca caaagtacgc cgattccgta aaggggccgtt ttacaacctc cagggataat 240  
 gctaagaaca ccgtatatact ccagatggaaat ttctctgaaccc ccgacgtatcc ggccgtataat 300  
 tactgtggc ctttcaggcc gactatgtgc ggcgtgttcc ctctgagcccc ttacggcaag 360  
 gacgactggg gcaaggggac ccttggtgacc gtatcctca ggcgtggggg gtctggggg 420  
 ggagggtctag ggggtgggagg cagccagggtg caactgggtt aatctggggg aggcttgta 480  
 caaacctgggg gatcccttag actctcttg gaggccctccg gattcacctt ggactactat 540  
 ggcatcggtt gtttccggca gccccccagg aaggagccggg aggccgttcc atacattatgt 600  
 gccagtgtccc ggaccataact gtacccggac tctgtgaagg gacgccttac catctctagg 660  
 gacaatgcca aaaatgtctgt gtacccgtcaa atgaacacgg tcaaggggg ggtataccca 720  
 gtgtactact ggcgcggacg ggcgttccatc gtttcttagc tgaatagatg gtcggccac 780  
 gactacgacg tggggggacg gggcacaacag gtcgtgttgc ctteccgttgg cggaaaggccg 840  
 gggggccggc ggggtggggag cgggtggggc agccaaactgc agctggtaga gacggggggc 900  
 ggcttagttc agccctgggg gtcctctcaga ctgtcatcgcc ctgcctctgg cttaaccttc 960  
 atgtgactactc tgatgacatg ggtccggccaa gtcgtggggg aggccgttgc gtggatcgct 1020  
 actatataataa cagatggccg cacaatggccg gacgacttca caaaggggggc gtccacatt 1080  
 tccagagaca acgccaagaa tactctgtac ctccagatga ccagtcgtaa accccaggac 1140  
 actgtctgt actattgtc aagaggccgg gtcgtctctg ctteccgttat cagaggccca 1200  
 gtaaaggggcc ctggaaacaca ggttaaccgtt tcatccgggg gaggggggttc agggcggtgg 1260  
 ggatctggccg ggggtggatc ccaagttcag ctggtcgaat cccggggggc actggtcac 1320  
 acaggggggct cccttgggtt ctccgtca tttccggaa gcatccggg ctteccggac 1380  
 gtgaccctggt ctgcggcggc tcccgaaag tctctgtcgtt gggctgttc catgactaa 1440  
 actaaacaaacg agatctactc tgacttcgtt aaaggccgtt tcattatcc tagagataac 1500  
 gctaaaaaaa cagtgtatctt gcatgttcaat gatgtccaaac ctgaagacac aggccgttat 1560  
 ttctgttaagg gtcctgtatctt gggggccggc ggcattccagg taacatgttc gatgtccct 1620  
 aaaaaagcca aacggagatg ggtccagaga gagaagaccc aataaa

1665

SEQ ID NO: 115      moltype = DNA length = 7257  
 FEATURE                Location/Qualifiers  
 misc\_feature          1..7257  
 note = Plasmid pCEV-G4-Km-TEF-AT-RSyABAB

source                1..7257  
 mol\_type = other DNA  
 organism = synthetic construct

**SEQUENCE: 115**  
 cacacaccat agcttcaaaa tgtttctact ctttttttac ttttccagat tttctcgac 60  
 tccgcgcata ggcgttccac ttcggaaacac ccaagcacaag catactaaat ttccccttt 120  
 ttttccctta ggggtgtcggtt aattaccgtt actaaagggtt tggaaaaaaa aaaagagacc 180  
 ggcctcggtt ttttttttcg tcggaaaaagg caataaaaaat ttatccatcg ttctttttc 240  
 ttggaaaattt ttttttttgc ttccatcgac ctcccatgg tattttatgtt 300  
 aataaaacggt cttcaatttc tcaagttca gtttcatttt ttttgcgttca ttacaacttt 360  
 ttttacttct tgctcatttag aaagaaaagca tagcaatctt atctaaaggtt taattacaag 420  
 gatccatgag atttccttca attttactgt ctgtttttatc cgcacatcc tccgcattag 480  
 ctatgcgtt acaatgggtt gaaaccgggtt gttttttatgtt tcaaccaggat ggttagttca

