

RECOMBINANT PHYTOGLOBINS

Technical field of the invention

[0001] The present invention relates to recombinant haem- or phytooglobin or haem-bound proteins comprising a C-helix and a D-helix covalently joint into a CD-loop domain and comprising
5 a hexa- or penta- coordinated prosthetic haem group, such as hexacoordinated globins and phytooglobins derived from plants in the Caryophyllidae family. The present invention also relates to compositions comprising the recombinant protein, to non-therapeutic and therapeutic method of using the recombinant protein or the compositions, to dosage regimes of the recombinant protein, to genes and gene constructs encoding the recombinant proteins, to host cells and
10 cultures expressing the recombinant proteins and to methods of producing the recombinant proteins by cultivation of such host cells and cultures. Further, the present invention also relates to a composition comprising the hexacoordinated haem- or phytooglobin protein and various uses of the hexacoordinated haem- or phytooglobin protein and/or composition comprising the recombinant haem- or phytooglobin protein for humans and animals.

15

Background of the invention

[0002] Iron is an essential element with important functions such as oxygen transport, DNA synthesis and muscle metabolism. According to WHO, iron deficiency is a major nutritional problem, affecting over 2 billion people. It is the main cause of anaemia, which is the most
20 prevalent nutritional deficiency worldwide, affecting 33% of non-pregnant women, 40% of pregnant women, and 42% of children worldwide.

[0003] Iron deficiency results from depletion of iron stores and occurs when iron absorption cannot keep pace over an extended period with the metabolic demands for iron to sustain growth and to replenish iron loss, which is primarily related to blood loss. The primary causes of iron
25 deficiency include low intake of bioavailable iron, increased iron requirements as a result of rapid growth, pregnancy, menstruation, and excess blood loss caused by disease or trauma.

[0004] When iron stores are depleted and insufficient iron is available for erythropoiesis, haemoglobin synthesis in erythrocyte precursors become impaired and hematologic signs of iron deficiency and anaemia (IDA) appear.

30 [0005] Currently, IDA is treated with iron supplements; however, the iron used in almost all of these products is inorganic *a.k.a.* elemental iron. Elemental iron is poorly absorbed and most of it passes through the digestive system, producing free radicals that damage cells, tissues, and organs, resulting in side effects —digestive discomfort, stomach pain, nausea, constipation to name a few— and leaves a metallic aftertaste in patients' mouth. As a result, fewer than four out

of ten patients finish their treatment, the rest remain iron deficient or anaemic. It is worth mentioning that IDA is a recurring condition that returns easily after been resolved; therefore, there is a constant need for iron by those affected with this condition.

[0006] For humans, the main and best source of iron is found in foods of animal origin, especially red meat. There, iron is present in its organic form, haem-iron, which is bound to a protein (haemoglobin and myoglobin). However, since haem-iron is loosely bound to the apoprotein - as they are pentacoordinated - it is easily released and initiates toxic tumorigenic reactions (as described in *Example 1*).

[0007] Pentacoordinated phytooglobins readily lose their haem which is oxidized at high rates already in the mouth which is the reason behind the "meat flavour" but also has unhealthy effects. The haem, which becomes free haem as it is released very early during digestion before arriving to the iron absorption region (the duodenum and upper jejunum), is toxic and plays a central role in colon cancer. This has been the main reason for the WHO to classify red meat as probably carcinogenic and processed red meat as carcinogenic, as the treatment of these meat products speeds up the release and accumulation of free haem in the final product.

[0008] Bai et al. (2016) describes recombinant expression of non-symbiotic hemoglobin family gene, SoHb in Arabidopsis. However, heme-bound globins or phytooglobin are expressed in plants for the purpose of alleviating cellular stress such as abiotic nitrate stress. Therefore, phytooglobins are only found in plants in very low concentrations (1–20 μmol per kg fresh weight of stressed plants) or otherwise they are absent, and plants are not a reliable source of phytooglobins. Further plants contain iron inhibitors such as phenolic compounds and phytates which are harmful to phytooglobins. Thus, there remains a need for a safe and tolerable source of iron suitable for fortification and/or supplementation, which avoids the discomfort of elemental iron supplementation as well as the unhealthy effects of iron oxidation.

25

Summary of the invention

[0009] One objective of the present invention is to provide for improved haem-iron containing proteins which can provide for improved effect and convenience when administered to a subject in need thereof.

[0010] It is another object of the present invention to meet the need for a safe and tolerable source of dietary iron suitable for supplementation which overcomes the mentioned problems of conventional sources of dietary iron.

[0011] The recombinant and modified recombinant proteins containing haem-iron provided for herein display improved and superior stability and bioavailability compared to traditional iron supplements. Moreover, the inventor has identified phytooglobins which display a surprisingly

35

improved stability while also displaying a bioavailability on par with myoglobin, such as phytooglobins comprising a hexacoordinated haem group and having at least 86% identity to amino acid sequence SEQ ID NO: 1, or a fragment, variant, or fusion thereof which retains the hexacoordinated haem group, as well as compositions comprising said hexacoordinated phyto-
5 globin(s) and uses thereof.

[0012] Accordingly in a first aspect the present invention provides a recombinant haem- or phyto-
globin protein comprising a C-helix and a D-helix covalently joint into a CD-loop domain and comprising a hexa-or penta-coordinated prosthetic haem group, wherein the haem- or phyto-
globin has an amino acid sequence which is at least 50% identical the haem- or phyto-
10 globin comprised in anyone of SEQ ID NO: 1 to 55.

[0013] In a further aspect the present invention provides a composition comprising the recombinant protein disclosed herein and one or more carriers, agents, additives and/or excipients.

[0014] In a further aspect the present invention provides a non-therapeutic method of improving
15 the stamina or low oxygen tolerance in a subject, comprising administering the recombinant protein or the composition disclosed herein to the subject in an amount effective to improve stamina or low oxygen tolerance in the subject.

[0015] In a further aspect the present invention provides a method or the recombinant protein or composition for use in a method for treating, ameliorating, or preventing of a disease or deficiency
20 in a subject, comprising administering the recombinant protein or the composition to the subject in an amount effective to ameliorate or prevent said disease or deficiency.

[0016] In a further aspect the present invention provides a gene encoding the recombinant protein disclosed herein.

[0017] In a further aspect the present invention provides a polynucleotide construct comprising
25 the nucleotide sequence described herein operably linked to a control sequence directing the transcription and/or translation of the gene into the recombinant protein.

[0018] In a further aspect the present invention provides a genetically modified host cell expressing the gene or polynucleotide construct and the recombinant protein disclosed herein.

[0019] In a further aspect the present invention provides a method for producing the recombinant
30 protein described herein comprising

- a) culturing the cell culture of the disclosure at conditions allowing the cell to produce the recombinant protein; and
- b) optionally recovering and/or isolating the recombinant protein.

[0020] In a further aspect the present invention provides a composition comprising the cell culture
35 disclosed herein.

[0021] In a further aspect the present invention relates to a recombinant haem- or phytoalbumin protein having a hexacoordinated haem group and at least 86% identity to amino acid sequence SEQ ID NO: 1, or a fragment, variant, or fusion thereof which retains the hexacoordinated haem group.

5 [0022] In a further aspect of present invention relates to a composition comprising at least one recombinant hexacoordinated haem- or phytoalbumin protein as described herein.

[0023] A further aspect relates to the recombinant hexacoordinated haem- or phytoalbumin proteins according to the invention, and/or compositions comprising said hexacoordinated haem- or phytoalbumin proteins, for various uses, for example use in treatment, amelioration, or
10 prevention of iron deficiency and/or anaemia in a subject.

[0024] A further aspect of the present invention relates to non-recombinant haem- or phytoalbumin proteins comprising a hexacoordinated haem group and having at least 86% identity to amino acid sequence SEQ ID NO: 1, or a fragment thereof which retains the hexacoordinated haem group, as well as compositions comprising said non-recombinant hexacoordinated haem- or phytoalbumin
15 protein(s) and uses thereof.

Brief description of the figures

[0025] Figure 1. Production of recombinant plant haemoglobin. (A) *E. coli* cells containing the recombinantly expressed haemoglobins. (B) Cell lysate after sonication. (C) Purified recombinant
20 haemoglobins. (D) SDS-PAGE showing the purification process, M: protein size ladder, 1: cell lysate, 2: IEX purification, 3: HIC purification, 4: IEX purification.

[0026] Figure 2. Autooxidation of plant haemoglobins. Time courses for the autooxidation of plant haemoglobins. The data shown corresponds to the change in absorbance monitored at 425 nm Bv: sugar beet (*Beta vulgaris*), Chq: quinoa (*Chenopodium quinoa*), so: spinach (*Spinacia oleracea*).
25 See Example 1.

[0027] Figure 3 Melting temperature (T_m) of plant haemoglobin. Graphical representation of the T_m observed by the formation of shoulders in the curves at the temperature where fluorescence increased due to scattering caused by protein denaturation. So: spinach, Bv: sugar beet, Zm: maize, BtMg: bovine myoglobin, Chq: quinoa. See Example 2.

30 [0028] Figure 4. DNA degradation caused by haemoglobin. A. Agarose gel showing the amount of degraded DNA. B. Percentage SC pDNA after incubation with oxidized haemoglobin. rHbA: recombinant adult human haemoglobin, Bv: sugar beet (*Beta vulgaris*), Cq: quinoa (*Chenopodium quinoa*), So: spinach (*Spinacia oleracea*). See Example 3 and Table 3.

[0029] Figure 5. Iron bioavailability of Plant Haemoglobin. Results from the Caco-2 cells assay. A, Ferritin formation with pure haemoglobin. B, Ferritin formation in a mixture of pure haemoglobin
35

and food. so, spinach; Chq, quinoa; Zm, corn; BtMg, beef myoglobin; BV, sugar beet; CBF, Canned Baby Food; PBS, Plant-based drink. $n = 3$, $p < 0.05$. See Example 4 and Table 4.

[0030] Figure 6 shows the synthesis of haemoglobin consists of two parts, first the synthesis of haem b group and then the synthesis of the apoprotein (globin protein) that will host the haem-group. For the haem synthesis the ring system is that of protoporphyrin IX. This prosthetic group is tightly, but non-covalently bound within its host protein, a globin (the apoprotein).

[0001] Figure 7 shows the structure of a hexacoordinated haem-bound globin. The CD-loop is enclosed inside the circle. The iron is in the centre of the Haem group coordinated by two histidine side chains.

[0002] Figure 8 shows the phylogenetic trees of the different types and classes of haem- or phytoglobin proteins. (A) This phylogenetic tree was built using only the CD-loops of haem- or phytoglobin proteins (SEQ ID NO: 56-110). (B). For this phylogenetic tree the haem- and phytoglobin proteins had their CD-loops substituted by the parent CD-loop SEQ ID NO: 56.

[0003] Figure 9 shows the expression and production of haem-bound proteins containing each a modified CD-loop. From the upper left corner and following from left to right and up and down the haem-bound proteins are: A1_H_01; A2_H_02; A3_H_03; A4_H_04; A5_H_05; A6_H_06; B1_H_07; B2_H_08; B3_H_09; B4_H_10; B5_H_11; B6_H_12; C1_H_13; C2_H_14; C3_H_15; C4_H_16; C5_H_17; C6_H_18; D1_H_19; D2_H_20; D3_H_21; D4_H_22; D5_H_23; D6_H_24 (SEQ ID NO 11 to 134).

20

Detailed description of the invention

Definitions

[0004] The term “phytoglobin” is used to demote globular plant (algae and land plant) proteins which contain a haem prosthetic group. These proteins are also known as plant haemoglobin, and the terms are used interchangeably herein.

[0005] The term “haem-bound protein” as used herein refers to globular proteins which contain a haem prosthetic group that contains at least one iron atom.

[0006] Haem group: A haem is a porphyrin molecule holding an iron molecule. Phytoglobins and haem-bound proteins typically have a haem b prosthetic group.

[0007] Iron deficiency: In the present context, the term “iron deficiency” should be understood as a condition when physiological requirements cannot be met by iron absorption from the diet or iron absorption cannot keep pace over an extended period with the metabolic demands for iron to sustain growth or pregnancy and to replenish iron loss, which is primarily related to blood loss due to menstruation, or caused by disease or trauma.. The symptoms may be mild and can include fatigue, weakness, headache, etc. If iron deficiency progresses, it leads to iron deficiency anaemia,

which is the most common form of anaemia.

[0008] The term "Anaemia" is defined as a haemoglobin concentration below a specified cut-off point; that cut-off point depends on the age, gender, physiological status, smoking habits, and altitude at which the population being assessed lives. WHO defines anaemia in children aged
5 under 5 years and pregnant women as a haemoglobin concentration <110 g/L at sea level, and anaemia in non-pregnant women as a haemoglobin concentration <120 g/L. Anaemia may result as nutritional deficiency but may also be caused by other chronic or acute disease or conditions. In context of this application, the terms "iron deficiency", "iron deficiency anaemia" and "anaemia" are used interchangeably, and all designate conditions where it would be benefit from
10 increased intake of dietary iron.

[0009] Sequence identity: The relatedness between two amino acid sequences or between two nucleotide sequences is described by the parameter "sequence identity". In the present context, the term "sequence identity" is here defined as the sequence identity between proteins at the amino acid level. The protein sequence identity may be determined by comparing the amino acid
15 sequence in a given position in each sequence when the sequences are aligned. To determine the percent identity of two amino acid sequences, the sequences are aligned for optimal comparison purposes. The amino acid residues at corresponding amino acid positions are then compared. When a position in the first sequence is occupied by the same amino acid residue as the corresponding position in the second sequence, then the molecules are identical at that position.
20 The percent identity between the two sequences is a function of the number of identical positions shared by the sequences (i.e., % identity = # of identical positions/total # of positions (e.g., overlapping positions) x 100). The sequence identity between two amino acid sequences may for example be determined using the Needleman-Wunsch algorithm (Needleman and Wunsch, 1970, J. Mol. Biol. 48: 443-453) as implemented in the Needle program of the EMBOSS package
25 (EMBOSS: The European Molecular Biology Open Software Suite, Rice et al., 2000, Trends Genet. 16: 276-277), preferably version 5.0.0 or later. The parameters used are gap open penalty of 10, gap extension penalty of 0.5, and the EBLOSUM62 (EMBOSS version of BLOSUM62) substitution matrix. The output of Needle labelled "longest identity" (obtained using the -no brief option) is used as the percent identity and is calculated as follows:

30

$$(\text{Identical Residues} \times 100) / (\text{Length of Alignment} - \text{Total Number of Gaps in Alignment}) .$$

Alignments in the present context were performed using Protein Blast (blastp) on NCBI with Matrix BLOSUM62.

35 [0010] Variant: The term "variant" means a polypeptide having hexacoordinated haem group

comprising an alteration, i.e., a substitution, insertion, and/or deletion, at one or more (e.g., several) positions.

[0011] A substitution means replacement of the amino acid occupying a position with a different amino acid; a deletion means removal of the amino acid occupying a position; and an insertion means adding one or more amino acids adjacent to and immediately following the amino acid occupying a position. In some embodiments of the invention, the variants comprise only conservative substitutions. Conservative substitutions are where an amino acid residue is replaced with an amino acid that has similar chemical properties. For example, the replacement of serine by threonine, or leucine by isoleucine, are both considered conservative substitutions.

10 [0012] Fragment: The term "fragment" as used herein, refers to a polypeptide having one or more (e.g., several) amino acids absent from the amino and/or carboxyl terminus of the mature polypeptide of any one of the parent sequences herein disclosed, such as SEQ ID NOs: 5, 7, 9, and 10; wherein the fragment has a hexacoordinated haem group. In one aspect, a fragment contains at least 200 contiguous amino acid residues of SEQ ID NOs: 1-55, for example at least 300
15 contiguous amino acid residues, or at least 60 contiguous amino acid residues, or at least 80 contiguous amino acid residues, or at least 100 contiguous amino acid residues of SEQ ID NOs: 5, 7, 9, and 10.

[0013] Fusion polypeptide: The term "fusion polypeptide" or "fusion" is a polypeptide in which one polypeptide is fused at the N-terminus or the C-terminus of a polypeptide of the present
20 invention. A fusion polypeptide is produced by fusing a polynucleotide encoding another polypeptide to a polynucleotide of the present invention. Techniques for producing fusion polypeptides are known in the art and include ligating the coding sequences encoding the polypeptides so that they are in frame and that expression of the fusion polypeptide is under control of the same promoter(s) and terminator. Fusion polypeptides may also be constructed
25 using intern technology in which fusion polypeptides are created post-translationally (Cooper et al., 1993, EMBO J. 12: 2575-2583; Dawson et al., 1994, Science 266: 776-779). A fusion polypeptide can further comprise a cleavage site between the two polypeptides. Upon secretion of the fusion protein, the site is cleaved releasing the two polypeptides. Examples of cleavage sites include, but are not limited to, the sites disclosed in Martin et al., 2003, J. Ind. Microbiol. Biotechnol. 3: 568-
30 576; Svetina et al., 2000, J. Biotechnol. 7Q: 245-251; Rasmussen- Wilson et al., 1997, Appl. Environ. Microbiol. 63: 3488-3493; Ward et al., 1995, Biotechnology 13: 498-503; and Contreras et al., 1991, Biotechnology 9: 378-381 ; Eaton et al., 1986, Biochemistry 25: 505-512; Collins-Racie et al., 1995, Biotechnology 13: 982-987; Carter et al., 1989, Proteins: Structure, Function, and Genetics 6: 240-248; and Stevens, 2003, Drug Discovery World 4: 35-48.

35 [0014] Recombinant: The term "recombinant," when used in reference to a cell, nucleic acid,

protein, or vector, means that it has been modified from its native state. Thus, for example, recombinant cells express genes that are not found within the native (non-recombinant) form of the cell, or express native genes at different levels or under different conditions than found in nature. Recombinant nucleic acids differ from a native sequence by one or more nucleotides and/or are operably linked to heterologous sequences, e.g., a heterologous promoter in an expression vector. Recombinant proteins may differ from a native sequence by one or more amino acids and/or are fused with heterologous sequences. A vector comprising a nucleic acid encoding a polypeptide is a recombinant vector. The term "recombinant" is synonymous with "genetically modified" and "transgenic". Recombinant polypeptides of the invention are obtained by use of recombinant DNA techniques. Such methods normally comprise cultivation of a host cell transformed with a recombinant DNA vector comprising a DNA sequence encoding the haem- or phytoalbumin protein according to the invention, and the DNA sequence being operationally linked with an appropriate expression signal such that it is capable of expressing the enzyme in a culture medium under conditions permitting the expression of the haem- or phytoalbumin protein and recovering the haem- or phytoalbumin protein from the culture. The DNA sequence may also be incorporated into the genome of the host cell. The DNA sequence may be of genomic, cDNA or synthetic origin or any combinations of these, and may be isolated or synthesized in accordance with methods known in the art.

[0015] Pharmaceutical acceptable carrier: In the present context, the term "pharmaceutically acceptable" refers to molecular entities and compositions that are suitable for use with humans and/or animals without undue adverse side effects (such as toxicity, irritation, and allergic response) commensurate with a reasonable benefit/risk ratio.

[0016] The use of the terms "a" and "an" and "the" and similar referents in the context of describing the invention (especially in the context of the following claims) are to be construed to cover both the singular and the plural, unless otherwise indicated herein or clearly contradicted by context.

[0017] The terms "comprising", "comprises", "consisting of", "having", "including", "at least" and "containing" are to be construed as open-ended terms (i.e., meaning "including, but not limited to,") unless otherwise noted. Recitation of ranges of values herein are merely intended to serve as a shorthand method for referring individually to each separate value falling within the range, unless otherwise indicated herein, and each separate value is incorporated into the specification as if it were individually recited herein. All methods described herein can be performed in any suitable order unless otherwise indicated herein or otherwise clearly contradicted by context. The use of any and all examples, or exemplary language (e.g., "such as") provided herein, is intended merely to better illuminate the invention and does not pose a limitation on the scope of the

invention unless otherwise claimed. No language in the specification should be construed as indicating any non-claimed element as essential to the practice of the invention.

[0018] Reference to "about" a value or parameter herein includes embodiments that are directed to that value or parameter per se. For example, description referring to "about X" includes the
5 embodiment "X". When used in combination with measured values, "about" includes a range that encompasses at least the uncertainty associated with the method of measuring the particular value and can include a range of plus or minus two standard deviations around the stated value.

[0019] Likewise, reference to a gene or polypeptide that is "derived from" another gene or polypeptide X, includes the gene or polypeptide X.

10 [0020] As used herein and in the appended claims, the singular forms "a," "or," and "the" include plural referents unless the context clearly dictates otherwise.

[0021] It is understood that the embodiments described herein include "consisting of" and/or "consisting essentially of" embodiments. As used herein, except where the context requires otherwise due to express language or necessary implication, the word "comprise" or variations
15 such as "comprises" or "comprising" is used in an inclusive sense, i.e. to specify the presence of the stated features but not to preclude the presence or addition of further features in various embodiments.

[0022] The term "CD-loop domain" or "CD variable loop domain" or "variable loop domain" as used herein interchangeably refers to the domain in a haem or phytoalbumin beginning at the end
20 of the C- helix and the beginning of E-helix.

[0023] The term "hexacoordinated" or "pentacoordinated" as used herein refers to coordination of haem iron in a haem- or phytoalbumin to one histidine (pentacoordinated) or two histidine amino acids or another amino acid that binds the haem-group to the apoprotein (hexacoordinated).

[0024] The terms "heterologous" or "recombinant" or "genetically modified" and their
25 grammatical equivalents as used herein interchangeably about nucleotides, polypeptides and cells refers to entities "derived from a different species or cell". For example, a heterologous or recombinant polynucleotide gene is a gene in a host cell not naturally containing that gene, i.e. the gene is from a different species or cell type than the host cell. A heterologous or recombinant polypeptide is a polypeptide produced in a host cell not naturally containing the polypeptide, i.e.
30 the polypeptide is from a different species or cell type than the host cell. Where the terms as used herein about host cells, they refer to host cells comprising and expressing heterologous or recombinant polynucleotides. In some embodiments "recombinant" or "non-naturally occurring" when used with reference to, e.g., a host cell, nucleic acid, or polypeptide, refers to a material, or a material corresponding to the natural or native form of the material, that has been modified in
35 a manner that would not otherwise exist in nature, or is identical thereto but produced or derived

from synthetic materials and/or by manipulation using recombinant techniques. Non-limiting examples include, among others, recombinant host cells expressing genes that are not found within the native (non-recombinant) form of the cell or express native genes that are otherwise expressed at a different level. "Heterologous" as used herein means that a polypeptide is normally not found in or made (i.e. expressed) by the host organism but derived from a different species.

[0025] The term "expression" includes any step involved in the production of a polypeptide (e.g., encoded enzyme) including, but not limited to, transcription, post-transcriptional modification, translation, post-translational modification, and secretion.

[0026] The term "expression vector" refers to a DNA molecule, either single- or double stranded, either linear or circular, which comprises a polynucleotide encoding a polypeptide and is operably linked to control sequences that provide for its expression. Expression vectors include expression cassettes for the integration of genes into a host cell as well as plasmids and/or chromosomes comprising such genes. Vectors capable of directing the expression of genes to which they are operatively linked are referred to herein as expression vectors. "Vectors" can also refer to a nucleic acid molecule capable of transporting another nucleic acid molecule to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded nucleic acid loop into which additional nucleic acid segments can be ligated. Certain other vectors are capable of facilitating the insertion of an exogenous nucleic acid molecule into a genome of a bacterium. Such vectors are referred to herein as "transformation vectors". In general, vectors of utility in recombinant nucleic acid techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is one of the most commonly used forms of a vector. Large numbers of suitable vectors are known to those of skill in the art and commercially available.

[0027] The term "host cell" refers to any cell type that is susceptible to transformation, transfection, transduction, or the like with a nucleic acid construct or expression vector comprising a polynucleotide to be expressed in the host cell. Host cell encompasses any progeny of a parent cell including those that are not identical to the parent cell due to mutations that occur during replication.

[0028] The term "Nucleic acid" or "polynucleotide" are used interchangeably herein to denote a polymer of at least two nucleic acid monomer units or bases (e.g., adenine, cytosine, guanine, thymine) covalently linked by a phosphodiester bond, regardless of length or base modification.

[0029] The term "polynucleotide construct" refers to a polynucleotide, either single- or double stranded, which is isolated from a naturally occurring gene or is modified to contain segments of nucleic acids in a manner that would not otherwise exist in nature, or which is synthetic, and which comprises a polynucleotide encoding a polypeptide and one or more control sequences.

[0030] The term "operably linked" refers to a configuration in which a control sequence is placed at an appropriate position relative to a coding polynucleotide such that the control sequence directs expression of the coding polynucleotide. More generally, "operably linked" refers to a juxtaposition wherein the components described are in a relationship permitting them to function in their intended manner. A control sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequence. A promoter sequence is "operably linked" to a gene when it is in sufficient proximity to the transcription start site of a gene to regulate transcription of the gene.

5 [0031] As used herein, "promoter" refers to a sequence of DNA, usually upstream (5') of the coding region of a structural gene, which controls the expression of the coding region by providing recognition and binding sites for RNA polymerase and other factors which may be required for initiation of transcription. The selection of the promoter will depend upon the nucleic acid sequence of interest. A suitable "promoter" is generally one which is capable of supporting the initiation of transcription in a bacterium of the invention, causing the production of an mRNA molecule.

15 [0032] "Polypeptide" and "protein" are used interchangeably herein to denote a polymer of at least two amino acids covalently linked by an amide bond, regardless of length or post-translational modification (e.g., glycosylation, phosphorylation, lipidation, myristylation, ubiquitination, etc.). Included within this definition are D- and L-amino acids, and mixtures of D- and L-amino acids.

[0033] The terms "nucleotide sequence" and "polynucleotide" are used herein interchangeably.

[0034] Terms like "preferably", "commonly", "particularly", and "typically" are not utilized herein to limit the scope of the itemed invention or to imply that certain features are critical, essential, or even important to the structure or function of the itemed invention. Rather, these terms are merely intended to highlight alternative or additional features that can or cannot be utilized in a particular embodiment of the present invention.

