TITLE: ALBUMIN VARIANTS AND CONJUGATES

Reference to sequence listing

This application contains a Sequence Listing in computer readable form. The computer readable form is incorporated herein by reference.

FIELD OF THE INVENTION

The present invention relates to conjugation competent albumins and albumin-related polypeptides, and their conjugates with at least one moiety, and to polynucleotides encoding them.

BACKGROUND OF THE INVENTION

Serum albumins provide valuable scaffolds to which bioactive molecules may be fused, either through genetic fusions or chemical fusions to improve the properties of the fused molecule(s) (Leger, R. *et al.* (2004). Bioorg Med Chem Lett 14(17): 4395-8; Thibaudeau, K., *et al.* (2005). Bioconjug Chem 16(4): 1000-8; Balan, V. *et al.* (2006). Antivir Ther 11(1): 35-45; EP 0 413 622; WO 90/13653; EP 1 681 304; WO 1997/024445; WO 01/79271). Albumins and albumin particles are also important for carrying and delivering drugs and prodrugs to their sites of action (Kratz (2008) Journal of Controlled Release, 132 (3), p.171-183). Fusion and particle technologies offer improved dosing regimes due to improved pharmacokinetic properties, such as half-life extension, and may improve bioavailability and protect the fused bioactive molecule from inactivation.

The biochemistry, genetics and medical applications of albumins are well characterized ("All about Albumin", T. Peters Jr., Academic Press N. Y., and see for example, the world wide web at http://www.albumin.org/). Human serum albumin (HSA, also referred to as HA) is the most abundant protein in human plasma at ~60g/L. The sequence of HSA is provided in SEQ ID NO: 1. Natural variants of HSA occur and a list of know polymorphisms is given in Minchiotti *et al.* (2008). Hum Mutat **29**(8): 1007-16 and see for example, the world wide web at http://www.uniprot.org/uniprot/P02768.

The production and purification of recombinant human albumins are well established (WO 95/23857; WO 00/44772; WO 2006/066595; EP 0 305 216; Sleep *et al.* 1990 Biotechnology (N Y). 1990 Jan; 8(1):42-6)) and include recombinant human albumin for pharmaceutical applications (Bosse *et al.* (2005). J Clin Pharmacol 45(1): 57-67). The three-dimensional structure of HSA has been elucidated by X-ray crystallography (Carter *et al.* (1989). Science 244(4909): 1195-8; Sugio *et al.* (1999). Protein Eng 12(6): 439-46). The HSA polypeptide chain has 35 cysteine residues, which form 17 disulphide bonds and one unpaired (free) cysteine at position 34 of the mature protein (SEQ ID NO: 1). Cysteine-34 has been used for

conjugation of molecules to albumin (Leger *et al.* (2004) Bioorg Med Chem Lett 14(17): 4395-8; Thibaudeau *et al.* (2005). Bioconjug Chem 16(4): 1000-8), and provides a precise, well defined site for conjugation. However, conjugation at cysteine-34 provides only one site for attachment of a single moiety thus there is no choice of conjugation site. Also, the provision of a single conjugation sites means that only one moiety can be conjugated to each albumin molecule. What is required is an albumin molecule which provides one or more alternative attachment sites.

SUMMARY OF THE INVENTION

Based on an analysis of the three-dimensional structure of a human serum albumin (HSA), conserved residues within albumin polypeptides and natural polymorphisms thereof, the inventors have designed variant polypeptides (muteins) of albumin which have one or more conjugation competent cysteine residues. The term 'thio-albumin' is used herein to describe an albumin variant which comprises one or more unpaired cysteine residues, particularly an albumin variant in which one or more of the unpaired cysteine residues does not occur in a naturally occurring variant of an albumin. Thus a thio-albumin is a 'conjugation competent albumin'. A thio-albumin may be referred to as a 'cysteine variant of an albumin'.

Throughout this specification, the term 'albumin' includes naturally occurring albumin, albumin-related proteins and variants thereof such as natural and engineered variants. Variants include polymorphisms, fragments such as domains and sub-domains, fragments and/or fusion proteins. The albumin may have at least 40, 50, 60, 70, 80, 90, 95, 96, 97, 98, 99 % similarity or identity to SEQ ID NO: 1. Thus a thio-albumin of the invention may be a derivative of, or be based on, any of such albumin.

The unpaired cysteine residues may be provided by insertion, deletion, substitution, addition or extension of an albumin sequence.

The invention also relates to a conjugate comprising at least one, for example 2, 3, 4, 5 or 6, conjugation partners such as bioactive compounds and a polypeptide according to the invention.

The invention also provides a method for designing conjugation-competent albumins.

BRIEF DESCRIPTION OF DRAWINGS

Fig. 1. is a table showing criteria used to select sites in human serum albumin (SEQ ID NO: 1) for amino acid substitutions, insertions and deletions for the generation of conjugation competent cysteines.

Fig. 2. is an alignment of the amino acid sequence of human serum albumin (SEQ ID No. 1 = Human – P02768.pro) with albumins from fifteen other mammalian species. 'Majority' shows the consensus sequence. '+ Majority' shows the relative homology between all sixteen sequences in bar chart form, where the height of the bar indicates the relative homology at 20, 40, 60, 80 and 100%. The protein sequences include the leader sequence. Hum.pro (SEQ ID NO: 109); Bov.pro (SEQ ID NO: 94); Cat.pro (SEQ ID NO: 95); Dog.pro (SEQ ID NO: 96); Don.pro (SEQ ID NO: 97); Ger.pro (SEQ ID NO: 98); Goa.pro (SEQ ID NO: 99); Gui.pro (SEQ ID NO: 100); Hor.pro (SEQ ID NO: 101); Mac.pro (SEQ ID NO: 103); Mou.pro (SEQ ID NO: 104); Pig.pro (SEQ ID NO: 105); Rat.pro (SEQ ID NO: 107); Rab.pro (SEQ ID NO: 106); and She.pro (SEQ ID NO: 108).