-continued

---

gattatcttgcgtatca	ggttttacat	tggatttattc	ttcaataggt	tggttcagac	600
aagctccctgg	taaagaaaaga	gaagggtttt	cggcatatc	cagttctgtt	660
aatatgtgtaa	ctccgttaag	ggtagatca	ctacttcaag	agataacgct	720
tctacttgc	aatgaactca	ttaaaaggccag	atgacacagc	agtcttattac	780
tttagagccac	catgtcggtt	gttcccccatt	tgctccctta	cggtaaaaat	840
aaggtagttt	atgttactgtc	teatccgggt	tggtgtggtt	cggtgggtt	900
gttgtgtgtt	tcaagttcaa	tttagtagaa	ccgggtgggtt	tttagtccaa	960
gtttaagatt	atccctcgaa	gcaagtgggtt	ttacattaga	ttattacggt	1020
ttagacaacc	acctggtaa	gaaagagaag	ctgtcttta	tatttccgtt	1080
ctatattgtt	cgcagatct	cgaaagggtt	gtttccacaat	ttcaagagac	1140
acgctgttta	tttgcataat	aactttgtt	agagagaaga	caccgcagg	1200
ccagaagaag	attttctgtt	tcttcgtca	acagatgggtt	agcagacgt	1260
ggggtagagg	tacacaatgc	gcccgttaat	ctgtgtgggtt	ttccgggtgt	1320
gtgtttctgg	ttgtgtgttca	cagetggc	tggtggagac	cggggggaggc	1380
ctggggggc	cttgcagact	tcctgtcag	cctctggatt	cacccctagt	1440
tgacctgggt	ccgccaagct	ccagggaagg	ggcctgagt	gatecgaact	1500
atgggagcac	aatgcgcgc	gactccacaa	aggggccgg	caccatccc	1560
ccaagaacac	tctgtatgt	caaataaca	gtctgaaac	cgaggacact	1620
actctgtca	aggccgggtt	atctctgtt	ccgcgtatc	aggcgcagtc	1680
gaacccagg	caccgtctcg	agccgtggcg	ggggtagtgg	cgccggcggt	1740
gtgttccaca	agtccaaat	gttagatctt	gtgtgtggttt	agttcaact	1800
tgagatttct	ctgcgttcc	agtgggttca	ttgcagggtt	cgaaaactt	1860
gacaagctcc	aggtaaatct	ttgcaatggg	tcgcctcaat	gaccagact	1920
tctatttgc	ttcagttaa	ggtagattca	ttatccaa	agataatgt	1980
tttatttgc	aatgaactca	ttgaaaggccag	aaagataactt	tggttactt	2040
ctgaattaat	agggtcaagg	attcaatgtt	cagtttcttc	agtcgcacat	2100
tgatttccg	agaagacctc	gtagatgtt	ggtaccgcgg	ctagctaaga	2160
ccgaaaagga	aggaggtaga	caacctgaag	tctaggtccc	tattttttt	2220
tgttagtatt	aaaacgttta	tttatatttt	aaatttttt	tttttttctg	2280
tgtacgcat	taacatata	ctgaaaacat	tggttggagaa	gggtttggga	2340
tccagctgca	ttatgaatc	ggccaaacgc	cggggggagg	cggtttcggt	2400
cttccgttcc	ctcgctca	gactcgctc	gctccgtcg	tccgtcg	2460
cagctca	aaaggccgtt	atacggttat	ccacagaatc	aggggataac	2520
acatgtgac	ggggccgg	caaaaaggca	ggaacgttta	aaaggccgc	2580
tttccatgt	gtccggcccc	cctgcagac	atcacaaaa	tcgacgtca	2640
ggcgaaacc	gacaggacta	taaagatacc	aggcgtttcc	ccctggaa	2700
gctctccgt	tcgcacccct	ccgcttacc	gatacctgtc	cgcccttctc	2760
gegtggcggt	tttcatcata	tcacggttca	ggtagtctt	ttccgtgtt	2820
ccaaagctgg	ctgtgtgcac	gaacccccc	ttcagccgc	ccgctgcgc	2880