25 [0035] The term "cell culture" as used herein refers to a culture medium comprising a plurality of the host cells described herein. A cell culture may comprise a single strain of host cells or may comprise two or more distinct host cell strains. The culture medium may be any medium that may comprise a recombinant host, e.g., a liquid medium (i.e., a culture broth) or a semi-solid medium, and may comprise additional components, e.g., a carbon source; a nitrogen source; a phosphate source; vitamins; trace elements; salts; amino acids; nucleobases; and the like.

[0036] Term "endogenous" or "native" as used herein refers to a gene or a polypeptide in a host cell which originates from the same host cell.

35 [0037] The terms "substantially" or "approximately" or "about", as used herein refers to a

reasonable deviation around a value or parameter such that the value or parameter is not significantly changed. These terms of deviation from a value should be construed as including a deviation of the value where the deviation would not negate the meaning of the value deviated from. For example, in relation to a reference numerical value the terms of degree can include a range of values plus or minus 10% from that value. For example, deviation from a value can include a specified value plus or minus a certain percentage from that value, such as plus or minus 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, or 1% from the specified value.

[0038] Where a numerical limit or range is stated herein, the endpoints are included. Also, all values and sub ranges within a numerical limit or range are specifically included as if explicitly written out.

[0039] The term "and/or" as used herein is intended to represent an inclusive "or". The wording X and/or Y is meant to mean both X or Y and X and Y. Further the wording X, Y and/or Z is intended to mean X, Y and Z alone or any combination of X, Y, and Z.

[0040] The term "isolated" as used herein about a compound, refers to any compound, which by means of human intervention, has been put in a form or environment that differs from the form or environment in which it is found in nature. Isolated compounds include but are not limited to compounds of the disclosure for which the ratio of the compounds relative to other constituents with which they are associated in nature is increased or decreased. In an important embodiment the amount of compound is increased relative to other constituents with which the compound is associated in nature. In an embodiment the compound of the disclosure may be isolated into a pure or substantially pure form. In this context a substantially pure compound means that the compound is separated from other extraneous or unwanted material present from the onset of producing the compound or generated in the manufacturing process. Such a substantially pure compound preparation contains less than 10%, such as less than 8%, such as less than 6%, such as less than 5%, such as less than 4%, such as less than 3%, such as less than 2%, such as less than 1%, such as less than 0.5% by weight of other extraneous or unwanted material usually associated with the compound when expressed natively or recombinantly. In an embodiment the isolated compound is at least 90% pure, such as at least 91% pure, such as at least 92% pure, such as at least 93% pure, such as at least 94% pure, such as at least 95% pure, such as at least 96% pure, such as at least 97% pure, such as at least 98% pure, such as at least 99% pure, such as at least 99.5% pure, such as 100 % pure by weight.

[0041] The term "deletion" as used herein in the context of polynucleotides and genes refers to the manipulation of a gene so that it is no longer expressed in a host cell. "Deletion" or "deleted", when used with reference to a polypeptide, refers to modification of the polypeptide by removal of one or more amino acids in the reference polypeptide. Deletions can comprise removal of 1 or

more amino acids, 2 or more amino acids, 5 or more amino acids, 10 or more amino acids, 15 or more amino acids, or 20 or more amino acids, up to 10% of the total number of amino acids, or up to 20% of the total number of amino acids making up the polypeptide while retaining enzymatic activity and/or retaining the improved properties of an engineered enzyme. Deletions can be directed to the internal portions and/or terminal portions of the polypeptide, in various embodiments, the deletion can comprise a continuous segment or can be discontinuous.

[0042] The terms "insertion" or "inserted" or "addition" when used with reference to a polypeptide, refers to modification of the polypeptide by addition of one or more amino acids to the reference polypeptide. Insertions can comprise the addition of 1 or more amino acids, 2 or more amino acids, 5 or more amino acids, 10 or more amino acids, 15 or more amino acids, or 20 or more amino acids. Insertions can be in the internal portions of the polypeptide, or to the carboxy or amino terminus. The insertion can be a contiguous segment of amino acids or separated by one or more of the amino acids in the reference polypeptide.

[0043] Further to the above "Substitution" or "substituted" refers to modification of the polypeptide by replacing one amino acid residue with another, for instance the replacement of a Serine residue with a Glycine or Alanine residue in a polypeptide sequence is an amino acid substitution. When used with reference to a polynucleotide, "substitution" or "substituted" refers to modification of the polynucleotide by replacing one nucleotide with another, for instance the replacement of a cytosine with a thymine in a polynucleotide sequence is a nucleotide substitution.

[0044] Further to the above "conservative substitution", when used with reference to a polypeptide, refers to a substitution of an amino acid residue with a different residue having a similar side chain, and thus typically involves substitution of the amino acid in the polypeptide with amino acids within the same or similar class of amino acids. By way of example and not limitation, an amino acid with an aliphatic side chain may be substituted with another aliphatic amino acid, e.g., alanine, valine, leucine, and isoleucine; an amino acid with hydroxyl side chain is substituted with another amino acid with a hydroxyl side chain, e.g., serine and threonine; an amino acid having an aromatic side chain is substituted with another amino acid having an aromatic side chain, e.g., phenylalanine, tyrosine, tryptophan, and histidine; an amino acid with a basic side chain is substituted with another amino acid with a basic side chain, e.g., lysine and arginine; an amino acid with an acidic side chain is substituted with another amino acid with an acidic side chain, e.g., aspartic acid or glutamic acid; and a hydrophobic or hydrophilic amino acid is replaced with another hydrophobic or hydrophilic amino acid, respectively.

[0045] Further to the above "non-conservative substitution", when used with reference to a polypeptide, refers to a substitution of an amino acid in a polypeptide with an amino acid with

significantly differing side chain properties. Non-conservative substitutions may use amino acids between, rather than within, the defined groups and affects (a) the structure of the peptide backbone in the area of the substitution (e.g., serine for glycine), (b) the charge or hydrophobicity, or (c) the bulk of the side chain. By way of example and not limitation, an exemplary non-conservative substitution can be an acidic amino acid substituted with a basic or aliphatic amino acid; an aromatic amino acid substituted with a small amino acid; and a hydrophilic amino acid substituted with a hydrophobic amino acid.

Recombinant haem- or phytoalbumin proteins

[0046] In the first aspect provided herein is a recombinant haem- or phytoalbumin protein comprising a C-helix and a D-helix covalently joint into a CD-loop domain and comprising a hexa- or penta-coordinated prosthetic haem group, wherein the haem- or phytoalbumin has an amino acid sequence which is at least 50% identical the haem- or phytoalbumin comprised in anyone of SEQ ID NO: 1 to 55. Preferably, the amino acid sequence of the haem- or phytoalbumin is at least 60%, such as at least 70%, such as at least 80%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99%, such as 100% identical the haem- or phytoalbumin comprised in anyone of SEQ ID NO: 1 to 55.

[0047] The haem prosthetic group in haemoglobin is most often attached to the globin apoprotein through coordination of either one or two histidine side chains or another amino acid that keeps the haem-group bound to the apoprotein. Those proteins with one histidine coordinating the haem iron are called "pentacoordinated" haemoglobins, a group represented by red blood cell haemoglobin and most other oxygen transporters. Those with two histidine amino acids chains or another amino acid that keeps the haem-group bound to the apoprotein are called "hexacoordinated" haemoglobins, or hexacoordinated phytoalbumins and have been found to be more stable than pentacoordinated phytoalbumins. Hexacoordination contributes to stabilizing the iron binding, keeping the iron group inside the protein, and protecting it from oxidation and from early release.

[0048] It has been found that the recombinant haem- or phytoalbumin proteins described herein have improved bioavailability compared to traditional iron supplements. Further it has been found that the amino acid sequence of the CD-loop domain has significant impact on how strong the prosthetic haem group is coordinated or bound to the protein and there how stable the haem group is placed in the complex and how well protected the haem group is protected against autooxidation. Autooxidation in haemoglobin results in the formation of superoxide and hydrogen peroxide (H_2O_2). The formed H_2O_2 reacts with the iron of oxy-ferrous haemoglobin, (Fe(II)) and undergoes a Fenton reaction producing the highly reactive hydroxyl radical (Fe(III)) which

produces cellular and tissue damage (see Sadrzadeh et al. (1984))⁴). This is a “spontaneous oxidation” that occurs at different rates giving an indication of the intrinsic reactivity of a haemoglobin molecule. In addition to producing damaging radicals, autooxidation results in the release of the haem group outside of the protein. Haem is an iron-coordinating porphyrin proposed to be the key molecule contributing to tumorigenesis. Haem is contained predominantly in red and processed meat in the form of haemoglobin and myoglobin. The role of dietary haem in cancer has been highlighted in different types of carcinomas. Indeed, high consumption of red and processed meat has been associated with increased incidence of oesophageal, gastric, breast, endometrial, pancreas and lung tumour. However, the majority of studies focus on the role of dietary haem in the pathogenesis of colorectal cancer (CRC), a leading cause of cancer deaths in Western Countries (Fiorito et al. (2020))⁵.

[0049] Accordingly, modifications in the CD-loop domain have been found to have an impact on the expression and the stability of the haem or phytoalbumin, and its modifications has been identified which increases the binding strength and stability of the haem in the protein decreasing the harmful effects of autooxidation and release of haem from the protein. Accordingly, in an embodiment the CD-loop domain in the recombinant protein comprises one or more modifications compared to a parent CD-loop domain, whereby the modified CD-loop domain binds the prosthetic group stronger than the parent unmodified CD-loop domain. The parent CD-loop domain is preferably native to the parent protein. Further the parent CD-loop domain is preferably positioned in the recombinant haem- or phytoalbumin protein corresponding to the position 52 to 78 of SEQ ID NO: 1. This portion includes the first portion extending from the CD loop, so in some embodiments, the parent CD-loop domain is preferably positioned in the recombinant haem- or phytoalbumin protein corresponding to the position 52 to 72 of SEQ ID NO: 1. In further embodiments the CD-loop domain of the recombinant protein has an amino acid sequence which is at least 50% identical to the CD-loop domain comprised in anyone of SEQ ID NO: 56 to 147. Preferably, the CD-loop domain of the recombinant protein has an amino acid sequence which is at least at least 60%, such as at least 70%, such as at least 80%, such as at least 90%, such as at least 95%, such as at least 98% , such as at least 99%, such as 100% identical to the CD-loop domain comprised in anyone of SEQ ID NO: 56 to 147. SEQ ID NO: 56 to 110 are the CD-loop domains identified in the phytoalbumins SEQ ID NO: 1 to 55, while SEQ ID NO: 123 to 147 are genetically modified CD-loop domains. SEQ ID NO: 111 to 123 represents CD loop domains which are also found in SEQ ID NO: 56 to 110, but also contains some amino acid residues before and/or after the CD loop domain.

[0050] The modifications in the CD-loop domain can be any modification that increases the binding or coordination strength of the haem group, including deletions, substitutions and/or

additions of one or more amino acids. In some embodiments, the modifications alter the flexibility of the flexibility of the CD loop domain to make it more rigid or flexible compared to the parent CD-loop domain. In some embodiments the modification in the CD-loop domain comprises substitution of a flexible and/or small amino acid (such as G, A, V, S, I, L, M) with a rigid and/or bulky amino acid (such as P, F, W, Y) and/or vice versa. The CD-loop domain may comprise, one, two, three, four, five, six, seven, eight, nine, ten, or more modifications such as substitutions compared to the parent CD-loop domain. Referring to corresponding amino acid positions of the CD-loop domain in the phytoalbumin of *Chenopodium quinoa* substituting S13 for a G increases the expression and production of the recombinant protein. Maintaining the position P21 or if not a P in the parent CD-loop domain then substituting it to a P also increases the expression and production of the recombinant protein. Position V15 can be substituted for a G and in particular together with a substitution of S13 to G, provide a beneficiary effect on the expression and production of the recombinant protein. Position A25 can be substituted with G or P. Position A27 can be substituted with a G. Some positions however should either not be substituted or should only be conservatively substituted in hexacoordinated recombinant proteins. Position P1 should not be substituted as it can increase the instability of the protein and affect its expression. The same applies for A3 and F7. Position S8 should not be substituted unless the substitution is accompanied by a second substitution around the His26 by substituting A25 and/or A27 with G to balance the effect of the substitution. P21 should be maintained or only substituted with conservative amino acids. A25 should not be substituted with a G or another conserved amino acid. Accordingly, in some embodiments the amino acid modification in the CD-loop domain is a modification corresponding to one or more modifications selected from S13G, S13P, V15G, A25P, and A27G of SEQ ID NO: 56. In other embodiments the amino acid in the CD loop corresponding to positions P1, A3, F7, S8, S13, P16, P21 and/or P25 of SEQ ID NO: 56 are conserved or only conservatively substituted. In further embodiments the CD-loop domain comprises the amino acid sequence as set forth in anyone of SEQ ID NO: 123, 127, 128, 129, 130, 131, 133, 134, 135, 137, 139, 141, 142, or 145. In particular the CD-loop domain comprises the amino acid sequence as set forth in anyone of SEQ ID NO: 127, 128, 130, 131, 135, 139, 141, 142, or 145.

[0051] In further embodiment the recombinant protein comprises a signal peptide, directing the recombinant protein for secretion from a host cell wherein it is produced. The signal peptide suitably directs expression in a microbial cell. In some embodiments the signal peptide is heterologous to the recombinant protein and optionally selected for optimized functionality in the selected host cell suitably being a eukaryotic, bacterial, or archaeal cell – preferably a bacterial cell or a fungal cell. The signal peptide preferably has an amino acid sequence which is at least 50% identical to the signal peptide comprised in anyone of SEQ ID NO: 203 to 213. Preferably, the

signal peptide has an amino acid sequence which is at least at least 60%, such as at least 70%, such as at least 80%, such as at least 90%, such as at least 95%, such as at least 98%, such as at least 99%, such as 100% identical to the signal peptide comprised in anyone of SEQ ID NO: 203 to 213.

[0052] In further embodiments the recombinant protein is non-natural or synthetic and optionally further modified, such as being a fusion protein. The fusion protein can for example comprise two of more prosthetic haem groups.

[0053] In further embodiments. the recombinant protein is a nonsymbiotic haemoglobin of Class 1. Further the recombinant protein may be derived from a plant of the genus Caryophyllidae. Phytoglobins isolated from this genus has been found to be particularly stabile. – especially phytoglobins isolated from spinach or quinoa.

[0054] The recombinant protein further can comprise one or more post translational modification such as glycosylations and/or phosphorylations compared to a corresponding natural protein.

[0055] As described, *supra*, the inventor of the present invention has surprisingly succeeded in developing hexacoordinated phytoglobins which display increased stability, while also displaying good bioavailability. These qualities render them particularly suitable for dietary iron supplementation and/or fortification and for use to prevent or treat iron deficiencies and/or anaemia in organisms.

[0056] Thus, in further embodiments the present invention relates to a recombinant haem- or phytoglobin comprising a hexacoordinated haem group, wherein said haem- or phytoglobin protein has at least 86% identity to amino acid sequence SEQ ID NO: 1; or a fragment, variant, or fusion thereof which retains the hexacoordinated haem group, for use in treatment or prevention of iron deficiency and anaemia in a subject.

A further embodiment relates to the recombinant haem- or phytoglobin protein according to the invention, wherein said haem- or phytoglobin globin has a sequence identity of at least 87%, such as at least 88%, such as at least 89%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98, such as at least 99% to sequence SEQ ID NO: 1.

[0057] A further embodiment relates to the recombinant haem- or phytoglobin protein according to the invention, wherein the haem- or phytoglobin protein is a nonsymbiotic haemoglobins of Class 1 and and/or Class 2. A further embodiment relates to the recombinant haem- or phytoglobin protein according to any of the invention, wherein the recombinant haem- or phytoglobin protein is derived from a plant in Caryophyllidae, such as from quinoa and/or spinach.

[0058] Thus, the invention in one embodiment relates to the recombinant haem- or phytoglobin protein according to the invention, wherein said haem- or phytoglobin protein has a sequence identity of at least 87%, such as at least 88%, such as at least 89%, such as at least 90%, such as at

least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98, such as at least 99% to sequence acid sequence selected from the group consisting of SEQ ID NO: 1 to 55, in particular the group consisting of SEQ ID NO: 1, 3, 4, 5, 7, 9, and 10.

5 [0059] A further embodiment relates to the recombinant haem- or phytoalbumin proteins according to the invention, comprising or consisting of an amino acid sequence selected from the group consisting of any of SEQ ID NO: 1 to 55, in particular the group consisting of SEQ ID NO: 1, 3, 4, 5, 7, 9, and 10. or a fragment, variant, or fusion of any of the group which retains the hexacoordinated haem group. SEQ ID NO: 1 is a hexacoordinated phytoalbumin from quinoa, while
10 SEQ ID NO: 5 is a hexacoordinated phytoalbumin from spinach.

[0060] Further embodiments relate to the recombinant phytoalbumin according to the invention comprising or consisting of an amino acid sequence SEQ ID NO 1, or a fragment, variant, or fusion thereof which retains the hexacoordinated haem group.

[0061] In one embodiment, the invention relates to a haem- or phytoalbumin globin according to
15 the invention where the phytoalbumin has an auto-oxidation rate of less than 0.2, such as less than 0.15, for example in the range from 0.01 to 0.18, or 0.01 to 0.15, such as about 0.1 h⁻¹

[0062] In one embodiment, the invention relates to a haem- or phytoalbumin globin according to the invention wherein said phytoalbumin has no metallic taste or has neutral taste.

[0063] In one embodiment, the invention relates to a phytoalbumin according to the invention
20 wherein said phytoalbumin has a melting temperature of above 68°C, such as in the range from 68 to 75°C; such as in the range from 70 to 76°C. In some embodiments the melting temperature can even be above of above 76°C, such as above 77°C, such as above 78°C, such as above 79°C, such as above 80°C, such as above 81°C, such as above 82°C, such as above 83°C, and even about 84°C. Increased stability is associated with improved taste.

25 [0064] As described, supra, the variable loop domains or CD-loop domains between alpha helices contribute to flexibility in the molecule and determine to some degree the access to the haem-group and its stability. Haem- and phytoalbumin proteins can display a great sequence diversity, however, the variable loop between alpha-helices B and E (CD Variable loop or CD-loop) is characterized in some instances by a high degree of conservation indicating the relative
30 importance of those residues. In other instances, the CD-loop can be characterized by great variability. See Example 6 Alignments on the CD variable loop.

[0065] Thus, the invention in one aspect relates to a recombinant a haem- or phytoalbumin protein comprising a hexacoordinated haem group, wherein said a haem- or phytoalbumin globin comprises a CD variable loop region which has at least 50% or at least 60% identity to amino acid sequence
35 SEQ ID NO: 56 to 147.

[0066] The invention in further embodiments relates to the recombinant a haem- or phytoalbumin protein according to the invention, wherein said a haem- or phytoalbumin protein comprises a CD variable loop region having a sequence identity of at least 65%, such as at least 70%, such as at least 78%, such as at least 80%, such as at least 84%, such as at least 86%, such as at least 90%,
5 such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99% to amino acid sequence SEQ ID NO: 56 to 147.

[0067] Particular embodiments relate to the recombinant a haem- or phytoalbumin protein according to the invention, wherein said haem- or phytoalbumin globin comprises a CD variable loop region having a sequence identity of at least 50%, such as at least 60%, such as at least 65%, such as
10 at least 70%, for example at least 80%, or such as at least 90% sequence identity to amino acid sequence SEQ ID NO: 56.

[0068] Without wishing to be bound by theory, the inventor has identified several residues that appear to be key to contributing to the improved stability of the haem- or phytoalbumin protein of the invention.

15 One embodiment relates to the recombinant haem- or phytoalbumin protein according to the invention, wherein said haem- or phytoalbumin protein has a sequence identity of at least 65%, such as at least 70%, such as at least 78%, such as at least 80%, such as at least 84%, such as at least 86%, such as at least 90%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99% to amino acid sequence SEQ ID NO:
20 1 to 55.

[0069] Further embodiments relate to a recombinant haem- or phytoalbumin protein comprising a hexacoordinated haem group, wherein said haem- or phytoalbumin protein comprises a region which has the amino acid sequence SEQ ID NO: 56 to 147.

[0070] It appears that this region is characterized by a high degree of variability but still with high
25 degree of conservation in specific points indicating the relative importance of those conserved residues.

[0071] Very conserved and small changes result in changes in stability. For example, the variable loop from Arabidopsis Thaliana (SEQ ID NO 119) differs from Spinach (SEQ ID NO 111) only in having two non-conservative substitutions. However, the replacing of Alanine with a Threonine
30 and an Aspartic acid with a Proline, appears to destabilize this phytoalbumin. Moreover, the variable loop in sugar beet phytoalbumin (SEQ ID NO: 62) differs from SEQ ID NO: 125 only in having two non-conservative substitutions. However, the replacing of P with a G in positions 01 and 16, according to the parent CD-loop from SEQ ID NO: 56, appears to destabilize this phytoalbumin.

35 SEQ ID NO: 111: FLRIFEIAPT AKKMFSFLRD SNVPLEQNPK LKAHAMSV

[0072] The variable loop region in corn has only 79% identity with SEQ ID NO 13, and this appears to correlate with lower temperature stability for corn phytoalbumin. The variable loop region in corn has only 70% identity with SEQ ID NO 56, and this also appears to correlate with lower temperature stability for corn phytoalbumin.

[0073] Conversely, the variable loop region of sugar beet BvHB1.2 has a SEQ ID NO: 116 or 62 and appears to show increased stability and bioavailability. Thus, a preferred embodiment, relates to a recombinant haem- or phytoalbumin protein comprising a hexacoordinated haem group, wherein said phytoalbumin comprises the SEQ ID NO 56 or 116. The improved stability of this embodiment is surprising, as it comprises the replacement of a highly conserved Proline with a Glycine. Prolines are bulky and introduce bends in a molecule, and where conserved are typically highly important for preserving the structure. It is therefore surprising that a non-conservative substitution would lead to improved function.

[0074] The invention in a further embodiment relates to a recombinant haem- or phytoalbumin protein comprising a hexacoordinated haem group, wherein said haem- or phytoalbumin protein comprises a region which has the amino acid sequence similar to SEQ ID NO: 123, 127, 128, 129, 130, 131, 133, 134, 135, 137, 139, 141, 142, or 145, in particular to SEQ ID NO: 127, 128, 130, 131, 135, 139, 141, 142, or 145.

[0075] Further embodiments relate to any of the preceding recombinant haem- or phytoalbumin proteins comprising a hexacoordinated haem group, wherein said phytoalbumin comprises a CD variable loop region described herein, for use in treatment of iron deficiency and/or anaemia, and/or for use in a food.

Compositions

[0076] In a further aspect provided herein is a composition comprising the recombinant protein disclosed herein and one or more carriers, agents, additives and/or excipients. In particular this aspect of the present invention relates to a composition comprising at least one recombinant hexacoordinated haem- or phytoalbumin protein as described herein.

[0077] In some embodiments the one or more carriers, agents, additives and/or excipients are selected from salts, antioxidants and/or reducing agents. Salts are suitably selected from NaCl, ammonium sulphate, CaCl_2 , KCl, MgCl_2 . and are in some embodiments added to the composition in amounts of between 10 μM to 2M or 10 μmole to 2 mole per kg. Antioxidants are suitably selected from ascorbic acid, 2-mercaptoethanol, dithiothreitol (DTT), superoxide dismutase (SOD), catalase, β -carotene, lycopene, glutathione, melatonin, oestrogen, and biquinol-10, N-acetyl cysteine, lipoic acid, salts of zinc, selenium, and/or copper, quercetin, catechin, cortisone,

oestradiol, estriol, and α -tocopherol, and are in some embodiments added to the composition in amounts between 10 μ M to 1M or 10 millimole to 2 mole per kg. Reducing agent are suitably selected from ascorbic acid (vitamin C), tocopherols, carotenoids, flavonoids, and glutathione and are in some embodiments added to the composition in amounts between 0,01 to 10 mg/g.

5 [0078] The composition may further comprise at least one vitamin and/or mineral selected from vitamin C, vitamin D3, vitamin E, vitamin B6, vitamin B12, folate, folic acid, biotin, zinc, copper, magnesium, selenium and/or combinations thereof.

[0079] In further embodiments the composition comprises at least 0.002 % w/w of the recombinant protein, such as from 0.1 to 100 % wt. – or alternatively comprising at least 0.1 μ M
10 of recombinant protein, such as for example from 0.1 μ M to 8 M recombinant protein.

The composition can in some embodiments be a food product, a feed product, a beverage, a food ingredient, a dietary or nutritional supplement, and/or a pharmaceutical product. In some embodiments the composition is a nutritional supplement, while in other embodiments the composition is a food ingredient. The food composition is in some embodiments suitable for and
15 nutritionally balanced for children, elderly or subjects undergoing medical or surgical treatment. The recombinant protein described herein has the improved property of having a lowered or no metal flavour or taste, as it is less prone to release haem and iron and accordingly, in some embodiments, the composition has lowered or no metallic flavour or taste.

[0080] The composition can be formulated as a dry formulation, a liquid formulation or a slurry or
20 dispersion. The composition can also be formulated into a tablet, a capsule, a liquid, a drop, a concentrate, a powder, a granule and/or combinations thereof.

[0081] In a further embodiment the carrier of the composition is a pharmaceutical acceptable carrier, and in additionally or alternatively the composition is a pharmaceutical composition selected from the group consisting of drugs, vaccines, personal care composition and
25 combinations thereof.

[0082] Compositions according to the invention will be able to be formulated with high concentrations of iron, due to neutral taste/lack of metallic taste. Thus, some embodiments relate to wherein the composition according to the invention comprises at least 0.1 μ M of recombinant haem- or phytoalbumin protein, such as for example from 0.1 μ M to 8 M recombinant haem- or
30 phytoalbumin protein – corresponding to 0.017 to 170 000 mg/g.

