cysteine residues, and any two adjacent cysteines are separated by no more than 15 non-cysteine amino acid residues. In some embodiments, at least one cysteine residue is located within 9 amino acid residues from the N- or C-terminus of the CCD. In some embodiments, the CCD sequence has at least 90%, or 91%, or 92%, or 93%, or 94%, or 95%, or 96%, or 97%, or 98%, or 99% identity or is identical to a sequence selected from the sequence set forth in Table 6.

[0009a] In one aspect, the present disclosure provides a fusion protein comprising the CCD of any one of the previous claims, wherein the CCD is fused to an extended recombinant polypeptide (XTEN), wherein the XTEN is characterized in that:

a. it has a molecular weight that is at least 4-fold, at least 5-fold, at least 6-fold, at least 7-fold, at least 8-fold, at last 10-fold, at least 20-fold, or at least 30-fold greater than the molecular weight of the CCD;
b. has at least 90%, or at least 91%, or at least 92%, or at least 93%, or at least 94%, or at least 95%, or at least 96%, or at least 97%, or at least 98%, or at least 99% sequence identity, or is identical to a sequence selected from the group of sequences set forth in Table 10 and Table 11.

[0009b] In one aspect, the present disclosure provides a pharmaceutical composition, comprising the fusion protein as described herein and a pharmaceutically acceptable carrier.

[0009c] In one aspect, the present disclosure provides use of the pharmaceutical composition as described herein in the manufacture of a medicament for the treatment of a disease in a subject, wherein the disease is selected from the group consisting of breast cancer, ER/PR+ breast cancer, Her2+ breast cancer, triple-negative breast cancer, liver carcinoma, lung cancer, non-small cell lung cancer, colorectal cancer, esophageal carcinoma, fibrosarcoma, choriocarcinoma, ovarian cancer, cervical carcinoma, laryngeal carcinoma, endometrial carcinoma, hepatocarcinoma, gastric cancer, prostate cancer, renal cell carcinoma, Kaposi's sarcoma, astrocytoma, melanoma, squamous cell cancer, basal cell carcinoma, head and neck cancer, thyroid carcinoma, Wilm's tumor, urinary tract carcinoma, thecoma, arrhenoblastoma, glioblastomoa, pancreatic cancer, leukemia, acute myeloid leukemia (AML), chronic myeloid leukemia (PCML), acute lymphocytic leukemia (ALL), chronic lymphocytic leukemia (CLL), T-cell acute lymphoblastic leukemia, lymphoblastic disease, multiple myeloma, Hodgkin's lymphoma, non-Hodgkin's lymphoma, acne vulgaris, asthma, autoimmune diseases, autoinflammatory diseases, celiac disease, chronic prostatitis, glomerulonephritis, hypersensitivity reaction, inflammatory bowel disease, Crohn's disease, pelvic inflammatory disease, reperfusion injury, rheumatoid arthritis, sarcoidosis, transplant rejection, vasculitis, psoriasis, fibromyalgia, irritable bowel syndrome, lupus erythematosis, osteoarthritis, scleroderma, and ulcerative colitis.

[0009d] In one aspect, the present disclosre provides an isolated nucleic acid comprising a polynucleotide sequence selected from (a) a polynucleotide encoding the CCD as described herein or the fusion protein as described herein and (b) the complement of the polynucleotide of (a).

[0009e] In one aspect, the present disclosure provides an expression vector comprising the polynucleotide sequence as described herein and a recombinant regulatory sequence operably linked to the polynucleotide sequence.

[0009f] In one aspect, the present disclosure provides an isolated prokaryotic host cell, comprising the expression vector as described herein.

