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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
8 December 2011 (08.12.2011)

PCT

(10) International Publication Number
WO 2011/152995 A1

(51) International Patent Classification:
A61K 38/00 (2006.01)

(21) International Application Number:
PCT/US2011/037049

(22) International Filing Date:
18 May 2011 (18.05.2011)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/351,404 4 June 2010 (04.06.2010) US

(71) Applicant *(for all designated States except US)*: **BOLT BIOPHARMA, INC.** [US/US]; 19223 - 55th Avenue N.E., Lake Forest Park, Washington 98155 (US).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: **ALLISON, Daniel Scott** [US/US]; 19223 - 55 Avenue, N.E., Lake Forest Park, Washington 98155 (US).

(74) Agents: **SLEATH, Janet** et al.; Speckman Law Group PLLC, 2014B East Union Street, Seattle, Washington 98122 (US).

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

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

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

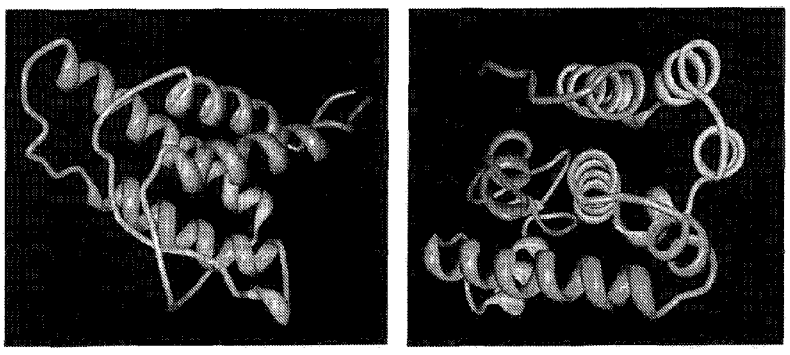
Published:

— *with international search report (Art. 21(3))*

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

(54) Title: SYNTHETIC PROTEINS CONTAINING T CELL EPITOPES FOR USE AS IMMUNOGENIC COMPOSITIONS

Fig. 2: Four-helix bundle proteins – two examples



IL-4 (PDB 1ITL) T4 lysozyme (PDB 1LYD)

(57) Abstract: Synthetic protein domains containing one or more T cell epitopes are provided, together with immunogenic compositions including such domains and methods for the treatment of infectious diseases and/or cancer using such immunogenic compositions. The synthetic proteins contain CD4 and/or CD8 T cell epitopes from the proteome of an infectious disease organism or from a tumor-associated antigen, together with an artificial, non-human-like polypeptide scaffold sequence which, when expressed in a recombinant expression system, generate well-folded, stable, and non-aggregated polypeptides.

WO 2011/152995 A1

SYNTHETIC PROTEINS CONTAINING T CELL EPITOPES FOR USE AS IMMUNOGENIC COMPOSITIONS

TECHNICAL FIELD OF THE INVENTION

The present invention relates to synthetic protein domains containing one or more T cell epitopes and their use in immunogenic compositions for the treatment of disorders including infectious diseases and cancer.

BACKGROUND OF THE INVENTION

Immune responses that occur in response to infection or other immune challenge can be divided into two main types - humoral immunity and cell-mediated immunity. Humoral immunity involves the secretion of antibodies by B cells in response to antigens found on the surfaces of invading microbes, such as viruses or bacteria. Cell-mediated immunity involves the activation of antigen specific cytotoxic T-lymphocytes (T cells), macrophages and natural killer cells, and the release of various cytokines in response to an antigen.

Protein-based vaccines, or immunogenic compositions, for use in the treatment and/or prevention of disorders such as infectious diseases have traditionally been focused on the generation of humoral immunity. Since many epitopes recognized by antibodies are conformational, not linear, vaccines of this type are usually composed of full-length recombinant versions of natural polypeptides (or intact domains) from infectious organisms or of tumor-associated antigens (TAA). However, in recent years there has been increased interest in developing vaccines that stimulate cell-mediated immunity, especially for diseases that are either not controlled (e.g., HIV) or incompletely controlled (e.g., influenza) by vaccines that only stimulate humoral immunity. In particular, there has recently been an increased emphasis on the use of specific T cell epitopes in compositions for stimulation of specific CD4+ and/or CD8+ T cells, especially T cell epitopes that are both located within conserved regions of proteins and have been demonstrated to generate strong T cell responses. One advantage of vaccines that carry only well-characterized T cell epitopes from conserved regions of a pathogenic proteome is that "decoy" epitopes, which are usually located in variable regions, are avoided. The above ideas have been discussed, for example, in the context of HIV vaccines (Schneidewind et al., J. Virol., 81:12382 (2007); Rolland et al., PLoS Pathogens, 3:e157 (2007); Yang, PLoS ONE, 4:e7388 (2009)), influenza vaccines (Doherty and Kelso, J. Clin. Invest., 118:3273 (2008); Heiny et al., PLoS ONE, 2(11):e1190

(2007)) and cancer vaccines (Disis et al., J. Immunol., 156:3151 (1996); Kalli et al., Cancer Res. 68:4893 (2008); Pavelko et al., Cancer Res. 69:3114 (2009)).

Although peptides carrying T cell epitopes have been used to stimulate cell-mediated immunity in animal models, they have not been as successful in human trials. This is at least partly due to the susceptibility of peptides to peptidases that cleave peptide substrates (see, e.g., Falciani et al., Chem. Biol. Drug Des. 69:216 (2007)). On the other hand, polyepitope proteins created by simple genetic fusion of multiple T cell epitopes are not expected to fold reproducibly to a specific conformation (Sfatos et al., Phys. Rev. E48:465 (1993); Pande et al., Phys. Rev. E51:3381 (1995)) and thus in most cases will be expected to form insoluble aggregates. Indeed, formation of insoluble protein was observed for poly-T-cell epitope proteins produced recombinantly (Levy et al., Cell Immunol. 250:24 (2007)) or synthetically (Alexander et al., J. Immunol. 168;6189 (2002)). Concerns about reproducibility of manufacturing and adequacy of characterization limit the usefulness of insoluble or aggregated proteins as vaccines for use in humans.

US patent application publication no. US 2010/0068217 describes computational protocols for the design of epitope-protein scaffolds designed to elicit neutralizing antibodies wherein the three-dimensional structure of the epitope-antibody complex is known. More specifically, this publication describes chimeric polypeptides comprising a scaffold from a non-HIV polypeptide and an epitope that is recognized by an HIV-1 neutralizing antibody.

SUMMARY

The present disclosure provides synthetic T cell epitope-containing protein domains (referred to herein as STEDs), together with immunogenic compositions comprising such domains, and methods employing such domains in the treatment and/or prevention of disorders, such as infectious diseases and cancer. Each STED contains at least one T cell epitope of interest and a non-human-like synthetic "scaffold" amino acid sequence. As detailed below, STEDs can be designed using a combination of rational protein engineering and computational protein design. They have several advantages over recombinant proteins currently employed as vaccines or immunogenic compositions, including the following: (1) only the disease T cell epitopes of greatest interest need be included in the STED; (2) they do not contain unwanted sequence found in intact, natural proteins, such as "decoy" T cell epitopes (Rolland et al., PLoS Pathogen 3:e157 (2007)); (3) as the scaffold sequence is non-human-like, immune responses to the scaffold should not cross-react with human proteins; (4) unlike viral vector-based vaccines, STEDs are custom designed for different diseases so

that immune responses against a STED used for treatment of one disease should not cross-react with a STED designed for treatment of a different disease; (5) STEDs can be easily linked, either genetically or chemically, to other protein domains or non-protein structures; (6) STEDs are proteins, and protein-based vaccines have an excellent safety profile; (7) unlike peptides and non-well-folded polyepitope proteins, STEDs should be resistant to extracellular proteases in tissues; and (8) as STEDs are proteins, their manufacture and characterization is straightforward and inexpensive.

In one aspect, synthetic protein domains, or polypeptides, comprising at least one T cell epitope and a heterologous non-human-like scaffold amino acid sequence are provided wherein the polypeptide is non-aggregated, stable and well-folded when expressed in a recombinant expression system. In certain embodiments, the T cell epitope is a CD4 or CD8 T cell epitope from a proteome of an infectious organism or from a tumor associated antigen.

In another aspect, compositions are provided that comprise at least one synthetic T cell epitope-containing domain, or polypeptide, disclosed herein together with a pharmaceutically acceptable carrier. In certain embodiments, such compositions comprise at least two polypeptides disclosed herein, the at least two polypeptides being connected by a linker molecule. In other embodiments, the disclosed compositions comprise at least one polypeptide connected to a targeting molecule, such as an antibody, or antigen-binding fragment thereof, that binds to a polypeptide expressed on the surface of a target cell, such as a dendritic cell.

In a further aspect, methods for the treatment and/or prevention of a disorder in a subject are provided, such methods comprising administering to the subject an effective amount of a composition disclosed herein. Disorders that may be treated using the presently disclosed compositions include, but are not limited to, infectious diseases, such as human immunodeficiency virus (HIV), influenza, herpes-viruses, hepatitis C, human T-lymphotropic virus (HTLV), Epstein-Barr virus (EBV), dengue viruses, malaria, and tuberculosis; and cancers, such as breast, colorectal, ovarian, prostate, bladder, gastric, and non-small cell lung cancers (NSCLC) that overexpress tumor-associated antigens (TAAs) such as Her-2, CEA, MUC1, EGFR, COA-1, IGF-1, or EpCAM.

In yet another aspect, the present disclosure provides polynucleotides encoding a synthetic T cell epitope-containing domain, or polypeptide, disclosed herein, together with expression constructs comprising such polynucleotides and host cells transformed or transfected with such constructs.

The above-mentioned and additional features of the present invention and the manner of obtaining them will become apparent, and the invention will be best understood by reference to the following more detailed description. All references disclosed herein are hereby incorporated by reference in their entirety as if each was incorporated individually.

BRIEF DESCRIPTION OF THE FIGURES

Fig. 1 shows the tertiary structure and sequence of two artificial proteins (M7 (SEQ ID NO: 1) and TOP7 (SEQ ID NO: 2)) that are nearly identical structurally but have only 29% sequence identity.

Fig. 2 shows the tertiary structures of two slightly different four-helix bundle domain proteins IL-2 and T4 lysozyme.

Fig. 3 shows three different alpha helices from T4 lysozyme plotted on a helical wheel plot.

Fig. 4 shows conserved CD8, or cytotoxic T lymphocyte (CTL), epitopes in influenza NP (SEQ ID NO: 3).

Fig. 5 shows the amino acid sequence and structure of influenza NP (SEQ ID NO: 3) compared to DpsB from *Lactococcus lactis* (PDB:1ZS3; SEQ ID NO: 4).

Fig. 6 shows the secondary structure of DpsB from *Lactococcus lactis* (SEQ ID NO: 4).

DETAILED DESCRIPTION

As outlined above, the present disclosure provides synthetic T cell epitope-containing domains, or STEDs, together with polynucleotides encoding such domains, immunogenic compositions comprising such domains, and methods of using such compositions. As used herein, STEDs are defined as protein domains containing at least one T cell epitope of interest genetically linked with non-human-like synthetic "scaffold" sequence and which together form a protein domain that is, by design, well-folded, stable, and non-aggregated.