[0083] As described, *supra*, the compositions of the invention may be a dietary composition, such as a feed, food, beverage, nutritional supplement, or a food ingredient, and may be formulated in any suitable manner. For example, the composition according to the invention has a formulation selected from the group consisting of tablet, capsule, liquid, drops, concentrate, powder, granule,
35 and combinations thereof.

[0084] As described, *supra*, the compositions of the invention may be formulated as a pharmaceutical, using the conventional pharmaceutically acceptable excipients and/or carriers.

[0085] As described, *supra*, the composition according to the invention may comprise one or more further ingredients, such as antioxidants, reducing agents or nutritive or any compound suitable for supporting nutrition and/or supporting uptake of the iron. For example, some embodiments of the invention relate to composition according to the invention, further comprising at least one vitamin and/or mineral, for example one or more selected from the group consisting of vitamin C, vitamin D3, vitamin E, vitamin B6, vitamin B12, folate, folic acid, biotin, zinc, copper, magnesium, selenium, and combinations thereof. Further embodiments relate to composition according to the invention, further comprising one or more reducing agent, such as for example one or more of ascorbic acid (vitamin C), tocopherols, carotenoids, flavonoids, and glutathione.

[0086] The compositions of the invention may further comprise any suitable food grade preservatives, excipients, texturiser, flavourings, or colourings deemed useful.

[0087] The phytoalbumin from sugar beet, spinach and quinoa have been shown to be a source of bioavailable iron in *Example 4*; therefore, they and molecules alike can be used to prevent and treat iron deficiency and anaemia (IDA) if used as organic, biobased iron supplements but also as food ingredients and or feed, as they do not affect the taste of the food they are added to.

[0088] One particular embodiment relates to a liquid composition comprising a non-recombinant and/or a non-recombinant haem- or phytoalbumin protein comprising a hexacoordinated haem group, wherein said haem- or phytoalbumin protein has at least 86% identity to amino acid sequence SEQ ID NO: 1; or a fragment, variant, or fusion thereof which retains the hexacoordinated haem group. Such liquid compositions may be for use to treat or prevent iron deficiency.

[0089] Liquid compositions according to the invention have the advantage of being even more readily absorbable than e-g- encapsulated formulations. The neutral taste of the haem- or phytoalbumin protein of the invention make it possible to provide liquid formulations.

[0090] The liquid compositions may further comprise one or more stabilisers, excipients, antioxidants, or other known formulation aids.

30 **Uses-conditions.**

[0091] Due to their stability and bioavailability, the recombinant haem- or phytoalbumin proteins as described herein and compositions comprising them are suitable for several uses.

[0092] Accordingly, in a further aspect provided herein is a non-therapeutic method of improving the stamina or low oxygen tolerance in a subject, comprising administering any of the recombinant proteins or the compositions disclosed herein to the subject in an amount effective to improve

stamina or low oxygen tolerance in the subject. In some embodiment the subject is performing or is preparing to perform a physical activity, optionally under low oxygen pressure, such as mountaineering or mountain climbing. In further embodiments the non-therapeutic method comprises administering the recombinant protein or the composition into skin of the subject in an amount effective to alter the skin colour or texture for beauty purposes.

[0093] In a further aspect provided herein is a method or the recombinant protein or composition described herein for use in treating, ameliorating, or preventing of a disease or deficiency and/or anaemia in a subject, comprising administering any of the recombinant haem- or phytoalbumin proteins or the compositions to the subject in an amount effective to treat, ameliorate or prevent said disease or deficiency.

[0094] It is to be understood that the instant application describes both methods for using the recombinant protein for therapeutic purposes and describes recombinant protein for use in therapeutic applications, and that the alternative disclosures are used interchangeably for describing the same mirroring subject matter in alternative language.

[0095] The disease or deficiency to be treated is preferably iron deficiency or any disease resulting from such iron deficiency. For therapy in some embodiments the recombinant protein is preferably a haem- or phytoalbumin protein, in particular a haem- or phytoalbumin protein comprising a hexacoordinated haem group and has an amino acid sequence which is at least 86% identical to the phytoalbumin comprised in SEQ ID NO: 1, especially where the recombinant phytoalbumin has an amino acid sequence which is at least 87%, such as at least 88%, such as at least 89%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99% identical to the phytoalbumin comprised in sequence SEQ ID NO: 1. The recombinant protein used to treat, ameliorate or prevent the deficiency and/or disease is preferably a nonsymbiotic haemoglobin or class 2. Additionally or alternatively the recombinant phytoalbumin is derived from a plant in Caryophyllidae, such as spinach or quinoa. The recombinant phytoalbumin for therapy preferably has an amino acid sequence comprised in or consisting of SEQ ID NO: 1 or 4.

[0096] In further embodiments recombinant protein comprises a hexacoordinated haem group and has an improved stability compared to a pentacoordinate haemoglobin. In particular the recombinant protein for therapy has an auto-oxidation rate of less than 0.2, such as less than 0.15, for example in the range from 0.01 h^{-1} to 0.18 h^{-1} , or 0.01 h^{-1} to 0.15 h^{-1} , such as about 0.1 h^{-1} . For therapy the recombinant protein has lowered or no metallic taste or has neutral taste.

[0097] In some embodiments for therapy the recombinant protein has a melting temperature of above 68°C , such as in the range from 68 to 75°C ; such as in the range from 70 to 76°C . In some

embodiments for use in therapy the melting temperature can even be above of above 76°C, such as above 77°C, such as above 78°C, such as above 79°C, such as above 80°C, such as above 81°C, such as above 82°C, such as above 83°C, and even about 84°C.

[0098] In further embodiments the recombinant haem- or phytoalbumin protein comprises a
5 hexacoordinated haem group, and further comprises a CD-loop domain which has an amino acid sequence which is at least 60% sequence identical to the CD-loop domain comprised in SEQ ID NO: 111 or 56, preferably at least 90% identical to the CD-loop domain comprised in SEQ ID NO: 111 or 56.

[0099] In other embodiments the CD-loop domain in the recombinant protein has an amino acid
10 sequence which comprises the CD-loop domain comprised in SEQ ID NO: 111 or 56 having one or more, two or more, three or more, or four or more amino acid substitutions. Additionally or alternatively, the CD-loop domain has an amino acid sequence which comprises or consists of SEQ ID NO: 56 to 147.

[0100] In some embodiments the subject is a blood donor. In other embodiments the subject is a
15 vegetarian, vegan and/or flexitarian. The subject can also be an animal, optionally a companion animal, poultry, or livestock. In particular, the subject is a woman optionally selected from the group of pregnant women, breast feeding women, and women in reproductive age.

[0101] The diseases to be treated include endometriosis, dysmenorrhea, menorrhagia, cancer, kidney disease, diabetes, obesity, metabolic syndrome, digestive disorders, psychological
20 disorders, genetic disorders, conditions associated with ageing, acute conditions, and/or infections. Digestive disorders include celiac disease, inflammatory bowel disease, and/or peptic ulcers. Genetic disorders include selected from haemolytic anaemia, and/or autoimmune disease. Psychological disorders include bulimia and/or anorexia. Infections include malaria and/or streptococcus infections. The disease may be chronic or acute. Acute conditions include trauma,
25 gastrointestinal surgery, gastrointestinal bleeding, kidney failure, and/or acute toxicity. Further embodiments relate to the recombinant hexacoordinated haem- or phytoalbumin protein according to the invention, and/or compositions comprising said hexacoordinated phytoalbumins, for various uses, for example use in treatment, amelioration, or prevention of iron deficiency and/or anaemia in a subject.

[0102] For example, the haem- or phytoalbumin protein as described herein and compositions comprising them are suitable for use in food, for use in supplementing dietary intake of iron (food supplement, nutraceutical), for fortification of foodstuffs, for treating, preventing or ameliorating iron deficiency; for treating, preventing or ameliorating anaemia; and/or for use in medicine.

[0103] In one embodiment, the present invention relates to a composition comprising the
35 recombinant hexacoordinated phytoalbumin as described herein for use as a medicament.

[0104] In particular, the haem- or phytoalbumin protein and compositions according to the invention are suitable for the treatment, amelioration or prevention of iron deficiency and/or anaemia in subjects with eg. chronic conditions and/or acute conditions where the subject is at risk for iron deficiency.

- 5 Certain populations are at particular risk for iron deficiency, and the haem- or phytoalbumin proteins and compositions comprising them according to the invention are suitable for use by these groups. For example, the neutral taste makes it suitable for use in particular in children's foods or food supplements.

[0105] Further embodiments relate to a haem- or phytoalbumin protein comprising a
10 hexacoordinated haem group, wherein said haem- or phytoalbumin protein has at least 86% identity to amino acid sequence SEQ ID NO: 1; or a fragment, variant, or fusion thereof which retains the hexacoordinated haem group, for use in treatment or prevention of iron deficiency in a subject.

[0106] The phytoalbumin according to the invention may be recombinantly produced or may be produced by non-recombinant methods.

- 15 [0107] Recombinant haem- or phytoalbumin proteins as described herein has the advantage that it may be produced in high quantities, and to a high degree of purity. Further, recombinant phytoalbumin would display less batch-to batch variability. Phytoalbumin content in plant tissues varies both in the distribution of the plant, and over time, making it difficult to produce commercially viable quantities. Further, plants may contain other compounds which would inhibit
20 the iron uptake.

[0108] Notwithstanding, the present invention also relates to a non-recombinant haem- or phytoalbumin protein comprising a hexacoordinated haem group, wherein said phytoalbumin has at least 86% identity to amino acid sequence SEQ ID NO: 1; or a fragment, variant, or fusion thereof which retains the hexacoordinated haem group, for use in food.

- 25 [0109] The non-recombinant haem- or phytoalbumin protein according to the invention will have all the same benefits as described herein regarding recombinant forms, and thus all embodiments described in relation to recombinant form also apply to non-recombinant haem- or phytoalbumin protein according to the invention, except where specifically stated otherwise.

[0110] The recombinant protein is suitably administered to the subject is between 0,1 mg to 350
30 mg per kg body weight of the subject, such as between 0.25 mg to 100 mg, such as between 0,5 to 50 mg, such as between 0.75 mg to 10 mg, such as between 0,75 mg to 3 mg per kg body weight. In an embodiment the subject is a male and optionally the amount of recombinant protein administered to the subject is between 0,75 mg to 1,25, such as about 1 mg per kg body weight. In another embodiment the subject is a female, and optionally amount of recombinant protein
35 administered to the subject is between 1,3 mg to 1,5, such as about 1,4 mg per kg body weight.

Genes and constructs

[0111] In a further aspect, provided herein is a gene encoding any of the recombinant proteins disclosed herein. The gene preferably has a nucleotide sequence which is at least 50% identical to the nucleotide sequence comprised in any of SEQ ID NO: 148 to 202 encoding the corresponding recombinant protein of item SEQ ID NO: 1 to 55.

[0112] In a further aspect, provided herein is a polynucleotide construct comprising the gene described herein, operably linked to a control sequence directing the transcription and/or translation of the gene in a host cell. The control sequence may be a promoter, which is a polynucleotide that is recognized by a host cell for expression of a polynucleotide. The promoter contains transcriptional control sequences that mediate the expression of the polypeptide. The promoter may be any polynucleotide that shows transcriptional activity in the host cell including mutant, truncated, and hybrid promoters, and may be obtained from genes encoding extracellular or intracellular polypeptides either homologous or heterologous to the host cell. The promoter can be an inducible or a constitutive promoter, and optionally the promoter has a nucleotide sequence which is at least 50 % identical to the promoter comprised in SEQ ID NO: 225 to 235. Further, the polynucleotide construct is an expression vector. The expression vector may be any vector (e.g., a plasmid or virus or chromosomal) that can be conveniently subjected to recombinant DNA procedures and can bring about expression of the gene encoding the recombinant protein. The choice of the vector will typically depend on the compatibility of the vector with the host cell into which the vector is to be introduced. The vector may be an autonomously replicating vector, i.e., a vector that exists as an extrachromosomal entity, the replication of which is independent of chromosomal replication, e.g., a plasmid (linear or closed circular plasmid), an extrachromosomal element, a mini-chromosome, or an artificial chromosome. The vector may contain any means for assuring self-replication. Alternatively, the vector may, when introduced into the host cell, integrate into the genome, and replicate together with the chromosome(s) into which it has been integrated. Furthermore, a single vector or plasmid or two or more vectors or plasmids that together contain the total DNA to be introduced into the genome of the host cell, or a transposon, may be used.

[0113] The vector may contain one or more selectable markers that permit easy selection of transformed, transfected, transduced, or the like cells. A selectable marker is a gene from which the product provides for biocide or viral resistance, resistance to heavy metals, prototrophy to auxotrophs, and the like. Useful selectable markers for fungal host cells include amdS (acetamidase), argB (ornithine carbamoyltransferase), bar (phosphinothricin acetyltransferase), hph (hygromycin phosphotransferase), niaD (nitrate reductase), pyrG (orotidine-5'-phosphate

decarboxylase), *sC* (sulfate adenylyltransferase), and *trpC* (anthranilate synthase), as well as equivalents thereof. Useful selectable markers for yeast host cells include, but are not limited to, *ADE2*, *HIS3*, *LEU2*, *LYS2*, *MET3*, *TRP1*, and *URA3*.

[0114] The vector may further contain element(s) that permits integration of the vector into
5 genome of the host cell or permits autonomous replication of the vector in the cell independent of the genome. For integration into the host cell genome, the vector may rely on the gene encoding the recombinant protein or any other element of the vector for integration into the genome by homologous or non-homologous recombination. Alternatively, the vector may contain additional polynucleotides for directing integration by homologous recombination into the
10 genome of the host cell at precise location(s) in the chromosome(s). To increase the likelihood of integration at a precise location, the integrational elements should contain a sufficient number of nucleic acids, such as 100 to 10,000 base pairs, such as 400 to 10,000 base pairs, and such as 800 to 10,000 base pairs, which have a high degree of sequence identity to the corresponding target sequence to enhance the probability of homologous recombination. The integrational elements
15 may be any sequence that is homologous with the target sequence in the genome of the host cell. Furthermore, the integrational elements may be non-encoding or encoding polynucleotides. On the other hand, the vector may be integrated into the genome of the host cell by non-homologous recombination. For autonomous replication, the vector may further comprise an origin of replication enabling the vector to replicate autonomously in the host cell in question. The origin
20 of replication may be any plasmid replicator mediating autonomous replication that functions in a cell. The term "origin of replication" or "plasmid replicator" refers to a polynucleotide that enables a plasmid or vector to replicate *in vivo*. Useful origins of replication for fungal cells include AMA 1 and ANS1 (Gems et al., 1991, Gene 98: 61-67; Cullen et al., 1987, Nucleic Acids Res. 15: 9163-9175; WO 00/24883). Isolation of the AMA 1 sequence and construction of plasmids or
25 vectors comprising the gene can be accomplished using the methods disclosed in WO2000/24883. Useful origins of replication for yeast host cells are the 2-micron origin of replication, ARS1, ARS4, the combination of ARS1 and CEN3, and the combination of ARS4 and CEN6.

[0115] More than one copy of a polynucleotide encoding the recombinant protein of the invention may be inserted into a host cell to increase production of the recombinant protein. An increase in
30 the copy number can be obtained by integrating one or more additional copies of the protein coding sequence into the host cell genome or by including an amplifiable selectable marker gene with the polynucleotide, so that cells containing amplified copies of the selectable marker gene - and thereby additional copies of the polynucleotide - can be selected by cultivating the cells in the presence of the appropriate selectable agent. The procedures used to ligate the elements
35 described above to construct the recombinant expression vectors of the present disclosure are

well known to one skilled in the art (see, e.g., Green & Sambrook, 2012, Molecular cloning: A laboratory Manual, Fourth Edition, Cold Spring Harbor Laboratory, New York, USA).

[0116] In alignment with the above also disclosed herein is host cells comprising the polynucleotide construct as described, *supra*.

5

Host cells expressing recombinant proteins

[0117] In a further aspect, provided herein is a genetically modified host cell expressing the gene or polynucleotide construct and the recombinant protein described herein.

The host cell can be eukaryotic, bacterial cell, or archaeal. Among eukaryotic cells fungal cells, plant cells, mammalian cells or insect cells are useful. Among plant cells those from the genus of *Arabidopsis*, *Lepidium*, *Nicotiana*, *Triticum*, *Hordeum*, *Oryza*, *Chenopodium*, *Beta*, or *Glycine* are particularly useful, such as the species of *Arabidopsis thaliana*, *Lepidium campestre* *Nicotiana tabacum*, *Triticum aestivum*, *Hordeum vulgare*, *Oryza sativa*, *Chenopodium quinoa*, *Beta vulgaris*, or *Glycine max*. Among fungal cells yeasts or a filamentous fungi are preferred. Preferred filamentous fungi are those from the genus of *Aspergillus*, *Trichoderma*, or *Rhizopus*, such as the species *Aspergillus* sp, *Aspergillus oryzae*, *Trichoderma* sp, or *Rhizopus* sp. Preferred yeasts are those from the genus of *Pichia*, *Saccharomyces*, *Yarrowia*, *Kluyveromyces*, *Ashbya* or *Hansenula*, such as the species of *P. pastoris*, *Pichia* sp., *S. cerevisiae*, *Yarrowia* sp., *Y. lipolytica*, *Kluyveromyces lactis*, *Saccharomyces carlsbergensis*, *Saccharomyces cerevisiae*, *Saccharomyces diastaticus*, *Saccharomyces douglasii*, *Saccharomyces kluyveri*, *Saccharomyces bensis*, *Saccharomyces oviformis*, *Ashbya gossypii*, *H. polymorpha*, or *Hansenula* sp. Among bacteria preferred ones are those from the genus of *Escherichia*, *Bacillus*, *Brevibacterium*, *Burkholderia*, *Campylobacter*, *Corynebacterium*, *Serratia*, *Lactobacillus*, *Lactococcus*, *Acinetobacter*, *Acetobacter* or *Pseudomonas*, such as the species *E. coli*, *C. glutamicum*, *B. subtilis*, *S. marcescens*, *P. putida*, *P. aeruginosa* and/or *P. mutabilis*. Archaeal host cells include algae. In particular ClearColi (endotoxin-free *E. coli*) and in *E. coli* Nissle 1917 (Schultz and Burton 2017) are useful.

25

Cell cultures comprising host cells

[0118] In a further aspect, provided herein is a cell culture comprising the host cell described herein and a growth medium. Suitable growth mediums for host cells such as mammalian, insect, plant, fungal and/or yeast cells are all well known in the art.

30

Methods of producing recombinant protein

[0119] In a further aspect, provided herein is a method for producing the recombinant protein described herein comprising:

35

- a) culturing the cell culture described herein at conditions allowing the cell to produce the recombinant protein; and
- b) optionally recovering and/or isolating the recombinant protein.

[0120] The cell culture can be cultivated in a nutrient medium and at conditions suitable for production of the recombinant protein described herein and/or propagating cell count using methods known in the art. For example, the culture may be cultivated by shake flask cultivation, or small-scale or large-scale fermentation (including continuous, batch, fed-batch, feed and draw, or solid-state fermentations) in laboratory or industrial fermentors in a suitable medium and under conditions allowing the host cells to grow and/or propagate, optionally to be recovered and/or isolated.

[0121] The cultivation can take place in a suitable nutrient medium comprising carbon and nitrogen sources and inorganic salts, using procedures known in the art. Suitable media are available from commercial suppliers or may be prepared according to published recipes (e.g. from catalogues of the American Type Culture Collection). The selection of the appropriate medium may be based on the choice of host cell and/or based on the regulatory requirements for the host cell. Such media are available in the art. The medium may, if desired, contain additional components favouring the transformed expression hosts over other potentially contaminating microorganisms. Accordingly, in an embodiment a suitable nutrient medium comprises a carbon source (e.g. glucose, maltose, molasses, starch, cellulose, xylan, pectin, lignocellolytic biomass hydrolysate, etc.), a nitrogen source (e.g. ammonium sulphate, ammonium nitrate, ammonium chloride, etc.), an organic nitrogen source (e.g. yeast extract, malt extract, peptone, etc.) and inorganic nutrient sources (e.g. phosphate, magnesium, potassium, zinc, iron, etc.). In some embodiments the medium contains elevated levels of iron to provide for the production of the iron containing recombinant protein.

The synthesis of haemoglobin in a host cell includes two parts, the synthesis of haem *b* and the synthesis of the apoprotein (globin protein) as shown in figure 6. The ring system of haem *b* is that of protoporphyrin IX. This prosthetic group is tightly, but non-covalently bound within its host protein (the apoprotein). For making high titres of the recombinant haem- or phyto globin protein, increased levels of iron can advantageously be supplied to the medium. The iron can be supplied to the medium for example in the form of FeSO₄, FeO, or Fe₂O₃ and in amounts preferably exceeding [0.01 g/L]. Further the medium may be kept low in glucose ([to avoid protein glycosilation), preferably 50 mM glucose, such as below 40 mM, such as below 30 mM, such as below 20 mM, such as below 10 mM, such as below 5 mM, such as below 4 mM, such as below 3 mM, such as below 2 mM, such as below 1mM glucose. To further optimize the cultivation and the expression and production of the recombinant haem- or phyto globin protein δ -aminolevulinic

acid (D-ALA) can be added, in case formation of ALA (see figure 6) by ALA synthase becomes the rate-limiting enzyme of haem synthesis. To further optimize the cultivation and the expression and production of the recombinant protein carbon monoxide may be added to the medium, supplied by feeding carbon monoxide gas to the medium. To further optimize the cultivation and the expression and production of the recombinant protein the level dissolved oxygen (DO) is controlled preferably kept below 20%, such as below 15%, such as below 10 %wt, such as below 9 %wt, such as below 8 %wt, such as below 7 %wt, such as below 6 %wt, such as below 5 %wt, such as below 4 %wt, such as below 3 %wt, such as below 2 %wt, such as below 1 %wt, such as below 0,1 %wt, such as below, such as below 0,01 %wt, such as below 0,001%.

[0122] The cultivation of the host cell may be performed over a period of time from about 0.5 to about 30 days. The cultivation process may be a batch process, continuous or fed-batch process, suitably performed at a temperature in the range of 0-100 °C or 0-80 °C, for example, from about 0 °C to about 50 °C and/or at a pH, for example, from about 2 to about 10. Preferred fermentation conditions for yeast and filamentous fungi are a temperature in the range of from about 25 °C to about 55 °C and at a pH of from about 3 to about 9. Preferred fermentation conditions for bacteria, such as *E. coli* are a temperature of 17-38 °C and a pH of 5-8.

[0123] The appropriate conditions are usually selected based on the choice of host cell. Accordingly, in an embodiment the method described herein further comprises one or more elements selected from:

- a) culturing the cell culture in a nutrient growth medium;
- b) culturing the cell culture under aerobic or anaerobic conditions
- c) culturing the cell culture under agitation;
- d) culturing the cell culture at a temperature of between 25 to 50 °C;
- e) culturing the cell culture at a pH of between 3-9;
- f) culturing the cell culture for between 10 hours to 30 days; and.

[0124] The recombinant protein may be recovered and or isolated using methods known in the art, including, but not limited to, centrifugation, filtration, spray-drying, or lyophilization. For example, in a particular embodiment the method includes a recovery and/or isolation step comprising separating a liquid phase of the cell or cell culture from a solid phase of the cell or cell culture to obtain a supernatant comprising the recombinant protein and subjecting the supernatant to one or more steps selected from:

- a) separating the supernatant from the solid phase of the cell culture, such as by filtration or gravity separation;
- b) contacting the supernatant with one or more adsorbent resins to obtain at least a portion of the produced recombinant protein;

- c) contacting the supernatant with one or more ion exchange or reversed-phase chromatography columns or filtration or ultrafiltration devices to obtain at least a portion of the recombinant protein;
 - d) extracting the recombinant protein; and/or
 - 5 e) precipitating the recombinant protein by crystallization or evaporating the solvent of the liquid phase, optionally through spray drying or lyophilization; and optionally isolating the recombinant protein by filtration or gravity separation;
- thereby recovering and/or isolating the recombinant protein.

[0125] In further embodiments the composition described, *supra*, can also be a composition comprising the cell culture described herein, containing the recombinant protein – in particular an edible composition, suitable for ingestion by a subject in need thereof for supplementing iron. In this composition the cell culture may be in the form of an edible mycelium distributed in a matrix such as a solid-state substrate, suitable for ingestion by humans or animals.

[0126] The listing or discussion of an apparently prior published document in this specification should not necessarily be taken as an acknowledgement that the document is part of the state of the art or is common general knowledge.

[0127] Preferences, options, and embodiments for a given aspect, feature or parameter of the invention should, unless the context indicates otherwise, be regarded as having been disclosed in combination with any and all preferences, options and embodiments for all other aspects, features and parameters of the invention. This is especially true for the description of both the recombinant and non-recombinant hexacoordinated haem- or phytoalbumin protein and all its features, which may readily be part of the final composition or use as described herein. Embodiments and features of the present invention are also outlined in the following items and also illustrated by the following non-limiting examples.

25

Examples

Materials and Methods

Sequences

[0128] The present application contains a list of sequences prepared in WIPO Sequence 2.2.0 or later including the sequences described below and submitted electronically in ST26 format which is hereby incorporated by reference in its entirety.