In one aspect, the present disclosure provides a fusion protein comprising any CCD disclosed herein. In some embodiments, the fusion protein comprises the CCD fused to an extended recombinant polypeptide (XTEN), wherein the XTEN is characterized in that: (a) it has a molecular weight that is at least 4-fold, at least 5-fold, at least 6-fold, at least 7-fold, at least 8-fold, at last 10-fold, at least 20-fold, or at least 30-fold greater than the molecular weight of the CCD; (b) it has between 100 to about 1200 amino acids wherein at least 90%, or at least 91%, or at least 92%, or at least 93%, or at least 94%, or at least 95%, or at least 96%, or at least 97%, or at least 98%, or at least 99%, or 100% of the amino acid residues are selected from 4 to 6 types of amino acids selected from the group consisting of glycine (G), alanine (A), serine (S), threonine (T), glutamate (E) and proline (P); (c) it is substantially non-repetitive such that (1) the XTEN sequence contains no three contiguous amino acids that are identical unless the amino acids are serine, (2) at least 90% of the XTEN sequence consists of non-overlapping sequence motifs, each of which comprise 12 amino acid residues, wherein any two contiguous amino acid residues does not occur more than twice in each of the sequence motifs; or (3) the XTEN sequence has an average subsequence score of less than 3; (d) it has greater than 90% random coil formation as determined by GOR algorithm; (e) it has less than 2% alpha helices and 2% beta-sheets as determined by Chou-Fasman algorithm; and (f) it lacks a predicted T-cell epitope when analyzed by TEPITOPE algorithm, wherein the TEPITOPE algorithm prediction for epitopes within the XTEN sequence is based on a threshold score of –9. In some embodiments, the sequence motifs are selected from the group consisting of the sequences set forth in Table 9. In some embodiments, the XTEN has at least 90% sequence identity, or at least 91%, or at least 92%, or at least 93%, or at least 94%, or at least 95%, or at least 96%, or at least 97%, or at least 98%, or at least 99% sequence identity, or is identical to a sequence selected from the group of sequences set forth in Table 10 or Table 11. In some embodiments, the fusion protein further comprises at least a first targeting moiety (TM) wherein the targeting moiety is capable of specifically binding a ligand associated with a target tissue. In some embodiments, the TM is joined to the N-terminus or the C-terminus of the CCD. In some embodiments, the fusion protein is configured from the N-terminus to the C-terminus as: (a) (TM)-(CCD)-(XTEN); or (b) (XTEN)-(CCD)-(TM). In some embodiments, the TM is fused to the CCD recombinantly. In some embodiments, the TM is conjugated to the CCD using a linker sequence selected from the group consisting of the sequences set forth in Table 12. In some embodiments, the ligand of the target tissue is associated with a tumor, a cancer cell, or a tissue with an inflammatory condition. In some embodiments, the fusion protein further comprises one or more drugs or biologically active proteins, wherein each drug or biologically

WHAT IS CLAIMED IS:

1. A cysteine containing domain (CCD) comprising at least 6 amino acid residues, wherein the domain is characterized in that:

a. it has at least one cysteine residue;

b. it has at least one non-cysteine residue, and at least 90%, or at least 91%, or at least 92%, or at least 93%, or at least 94%, or at least 95%, or at least 96%, or at least 97%, or at least 98%, or at least 99%, or 100% of the non-cysteine residues are selected from 3 to 6 types of amino acids selected from the group consisting of glycine (G), alanine (A), serine (S), threonine (T), glutamate (E) and proline (P);

c. no three contiguous amino acids are identical unless the amino acid is cysteine or serine; and

d. no glutamate residue is adjacent to a cysteine residue.

2. The CCD of claim 1, wherein the CCD has between 6 to about 144 amino acid residues and between 1 to about 10 cysteine residues.

3. The CCD of claim 1 or claim 2, wherein at least one cysteine residue is located within 9 amino acid residues from the N- or C-terminus of the CCD.

4. The CCD of claim 1, wherein the CCD sequence has at least 90%, or 91%, or 92%, or 93%, or 94%, or 95%, or 96%, or 97%, or 98%, or 99% identity or is identical to a sequence selected from the sequences set forth in Table 6.