As used herein, a "well-folded" protein or protein domain is defined as a protein or protein domain capable of folding reproducibly to a specific conformation, including secondary structural elements and a tertiary structure, after its synthesis in a recombinant expression system or using a chemical synthetic process. This definition of well-folded protein or protein domain excludes not only proteins that are inherently unfolded, or partially unfolded, but also those that at equilibrium form a molten globule-like structure, which is a relatively compact structure but which is disordered (Ohgushi and Wada, Adv. Biophys.

18:75 (1984); Pande and Rokhsar, PNAS 95:1490 (1998)) and significantly less stable than well-folded (native) proteins (Kuwajima, Proteins 6:87 (1989); Yutani et al., J. Mol. Biol. 228:347 (1992)). A protein that is well-folded has a well-defined thermal denaturation transition, with a significant enthalpic component to its stability, unlike unfolded proteins that have no thermal transition, and molten globules which have either a shallow or no thermal denaturation transition (Munson et al., Protein Science 5:1584 (1996)).

A "non-aggregated" protein or protein domain is defined herein as one that can be prepared in a monomeric or monodisperse polymeric form in an aqueous solution with appropriate buffering agent, salts and/or other excipients as needed provided they are approved by the FDA for use with parenteral biologics, and at a concentration of at least 1 mg/ml. It will be understood by those in the art that many proteins will form aggregates when synthesized at high levels in a recombinant expression system. However, using appropriate expression and purification techniques, it will be possible in most cases to purify and concentrate to at least 1 mg/ml a non-aggregated form of the protein or protein domain from such an expression system provided the protein or protein domain is capable of forming a well-folded protein.

Methods of producing a protein domain or STED disclosed herein include, but are not limited to, chemical synthesis or using an *in vitro* translation system, which would be expected to produce a well-folded protein in a substantially non-aggregated and stable form. In contrast to a STED, a protein consisting of a simple genetic fusion of multiple T cell epitopes is not designed to form a particular structure and thus will, in most cases, have exposed hydrophobic groups that inherently generate aggregated forms of the protein no matter how the protein is expressed or purified.

From a regulatory perspective, the stability of a drug candidate or composition containing a STED must be determined by measurement of its chemical, biophysical and biological properties at multiple time points, as is typical for characterization of a biopharmaceutical. However, in the context of the present disclosure it is conformational stability that is of primary interest. Many synthetic proteins have been found to be conformationally unstable (Munson et al., Protein Science 5:1584 (1996)), and as such are more prone to formation of aggregates upon long term storage. Protein conformational stability can be measured directly using a variety of techniques such as differential scanning calorimetry (DSC), circular dichroism (CD), intrinsic fluorescence spectroscopy and NMR (Pace and Scholtz, In: Creighton, TE., editor. Protein structure: A practical approach. IRL Press; Oxford: 1997. p. 299-321). Analytical techniques commonly used to measure protein

aggregation include size exclusion chromatography (SEC), dynamic light scattering, and analytical ultracentrifugation. A common guideline for minimum stability of a biopharmaceutical candidate is no more than 5% aggregate observed when the product is stored at its designated temperature (usually 4 degrees centigrade) over the course of the anticipated shelf life of the product, which is typically greater than 1 year (Brorson and Phillips, BioProcess International, Nov. 2005 p.50). However, accelerated methods can be used to assess protein stability, such as DSC (Bruylants et al., Curr Med Chem. 12:2011 (2005)) or by use of chemical denaturants coupled with an analytical method, such as CD, to measure unfolding (Pace and Shaw, Proteins. Suppl 4:1 (2000)).

In certain embodiments, the STEDs disclosed herein have a melting temperature (T_m) greater than 60°C, as determined, for example, using DSC. In other embodiments, the STEDs have a T_m greater than 70°C or greater than 80°C. In certain embodiments, the STEDs disclosed herein are stable to the presence of at least 2 M guanidinium chloride, with unfolding typically measured by intrinsic fluorescence of the protein. Another approach to measuring stability is to incubate at temperatures below the Tm for extended periods of time, followed by determination of the fraction of folded protein using intrinsic fluorescence (Jung et al., J. Mol. Biol. 294:163 (1999)) or followed by SEC when unfolding of the protein is irreversible and results in aggregation (Willuda et al., Cancer Res. 59:5758 (1999)). In certain embodiments, a STED disclosed herein is stable to incubation at 37°C for greater than 1 hour, greater than 1 day, or greater than 5 days, with less than 5% of the protein being unfolded or in the form of aggregates, measured as described above. The stability of a STED is further analyzed using a biological potency assay, such as uptake by antigen presenting cells (APCs) and presentation on MHC molecules, as discussed further below. It is likely that stability of a STED as measured through the biophysical techniques described in the above (unfolding or aggregation) will correlate to biological potency. Verification that STEDs with a very high thermal stability (>80°C) are as efficiently proteolytically processed and presented on APCs as are STEDs with a lower thermal stability can also be performed using methods well known to those of skill in the art, including those described herein.

In the embodiments detailed below, the STEDs contain CD4 and/or CD8 T cell epitope domains from infectious organisms or tumor associated antigens (TAA). Such STEDs can be employed in the treatment and/or prevention of infectious diseases, such as human immunodeficiency virus (HIV), influenza, herpes-viruses, hepatitis C, human T-lymphotropic virus (HTLV), Epstein-Barr virus (EBV), dengue viruses, malaria, and tuberculosis; and

cancers, such as breast, colorectal, ovarian, prostate, bladder, gastric, and non-small cell lung cancers (NSCLC) that overexpress tumor-associated antigens (TAAs) such as Her-2, CEA, MUC1, EGFR, COA-1, IGF-1, or EpCAM.

Those of skill in the art will appreciate that STEDs can also, or alternatively, incorporate linear B cell epitopes. For example, linear B cell epitopes, which typically consist of 6-10 consecutive residues, can be incorporated into the loop regions of a scaffold structure in such a way that the structure is immunogenically similar to the epitope in its native context in the pathogenic or tumor protein.

All of the STEDs, polypeptides and polynucleotides described herein are isolated and purified, as those terms are commonly used in the art. Preferably, the STEDs, polypeptides and polynucleotides are at least about 80% pure, more preferably at least about 90% pure, and most preferably at least about 99% pure. Various techniques suitable for achieving such purification are well known to those of skill in the art. For STEDs and polypeptides, these include, for example, precipitation with ammonium sulfate, PEG, antibodies and the like, or by heat denaturation, followed by centrifugation; chromatography steps such as ion exchange, gel filtration, reverse phase, hydroxylapatite and affinity chromatography; isoelectric focusing; gel electrophoresis; and combinations of these and other techniques.

T Cell Epitopes

As noted above, in one aspect, the present disclosure provides STEDs that each comprise at least one T cell epitope of interest and a non-human-like synthetic "scaffold" amino acid sequence. In certain embodiments, such STEDs each comprise at least one epitope selected from the group consisting of: CD4 epitopes, CD8 epitopes, and combinations thereof. STEDs containing at least two, for example, two, three, four or five, epitopes are encompassed by the present disclosure, as are STEDs containing ten or fewer epitopes. The epitopes contained within the STED can be separated by scaffold sequence or can be contiguous. For example, when two or more epitopes are contiguous with each other in a naturally occurring protein, such epitopes can be incorporated as such into a STED. In certain embodiments, the STEDs disclosed herein are between 90 and 210 amino acids, for example 100, 110, 120, 130, 140, 150, 160, 170, 180, 190 or 200, amino acids in length.

a) CD8 epitopes

In certain embodiments, the T cell epitopes included in the disclosed STEDs are CD8, or cytotoxic T lymphocyte (CTL), epitopes. CD8 epitopes within protein sequences of

specific infectious disease organisms or in tumor associated antigens can be identified using techniques well known to those of skill in the art, including database and/or literature searches.

In STEDs for use in the treatment or prevention of infectious disorders, the CD8 epitope is preferably one that is found within conserved regions of a pathogen proteome. CD8 epitopes in conserved sequences are less likely to mutate so as to no longer be recognized by the cognate CD8+ T cell (i.e. are less likely to undergo "escape mutations"). CD8 epitopes that are MHC Class I epitopes for which corresponding CD8+ T cell populations have been experimentally demonstrated in persons infected with the pathogen are preferred for inclusion in a STED. Techniques for identifying the presence of corresponding CD8+ T cell populations are well known in the art and include, for example, memory T cell recall assays (Lee et al., J. Clin. Invest. 118:3478-3490 (2008)). In one embodiment, the CD8 epitope is a MHC Class I epitope for which escape mutations hardly ever occur, or for which escape mutations that do occur result in a pathogen with significantly reduced fitness. For example, certain mutations in the p24 GAG protein of HIV result in escape from CTLs specific to those epitopes but generate a less fit virus as demonstrated by lower viral titers (Allen and Altfield, J. Exp. Med. 205:1003 (2008)).

In another embodiment, the STED comprises multiple CD8 epitopes corresponding to a collection of HLA haplotypes that would cover a large percentage of the human population.

In STEDs for use in the treatment or prevention of a cancer, the CD8 epitope is preferably from a tumor associated antigen and is one for which cognate CTLs have been shown to be present in patients with the cancer of interest (Pavelko et al., Cancer Res 69(7):3114 (2009)). In certain embodiments, it is also desirable to incorporate subdominant or low-affinity T cell epitopes since the corresponding CTLs may be less likely to have been deleted (Pavelko et al., Cancer Res 69(7):3114 (2009)).

In certain embodiments, the STED contains CD8 epitopes that do not normally reside in the same protein. In other embodiments the STED contains protein segments from an infectious organism or from a tumor associated antigen, wherein the protein segment contains multiple CD8 epitopes.

b. CD4 epitopes

In certain embodiments, the T cell epitopes included in the disclosed STEDs are CD4, or helper T lymphocyte (HTL), epitopes. CD4 epitopes for inclusion in a STED are selected based on similar factors to those discussed above with respect to CD8 epitopes. CD4

epitopes within protein sequences of specific infectious disease organisms can be identified using databases or by doing literature searches. For example, HIV helper/CD4+ epitopes are listed in "HIV Molecular Immunology", 2008, Ed. Korber et al. In certain embodiments, the STED comprises a universal, or immunodominant, CD4 epitope, for example from tetanus and/or diphtheria toxins (Diethelm-Okita et al., J. Infect. Dis. 181:1001 (2000)).

The CD4 epitopes need to be incorporated in the STED in such a way that they are efficiently processed and displayed on MHC Class II molecules after uptake by dendritic cells. This can be verified experimentally using methods known to those of skill in the art, such as an *ex vivo* memory recall assay (e.g., Lee et al., J. Clin. Invest. 118:3478-3490 (2008)). For example, tetanus toxin or diphtheria toxin-specific HTLs are present in the blood of most people who have been immunized with these vaccines. Preferred CD4 T cell epitopes for inclusion in a STED for the treatment or prevention of an infectious disease are those for which HTL specific to the CD4 T cell epitope are known to be present in the blood of persons who have been infected with that disease organism.

Scaffold Sequence

As used herein, the term "scaffold sequence" refers to protein sequences with which T cell epitopes from an infectious organism or a tumor associated antigen are combined to form a STED that is well-folded, non-aggregated and stable as discussed in detail above.