Fig. 3. Is an alignment of the amino acid sequence of human serum albumin (P02768.pro = SEQ ID NO: 109) with albumins from thirty two other species, some of which are mammalian. 'Majority' shows the consensus sequence. '+ Majority' shows the relative homology between all thirty three sequences in bar chart form, where the height of the bar indicates the relative homology at 20, 40, 60, 80 and 100%. The protein sequences include the leader sequence. P02768.pro: human (SEQ ID NO: 109); P02769.pro: bovine(SEQ ID NO: 94); P49064.pro: cat(SEQ ID NO: 95); P49822.pro: dog(SEQ ID NO: 96); Q5XLE4.pro: donkey(SEQ ID NO: 97); JC5838.pro: gerbil(SEQ ID NO: 98); ACF10391,1.pro: goat fragment (SEQ ID NO: 99); AAQ20088.pro: guinea pig(SEQ ID NO: 100); P35747.pro: horse(SEQ ID NO: 101); Q28522.pro: macaque(SEQ ID NO: 103); P07724.pro: mouse(SEQ ID NO: 104); P08835.pro: pig(SEQ ID NO: 105); P02770.pro: rat(SEQ ID NO: 107); P49065.pro: rabbit(SEQ ID NO: 106); Q28522.pro: Rhesus monkey(SEQ ID NO: 103); P14639.pro: sheep(SEQ ID NO: 108); NP_001127106.pro: orangutan(SEQ ID NO: 110); P19121.pro: chicken(SEQ ID NO: 111); P01012.pro: chicken ovalbumin (SEQ ID NO: 112); O73860.pro: turkey ovalbumin (SEQ ID NO: 113); AAC63407.pro: sea lamprey(SEQ ID NO: 114); Q91274.pro: sea lamprey(SEQ ID NO: 115); P21847.pro: bullfrog(SEQ ID NO: 116); AAD09358.pro: Rana shqiperica(SEQ ID NO: 117); ABXL68.pro: Xenopus(SEQ ID NO: 118); NP_001004887.pro: Xenopus (SEQ ID NO: 119); AAL56646.pro: Spotted Salamander(SEQ ID NO: 120); Q03156.pro: Atlantic salmon(SEQ ID NO: 138); P21848.pro: Atlantic salmon(SEQ ID NO: 121); AAM46104.pro: Sphenodon punctatus (SEQ ID NO: 122); P83517.pro: Australian lungfish (SEQ ID NO: 123); S59517.pro: monocled cobra (SEQ ID NO: 124); AAL08579.pro: Schistosoma mansoni (SEQ ID NO: 125).

Fig. 4. is a Venn diagram showing the classes of and relationship between twenty amino acids.

Figs. 5A, 5B, 5C and 5D are tables showing groups of preferred sites in human serum albumin (SEQ ID No. 1) for amino acid substitutions, insertions and deletions for the generation of one or more conjugation competent cysteine.

Figs. 6A and 6B are tables showing groups of preferred sites in human serum albumin (SEQ ID No. 1) for disruption of one or more disulphide bonds for the generation of one or more conjugation competent cysteines.

Fig. 7. is a map of plasmid pDB2244.

Fig. 8 is a map of plasmid pDB2243.

Fig. 9. is a map of plasmid pDB2713.

Fig. 10 is a table showing preferred sites for conjugation grouped according to their relative position on a folded albumin of SEQ ID No. 1.

DETAILED DESCRIPTION OF THE INVENTION

A first aspect of the invention provides a method for designing and/or preparing variant albumins comprising one or more conjugation competent cysteine residues. Therefore, the polypeptide may be considered to be conjugation-competent. Such an albumin may be referred to as a 'thio-albumin' or as a 'cysteine varant' of an albumin. The term 'conjugation competent cysteine' includes a cysteine which has a thiol which is not disulphide bonded to another cysteine and which is, preferably, not blocked from conjugating to another molecule (which may be referred to as a 'conjugation partner') due to unfavourable steric hindrances. That is, preferably the location of the cysteine within or on a folded polypeptide is such that it is available for conjugation.

A number of selection criteria may or may not be used alone or in any combination in order to identify suitable sites for introduction of a conjugation competent cysteine residue. Therefore, the invention provides a method and/or rules for *a priori* identification of sites of an amino acid sequence of albumin at which a conjugation competent cysteine may be introduced. Such sites may be referred to as 'candidate residues'. The albumin sequence on which the variant albumin is based may be SEQ ID No. 1 or any other albumin. For example, the variant albumin may be based on an albumin which does or does not have a cysteine at position 34 ot the amino acid sequence, or an equivalent position. Cysteine residues may or may not be introduced by one or more of substitution, insertion, deletion, extension and addition. Sites may or may not be selected with reference to a 3-dimensional structure of an albumin or variant thereof. The following criteria may or may not be used to select suitable sites:

(a) Solvent Accessible Surface Area ("Surface accessibility" (% SASA)).

Preferably the surface accessibility is high. For example, preferably the surface accessibility is at least 60%, more preferably, from 60, 65, 70, 75, 80, 85, 90, 95, 96, 97, 98 or 99 % to 65, 70, 75, 80, 85, 90, 95, 96, 97, 98, 99 or 100%. % SASA may be determined as a 'raw score' using the methods described herein or may be calculated relative to the score of the residue which has the maximum surface accessibility in the protein. For example, the albumin of HSA 1AO6 has a maximum surface accessibility of 229.0 and this is the highest scoring residue in HSA. A higher surface accessibility indicates that the residue is on the surface of the protein and is therefore available for binding. Such accessibility may be calculated using a method as described herein.

