Claims

1. A nucleic acid molecule, wherein said nucleic acid molecule comprises a nucleotide sequence encoding a polypeptide consisting of proline, alanine and, optionally, serine, wherein said polypeptide forms a random coil,
wherein the nucleotide sequence of said nucleic acid has a length of at least 300 nucleotides,
wherein said nucleotide sequence has a Nucleotide Repeat Score (NRS) lower than 1000,
wherein said Nucleotide Repeat Score (NRS) is determined according to the formula:

$$NRS = \frac{\sum_{n=4}^{N_{tot}-1} n^2 \sqrt{\sum_{i=1}^{k(n)} f_i(n)}}{N_{tot}},$$

wherein
N_{tot} is the length of said nucleotide sequence,
n is the length of a repeat within said nucleotide sequence, and
$f_i(n)$ is the frequency of said repeat of length n,
wherein if there is more than one repeat of length n, k(n) is the number of said different sequences of said repeat of length n, otherwise k(n) is 1 for said repeat of length n.

2. The nucleic acid molecule of claim 1, wherein said encoded polypeptide consists of proline and alanine.

3. The nucleic acid molecule of claim 2, wherein said proline residues constitute more than about 10 % and less than about 75 % of said encoded polypeptide.

4. The nucleic acid molecule of claim 1, wherein said encoded polypeptide consists of proline, alanine and serine.

5. The nucleic acid molecule of claim 4, wherein said proline residues constitute more than 4% and less than 40% of said encoded polypeptide.

6. The nucleic acid molecule of any one of claims 1 to 5, wherein said Nucleotide Repeat Score (NRS) is lower than 100.

7. The nucleic acid molecule of any one of claims 1 to 6, wherein said Nucleotide Repeat Score (NRS) is lower than 50.

8. The nucleic acid molecule of any one of claims 1 to 7, wherein said Nucleotide Repeat Score (NRS) is lower than 35.

9. The nucleic acid molecule of any one of claims 1 to 8, wherein the nucleotide sequence of said nucleic acid has a length of at least 900 nucleotides.

10. The nucleic acid molecule of any one of claims 1 to 9, wherein said nucleic acid molecule has an enhanced genetic stability.

11. The nucleic acid molecule of any one of claims 1 to 10, wherein said nucleotide sequence comprises said repeats, wherein said repeats have a maximum length n_{max}, wherein n_{max} is determined according to the formula:

$$n_{\max} \leq 17 + \frac{N_{tot}}{600}$$

and wherein N_{tot} is the length of said nucleotide sequence.

12. The nucleic acid molecule of any one of claims 1 to 11, wherein said repeats have a maximum length of about 14, 15, 16, or 17 nucleotides to about 55 nucleotides.

13. The nucleic acid molecule of any one of claims 1 to 12, wherein said encoded polypeptide comprises a repetitive amino acid sequence with a plurality of amino acid repeats, wherein no more than 9 consecutive amino acid residues are identical and wherein said polypeptide forms a random coil.

14. The nucleic acid molecule of any one of claims 1 to 3 and 6 to 13, wherein said nucleic acid molecule is selected from the group consisting of:

(a) the nucleic acid molecule comprising at least one nucleotide sequence selected from the group consisting of SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 87, SEQ ID NO: 88, SEQ ID NO: 89, SEQ ID NO: 90, SEQ ID NO: 91, SEQ ID NO: 92, SEQ ID NO: 93, SEQ ID NO: 94, SEQ ID NO: 95, SEQ ID NO: 96, SEQ ID NO: 97, SEQ ID NO: 98, SEQ ID NO: 99, SEQ ID NO: 100, SEQ ID NO: 101, SEQ ID NO: 102, SEQ ID NO: 103, SEQ ID NO: 104, SEQ ID NO: 105, SEQ ID NO: 106, SEQ ID NO: 107, SEQ ID NO: 108, SEQ ID NO: 109, SEQ ID NO: 110, SEQ ID NO: 111, SEQ ID NO: 112, SEQ ID NO: 113, SEQ ID NO: 114, SEQ ID NO: 115, SEQ ID NO: 116, SEQ ID NO: 117, SEQ ID NO: 118, SEQ ID NO: 119, SEQ ID NO: 120, SEQ ID NO: 121, SEQ ID NO: 122, SEQ ID NO: 192 and SEQ ID NO: 193;