actatgtct	tgagtccaa	ccgcgttac	acgacttatt	gccactggca	2940
gtaacaggat	tagcagagc	aggtatgtt	ggccgtgtac	agagttctt	3000
ctaaactacgg	ctacactaga	aggacatgt	ttgttatctt	cgctctgc	3060
ccttcggaaa	aaaggtgg	agcttctgtt	ccgcgttac	ttccggct	3120
gtttttttgt	ttcaagcagc	cagattacgc	gcagaaaaaa	aggatctt	3180
tgatctttc	tacgggttct	gacgttgcgt	ggaacgaaaa	ctcacgtt	3240
tcatgaggat	atcttcac	atgtccctt	aaataaaaaa	tgaagttt	3300
aatcaatcta	aaatataat	gatgtttttt	ggtctgcac	tttccatgt	3360
aggcacctat	ctcagcgatc	tgtcttattt	gttccatcat	agttgcgt	3420
tgttagataac	tacgatacgg	ggggcttac	catctggccc	cgatgtc	3480
gagaccaccc	ctccacccgt	ccagatattt	cagcaataaa	ccagccaggc	3540
agccgagaag	ttgtctgc	actttatcc	cctccatcca	gtcttataat	3600
aaqcttagat	aaatgttgc	ccgttata	gttgcgtt	cggtttttt	3660
gcatecggtt	gtcagcgtc	tcgttggta	tggcttcat	cgatccgg	3720
caaggcgagt	tacatgtatc	cccatgttgc	gcaaaaaaa	ggtttagtcc	3780
cgatcggtt	cagaatgtt	ttggcgcgc	ttgttgcgc	ttgtacttgc	3840
ataatttctt	tactgtcat	ccatccgtt	gatgtttt	tgtactgtt	3900
ccaaagtctt	ctgagaatag	tgtatgcgc	gaccgatgtt	ctttgc	3960
gggataatcc	ccgcgcacat	agcagaactt	taaaatgtt	catcttgc	4020
cggggcgaaa	actctcaagg	atcttaccgc	ttgttgcgc	tttccgc	4080
gtgcacccaa	ctgtatcttca	gcatctttt	catttcac	cggttttgg	4140
caggaaggca	aatgcgc	aaaaaggaa	taagggcgc	acggaaatgt	4200
tactcttctt	tttcaat	tatgttgc	tttttgcgg	tttgcgtt	4260
acatatttgc	atgttatttgc	aaaatgggg	ttccgcgc	caca	4320
aaatgtccacc	tgaacaa	gttgcgtt	catttttgc	taacc	4380
gctgtatattt	ttcaacaaa	aatctgttgc	ttgttgcgc	ttgttgcgc	4440
aaagcgctat	tttcaacac	aaatgttgc	gttgcgtt	tttgcgtt	4500
cgagagccgt	attttccaa	aaaaagaatc	tgatgttgc	tttttgcgg	4560
acgcgcagag	gttattttac	caacaaagaa	tctataattt	tttttgcgtt	4620
catcccgaga	gogcttattt	tctaaat	catcttgcgt	tttttgcgtt	4680
cgctctata	tgcgttgc	tgataactt	ttgcactgt	ggtccgtt	4740
aggctacttt	ttgttgcattt	tttcttccca	taaaaaaa	ctgactccac	4800
tactgttac	tagcgaatgc	ggccgttgc	tttttcaaga	cccgattata	4860
ttctatacc	atgtggattt	cgcatat	gtgaacagaa	agtgtatgc	4920
ttcatttgc	agaaaaattat	gaacgggtt	ttcttatttt	tctctatata	4980
gaaatgttta	cattttcgta	ttgttttgc	ttcactctat	gaatgttct	5040
tttttgcata	aaggtataat	ctagatata	acataaaaa	tgttaggtt	5100
gcaagttca	ggagcgaaag	gtggatgggt	aggttatata	gggatata	5160
atagcaaaga	gataattttgc	agcaatgtt	gtggaaagcgg	tattcgat	5220
ctcggttac	cccggtcg	ttttggttt	ttgaaagtgc	gttccagag	5280