Most hydrophobic side groups found within the context of an intact natural protein interact with other hydrophobic residues within a central, water-excluding, core of the folded protein. These hydrophobic interactions are key to the formation of native well-folded structures. When segments of a protein are removed from their natural context, hydrophobic residues may become exposed to the aqueous environment, either because the protein is not well-folded or because the folded structure exposes many of the previously-buried hydrophobic side chains to the aqueous environment. This will usually result in the formation of protein aggregates, generated by the intermolecular interactions between hydrophobic groups on different protein molecules. Due to the heterogeneous nature, uncertain structure, probable reduced efficacy, and potential batch to batch inconsistency of these aggregates, they are unsuitable for use as biopharmaceuticals. To minimize these problems, scaffold sequences are engineered into the STED sequences in order to generate well-folded structures in which most of the hydrophobic side groups are clustered with other hydrophobic groups in a water-excluding environment.

Aggregation may also result from disruption of other types of interactions within a folded protein. For example, disruption of charged pairs could also result in protein aggregation. Thus, in some embodiments scaffold sequences with charged residues are included in the STED to provide pairing partners for charged groups and/or regions that might otherwise generate aggregates due to these unpaired charged groups.

As discussed above, preferably the STED is stable, i.e. the potency of an immunogenic composition comprising the STED does not diminish significantly during the time period and/or storage conditions anticipated for the composition.

In certain embodiments, the scaffold sequence is chosen such that the STED is monomeric whereby cross-linking of dendritic cell targets is avoided. On the other hand, if experimental results suggest that a dimeric or multimeric structure is not disadvantageous, then a dimer or higher multimers of the STED may be acceptable. Preferably the STED is resistant to digestion by extra-cellular proteases present at the site of administration. Well-folded proteins are generally more resistant to digestion by protease than unfolded or partially-folded proteins (Hetzel et al., J. Immunother. 31:370 (2008)).

A scaffold sequence employed in the STEDs disclosed herein can be either completely non-natural (artificial) or at least partially identical to a sequence from a non-human organism, provided that the scaffold sequence has no significant identity to a human sequence. In certain embodiments, the scaffold sequence is non-human-like. As used herein, "non-human-like" refers to a sequence having no more than five consecutive residues that are identical to a human sequence. In certain embodiments, a non-human-like sequence has no more than four, or no more than three, consecutive residues that are identical to a human sequence.

Scaffold sequences should not generate either a humoral or cell-mediated immune response that cross-reacts with human proteins. Given that the scaffold sequences are non-human-like, it is unlikely that they will generate immune responses that are cross-reactive to human sequences. To reduce the flux of competing T cell epitopes deriving from the scaffold sequences, the total length of scaffold sequence should be minimized to the extent possible. In certain embodiments, the length of scaffold sequence relative to the total length of T cell epitopes will be at a ratio of less than or equal to 20:1, less than or equal to 15:1, or less than or equal to 10:1.

In certain embodiments, the scaffold sequences lack immunodominant epitopes that might inhibit priming, expansion and/or memory cell generation of CD4 and/or CD8+ T cells specific for the infectious pathogen or tumor associated antigen epitopes employed in the

STED. Predictive algorithms that can be used to assess the potential for immunodominant T cell epitopes within a scaffold sequence are well known to those in the art and include, for example, the Immune Epitope Database and NetCTL 1.2 Server, both of which are available on the internet. If a high affinity T cell epitope was predicted within a proposed scaffold sequence, it can be mutated such that the structure of the STED is not compromised, by, for example, using a computational program such as Rosetta to design an alternative sequence (Dantas et al., J. Mol. Biol. 332:449 (2003)) but yet remove the T cell epitope of concern.

As detailed below, T cell epitopes are genetically fused with a scaffold sequence in such a way that leads to a well-folded structure.

Sequences flanking T cell epitopes are found at the interfaces between scaffold sequence and T cell epitopes. These flanking sequences, or regions, can be important for the proteolytic processing, within antigen presenting cells, of proteins that will be presented on MHC Class II molecules or cross-presented on MHC Class I molecules. In some embodiments, optimization of the flanking sequences is performed based on which protease system is expected to process the STED. For example, for presentation on MHC Class II molecules, the relevant proteases would be endosomal cathepsins such as CatS, CatL and CatB (Lennon-Duménil et al., J. Exp. Med. 196:529 (2002)), while for cross-presentation on MHC Class I molecules the relevant protease could either be the cytoplasmic-localized proteasome or endosomal-localized proteases such as CatS (Shen et al., Immunity 21:155 (2004)). It is thus advantageous to include sequences in the flanking regions for which the relevant protease(s) has some demonstrated specificity.

Design of synthetic T cell epitope-containing domains (STEDs)

The design of STEDs is guided by protein structure and folding principles and considerations, as will be well-understood by persons of ordinary skill in the field of protein structure and design. Some examples of these principles and considerations are as follows:

a. Burial and dense packing of hydrophobic side chains in the interior of proteins is a primary driving force for protein folding (Yue and Dill, Proc. Natl. Acad. Sci. USA 89:4163 (1992); Kauzman, Adv. Protein Chem. 14:1 (1959); Lau and Dill, Macromolecules 22:3986 (1989); Dill et al., Proc. Natl. Acad. Sci. USA 90:1942 (1993)).

b. Well-folded protein structures have an abundance of hydrogen-bonded secondary structural elements, with these bonds residing primarily in alpha helices and beta-strands (Dill et al., Biochem. 29:7133 (1990)).

c. Structure-determining properties, such as periodicity of polar and nonpolar amino acids (Xiong et al., Proc. Natl. Acad. Sci. USA, 92:6349 (1995)).
d. Mechanisms utilized by natural proteins to prevent inter-polypeptide association or aggregation through beta strand edge-edge associations (Richardson and Richardson, Proc. Natl. Acad. Sci. USA, 99:2754 (2002)).
e. Structures prone to aggregation even when properly folded should be avoided. Such structures can be predicted using algorithms such as TANGO (Fernandez-Escamilla et al., Nat. Biotechnol. 22:1302 (2004)), and screened for using techniques described above.
f. Use of capping amino acids at the ends of alpha helices that stabilize these secondary structures (Doig and Baldwin, Protein Sci. 4:1325 (1995)).
g. Use of stability-enhancing strategies such as optimization of beta-turn sequences (Trevino et al., J. Mol. Biol. 373:211 (2007)).
h. Negative design elements can be used to avoid competing folding states or increase the solubility of a protein (Suárez and Jaramillo, J. R. Soc. Interface 6:S477 (2009).

The main steps in the design of a STED are: selection of the scaffold template, computational protein design and, optionally, use of combinatorial libraries and screening.

a. Selection of scaffold templates

The term "scaffold template" as used here refers to a protein used as the starting structure for incorporation of T cell epitopes. It is the tertiary structure of the scaffold that is of interest, not the sequence per se, and the sequence of the final STED may be very different from the sequence of the scaffold template. This is because, as discussed above, the sequence of the final STED should be non-human-like whereas the scaffold template used to design a STED may have regions with significant identity to human proteins. In this respect, a scaffold template used to create a STED is very different from, for example, a scaffold used as a binding protein (Nygren and Skerra, J. Immunol. Methods 290:3 (2004)). Using modern computational protein design methods, it is now possible to dramatically change the sequence of a protein while retaining the same structure, as exemplified in Figure 1, which shows the tertiary structure and amino acid sequence for two artificial proteins, TOP7 (SEQ ID NO: 1) and M7 (SEQ ID NO: 2). M7 is an artificial protein that was designed and successfully synthesized and that was nearly identical structurally to the previously designed artificial

sequence TOP7 but had only 29% sequence identity to TOP7 (Dallüge et al., Proteins 68:839 (2007)). The TOP7 and M7 sequences, which were both designed using computational methods in an analogous way to that which is used to design STEDs, both fit the definition of a preferred scaffold sequence in that both have, at most, a three amino acid match to human sequence as determined using a standard BLAST search.

The scaffold template can be a natural protein fold or a known artificial fold (e.g., TOP7) or a newly designed artificial fold. For reasons of computation, folding/expression, and minimization of competing T cell epitopes derived from scaffold sequence, scaffold templates less than 200 amino acids in length are preferred for use in certain embodiments. However, in certain embodiments larger scaffold templates are desirable, for example, when the scaffold consists largely of a natural protein from a pathogen or tumor associated antigen.

A STED can serve its function as a vaccine, or immunogenic composition, without necessarily having a specific structure, i.e., it is not necessary that the actual structure of the STED match the structure of the scaffold template used in its design, or that the predicted structure of the STED match its actual structure, as long as the structure of the STED fulfils the requirements of being well-folded, stable and non-aggregated.

There are a large number of candidate scaffold templates that can be found in the protein structure databases such as SCOP (Murzin et al., J. Mol. Biol. 247:536 (1995)) or CATH (Orengo et al., Nucleic Acids Res. 27:275 (1999)). To identify appropriate scaffold template candidates for use with a particular set of T cell epitopes, it is necessary to first characterize the structural elements within which the T cell epitopes are contained, beginning with secondary structure (alpha helix, beta-strand, or loop). In many cases the structure of the protein from the infectious agent or tumor associated antigen is known and this characterization will then be relatively straightforward. If the structure is not known, it can be predicted using structure prediction algorithms such as Rosetta (Das and Baker, Annu. Rev. Biochem. 77:363 (2008)).

Characterization of alpha helical segments containing T cell epitopes of interest can include the degree of amphiphilicity, as measured using a helical wheel plot (Schiffer and Edmundson, Biophys. J. 7:121 (1967)) or using more recent methods for characterizing amphiphilicity (Phoenix et al., Curr. Protein Pept Sci. 3(2):201 (2002)). For example, an alpha helix buried within a protein structure will have low amphiphilicity and contain hydrophobic amino acids distributed relatively densely all around the helix (e.g., LSFAAAMNGLA (SEQ ID NO: 5) in citrate synthase, residues 260-270) whereas an entirely solvent-exposed helix will also have relatively low amphiphilicity but contain few

hydrophobic amino acids (e.g., KEDAKGKSEEE (SEQ ID NO: 6) in troponin C, residues 87-97). In contrast, a partially buried helix typically would be amphiphilic, with its hydrophobic side facing the core of the domain and its hydrophilic side exposed to solvent (e.g., INEGFDLLRSG (SEQ ID NO: 7) in alcohol dehydrogenase, residues 355-365). The presence and location of bulky side groups within the alpha helix is another element of characterization that is considered in choosing a scaffold into which the T cell epitope-containing alpha helical segment is grafted.

Although not limited to these structures, alpha helical segments containing T cell epitopes of interest can, in many cases, be incorporated into a four-helix bundle-based scaffold. The crystal structures of two slightly different four-helix bundle domain proteins (IL-2 and T4 lysozyme) are shown in Figure 2. These structures demonstrate that even in this one family of protein domains there is significant variability, which thus offers a variety of structures from which to choose a possible scaffold template within which to incorporate T cell epitopes. For example, Figure 3 shows three different alpha helices from T4 lysozyme plotted on a helical wheel plot. Each is distinct with respect to its hydrophobicity, amphiphilicity and length of helix. Thus, for example, segment 96-106 in T4 lysozyme (Figure 3) is a candidate target site for grafting a T cell epitope residing in a short alpha helical segment buried within the pathogenic protein, assuming T7 lysozyme protein is a candidate scaffold template. Similar to the example provided below using DpsB as a scaffold template, using the T4 lysozyme scaffold template involves considerably more protein engineering than simply grafting of the T cell epitope.