(b) the nucleic acid molecule comprising the nucleotide sequence consisting of SEQ ID NO: 42, SEQ ID NO: 43, SEQ ID NO: 44, SEQ ID NO: 45, SEQ ID NO: 153, SEQ ID NO: 154, SEQ ID NO: 155, SEQ ID NO: 156, SEQ ID NO: 157, SEQ ID NO: 158, SEQ ID NO: 159, SEQ ID NO: 160, SEQ ID NO: 161, SEQ ID NO: 162, SEQ ID NO: 163, SEQ ID NO: 164, SEQ ID NO: 165, SEQ ID NO: 166, SEQ ID NO: 167, SEQ ID NO: 168, SEQ ID NO: 169, SEQ ID NO: 170, SEQ ID NO: 171, SEQ ID NO: 172, and/or SEQ ID NO: 173;

(c) the nucleic acid molecule hybridizing under stringent conditions to the complementary strand of the nucleotide sequence as defined in (a) or (b);

(d) the nucleic acid molecule comprising the nucleotide sequence having at least 66.7% identity to the nucleotide sequence as defined in any one of (a), (b) and (c); and

(e) the nucleic acid molecule being degenerate as a result of the genetic code to the nucleotide sequence as defined in (a) or (b).

15. The nucleic acid molecule of any one of claims 1 and 4 to 13, wherein said nucleic acid molecule is selected from the group consisting of:

(a) the nucleic acid molecule comprising at least one nucleotide sequence selected from the group consisting of SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 123, SEQ ID NO: 124, SEQ ID NO: 125, SEQ ID NO: 126, SEQ ID NO: 127, SEQ ID NO: 128, SEQ ID NO: 129, SEQ ID NO: 130, SEQ ID NO: 131, SEQ ID NO: 132, SEQ ID NO: 133, SEQ ID NO: 134, SEQ ID NO: 135, SEQ ID NO: 136, SEQ ID NO: 137, SEQ ID NO: 138, SEQ ID NO: 139, SEQ ID NO: 140, SEQ ID NO: 141, SEQ ID NO: 142, SEQ ID NO: 143, SEQ ID NO: 144, SEQ ID NO: 145, SEQ ID NO: 146, SEQ ID NO: 147, SEQ ID NO: 148, SEQ ID NO: 149, SEQ ID NO: 150, SEQ ID NO: 151, SEQ ID NO: 152, SEQ ID NO: 194 and SEQ ID NO: 195;

(b) the nucleic acid molecule comprising the nucleotide sequence selected from the group consisting of SEQ ID NO: 38, SEQ ID NO: 39, SEQ ID NO: 40, SEQ ID NO: 41, SEQ ID NO: 174, SEQ ID NO: 175, SEQ ID NO: 176, SEQ ID NO: 177, SEQ ID NO: 178, SEQ ID NO: 179, SEQ ID NO: 180, SEQ ID NO: 181, SEQ ID NO: 182, SEQ ID NO: 184, SEQ ID NO: 185, SEQ ID NO: 186, SEQ ID NO: 187, SEQ ID NO: 188, SEQ ID NO: 189, SEQ ID NO: 190, and SEQ ID NO: 191;

(c) the nucleic acid molecule hybridizing under stringent conditions to the complementary strand of the nucleotide sequence as defined in (a) or (b);

(d) the nucleic acid molecule comprising the nucleotide sequence having at least 56% identity to the nucleotide sequence as defined in any one of (a), (b) and (c);

(e) the nucleic acid molecule being degenerate as a result of the genetic code to the nucleotide sequence as defined in (a) or (b).

16. The nucleic acid molecule of any one of claims 1 to 15 operably linked in the same reading frame to a nucleic acid encoding a biologically active protein.

17. The nucleic acid molecule of claim 16, wherein said biologically active protein is a therapeutically effective protein.

18. The nucleic acid molecule of claim 16 or 17, wherein said biologically active protein is selected from the group consisting of a binding protein, an antibody fragment, a cytokine, a growth factor, a hormone, an enzyme, a protein vaccine, a peptide vaccine, a peptide which consists of up to 50 amino acid residues or a peptidomimetic.