30

SEQ ID NO: 1	protein sequence of globin from Chenopodium quinoa
SEQ ID NO: 2	protein sequence of globin from Chenopodium quinoa
SEQ ID NO: 3	protein sequence of globin from Chenopodium quinoa

SEQ ID NO: 4	protein sequence of globin from <i>Chenopodium quinoa</i>
SEQ ID NO: 5	protein sequence of globin from <i>Spinacia oleracea</i>
SEQ ID NO: 6	protein sequence of globin from <i>Spinacia oleracea</i>
SEQ ID NO: 7	protein sequence of globin from <i>Beta vulgaris</i>
SEQ ID NO: 8	protein sequence of globin from <i>Beta vulgaris</i>
SEQ ID NO: 9	protein sequence of globin from <i>Beta vulgaris</i>
SEQ ID NO: 10	protein sequence of globin from <i>Beta vulgaris</i>
SEQ ID NO: 11	protein sequence of globin from <i>Amaranthus tricolor</i>
SEQ ID NO: 12	protein sequence of globin from <i>Amaranthus tricolor</i>
SEQ ID NO: 13	protein sequence of globin from <i>Vitis vinifera</i>
SEQ ID NO: 14	protein sequence of globin from <i>Vitis vinifera</i>
SEQ ID NO: 15	protein sequence of globin from <i>Citrus sinensis</i>
SEQ ID NO: 16	protein sequence of globin from <i>Citrus sinensis</i>
SEQ ID NO: 17	protein sequence of globin from <i>Solanum_lycopersicum</i>
SEQ ID NO: 18	protein sequence of globin from <i>Solanum_tuberosum</i>
SEQ ID NO: 19	protein sequence of globin from <i>Ipomoea batatas</i>
SEQ ID NO: 20	protein sequence of globin from <i>Gossypium_hirsutum</i>
SEQ ID NO: 21	protein sequence of globin from <i>Arabidopsis thaliana</i>
SEQ ID NO: 22	protein sequence of globin from <i>Arabidopsis thaliana</i>
SEQ ID NO: 23	protein sequence of globin from <i>Malus domestica</i>
SEQ ID NO: 24	protein sequence of globin from <i>Populus trichocarpa</i>
SEQ ID NO: 25	protein sequence of globin from <i>Avena sativa</i>
SEQ ID NO: 26	protein sequence of globin from <i>Avena sativa</i>
SEQ ID NO: 27	protein sequence of globin from <i>Avena sativa</i>
SEQ ID NO: 28	protein sequence of globin from <i>Zea mays</i>
SEQ ID NO: 29	protein sequence of globin from <i>Zea mays</i>
SEQ ID NO: 30	protein sequence of globin from <i>Oryza sativa</i>
SEQ ID NO: 31	protein sequence of globin from <i>Oryza sativa</i>
SEQ ID NO: 32	protein sequence of globin from <i>Oryza sativa</i>
SEQ ID NO: 33	protein sequence of globin from <i>Hordeum vulgare</i>
SEQ ID NO: 34	protein sequence of globin from <i>Ananas comosus</i>
SEQ ID NO: 35	protein sequence of globin from <i>Glycine max</i>
SEQ ID NO: 36	protein sequence of globin from <i>Glycine max</i>
SEQ ID NO: 37	protein sequence of globin from <i>Glycine max</i>
SEQ ID NO: 38	protein sequence of globin from <i>Glycine max</i>
SEQ ID NO: 39	protein sequence of globin from <i>Glycine max</i>
SEQ ID NO: 40	protein sequence of globin from <i>Medicago sativa</i>
SEQ ID NO: 41	protein sequence of globin from <i>Vicia faba</i>
SEQ ID NO: 42	protein sequence of globin from <i>Vicia faba</i>
SEQ ID NO: 43	protein sequence of globin from <i>Pisum sativum</i>
SEQ ID NO: 44	protein sequence of globin from <i>Pisum sativum</i>
SEQ ID NO: 45	protein sequence of globin from <i>Phaseolus vulgaris</i>
SEQ ID NO: 46	protein sequence of globin from <i>Lotus japonicus</i>
SEQ ID NO: 47	protein sequence of globin from <i>Lotus japonicus</i>
SEQ ID NO: 48	protein sequence of globin from <i>Bos taurus</i>
SEQ ID NO: 49	protein sequence of globin from <i>Bos taurus</i>

SEQ ID NO: 50	protein sequence of globin from Bos taurus
SEQ ID NO: 51	protein sequence of globin from Homo sapiens
SEQ ID NO: 52	protein sequence of globin from Homo sapiens
SEQ ID NO: 53	protein sequence of globin from Homo sapiens
SEQ ID NO: 54	protein sequence of globin from Homo sapiens
SEQ ID NO: 55	protein sequence of globin from Vitreoscilla sp.
SEQ ID NO: 56	protein sequence of CD-loop domain from Chenopodium quinoa
SEQ ID NO: 57	protein sequence of CD-loop domain from Chenopodium quinoa
SEQ ID NO: 58	protein sequence of CD-loop domain from Chenopodium quinoa
SEQ ID NO: 59	protein sequence of CD-loop domain from Chenopodium quinoa
SEQ ID NO: 60	protein sequence of CD-loop domain from Spinacia oleracea
SEQ ID NO: 61	protein sequence of CD-loop domain from Spinacia oleracea
SEQ ID NO: 62	protein sequence of CD-loop domain from Beta vulgaris
SEQ ID NO: 63	protein sequence of CD-loop domain from Beta vulgaris
SEQ ID NO: 64	protein sequence of CD-loop domain from Beta vulgaris
SEQ ID NO: 65	protein sequence of CD-loop domain from Beta vulgaris
SEQ ID NO: 66	protein sequence of CD-loop domain from Amaranthus tricolor
SEQ ID NO: 67	protein sequence of CD-loop domain from Amaranthus tricolor
SEQ ID NO: 68	protein sequence of CD-loop domain from Vitis vinifera
SEQ ID NO: 69	protein sequence of CD-loop domain from Vitis vinifera
SEQ ID NO: 70	protein sequence of CD-loop domain from Citrus sinensis
SEQ ID NO: 71	protein sequence of CD-loop domain from Citrus sinensis
SEQ ID NO: 72	protein sequence of CD-loop domain from Solanum_lycopersicum
SEQ ID NO: 73	protein sequence of CD-loop domain from Solanum_tuberosum
SEQ ID NO: 74	protein sequence of CD-loop domain from Ipomoea batatas
SEQ ID NO: 75	protein sequence of CD-loop domain from Gossypium_hirsutum
SEQ ID NO: 76	protein sequence of CD-loop domain from Arabidopsis thaliana
SEQ ID NO: 77	protein sequence of CD-loop domain from Arabidopsis thaliana
SEQ ID NO: 78	protein sequence of CD-loop domain from Malus domestica
SEQ ID NO: 79	protein sequence of CD-loop domain from Populus trichocarpa
SEQ ID NO: 80	protein sequence of CD-loop domain from Avena sativa
SEQ ID NO: 81	protein sequence of CD-loop domain from Avena sativa
SEQ ID NO: 82	protein sequence of CD-loop domain from Avena sativa
SEQ ID NO: 83	protein sequence of CD-loop domain from Zea mays
SEQ ID NO: 84	protein sequence of CD-loop domain from Zea mays
SEQ ID NO: 85	protein sequence of CD-loop domain from Oryza sativa
SEQ ID NO: 86	protein sequence of CD-loop domain from Oryza sativa
SEQ ID NO: 87	protein sequence of CD-loop domain from Oryza sativa
SEQ ID NO: 88	protein sequence of CD-loop domain from Hordeum vulgare
SEQ ID NO: 89	protein sequence of CD-loop domain from Ananas comosus
SEQ ID NO: 90	protein sequence of CD-loop domain from Glycine max
SEQ ID NO: 91	protein sequence of CD-loop domain from Glycine max
SEQ ID NO: 92	protein sequence of CD-loop domain from Glycine max
SEQ ID NO: 93	protein sequence of CD-loop domain from Glycine max
SEQ ID NO: 94	protein sequence of CD-loop domain from Glycine max
SEQ ID NO: 95	protein sequence of CD-loop domain from Medicago sativa

SEQ ID NO: 96	protein sequence of CD-loop domain from <i>Vicia faba</i>
SEQ ID NO: 97	protein sequence of CD-loop domain from <i>Vicia faba</i>
SEQ ID NO: 98	protein sequence of CD-loop domain from <i>Pisum sativum</i>
SEQ ID NO: 99	protein sequence of CD-loop domain from <i>Pisum sativum</i>
SEQ ID NO: 100	protein sequence of CD-loop domain from <i>Phaseolus vulgaris</i>
SEQ ID NO: 101	protein sequence of CD-loop domain from <i>Lotus japonicus</i>
SEQ ID NO: 102	protein sequence of CD-loop domain from <i>Lotus japonicus</i>
SEQ ID NO: 103	protein sequence of CD-loop domain from <i>Bos taurus</i>
SEQ ID NO: 104	protein sequence of CD-loop domain from <i>Bos taurus</i>
SEQ ID NO: 105	protein sequence of CD-loop domain from <i>Bos taurus</i>
SEQ ID NO: 106	protein sequence of CD-loop domain from <i>Homo sapiens</i>
SEQ ID NO: 107	protein sequence of CD-loop domain from <i>Homo sapiens</i>
SEQ ID NO: 108	protein sequence of CD-loop domain from <i>Homo sapiens</i>
SEQ ID NO: 109	protein sequence of CD-loop domain from <i>Homo sapiens</i>
SEQ ID NO: 110	protein sequence of CD-loop domain from <i>Vitreoscilla</i> sp.
SEQ ID NO: 111	protein sequence of CD-loop domain from <i>Spinacia oleracea</i>
SEQ ID NO: 112	protein sequence of CD-loop domain from <i>Chenopodium quinoa</i>
SEQ ID NO: 113	protein sequence of CD-loop domain from <i>Chenopodium quinoa</i>
SEQ ID NO: 114	protein sequence of CD-loop domain from <i>Chenopodium quinoa</i>
SEQ ID NO: 115	protein sequence of CD-loop domain from <i>Beta vulgaris</i>
SEQ ID NO: 116	protein sequence of CD-loop domain from <i>Beta vulgaris</i>
SEQ ID NO: 117	protein sequence of CD-loop domain from <i>Beta vulgaris</i>
SEQ ID NO: 118	protein sequence of CD-loop domain from <i>Beta vulgaris</i>
SEQ ID NO: 119	protein sequence of CD-loop domain from <i>Arabidopsis thaliana</i>
SEQ ID NO: 120	protein sequence of CD-loop domain from <i>Glycine max</i>
SEQ ID NO: 121	protein sequence of CD-loop domain from <i>Zea mays</i>
SEQ ID NO: 122	protein sequence of CD-loop domain from <i>Bos taurus</i>
SEQ ID NO: 123	protein sequence of modified CD-loop
SEQ ID NO: 124	protein sequence of modified CD-loop
SEQ ID NO: 125	protein sequence of modified CD-loop
SEQ ID NO: 126	protein sequence of modified CD-loop
SEQ ID NO: 127	protein sequence of modified CD-loop
SEQ ID NO: 128	protein sequence of modified CD-loop
SEQ ID NO: 129	protein sequence of modified CD-loop
SEQ ID NO: 130	protein sequence of modified CD-loop
SEQ ID NO: 131	protein sequence of modified CD-loop
SEQ ID NO: 132	protein sequence of modified CD-loop
SEQ ID NO: 133	protein sequence of modified CD-loop
SEQ ID NO: 134	protein sequence of modified CD-loop
SEQ ID NO: 135	protein sequence of modified CD-loop
SEQ ID NO: 136	protein sequence of modified CD-loop
SEQ ID NO: 137	protein sequence of modified CD-loop
SEQ ID NO: 138	protein sequence of modified CD-loop
SEQ ID NO: 139	protein sequence of modified CD-loop
SEQ ID NO: 140	protein sequence of modified CD-loop
SEQ ID NO: 141	protein sequence of modified CD-loop

SEQ ID NO: 142	protein sequence of modified CD-loop
SEQ ID NO: 143	protein sequence of modified CD-loop
SEQ ID NO: 144	protein sequence of modified CD-loop
SEQ ID NO: 145	protein sequence of modified CD-loop
SEQ ID NO: 146	protein sequence of modified CD-loop
SEQ ID NO: 147	protein sequence of modified CD-loop
SEQ ID NO: 148	DNA sequence encoding globin from <i>Chenopodium quinoa</i>
SEQ ID NO: 149	DNA sequence encoding globin from <i>Chenopodium quinoa</i>
SEQ ID NO: 150	DNA sequence encoding globin from <i>Chenopodium quinoa</i>
SEQ ID NO: 151	DNA sequence encoding globin from <i>Chenopodium quinoa</i>
SEQ ID NO: 152	DNA sequence encoding globin from <i>Spinacia oleracea</i>
SEQ ID NO: 153	DNA sequence encoding globin from <i>Spinacia oleracea</i>
SEQ ID NO: 154	DNA sequence encoding globin from <i>Beta vulgaris</i>
SEQ ID NO: 155	DNA sequence encoding globin from <i>Beta vulgaris</i>
SEQ ID NO: 156	DNA sequence encoding globin from <i>Beta vulgaris</i>
SEQ ID NO: 157	DNA sequence encoding globin from <i>Beta vulgaris</i>
SEQ ID NO: 158	DNA sequence encoding globin from <i>Amaranthus tricolor</i>
SEQ ID NO: 159	DNA sequence encoding globin from <i>Amaranthus tricolor</i>
SEQ ID NO: 160	DNA sequence encoding globin from <i>Vitis vinifera</i>
SEQ ID NO: 161	DNA sequence encoding globin from <i>Vitis vinifera</i>
SEQ ID NO: 162	DNA sequence encoding globin from <i>Citrus sinensis</i>
SEQ ID NO: 163	DNA sequence encoding globin from <i>Citrus sinensis</i>
SEQ ID NO: 164	DNA sequence encoding globin from <i>Solanum_lycopersicum</i>
SEQ ID NO: 165	DNA sequence encoding globin from <i>Solanum_tuberosum</i>
SEQ ID NO: 166	DNA sequence encoding globin from <i>Ipomoea batatas</i>
SEQ ID NO: 167	DNA sequence encoding globin from <i>Gossypium_hirsutum</i>
SEQ ID NO: 168	DNA sequence encoding globin from <i>Arabidopsis thaliana</i>
SEQ ID NO: 169	DNA sequence encoding globin from <i>Arabidopsis thaliana</i>
SEQ ID NO: 170	DNA sequence encoding globin from <i>Malus domestica</i>
SEQ ID NO: 171	DNA sequence encoding globin from <i>Populus trichocarpa</i>
SEQ ID NO: 172	DNA sequence encoding globin from <i>Avena sativa</i>
SEQ ID NO: 173	DNA sequence encoding globin from <i>Avena sativa</i>
SEQ ID NO: 174	DNA sequence encoding globin from <i>Avena sativa</i>
SEQ ID NO: 175	DNA sequence encoding globin from <i>Zea mays</i>
SEQ ID NO: 176	DNA sequence encoding globin from <i>Zea mays</i>
SEQ ID NO: 177	DNA sequence encoding globin from <i>Oryza sativa</i>
SEQ ID NO: 178	DNA sequence encoding globin from <i>Oryza sativa</i>
SEQ ID NO: 179	DNA sequence encoding globin from <i>Oryza sativa</i>
SEQ ID NO: 180	DNA sequence encoding globin from <i>Hordeum vulgare</i>
SEQ ID NO: 181	DNA sequence encoding globin from <i>Ananas comosus</i>
SEQ ID NO: 182	DNA sequence encoding globin from <i>Glycine max</i>
SEQ ID NO: 183	DNA sequence encoding globin from <i>Glycine max</i>
SEQ ID NO: 184	DNA sequence encoding globin from <i>Glycine max</i>
SEQ ID NO: 185	DNA sequence encoding globin from <i>Glycine max</i>
SEQ ID NO: 186	DNA sequence encoding globin from <i>Glycine max</i>
SEQ ID NO: 187	DNA sequence encoding globin from <i>Medicago sativa</i>

SEQ ID NO: 188	DNA sequence encoding globin from <i>Vicia faba</i>
SEQ ID NO: 189	DNA sequence encoding globin from <i>Vicia faba</i>
SEQ ID NO: 190	DNA sequence encoding globin from <i>Pisum sativum</i>
SEQ ID NO: 191	DNA sequence encoding globin from <i>Pisum sativum</i>
SEQ ID NO: 192	DNA sequence encoding globin from <i>Phaseolus vulgaris</i>
SEQ ID NO: 193	DNA sequence encoding globin from <i>Lotus japonicus</i>
SEQ ID NO: 194	DNA sequence encoding globin from <i>Lotus japonicus</i>
SEQ ID NO: 195	DNA sequence encoding globin from <i>Bos taurus</i>
SEQ ID NO: 196	DNA sequence encoding globin from <i>Bos taurus</i>
SEQ ID NO: 197	DNA sequence encoding globin from <i>Bos taurus</i>
SEQ ID NO: 198	DNA sequence encoding globin from <i>Homo sapiens</i>
SEQ ID NO: 199	DNA sequence encoding globin from <i>Homo sapiens</i>
SEQ ID NO: 200	DNA sequence encoding globin from <i>Homo sapiens</i>
SEQ ID NO: 201	DNA sequence encoding globin from <i>Homo sapiens</i>
SEQ ID NO: 202	DNA sequence encoding globin from <i>Vitreoscilla</i> sp.
SEQ ID NO: 203	protein sequence of Signal peptide from <i>Aspergillus terreus</i>
SEQ ID NO: 204	protein sequence of Signal peptide from <i>Aspergillus niger</i>
SEQ ID NO: 205	protein sequence of Signal peptide from <i>Aspergillus niger</i>
SEQ ID NO: 206	protein sequence of Signal peptide from <i>Aspergillus niger</i>
SEQ ID NO: 207	protein sequence of Signal peptide from <i>Aspergillus niger</i>
SEQ ID NO: 208	protein sequence of Signal peptide from <i>Aspergillus niger</i>
SEQ ID NO: 209	protein sequence of Signal peptide from <i>Aspergillus awamori</i>
SEQ ID NO: 210	protein sequence of Signal peptide from <i>Aspergillus awamori</i>
SEQ ID NO: 211	protein sequence of Signal peptide from <i>Aspergillus nidulans</i>
SEQ ID NO: 212	protein sequence of Signal peptide from <i>Trichoderma reesei</i>
SEQ ID NO: 213	protein sequence of Signal peptide from <i>Trichoderma reesei</i>
SEQ ID NO: 214	DNA sequence encoding Signal peptide from <i>Aspergillus terreus</i>
SEQ ID NO: 215	DNA sequence encoding Signal peptide from <i>Aspergillus niger</i>
SEQ ID NO: 216	DNA sequence encoding Signal peptide from <i>Aspergillus niger</i>
SEQ ID NO: 217	DNA sequence encoding Signal peptide from <i>Aspergillus niger</i>
SEQ ID NO: 218	DNA sequence encoding Signal peptide from <i>Aspergillus niger</i>
SEQ ID NO: 219	DNA sequence encoding Signal peptide from <i>Aspergillus niger</i>
SEQ ID NO: 220	DNA sequence encoding Signal peptide from <i>Aspergillus awamori</i>
SEQ ID NO: 221	DNA sequence encoding Signal peptide from <i>Aspergillus awamori</i>
SEQ ID NO: 222	DNA sequence encoding Signal peptide from <i>Aspergillus nidulans</i>
SEQ ID NO: 223	DNA sequence encoding Signal peptide from <i>Trichoderma reesei</i>
SEQ ID NO: 224	DNA sequence encoding Signal peptide from <i>Trichoderma reesei</i>
SEQ ID NO: 225	DNA sequence of promoter from <i>Aspergillus oryzae</i> , PsorA
SEQ ID NO: 226	DNA sequence of promoter from <i>Aspergillus oryzae</i> , PsorB
SEQ ID NO: 227	DNA sequence of promoter from <i>Aspergillus nidulans</i> , PniiA
SEQ ID NO: 228	DNA sequence of promoter from <i>Aspergillus nidulans</i> , PniaD
SEQ ID NO: 229	DNA sequence of promoter from <i>E. coli</i> , T7
SEQ ID NO: 230	DNA sequence of promoter from <i>Pichia pastoris</i> , PFMT
SEQ ID NO: 231	DNA sequence of promoter from <i>Pichia pastoris</i> , HpMOX
SEQ ID NO: 232	DNA sequence of promoter from <i>Pichia pastoris</i> , CbFLD1
SEQ ID NO: 233	DNA sequence of promoter from <i>Pichia pastoris</i> , CbAOD1

SEQ ID NO: 234	DNA sequence of promoter from <i>Pichia pastoris</i> , PmMOD1
SEQ ID NO: 235	DNA sequence of promoter from <i>Pichia pastoris</i> , PmMOD2

Cloning

[0129] Genes for proteins were found using the BLASTP online tool (<https://blast.ncbi.nlm.nih.gov>) using the sequence with GenBank Accession No. KF549981 (Sugar beet haemoglobin BvHb1.2) as the template. Genes with a sequence identity of >80% that came from eatable plants were selected: Three class 1-haemoglobin from quinoa, *Chenopodium quinoa* (Accession Numbers XP_021738545.1, XP_021738543.1, and XP_021731656.1) and one (1) class 1-haemoglobin from spinach, *Spinacia oleracea* (Accession No. XP_021850852.1). The genes of these selected sequences were custom synthesized by IDT technologies according to the suppliers' instructions (www.idtdna.com). Recombination-based cloning, and transformation of the genes were carried out according to the manufacturer's instructions (Gateway™, Invitrogen). The genes were individually cloned into a pET-DEST42 plasmid, and the final expression vectors transformed into BL21-DE3 *Escherichia coli* cells.

15 Growth conditions for expression

[0130] Cells containing the expression vectors pET-DEST42 with the correspondent haemoglobin genes were grown in sterile TB medium containing 100 mg/ml carbenicillin at 37°C and 150 rpm until OD₆₀₀ ≥ 2. The expression of the recombinant proteins was induced by adding 0.5 mM IPTG, and 0.3 mM δ-aminolevulinic acid. The cells were also briefly bubbled with carbon monoxide (CO) to obtain a stable CO-Haemoglobin compound. After induction, cells were grown overnight at 22°C, 150 rpm. The cells expressing the recombinant haemoglobin were collected by centrifugation and snap frozen with liquid nitrogen and then stored at -80°C until used.

Purification of recombinant haemoglobins and/or phytooglobins

25 [0131] All procedures were performed at room temperature and all buffers were pH adjusted. The purification was made according to Leiva Eriksson *et al.* 2019² The recombinant cells were resuspended in extraction buffer (50 mM Tris-HCl, pH 8.5, 50 mM NaCl, 25 mM glucose, 1 mM EDTA, 10 mM ascorbic acid, and 5% glycerol) at a concentration of 2ml of buffer per gram of cells). The resuspended cells were disrupted by sonication (Q500 Qsonica) and the lysate was clarified twice by centrifugation at 27 000g, first 45 min and then 30 min.

[0132] To purify the recombinant haemoglobin from the lysate, an ÄKTA purification system (Cytiva) was used. The recombinant proteins were first loaded into a Q-Sepharose FF column and eluted with start buffer (50 mM Tris-HCl, pH 8.5) plus 50 mM NaCl. The red fractions were

concentrated using 10 kDa Viva-Spin columns (Vivascience). Ammonium sulphate was then added to the pooled sample to a final concentration of 0.8 M. The sample was then loaded onto a Butyl-Sepharose HP column previously equilibrated with 50 mM Tris-HCl pH 8.0 buffer plus 0.8 M ammonium sulphate. The proteins were eluted with a 10 CV (column volume) linear gradient ranging from 0.8 to 0.0 M ammonium sulphate in 50 mM Tris-HCl pH 8.0 buffer. Finally, the proteins were loaded onto a Q-Sepharose HP column, and a linear gradient of NaCl from 0 to 100 mM in 50 mM Tris buffer at pH 8.5 was used for elution. The purified proteins were analysed by SDS-PAGE (see Figure 1-D) and total haem was determined using the pyridine haemochrome haem assay (see Sinclair et al (2001))³. They were then flash-frozen in liquid nitrogen and stored at -80°C until needed.

Molecular mass of recombinant haemoglobin

[0133] The native molecular masses of the purified recombinant proteins were determined by gel filtration using a prepacked column on a HiLoad 16/60 Superdex 200 pg (GE Healthcare) using 50 mM sodium phosphate pH 7.2 plus 150 mM NaCl as column buffer. The standard proteins for mass determination were processed according to the instructions in the Gel Filtration Calibration Kit (Cytiva) using the native protein standards: Ferritin (444 kDa), Aldolase (158 kDa), Conalbumin (75 kDa), Ovalbumin (43 kDa), Carbonic anhydrase (29 kDa), RNase (13.7 kDa), Aprotinin (6.5 kDa).

[0134] The purified proteins were determined to be homodimer with approximate molecular mass between 35 – 40 KDa.

Example 1. Assessment of intrinsic reactivity — Autooxidation assays

[0135] To study the autooxidation of the recombinant proteins, a stock solution of 400µM of each protein was first reduced with a slight excess of sodium dithionite to obtain the deoxy-ferrous form of the protein. The oxy-ferrous form of the protein was produced by passing the deoxy-ferrous protein through a PD-10 desalting column packed with Sephadex G-25 resin (Cytiva) using the buffer Tris-HCl 0.1 M pH 8.5 as the mobile phase. An aliquot was then diluted to 5µM final concentration with the same buffer. Oxidation of the samples was followed by recording the entire spectra (from 400 to 600 nm) every 15 min for the first two hours and then every 60 min for the next 72 hours. Time courses were obtained from the spectra change at 421 nm. The absorption spectra were then recorded versus time. The reaction was followed in a Cary 60 UV-VIS spectrophotometer (Agilent) equipped with a water thermostat 18-cell holder set to 20°C. The OriginPro 8.6 program (OriginLab) was employed for data analysis and generation of figures (see Figure 2).

[0136] Autooxidation in haemoglobin results in the formation of superoxide and hydrogen

peroxide (H_2O_2). The formed H_2O_2 reacts with the iron of oxy-ferrous haemoglobin, (Fe(II)) and undergoes a Fenton reaction producing the highly reactive hydroxyl radical (Fe(III)) which produces cellular and tissue damage (see Sadrzadeh et al. (1984))⁴). This is a “spontaneous oxidation” that occurs at different rates giving an indication of the intrinsic reactivity of a haemoglobin molecule.