5. A fusion protein comprising the CCD of any one of the previous claims, wherein the CCD is fused to an extended recombinant polypeptide (XTEN), wherein the XTEN is characterized in that:

a. it has a molecular weight that is at least 4-fold, at least 5-fold, at least 6-fold, at least 7-fold, at least 8-fold, at last 10-fold, at least 20-fold, or at least 30-fold greater than the molecular weight of the CCD;

b. has at least 90%, or at least 91%, or at least 92%, or at least 93%, or at least 94%, or at least 95%, or at least 96%, or at least 97%, or at least 98%, or at least 99% sequence identity, or is identical to a sequence selected from the group of sequences set forth in Table 10 and Table 11.

6. The fusion protein of claim 5, further comprising at least a first targeting moiety (TM), wherein the targeting moiety is capable of specifically binding a ligand associated with a target tissue.

7. The fusion protein of claim 6, wherein the TM is joined to the N-terminus or the C- terminus of the CCD.

8. The fusion protein of claim 6, wherein the ligand of the target tissue is associated with a tumor, a cancer cell, or a tissue with an inflammatory condition.

9. The fusion protein of claim 6, further comprising one or more drugs or biologically active proteins, wherein each drug or biologically active protein is conjugated to a thiol group of a cysteine residue of the CCD.

10. The fusion protein of claim 9, wherein the drug is a cytotoxic drug selected from the group consisting of doxorubicin, nemorubicin, PNU-159682, paclitaxel, docetaxel, auristatin E, auristatin F, dolastatin 10, dolastatin 15, monomethyl auristatin E (MMAE), monomethyl auristatin F (MMAF), monomethyl auristatin D (MMAD), maytansine, mertansine (DM1), maytansinoid DM4, calicheamicin, N-acetyl-calicheamicin, vinblastine, vincristine, vindesine, vinorelbine, camptothecin, topotecan, irinotecan, SN-38, duocarmycin A, duocarmycin B1, duocarmycin B2, duocarmycin C1, duocarmycin C2, duocarmycin D, duocarmycin SA, duocarmycin TM, duocarmycin MB, duocarmycin DM, mitomycin C, rachelmycin, epothilone A, epothilone B, epothilone C, tubulysin B, tubulysin M, pyrrolobenzodiazepine (PBD), bortezomib, hTNF, IL-12, ranpirnase, and Pseudomonas exotoxin A.

11. The fusion protein of claim 10, wherein the drug is monomethyl auristatin E (MMAE) or monomethyl auristatin F (MMAF).

12. The fusion protein of claim 10, wherein the drug is mertansine (DM1).

13. The fusion protein of claim 9, wherein the target tissue is a tumor or a cancer cell and the biologically active protein is selected from the group consisting of TNFα, IL-12, ranpirnase, human ribonuclease (RNAse), bovine pancreatic RNase, pokeweed antiviral protein, Pseudomonas exotoxin A, gelonin, ricin-A, interferon-alpha, interferon-lambda, urease, amatoxin, alpha-amanitin, beta-amanitin, gamma-amanitin, epsilon-amanitin, bouganin, and staphylococcal enterotoxin.

14. The fusion protein of any one of claims 6-13, wherein the at least first TM is selected from the group consisting of an IgG antibody, a Fab fragment, a F(ab')2 fragment, a scFv, a scFab, a dAb, a single domain heavy chain antibody, and a single domain light chain antibody.

15. The fusion protein of claim 14, wherein the at least first targeting moiety is an antibody.

16. The fusion protein of claim 15, wherein the antibody comprises heavy chain CDR segments HCDR1, HCDR2, HCDR3, light chain CDR segments LCDR1, LCDR2, LCDR3, and framework regions (FR) from an antibody, wherein the heavy chain CDR and FR are fused together in the order FR1-HCDR1-FR2-HCDR2-FR3-HCDR3-FR4 and the light chain CDR and FR are fused together in the order FR1-LCDR1-FR2-LCDR2-FR3-LCDR3-FR4.