-continued

---

```

tttcaaaagc gctctgaagt tcctatactt tctagagaat aggaacttcg gaataggAAC 5340
ttcaaaagcg ttccggaaaac gagcgcttcc gaaaatgcAA cgcgagctgc gcacatacAG 5400
ctcaactgttc acgtcgccacc tatactcgCG tgTTGCTGT atatatataat acatgagaAG 5460
aacggcatag tgcggttttA tgcttaaatC cgtacttttA tgcgcttatt tatgttagGA 5520
gaaaggtagt ctatgtacCC ctgtgtatTT atccccattCC atgcggggTA tcgtatGCT 5580
ccttcagcac taccctttAG ctgttctata tgctgcccAC ctcacattGG attagtctCA 5640
tccttcatagt ctatcatTC ctgttgatTT ggatcatGGT agacaacCC taataataCT 5700
tcgtataatg tatgtataAC gaaggtaTTA ggtctagAGA tctgtttAGC ttgcctcGTC 5760
cccgccggT caccggGCCA ggcacatGGA ggcccAGAA acccttcTT acagtcttGA 5820
cgtgcgeAGC tcaggggCAT gtatgtACTC tgccccGTtG atttAGCCa tacatccccA 5880
tgataatacA ttgcatacCA taccatTTA tggtccACG ggcgaAGCA aaaattACGG 5940
cttcctcgCTG cagacCTGCG agcaggGGAA cgcggccCTC acagacGCG tgaattGTC 6000
ccacgcggCG ccctctgtAGA gaaatataAC aggtttagGA ttgcctAGA ggttcttCT 6060
tcataactt ctttgttAAA atacagtTC cacatCACAT ccgaacataA 6120
acaacccatGG gtaaggAAAA gactcACGTT tcgaggGCCG gatTAATTC caacatGGAT 6180
gctgatttat atgggtataA atgggctcGC gataatgtCG ggcaatCAGG tgCgacaATC 6240
tatcgatgtt atgggaAGCC cgtatcgCCa gagttgttTC tgaAAACATGG cAAAGGTAGC 6300
gttgcctaTG atgttacAGA tgatgtGGt agactaaACt ggctgacGGa atttatGCT 6360
cttcgcacCA tcaagcATT tatecgACT cctgtatGtg catgttACT caccatGCG 6420
atccccGGCA aaacacGcATT ccaggtaTTA gaagaatATC ctgattCAGG tgAAAtATt 6480
gttgcgcgcG tggcagtGTT cctgcgcggT tgccatTCG ttccTGTtG taatttGCT 6540
tttaacacGC atcgatgtt ctcgtatCET caggcgcAAat cacatGAA taacGGTTG 6600
gttgcgcgcG tgatgttGA tgacgagCtG atggctGCGC ctgttGAAca agtctggAAA 6660
gaaatgcATA agctttGCC attcTcAccG gattcAGtG tcactcatGG tgatttCTCA 6720
cttgataacc ttatTTtGA cgaggGGAAAt ttaatAGGtG gtattGATGt tgGAcAGtC 6780
ggaatcGcAG accgataccA ggtatGcC attcTatGGA actcCTCtGG tgatTTtCT 6840
ccttcattAC agaaacGGCT ttttcaaaaaA tattGtGatC tataatCtGA tatGatAaaa 6900
ttgcagttc atttgatGtC cgtatgttG ttctaatCAG tactGACAat aaaaAGATtC 6960
ttgttttcaA gaaatGtCA ttGtGatAGt tttttatAt tgtagtGtGt ctatTTtAAt 7020
caaatGtGtAG cgtgttAt attttGtGtGt GcCtGAcAt catGtGcCA gatGcGAgt 7080
taagtGcGA gaaAGAtAAta tcatGcGtCA atcgatGtGt aatGtGtGtC gctatacTG 7140
tgtcgatTCG atactAACGc CGCCatCCAG tgtcgaaACG gagctcTGA gaACCtTA 7200
tataacttCG tataatGtA gctatacGA gttttagGt gatAtcAGAt ccactAG 7257