19. The nucleic acid molecule of claim 18, wherein said binding protein is selected from the group consisting of antibodies, Fab fragments, Fab' fragments, $F(ab')_2$ fragments, single chain variable fragments (scFv), (single) domain antibodies, isolated variable regions of antibodies (VL and/or VH regions), CDRs, immunoglobulin domains, CDR-derived peptidomimetics, lectins, protein scaffolds, fibronectin domains, tenascin domains, protein A domains, SH3 domains, ankyrin repeat domains, and lipocalins.

20. The nucleic acid molecule of any one of claims 16 to 18, wherein said biologically active protein is selected from the group consisting of interleukin 1 receptor antagonist, leptin, acid sphingomyelinase, adenosine deaminase, agalsidase alfa, alpha-1 antitrypsin, alpha atrial natriuretic peptide, alpha-galactosidase, alpha-glucosidase, alpha-N-acetylglucosaminidase, alteplase, amediplase, amylin, amylin analog, anti-HIV peptide fusion inhibitor, arginine deiminase, asparaginase, B domain deleted factor VIII, bone morphogenetic protein, bradykinin antagonist, B-type natriuretic peptide, bouganin, growth hormone, chorionic gonadotropin, CD3 receptor antagonist, CD19 antagonist, CD20 antagonist, CD40 antagonist, CD40L antagonist, cerebroside sulfatase, coagulation factor VIIa, coagulation factor XIII, coagulation factor IX, coagulation factor X, complement component C3 inhibitor, complement component 5a antagonist, C-peptide, CTLA-4 antagonist, C-type natriuretic peptide, defensin, deoxyribonuclease I, EGFR receptor antagonist, epidermal growth factor, erythropoietin, exendin-4, ezrin peptide 1, FcγIIB receptor antagonist, fibroblast growth factor 21, follicle-stimulating hormone, gastric inhibitory polypeptide (GIP), GIP analog, glucagon, glucagon receptor agonist, glucagon-like peptide 1 (GLP-1), GLP-1 analog, glucagon-like peptide 2 (GLP-2), GLP-2 analog, gonadorelin, gonadotropin-releasing hormone agonist, gonadotropin-releasing hormone antagonist, gp120, gp160, granulocyte colony stimulating factor (G-CSF), granulocyte macrophage colony

stimulating factor (GM-CSF), grehlin, grehlin analog, growth hormone, growth hormone-releasing hormone, hematide, hepatocyte growth factor, hepatocyte growth factor receptor (HGFR) antagonist, hepcidin antagonist, hepcidin mimetic, Her2/neu receptor antagonist, histrelin, hirudin, hsp70 antagonist, humanin, hyaluronidase, hydrolytic lysosomal glucocerebroside-specific enzyme, iduronate-2-sulfatase, IgE antagonists, insulin, insulin analog, insulin-like growth factor 1, insulin-like growth factor 2, interferon-alpha, interferon-alpha antagonist, interferon-alpha superagonist, interferon-alpha-n3, interferon-beta, interferon-gamma, interferon-lambda, interferon tau, interleukin, interleukin 2 fusion protein, interleukin-22 receptor subunit alpha (IL-22ra) antagonist, irisin, islet neogenesis associated protein, keratinocyte growth factor, Kv1.3 ion channel antagonists, lanthipeptide, lipase, luteinizing hormone, lutropin alpha, lysostaphin, mannosidase, N-acetylgalactosamine-6-sulfatase, N-acetylglucosaminidase, neutrophil gelatinase-associated lipocalin, octreotide, ω-conotoxin, Ornithodoros moubata complement inhibitor, osteogenic protein-1, osteoprotegerin, oxalate decarboxylase, P128, parathyroid hormone, Phylomer, PD-1 antagonist, PDGF antagonist, phenylalanine ammonia lyase, platelet derived growth factor, proinsulin, protein C, relaxin, relaxin analog, secretin, RGD peptide, ribonuclease, senrebotase, serine protease inhibitor, soluble complement receptor type 1, soluble DCC receptor, soluble TACI receptor, soluble tumor necrosis factor I receptor (sTNF-RI), soluble tumor necrosis factor II receptor (sTNF-RII), soluble VEGF receptor Flt-1, soluble FcγIIB receptor, somatostatin, somatostatin analog, streptokinase, T-cell receptor ligand, tenecteplase, teriparatide, thrombomodulin alpha, thymosin alpha 1, toll like receptor inhibitor, tumor necrosis factor (TNFα), tumor necrosis factor α antagonist, uricase, vasoactive intestinal peptide, vasopressin, vasopressin analog, VEGF antagonist, von Willebrand factor.