17. The fusion protein of any one of claims 5-16, further comprising a peptidic cleavage moiety (PCM), wherein the PCM is a capable of being cleaved by one, two, or more mammalian proteases.

18. The fusion protein of any one of claims 5-17, further comprising a peptidic cleavage moiety (PCM), wherein the PCM is a capable of being cleaved by one, two, or more mammalian proteases, and wherein the fusion protein is configured from the N-terminus to the C-terminus as:

a. (TM)-(CCD)-(PCM)-(XTEN);

b. (XTEN)-(PCM)-(CCD)-(TM);

c. (XTEN)-(PCM)-(TM)-(CCD); or

d. (CCD)-(TM)-(PCM)-(XTEN).

19. The fusion protein of any one of claims 9-18, wherein upon performing a conjugation reaction between the drug molecule and the cysteine residues of the CCD of the fusion protein, a heterogeneous population of conjugate products is obtained, wherein fully conjugated CCD-drug conjugate product is capable of achieving a peak separation ≥ 6, wherein: a) the fusion protein comprises a polypeptide having 600 or more cumulative amino acid residues comprising a CCD with between 3 to 9 cysteine residues; b) the heterogeneous conjugate products have a mixture of at least 1, 2, and 3 or more payloads linked to the CCD; and c) the conjugation products are analyzed under reversed-phase HPLC chromatography conditions.

20. A pharmaceutical composition, comprising the fusion protein of any one of claims 9-19 and a pharmaceutically acceptable carrier.

21. Use of the pharmaceutical composition of claim 20 in the manufacture of a medicament for the treatment of a disease in a subject, wherein the disease is selected from the group consisting of breast cancer, ER/PR+ breast cancer, Her2+ breast cancer, triple-negative breast cancer, liver carcinoma, lung cancer, non-small cell lung cancer, colorectal cancer, esophageal carcinoma, fibrosarcoma, choriocarcinoma, ovarian cancer, cervical carcinoma, laryngeal carcinoma, endometrial carcinoma, hepatocarcinoma, gastric cancer, prostate cancer, renal cell carcinoma, Kaposi's sarcoma, astrocytoma, melanoma, squamous cell cancer, basal cell carcinoma, head and neck cancer, thyroid carcinoma, Wilm's tumor, urinary tract carcinoma, thecoma, arrhenoblastoma, glioblastomoa, pancreatic cancer, leukemia, acute myeloid leukemia (AML), chronic myeloid leukemia (PCML), acute lymphocytic leukemia (ALL), chronic lymphocytic leukemia (CLL), T-cell acute lymphoblastic leukemia, lymphoblastic disease, multiple myeloma, Hodgkin's lymphoma, non-Hodgkin's lymphoma, acne vulgaris, asthma, autoimmune diseases, autoinflammatory diseases, celiac disease, chronic prostatitis, glomerulonephritis, hypersensitivity reaction, inflammatory bowel disease, Crohn's disease, pelvic inflammatory disease, reperfusion injury, rheumatoid arthritis, sarcoidosis, transplant rejection, vasculitis, psoriasis, fibromyalgia, irritable bowel syndrome, lupus erythematosis, osteoarthritis, scleroderma, and ulcerative colitis.

22. An isolated nucleic acid comprising a polynucleotide sequence selected from (a) a polynucleotide encoding the CCD of any one of claims 1-4 or the fusion protein of any one of claims 5-8 and (b) the complement of the polynucleotide of (a).

23. An expression vector comprising the polynucleotide sequence of claim 22 and a recombinant regulatory sequence operably linked to the polynucleotide sequence.

24. An isolated prokaryotic host cell, comprising the expression vector of claim 23.

25. The host cell of claim 24, wherein the host cell is *E. coli*.