[0137] In addition to producing damaging radicals, autooxidation results in the release of the haem group outside of the protein. Haem is an iron-coordinating porphyrin proposed to be the key molecule contributing to tumorigenesis. Haem is contained predominantly in red and processed meat in the form of haemoglobin and myoglobin. The role of dietary haem in cancer has been highlighted in different types of carcinomas. Indeed, high consumption of red and processed meat has been associated with increased incidence of oesophageal, gastric, breast, endometrial, pancreas and lung tumour. However, the majority of studies focus on the role of dietary haem in the pathogenesis of colorectal cancer (CRC), a leading cause of cancer deaths in Western Countries (Fiorito et al. (2020))⁵.

[0138] Thus, in conclusion, the autooxidation rate of haemoglobin is measured to determine their stability and intrinsic reactivity. A high autooxidation rate makes haemoglobin highly toxic, especially when consumed in high amounts; while a low autooxidation rate means the protein is stable and safe.

20 *Method*

[0139] Autooxidation was measured as described above. The haemoglobin proteins from sugar beet, quinoa and spinach in their oxy-ferrous state ($5\mu\text{M}$) were incubated at 37°C at pH 8.0 for 72h (Figure 2). Time courses (400–600 nm) were fitted to single exponential function to determine their autooxidation rate.

25

Result and Conclusion

[0140] The autooxidation rate (k_{autoox}) for the haemoglobin of sugar beet, quinoa and spinach were determined to be 0.1 h^{-1} at pH 8.0 and 37°C . This result means that these proteins autooxidize at a lower rate than other common haemoglobin (Table 1). Therefore, the haemoglobin from sugar beet, quinoa and spinach are stable and have low intrinsic reactivity. This (stability) is achieved because they are hexacoordinated proteins (the haem-iron group is coordinated by two histidine side chains). On the other side, the iron-proteins found in meat (haemoglobin and myoglobin) are pentacoordinated proteins. This means that their haem-iron is coordinated by a single histidine side chain, leaving the other axial site open and making them highly reactive and easy to detach from the apoprotein. In fact, the International Agency for Research on Cancer (IARC), following an

assessment of over 800 studies performed world-wide, classified processed meat as group 1 “carcinogenic to humans” and fresh red meat as group 2A “probably carcinogenic to humans”. This feature is shared by soy leg haemoglobin which is also a pentacoordinated haemoglobin that is being used in a meat-analogue for colour and flavour purposes (FAQ Impossible Foods ⁶). Soy leghaemoglobin has been shown to have a haem dissociation rate 600 times faster than that for sperm whale myoglobin, under the same conditions and, which in combination with its high auto-oxidation rate, results in a much less stable holoprotein (Hargrove and Olson (1996)⁷). [0141] Both temperature and pH are found to affect and have a direct effect on autooxidation. The autooxidation rates are usually accelerated at more acidic pHs and at high temperatures. At these haemoglobins are potentially to be used for human consumption, autooxidation is tested at or around physiological conditions (pH 7 and 37°C).

Table 1: Autoxidation rates for hexa- and penta-coordinated haemoglobin:

Haemoglobin*	SEQ ID NO:	Autooxidation rate (h ⁻¹)
Sugar beet, spinach and quinoa haemoglobin	1-10	0.1(pH 7, 37°C), this patent
Soy leghaemoglobin	35-39	0.2 (pH 7, 37°C); 3.0 (pH 5, 37°C) ⁸
Bovine myoglobin	48	0,78 (pH 6, 37°C); 0,3 (pH 5.7, 30°C) ⁹
Human adult haemoglobin	51-52	1.8 (pH 7.2, 37°C) ¹⁰
Human foetal haemoglobin	51 and 53	1,6 (pH 7.2, 37°C) ⁸

*See Sequence table or listing for sequences

Example 2: Assessment of Temperature stability

[0142] Proteins are amphoteric molecules with positive and negative charges which implies electrostatic repulsion between groups when they are below or above their isoelectric point. These intra-molecular repulsive interactions between charged groups can be enhanced at higher temperatures causing instability. The protein melting point (T_m) is defined as the temperature at which the concentration of the protein in its folded state equals the concentration in the unfolded state. At temperatures below T_m, the size and scattering intensity of a protein are constant, suggesting a stable tertiary structure. At temperatures equal to its T_m and higher, both the size and scattering intensity increase exponentially, indicating the presence of denatured protein and aggregates. [0143] The thermal stability is usually determined through a melt curve analysis in which the protein is exposed to incremental temperature during a specified time while assessing its integrity. Proteins are classified as mesostable or thermostable if their T_m are, respectively, lower or higher than 70°C.

[0144] Haemoglobin are proteins that exist as monomers (e.g., human and bovine myoglobin and soy leghaemoglobin), dimers (e.g., nonsymbiotic haemoglobins from sugar beet, spinach, and

quinoa), and tetramers (e.g., human adult and foetal haemoglobin). Therefore, the T_m will vary from haemoglobin to haemoglobin. In the case of haemoglobin proteins, the T_m not only indicates the denaturation of the protein but also the loss of the haem-iron, which as explained in *Example 1*, is a molecule recognized as tumorigenic. Therefore, a high T_m will be preferable as the protein will be stable at a wider temperature range and better tolerate sudden temperature changes.

Method

[0145] The thermal stability of the proteins (10-20 µL of sample at a concentration of 10-50 µM in 50mM Tris-HCl pH 8.5, n>3) were analysed through nano-differential scanning fluorimetry (nanoDSF) using a Prometheus NT.48 instrument containing aggregation optics (NanoTemper Technologies). The LED intensity for excitation at 280 nm was set to 20%. Sample fluorescence intensities at 350 nm (F350) and 330 nm (F330) were recorded whilst running a temperature ramp from 20 to 95°C at 1°C/min increments (Figure 3). The melting point (T_m) for the proteins were acquired by a first-order derivative of the F350/F330 fluorescence ratio plotted against the temperature. The standard deviations were calculated using Microsoft Excel.

Results and conclusions

[0146] The haemoglobin from sugar beet, spinach and quinoa have, together, an average melting T° (T_m) above 70 °C about 72 °C, which makes them thermostable proteins (Table 2). In contrast, beef myoglobin has a melting temperature of 67,4 °C. This is expected as beef myoglobin is pentacoordinated, and therefore is less intrinsically stable molecule. See Figure 2 and Table 2.

Table 2. Melting T° (T_m) for three hexacoordinated haemoglobin from the amaranthacea family (spinach, quinoa, and sugar beet) compared to the T_m of corn and bovine haemoglobin.

Haemoglobin	T _m (degrees Celsius)
Spinach (So) SEQ ID NO: 5	71.7
Quinoa (Chq) SEQ ID NO: 1	74,1
Sugar beet (Bv) SEQ ID NO: 7	72,4
Corn (Zm) SEQ ID NO: 28	66,3
Beef myoglobin (BtMg) SEQ ID NO: 48	67,4

[0147] This example demonstrates that recombinant hexacoordinated haem- or phyto globin proteins of the invention, spinach, quinoa, and sugar beet display similar stability, in keeping with their hexacoordinated structure.

[0148] The stability the recombinant hexacoordinated haem- or phyto globin protein is found to

be attributed to their hexa-coordination. In this form, the haem-iron is stabilized within the apoprotein keeping the structure of the holoprotein intact and less vulnerable to changes in temperature. Previously it was demonstrated that haem-free globin is unstable and precipitates rapidly at room temperature. When cyanide is added to bind to the haem group of a pentacoordinated haemoglobin, a remarkable reduction in the precipitation of haemoglobin is observed even after 20 hr of exposure to 50 °C. The ability of ligands to stabilize haemoglobin - by locking the sixth bond of the haem-iron - during severe, prolonged heat stress was demonstrated by the effect of carbon monoxide and cyanide on haemoglobin solutions heated to 65°C for 2.5 hr. (Rossi-Fanelli et al. 1958¹¹). In the case of hexacoordinated haemoglobin, the sixth bond of the haem-iron is already “locked” as it is forming a covalent bond to the apoprotein, hence these proteins are more stable while keeping the haem-group safely inside the protein.

[0149] Further, the pentacoordinated myoglobin of beef displays the expected lower stability.

[0150] However, the hexacoordinated phytoalbumin of corn (SEQ ID NO: 28 and 29) has a lower stability, on par with that of myoglobin. This demonstrates that stability does vary also within hexacoordinated haem- or phytoalbumin protein, and that the three phytoalbumins derived from Cayrophyllidae have similar, improved stability.

[0151] In contrast, the pentacoordinated soy leghaemoglobin (SEQ ID NO: 35) denatured at 64°C as shown by UV-visual spectroscopy, releasing the haem-iron. The same phenomenon was observed in another pentacoordinated haemoglobin from meat, bovine haemoglobin (SEQ ID NO: 48), which had a significantly higher storage stability at 4 °C than at 23 °C; however, after 40 days the haemoglobin is denatured irrespective of storage temperature. Also, when exposed to cycles between 25°C and 65°C structural changes that end in denaturation are quickly observed (Bhomia et al. (2016)¹²). The thermal instability of these two pentacoordinated haem- or phytoalbumin protein is behind the release of haem-iron which in turn plays a role in catalysing the production of the flavours and aromas characteristic of meat and analogue meat¹³. However, as above mentioned, pentacoordinated haemoglobins are more toxic due to their instability and reactive nature; therefore, they aren’t healthy if consumed in higher amounts.

Example 3: Assessment of stability – DNA Cleavage

[0152] Free haemoglobin is a biologically hazardous molecule because it acts as a “Fenton” reagent. As such, it catalyses the generation of hydroxyl-radicals that act directly on other

biomolecules¹⁴, cells¹⁵, and tissues.¹⁶ The rate of radical formation depends on the intrinsic reactivity of the haemoglobin molecule. One way to measure this reactivity is by exposing supercoiled DNA (SC) to free haemoglobin and then follow its degradation to open circular DNA (OC) and linear DNA (L). A highly reactive and unstable haemoglobin will degrade DNA faster than
5 a stable one.

Method

[0153] The DNA cleavage experiments were conducted in PCR-tubes with a final volume of 20 μ l. The cleavage reaction consisted of buffer sodium phosphate 20 mM pH 7.2, supercoiled plasmid
10 DNA (pUC18 at 1.25 ng/ μ l) and 25 μ M haemoglobin in its oxidized form (Fe(III)). The temperature used for the reaction was 37°C.

[0154] The oxidized haemoglobin was obtained by the addition of potassium ferrocyanide. The excess of potassium ferrocyanide was removed by running the sample through a PD-10 desalting column packed with Sephadex G-25 resin (Cytiva).

15 [0155] Results of the DNA cleavage were analysed using agarose electrophoresis. Samples were removed after 4h of reaction then mixed with DNA loading dye 6x (Thermo Fisher Scientific) and loaded onto a 1 % agarose gel containing gel red dye (Biotium). Electrophoresis was run at 100 V for 90 min in TAE buffer. The gels were photographed using a GelDoc XR system (BioRad) and the relative quantities of entire plasmid *a.k.a.* supercoiled (SC) versus partially or totally degraded
20 plasmid DNA *a.k.a.* open circular (OC) and linear (L) was estimated using the Quantity One software (BioRad). The comparison of remaining SC DNA after incubation with the hexacoordinated haemoglobin was compared to the amount remaining after incubation with recombinant adult haemoglobin (pentacoordinated), using unpaired, unequal variance t-test (Welch t-test). A result of $p < 0.05$ were considered significant. As a control, incubation of SC DNA
25 without haemoglobin was used.

Results and conclusions

[0156] The cleavage of DNA was determined by quantifying the percentage of SC versus OC/L plasmid DNA (pDNA) remaining after 4h incubation at 37°C. A highly reactive haemoglobin will
30 produce high levels of radicals that will damage the DNA and result in low amount of remaining SC pDNA. The cleavage capacity of the three hexacoordinated haemoglobin was compared to that of recombinant pentacoordinated haemoglobin (Adult haemoglobin, rHbA) (Table 3 and figure 4). The result show that significantly less SC DNA remains after incubation with recombinant adult human haemoglobin (pentacoordinated) than with the three other hexacoordinated
35 haemoglobin, $p < 0.05$ (Figure 4). No significant difference among the three hexacoordinated

haemoglobin was observed ($p > 0.05$). The DNA cleavage experiment showed that hexacoordinated haemoglobin from sugar beet, spinach and quinoa (all members of the Caryophyllales order, Amaranthaceae family) degrade significantly less DNA than the recombinant human adult haemoglobin, which is a pentacoordinated protein.

5 [0157] These results provide new insights into the positive effect of hexacoordination on the stability of haemoglobin and point towards a lower reactivity of hexacoordinated plant haemoglobin compared to pentacoordinated haemoglobin.

Table 3. Percentage SC pDNA after incubation with oxidized Hb for 4 h under standard conditions (20 mM phosphate buffer pH 7.2, 37 °C, 1.25 ng/μl pDNA, 25 μM Hb).

Control	Sugar beet SEQ ID NO: 7	Quinoa SEQ ID NO: 1	Spinach SEQ ID NO: 5	rHbA SEQ ID NO: 51-52
99 ± 7	60 ± 26	55 ± 13	59 ± 14	22 ± 7

Example 4: Bioavailability of hexacoordinated phytoglobin in plants from the Caryophyllales order (Amaranthaceae family)

[0158] Iron deficiency and anaemia (IDA) is a condition that affects 2 billion people around the world. Currently, IDA is treated with known iron supplements; however, the iron used in these known products is inorganic *a.k.a.* elemental iron. Elemental iron is poorly absorbed and most of it passes through the digestive system, producing free radicals that damage cells, tissues, and organs, resulting in side effects —digestive discomfort, stomach pain, nausea, constipation to name a few— and leaves a metallic aftertaste in patients’ mouth. As a result, fewer than four out of ten patients finish their treatment, the rest remain iron deficient or anaemic. It is worth mentioning that IDA is a recurring condition that returns easily after been resolved; therefore, there is a constant need for iron by those affected with this condition.

25 [0159] For humans, the main and best source of iron is found in foods of animal origin, especially red meat. There, iron is present in its organic form, haem-iron, which is bound to a protein (haemoglobin and myoglobin). However, since haem-iron is loosely bound to the apoprotein —as they are pentacoordinated— it is easily released and initiates toxic tumorigenic reactions as described in *Example 1*.

30 [0160] Plant haemoglobin such as those of the order Caryophyllales (family Amaranthaceae), on the other hand, keep haem-iron within the apoprotein thanks to their strong hexacoordination, making them less reactive and more stable.

[0161] However, stability is not enough to become a good source of iron. The iron contained in plant haemoglobin must also be bioavailable, that is, easy to be absorbed by the body; therefore, the bioavailability of iron in these plant proteins was tested.

[0162] Next, the method used to test the bioavailability of haem-iron from plant haemoglobin in
5 Caco-2 cells is described. These cells are the standard *in vitro* model for iron absorption, and the method mimics the entire digestion process —starting in the mouth— down to the duodenum and proximal jejunum, which are the regions of iron absorption. This method is robust and produces results that consistently agree with human studies.

10 *Method*

[0163] The bioavailability of iron in the haemoglobin from sugar beet, spinach and quinoa was determined using the Caco-2 cell line in conjunction with *in vitro* digestion techniques so the proteins will undergo simulated peptic digestion followed by intestinal digestion in the presence of Caco-2 cell monolayers. This model measures iron solubility in addition to providing a measure
15 of uptake via a living component, the Caco-2 cell monolayer (Glahn et al. (1998)¹⁷). This approach utilizes Caco-2 cell ferritin formation as a marker of Fe uptake, avoiding the cost and methodological issues associated with isotopic labelling. Moreover, the ferritin formation approach is a sensitive, cost-effective marker of Fe uptake that dramatically increased the throughput of the system.

[0164] The procedure consists of four parts and has been previously published in detail (Glahn et al. (2017)¹⁸). What follows is a brief summary. First, the content of iron in the samples is determined. For this a portion of the sample is freeze-dried and treated with acid (HNO₃ and HClO₄) and heat to destroy organic matter. Iron is determined using inductively coupled plasma atomic emission spectrometer (ICP-AES) analysis. The second part is *in vitro* digestion. This starts
25 with the gastric phase where the samples are exposed to acidic pH 2.0 and peptide digestion followed by raise of the pH to 5.5–6.0 and addition of pancreatin–bile extract solution. At this point, the mixture is referred to as a “digest”. The third part is the addition of the digest to the monolayer of Caco-2 cells followed by incubation at 37°C for 22h. After that, the cells were harvested by aspiration and sonicated. Cell protein was measured using a semimicro adaptation
30 of the Bio-Rad DC protein assay kit (Bio-Rad Laboratories). Finally, the fourth part is the determination of ferritin in Caco-2 by an enzyme-linked immunosorbent assay (ELISA) using commercially available kits, the same kits used for human ferritin measurements in clinical practice.

[0165] Data were analysed using analysis of variance and Tukey’s post test to determine
35 significant differences ($p < 0.05$) between samples. Values are expressed as mean standard error

of the mean (SEM); n = 3 independent replications.

Results and conclusions

[0166] The results were expressed in the amount of ferritin that was formed by the Caco-2 cells per mg of total protein and the amount of ferritin per unit of iron (μg). Regarding the amount of ferritin/ mg protein, differences were found between the different samples, with spinach (SO) having the lowest ferritin formed in proportion to total protein and sugar beet (Bv) having the highest (Table 4). This value (ferritin/protein) is calculated based on the amount of protein in the cells in the assay, not on the protein applied to the cells. This number might therefore be subject to variation between samples, as the amount of ferritin formed depends on the amount of iron that is in the sample and is absorbed. Therefore, if the cells have received different amounts of iron, they will have different ferritin levels per protein (amount of cells). To have a direct assessment of how much iron has been absorbed it's better to use μg ferritin per unit (μg) of iron. If the iron is not bioavailable (not absorbed), this value will be low. When it comes to the amount of ferritin formed per μg of iron, there was not significant differences among the different samples. This indicates that the iron contained in the plant haemoglobin is absorbed by the mammalian cells in the same rate as the iron from the myoglobin from beef (Figure 5).

Table 4. Bioavailability of iron tested in Caco-2 cells.

Haemglobin	ng ferritin/mg protein	ng ferritin/ μg Fe
Spinach (So) SEQ ID NO 5	16,08	1,57
Quinoa (Chq) SEQ ID NO 1	14,92	1,75
Corn (Zm) SEQ ID NO 28	29,84	1,49
Beef myoglobin (BtMg) SEQ ID NO 48	44,8	1,78
Sugar beet (Bv) SEQ ID NO 7	49,3	1,45
So + Canned baby food (CBF)	1,31	0,128431
Chq + CBF	1,85	0,217136
Zm + CBF	2,01	0,10055
BtMg + CBF	4,1	0,163217
Bv + CBF	1,67	0,049306
So + Plant based drink (PBD)	14,09	1,381373
Chq + PBD	11,93	1,400235
Zm + PBD	23,52	1,176588
BtMg + PBD	53,51	2,130175
Bv + PBD	32,04	0,94597

[0167] Iron is a nutrient whose absorption is affected by the food components. Those components can enhance or inhibit the bioavailability of iron. In Figure 5B, the results of the effect of two common foods on the absorption of iron are shown. The foods were a plant-based drink and a sweet potato baby food.

[0168] The results show that the hexacoordinated phytooglobins from spinach, quinoa and sugar beet all display similar bioavailability to that of the myoglobin in beef which the most common source of iron in red meat, and therefore that they are suitable as iron supplement.

[0169] Corn also shows a bioavailability on par with myoglobin, however, as shown in temperature stability example above, the phytoglobin from corn is less suitable as a supplement. The bioavailability of the iron in plant haemoglobin is comparable to the bioavailability of the myoglobin in beef, the most common source of iron in red meat. Taken together, the experiments show that the phytooglobins according to the invention are both stable and bioavailable.

10 **Example 5: Taste**

[0170] The phytooglobins from sugar beet, spinach and quinoa have been shown to be a source of bioavailable iron in *Example 4*; therefore, they can be used to prevent and treat iron deficiency and anaemia (IDA) if used as organic, biobased iron supplements but also as food ingredients, if they do not affect the taste of the food they are added to.

15 [0171] To evaluate if these proteins had any taste or modify the taste of other foods, a sensorial test was set up. The phytoglobin proteins were tested in three different food matrices: blueberry juice, rosehip juice, plum juice, and as a control water was used.

Method

20 [0172] Two sets of non-transparent cups with 200 ml of juice were presented to the members of the panel. To one set, 1 ml of 0.3 mM phytoglobin was added while to the other set 1 ml of water was added. This was to avoid the panellists to notice the difference between the cups with added and non-added phytoglobin due to the dilution of the sweetness in the juice. The same set up was prepared with the control (water), one cup had only water while the other had 1ml of phytoglobin added. In total, four pair of cups were presented to the panellists, and they had to rate the taste of each pair. The goal was to determine if the panellist could detect any change in taste due to added phytoglobin.

[0173] The panel members were asked to rate the flavour of the drinks from 1 to 5, according to the following description: 1= very bad taste, 3 = neutral taste, and 5 = very good taste. The results were compared in pairs, as the goal was to determine if the addition of the phytoglobin could affect the perception of the taste when compared to the same juice that had added water. The delta (Δ value) was determined by calculating the difference between the rates for each pair of beverages (with added phytoglobin or only water).

Results and conclusion

[0174] In all cases and for all the drinks, including the control water, the panel members gave the same rate to the pair of cups presented to them. This means that the Δ Value for all beverages was zero (0). These results show convincingly that taste of the juices was not changed due to the added phytoalbumin as the panel members could not detect any change on the taste of the juices. There were comments that indicated that they noticed that something was added to both cups presented to them, as they recalled having tasted the juice before but that there were sweeter or had more intense taste. They said the taste of the juice felt diluted. Apart from this detail, they could not find a difference in the taste even after they were told which cup had the phytoalbumin added.

Table 5. Taste analysis of plant phytoalbumins when added to fruit juices. W = with phytoalbumin, W/O = without phytoalbumin

Water	Sugar beet phytoalbumin SEQ ID NO: 7			Spinach phytoalbumin SEQ ID NO: 5			Quinoa phytoalbumin SEQ ID NO: 1		
Subject	W	W/O	DELTA	W	W/O	DELTA	W	W/O	DELTA
A	3	4	1	3	2	1	3	2	1
B	3	3	0	2	2	0	2	2	0
C	2	2	0	2	2	0	2	2	0
D	4	3	1	4	4	0	4	4	0
E	1	1	0	0	0	0	1	0	1
Blueberry	Sugar beet phytoalbumin			Spinach phytoalbumin			Quinoa phytoalbumin		
Subject	W	W/O	DELTA	W	W/O	DELTA	W	W/O	DELTA
A	4	4	0	5	5	0	4	4	0
B	4	4	0	4	4	0	4	4	0
C	5	5	0	5	5	0	4	4	0
D	4	4	0	5	5	0	4	4	0
E	2	2	0	3	3	0	2	2	0
Rosehip	Sugar beet phytoalbumin			Spinach phytoalbumin			Quinoa phytoalbumin		
SUBJECT	W	W/O	DELTA	W	W/O	DELTA	W	W/O	DELTA
A	3	3	0	3	3	0	3	3	0
B	4	4	0	4	4	0	4	4	0
C	2	2	0	3	3	0	2	2	0
D	5	5	0	4	4	0	5	5	0
E	3	3	0	4	4	0	4	4	0
Plum	Sugar beet phytoalbumin			Spinach phytoalbumin			Quinoa phytoalbumin		
SUBJECT	W	W/O	DELTA	W	W/O	DELTA	W	W/O	DELTA
A	3	3	0	3	3	0	3	3	0
B	4	4	0	4	4	0	4	4	0
C	2	2	0	2	2	0	3	3	0
D	5	5	0	5	5	0	5	5	0
E	3	3	0	3	3	0	3	3	0

[0175] None of the panellists could distinguish a difference in taste between the cup of juice that had phytoalbumin or water added to. Therefore, we concluded that the phytoalbumin from sugar

beet, spinach, and quinoa do not affect the taste of the food they are added to.

[0176] This means that these proteins can be used as food ingredients to improve their nutritional quality. They can also be used in a most concentrated form without a cover to mask a metallic flavour as it is the case with current iron supplements that leave a metallic taste after their intake.

5 Iron supplements based on these proteins can be given in a liquid form (e.g. drops or syrup) or as a tablet, which means their absorption will not be limited by an external cover as it happens with capsules.

[0177] This lack of taste constitutes a big difference when compared to pentacoordinated haemoglobin that are in meat analogues and animal food, e.g., soy leghaemoglobin and beef myoglobin and haemoglobin which catalyse the production of flavours and aromas characteristic of meat as a result of the loss of their haem-iron.

[0178] These proteins according to the present invention have no taste since the iron is only released from the protein by the iron-transport-system of the digestive system in the region where iron is absorbed, namely the duodenum and upper jejunum. The proteins of the present invention
15 therefore have very low oxidation rate.

Example 6: Alignments Variable loop (CD region)

Table 6. Sequence alignments Percent Identity Matric by Clustal2.1

Percent Identity Matrix - created by Clustal2.1												
1: BtMg	100.00	11.11	15.79	20.59	23.68	21.62	18.42	18.42	21.05	21.05	21.05	21.05
2: SoyLb	11.11	100.00	54.29	43.33	50.00	45.45	44.12	44.12	47.06	44.12	44.12	44.12
3: BvHb2	15.79	54.29	100.00	60.61	61.11	61.11	59.46	59.46	67.57	64.86	64.86	64.86
4: ZmHb1	20.59	43.33	60.61	100.00	75.00	63.64	69.70	69.70	78.79	75.76	75.76	75.76
5: BvHb1.1	23.68	50.00	61.11	75.00	100.00	68.57	75.00	75.00	83.33	80.56	80.56	80.56
6: AtHb1	21.62	45.45	61.11	63.64	68.57	100.00	78.38	78.38	81.08	78.38	78.38	78.38
7: BvHb1.2	18.42	44.12	59.46	69.70	75.00	78.38	100.00	100.00	86.84	89.47	89.47	89.47
8: C86A	18.42	44.12	59.46	69.70	75.00	78.38	100.00	100.00	86.84	89.47	89.47	89.47
9: SoHb	21.05	47.06	67.57	78.79	83.33	81.08	86.84	86.84	100.00	97.37	97.37	97.37
10: Chq_38543a	21.05	44.12	64.86	75.76	80.56	78.38	89.47	89.47	97.37	100.00	100.00	100.00
11: Chq31656	21.05	44.12	64.86	75.76	80.56	78.38	89.47	89.47	97.37	100.00	100.00	100.00
12: Chq_38543b	21.05	44.12	64.86	75.76	80.56	78.38	89.47	89.47	97.37	100.00	100.00	100.00

20 [0179] Sequence alignments were made between the sequences of the variable loop (CD region). The region is highly conserved, and differences in the region are thought to reflect in the variations in molecule stability.