In one embodiment, a beta strand segment containing a T cell epitope of interest is characterized firstly by its length and composition of R groups on each side of the strand. The vast majority of beta strands have at least one side composed largely of hydrophobic residues. The hydrophobic face of a beta strand may contain residues that are amphiphilic, such as lysine or arginine, which have long aliphatic chains that can be buried with other hydrophobic groups but are terminated with a charged group that is usually exposed to solvent. The position of these amphiphilic side groups relative to other parts of the protein is a critical design consideration since in most cases a charged group cannot be accommodated inside the hydrophobic core of a protein. Some beta strands are largely hydrophobic on both sides of the strand, and this environment will most likely need to be replicated when the T cell epitope peptide is grafted into the scaffold template since exposure of a hydrophobic surface that was previously buried or mostly buried may result in protein aggregation.

T cell epitopes that reside in loops or unstructured regions are usually located at the surface of a protein and lack the periodic structure of alpha helices or beta strands. However, loop segments can be grouped according to their various conformations (Perskie et al., Protein Sci. 17:1151 (2008)). Most T cell epitopes that fall within loop regions can be incorporated into specific loop regions connecting alpha helical or beta strand segments within a scaffold, for example in a similar way that CDRs or CDR-like peptides are engineered in antibodies or antibody-like proteins, respectively.

Appropriate scaffold templates are then selected based on identification of small, well-folded proteins that contain structural elements matching those that contain the T cell epitopes, as described above. Preferred scaffold templates are those known to be produced in a non-aggregated form in recombinant expression systems. Such scaffold templates can be identified, for example, manually by searching databases and the scientific literature. For this purpose, it is useful to assemble a custom set of well-characterized proteins that together comprise a diverse set of structures with a diverse set of structural subdomains. Examples of proteins that can be included in this custom set of scaffold templates include those described by Nygren and Skerra as non-antibody binding protein scaffolds such as those based on the lipocalins or fibronectin type III domain (Nygren and Skerra, J. Immunol. Methods 290:3 (2004)).

In addition to manual assembly of candidate scaffold templates, it is advantageous to use algorithms capable of searching the protein structure databases for domains with similar backbone structures as those containing the T cell epitopes of interest. Direct comparison of protein structures can be done using a variety of programs including CE, (Shindyalov and Bourne, Protein Eng., 11:739 (1998)), MAMMOTH (Ortiz et al., Protein Sci., 11: 2606 (1998)), DALI (Holm and Sander, J. Mol. Biol. 233:123 (1993)), or VAST (Madej et al., Proteins 23, 356 (1995)). Recent algorithms that are index-based and that more quickly identify similar structures include ProtDex2 (Aung and Tan, Bioinformatics, 20:1045 (2004)) and ProteinDBS (Shyu et al., Nucleic Acids Res., 32:W572 (2004)). Other search and compare programs first transform three-dimensional structural data into one-dimensional text strings, using a structural alphabet, which allows very rapid searches. Examples in this latter category include TOPSCAN (Martin, Protein Eng. 13:829 (2000)), YAKUSA (Carpentier et al., Proteins 61:137 (2005)), RST (Lo et al., BMC Bioinformatics 8:307 (2007)), iSARST (Lo et al., Nucleic Acids Res. 37:W545 2009)), 3D-BLAST (Yang et al., Nucleic Acids Res. 34:3647 (2006)), SA-Search (Guyon et al., Nucleic Acids Res. 32:W545 (2004)), and GANGSTA (Guerler and Knapp, Protein Sci. 17:1374 (2008)). Additional algorithms can be

found at Wikipedia under the heading "structural alignment software". Many of these computational search methods are freely available for use on a web-based server. In addition, it may be beneficial to develop a custom structure search algorithm or a custom structure database (e.g., < 200 residue domains) for the purposes of this invention.

The following example illustrates how the programs discussed above can be used to identify candidate scaffolds, using the influenza NP protein (SEQ ID NO: 3) as the search structure. Because it is relatively conserved, the influenza NP protein is a target for cell mediated immunity using vaccines or immunogenic compositions (Doherty and Kelso, J. Clin. Invest. 118:3273 (2008)). Multiple CD8, or cytotoxic T lymphocyte (CTL), epitopes have been identified within the conserved regions of NP, with some of the best characterized shown in Figure 4.

The crystal structure of NP (PDB: 2IQH) was used to search for proteins containing substructures similar to those in NP using the 3D-BLAST algorithm (http://3d-blast.life.nctu.edu.tw/). One of the proteins identified by this search, DpsB from *Lactococcus lactis* (PDB:1ZS3; SEQ ID NO: 4), is a candidate as a scaffold because of its small size (182 residues) and similarity of structure over a lengthy region containing four of the NP CTL epitopes, three of which reside largely within highly similar structures, as shown in Figure 5. Of note is the complete absence of sequence similarity between NP and DpsB sequence in regions of very similar structure (shaded regions in Figure 5), which again demonstrates that very different sequences can form the same backbone structures and illustrates the great flexibility a protein engineer has to change residues without changing structure, provided those changes are made using appropriate structure and computational tools as discussed in above.

It should be noted that just because a domain with a similar structure can be identified using this approach, it may not be suitable if, for example, the side groups are very different between the T cell epitope peptide and the segment which it is to replace in the scaffold. Some simple modelling of the T cell epitope in place of the corresponding peptide segment in the scaffold template would help answer this question, also taking into account that residues on an adjacent structure but not containing the desired T cell epitopes can be freely changed to accommodate differences between the grafted peptide and the one it replaces in the scaffold.

Assuming that the DpsB scaffold template works for the set of T cell epitopes shown in Figure 5, the DpsB protein itself is only a starting point and many changes need to be made

before a final STED is generated. For example, some of the changes that could be made to DpsB include: (1) grafting of the influenza NP CTL sequences in place of the homologous DpsB sequences as indicated by the underlined sequence in Figure 5; (2) mutating sequences with greater than four consecutive matches to human sequence (the human sequence Zinc finger protein 211 has a 7 residue match to DpsB); (3) mutation of residues responsible for homo-dodecamer formation of DpsB (e.g., Chien et al., J. Biol. Chem. 279:52338 (2004); Jónsson et al., EMBO J. 14(22):5745 (1995)) and, if the protein aggregates after generation of the monomer, mutation of residues at the hydrophobic face revealed on the monomeric form; and (4) changes in the DpsB scaffold potentially needed to accommodate the grafted CTL epitopes (e.g., steric interference between buried hydrophobic R groups resulting from the NP CTL peptide grafts). Other alterations may include: (1) deletion of unnecessary sequence from the DpsB scaffold, such as most of the N-terminal DNA-binding helix (N-terminal 18 residues of DpsB), since most of this sequence is not needed for the STED to function as a vaccine protein; and (2) incorporation of additional CTL epitopes into the structure, such as replacing the surface-localized region in DpsB approximately between residues 96 and 108 (see Figure 6) with additional sequence from NP such as the mostly-unstructured CTL epitope-containing segments centered around position 269 or position 385 (see Figure 4). Additional engineering of the protein can be done for the purpose of optimizing the final STED protein for its function as a vaccine or immunogenic composition such as, as discussed above, modification of sequence flanking the CTL epitopes for the purpose of optimizing intracellular proteolytic processing to facilitate MHC-mediated presentation of CTL epitopes by antigen presenting cells.

It may not be necessary in all cases to incorporate a T cell epitope into the same structural environment within which it normally resides. For example, Lunde et al. described grafting four different T cell epitopes of different lengths and derived from different structural elements – including beta strand, alpha helix and loop structures – into the loop region of an IgG constant domain, apparently without negatively affecting folding or secretion of the modified antibody (Lunde et al., J. Immunol. 168:2154 (2002)). However, the work in this publication did not include a complete characterization of these molecules from a biopharmaceutical perspective, and it remains to be seen to what extent T cell epitopes can be incorporated into the loop regions of scaffold proteins without causing destabilization or aggregation of the resulting protein. Likely, success in transferring a peptide from a structured region (e.g., alpha helix or beta strand) into a loop region of a scaffold will be sequence-specific.

Scaffold templates that could be used for this purpose include, for example, those with a beta sandwich structure for which some of the connecting loop regions could be replaced with peptide segments containing T cell epitopes of interest. Two such structures include the lipocalin family, in which for one such protein some of the beta strand-connecting loops were replaced to create antibody-like molecules (Schonfeld et al., Proc. Natl. Acad. Sci. USA 106:8198 (2009)), and the fibronectin type III domain family, in which similar loop replacements have been described (Olson and Roberts, Protein Sci. 16:476 (2007)). There are many such beta sandwich-containing or other structures that can be identified in the SCOP or CATH databases and that could be employed as scaffold templates into which T cell epitopes could be incorporated by replacement of loop regions. The actual sequence of these scaffolds would be altered to become non-human-like, as described above.

b. Computational methods

STEDs designed using computational methods will differ in several respects from the examples of computationally-designed proteins given in the examples below. One difference is that, within STEDs, the T cell epitopes are invariant, with the scaffold sequences being built around the T cell epitopes. Thus, T cell epitope sequences are fixed within a STED (although not necessarily their secondary structures as they exist in the context of their native proteins). In contrast, there is significant latitude to change scaffold sequences provided the end result is a protein domain that meet the various criteria outlined above.

Computational protein design has undergone rapid development in recent years, and there are now multiple examples of synthetic proteins being designed from scratch. The increased accuracy and speed with which protein structure can be determined using computational methods has played a large part in these advances. For a review of computational protein design see Butterfoss and Kuhlman, Ann. Rev. Biophys. Biomol. Struct. 35:49 (2006), and Suárez and Jaramillo, J. R. Soc. Interface 6:S477 (2009). Additional examples of *de novo* protein design are provided in the following.

Using the RosettaDesign program, Kuhlman et al. designed an artificial 92 residue alpha/beta protein with a novel fold and highly stable structure. After expression and crystallization, this protein was found to have a structure nearly identical to that predicted by computational methods (Science 302:1364 (2003)). The protein design process employed by Kuhlman et al. iterates between design and prediction until the desired structure is obtained.

Using an automated sequence selection algorithm based on dead-end elimination, Offredi et al. designed a synthetic 216 residue protein with an idealized alpha/beta barrel

structure and presented evidence that the expressed protein was folded as designed (J. Mol. Biol. 325:163 (2003)).

Dantas et al. described the redesign of nine small protein domains, using RosettaDesign, with an average of only 35% sequence identity between the original protein and the redesigned one. Eight of the redesigned proteins were folded, six with stabilities equal to or greater than the parent structure, and four of these had a monomeric, well-packed and rigid structure (J. Mol. Biol. 332:449 (2003)).