```

SEQ ID NO: 116 moltype = DNA length = 8202  
 FEATURE Location/Qualifiers  
 misc\_feature 1..8202  
 note = Plasmid pCEV-G4-Km-TEF-AT-yABAB AA6T83N  
 source 1..8202  
 mol\_type = other DNA  
 organism = synthetic construct

SEQUENCE: 116  
 ggagcgcacCT catgtatac ctgagaaAGC aacctgacCT acaggAAAGA gttactcaAG 60  
 aataagaattt ttcgtttAA aacctaAGAG tcactttAA atttGtatac acttttttt 120  
 ttataactt atttaataAAt aaaaatcata aatcataAGA aattcgttta tttagaAGtG 180  
 tcaacaacGT atctaccaAC gatttGACCC ttTccatCtC ttTcgtaat ttctGcAAG 240  
 gtagacaAGC cgacaACCTT gatttggAGC ttgaccaAAc ctctggGAA gaatttGtAA 300  
 ttaaagatGtC agattttACtC tgcgtcatCC ttgtatCtCA tcgataACTG ttTTTgtt 360  
 aaaaattttAA aaactttttG ttTTTgttG tattttttGt tcttagAAAGa gacaAGtG 420  
 gcttggTTGt ttttGatGtt ttatTTtAt acaatAGtGt cAAatGAAAG atagattcGA 480  
 acatttGgAA gtttggCAt atatGtCtC tattttatC ttTTTTtGt cagtGtAGt 540  
 atatTTtGtA ttTtcTtCtC ttTcatttCtC ttGatGtCC tcaatGtGtA gggGatCtG 600  
 aaagatGtGt gggattttC ttGaaAGAC gacttttGtC catGtAattt ttccTttGtt 660  
 cctctGaaaA ttatccAGCA gaAGcaAAtG taaaAGatGt ACtCAGAAG aacacGcAG 720  
 ggcccGaaAt ttGtCtAcG agaaAGatGt ggtcataAAAGtGt ttttttttCtGggAAA 780  
 aattttGcGt tGccTTtCtG gagaattttC tcaGatTTtC tGtGtGcAc tGcAtGtC 840  
 tctGagaAGt gtggGatTC ttccaccAGt tgTTccTtCtC aaaaaaaaaa agatttCtA 900  
 ccccgacAA attcGtGat ACCCTtCtC tccacGggGc CggccGcAcA caccatAGt 960  
 tcaaaatGtT tcaactCtCtt ttTtacttCt ccGatTTtC tccGactCtG cGcAtGccG 1020  
 taccacttCA aAACACCCAd GcAcGcAtA ctaaaatCtC ccttttttCtCtGAGGt 1080  
 gtcgttaattt accCGtActA aaggTTGGA aaaaAGAAAG gagaccGcT cGtttCttt 1140  
 ttctcgtGA aaaaGGCAAt aaaaatTTtCtCtC ttttttttGtA aaaaatTTtCtCtC 1200  
 ttGtGatTTtTtCtCtCtC tGtGacCtCtC tattGatAt taatGttaAt aacGGtCtC 1260  
 aatttCtCAAt gttcGatTTtCtCtCtC ttttttttCtC ttttttttCtCtCtCtCtC 1320  
 cattGaaAG aaaaGcAtAGC aatctAatCt aatGtttAtA tacaAGGAtC tGtGAGAtt 1380  
 ccttcaAttT ttactGtCtG ttttattCtCA gcatCtCCG cattAGtCtA gcaAGtAcA 1440  
 ttGtGtGAAAG ccGGtGtGtG ccGatGtGtA gtttGatGtGtGtGtGtGtGtGtGtGt 1500  
 gcatcaggGt ttacatGtGA ttttattCtCA ataggGtGtGtGtGtGtGtGtGtGtGtGt 1560  
 gaaAGAGAAG gttGtGtGtG ttttattCtCA ataggGtGtGtGtGtGtGtGtGtGtGtGt 1620  
 gttaaGGtGA gattcaACTAC ttCAAGAGAt aacGtCtAAAt atacAGtCtA cttGcAAAtG 1680  
 aactcattAA aGCCAGAtG CAcAGCAGtC tattactGtG cGcGtttGtGtGtGtGtGtGt 1740  
 tGcGgtGtGtG tcccatGtGtG tcccttacGGt aaAGatGtAct gggGtAAAGtGtGtGt 1800  
 actGtCtCtC tccGgtGtGtG tggGtCtCtC tccGactCtG tcaAGACtG tccGtGtGtGt 1860  
 gttcaattAG tGAGAtCtGtG tggGtGtGtG tggGtGtGtGtGtGtGtGtGtGtGtGt 1920  
 tGcGAGcAA gttGtGtGtG ttttattCtCA tGtGtGtGtGtGtGtGtGtGtGtGtGt 1980  
 ggttaaAGAAAG gagaAGt 2040  
 gattctGtAA AgggtagAttCtCA aGAGAtCtGtGtGtGtGtGtGtGtGtGtGtGtGtGt 2100  
 caaatGtAAAct ttttGt 2160  
 tctGtCtCtCtC tGt 2220