(b) Presence or absence of crystallographic B-factor(s).

B-factor indicates relative flexibility of an amino acid residue within a 3-dimensional structure. Preferably the B-factor is from at least 30, 40, 50, 60, 70, 80 or 90 % to at least 40, 50, 60, 70, 80, 90 or 100 % which may or may not be relative to the maximal B-factor score of any amino acid residue within the molecule. For HSA (*e.g.* 1AO6), preferably the B-factor score is high, for example from at least 30, 40, 50, 60, 70, 80, 90, or 100 to at least 40, 50, 60, 70, 80, 90, 100 or 106 (for example using the B-factor scoring system described herein). Alternatively the B-factor score may be less than or equal to 100, 90, 80, 70, 60, 50, 40, 30, 20, or 10 %, as described herein.

The B-Factor (root mean square fluctuations) of the C-alpha carbon atoms during the last nanosecond of the simulation may be calculated using the Gromacs tool "g_rmsf", version 3.3, based on D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. C. Berendsen: GROMACS: Fast, Flexible and Free, J. Comp. Chem. 26 pp. 1701-1718 (2005).

(c) Presence or absence of Secondary Structure (SS).

The candidate residue may or may not be located within secondary structure for example H (Helix), B (isolated beta bridge) or E (Extended sheet). Location of the residue outside of secondary structure indicates that the residue is less likely to be important to secondary structure and/or is more likely to be available for binding than a residue located within secondary structure.

(d) Relative homology with other albumins.

Within a given protein sequence, an amino acid residue showing high homology with other similar sequences is likely to indicate such a residue or region is likely to be important to the structure and/or function of the protein. Therefore it is preferred that a candidate residue shows a homology of less than 100 % relative to alignment of the albumin in which the residue is located with known albumins (*e.g.* mammalian albumins such as those shown in Fig. 2 or a combination of mammalian and non-mammalian albumins such as those shown in Fig. 3). A homology of less than 100, 98, 96, 95, 94, 92, 90, 85, 80, 75, 70, 65, 60, 55, 50, 45, 40, 35, 30, 25, 20, 15, 10, 5 is preferred. Homology can be determined using algorithms known in the art such as Clustal, *e.g.* Clustal W (Thompson *et al.* (1994). Nucleic Acids Res 22(22): 4673-80) or Clustal V (Higgins, D. G. and P. M. Sharp (1989). "Fast and sensitive multiple sequence alignments on a microcomputer." Comput Appl Biosci 5(2): 151-3.). Lower homology indicates that the residue is not particularly important or critical to the structure and/or function of the protein. Preferably the homology is determined with reference to the sixteen mammalian albumins of Fig. 2 or the thirty three mammalian and non-mammalian albumins of Fig. 3.

(e) Presence or absence of adjacent conserved residues.

Within an amino acid sequence, each residue has one or two adjacent residues. If a candidate residue is immediately adjacent one or more residues having a low homology, relative to known albumins, this indicates that the candidate residue is unlikely to be particularly important or critical to the structure and/or function of the protein. This is because the candidate residue is likely to be located within a relatively unconserved region of the protein. It is therefore preferred that the candidate residue is not adjacent a residue which has 100 % homology relative to alignment of the albumin with known albumins. Homology may be determined as described herein. The candidate residue may be adjacent two residues (*i.e.* one residue C-terminal relative to the candidate residue and one residue N-terminal relative to the candidate residue) which each have 100 % homology relative to alignment of the albumin with known albumins (*e.g.* Fig. 2 or Fig. 3). It is preferred that the candidate residue is adjacent one or two residues having a homology of less than 100, 98, 96, 95, 94, 92, 90, 85, 80, 75, 70, 65, 60, 55, 50, 45, 40, 35, 30, 25, 20, 15, 10, 5 – these levels of homology may be referred to as 'thresholds'. Homology may be determined as described herein. Taking into account the homology threshold, the location of an amino acid relative to a conserved region may be quantified, for example by scoring an amino acid which is not adjacent any amino acid exceeding the homology threshold as 0, scoring an amino acid which is adjacent one amino acid exceeding the threshold as 1 and scoring an amino acid which adjacent two amino acids exceeding the threshold as 2.

(f) Evidence for polymorphism(s).

A polymorphism is a genetic variation, a polymorphism may or may not cause a phenotypic change to the resultant protein. Preferably the candidate residue is not at a position for which a polymorphism causing a phenotypic change is known. More preferably, the candidate residue is not at a position for which a polymorphism causes, or is known to cause, thermal instability. Polymorphisms known for HSA (SEQ ID No. 1) are detailed in Fig. 1 and are also discussed in Minchiotti *et al.* (2008). Hum Mutat 29(8): 1007-16 and see for example, the world wide web at http://www.uniprot.org/uniprot/P02768. The presence, absence and/or effect of a polymorphism may be quantified, for example by scoring a known polymorphism that has no phenotypic change as 0, scoring a polymorphism where a phenotypic change is known (but not known to cause thermal instability) as 1 and scoring a polymorphism which is known to cause thermal instability as 2.

(g) Relative conservation of candidate amino acid and cysteine.

Amino acids fall into various well known classes. Therefore, some amino acids are more closely related than others. The introduced cysteine residue may or may not maintain a

relatively high level of conservation with the candidate amino acid. Fig. 4 is a Venn diagram which provides one system by which conservation level can be quantified. The scoring system of Fig. 4 uses a scale of 0 to 5 in which substitutions of high conservation have a score of 0, substitutions of low conservation have a score of 5 and substitutions of intermediate conservation have a score of 1, 2, 3 or 4. Preferably substitution of the candidate residue is not an unconserved substitution, that is preferably (using the scoring system of Fig. 4) the candidate residue does not have conservation score (relative to cysteine) of 5. More preferably the candidate residue has a higher conservation relative to cysteine (*e.g.* a score of 4, 3, 2 and, more preferably, 1). The scoring system is described in the section entitled 'Conservative Substitution' (below).