21. A vector comprising the nucleic acid molecule of any one of claims 1 to 20.

22. A host or host cell comprising the nucleic acid molecule of any one of claims 1 to 20, a host or host cell comprising the vector of claim 21, or a host or host cell transformed with a vector of claim 21.

23. A method for preparing the nucleic acid molecule of any one of claims 1 to 20, wherein the method comprises culturing the host or host cell of claim 22 and optionally isolating the produced nucleic acid molecule.

24. A method for preparing the vector of claim 21, wherein the method comprises culturing the host or host cell of claim 22 and optionally isolating the produced vector.

25. A method for preparing a polypeptide encoded by the nucleic acid molecule of any one of claims 1 to 20, wherein the method comprises culturing/raising the host or host cell of claim 22 and optionally isolating the produced polypeptide.

26. A method for preparing a drug conjugate, wherein said drug conjugate comprises the polypeptide encoded by the nucleic acid molecule of any one of claims 1 to 15 and further comprises (i) a biologically active protein and/or (ii) a small molecule and/or (iii) a carbohydrate, wherein the method further comprises culturing the host or host cell of claim 22 and optionally isolating the produced polypeptide and/or drug conjugate.

27. The method for preparing the drug conjugate of claim 26, wherein said biologically active protein is a therapeutically effective protein.

28. The method for preparing the drug conjugate of claim 26 or 27, wherein said biologically active protein is selected from the group consisting of a binding protein, an antibody fragment, a cytokine, a growth factor, a hormone, an enzyme, a protein vaccine, a peptide vaccine, a peptide which consists of up to 50 amino acid residues or a peptidomimetic.

29. The method for preparing the drug conjugate of claim 28, wherein said binding protein is selected from the group consisting of antibodies, Fab fragments, Fab' fragments, $F(ab')_2$ fragments, single chain variable fragments (scFv), (single) domain antibodies, isolated variable regions of antibodies (VL and/or VH regions), CDRs, immunoglobulin domains, CDR-derived peptidomimetics, lectins, protein scaffolds, fibronectin domains, tenascin domains, protein A domains, SH3 domains, ankyrin repeat domains, and lipocalins.

30. The method for preparing the drug conjugate of any one of claims 26 to 28, wherein said biologically active protein is selected from the group consisting of interleukin 1 receptor antagonist, leptin, acid sphingomyelinase, adenosine deaminase, agalsidase alfa, alpha-1 antitrypsin, alpha atrial natriuretic peptide, alpha-galactosidase, alpha-glucosidase, alpha-N-acetylglucosaminidase, alteplase, amediplase, amylin, amylin analog, anti-HIV peptide fusion inhibitor, arginine deiminase, asparaginase, B domain deleted factor VIII, bone morphogenetic protein, bradykinin antagonist, B-type natriuretic peptide, bouganin, growth hormone, chorionic gonadotropin, CD3 receptor antagonist, CD19 antagonist, CD20 antagonist, CD40 antagonist, CD40L antagonist, cerebroside sulfatase, coagulation factor VIIa, coagulation factor XIII, coagulation factor IX, coagulation factor X, complement component C3 inhibitor, complement component 5a antagonist, C-peptide, CTLA-4 antagonist, C-type natriuretic peptide, defensin, deoxyribonuclease I, EGFR receptor antagonist, epidermal growth factor, erythropoietin, exendin-4, ezrin peptide 1, FcγIIB receptor antagonist, fibroblast growth factor 21, follicle-stimulating hormone, gastric inhibitory polypeptide (GIP), GIP analog, glucagon, glucagon receptor agonist, glucagon-like peptide 1 (GLP-1), GLP-1 analog, glucagon-like peptide 2 (GLP-2), GLP-2 analog, gonadorelin, gonadotropin-releasing hormone agonist, gonadotropin-releasing hormone antagonist, gp120, gp160, granulocyte colony stimulating factor (G-CSF), granulocyte macrophage colony stimulating factor (GM-CSF), grehlin, grehlin analog, growth hormone, growth hormone-releasing hormone, hematide, hepatocyte growth factor, hepatocyte growth factor receptor (HGFR) antagonist, hepcidin antagonist, hepcidin mimetic, Her2/neu receptor antagonist, histrelin, hirudin, hsp70 antagonist, humanin, hyaluronidase, hydrolytic lysosomal glucocerebroside-specific enzyme, iduronate-2-sulfatase, IgE antagonists, insulin, insulin analog, insulin-like growth factor 1, insulin-like growth factor 2, interferon-alpha, interferon-alpha antagonist, interferon-alpha superagonist, interferon-alpha-n3, interferon-beta, interferon-gamma, interferon-lambda, interferon tau, interleukin, interleukin 2 fusion protein, interleukin-22 receptor subunit alpha (IL-22ra) antagonist, irisin, islet neogenesis associated protein, keratinocyte growth factor, Kv1.3 ion channel antagonists, lanthipeptide, lipase, luteinizing hormone, lutropin alpha, lysostaphin, mannosidase, N-acetylgalactosamine-6-sulfatase, N-