-continued

---

caagtcgcgg	taagttctgg	tgggtggg	ttctgggtt	2280
ggttcacaat	tgcaattagt	agaaactgg	gggtttgg	2340
agattaagg	gtgctgcac	tggtttact	tctctgatt	2400
caagtcgcgg	gtaaagg	tgaatgg	gctacaatta	2460
agagatgaca	gtaccaagg	tagatct	attcaagag	2520
tacttacaaa	tgaactt	gaaaagg	ataacgctaa	2580
agagtcata	ccgcagg	tatcagag	gacaccgcat	2640
gtcttccag	gtggcg	tagtgg	gggggtct	2700
caatttgtt	aatctgtt	tgggttagt	caaactgg	2760
gcttcagg	gttccat	agggttgc	gttcattg	2820
aatcttgc	aatgggtc	ctcaat	aaactaaca	2880
gttaagggtt	gattcatt	ttcaagag	aatgctaa	2940
aactcattt	agccaa	tacttgc	agggttgc	3000
caaggat	aagaat	tcttcag	gacatgg	3060
gacctcg	aacttgg	cccg	ggtagt	3120
gttagaca	ctgaat	ggte	ttttttt	3180
acgttattt	tatccaat	ttttttt	tttgcata	3240
attatactg	aaccc	tgagaagg	gacgcgtt	3300
tgaatcgcc	aac	gagagg	ttcgat	3360
ctca	ctgc	ggte	ggcg	3420
gcccgtat	gttattcc	agaat	cggat	3480
ggccag	ccccc	aggccg	ccgtt	3540
cgccccctt	acgacat	aaaaat	cgat	3600
ggactataa	gatacc	gttcccc	ggaag	3660
acccgtcgc	ttaccg	cctgt	cggtt	3720
catagctc	gtctgt	tctcag	gttgc	3780
gtgcac	cccccc	ccgc	ccgc	3840
tccaa	taag	cttat	ctgg	3900
agagcgg	atgt	tgcata	ttctt	3960
actaga	cgtat	tatctgc	ctgt	4020
gttggat	cttgc	tttgc	cggtt	4080
aac	ttac	tttgc	gggtt	4140
gggtct	cccg	tttgc	tttgc	4200
aaaaggat	tccat	ttttt	tttgc	4260
atata	aaat	tttgc	tttgc	4320
gcat	tat	tttgc	tttgc	4380
atacgg	gttatt	tttgc	tttgc	4440
cccg	cccg	tttgc	tttgc	4500
cctgc	attt	tttgc	tttgc	4560
agttcg	ttat	tttgc	tttgc	4620
cgctcg	ttgg	tttgc	tttgc	4680
tgtat	tttgc	tttgc	tttgc	4740
agta	tttgc	tttgc	tttgc	4800
gtcat	tttgc	tttgc	tttgc	4860
gaat	tttgc	tttgc	tttgc	4920
ccacat	tttgc	tttgc	tttgc	4980
tcaaggat	tacc	tttgc	tttgc	5040
tctt	tttgc	tttgc	tttgc	5100
gccg	tttgc	tttgc	tttgc	5160
caat	tttgc	tttgc	tttgc	5220
attt	tttgc	tttgc	tttgc	5280
cga	tttgc	tttgc	tttgc	5340
aaca	tttgc	tttgc	tttgc	5400
ccaa	tttgc	tttgc	tttgc	5460
tttca	tttgc	tttgc	tttgc	5520
tttacc	tttgc	tttgc	tttgc	5580
tat	tttgc	tttgc	tttgc	5640
gtct	tttgc	tttgc	tttgc	5700
tctt	tttgc	tttgc	tttgc	5760
gaag	tttgc	tttgc	tttgc	5820
ggat	tttgc	tttgc	tttgc	5880
attt	tttgc	tttgc	tttgc	5940
ttcg	tttgc	tttgc	tttgc	6000
gtta	tttgc	tttgc	tttgc	6060
cgaa	tttgc	tttgc	tttgc	6120
cttt	tttgc	tttgc	tttgc	6180
gtcg	tttgc	tttgc	tttgc	6240
tga	tttgc	tttgc	tttgc	6300
gaaa	tttgc	tttgc	tttgc	6360
cgc	tttgc	tttgc	tttgc	6420
ctat	tttgc	tttgc	tttgc	6480
tg	tttgc	tttgc	tttgc	6540
ctt	tttgc	tttgc	tttgc	6600
catt	tttgc	tttgc	tttgc	6660
ctat	tttgc	tttgc	tttgc	6720
cggc	tttgc	tttgc	tttgc	6780
gggc	tttgc	tttgc	tttgc	6840
catcc	tttgc	tttgc	tttgc	6900
cctgc	tttgc	tttgc	tttgc	6960