25	BtMg	LIRLFTGHPETLEKFDKFKHLKTEAEMKASEDLKKHGNTV	40 (SEQ IN NO: 122)
	SoyLb	YTSILEKAPAAKDLFSFLANGVD----PTNPKLTGHAEL	36 (SEQ IN NO: 120)
	BvHb2	FSIILEIAPAAKNMFSFLRDSE--EVPQNNPKLKAHAIV	38 (SEQ IN NO: 117)
	ZmHb1	FLRIFEIAPGAKQMFSFLRDAG-DAPLEKHPKLKA-----	34 (SEQ IN NO: 121)
	BvHb1.1	--RVFEIAPTAKKLFSLRDSSDDVPLEKNPKLKAHALTV	38 (SEQ IN NO: 115)
30	AtHb1	FIKIFEIAPTTKKMFSFLRDS--PIPAEQNPCLKPHAMS-	37 (SEQ IN NO: 119)
	BvHb1.2	FLKIFEIAPTAKKMFSFVRDS--DVPLEQNQKLKGHAMSV	38 (SEQ IN NO: 116)
	C86A	FLKIFEIAPTAKKMFSFVRDS--DVPLEQNQKLKGHAMSV	38 (SEQ IN NO: 118)
	SoHb	FLRIFEIAPTAKKMFSFLRDS--NVPLEQNPKLKAHAMSV	38 (SEQ IN NO: 123)
	Chq_38543a	FLRIFEIAPTAKKMFSFVRDS--NVPLEQNPKLKAHAMSV	38 (SEQ IN NO: 112-114)
	Chq31656	FLRIFEIAPTAKKMFSFVRDS--NVPLEQNPKLKAHAMSV	38 (SEQ IN NO: 112-114)

Chq_38543b FLRIFEIAPTAKKMFSEVRDS--NVPLEQNPKLKAHAMSV 38 (SEQ IN NO: 112-114)
 :: * : . * *

Example 7: Further alignments of CD-loops, a key structural domain

5 [0180] Without being bound by theory it was hypothesized that the insertion of a haem group and its stability inside globin proteins depend on the degree of flexibility and moiety offered by the CD-loop and its key amino acids that lock and stabilize the haem inside their pocket. The CD-loop governs the opening and closing of helix-E when the haem is inserted during proteins synthesis, i.e. the formation of the holoprotein by inserting the haem group inside the apoprotein (a globin
 10 protein). The CD-loop also governs the rates of haem loss and autooxidation through the lifetime of the holoprotein. Furthermore, the CD-loop influences the binding of external ligands (for example, oxygen, carbon monoxide, carbon dioxide, nitric oxide, sulphate, nitrite, nitrate, cyanide, among others) to the haem containing pocket, determining its final function, e.g. oxygen transporter, nitrite oxidase, nitrite reductase, peroxidase, sulphide reductant, carbon monoxide
 15 carrier, and other gas carriers. Given the relevance of the CD-Loop, an extended sample of known and hypothetical CD-loops were compared to each other.

[0181] Haem- or phytoglobin proteins are divided in two three main groups, pentacoordinated and hexacoordinated, and truncated haemoglobins. Usually, pentacoordinated proteins will be responsible for oxygen transport and oxygen scavenging inside the cells and tissues where they
 20 are expressed. They are also quite recent in terms of evolution. On the other hand, hexacoordinated proteins will not necessarily transport oxygen but sense it. These proteins are mostly associated with the binding of nitric oxide, and they are more ancient in terms of evolution.

[0182] Pentacoordinated haem- and phytoglobin proteins will be found in aerobic animals (e.g. mammals) and in leguminous plants where they are collectively called Leg-haemoglobins or
 25 symbiotic haemoglobins.

[0183] Hexacoordinated proteins, also known as nonsymbiotic haemoglobins, are found in every single living organism, from archaea to humans, and they will be divided into two groups: class-1 and class-2. This differentiation is based on their protein sequence and their differential oxygen affinity. Class-1 proteins are believed to have a higher affinity for oxygen than Class-2². Also, Class-
 30 2 proteins are phylogenetically closer to leg-haemoglobins and oxygen transporting-haemoglobins compared to class-1 proteins. To our knowledge, no structural domain or sequence has been identified that could differentiate one class from the other. Here it is hypothesized that the CD-loop could be a key domain differentiating class-1 and class-2 haem- and phytoglobin proteins and having a great contribution to the formation of hexa- or pentacoordinated haem- or phytoglobin
 35 proteins.

Table 7. Sequence alignments Percent Identity Matric by Clustal2.1

Percent Identity Matrix - created by Clustal2.1

1: BtMg	100.00	11.11	15.79	20.59	23.68	21.62	18.42	18.42	21.05	21.05	21.05	21.05
2: SoyLb	11.11	100.00	54.29	43.33	50.00	45.45	44.12	44.12	47.06	44.12	44.12	44.12
3: BvHb2	15.79	54.29	100.00	60.61	61.11	61.11	59.46	59.46	67.57	64.86	64.86	64.86
4: ZmHb1	20.59	43.33	60.61	100.00	75.00	63.64	69.70	69.70	78.79	75.76	75.76	75.76
5: BvHb1.1	23.68	50.00	61.11	75.00	100.00	68.57	75.00	75.00	83.33	80.56	80.56	80.56
6: AtHb1	21.62	45.45	61.11	63.64	68.57	100.00	78.38	78.38	81.08	78.38	78.38	78.38
7: BvHb1.2	18.42	44.12	59.46	69.70	75.00	78.38	100.00	100.00	86.84	89.47	89.47	89.47
8: C86A	18.42	44.12	59.46	69.70	75.00	78.38	100.00	100.00	86.84	89.47	89.47	89.47
9: SoHb	21.05	47.06	67.57	78.79	83.33	81.08	86.84	86.84	100.00	97.37	97.37	97.37
10: Chq_38543a	21.05	44.12	64.86	75.76	80.56	78.38	89.47	89.47	97.37	100.00	100.00	100.00
11: Chq31656	21.05	44.12	64.86	75.76	80.56	78.38	89.47	89.47	97.37	100.00	100.00	100.00
12: Chq_38543b	21.05	44.12	64.86	75.76	80.56	78.38	89.47	89.47	97.37	100.00	100.00	100.00

Method

[0184] The sequences of known and hypothetical haem-bound proteins were retrieved from public databases. The sequences came from a wide range of species spanning from animals to plants to bacteria. The sequences were analysed to identify the location of the CD-loops according to the general structure of haemoglobin proteins. Once identified, they were retrieved and compared to each other using the public online software Clustal Omega which is a multiple sequence alignment program that uses seeded guide trees and HMM profile-profile techniques to generate alignments between sequences.

Result and Discussion

[0185] The alignment showed that the CD-loop contains at least two conserved amino acids, a P, and a F in positions 01 and 08 as well as a semiconserved amino acid in position 03, with respect to the reference CD-Loop. In addition to those, we have identified at least four key positions in the region important for the expression and stability of a globin protein that contains a haem group. These positions are positions 08, 13, 16, and 21 (positions 08, 13, 18, and 23 in some species with longer loops). (Figure 7).

[0186] The percent identity matrix revealed a great variability and in some cases the similarity between the CD-loop sequences was as low as 11% indicating that, except for key positions (above mentioned), this region might be a key differentiating structure between the different types and classes of haem- or phyto globin proteins. This was confirmed by doing a phylogenetic tree only using the CD-loop (Figure 8A). This tree shows, how by only using the CD-loop was still possible to group the different types or haem- or phyto globin proteins according to their classification. In order to test this further, the next step was to substitute the CD-loop of haem- or phyto globin proteins with that of the CD-loop of the parent sequence (SEQ ID NO: 56) to see if the entire proteins carrying the CD-loop of SEQ ID NO: 56 could still be differentiated into their different types and classes. The results in Figure 8B demonstrate that the differentiation between the

different classes of haem- and iron phytoglobins was not as clear-cut as in Figure 8A. Therefore, it is concluded that the CD-loop could be a key domain differentiating pentacoordinated from hexacoordinated haemoglobins. Therefore, its manipulation could reversibly render the haem- or phyto- globin proteins, partially or totally, from class-1 to class-2 and from penta- to hexacoordinated proteins (Figure 8B).

Table 8. CD-loop amino acids (SEQ ID NO: 56) quinoa phyto- globin and their position in the parent quinoa phyto- globin sequence (SEQ ID NO 1).

Position	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78
SEQ_ID NO: 56	P	T	A	K	K	M	F	S	F	V	R	D	S	N	V	P	L	E	Q	N	P	K	L	K	A	H	A

[0187] Sequence alignments were made between the sequences of the variable loop (CD-loop) of globin proteins containing a haem group.

BtMg	LIRLFTGHPETLEKFDKFKHLKTEAEMKASEDLKKHGNTV	40 (SEQ IN NO: 122)
SoyLb	YTSILEKAPAAKDLFSFLANGVD----PTNPKLTGHAEL	36 (SEQ IN NO: 120)
BvHb2	FSIILEIAPAAKNMFSFLRDSE--EVPQNNPKLKAHAIV	38 (SEQ IN NO: 117)
ZmHb1	FLRIFEIAPGAKQMFSFLRDAG-DAPLEKHPKLKA----	34 (SEQ IN NO: 121)
BvHb1.1	--RVFEIAPTAKKLFSLRDSSDDVPLEKNPKLKAHALTV	38 (SEQ IN NO: 115)
AtHb1	FIKIFEIAPTTHKKMFSFLRDS--PIPAEQNPKLKPHAMS-	37 (SEQ IN NO: 119)
BvHb1.2	FLKIFEIAPTAKKMFSLRDS--DVPLEQNQKLKGHAMSV	38 (SEQ IN NO: 116)
C86A	FLKIFEIAPTAKKMFSLRDS--DVPLEQNQKLKGHAMSV	38 (SEQ IN NO: 118)
SoHb	FLRIFEIAPTAKKMFSLRDS--NVPLEQNPKLKAHAMSV	38 (SEQ IN NO: 123)
Chq_38543a	FLRIFEIAPTAKKMFSLRDS--NVPLEQNPKLKAHAMSV	38 (SEQ IN NO: 112-114)
Chq31656	FLRIFEIAPTAKKMFSLRDS--NVPLEQNPKLKAHAMSV	38 (SEQ IN NO: 112-114)
Chq_38543b	FLRIFEIAPTAKKMFSLRDS--NVPLEQNPKLKAHAMSV	38 (SEQ IN NO: 112-114)
	::*:..*...*.*	
SEQ_110	PEVRPLFDMGRQESLEQP-----KALA-----	22
SEQ_103	PTTKTYFPHFDLS-----HGSAQVKGHG--	23
SEQ_106	PTTKTYFPHFDLS-----HGSAQVKGHG--	23
SEQ_105	PETLEKFDKFKHLKTEAEMKASEDLKKHG--	29
SEQ_109	PETLEKFDKFKHLKSEDEMASEDLKKHG--	29
SEQ_104	PWTQRFFESFGDLSTADAVMNNPKVKAHG--	29
SEQ_107	PWTQRFFESFGDLSTPDVAMGNPKVKAHG--	29
SEQ_108	PWTQRFFDSFGNLSSASAIMGNPKVKAHG--	29
SEQ_63	PTAKKLFSLRDSDDVPLEKNPKLKAHA--	29
SEQ_76	PTTKKMFSFLRDS--PIPAEQNPKLKPHA--	27
SEQ_72	PSAKKMFSFLKDS--NVPLDQNPKLKIHA--	27
SEQ_73	PSAKKMFSFLKDS--NVPLDQNPKLKVHA--	27
SEQ_66	PTAKKLFSLRES--DVPLEQNSKLKAHA--	27
SEQ_62	PTAKKMFSFLRDS--DVPLEQNQKLKGHA--	27
SEQ_65	PTAKKMFSFLRDS--DVPLEQNQKLKGHA--	27
SEQ_56	PTAKKMFSFLRDS--NVPLEQNPKLKAHA--	27
SEQ_57	PTAKKMFSFLRDS--NVPLEQNPKLKAHA--	27
SEQ_58	PTAKKMFSFLRDS--NVPLEQNPKLKAHA--	27
SEQ_60	PTAKKMFSFLRDS--NVPLEQNPKLKAHA--	27
SEQ_83	PGAKQMFSFLRDA--GDAPLEKHPKLKAHA--	28
SEQ_74	PSAKKLFSLRDDSSAAPLHQNAKLKTHA--	29
SEQ_89	PSARQLFSFLRDS--DVPLDKNPKLKAHA--	27
SEQ_87	PSARQLFSFLRNS--DVPLEKNPKLKIHA--	27
SEQ_85	PSASQMFSFLRNS--DVPLEKNPKLKTHA--	27

	SEQ_86	PSASQMFSFLRNS--DVPLEKNPKLKTHA--	27
	SEQ_84	PSAKQMFSFLRDS--DVPLEKNPKLKTHA--	27
	SEQ_88	PSARQMFPFLRDS--DVPLETNPKLKTHA--	27
	SEQ_80	PSAKQMFPFLRNS--DVPLETNPKLKTHA--	27
5	SEQ_81	PSAKQMFPFLRNS--DVPLETNPKLKTHA--	27
	SEQ_82	PSAKQMFPFLRDS--DVPLETNPKLKTHA--	27
	SEQ_68	PSAQKLFSFLKDS--KVPLEKNPKLKSHA--	27
	SEQ_69	PSAQKLFSFLKDS--KVPLEKNPKLKSHA--	27
	SEQ_79	PSAQKLFPFLRDS--SVPVEQNPKLKPHA--	27
10	SEQ_78	PSAQKLFSFLRDS--DIPLEKNPKLKPHA--	27
	SEQ_90	PSAQKLFSFLRDS--TVPLEQNPKLKPHA--	27
	SEQ_70	PSAQKLFSFLRDS--NVPLEQNQKLKPHA--	27
	SEQ_71	PSAQKLFSFLRDS--NVPLEQNQKLKPHA--	27
	SEQ_75	PSAKKLFSFLRDS--NVPLEQNTKLKPHA--	27
15	SEQ_95	PAAKGMFSFL-DS--AGVQ-DSPKLQSHA--	25
	SEQ_96	PTAKGMFSFLKDS--AGVV-DSPKLQAHA--	26
	SEQ_98	PAAKGMFSFLKDS--AEVV-DSPKLQAHA--	26
	SEQ_99	PAAKGMFSFLKDS--AEVV-DSPKLQAHA--	26
	SEQ_97	PAAKGLFSFLKDT--AGVE-DSPKLQAHA--	26
20	SEQ_100	PAAKNLFSFLA----NGVDPTNPKLTAHA--	25
	SEQ_92	PAAKDLFSFLS----NGVDPSNPKLTGHA--	25
	SEQ_91	PAAKDLFSFLA----NGVDPTNPKLTGHA--	25
	SEQ_93	PVAKDLFSFLA----NGVDPTNPKLTGHA--	25
	SEQ_94	PAAKDLFSFLKGT--SEVPQNNPELQAHA--	27
25	SEQ_77	PAAKGLFSFLRDS--DEVPHNNPKLKAHA--	27
	SEQ_59	PAAKAMFSFLRDS--EEIPQNNPKLKAHA--	27
	SEQ_64	PAAKNMFSFLRDS--EEVPQNNPKLKAHA--	27
	SEQ_61	PAAKNMFSFLRDS--DEIPHNNPKLKAHA--	27
	SEQ_67	PAAKDMFSFLRDS--DEIPQNNPKLKAHA--	27
30	SEQ_101	PEAKAMFSFLKDS--DGVPKDNLDEAH---	26
	SEQ_102	PEVKDMFFFLKDC--DGVPHNNLKLEAQAEK	29
		* . *	.

Example 8. CD-loop modifications and effect on expression of haem-bound proteins.

- 35 [0188] Without being bound to theory, it was hypothesized that the CD-loop is relevant for the expression and synthesis of haem-bound holoproteins (haemoglobin and phytooglobins formed by a globin apoprotein bound to a haem-iron group) when produced outside their natural source.

Method

- 40 [0189] To test this hypothesis, the CD-loop was modified by inserting mutations i.e. amino acid substitution, in different positions including those found to be consistent during the alignment. The modified CD-loops were added into the reference sequence (SEQ ID NO: 1) substituting its own CD loop. The resulting genes were then synthesized and cloned into expression vectors that were transformed into competent *E. coli* cells. The mutant proteins containing the modified CD-
- 45 loops were produced by cultivating the transformed cells in enriched liquid media.

Result and Discussion

- [0190] The haem group attached to the globin apoprotein contains iron; as a result, an expressed protein with iron will be red which will result in a red cell pellet. As expected, the results show that

the modification of the CD-loop affects the expression and production of the haemoglobin proteins as observed in the different degrees of red colour of the cell pellets. This confirmed that some mutations have a strengthening or deteriorating effect on the expression and production of haem-bound proteins. A very bad effect resulted in very little to no red pellets while very good resulted in a very intense red pellet. Bad and good were colour grades in between very good and very bad (Figure 9), where good were more red than bad.

Table 9. Effect of the mutations on the production of haem-bound proteins.

SEQ ID NO	Substitution	Sequence	Effect
SEQ ID NO: 56	Reference sequence	PTAKKMFSFVRDSNVPLEQNPKLKAHA	
SEQ ID NO: 111	A3_H_03	GTAKKMFSFVRSDVGLQNGKLKGHA	Very Bad
SEQ ID NO: 112	B3_H_09	PTAKKMFSFVRDGDGPLEQNQKLKAHA	
SEQ ID NO: 113	C3_H_15	GTAKKMFSFVRSDVGLQNGKLKAHA	
SEQ ID NO: 114	D2_H_20	PTAKKMFPFVRSDVPLEQNQKLKGHA	
SEQ ID NO: 115	D5_H_23	GTAKKMFPFVRSDVPLEQNPKLKGHG	
SEQ ID NO: 116	D6_H_24	PTAKKMFSFVRDGDVPLEQNPKLKPHA	
SEQ ID NO: 117	A1_H_01	GTAKKMFSFVRSDVPLEQNQKLKGHA	Bad
SEQ ID NO: 118	A2_H_02	GTAKKMFSFVRSDVGLQNGKLKGHA	
SEQ ID NO: 119	C1_H_13	GTAKKMFSFVRDGDVPLEQNQKLKGHA	
SEQ ID NO: 120	C5_H_17	PTAKKMFSFVRSDVPLEQNQKLKAHG	
SEQ ID NO: 121	D3_H_21	PTAKKMFGFVRSDVPLEQNPKLKGHA	Good
SEQ ID NO: 122	A6_H_06	PTAKKMFSFVRSDVPLEQNPKLKPHA	
SEQ ID NO: 123	B4_H_10	PTAKKMFSFVRDGDVPLEQNQKLKAHA	
SEQ ID NO: 124	B5_H_11	PTAKKMFSFVRDGDVPLEQNPKLKGHA	
SEQ ID NO: 125	C2_H_14	PTAKKMFSFVRDPDVPLEQNPKLKPHA	Very Good
SEQ ID NO: 126	A4_H_04	PTAKKMFSFVRSDVPLEQNPKLKGHA	
SEQ ID NO: 127	A5_H_05	PTAKKMFSFVRSDVPLEQNQKLKPHA	
SEQ ID NO: 128	B1_H_07	PTAKKMFSFVRDGDVPLEQNQKLKGHA	
SEQ ID NO: 129	B2_H_08	PTAKKMFSFVRDGDGPLEQNQKLKGHA	
SEQ ID NO: 130	B6_H_12	PTAKKMFSFVRDGDVPLEQNPKLKAHA	
SEQ ID NO: 131	C4_H_16	PTAKKMFSFVRSDVPLEQNQKLKGHG	
SEQ ID NO: 132	C6_H_18	PTAKKMFSFVRSDVPLEQNPKLKAHG	
SEQ ID NO: 133	D1_H_19	PTAKKMFSFVRSDVPLEQNPKLKGHG	
SEQ ID NO: 134	D4_H_22	PTAKKMFPFVRSDVPLEQNPKLKGHG	

Example 9: CD-loop modifications and effect on the stability of haem

[0191] The insertion and stabilization of the haem group in a globin apoprotein affects its expression and synthesis. This effect can be due to an increased molecular stability of the haem-bound protein. This positive effect can offer tolerance to physicochemical events such as tolerance to high temperature as a stable CD-loop will contribute to a steadiness of the haem-group inside the active site of the haemoglobin protein.

Method

[0192] Three of the mutants who gave Very Good expression levels and one mutant that had Very Bad expression results were chosen to be produced in 2 Liters of enriched cultivation media. After production, the cells were lysed, and the protein purified using chromatography. The thermal stability of the proteins (10-20 µL of sample at a concentration of 10-50 µM in 50mM Tris-HCl pH 8.5, n>3) were analysed through nano-differential scanning fluorimetry (nanoDSF) using a Prometheus NT.48 instrument containing aggregation optics (NanoTemper Technologies). The LED intensity for excitation at 280 nm was set to 20%. Sample fluorescence intensities at 350 nm (F350) and 330 nm (F330) were recorded whilst running a temperature ramp from 20 to 95°C at 7°C/min increments. The melting point (T_m) for the proteins were acquired by a first-order derivative of the F350/F330 fluorescence ratio plotted against the temperature.

Result and Discussion

[0193] The very good haemoglobin proteins had, together, an average melting T° (T_m) above 80 °C, which makes them thermostable proteins compared to other proteins (compared to Table 2 above). In contrast, the very bad haemoglobin protein has a melting temperature below 70 °C. This is expected as this very bad protein had mutations (replacement of the rigid Proline amino acid by the small and flexible Glycine) that increased the flexibility of the CD-loop making it unable to stabilize the haem-group inside the globin apoprotein. The lack of steadiness decreased the intrinsic stability of the very bad haemoglobin.

Table 8. Melting point of proteins containing a modified CD-loop..

Haem- or phyto globin protein	Tm (degrees Celsius)	
Beta vulgaris SEQ ID NO: 7	72,4	Reference
Modified CD-loop A4_H_04	77,0	Very Good
Modified CD-loop B2_H_08	84,5	Very Good
Modified CD-loop C4_H_16	88,1	Very Good
Modified CD-loop A3_H_03	61,2	Very Bad

Example 10: Alternative signal peptide

[0194] Various signal peptides, in particular signal peptides heterologous to the recombinant haem- or phyto globin, can be added to the N-terminal of the recombinant proteins to transport the protein to a specific location inside and outside the cell. In those examples where the molecules already carry a signal peptide (for example SEQ ID NO: 3, SEQ ID NO: 9, SEQ ID NO 16), these can be replaced by other signal peptides. Among the applications of adding signal peptides

is the transport of the haem-protein outside the producing cell to easy its down-stream processing (i.e. isolation from other non-haem molecules). Signal peptides which can be used to direct transport of the recombinant haem- or phytoalbumin include of SEQ ID NO: 203 to 213.

5 **Example 11: Expressing recombinant haem- or phytoalbumin in alternative host strains**

[0195] In addition to E. coli the recombinant haem- or phytoalbumin is successfully expressed in *Arabidopsis thaliana*, *Lepidium campestre*, *Nicotiana tabacum*, *Triticum aestivum*, *Hordeum vulgare*, *Oryza sativa*, *Chenopodium quinoa*, *Beta vulgaris*, *Glycine max*, *P. pastoris*, *Pichia* sp., *S. cerevisiae*, *Yarrowia* sp., *Y. lipolytica*, *Kluyveromyces lactis*, *Saccharomyces carlsbergensis*,
10 *Saccharomyces cerevisiae*, *Saccharomyces diastaticus*, *Saccharomyces douglasii*, *Saccharomyces kluyveri*, *Saccharomyces bensis*, *Saccharomyces oviformis*, *Ashbya gossypii*, *H. polymorpha*, *Hansenula* sp, *C. glutamicum*, *B. subtilis*, *S. marcescens*, *P. putida*, and *P. aeruginosa* using the genes disclosed herein using methods known in the art.

15 **Example 12: Stabilizing recombinant protein with salts and/or antioxidants**

[0196] Addition of 500 mM of a salt selected from NaCl, Ammonium sulphate, CaCl₂, KCl, MgCl₂ or 25 mM antioxidants selected from ascorbic acid, *2-Mercaptoethanol*, *Dithiothreitol (DTT)*, *Superoxide dismutase (SOD)*, *Catalase*, β -carotene, Lycopene, glutathione, melatonin, oestrogen, and biquinol-10, N-acetyl cysteine, lipoic acid, zinc, selenium, copper, quercetin, catechin,
20 cortisone, oestradiol, estriol, and α -tocopherol to a powder or liquid preparation of recombinant protein enhances its stability and the absorption of the iron contained inside the haem group.