(h) Expression level.

The thio-albumin may or may not be capable of being expressed at a level of at least 10, 20, 30, 40, 50, 60, 70, 80, 90 or 100% relative to the expression of an unmodified albumin (such as SEQ ID No. 1) from a suitable expression system, such as yeast (*e.g. Saccharomyces, e.g. S. cerevisiae*) or an Aspergillus. Relative expression levels can be determined, for example, by expression of the protein followed by quantification by SDS-PAGE, HPLC or Western Blotting.

(i) Conjugation competence.

The thio-albumin may or may not have a high level of conjugation competence, for example at least 50, 60, 70, 80, 90, 95 or 100 % relative to the conjugation competence of an albumin consisting of SEQ ID No. 1 having only one conjugation competent cysteine at Cys-34. Conjugation competence may be determined relative to any conjugatable molecule (conjugation partner) of interest, for example a bioactive molecule or a fluorescent dye. Determination may be through mass spectrometry analysis or quantification of the activity of the bioactive compound such as its fluorescence. An advantage of a thio-albumin having a high conjugation competence is that it may allow efficient conjugation of molecules to the thio-albumin. Conjugation competence may be measured with respect to time. Favoured thio-albumins may be (a) those which achieve maximal conjugation quickly or (b) slowly.

(j) Activity of conjugated compound.

The thio-albumin of the invention may be conjugated to a compound (conjugation partner), for example a bioactive compound, such that the compound has a high level of activity relative to its activity in an unconjugated state. Preferably, the conjugated compound shows at least 1, 10, 20, 40, 50, 60, 70, 80, 80 and most preferably 100 % of its activity relative to its unconjugated state. An advantage of a conjugated compound with a high level of

activity is that it reduces the quantity of conjugated compound required to achieve a desire effect, *e.g.* a desired therapeutic effect.

(k) Receptor binding capacity of albumin

The conjugated- and/or non-conjugated thio-albumin may or may not have a receptor binding activity of at least 10, 20, 30, 40, 50, 60, 70, 80, 90 or 100 % of the receptor binding activity of human serum albumin (SEQ ID No. 1). Alternatively, the conjugated- and/or non-conjugated thio-albumin may or may not have a lower receptor binding activity for example at most 0, 10, 20, 30, 40, 50, 60, 70, 80 or 90 % than human serum albumin. Receptor binding activity may be determined by assay, such as in relation to binding to FcRn.

Fig. 1. shows the scores of each amino acid residue of HSA (SEQ ID No. 1) for each of parameters (a) to (g). For clarity, *in vivo*, HSA is initially produced as a 609 amino acid protein in which the first twenty four amino acids are a leader sequence. The leader sequence is cleaved off to generate a 585 amino acid mature protein. Throughout this specification, the mature protein is referred to as SEQ ID No. 1. The structure of HSA model A106 disregards the first four residues and the last three residues of SEQ ID No. 1 because these are unresolved in the 3D model. Therefore, residue 1 of model A106 is equivalent to residue 5 of SEQ ID No. 1. Throughout this specification, all residues are cited with reference to SEQ ID No. 1, unless stated otherwise. The immature sequence of HSA (*i.e.* HSA with its natural C-terminal leader sequence) is provided in SEQ ID No. 102.

The column labels of Fig. 1. are detailed below:

Position in 1AO6: Refers to the amino acid position in the crystal structure of human serum albumin from the RCSD Protein Databank (PDB, see, for example, the world wide web at http://www.rcsb.org/pdb/) with the entry with PDB identity 1AO6 or 1ao6 (Sugio, S., A. Kashima, *et al.* (1999). Protein Eng **12**(6): 439-46). Note that compared to the mature HSA sequence (SEQ ID NO:1), the 1AO6 structure starts at residue 5S (with the first 4 amino acids absent from the structure) and finishes at 582A of SEQ ID NO:1 (with the last 3 amino acids absent from the structure). The amino acid positions used herein to describe positions to alter to generate conjugation competent cysteines are referring to the positions in SEQ ID No1, not 1ao6.

Position in Mature HSA: The amino acid position in SEQ ID NO:1 taken from the 585 residue secreted form of HSA, National Center for Biotechnology Information, ACCESSION: 1AO6_A VERSION GI:3212456 (24-SEP-2008), Chain A, Crystal Structure Of Human Serum Albumin. (Sugio *et al.* (1999). Protein Eng 12(6): 439-46).

Align 2 (Var. Sps. V): The homology level for an alignment of various albumin family proteins with HSA (SEQ ID NO: 1), identified as P02768 compiled using MegAlign program (DNASTAR, Lasergene, version 8.0.2) based on Clustal **V**; six levels of homology are determined with the highest =100%, decreasing in 20% increments, to the lowest =0% (Fig. 3).

Polymorph.: This identifies whether or not a polymorphism is known at the amino acid residue. Single amino acid polymorphisms of human serum albumin (SEQ ID NO: 1) were taken from Minchiotti *et al.* (2008). Hum Mutat 29(8): 1007-16 and the world wide web at http://www.uniprot.org/uniprot/P02768, with amino acid positions taken from the unprocessed form of human serum albumin containing the 24 amino acid secretory leader sequence, and described using the standard one letter amino acid code (*e.g.* D25V refers to an aspartic acid being changed to a valine at position 1 in SEQ ID NO: 1).