-continued

```
tgttagaaaa tataaaaggt taggattgc cactgagggt cttcttcat atacttcctt 7020
ttaaaatctt gctaggatac agttctaca tcacatccga acataaaacaa ccatgggtaa 7080
ggaaaagact cacgttcga ggccgcgatt aaattccaac atggatgcgt atttatatgg 7140
gtataaatgg gtcgcgata atgtcggtca atcagggtgcg acaatctatc gattgtatgg 7200
gaagccccat gogccagact tggttctgaa acatggcaaa ggtagegttg ccaatgtatgg 7260
tacagatgg atggcagac taaaatgggt gacggattt atgccttc cgaccatcaa 7320
gcattttatc cgtacttcgt atgatgcgt gttactcacc actgcgtatcc cggcaaaac 7380
agcattcccg gtattagaag aatatcttgc ttcaagggtaa aatattgttg atgcgttgc 7440
agtgttccgtc cgccgggtgc attcgatcc tggttgcata tgcgttgcata acaggatcg 7500
cgttatttcgt ctgcgtatcgc caaataatcgt aatgataataa ggtttgggtg atgcgtatgg 7560
tttgcgtatc gacgtatcgt gctggccgt tgaacaatgc tggaaagaaa tgcataatgg 7620
tttgcgtatc tcaacggatt cagtcgtcgtc tcatgggtat ttcttcactt ataaccttat 7680
tttgcgtatc gggaaatattaa taggttgcata tgatgttgc ctagtgcgttgc tgcagaccg 7740
ataccaggat ttgcgtatcc tgcgtatcgt cttccgttgc tttttccattt cattacagaa 7800
acggctttt caaaaatatg gtattgtataa tccgtatgc aataaaatgc agtttcattt 7860
gatgtctcgat gagtttttctt aatcgtatcgt gacaataaaa agattttgtt tttcaagaac 7920
tttgcgtatc tatagtttt tttatattgtt gttgttctat ttaatcaaa ttttagcgttgc 7980
atttatattt ttttcgtatcgt cgcacatcgt tgcgtatcgt ctagtgcgttgc gtttcgtatc 8040
gtataatcat gctgtatcgt tgcgtatcgt ctagtgcgttgc gtttcgtatc 8100
taacggccgc atccagtgtc gaaacgago tctcgagaaac ccttaataata acttcgtata 8160
atgtatgtata tgcgtatcgt ttaggtatcgt tgcgtatcgt 8202
```