As discussed above, a tetrapeptide fragment-based computational approach was used successfully (Dallüge, Proteins 68:839 (2007)), to design an artificial protein (M7) that was nearly identical structurally to a structure (TOP7) previously designed using RosettaDesign (Kuhlman et al., Science 302:1364 (2003)), yet the M7 sequence has only 29% identity to that of TOP7 (Figure 1). There are many examples of natural proteins with highly similar structures but low sequence identity (Brenner and Levitt, Protein Sci. 9:197 (2000); Koppensteiner et al., J. Mol. Biol. 296:1139 (2000); Tian and Scolnick, J. Mol. Biol. 333:863 (2003)), and the successful design of the M7 structure as well as the protein redesigns of Dantas et al. described above illustrate the power of modern computational methods to "copy" the structure of a (small) protein but using an amino acid sequence very different from that of the original protein. This permits the design of synthetic proteins containing T cell epitopes based on the structures of natural proteins, including human proteins, but which have a sequence that is very different from any human protein sequence.

Other examples of computational protein design are provided in the following publications: MacDonald et al., Proteins 78:1311 (2010); Faiella et al., Nat Chem Biol. 5:882 (2009); Fung et al., Biophys J. 94:584 (2008); and Fry et al., J Am Chem Soc. 132:3997 (2010). Detailed methods for use of computational programs to design proteins can be found in various publications, including the patent literature. For example, US Patents 6,804,611 and 7,574,306, the disclosures of which are hereby incorporated by reference, each describe the use of a computer-based algorithm for protein design.

As mentioned previously, the designed protein should be non-human-like – i.e., it should not have regions with similarity to human protein sequence from the perspective of having either CD4 or CD8 T cell epitopes or B cell epitopes. As an example of a sequence that would fulfill this criteria, when the TOP7 sequence (cited above) is aligned to human protein sequence using standard BLAST parameters, there are only three human sequences that come up as hits, with the longest stretch of identity being three amino acids and the maximum number of identities in an eight amino acid moving window being only four. This

sequence is highly unlikely to generate either antibodies or T cell clones that recognize human protein sequences. To further ensure there are no cross-reacting immune epitopes between the designed STED sequence and human sequences, T and B cell epitope prediction algorithms can be used for both the STED and the closest human sequences.

c. Combinatorial libraries

In some embodiments, STEDs are designed using a combinatorial library approach in concert with computational design. For example, a library can be designed, constructed and screened in cases where there is a particularly problematic region of the protein that is, for example, responsible for aggregation. Such a screen can be employed to determine which sequences generate well-folded proteins, with the results being used in an iterative fashion with computational design to further optimize design of the protein domain of interest.

In fact, synthetic proteins have been successfully designed using combinatorial libraries in the absence of computational methods. For example, the Hecht group described a "binary code" strategy (Hecht et al., Protein Science 13:1711 (2004)), which is based on the observation that the periodicity of polar and nonpolar amino acids is a major determinant of secondary structure (Vazquez et al., Proc. Natl. Acad. Sci. USA, 90:9100 (1993); Xion et al., Proc. Natl. Acad. Sci. USA, 92:6349 (1995)). This strategy was used to construct combinatorial libraries from which artificial proteins were identified that were well-folded four-helix bundle proteins (Wei et al., Proc. Natl. Acad. Sci. USA 100:13270 (2003)) or soluble, monomeric six-stranded beta proteins (Wang and Hecht, Proc. Natl. Acad. Sci. USA 99:2760 (2002)). Screening members in such a library for those produced in a soluble form can be facilitated by the use of genetic fusions such as to GFP (Olson and Roberts, Protein Sci. 16:476 (2007)).

Genetic or chemical linkage of STEDs to other structures

STEDs can readily be genetically linked to a variety of other protein domains, including other STEDs, other T cell epitope-containing domains, virus-like particle-forming proteins such as the Hepatitis B surface antigen (Stoute et al., N Engl J Med., 336:86 1997), or antibody-like molecules targeting dendritic cells. Particularly useful are molecules that target STEDs to receptors on dendritic cells that internalize bound proteins through one of several endocytic pathways for efficient uptake and presentation of T cell epitopes, including both presentation of CD4 T cell epitopes on MHC Class I molecules and cross-presentation of CD8 T cell epitopes on MHC Class I molecules (Caminschi et al, Eur. J. Immunol. 39:931

(2008); Keler et al., Oncogene 26, 3758 (2007)). STEDs can be connected to other STEDs or other domains as described in the above by use of flexible linkers (e.g., a short peptide segment containing glycines and serines) designed to permit independent folding of each domain but which is not likely to be a target for proteases (Hetzel et al., J. Immunother. 31:370 (2008)).

STEDs can also be chemically conjugated to other protein domains or to non-protein structures, such as an endosomolytic polymer that facilitates entry of the conjugated protein into the cytoplasm of antigen presenting cells (Flanary et al., Bioconjug. Chem. 20(2):241 (2009)).

As used herein, the term "targeting molecule", refers to a molecule that specifically binds to a polypeptide, such as a receptor, expressed on the surface of a target cell, such as a dendritic cell, and includes, but is not limited to, antibodies, including monoclonal antibodies and polyclonal antibodies; antigen-binding fragments thereof, such as F(ab) fragments, $F(ab')_2$ fragments, variable domain fragments (Fv), small chain antibody variable domain fragments (scFv), and heavy chain variable domains (V_{HH}); small molecules; hormones; cytokines; and viruses (either native or modified). Antibodies, and fragments thereof, may be derived from any species, including humans or may be formed as chimeric proteins which employ sequences from more than one species. The term "targeting molecule" as used herein thus encompasses humanized antibodies and veneered antibodies.

A targeting molecule is said to "specifically bind," to a polypeptide if it reacts at a detectable level (within, for example, an ELISA assay) with the polypeptide, and does not react detectably with unrelated polypeptides under similar conditions. In order to minimize any off-target effects, cell-type specific targeting molecules can be employed. Antibodies, and fragments thereof, may be prepared by any of a variety of techniques known to those of ordinary skill in the art (*see, e.g.,* Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, 1988). In general, antibodies can be produced by cell culture techniques, including the generation of monoclonal antibodies as described, for example, by Kohler and Milstein, *Eur. J. Immunol.* 6:511-519, 1976, and improvements thereto, via transfection of antibody genes into suitable bacterial or mammalian cell hosts, in order to allow for the production of recombinant antibodies, or by protein synthesis.

Polynucleotides, Expression Vectors and Host Cells

The present disclosure further provides polynucleotides encoding the STEDs of this disclosure, vectors (including cloning vectors and expression vectors) comprising such

polynucleotides, and cells (e.g., host cells) transformed or transfected with such polynucleotides or vectors.

The term "polynucleotide(s)," as used herein, means a single or double-stranded polymer of deoxyribonucleotide or ribonucleotide bases and includes DNA and corresponding RNA molecules, including HnRNA and mRNA molecules, both sense and anti-sense strands, and comprehends cDNA, genomic DNA and recombinant DNA, as well as wholly or partially synthesized polynucleotides. An HnRNA molecule contains introns and corresponds to a DNA molecule in a generally one-to-one manner. An mRNA molecule corresponds to an HnRNA and DNA molecule from which the introns have been excised.

In certain embodiments, vectors, in particular recombinant expression constructs, are provided that include a polynucleotide encoding a STED. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid molecule to which it has been linked. Exemplary vectors include plasmids, yeast artificial chromosomes, and viral genomes. Certain vectors can autonomously replicate in a host cell, while other vectors can be integrated into the genome of a host cell and thereby are replicated with the host genome. In addition, certain vectors, referred to herein as "recombinant expression vectors" or "expression vectors" contain one or more nucleic acid sequences, such as a polynucleotide encoding a STED, that are operatively linked to an expression control sequence and, therefore, are capable of directing the expression of those sequences.

Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described, for example, in Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor, NY, (1989). Exemplary cloning/expression vectors include cloning vectors, shuttle vectors, and expression constructs, that may be based on plasmids, phagemids, phasmids, cosmids, viruses, artificial chromosomes, or any nucleic acid vehicle known in the art suitable for amplification, transfer, and/or expression of a polynucleotide contained therein

Generally, recombinant expression vectors include origins of replication, selectable markers permitting transformation of a host cell, and a promoter derived from a highly-expressed gene to direct transcription of a downstream nucleic acid sequence, such as polynucleotide disclosed herein. Additional expression control elements, such as enhancers, factor-specific binding sites, terminators, and ribosome binding sites may also be included in the expression vectors of this disclosure. A polynucleotide disclosed herein is assembled in appropriate phase with translation initiation and termination sequences. Thus, for example,

the STED-encoding polynucleotides as provided herein may be included in any one of a variety of expression vector constructs for expressing the STED in a host cell.

The polynucleotide may be inserted into a vector by a variety of procedures well known to those in the art. In general, a DNA sequence is inserted into an appropriate restriction endonuclease cleavage site(s) using known procedures. Standard techniques for cloning, DNA isolation, amplification and purification, for enzymatic reactions involving DNA ligase, DNA polymerase, restriction endonucleases and the like, and various separation techniques may be employed. A number of standard techniques are described, for example, in Ausubel et al. (1993 Current Protocols in Molecular Biology, Greene Publ. Assoc. Inc. & John Wiley & Sons, Inc., Boston, MA); Sambrook et al. (1989 Molecular Cloning, Second Ed., Cold Spring Harbor Laboratory, Plainview, NY); Maniatis et al. (1982 Molecular Cloning, Cold Spring Harbor Laboratory, Plainview, NY); Glover (Ed.) (1985 DNA Cloning Vol. I and II, IRL Press, Oxford, UK); and Hames and Higgins (Eds.), (1985 Nucleic Acid Hybridization, IRL Press, Oxford, UK).

The DNA sequence in the expression vector is operatively linked to at least one appropriate expression control sequence, such as a constitutive promoter or a regulated promoter, to direct mRNA synthesis. Representative examples of such expression control sequences include promoters of eukaryotic cells or their viruses, as described above. Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art, and preparation of certain particularly preferred recombinant expression constructs comprising at least one promoter or regulated promoter operably linked to a nucleic acid encoding a protein or polypeptide according to this disclosure is described herein.

In other embodiments, the present disclosure provides a host cell that is transformed or transfected with, or otherwise contains, a polynucleotide or vector/expression construct disclosed herein. The term "recombinant host cell", or "host cell", refers to a cell containing a recombinant expression vector. It should be understood that such terms are intended to refer not only to the particular subject cell but also to the progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term "host cell" as used herein.

The polynucleotides or cloning/expression constructs of this disclosure are introduced into suitable host cells using any method known in the art, including transformation, transfection and transduction. Examples of host cells that may be transformed with an expression vector disclosed herein include eukaryotic host cells, such as VERO cells, HeLa cells, Chinese hamster ovary (CHO) cell lines (including modified CHO cells capable of modifying the glycosylation pattern of expressed multivalent binding molecules, *see* US Patent Application Publication No. 2003/0115614), COS cells (such as COS-7), W138, BHK, HepG2, 3T3, RIN, MDCK, A549, PC12, K562, HEK293 cells, HepG2 cells, N cells, 3T3 cells, *Spodoptera frugiperda* cells (*e.g.*, Sf9 cells), *Saccharomyces cerevisiae* cells, and any other eukaryotic cell known in the art to be useful in expressing, and optionally isolating, a protein domain according to this disclosure. In other embodiments, the host cells are prokaryotic cells, such as *Escherichia coli*, *Bacillus subtilis*, *Salmonella typhimurium*, Streptomycete, or any prokaryotic cell known in the art to be suitable for expressing, and optionally isolating, a protein domain disclosed herein.