Phenotype Change: A score representing the 'severity' of phenotypic change derived from the sources of known phenotypic changes ('Polymorph.', referenced above) where; 0 = no known phenotypic change, 1 = a phenotypic change has been described for any of the mutations at this position compared to the residue in SEQ ID NO: 1, excluding a change resulting in decreased thermal stability, 2 = a mutation at this position in SEQ ID NO: 1 is described as causing reduced thermal stability.

Conserved Mutation *vs.* Cysteine: A score referring to how well conserved the amino acid is compared to cysteine, as derived from Figure 4 (described herein), and ranging from 1 to 5 for mutations to cysteine. A score of 1 is assigned to the most conservative changes possible (*e.g.* alanine to cysteine), and ranging to a score of 5 for the lowest of conservation compared to a cysteine (*e.g.* histidine to cysteine).

Although the selection criteria can be used in any desired combination, four preferred groups of selection criteria (A, B, C, D) are described, by way of example only, below. Of these (A) and (B) may also be referred to as Selection Groups 1 and 2 (respectively):

(A) A particularly preferred embodiment of the first aspect of the invention provides a method for designing and/or preparing a thio-albumin, the method comprising:

providing a three-dimensional model comprising at least one instance of an albumin sequence (preferably the three dimensional model relates to an amino acid sequence of an albumin and most preferably the amino acid sequence of the albumin sequence is also provided, the amino acid sequence may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 amino acids which are not resolved at the C- and/or N-terminus of the three dimensional model, preferably the amino acid sequence is 'full length', i.e. the mature amino acid sequence of the albu-

min);

selecting a candidate amino acid residue in the albumin sequence which corresponds to the first, second, third, fourth or fifth residue relative to the N- or C- terminus of the albumin sequence (of the model or of the amino acid sequence) or which (preferably in relation to the three dimensional model) fulfils the following conditions: not present within secondary structure; surface accessibility (SASA) of at least 90 %; B-factor score of at least 60; less than 80% homology to known mammalian albumins (*e.g.* Fig. 2); no adjacent residues with 100 % homology to known mammalian albumins (*e.g.* Fig. 2); no polymorphism with a known phenotypic change; and no unconserved amino acid change to cysteine with a score of 5 or above;

substituting one or more of the selected amino acid residues with cysteine or inserting cysteine at the N-side or C-side of the selected residue,

optionally making one or more additional alterations to the albumin sequence where each alteration is an amino acid deletion, substitution, extension, addition or insertion; and

optionally preparing a polypeptide having the required amino acid sequence.

With reference to model 1A06 and SEQ ID No. 1, candidate residues identified by selection criteria (A) include L585, D1, A2, D562, A364, A504, E505, T79 and E86 (in descending order of solvent accessibility) and are also shown in Fig. 5A.

(B) Another preferred embodiment of the first aspect of the invention provides a method for designing and/or preparing a thio-albumin, the method comprising:

providing a three-dimensional model comprising at least one instance of an albumin sequence (preferably the three dimensional model relates to an amino acid sequence of an albumin and most preferably the amino acid sequence of the albumin sequence is also provided, the amino acid sequence may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 amino acids which are not resolved at the C- and/or N-terminus of the three dimensional model, preferably the amino acid sequence is 'full length', i.e. the mature amino acid sequence of the albumin);

selecting a candidate amino acid residue in the albumin sequence (of the model or of the amino acid sequence) which (preferably in relation to the three dimensional model) fulfils the following conditions: present within secondary structure; surface accessibility of at least 90 %; B-factor score of at least 40; less than 80% homology to known mammalian albumins (*e.g.* Fig. 2); no adjacent residues with 100 % homology to known mammalian albumins (*e.g.* Fig. 2); no polymorphism with a known phenotypic change; and no unconserved amino acid change to cysteine with a score of 5 or above;

substituting one or more of the selected amino acid residues with cysteine or inserting cysteine at the N-side or C-side of the selected residue,

optionally making one or more additional alterations to the albumin sequence where each alteration is an amino acid deletion, substitution, extension, addition or insertion; and

optionally preparing a polypeptide having the required amino acid sequence.

With reference to model 1A06 and SEQ ID No. 1, candidate residues identified by selection criteria (B) include D129, D549, A581, D121, E82, S270, Q397 and A578 (in descending order of solvent accessibility) and are also shown in Fig. 5B.

(C) Another preferred embodiment of the first aspect of the invention provides a method for designing and/or preparing a thio-albumin, the method comprising:

providing a three-dimensional model comprising at least one instance of an albumin sequence (preferably the three dimensional model relates to an amino acid sequence of an albumin and most preferably the amino acid sequence of the albumin sequence is also provided, the amino acid sequence may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 amino acids which are not resolved at the C- and/or N-terminus of the three dimensional model, preferably the amino acid sequence is 'full length', i.e. the mature amino acid sequence of the albumin);

selecting a candidate amino acid residue in the albumin sequence (of the model or of the amino acid sequence) which (preferably in relation to the three dimensional model) fulfils the following conditions: not present within secondary structure; surface accessibility of at least 80 %; B-factor score of at least 50; less than 100% homology to known mammalian albumins (*e.g.* Fig. 2); less than 80% homology to the various albumins aligned in Fig. 3; no polymorphism known to cause thermal instability; and no unconserved amino acid change to cysteine with a score of 4 or above;

substituting one or more of the selected amino acid residues with cysteine or inserting cysteine at the N-side or C-side of the selected residue;

optionally making one or more additional alterations to the albumin sequence where each alteration is an amino acid deletion, substitution, extension, addition or insertion; and

optionally preparing a polypeptide having the required amino acid sequence.

With reference to model 1A06 and SEQ ID No. 1, candidate residues identified by selection criteria (C) are shown in Fig. 5C.