SEQ ID NO: 117 moltype = AA length = 538

FEATURE Location/Qualifiers  
 REGION 1..538  
 note = AT-yABAB hAA6 T83N binding agent  
 source 1..538  
 mol\_type = protein  
 organism = synthetic construct

SEQUENCE: 117

```
MQVQLVETGG GLVQPGGSLR LSCAASGFTL DYSSIGWFRQ APGKEREGRVS CISSSGDSTK 60
YADSVKGRFT TSRDNAKNTV YLQMNSLKPQ DTAVYYCAAF RATMCGVFPL SPYGKDDWGK 120
GTLVTVSSGG CGGGGGGGGG GGSQVQLVES GGGLVQPGGS LRLSCEASGF TLDYYGIGWF 180
RQPPGKEREAA VSYISASART ILYADSVKGR FTISRDNAKVN AVYLQMNSLK REDTAVYYCA 240
RRRFSASSVN RWLADDYDVW GRGTQVAVSS GGGGGGGGG GSQQGSQLQL VETGGGLVQP 300
GGSLRLSCAA SGFTFSVDYVM TWVRQAPGKG PEWIATINTD GSTMRDDSTK GRFTISRDNA 360
KNFLYLMQNS LKPEDTALYY CARGRVISAS AIRGAVRGPG TQTVTSSGGG GGGGGGGGG 420
GSQVQLVESG GGLVQTGGSL RLSCASSGSI AGFETVTVWSR QAPGKSLQWV ASMTKTNNIEI 480
YSDSVKGRFI ISRDNAKNTV YLQMNSLKPQ DTGVYFCKGP ELRGQGIQVT VSSVDAAS 538
```

The invention claimed is:

1. A method of treating a *Clostridium difficile* (*C. difficile*) infection in a subject, comprising administering an effective amount of an engineered strain of *Saccharomyces boulardii* yeast to the subject, wherein the engineered strain of *Saccharomyces boulardii* yeast produces a tetra-specific, tetrameric ABAB binding agent comprising: (i) a first, a second, a third, and a fourth linked  $V_H$ H peptide monomer each independently having binding specificity for an epitope of *C. difficile* toxin A (TcdA) or *C. difficile* toxin B (TcdB), and (ii) an amino acid sequence of SEQ ID NO: 109 or an amino acid sequence that is at least 95% identical to SEQ ID NO: 109.

2. The method of claim 1, wherein two of the monomers have binding specificity for epitopes of TcdA and two of the monomers have binding specificity for epitopes of TcdB.

3. The method of claim 1, wherein the monomers independently have binding specificity for an epitope in the glucosyltransferase domain, cysteine protease domain, translocation domain, or receptor binding domain of TcdA or TcdB.

4. The method of claim 1, wherein the first  $V_H$ H peptide monomer comprises an amino acid sequence of SEQ ID NO: 7, the second  $V_H$ H peptide monomer comprises an amino acid sequence of SEQ ID NO: 1, the third  $V_H$ H peptide monomer comprises an amino acid sequence of SEQ ID NO:

5, and the fourth  $V_H$ H peptide monomer comprises an amino acid sequence of SEQ ID NO: 3.

5. The method of claim 1, wherein the ABAB binding agent comprises an amino acid sequence of SEQ ID NO 19.

6. The method of claim 1, wherein the ABAB binding agent further comprises an N-terminal secretion signal selected from SEQ ID NO:99 and SEQ ID NO: 103.

7. The method of claim 1, wherein the ABAB binding agent comprises an amino acid sequence of SEQ ID NO: 107.

8. The method of claim 1, wherein the ABAB binding agent comprises the amino acid sequence of SEQ ID NO: 109.

9. The method of claim 1, wherein the engineered strain of *Saccharomyces boulardii* yeast is administered in an amount between 10 µg/kg and 100 mg/kg per body weight of the subject.

10. The method of claim 1, wherein the engineered strain of *Saccharomyces boulardii* yeast is administered to the subject orally, nasally or rectally.

11. The method of claim 1, wherein the engineered strain of *Saccharomyces boulardii* yeast is in a pharmaceutical formulation comprising a pharmaceutically acceptable carrier or diluent.

12. The method of claim 1 further comprising administering an antibiotic to the subject.

\* \* \* \* \*