References

1. BLASTP online tool (<https://blast.ncbi.nlm.nih.gov>)
- 25 2. Leiva Eriksson N, Reeder B, Wilson M, Bülow L. Sugar beet haemoglobin: reactions with nitric oxide and nitrite reveal differential roles for nitrogen metabolism. *Biochem J* 31 2019; 476 (14): 2111–2125.
3. Sinclair, P.R., Gorman, N. and Jacobs, J.M. (2001) Measurement of haem concentration. *Curr. Prot. Toxicol.* Chapter 8, Unit 8.3.
- 30 4. Sadrzadeh, S. M., Graf, E., Panter, S. S., Hallaway, P. E., and Eaton, J. W. (1984). Haemoglobin. A biologic fenton reagent. *J. Biol. Chem.* 259, 14354–14356.
5. Fiorito V, Chiabrando D, Petrillo S, Bertino F and Tolosano E (2020) The Multifaceted Role of Haem in Cancer. *Front. Oncol.* 9:1540.
6. FAQ_Impossible Foods webpage; <https://faq.impossiblefoods.com/hc/en-us>

7. M.S Hargrove, J.S Olson. The stability of holomyoglobin is determined by haem affinity. *Biochemistry*, 35 (1996), pp. 11310-11318.
8. M.S. Hargrove, et al. Characterization of recombinant soybean leghaemoglobin a and apolar distal histidine mutants. *J. Mol. Biol.*, 266 (1997), pp. 1032-1042.
- 5 9. M. Renerre, M. Anton, P. Gatellier. Autoxidation of purified myoglobin from two bovine muscles. *Meat Science*, 32 (3) (1992), pp. 331-342.
10. M. Simons et al. Comparison of the oxidative reactivity of recombinant fetal and adult human haemoglobin: implications for the design of haemoglobin-based oxygen carriers. *Biosci Rep* 31 August 2018; 38 (4): BSR20180370.
- 10 11. Rossi-Fanelli, A., E. Antonini, and A. Caputo. 1958. Studies on the structure of haemoglobin. I. Physico-chemical properties of human globin. *Biochim. Biophys. Acta*. 30: 608.
12. R. Bhomia, V. Trivedi, N.J. Coleman, J.C. Mitchell. The thermal and storage stability of bovine haemoglobin by ultraviolet-visible and circular dichroism spectroscopies. *J. Pharm. Anal.*, 6 (4) (2016), pp. 242-248.
- 15 13. <https://www.fda.gov/media/124351/download>
14. Chakane S, Matos T, Kettisen K, Bulow L. Fetal haemoglobin is much less prone to DNA cleavage compared to the adult protein. *Redox Biol.* 2017, 12:114-120.
15. Glei M, Klenow S, Sauer J, Wegewitz U, Richter K, Pool-Zobel BL. Haemoglobin and hemin induce DNA damage in human colon tumor cells HT29 clone 19A and in primary human colonocytes. *Mutat Res.* 2006, 22; 594(1-2):162-71.
- 20 16. Alayash AI, Patel RP, Cashon RE. Redox reactions of haemoglobin and myoglobin: biological and toxicological implications. *Antioxid Redox Signal.* 2001.3(2):313-27.
17. Glahn RP, Lee OA, Yeung A, Goldman MI, Miller DD. Caco-2 cell ferritin formation predicts nonradiolabeled food iron availability in an in vitro digestion/Caco-2 cell culture model. *J Nutr.* 1998 128(9):1555-61.
- 25 18. Glahn R, Tako E, Hart J, Haas J, Lung'aho M, Beebe S. Iron Bioavailability Studies of the First Generation of Iron-Biofortified Beans Released in Rwanda. *Nutrients.* 2017 21;9(7):787.
19. M. Schultz and J.P. Burton: Chapter 5 - *Escherichia coli* Nissle 1917; The Microbiota in Gastrointestinal Pathophysiology; Implications for Human Health, Prebiotics, Probiotics, and Dysbiosis 2017, Pages 59-69Items
- 30

Items

[0197] The present disclosure further discloses the following first set of itemized embodiments:

1. A recombinant phytoalbumin comprising a hexacoordinated heme group, wherein said
- 35 phytoalbumin has at least 86% identity to amino acid sequence SEQ ID NO: 5; or a fragment, variant,

or fusion thereof retaining the hexacoordinated heme group; for use in treatment or prevention of iron deficiency in a subject.

2. The recombinant phytoalbumin according to item 1, wherein said phytoalbumin has a sequence
5 identity of at least 87%, such as at least 88%, such as at least 89%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99% to sequence SEQ ID NO: 5.
- 10 3. The recombinant phytoalbumin according to any one of the preceding items, wherein the phytoalbumin is a nonsymbiotic haemoglobin of Class 1.
4. The recombinant phytoalbumin according to any one of the preceding items, wherein the recombinant phytoalbumin is derived from a plant in Caryophyllidae, such as derived from spinach,
15 quinoa.
5. The recombinant phytoalbumin according to any one of the preceding items comprising or consisting of an amino acid sequence SEQ ID NO 5, or a fragment, variant, or fusion thereof retaining the hexacoordinated heme group.
20
6. The recombinant phytoalbumin according to any one of the preceding items comprising or consisting of an amino acid sequence SEQ ID NO 4, or a fragment, variant, or fusion thereof retaining the hexacoordinated heme group.
- 25 7. The recombinant phytoalbumin according to any one of the preceding items, wherein said phytoalbumin has an improved stability compared to a pentacoordinate haemoglobin.
8. The recombinant phytoalbumin according to any one of the preceding items, wherein the phytoalbumin has an auto-oxidation rate of less than 0.2, such as less than 0.15, for example in the
30 range from 0.01 to 0.18, or 0.01 to 0.15, such as about 0.1 h^{-1}
9. The recombinant phytoalbumin according to any one of the preceding items, wherein said phytoalbumin has no metallic taste, or has neutral taste.

10. The recombinant phytoalbumin according to any one of the preceding items, wherein said phytoalbumin has a melting temperature of above 68 °C, such as in the range from 68 to 75 °C; such as in the range from 70 to 76 °C.
- 5 11. A recombinant phytoalbumin comprising a hexacoordinated heme group, and further comprising a CD variable loop according to the preceding items, for use in treatment or prevention of iron deficiency in a subject, wherein said region or variable domain has at least 60% sequence identity to the SEQ ID NO 111.
- 10 12. A recombinant phytoalbumin comprising a hexacoordinated heme group, and further comprising a variable domain, for use in treatment or prevention of iron deficiency in a subject, wherein said CD variable loop has at least 90% sequence identity to the SEQ ID NO 111.
13. A recombinant phytoalbumin comprising a hexacoordinated heme group, and further
15 comprising a CD variable loop, for use in treatment or prevention of iron deficiency in a subject, wherein said CD variable loop comprises the SEQ ID NO 111, with one substitution, or with two substitutions.
14. A recombinant phytoalbumin having a hexacoordinated heme group for use in treatment or
20 prevention of iron deficiency in a subject, wherein the variable domain comprises or consists of SEQ ID NO 111, SEQ ID NO 112, SEQ ID NO 113, SEQ ID NO 114, SEQ ID NO 115 or SEQ ID NO 116.
15. A composition comprising at least one recombinant phytoalbumin according to any of items 1
to 14, for use in food.
- 25 16. A composition comprising at least one recombinant phytoalbumin according to any of items 1 to 14, for use in treatment or prevention of iron deficiency.
17. The composition according to item 15 or 16, wherein the composition comprises at least 0.002
30 % w/w of said recombinant phytoalbumin.
18. The composition according to item 15 or 16, wherein the composition comprises at least 0.1
uM of recombinant phytoalbumin, such as for example from 0.1 uM to 8 M recombinant
phytoalbumin.

19. The composition according to any of preceding items 15 to 18, wherein the composition has no metallic taste.
20. The composition according to any one of the preceding items 15 to 19, further comprising one
5 or more reducing agent, such as for example one or more of ascorbic acid (vitamin C), tocopherols, carotenoids, flavonoids, and glutathione.
21. The composition according to any one of the preceding items 15- 20, further comprising at least one vitamin and/or mineral, for example one or more selected from the group consisting of
10 vitamin C, vitamin D3, vitamin E, vitamin B6, vitamin B12, folate, folic acid, biotin, zinc, copper, magnesium, selenium and combinations thereof.
22. The composition according to any one of the preceding items 15 to 21, wherein the composition has a formulation selected from the group consisting of tablet, capsule, liquid, drops,
15 concentrate, powder, granule and combinations thereof.
23. The composition according to any one of the preceding items 15 to 22, wherein the composition is a feed, food, beverage, food ingredient and/or nutritional supplement.
- 20 24. The composition according to item 23, wherein the composition is a nutritional supplement.
25. The composition according to item 23, wherein the composition is a food ingredient.
26. The composition according to any one of the preceding items 15 to 21, wherein the composition is a pharmaceutical composition selected from the group consisting of drugs,
25 vaccines, personal care composition and combinations thereof.
27. The composition according to item 26, wherein the composition further comprises a pharmaceutical acceptable carrier.
- 30 28. The composition according to any one of the preceding items 15 to 23 use in a food composition.
29. A recombinant phytoalbumin comprising a hexacoordinated heme group, wherein said phytoalbumin has at least 60% identity to amino acid sequence SEQ ID NO: 5, or a fragment, variant,
35 or fusion thereof retaining the hexacoordinated heme group; or a composition comprising said

recombinant phytoglobin; for use as a medicament.

30. The recombinant phytoglobin and/or the composition comprising the recombinant phytoglobin according to any of preceding items 15 to item 28 for use in treatment, amelioration
5 or prevention of iron deficiency in humans and animals.

31. The recombinant phytoglobin and/or the composition comprising the recombinant phytoglobin for use according to items 15 to 28, wherein said use is for treatment, amelioration or prevention of iron deficiency in women selected from the group consisting of pregnant women,
10 breastfeeding women, women in reproductive age.

32. The recombinant phytoglobin and/or the composition comprising the recombinant phytoglobin for use according to items 15 to 28, wherein said use is for treatment, amelioration or prevention of iron deficiency in a patient with a chronic condition, such as a chronic condition
15 selected from the group consisting of endometriosis, dysmenorrhea, menorrhagia, cancer, chronic kidney disease, diabetes, obesity, digestive disorders such as celiac disease, inflammatory bowel disease, peptic ulcers, genetic disorders such as haemolytic anaemia, autoimmune disease and/or conditions associated with ageing.

20 33. The recombinant phytoglobin and/or the composition comprising the recombinant phytoglobin for use according to items 15 to 28, wherein said use is for treatment, amelioration or prevention of iron deficiency in a subjects with an acute condition, such as an acute condition selected from the group consisting of trauma, gastrointestinal surgery, gastrointestinal bleeding, kidney failure, acute toxicity.

25 34. The recombinant phytoglobin and/or the composition comprising the recombinant phytoglobin for use according to items 15 to 28, wherein said use is for treatment, amelioration or prevention of iron deficiency in humans with plant-based diets selected from the group consisting of vegetarian, vegan and/or flexitarian.

30 35. The recombinant phytoglobin and/or the composition comprising the recombinant phytoglobin for use according to items 15 to 28, wherein said use is for treatment, amelioration or prevention of iron deficiency in animals, such as in companion animals, poultry or livestock.

35 36. A composition comprising at least one phytoglobin, wherein said at least one phytoglobin has

at least 60% identity to amino acid sequence SEQ ID NO: 5, or a fragment, variant, or fusion thereof retaining the hexacoordinated heme group; and further wherein said composition comprises at least 0.002% w/w of said at least one phytoalbumin; for use in treatment or prevention of iron deficiency.

5

37. A liquid composition comprising at least one phytoalbumin comprising a hexacoordinated heme group, wherein said phytoalbumin has at least 86% identity to amino acid sequence SEQ ID NO: 5; or a fragment, variant, or fusion thereof which retains the hexacoordinated heme group, for use in treatment or prevention of iron deficiency in a subject.

10

[0198] The present disclosure further discloses the following second set of itemized embodiments:

1. A recombinant phytoalbumin comprising a hexacoordinated heme group, wherein said phytoalbumin has at least 86% identity to amino acid sequence SEQ ID NO: 5; or a fragment, variant, or fusion thereof retaining the hexacoordinated heme group; for use in food.

15

2. The recombinant phytoalbumin according to item 1, wherein said phytoalbumin has a sequence identity of at least 87%, such as at least 88%, such as at least 89%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99% to sequence SEQ ID NO: SEQ ID NO: 5.

20

3. The recombinant phytoalbumin according to any one of the preceding items comprising or consisting of an amino acid sequence SEQ ID NO: 5, or a fragment, variant, or fusion thereof retaining the hexacoordinated heme group.

25

4. The recombinant phytoalbumin according to any one of the preceding items comprising or consisting of an amino acid sequence SEQ ID NO: 4, or a fragment, variant, or fusion thereof retaining the hexacoordinated heme group.

30

5. A recombinant phytoalbumin comprising a hexacoordinated heme group, and further comprising a CD variable loop according to the preceding items, for use in treatment or prevention of iron deficiency in a subject, wherein said region or variable domain has at least 60% sequence identity to the SEQ ID NO: 111.

35

6. A composition comprising at least one recombinant phytoalbumin according to any of items 1 to 5, for use in food.
7. A composition comprising at least one recombinant phytoalbumin according to any of items 1 to 5, for use in treatment or prevention of iron deficiency.
8. The composition according to any of items 6 to 7, wherein the composition comprises at least 0.1 μ M of recombinant phytoalbumin, such as for example from 0.1 μ M to 8 M recombinant phytoalbumin.
9. The composition according to any of preceding items 6 to 8, wherein the composition has no metallic taste.
10. A liquid composition comprising at least one phytoalbumin comprising a hexacoordinated heme group, wherein said phytoalbumin has at least 86% identity to amino acid sequence SEQ ID NO: 5; or a fragment, variant, or fusion thereof which retains the hexacoordinated heme group, for use in treatment or prevention of iron deficiency in a subject.
- [0199] The present disclosure further discloses the following third set of itemized embodiments:
1. A recombinant haem- or phytoalbumin protein comprising a C-helix and a D-helix covalently joint into a CD-loop domain and comprising a hexa- or penta-coordinated prosthetic haem group, wherein the haem- or phytoalbumin has an amino acid sequence which is at least 50% identical the haem- or phytoalbumin comprised in anyone of SEQ ID NO: 1 to 55.
2. The recombinant protein of item 1, wherein the CD-loop comprises one or more modifications compared to a parent CD-loop domain, whereby the modified CD-loop domain binds the prosthetic group stronger than the parent unmodified CD-loop domain.
3. The recombinant protein of any preceding item wherein the parent CD-loop domain is native to the protein.
4. The recombinant protein of any preceding item wherein the parent CD-loop domain is positioned in the recombinant haem- or phytoalbumin protein corresponding to the position 52 to 78 of SEQ ID NO: 1, optionally position 52 to 72 of SEQ ID NO: 1.

5. The recombinant protein of any preceding item wherein the CD-loop domain has an amino acid sequence which is at least 50% identical to the CD-loop domain comprised in anyone of SEQ ID NO: 56 to 147.
- 5 6. The recombinant protein of item 2 to 5 wherein the modification in the CD-loop domain is selected from deletion, substitution and/or addition of one or more amino acids.
7. The recombinant protein of item 2 to 6 wherein the modification of the CD-loop domain changes the flexibility or rigidity of the CD-loop domain compared to the parent CD-loop domain.
- 10 8. The recombinant protein of item 2 to 7 wherein the CD-loop domain comprises two or more, such as three or more, such as four or more substitutions compared to the parent CD-loop domain.
9. The recombinant protein of item 2 to 8 wherein the amino acid modification in the CD-loop
15 corresponds to one or more modifications is selected from S13G, S13P, V15G, A25P, and A27G of the CD-loop as set forth in SEQ ID NO: 56.
10. The recombinant protein of item 2 to 8 wherein the amino acids corresponding to positions P1, A3, F7, S8, S13, P16, P21 and/or P25 of SEQ ID NO: 56 are conserved or conservatively
20 substituted.
11. The recombinant protein of item 2 to 10 wherein the CD-loop comprises the amino acid sequence as set forth in anyone of SEQ ID NO: 123, 127, 128, 129, 130, 131, 133, 134, 135, 137, 139, 141, 142, or 145.
- 25 12. The recombinant protein of item 11 wherein the CD-loop comprises the amino acid sequence as set forth in anyone of SEQ ID NO: 127, 128, 130, 131, 135, 139, 141, 142, or 145.
13. The recombinant protein of any preceding item further comprising a signal peptide, directing
30 the recombinant proteins secretion from a cell.
14. The recombinant protein of item 13 wherein the signal peptide directs expression in a microbial cell.

15. The recombinant protein of item 13 to 14 wherein the signal peptide is heterologous to the recombinant protein.
16. The recombinant protein of item 13 to 15 wherein the microbial cell is a bacterial cell or a
5 fungal cell.
17. The recombinant protein of item 13 to 16 wherein the signal peptide has an amino acid sequence which is at least 50% identical to the signal peptide comprised in anyone of SEQ ID NO: 203 to 213.
10
18. The recombinant protein of any preceding item wherein the recombinant protein is non-natural or synthetic.
19. The recombinant protein of item 18 wherein the non-natural or synthetic recombinant protein
15 is a fusion protein.
20. The recombinant protein of item 19 wherein the fusion protein comprises two of more prosthetic groups.
- 20 21. The recombinant protein of any preceding wherein the recombinant protein is a nonsymbiotic haemoglobin of Class 1 or Class 2.
22. The recombinant protein of any preceding wherein the recombinant protein is derived from a plant of the genus Caryophyllidae.
25
23. The recombinant protein of item 22 wherein the recombinant protein is derived from spinach, or quinoa.
24. The recombinant protein of any preceding item wherein the recombinant protein comprises
30 at least one post translational modification selected from glycosylations or phosphorylations compared to a corresponding natural protein.
25. A composition comprising the recombinant protein of any preceding and one or more carriers, agents, additives and/or excipients.
35

26. The composition of item 25 wherein the one or more carriers, agents, additives and/or excipients are selected from salts, antioxidants and/or reducing agents.
27. The composition of item 26 wherein the salt is selected from NaCl, ammonium sulphate, CaCl₂,
5 KCl, MgCl₂.
28. The composition of item 26 to 27 wherein the amount of salt in the composition is between 10 µM to 2M.
- 10 29. The composition of item 26 to 28 wherein the antioxidant is selected from ascorbic acid, 2-mercaptoethanol, dithiothreitol (DTT), superoxide dismutase (SOD), catalase, β-carotene, lycopene, glutathione, melatonin, oestrogen, and biquinol-10, N-acetyl cysteine, lipoic acid, salts of zinc, selenium, and/or copper, quercetin, catechin, cortisone, oestradiol, estriol, and α-tocopherol.
- 15 30. The composition of item 26 to 29 wherein the amount of antioxidant in the composition is between 10 µM to 1M.
- 20 31. The composition of items 26 wherein the reducing agent is selected from ascorbic acid (vitamin C), tocopherols, carotenoids, flavonoids, and glutathione.
32. The composition of item 25 to 31, further comprising at least one vitamin and/or mineral selected from vitamin C, vitamin D3, vitamin E, vitamin B6, vitamin B12, folate, folic acid, biotin, zinc, copper, magnesium, selenium and combinations thereof.
- 25 33. The composition according to item 25 to 30 comprising at least 0.002 % w/w of said recombinant protein.
34. The composition according to item 25 to 30 comprising at least 0.1 uM of recombinant protein,
30 such as for example from 0.1 uM to 8 M recombinant protein.
35. The composition of item 25 to 34 wherein the composition is selected from a food product, a feed product, a beverage, a food ingredient, a dietary or nutritional supplement, and/or a pharmaceutical product.

36. The composition according to item 35, wherein the composition is a nutritional supplement.
37. The composition according to item 35, wherein the composition is a food ingredient.
- 5 38. The composition of item 35 wherein the composition is a food composition suitable for and nutritionally balanced for children, elderly or subjects undergoing medical or surgical treatment.
39. The composition of item 25 to 38 wherein the composition has no metallic flavour or taste.
- 10 40. The composition of item 25 to 39, wherein the composition is formulated as a dry formulation, a liquid formulation or a slurry or dispersion.
41. The composition of item 40 having a formulation selected from the group consisting of tablet, capsule, liquid, drop, concentrate, powder, granule and combinations thereof.
- 15 42. The composition according to item 25 to 41 further comprising a pharmaceutical acceptable carrier.
43. The composition according to item 25 to 42, wherein the composition is a pharmaceutical
- 20 composition selected from the group consisting of drugs, vaccines, personal care composition and combinations thereof.
44. A non-therapeutic method of improving the stamina or low oxygen tolerance in a subject, comprising administering the recombinant protein of item 1 to 24 or the composition of item 25
- 25 to 43 to the subject in an amount effective to improve stamina or low oxygen tolerance in the subject.
45. The method of item 44 wherein the subject is performing a physical activity.
- 30 46. The method of item 45 wherein the subject is performing mountaineering or mountain climbing operating under low oxygen conditions.
47. A non-therapeutic method of altering the skin colour or texture in a subject comprising administering the recombinant protein of item 1 to 24 or the composition of item 25 to 43 into
- 35 skin of the subject in an amount effective to alter the skin colour or texture.

48. A method for treating, ameliorating or preventing of a disease or deficiency in a subject, comprising administering the recombinant protein of item 1 to 24 or the composition of item 25 to 43 to the subject in an amount effective to treat, ameliorate or prevent said disease or deficiency.

49. The method of item 48 wherein the disease or deficiency is iron deficiency or any disease resulting from iron deficiency.

50. The recombinant protein of item 1 to 24 or the composition of item 25 to 43 for the use in treating, ameliorating or preventing of a disease or deficiency in a subject, comprising administering the recombinant protein or the composition to the subject in an amount effective to ameliorate or prevent said disease or deficiency.

51. The recombinant protein composition of item 50 wherein the disease or deficiency is iron deficiency or any disease resulting from iron deficiency.

52. The recombinant protein or composition of item 50 to 51 wherein the recombinant protein is a phytohemoglobin.

53. The recombinant protein or composition of item 50 to 52 wherein the recombinant phytohemoglobin comprises a hexacoordinated haem group and has an amino acid sequence which is at least 86% identical to the phytohemoglobin comprised in SEQ ID NO: 1 or 5.

54. The recombinant protein or composition of item 53 wherein the recombinant phytohemoglobin has an amino acid sequence which is at least 87%, such as at least 88%, such as at least 89%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99% identical to the phytohemoglobin comprised in sequence SEQ ID NO: 1 or 5.

55. The recombinant protein or composition of item 52 to 54 wherein the recombinant phytohemoglobin is a nonsymbiotic hemoglobin of Class 1.

56. The recombinant protein or composition of item 52 to 55 wherein the recombinant phytohemoglobin is derived from a plant in Caryophyllidae, such as derived from spinach, quinoa.

57. The recombinant protein or composition of item 52 to 56 wherein the recombinant phytoalbumin has an amino acid sequence comprised in or consisting of SEQ ID NO: 4.
- 5 58. The recombinant protein or composition of item 50 to 57 wherein the recombinant protein comprises a hexacoordinated haem group and has an improved stability compared to a pentacoordinate haemoglobin.
59. The recombinant protein or composition of item 50 to 58 wherein the recombinant protein
10 has an auto-oxidation rate of less than 0.2, such as less than 0.15, for example in the range from 0.01 to 0.18, or 0.01 to 0.15, such as about 0.1 h⁻¹
60. The recombinant protein or composition of item 50 to 59 wherein the recombinant protein has no metallic taste or has neutral taste.
- 15 61. The recombinant protein or composition of item 50 to 60 wherein recombinant protein has a melting temperature of above 68°C, such as in the range from 68 to 75°C; such as in the range from 70 to 76°C, such as above 76°C, such as above 77°C, such as above 78°C, such as above 79°C, such as above 80°C, such as above 81°C, such as above 82°C, such as above 83°C, such as about
20 84°C.
62. The recombinant protein or composition of item 52 wherein the recombinant phytoalbumin comprises a hexacoordinated haem group, and further comprises a CD-loop domain which has an amino acid sequence which is at least 60% sequence identical to the CD-loop domain comprised
25 in SEQ ID NO: 56, 60 or 111.
63. The recombinant protein or composition of item 62 wherein the CD-loop domain has an amino acid sequence which is at least 90% identical to the CD-loop domain comprised in SEQ ID NO: 56, 60 or 111.
- 30 64. The recombinant protein or composition of item 62 to 63 wherein the CD-loop domain has an amino acid sequence which comprises the DC loop domain comprised in SEQ ID NO: 56, 60 or 111 having one or more, two or more, three or more, or four or more amino acid substitutions.

65. The recombinant protein or composition of item 62 to 64 wherein the CD-loop domain has an amino acid sequence which comprises or consists of SEQ ID NO: 111, SEQ ID NO: 112, SEQ ID NO: 113, SEQ ID NO: 114, SEQ ID NO: 115 or SEQ ID NO: 116.
- 5 66. The recombinant protein or composition of item 50 to 65 wherein the subject is a blood donor.
67. The recombinant protein or composition of item 50 to 66 wherein the subject is a vegetarian, vegan and/or flexitarian.
- 10 68. The recombinant protein or composition of item 50 to 67 wherein the subject is an animal, optionally a companion animal, poultry or livestock.
69. The recombinant protein or composition of item 50 to 67 wherein the subject is a woman optionally selected from the group of pregnant women, breastfeeding women, and women in reproductive age.
- 15 70. The recombinant protein or composition of item 50 to 69 wherein the disease is selected from endometriosis, dysmenorrhea, menorrhagia, cancer, kidney disease, diabetes, obesity, metabolic syndrome, digestive disorders, psychological disorders, genetic disorders, conditions associated with ageing, acute conditions, and/or infections.
- 20 71. The recombinant protein or composition of item 70 wherein the digestive disorder is selected from celiac disease, inflammatory bowel disease, and/or peptic ulcers
- 25 72. The recombinant protein or composition of item 70 wherein the genetic disorder is selected from haemolytic anaemia, and/or autoimmune disease.
73. The recombinant protein or composition of item 70 wherein the psychological disorder is selected from bulimia and/or anorexia.
- 30 74. The recombinant protein or composition of item 70 wherein the infection is selected from malaria and/or streptococcus infections.
75. The recombinant protein or composition of item 70 wherein the disease is chronic.
- 35

76. The recombinant protein or composition of item 70 wherein the acute condition is selected from trauma, gastrointestinal surgery, gastrointestinal bleeding, kidney failure, and/or acute toxicity.
- 5 77. The method or the recombinant protein or composition of item 50 to 76 wherein the amount of recombinant protein administered to the subject is between 0,1 mg to 350 mg per kg body weight of the subject, such as between 0.25 mg to 100 mg, such as between 0,5 to 50 mg, such as between 0.75 mg to 10 mg, such as between 0,75 mg to 3 mg per kg body weight.
- 10 78. The method or the recombinant protein or composition of item 77 wherein the amount of recombinant protein administered to the subject is between 0,75 mg to 3 mg per kg body weight.
79. The method or the recombinant protein or composition of item 78 wherein the subject is a male.
- 15 80. The method or the recombinant protein or composition of item 79 wherein the amount of recombinant protein administered to the subject is between 0,75 mg to 1,25, such as about 1 mg per kg body weight.
- 20 81. The method or the recombinant protein or composition of item 78 wherein the subject is a female.
82. The method or the recombinant protein or composition of item 81 wherein the amount of recombinant protein administered to the subject is between 1,3 mg to 1,5, such as about 1,4 mg
- 25 per kg body weight.
83. A gene encoding the recombinant protein of item 1 to 24.
84. The gene of item 83 having a nucleotide sequence which is at least 50% identical to the
- 30 nucleotide sequence comprised in any of SEQ ID NO: 148 to 202 encoding the corresponding recombinant protein of SEQ ID NO: 1 to 55.
85. A polynucleotide construct comprising the nucleotide sequence of item 83 or 84, operably linked to a control sequence directing the transcription and/or translation of the gene in a host
- 35 cell.