(D) Another preferred embodiment of the first aspect of the invention provides a method for designing and/or preparing a thio-albumin, the method comprising:

providing a three-dimensional model comprising at least one instance of an albumin sequence (preferably the three dimensional model relates to an amino acid sequence of an albumin and most preferably the amino acid sequence of the albumin sequence is also provided, the amino acid sequence may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 amino acids which are not resolved at the C- and/or N-terminus of the three dimensional model);

selecting a candidate amino acid residue in the albumin sequence which (preferably in relation to the three dimensional model) fulfils the following conditions: present within secondary structure; surface accessibility of at least 80 %; B-factor score of at least 30; less than 100% homology to known mammalian albumins (*e.g.* Fig. 2); less than 80% homology to the various albumins aligned in Fig. 3; no polymorphism known to cause thermal instability; and no unconserved amino acid change to cysteine with a score of 4 or above;

substituting one or more of the selected amino acid residues with cysteine or inserting cysteine at the N-side or C-side of the selected residue,

optionally making one or more additional alterations to the albumin sequence where each alteration is an amino acid deletion, substitution, extension, addition or insertion; and

optionally preparing a polypeptide having the required amino acid sequence.

With reference to model 1A06 and SEQ ID No. 1, candidate residues identified by selection criteria (D) are shown in Fig. 5D.

Since Figures 5A, 5B, 5C and 5D are selections from Figure 1, the column headings are the same.

A candidate residue may be one or more of the cysteine residues involved in disulphide bonding present in the albumin molecule (in the case of HSA, SEQ ID No. 1, there are 17 disulphide bonds and therefore 34 cysteines involved in disulphide bonding). Two cysteines which are linked by a disulphide bond may be referred to as 'counterparts'. In order to generate a conjugation competent cysteine, the candidate residue may be deleted or may be substituted with a different amino acid, particularly Ser, Thr, Val or Ala in order to create a free thiol at the partner cysteine. The 34 cysteine residues of SEQ ID No. 1 which are involved in disulphide bonding are C53, C62, C75, C91, C90, C101, C124, C169, C168, C177, C200, C246, C245, C253, C265, C279, C278, C289, C316, C361, C360, C369, C392, C438, C437, C448, C461, C477, C476, C487, C514, C559, C558 and C567. In relation to the invention, some of these 34 candidate residues are more favoured than others.

Cysteine residues were visually inspected using the PyMOL software (Warren L. DeLano "The PyMOL Molecular Graphics System." DeLano Scientific LLC, San Carlos, CA, USA), and the cysteines in the disulphide bonds were divided into 3

categories:

- those that can be replaced, for example, by serine, leaving its counterpart cysteine as a free thiol that has a high probability of being a conjugation site. These correspond to C75, C91, C124, C168, C169, C316, C360, C361, C567, C558.
- those that can be replaced by serine, leaving its counterpart as a free thiol that has a medium probability of being a conjugation site. These correspond to C90, C101, C177, C265, C279, C278, C289, C369, C392, C438, C476, C487, C514, C559.
- those that can be replaced by serine, leaving its counterpart as a free thiol that has a low probability of being a conjugation site. These correspond to C53, C62, C200, C246, C245, C253, C437, C448, C461, C477.

The judgment is based on surface accessibility and the orientation of the C-alpha – C-beta bond of the potential free thiol relative to the folded polypeptide. Using this judgment each of the cysteine residues of HSA were given a modification score and ranked as high, medium or low.

Figure 6A, provides a list of all the cysteine residues which have a high modification score (right hand column), indicating that modification of a cysteine residue at this position would result in its counterpart cysteine providing a free thiol that has a high probability of being suitable for use as a conjugation site.

Figure 6B, provides a list of the counterpart cysteines that that, when unpaired (thus providing a free thiol), have a high probability of being suitable for use as a conjugation site.

The column labels for Fig. 6 are the same for those described for Fig. 1 with the addition of:

Modification Score: defined as 'high', 'medium' or 'low' as described herein.

Disulphide Information: summarises disulphide pairing in SEQ ID No. 1.

(Polymorp.) Phenotype Change: summarises the columns labelled 'Polymorph.' And 'Phenotype Change' in Fig. 1.

Preferred cysteine residues for modification could be further selected based on the other information provided in Figure 6A, such as assigned secondary structure, cysteine residues with no adjacent conserved residues (100% amongst mammalian albumins (aligned by Clustal W), no known polymorphisms causing phenotypic changes.

Alternatively cysteine residues for modification could be selected by examining the environment of the cysteine residue (containing a free thiol group) generated by modification of its counterpart cysteine residue provided in Figure 6A, characteristics such as high

tide sequence is increased relative to the polypeptide prior to the substitution, insertion, deletion and addition events. Within (a) to (d), above, the residues all of the residues are preferred. However, within each of (a), (b), (c) and (d), the residues are listed in order of decreasing preference.

A thio-albumin may or may not include a polypeptide where one or more naturally occurring free-thiol group(s), such as cysteine-34 in HSA (SEQ ID No. 1), is modified to an amino acid which is not cysteine. For example, cysteine may or may not be replaced by an amino acid which has a relatively high conservation score (*e.g.* 1, 2 or 3 as calculated according to Fig. 4) such as alanine or serine. A thio-albumin may or may not include a polypeptide where one or more naturally occurring free-thiol group(s), such as cysteine-34 in HSA (SEQ ID No. 1) are present.

As detailed herein, the invention may be achieved by introducing cysteine residues by one or more of extension, addition, insertion, substitution or deletion.

An *addition* may be made by extension and/or insertion.