A recombinant host cell can be cultured in a conventional nutrient medium modified as appropriate for activating promoters, selecting transformants, or amplifying particular genes. The culture conditions for particular host cells selected for expression, such as temperature, pH and the like, will be readily apparent to those of skill in the art.

Those of skill in the art will appreciate that other well known methods may be used to produce a STED including, but not limited to, chemical synthesis and *in vitro* translation systems.

Variant STEDs, Polypeptides and Polynucleotides

The compositions and methods of the present invention also encompass variants of the above protein domains, polypeptides and polynucleotides.

As used herein, the term "variant" comprehends nucleotide or amino acid sequences different from the specifically identified sequences, wherein one or more nucleotides or amino acid residues is deleted, substituted, or added. Variants may be naturally occurring allelic variants, or non-naturally occurring variants. Variant sequences (polynucleotide or polypeptide) preferably exhibit at least 90%, 95%, 96%, 97%, 98% or 99% identity to a sequence of the present invention. In one embodiment, variant STED sequences are provided wherein the T cell epitope has from zero to three amino acid changes.

The percentage identity is determined by aligning the two sequences to be compared as described below, determining the number of identical residues in the aligned portion,

dividing that number by the total number of residues in the inventive (queried) sequence, and multiplying the result by 100. In addition to exhibiting the recited level of sequence identity, variant sequences provided herein preferably exhibit a functionality that is substantially similar to the functionality of the parent, or reference, sequence. Variant STED sequences thus preferably retain the immunogenic properties of the STEDs disclosed herein. Preferably a variant STED sequence will generate at least 80%, at least 90%, at least 95% or 100% of the response generated by the specifically identified STED in an assay, such as one of the assays described below. Such variants may generally be identified by modifying the parent STED sequence, and evaluating the antigenic and/or diagnostic properties of the modified polypeptide using, for example, the representative procedures described herein.

Variant sequences generally differ from the specifically identified sequence only by conservative substitutions, deletions or modifications. As used herein, a "conservative substitution" is one in which an amino acid is substituted for another amino acid that has similar properties, such that one skilled in the art of peptide chemistry would expect the secondary structure and hydropathic nature of the polypeptide to be substantially unchanged. In general, the following groups of amino acids represent conservative changes: (1) ala, pro, gly, glu, asp, gln, asn, ser, thr; (2) cys, ser, tyr, thr; (3) val, ile, leu, met, ala, phe; (4) lys, arg, his; and (5) phe, tyr, trp, his. Variants may also, or alternatively, contain other modifications, including the deletion or addition of amino acids that have minimal influence on the antigenic properties, secondary structure and hydropathic nature of the polypeptide. For example, a polypeptide may be conjugated to a signal (or leader) sequence at the N-terminal end of the protein which co-translationally or post-translationally directs transfer of the protein. The polypeptide may also be conjugated to a linker or other sequence for ease of synthesis, purification or identification of the polypeptide (*e.g.*, poly-His), or to enhance binding of the polypeptide to a solid support. For example, a polypeptide may be conjugated to an immunoglobulin Fc region.

Polypeptide and polynucleotide sequences may be aligned, and percentages of identical amino acids or nucleotides in a specified region may be determined against another polypeptide or polynucleotide, using computer algorithms that are publicly available. Two exemplary algorithms for aligning and identifying the identity of polynucleotide sequences are the BLASTN and FASTA algorithms. The alignment and identity of polypeptide sequences may be examined using the BLASTP and algorithm. BLASTX and FASTX algorithms compare nucleotide query sequences translated in all reading frames against polypeptide sequences. The FASTA and FASTX algorithms are described in Pearson and

Lipman, *Proc. Natl. Acad. Sci. USA* 85:2444-2448, 1988; and in Pearson, *Methods in Enzymol.* 183:63-98, 1990. The FASTA software package is available from the University of Virginia, Charlottesville, VA 22906-9025. The FASTA algorithm, set to the default parameters described in the documentation and distributed with the algorithm, may be used in the determination of polynucleotide variants. The readme files for FASTA and FASTX Version 2.0x that are distributed with the algorithms describe the use of the algorithms and describe the default parameters.

The BLASTN software is available on the NCBI anonymous FTP server and is available from the National Center for Biotechnology Information (NCBI), National Library of Medicine, Building 38A, Room 8N805, Bethesda, MD 20894. The BLASTN algorithm Version 2.0.6 [Sep-10-1998] and Version 2.0.11 [Jan-20-2000] set to the default parameters described in the documentation and distributed with the algorithm, is preferred for use in the determination of variants according to the present invention. The use of the BLAST family of algorithms, including BLASTN, is described at NCBI's website and in the publication of Altschul, *et al.*, "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs," *Nucleic Acids Res.* 25:3389-3402, 1997.

The "hits" to one or more database sequences by a queried sequence produced by BLASTN, BLASTP, FASTA, or a similar algorithm, align and identify similar portions of sequences. The hits are arranged in order of the degree of similarity and the length of sequence overlap. Hits to a database sequence generally represent an overlap over only a fraction of the sequence length of the queried sequence.

The percentage identity of a polynucleotide or polypeptide sequence is determined by aligning polynucleotide and polypeptide sequences using appropriate algorithms, such as BLASTN or BLASTP, respectively, set to default parameters; identifying the number of identical nucleic or amino acids over the aligned portions; dividing the number of identical nucleic or amino acids by the total number of nucleic or amino acids of the polynucleotide or polypeptide of the present invention; and then multiplying by 100 to determine the percentage identity.

In other embodiments, variant polynucleotides hybridize to a parent, or reference, sequence under stringent hybridization conditions. The polynucleotide variants retain the capacity to encode a protein domain having the same functionality as that of the protein domain encoded by the parent, or reference, polynucleotide. In certain embodiments, such variant polynucleotides encode a polypeptide having the same amino acid sequence as a protein domain disclosed herein.

The term "stringent" is used to refer to conditions that are commonly understood in the art as stringent. Hybridization stringency is principally determined by temperature, ionic strength, and the concentration of denaturing agents such as formamide. Examples of stringent conditions for hybridization and washing are 0.015M sodium chloride, 0.0015M sodium citrate at about 65-68°C or 0.015M sodium chloride, 0.0015M sodium citrate, and 50% formamide at about 42°C (*see* Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. 1989).

Testing of STED Activity

The STEDs disclosed herein will generally have the same functional activity as the T cell epitope which they contain. CD4 epitopes are incorporated in the STED in such a way that they are efficiently processed and displayed on MHC Class II molecules after uptake by dendritic cells. Similarly, CD8 epitopes are incorporated into the STED in such a way that they are efficiently processed and displayed on MHC Class I molecules, especially by targeting the STED to appropriate receptors on antigen presenting cells (Caminschi et al, Eur. J. Immunol. 39:931 (2008); Keler et al., Oncogene 26, 3758 (2007)). Presentation of T cell epitopes on APCs can be verified experimentally using methods known to those of skill in the art. One such method is an *ex vivo* memory recall assay using human immune cells in which the antigen is first taken up by monocyte-derived DCs, which are then used to stimulate proliferation of pre-existing autologous T cells specific for T cell epitopes carried on the antigen (Lee et al., J. Clin. Invest. 118:3478-3490 (2008)). An assay that does not require pre-existing antigen-specific T cells is an *ex vivo* priming assay in which uptake of the antigen by human DCs is followed by long-term stimulation of autologous T cells, resulting in priming and proliferation of T cells specific for both CD4 and CD8 T cell epitopes (Colleton et al., J. Virol., 83:6288 (2009); Montagna et al., Cancer Res 66:7310 (2006); Rech et al., Med Immunol. 5:1 (2006)).

Compositions and Treatment Methods

In another aspect of this invention, compositions, such as immunogenic compositions, are provided for the prevention or treatment of a disorder, such as an infectious disease or a cancer, in a subject. Such compositions generally comprise one or more STEDs disclosed herein, or a pharmaceutically acceptable salt thereof, together with a physiologically acceptable carrier, excipient and/or diluent.

As used herein, the term "subject" refers to a mammal, preferably a human. The subject may be afflicted with a disorder or may be free of a detectable disorder. In other words, the methods may be employed for the prevention or treatment of a disorder.

The compositions of this disclosure may be in any form that allows for administration to a patient, such as, for example, in the form of a solid, liquid, or gas (aerosol). In one embodiment, the composition is in the form of a liquid, e.g., a solution, emulsion, suspension, elixir or syrup. The liquid may be formulated for delivery by injection or for oral administration, as two examples.

Pharmaceutically acceptable carriers, excipients and diluents for therapeutic use are well known in the pharmaceutical art, and include those described in Remington's Pharmaceutical Sciences, Mack Publishing Co. (A.R. Gennaro (Ed.) 1985). For example, sterile saline and phosphate buffered saline at physiological pH may be used. Preservatives, stabilizers, dyes and the like may also be provided in the composition. For example, sodium benzoate, sorbic acid, or esters of p-hydroxybenzoic acid may be added as preservatives. The disclosed compositions may also contain diluents such as buffers, antioxidants such as ascorbic acid, low molecular weight (less than about 10 residues) polypeptides, proteins, amino acids, carbohydrates (e.g., glucose, sucrose or dextrins), chelating agents (e.g., EDTA), glutathione and other stabilizers and excipients.

In a composition intended to be administered by injection, one or more of a surfactant, preservative, wetting agent, dispersing agent, suspending agent, buffer, stabilizer, isotonic agent, or any combination thereof may be included. For oral administration, an excipient and/or binder may be present, such as sucrose, kaolin, glycerin, starch dextrans, cyclodextrins, sodium alginate, carboxy methylcellulose, and/or ethyl cellulose. Sweetening agents, preservatives, dyes/colorants, flavor enhancers, or any combination thereof may optionally be present. A coating shell may also optionally be employed.

A liquid composition, whether in the form of a solution, suspension or other like form, may include one or more of the following components: sterile diluents such as water for injection, saline solution, preferably physiological saline, Ringer's solution, isotonic sodium chloride, fixed oils such as synthetic mono or digylcerides that may serve as the solvent or suspending medium, polyethylene glycols, glycerin, propylene glycol or other solvents; antibacterial agents such as benzyl alcohol or methyl paraben; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium, chloride, or dextrose. The liquid composition can be enclosed in ampoules,

disposable syringes or multiple dose vials made of glass or plastic. An injectable composition is preferably sterile.

It may also be desirable to include other components in the composition, such as delivery vehicles including aluminum salts, water-in-oil emulsions, biodegradable oil vehicles, oil-in-water emulsions, biodegradable microcapsules and liposomes. The composition may also contain an adjuvant or immunostimulant.

In one embodiment, a dosage unit comprising a composition disclosed herein is provided. Such dosage units include, for example, a single-dose or multi-dose vial or syringe, including a two-compartment vial or syringe, one comprising the composition of this disclosure in lyophilized form and the other a diluent for reconstitution. A multi-dose dosage unit can also be, for example, a bag or tube for connection to an intravenous infusion device.