86. The polynucleotide construct of item 85 wherein the control sequence is a promoter.

87. The polynucleotide construct of item 86 wherein the promoter is an inducible promoter.

5

88. The polynucleotide construct of item 87 wherein the promoter is at least 50 % identical to the promoter comprised in SEQ ID NO: 225 to 235.

89. The polynucleotide construct of item 85 to 88 wherein the construct is an expression vector.

10

90. A genetically modified host cell expressing the gene or polynucleotide construct of item 83 to 89 and the recombinant protein of item 1 to 24.

91. The host cell of item 90 wherein the cell is a eukaryotic cell, a bacterial cell, or an archaeal cell.

15

92. The host cell of item 91 wherein the eukaryotic cell is a fungal cell, a plant cell, a mammalian cell or an insect cell.

93. The genetically modified host cell of items 92 wherein the plant cell is selected from the genus of *Arabidopsis*, *Lepidium*, *Nicotiana*, *Triticum*, *Hordeum*, *Oryza*, *Chenopodium*, *Beta*, or *Glycine*.

20

94. The genetically modified host cell of items 93 wherein the plant cell is selected from the species of *Arabidopsis thaliana*, *Lepidium campestre*, *Nicotiana tabacum*, *Triticum aestivum*, *Hordeum vulgare*, *Oryza sativa*, *Chenopodium quinoa*, *Beta vulgaris*, or *Glycine max*.

25

95. The genetically modified host cell of items 92 wherein the fungal cell is a yeast or a filamentous fungus.

96. The genetically modified host cell of items 95 wherein the filamentous fungus is selected from the genus of *Aspergillus*, *Trichoderma*, or *Rhizopus*.

30

97. The genetically modified host cell of items 96 wherein the filamentous fungus is selected from the species *Aspergillus* sp, *Aspergillus oryzae*, *Trichoderma* sp, or *Rhizopus* sp.

98. The genetically modified host cell of items 95 wherein the yeast is selected from the genus of *Pichia*, *Saccharomyces*, *Yarrowia*, *Kluyveromyces*, *Ashbya* or *Hansenula*.
99. The genetically modified host cell of items 98 wherein the yeast is selected from the species
5 of *P. pastoris*, *Pichia* sp., *S. cerevisiae*, *Yarrowia* sp., *Y. lipolytica*, *Kluyveromyces lactis*,
Saccharomyces carlsbergensis, *Saccharomyces cerevisiae*, *Saccharomyces diastaticus*,
Saccharomyces douglasii, *Saccharomyces kluyveri*, *Saccharomyces bensis*, *Saccharomyces*
oviformis, *Ashbya gossypii*, *H. polymorpha*, or *Hansenula* sp.
100. The genetically modified host cell of items 91 wherein the bacterial cell is selected from the
10 genus of *Escherichia*, *Bacillus*, *Brevibacterium*, *Burkholderia*, *Campylobacter*, *Corynebacterium*,
Serratia, *Lactobacillus*, *Lactococcus*, *Acinetobacter*, *Acetobacter* or *Pseudomonas*.
101. The genetically modified host cell of items 100 wherein the bacterial cell is selected from the
15 species *E. coli*, *C. glutamicum*, *B. subtilis*, *S. marcescens*, *P. putida*, *P. aeruginosa* and/or *P.*
mutabilis.
102. The genetically modified host cell of items 91 wherein the archaeal cell is an algae.
103. A cell culture comprising the host cell of item 90 to 102 and a growth medium.
20
104. A method for producing the recombinant protein of item 1 to 24 comprising:
a. culturing the cell culture of item 103 at conditions allowing the cell to produce the
recombinant protein; and
25 b. optionally recovering and/or isolating the recombinant protein.
105. The method of item 104, further comprising one or more elements selected from:
a) culturing the cell culture in a nutrient medium;
b) culturing the cell culture under aerobic or anaerobic conditions
30 c) culturing the cell culture under agitation;
d) culturing the cell culture at a temperature of between 10 to 50 °C;
e) culturing the cell culture at a pH of between 3-9;
f) culturing the cell culture for between 10 hours to 30 days; and
g) culturing the cell culture under fed-batch, repeated fed-batch, continuous, or semi-
35 continuous conditions.

106. The method of item 105 wherein the nutrient medium has the properties of:

- a) pH 5-8,
- b) Temperature 17-38 °C,

5

107. The method of item 104 to 106 wherein the recovery and/or isolation step comprises separating a liquid phase of the cell or cell culture from a solid phase of the cell or cell culture to obtain a supernatant comprising the recombinant protein and/or subjecting the supernatant to one or more steps selected from:

- 10 a) separating the supernatant from the solid phase of the cell culture, such as by filtration or gravity separation;
 - b) contacting the supernatant with one or more adsorbent resins to obtain at least a portion of the produced recombinant protein;
 - c) contacting the supernatant with one or more ion exchange or reversed-phase
15 chromatography columns or filtration or ultrafiltration devices to obtain at least a portion of the recombinant protein;
 - d) extracting the recombinant protein; and/or
 - e) precipitating the recombinant protein by crystallization or evaporating the solvent of the liquid phase; and optionally isolating the recombinant protein by filtration or gravity
20 separation;
- thereby recovering and/or isolating the recombinant protein.

108. A composition comprising the cell culture of item 103.

- 25 109. The composition of item 108 wherein the composition is edible, and the cell culture is in the form of an edible biomass and mycelium.

* * *

Claims

1. A recombinant haem- or phytoalbumin protein comprising a C-helix and a D-helix covalently joint by a CD-loop domain and comprising a hexa- or penta-coordinated prosthetic haem group, wherein the haem- or phytoalbumin has an amino acid sequence which is at least 50% identical the haem- or phytoalbumin comprised in anyone of SEQ ID NO: 1 to 55, wherein the CD-loop domain comprises one or more modifications compared to a parent CD-loop domain, whereby the modified CD-loop domain binds the prosthetic group stronger than the parent unmodified CD-loop domain and wherein the parent CD-loop domain, wherein the parent CD-loop domain is positioned in the recombinant haem- or phytoalbumin protein corresponding to the position 52 to 78 of SEQ ID NO: 1, optionally position 52 to 72 of SEQ ID NO: 1 and wherein the CD-loop domain has an amino acid sequence which is at least 50% identical to the CD-loop domain comprised in anyone of SEQ ID NO: 56 to 147.
2. The recombinant protein of any preceding claim wherein the parent CD-loop domain is native to the protein.
3. The recombinant protein of any preceding claim wherein the modification in the CD-loop domain is selected from deletion, substitution and/or addition of one or more amino acids.
4. The recombinant protein of any preceding claim wherein the modification of the CD-loop domain changes the flexibility or rigidity of the CD-loop domain compared to the parent CD-loop domain.
5. The recombinant protein of any preceding wherein the CD-loop domain comprises two or more, such as three or more, such as four or more substitutions compared to the parent CD-loop domain.
6. The recombinant protein of any preceding claim wherein the amino acid modification in the CD-loop corresponds to one or more modifications is selected from S13G, S13P, V15G, A25P, and A27G of the CD-loop as set forth in SEQ ID NO: 56.
7. The recombinant protein of claim 1 to 5 wherein the amino acids corresponding to positions P1, A3, F7, S8, S13, P16, P21 and/or P25 of SEQ ID NO: 56 are conserved or conservatively substituted.

8. The recombinant protein of any preceding wherein the CD-loop comprises the amino acid sequence as set forth in anyone of SEQ ID NO: 123, 127, 128, 129, 130, 131, 133, 134, 135, 137, 139, 141, 142, or 145.
- 5 9. The recombinant protein of claim 8 wherein the CD-loop comprises the amino acid sequence as set forth in anyone of SEQ ID NO: 127, 128, 130, 131, 135, 139, 141, 142, or 145.
10. The recombinant protein of any preceding claim further comprising a signal peptide, directing the recombinant proteins secretion from a cell.
- 10 11. The recombinant protein of claim 10 wherein the signal peptide directs expression in a microbial cell.
12. The recombinant protein of claim 10 to 11 wherein the signal peptide is heterologous to the recombinant protein.
- 15 13. The recombinant protein of claim 10 to 12 wherein the microbial cell is a bacterial cell or a fungal cell.
- 20 14. The recombinant protein of claim 10 to 13 wherein the signal peptide has an amino acid sequence which is at least 50% identical to the signal peptide comprised in anyone of SEQ ID NO: 203 to 213.
- 25 15. The recombinant protein of any preceding claim wherein the recombinant protein is non-natural or synthetic.
16. The recombinant protein of claim 15 wherein the non-natural or synthetic recombinant protein is a fusion protein.
- 30 17. The recombinant protein of claim 16 wherein the fusion protein comprises two of more prosthetic groups.
18. The recombinant protein of any preceding wherein the recombinant protein is a nonsymbiotic haemoglobin of Class 1 or Class 2.
- 35

19. The recombinant protein of any preceding wherein the recombinant protein is derived from a plant of the genus Caryophyllidae.
20. The recombinant protein of claim 19 wherein the recombinant protein is derived from spinach,
5 or quinoa.
21. The recombinant protein of any preceding claim wherein the recombinant protein comprises at least one post translational modification selected from glycosylations or phosphorylations compared to a corresponding natural protein.
10
22. A composition comprising the recombinant protein of any preceding and one or more carriers, agents, additives and/or excipients.
23. The composition of claim 22 wherein the one or more carriers, agents, additives and/or
15 excipients are selected from salts, antioxidants and/or reducing agents.
24. The composition of claim 23 wherein the salt is selected from NaCl, ammonium sulphate, CaCl₂, KCl, MgCl₂.
- 20 25. The composition of claim 23 to 24 wherein the amount of salt in the composition is between 10 µM to 2M.
26. The composition of claim 23 to 25 wherein the antioxidant is selected from ascorbic acid, 2-mercaptoethanol, dithiothreitol (DTT), superoxide dismutase (SOD), catalase, β-carotene,
25 lycopene, glutathione, melatonin, oestrogen, and biquinol-10, N-acetyl cysteine, lipoic acid, salts of zinc, selenium, and/or copper, quercetin, catechin, cortisone, oestradiol, estriol, and α-tocopherol.
27. The composition of claim 23 to 26 wherein the amount of antioxidant in the composition is
30 between 10 µM to 1M.
28. The composition of claims 23 wherein the reducing agent is selected from ascorbic acid (vitamin C), tocopherols, carotenoids, flavonoids, and glutathione.

29. The composition of claim 22 to 28, further comprising at least one vitamin and/or mineral selected from vitamin C, vitamin D3, vitamin E, vitamin B6, vitamin B12, folate, folic acid, biotin, zinc, copper, magnesium, selenium and combinations thereof.
- 5 30. The composition according to claim 22 to 27 comprising at least 0.002 % w/w of said recombinant protein.
31. The composition according to claim 22 to 27 comprising at least 0.1 uM of recombinant protein, such as for example from 0.1 uM to 8 M recombinant protein.
- 10 32. The composition of claim 22 to 31 wherein the composition is selected from a food product, a feed product, a beverage, a food ingredient, a dietary or nutritional supplement, and/or a pharmaceutical product.
- 15 33. The composition according to claim 32, wherein the composition is a nutritional supplement.
34. The composition according to claim 32, wherein the composition is a food ingredient.
35. The composition of claim 32 wherein the composition is a food composition suitable for and
20 nutritionally balanced for children, elderly or subjects undergoing medical or surgical treatment.
36. The composition of claim 22 to 35 wherein the composition has no metallic flavour or taste.
37. The composition of claim 22 to 36, wherein the composition is formulated as a dry formulation,
25 a liquid formulation or a slurry or dispersion.
38. The composition of claim 37 having a formulation selected from the group consisting of tablet, capsule, liquid, drop, concentrate, powder, granule and combinations thereof.
- 30 39. The composition according to claim 22 to 38 further comprising a pharmaceutical acceptable carrier.
40. The composition according to claim 22 to 39, wherein the composition is a pharmaceutical composition selected from the group consisting of drugs, vaccines, personal care composition and
35 combinations thereof.

41. A non-therapeutic method of improving the stamina or low oxygen tolerance in a subject, comprising administering the recombinant protein of claim 1 to 21 or the composition of claim 22 to 40 to the subject in an amount effective to improve stamina or low oxygen tolerance in the
5 subject.

42. The method of claim 41 wherein the subject is performing a physical activity.

43. The method of claim 42 wherein the subject is performing mountaineering or mountain
10 climbing operating under low oxygen conditions.

44. A non-therapeutic method of altering the skin colour or texture in a subject comprising administering the recombinant protein of claim 1 to 21 or the composition of claim 22 to 40 into skin of the subject in an amount effective to alter the skin colour or texture.
15

45. The recombinant protein of claim 1 to 21 or the composition of claim 22 to 40 for the use in treating, ameliorating or preventing of a disease or deficiency in a subject, comprising administering the recombinant protein or the composition to the subject in an amount effective to ameliorate or prevent said disease or deficiency.
20

46. The recombinant protein composition of claim 45 wherein the disease or deficiency is iron deficiency or any disease resulting from iron deficiency.

47. The recombinant protein or composition of claim 45 to 46 wherein the recombinant protein is
25 a phytohemoglobin.

48. The recombinant protein or composition of claim 45 to 47 wherein the recombinant phytohemoglobin comprises a hexacoordinated haem group and has an amino acid sequence which is at least 86% identical to the phytohemoglobin comprised in SEQ ID NO: 1 or 5.
30

49. The recombinant protein or composition of claim 48 wherein the recombinant phytohemoglobin has an amino acid sequence which is at least 87%, such as at least 88%, such as at least 89%, such as at least 90%, such as at least 91%, such as at least 92%, such as at least 93%, such as at least 94%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as
35 at least 99% identical to the phytohemoglobin comprised in sequence SEQ ID NO: 1 or 5.

50. The recombinant protein or composition of claim 47 to 49 wherein the recombinant phytoalbumin is a nonsymbiotic haemoglobin of Class 1.

5 51. The recombinant protein or composition of claim 47 to 50 wherein the recombinant phytoalbumin is derived from a plant in Caryophyllidae, such as derived from spinach, quinoa.

52. The recombinant protein or composition of claim 47 to 51 wherein the recombinant phytoalbumin has an amino acid sequence comprised in or consisting of SEQ ID NO: 4.

10

53. The recombinant protein or composition of claim 45 to 52 wherein the recombinant protein comprises a hexacoordinated haem group and has an improved stability compared to a pentacoordinate haemoglobin.

15 54. The recombinant protein or composition of claim 45 to 53 wherein the recombinant protein has an auto-oxidation rate of less than 0.2, such as less than 0.15, for example in the range from 0.01 to 0.18, or 0.01 to 0.15, such as about 0.1 h⁻¹

55. The recombinant protein or composition of claim 45 to 54 wherein the recombinant protein
20 has no metallic taste or has neutral taste.

56. The recombinant protein or composition of claim 45 to 55 wherein recombinant protein has a melting temperature of above 68°C, such as in the range from 68 to 75°C; such as in the range from 70 to 76°C, such as above 76°C, such as above 77°C, such as above 78°C, such as above 79°C,
25 such as above 80°C, such as above 81°C, such as above 82°C, such as above 83°C, such as about 84°C.

57. The recombinant protein or composition of claim 47 wherein the recombinant phytoalbumin comprises a hexacoordinated haem group, and further comprises a CD-loop domain which has an
30 amino acid sequence which is at least 60% sequence identical to the CD-loop domain comprised in SEQ ID NO: 56, 60 or 111.

58. The recombinant protein or composition of claim 57 wherein the CD-loop domain has an amino acid sequence which is at least 90% identical to the CD-loop domain comprised in SEQ ID
35 NO: 56, 60 or 111.

59. The recombinant protein or composition of claim 57 to 58 wherein the CD-loop domain has an amino acid sequence which comprises the DC loop domain comprised in SEQ ID NO: 56, 60 or 111 having one or more, two or more, three or more, or four or more amino acid substitutions.
- 5 60. The recombinant protein or composition of claim 57 to 59 wherein the CD-loop domain has an amino acid sequence which comprises or consists of SEQ ID NO: 111, SEQ ID NO: 112, SEQ ID NO: 113, SEQ ID NO: 114, SEQ ID NO: 115 or SEQ ID NO: 116.
- 10 61. The recombinant protein or composition of claim 45 to 60 wherein the subject is a blood donor.
62. The recombinant protein or composition of claim 45 to 61 wherein the subject is a vegetarian, vegan and/or flexitarian.
- 15 63. The recombinant protein or composition of claim 45 to 62 wherein the subject is an animal, optionally a companion animal, poultry or livestock.
64. The recombinant protein or composition of claim 45 to 62 wherein the subject is a woman
- 20 optionally selected from the group of pregnant women, breastfeeding women, and women in reproductive age.
65. The recombinant protein or composition of claim 45 to 64 wherein the disease is selected from endometriosis, dysmenorrhea, menorrhagia, cancer, kidney disease, diabetes, obesity, metabolic
- 25 syndrome, digestive disorders, psychological disorders, genetic disorders, conditions associated with ageing, acute conditions, and/or infections.
66. The recombinant protein or composition of claim 65 wherein the digestive disorder is selected from celiac disease, inflammatory bowel disease, and/or peptic ulcers.
- 30 67. The recombinant protein or composition of claim 65 wherein the genetic disorder is selected from haemolytic anaemia, and/or autoimmune disease.
68. The recombinant protein or composition of claim 65 wherein the psychological disorder is
- 35 selected from bulimia and/or anorexia.

69. The recombinant protein or composition of claim 65 wherein the infection is selected from malaria and/or streptococcus infections.

5 70. The recombinant protein or composition of claim 65 wherein the disease is chronic.

71. The recombinant protein or composition of claim 65 wherein the acute condition is selected from trauma, gastrointestinal surgery, gastrointestinal bleeding, kidney failure, and/or acute toxicity.

10

72. The method or the recombinant protein or composition of claim 45 to 71 wherein the amount of recombinant protein administered to the subject is between 0,1 mg to 350 mg per kg body weight of the subject, such as between 0.25 mg to 100 mg, such as between 0,5 to 50 mg, such as between 0.75 mg to 10 mg, such as between 0,75 mg to 3 mg per kg body weight.

15

73. The method or the recombinant protein or composition of claim 72 wherein the amount of recombinant protein administered to the subject is between 0,75 mg to 3 mg per kg body weight.

20 74. The method or the recombinant protein or composition of claim 73 wherein the subject is a male.

75. The method or the recombinant protein or composition of claim 74 wherein the amount of recombinant protein administered to the subject is between 0,75 mg to 1,25, such as about 1 mg per kg body weight.

25

76. The method or the recombinant protein or composition of claim 73 wherein the subject is a female.

30 77. The method or the recombinant protein or composition of claim 76 wherein the amount of recombinant protein administered to the subject is between 1,3 mg to 1,5, such as about 1,4 mg per kg body weight.

78. A gene encoding the recombinant protein of claim 1 to 21.

79. The gene of claim 78 having a nucleotide sequence which is at least 50% identical to the nucleotide sequence comprised in any of SEQ ID NO: 148 to 202 encoding the corresponding recombinant protein of SEQ ID NO: 1 to 55.

5 80. A polynucleotide construct comprising the nucleotide sequence of claim 78 or 79, operably linked to a control sequence directing the transcription and/or translation of the gene in a host cell.

81. The polynucleotide construct of claim 80 wherein the control sequence is a promoter.
10

82. The polynucleotide construct of claim 81 wherein the promoter is an inducible promoter.

83. The polynucleotide construct of claim 82 wherein the promoter is at least 50 % identical to the promoter comprised in SEQ ID NO: 225 to 235.
15

84. The polynucleotide construct of claim 80 to 83 wherein the construct is an expression vector.

85. A genetically modified host cell expressing the gene or polynucleotide construct of claim 78 to 84 and the recombinant protein of claim 1 to 21.
20

86. The host cell of claim 85 wherein the cell is a eukaryotic cell, a bacterial cell, or an archaeal cell.

87. The host cell of claim 86 wherein the eukaryotic cell is a fungal cell, a plant cell, a mammalian
25 cell or an insect cell.

88. The genetically modified host cell of claims 87 wherein the plant cell is selected from the genus of *Arabidopsis*, *Lepidium*, *Nicotiana*, *Triticum*, *Hordeum*, *Oryza*, *Chenopodium*, *Beta*, or *Glycine*.

30 89. The genetically modified host cell of claims 88 wherein the plant cell is selected from the species of *Arabidopsis thaliana*, *Lepidium campestre*, *Nicotiana tabacum*, *Triticum aestivum*, *Hordeum vulgare*, *Oryza sativa*, *Chenopodium quinoa*, *Beta vulgaris*, or *Glycine max*.

90. The genetically modified host cell of claims 87 wherein the fungal cell is a yeast or a
35 filamentous fungus.

91. The genetically modified host cell of claims 90 wherein the filamentous fungus is selected from the genus of *Aspergillus*, *Trichoderma*, or *Rhizopus*.
- 5 92. The genetically modified host cell of claims 91 wherein the filamentous fungus is selected from the species *Aspergillus* sp, *Aspergillus oryzae*, *Trichoderma* sp, or *Rhizopus* sp.
93. The genetically modified host cell of claims 90 wherein the yeast is selected from the genus of *Pichia*, *Saccharomyces*, *Yarrowia*, *Kluyveromyces*, *Ashbya* or *Hansenula*.
- 10 94. The genetically modified host cell of claims 93 wherein the yeast is selected from the species of *P. pastoris*, *Pichia* sp., *S. cerevisiae*, *Yarrowia* sp., *Y. lipolytica*, *Kluyveromyces lactis*, *Saccharomyces carlsbergensis*, *Saccharomyces cerevisiae*, *Saccharomyces diastaticus*, *Saccharomyces douglasii*, *Saccharomyces kluyveri*, *Saccharomyces bensis*, *Saccharomyces*
- 15 *oviformis*, *Ashbya gossypii*, *H. polymorpha*, or *Hansenula* sp.
95. The genetically modified host cell of claims 86 wherein the bacterial cell is selected from the genus of *Escherichia*, *Bacillus*, *Brevibacterium*, *Burkholderia*, *Campylobacter*, *Corynebacterium*, *Serratia*, *Lactobacillus*, *Lactococcus*, *Acinetobacter*, *Acetobacter* or *Pseudomonas*.
- 20 96. The genetically modified host cell of claims 95 wherein the bacterial cell is selected from the species *E. coli*, *C. glutamicum*, *B. subtilis*, *S. marcescens*, *P. putida*, *P. aeruginosa* and/or *P. mutabilis*.
- 25 97. The genetically modified host cell of claims 86 wherein the archaeal cell is an algae.
98. A cell culture comprising the host cell of claim 85 to 97 and a growth medium.
99. A method for producing the recombinant protein of claim 1 to 21 comprising:
- 30 c. culturing the cell culture of claim 98 at conditions allowing the cell to produce the recombinant protein; and
- d. optionally recovering and/or isolating the recombinant protein.
100. The method of claim 99, further comprising one or more elements selected from:
- 35 h) culturing the cell culture in a nutrient medium;

- i) culturing the cell culture under aerobic or anaerobic conditions
- j) culturing the cell culture under agitation;
- k) culturing the cell culture at a temperature of between 10 to 50 °C;
- l) culturing the cell culture at a pH of between 3-9;
- 5 m) culturing the cell culture for between 10 hours to 30 days; and
- n) culturing the cell culture under fed-batch, repeated fed-batch, continuous, or semi-continuous conditions.

101. The method of claim 100 wherein the nutrient medium has the properties of:

- 10 c) pH 5-8,
- d) Temperature 17-38 °C,

102. The method of claim 99 to 101 wherein the recovery and/or isolation step comprises separating a liquid phase of the cell or cell culture from a solid phase of the cell or cell culture to
15 obtain a supernatant comprising the recombinant protein and/or subjecting the supernatant to one or more steps selected from:

- a) separating the supernatant from the solid phase of the cell culture, such as by filtration or gravity separation;
- b) contacting the supernatant with one or more adsorbent resins to obtain at least a portion
20 of the produced recombinant protein;
- c) contacting the supernatant with one or more ion exchange or reversed-phase chromatography columns or filtration or ultrafiltration devices to obtain at least a portion of the recombinant protein;
- d) extracting the recombinant protein; and/or
- 25 e) precipitating the recombinant protein by crystallization or evaporating the solvent of the liquid phase; and optionally isolating the recombinant protein by filtration or gravity separation;

thereby recovering and/or isolating the recombinant protein.

- 30 103. A composition comprising the cell culture of claim 98.

104. The composition of claim 103 wherein the composition is edible, and the cell culture is in the form of an edible biomass and mycelium.