For example, one or more conjugation competent cysteines may or may not be created in an albumin by *extension*; *e.g.* by adding an extra cysteine residue to the N-terminus or C-terminus of the molecule, which may or may not be added as a single cysteine residue, or as a longer polypeptide which contains one or more conjugation competent cysteines. The cysteine residue(s) may be added immediately adjacent the N- or C-terminus of the albumin. Alternatively, there may be one or more other amino acid residues located between the N- and/or C-terminus of the albumin and the cysteine residue(s). When two or more cysteine residues are added, some or all of the added cysteines may be separated from each other by one or more other amino acids, for example by from 1 to 50 amino acids, such as from 1, 10, 20, 30, or 40 amino acids to from 10, 20, 30, 40, or 50 amino acids. A preferred N-terminal extension is the addition of Cys immediately adjacent the N-terminal of a mature albumin (*i.e.* albumin cleaved from its leader sequence). For example, for an albumin comprising or consisting of SEQ ID No. 1, Cys is preferably immediately N-terminal to the first Asp (D1). Such an albumin may be referred to as 'Cys-albumin', *e.g.* 'Cys-HSA' (where HSA is Human Serum Albumin). Other preferred N-terminal extensions of albumins such as SEQ ID No. 1 include Cys-Ala-albumin such as Cys-Ala-HSA. A preferred C-terminal extension is the addition of Cys immediately adjacent the C-terminal of an albumin, such as a mature albumin. For example, for an albumin comprising or consisting of SEQ ID No. 1, Cys is preferably immediately C-terminal to the last Leu (L585) residue. Such an albumin may be referred to as 'albumin-Cys', *e.g.* HSA-Cys. Other preferred C-terminal extensions of albumins such as SEQ ID No. 1 include albumin-Ala-Cys, such as HSA-Ala-Cys. Polypeptides suitable for providing extensions, as described above, may be added or inserted to the C- or, N- side of

clear that the counterpart cysteine (*e.g.* Fig. 6A) in the disulphide bond should be removed for example by deletion or substitution.

When a combination of two or more mutations is desired, amino acid residues which occur in different 'proximity' groups (*e.g.* with reference to SEQ ID No.1) may be preferred over those that occur within the same proximity group. For SEQ ID No. 1, there are 14 proximity groups (*i.e.* A to N). It is preferred that, for a thio-albumin having two or more conjugation competent cysteines, there is zero or one conjugation competent cysteine defined from each of the 14 groups. That is, it is preferred that such a thio-albumin does not contain two or more conjugation competent cysteines falling within the same group. A large number of permutations exist which meet this criterion.

For example, for a thio-albumin variant containing two free thiol groups residues based on selection criteria that generated Selection Group 1, then T79 + A364, in which one residue is selected from proximity group A to combine with A364 in proximity group C, would be preferred over T79 + E86 which both occur in proximity group A.

For combinations including cysteine-34, it is preferred not to select residues from proximity groups A, F or K. That is, it is preferred to select residues from one or more of proximity groups B to E, G to J and L to N.

Examples of preferred mutations selected from within Selection Group 1 may include the following:

- For 2 amino acid residues selected from Selection Group 1; amino acid residues from proximity groups A + C are preferred, such as T79 + A364, Similarly, amino acid residues selected from proximity groups C + E, such as A364 + D1 are also preferred. Also, amino acid residues from proximity groups D + E, such as L585 + A2 or + the C-side of L585 + A2, or from G + I, such as S270 + A581, are preferred.
- For 3 amino acid residues selected from Selection Group 1; amino acid residues which occur in proximity groups A + C + B are preferred such as T79 + A364 + D562. Similarly, amino acid residues selected from proximity groups B + C + E, such as D562 + A364 + A2 or D562 + A364 + D1, are also preferred.
- For 4 amino acid residues selected from Selection Group 1; amino acid residues which occur in proximity groups A + C + B + D such as such as T79 + A364 + D562 + A504, or alternatively T79 + D562 + A364 + L585 are preferred. Even more preferred are amino acid residues selected from proximity groups A + B + C + E, such as C34 + D562 + A364 + A2 T79 + D562 + A364 + D1.
- For 5 amino acid residues selected from Selection Group 1; amino acid residues which occur in proximity groups A + C + B + D + E such as T79 + D562 + A364 +L585 + D1 are preferred. Similarly, E86 + D562 + A364 +A504 + A2 are also preferred.

C177 are preferred. Alternatively, residues from proximity groups B + M, such as D562 + C369 are preferred.

Examples of sites from Selection Groups 2 + 3 may include residues from proximity groups H + M, such as D129 + C369 are preferred. Alternatively, residues from proximity groups I + M, such as A581 + C369 are preferred.

Examples of sites from Selection Groups 1 + 2 + 3 may include residues from proximity groups A + H + M + D, such as C34 + D129 + C360 + L585, from proximity groups B + H + M, such as D562 + D129 + C369 are preferred.

The above combinations are generally more preferred than combinations of residues from the following sets of proximity groups: (i) A, K and F; (ii) B, D, I, J and N; (iii); C and M; (iv) H and L.

The above albumin variants of the invention may or may not further comprise a cysteine at Cys34 of SEQ ID No. 1, or at an equivalent potisionm, if based on an albumin other than SEQ ID No. 1.

A skilled person will appreciate that the sites for introduction of more than one free thiol group (conjugation competent cysteine) have been selected from Selection Groups 1, 2 and 3. However, this approach may also be used to select sites for the introduction of more than one free thiol group from other residues selected from SEQ ID NO:1, *i.e.* he could use the disclosed method to generate other useful selection groups.