acetylglucosaminidase, neutrophil gelatinase-associated lipocalin, octreotide, ω-conotoxin, Ornithodoros moubata complement inhibitor, osteogenic protein-1, osteoprotegerin, oxalate decarboxylase, P128, parathyroid hormone, Phylomer, PD-1 antagonist, PDGF antagonist, phenylalanine ammonia lyase, platelet derived growth factor, proinsulin, protein C, relaxin, relaxin analog, secretin, RGD peptide, ribonuclease, senrebotase, serine protease inhibitor, soluble complement receptor type 1, soluble DCC receptor, soluble TACI receptor, soluble tumor necrosis factor I receptor (sTNF-RI), soluble tumor necrosis factor II receptor (sTNF-RII), soluble VEGF receptor Flt-1, soluble FcγIIB receptor, somatostatin, somatostatin analog, streptokinase, T-cell receptor ligand, tenecteplase, teriparatide, thrombomodulin alpha, thymosin alpha 1, toll like receptor inhibitor, tumor necrosis factor (TNFα), tumor necrosis factor α antagonist, uricase, vasoactive intestinal peptide, vasopressin, vasopressin analog, VEGF antagonist, von Willebrand factor.

31. The method for preparing the drug conjugate of claims 26 to 30, wherein said small molecule is selected from the group consisting of angiogenesis inhibitors, anti-allergic drugs, anti-emetic drugs, anti-depressant drugs, anti-hypertensive drugs, anti-inflammatory drugs, anti-infective drugs, anti-psychotic drugs, anti-proliferative (cytotoxic and cytostatic) drugs, calcium antagonists and other circulatory organ drugs, cholinergic agonists, drugs acting on the central nervous system, drugs acting on the respiratory system, hormones, steroids, polyketides, carbohydrates, oligosaccharides, nucleic acids, nucleic acid derivatives, antisense nucleic acids, small interference RNAs (siRNAs), micro RNA (miR) inhibitors, microRNA mimetics, DNA aptamers and RNA aptamers.

32. A method for selecting a genetically stable nucleic acid molecule, wherein said nucleic acid molecule comprises a nucleotide sequence encoding a polypeptide consisting of proline, alanine and, optionally, serine, wherein said polypeptide forms a random coil, wherein said nucleotide sequence has a length of at least 300 nucleotides,
the method comprising a step of selecting a nucleic acid molecule comprising a nucleotide sequence having a Nucleotide Repeat Score (NRS) lower than 1000,
wherein said Nucleotide Repeat Score (NRS) is determined according to the formula:

$$NRS = \frac{\sum_{n=4}^{N_{tot}-1} n^2 \sqrt{\sum_{i=1}^{k(n)} f_i(n)}}{N_{tot}},$$

wherein

N_{tot} is the length of said nucleotide sequence,

n is the length of a repeat within said nucleotide sequence, and

$f_i(n)$ is the frequency of said repeat of length n,

wherein if there is more than one repeat of length n, k(n) is the number of said different sequences of said repeat of length n, otherwise k(n) is 1 for said repeat of length n.