The present disclosure also provides a kit comprising a composition disclosed herein in a unit dose, or multi-dose, container (e.g., a vial) and a set of instructions for administering the composition to a subject suffering from a disorder.

The compositions are generally administered to the subject in an amount that is effective to ameliorate symptoms of the disease following a course of one or more administrations. An effective dose is that dose required to prevent, inhibit the occurrence of, or treat (alleviate a symptom to some extent, preferably all symptoms of) a disease. The effective dose depends on the type of disease, the composition used, the route of administration, the type of subject being treated, the physical characteristics of the specific subject under consideration for treatment, concurrent medication, and other factors that those skilled in the medical arts will recognize. For example, an amount between 0.1 mg/kg and 100 mg/kg body weight (which can be administered as a single dose, daily, weekly, monthly, or at any appropriate interval) of active ingredient may be administered depending on the potency of a STED of this disclosure.

Routes of administration include intravenous, oral, topical, parenteral (*e.g.*, sublingually or buccally), sublingual, rectal, vaginal, and intranasal. The term parenteral as used herein includes subcutaneous injections, intravenous, intramuscular, intrasternal, intracavernous, intrathecal, intrameatal, intraurethral injection or infusion techniques. In particular embodiments, a protein domain of this disclosure is administered intravenously by, for example, bolus injection or infusion. The composition is formulated so as to allow the active ingredients contained therein to be bioavailable upon administration of the composition to a patient. Compositions that will be administered to a patient take the form of one or more dosage units, where for example, a tablet may be a single dosage unit, and a

container of one or more compounds of this disclosure in aerosol form may hold a plurality of dosage units.

The compositions disclosed herein may be administered in combination (either sequentially or simultaneously) with other known therapies or therapeutic agents. For example, the disclosed compositions may be employed in conjunction with cytokines, growth factors, steroids, NSAIDs, DMARDs, chemotherapeutics, radiotherapeutics, or a combination thereof.

In yet further embodiments, methods are provided for treating autoimmune or allergic diseases, such as multiple sclerosis, type I diabetes, rheumatoid arthritis, allergies to cat or bee proteins, or organ transplantation. This is achieved by, for example, targeting dendritic cells (DC) with a STED protein carrying T cell epitopes corresponding to a particular autoimmune disease in the absence of agents that activate or mature DCs, thus generating T cell tolerance to the disease-specific epitopes contained within the STED. Such an approach has been described using proteins or peptides targeted to DCs (Finkelman et al., J. Immunol. 157:1406 (1996); Caminschi et al., Eur. J. Immunol. 39:931 (2009); Larche and Wraith, Nat. Med. 11:S69 (2005); Miller et al., Nat. Rev. Immunol. 7:665 (2007); Chiang et al., J Immunol. 182:4282 (2009)).

It will be appreciated that the methods and systems of the present invention may be embodied in a variety of different forms, and that the specific embodiments shown in the figures and described herein are presented with the understanding that the present disclosure is considered exemplary of the principles of the invention, and is not intended to limit the invention to the illustrations and description provided herein.

I claim:

1. A polypeptide comprising at least one CD4 or CD8 T cell epitope from a proteome of an infectious organism or from a tumor associated antigen, and a heterologous non-human-like scaffold amino acid sequence, wherein the polypeptide is non-aggregated, stable and well-folded when expressed in a recombinant expression system.

2. A polynucleotide that encodes a polypeptide according to claim 1.

3. An expression vector comprising a polynucleotide according to claim 2.

4. A host cell transformed or transfected with an expression vector of claim 3.

5. A composition comprising at least one polypeptide according to claim 1 and at least one physiologically acceptable carrier.

6. The composition of claim 5, wherein the composition comprises at least two polypeptides according to claim 1, the at least two polypeptides being connected by a linker molecule.

7. The composition of claim 5, wherein the at least one polypeptide is connected to a targeting molecule.

8. The composition of claim 7, wherein the targeting molecule is an antibody, or antigen-binding fragment thereof, that binds to a polypeptide expressed on the surface of dendritic cells.

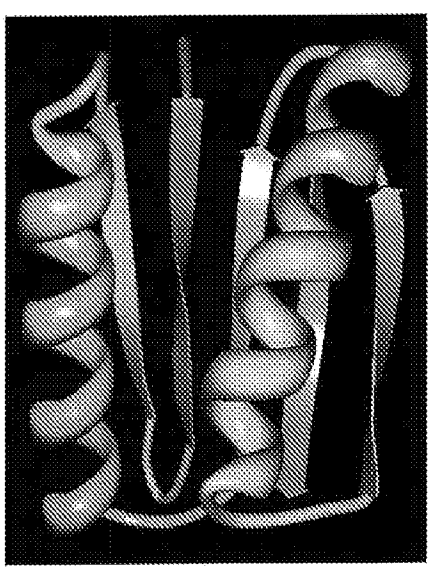
9. A method of treating or preventing a disorder in a subject, comprising administering to the subject an effective amount of a composition according to claim 5, wherein the disorder is an infectious disorder or a cancer.

10. The method of claim 9, wherein the infectious disorder is selected from the group consisting of: human immunodeficiency virus (HIV), influenza, herpes-viruses, hepatitis

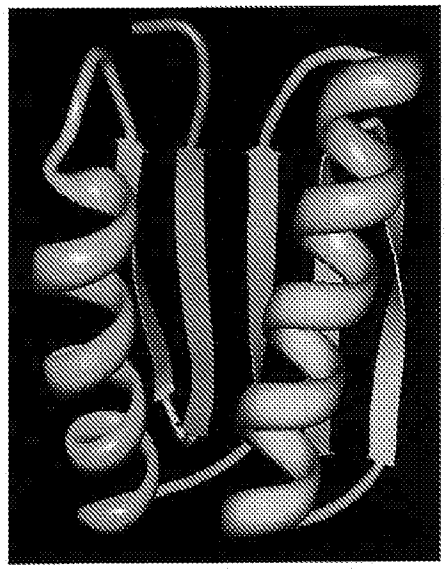
C, human T-lymphotropic virus (HTLV), Epstein-Barr virus (EBV), dengue viruses, malaria, and tuberculosis.

11. The method of claim 9, wherein the cancer is selected from the group consisting of: breast, colorectal, ovarian, prostate, bladder, gastric and non-small cell lung cancers.

Fig. 1: Proteins can be designed to have the highly similar structures but low sequence identity



TOP7 (PDB: 1QYS)



M7 (PDB: 2JVF)

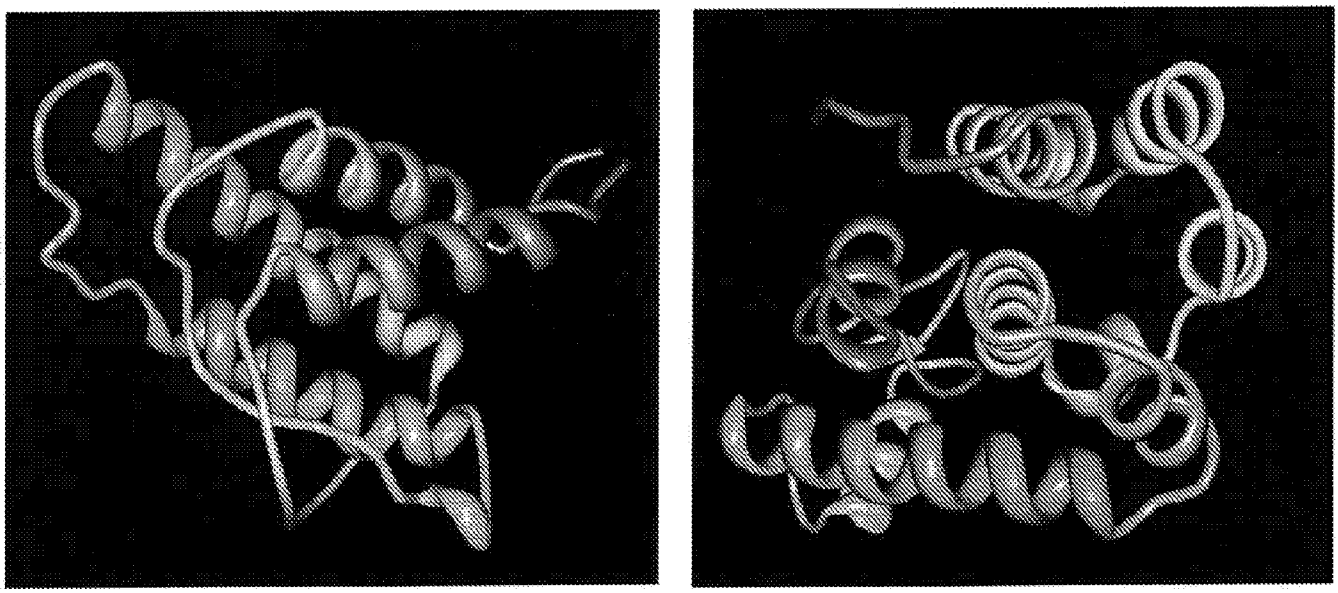
Example:
Two synthetic proteins, TOP7 and M7, both of which were designed using computational methods, have a nearly identical structure but have only 29% sequence identity

```
                  10        20        30        40        50
TOP7      DIQVQVNIDDNGKNFDYTYTVTTESELQKVLNELKDYIKKQGAKRVRISITARTKK
          ... ..:. .:.... .:f .:::..:.::. . . ::. :.:.:.:.. .
M7        GSHMKVDITIKIQRDGQEIRIDIRVSTGKELERALQELEKALARAGARNVQITISAENDE
                  10        20        30        40        50        60

            60        70        80        90
TOP7      EAEKFAAILIKVFAELGYNDINVTFDGDTVTVEGQLE
          .:... .. ... .:::.:::: .: : .:
M7        QAKELLELIARLLQKLGYKDINVRVNGTEVKIEVRV
                  70        80        90
```

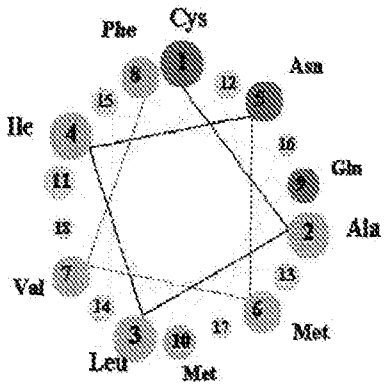
CONFIDENTIAL

Fig. 2: Four-helix bundle proteins – two examples

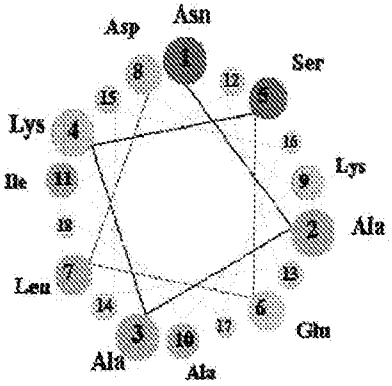


IL-4 (PDB 1ITL) T4 lysozyme (PDB 1LYD)

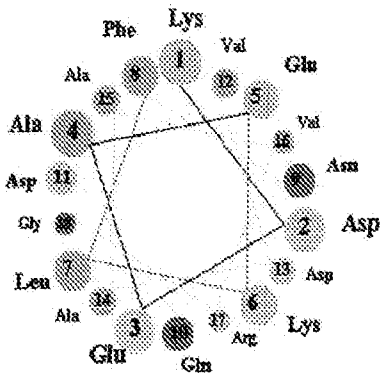
Fig. 3: Three alpha helices with distinct features in a four-helix bundle protein (T4 lysozyme, 164 residues)



Segment 97-106:
Mostly hydrophobic

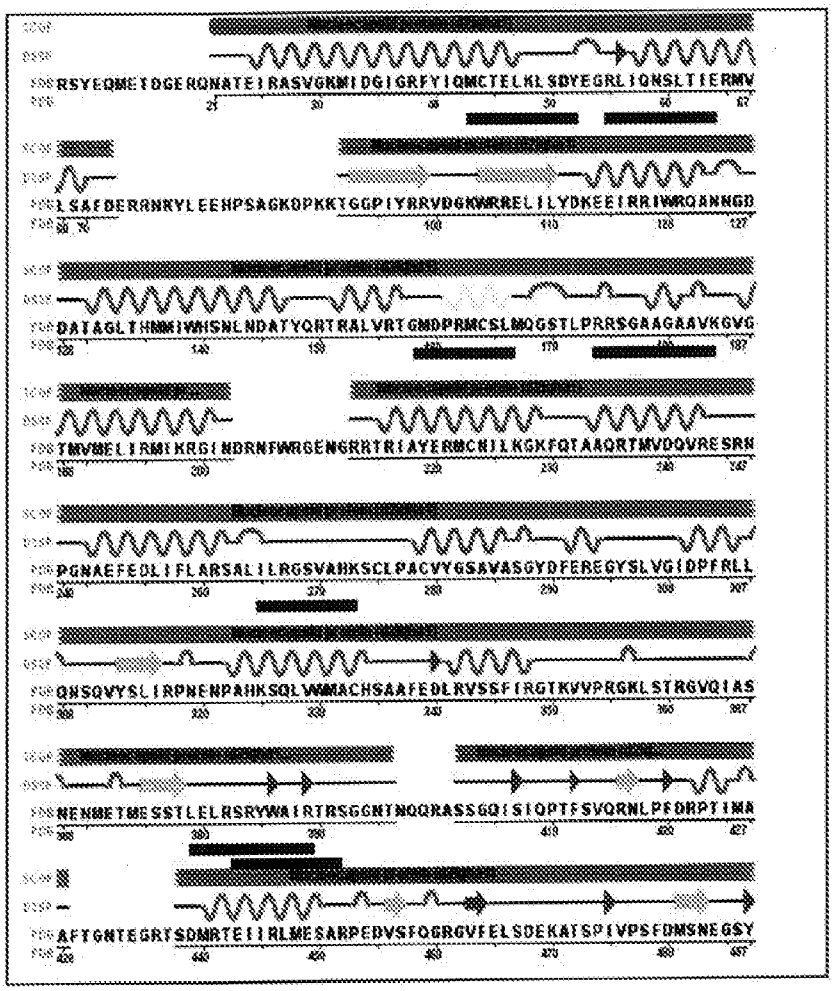


Segment 40-50:
Amphiphilic



Segment 60-79:
Mostly hydrophilic

Fig. 4: Conserved CD8 epitopes in influenza NP



Conserved CTL epitopes

Legend: Conserved CD8 (CTL) epitopes are indicated by bars under the corresponding sequence. CTL epitopes were identified using the immuneepitope database (www.immuneepitope.org), Heiny et al. (PLoS ONE, November 2007, Issue 11, e1190) and Lee et al. (J. Clinical Invest. (2008) 118:3478)

Fig. 5: Influenza NP has structural homology to L. lactis DpsB

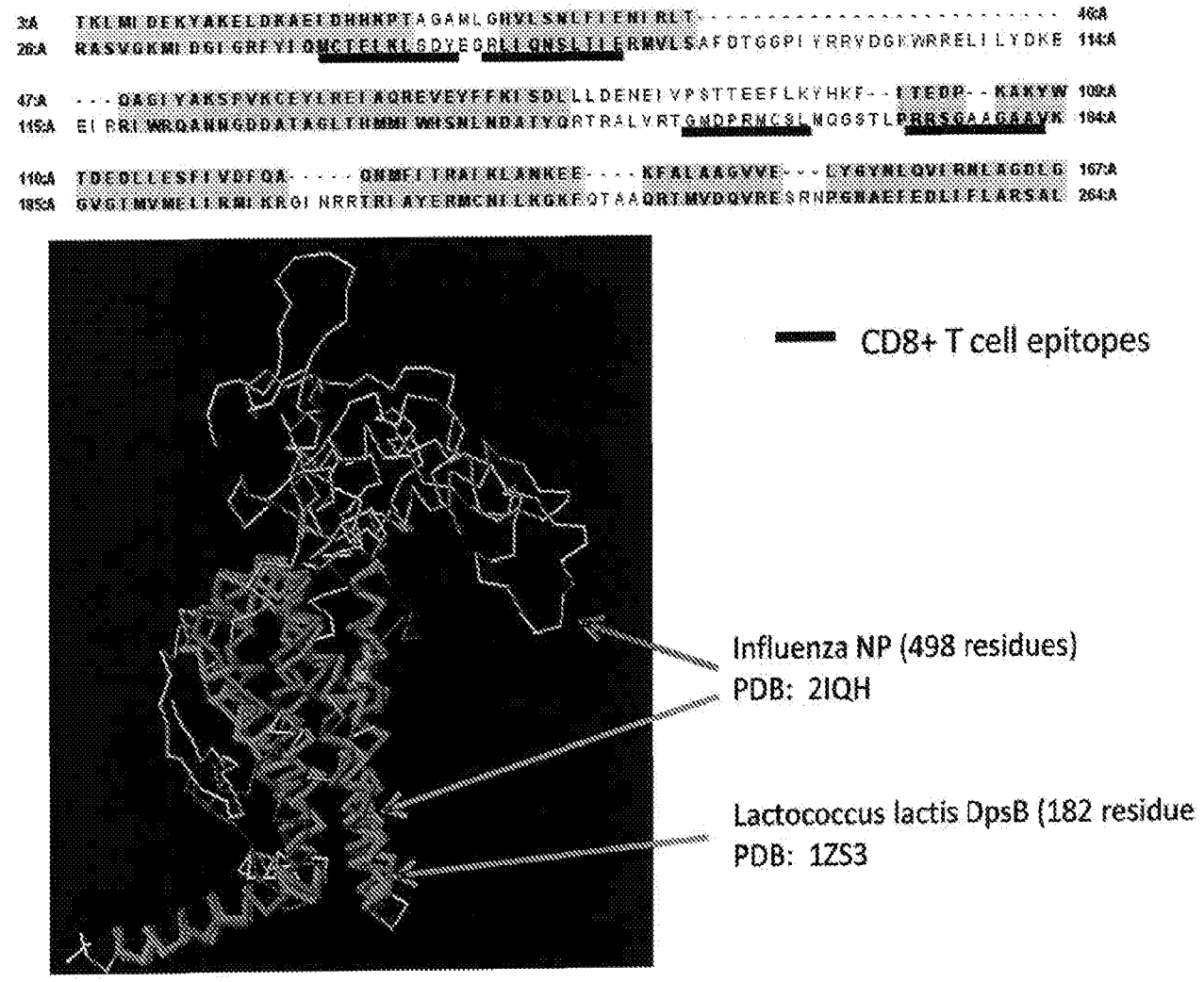
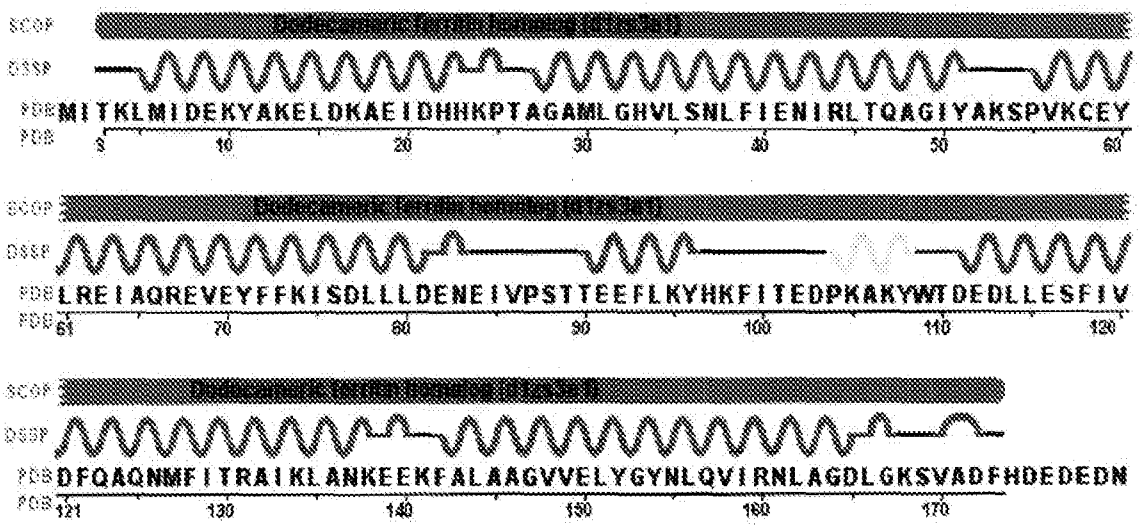


Fig. 6: Secondary structure of DpsB from L. lactis



INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 11/37049

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - A61K 38/00 (2011.01)
USPC - 530/300
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
USPC - 530/300

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
USPC - 514/21.3 (search terms below)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
PubWEST(PGPB,USPT,EPAB,JPAB); Google/Scholar: T cell epitope, epitope, fusion, chimeric, scaffold, solub$, non-aggreg$, aggreg$, protein, peptide, polypeptide, non human

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X ---- Y	Laplagne et al. Engineering of a polymeric bacterial protein as a scaffold for the multiple display of peptides. Proteins 2004, 57(4):820-828 (pg 1-9 in attached copy); Abstract, Table I, (pg 2, col 2, para 1 - pg 3, col 1, para 2), (pg 3, col 1, para 5), (pg 3, col 2, para 2), (pg 4, col 1, para 2), (pg 4, col 2, para 2), (pg 5, col 1, para 1)	1-5 -------- 7-11
Y	US 2007/0014807 A1 (Maida) 18 January 2007 (18.01.2007) para [0019]-[0020], [0025], [0087]-[0088], [0131]-[0132]	7-11
A	Goldbaum et al. The 18-kDa cytoplasmic protein of Brucella species - an antigen useful for diagnosis - is a lumazine synthase. J. Med Microbiol 1999, 48:833-839; Figure 3 - sequence a	1-5, 7-11
A,T	NCBI BLAST search of the amino acid sequence disclosed in Fig 3a of Goldbaum et al. [Performed and Retrieved from the Internet on 02 August 2011: <http://blast.ncbi.nlm.nih.gov/Blast.cgi>]	1-5, 7-11

☐ Further documents are listed in the continuation of Box C. ☐

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
02 August 2011 (02.08.2011)	17 AUG 2011
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Authorized officer: Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (July 2009)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 11/37049

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 6
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.