A preferred thio-albumin comprises SEQ ID No. 1 with Cys at positions 2 and 585 in addition to the naturally occurring Cys at position 34 (SEQ ID No. 78, construct 'TA33'). A more preferred thio-albumin comprises SEQ ID No. 1 with Cys at positions 2, 364, 562, 585 in addition to the naturally occurring Cys at position 34 (SEQ ID No. 82, construct 'TA38'). Thio-albumins comprising three or four of the Cys at positions 2, 364, 562 and 585 may also be preferred.

The polypeptide may or may not comprise at least one mutation that reduces glycosylation.

A third aspect of the invention provides a polynucleotide which encodes the polypeptide according to the invention. The polynucleotide may or may not be codon-optimised relative to the host from which it is to be expressed. SEQ ID No. 2 provides the usual coding sequence of HSA (SEQ ID No. 1). SEQ ID No. 3 provides a coding sequence of HSA (SEQ ID No. 1) which is codon-optimised for expression from *S. cerevisiae*. SEQ ID No. 2 or 3 may be mutated in order to provide a polynucleotide which encodes a polypeptide according to the invention. Preferably the polynucleotide is synthetic and/or recombinant. Preferably the polynucleotide is an isolated polynucleotide. The polynucleotide may encode an HSA with or without a leader sequence. For example, the polynucleotide may encode an HSA with the

		98)			
SA	Mongolian Gerbil	*Meriones unguiculatus* (SEQ ID No. 98)	O35090-1 / JC5838	73.2	609
SA	Goat	*Capra hircus* (SEQ ID No. 99)	B3VHM9-1	74.3	607
SA	Guinea Pig	*Cavia porcellus* (SEQ ID No. 100)	Q6WDN9-1	73.0	608
SA	Horse	*Equus caballus* (SEQ ID No. 101)	P35747-1	76.4	607
SA	Human	*Homo sapiens* (SEQ ID No. 102)	P02768-1	100.0	609
SA	Australian Lung-fish	*Neoceratodus forsteri*	P83517-1	22.8	101 (NL)
SA	Macaque (Rhesus Monkey)	*Macaca mulatta* (SEQ ID No. 103)	Q28522-1	93.5	608
SA	Mouse	*Mus musculus* (SEQ ID No. 104)	P07724-1 Version 3.	72.4	608
SA	North American bullfrogs	*Rana catesbeiana*	P21847-1	36.1	382 (NL)
SA	Pig	*Sus scrofa* (SEQ ID No. 105)	P08835-1 Version 2	75.6	607
SA	Rabbit	*Oryctolagus cuniculus* (SEQ ID No. 106)	P49065-1 Version 2	75.3	608
SA	Rat	*Rattus norvegicus* (SEQ ID No. 107)	P02770-1 Version 2.	73.4	608
SA	Salamander	*Ambystoma maculatum*	Q8UW05-1	38.8	626
SA	Salmon ALB1	*Salmo salar*	P21848-1	25.0	608
SA-2	Salmon ALB2	*Salmo salar*	Q03156-1	24.8	608
SA	Sea lamprey	*Petromyzon marinus*	Q91274-1	16.9	1423
SA	Sea lamprey	*Petromyzon marinus –AS*	O42279-1	17.4	551

Protein	Common Name	Species	SwissProt Accession No	Identity to SEQ ID NO: 1 (Clustal V)	Length (aa)
SA	Sheep	*Ovis aries* (SEQ ID No. 108)	P14639-1	75.3	607
SA	Sumatran Orangutan	*Pongo abelii*	Q5NVH5-1	100.0	609
SA	Tuatara	*Sphenodon punctatus*	Q8JIA9-1	43.8	527 (NL)
SA	Western clawed frog	*Xenopus (Silurana) tropicalis*	Q6DJ95-1	10.5	572 (NL)
OA	Chicken	*Gallus gallus*	P01012-1 Version 2	10.9	383 (NL)
OA	Turkey	*Meleagris gallopavo*	O73860-1 Version 3.	11.4	386 (NL)

SA: Serum albumin, SA-2: Serum albumin-2, OA: Ovalbumin, NL: No Leader sequence

Many naturally occurring mutant forms of albumin are known. Many are described in Peters, (1996, *All About Albumin: Biochemistry, Genetics and Medical Applications*, Academic Press, Inc., San Diego, California, p.170-181). A variant as defined herein may be one of these naturally occurring mutants such as those described in Minchiotti *et al.* (2008). Hum Mutat 29(8): 1007-16 and http://www.uniprot.org/uniprot/P02768.

A "variant albumin" refers to an albumin protein wherein at one or more positions there have been amino acid insertions, deletions, or substitutions, either conservative or non-conservative, provided that such changes result in an albumin protein for which at least one basic property, for example binding activity (type of and specific activity *e.g.* binding to bilirubin or a fatty acid such as a long-chain fatty acids, for example oleic (C18:1), palmitic (C16:0), linoleic (C18:2), stearic (C18:0), arachidonic (C20:4) and/or palmitoleic (C16:1)), osmolarity (oncotic pressure, colloid osmotic pressure), behaviour in a certain pH-range (pH-stability) has not significantly been changed. "Significantly" in this context means that one skilled in the art would say that the properties of the variant may still be different but would not be unobvious over the ones of the original protein, *e.g.* the protein from which the variant is derived. Such characteristics may be used as additional selection criteria in the invention.

Typically an albumin variant will have more than 40%, usually at least 50%, more typically at least 60%, preferably at least 70%, more preferably at least 80%, yet more preferably at

least 90%, even more preferably at least 95%, most preferably at least 98% or more sequence identity with a naturally occurring albumin such as SEQ ID No. 1. The percent sequence identity between two polypeptides may be determined using suitable computer programs, for example the GAP program of the University of Wisconsin Genetic Computing Group and it will be appreciated that percent identity is calculated in relation to polypeptides whose sequence has been aligned optimally. The alignment may alternatively be carried out using the Clustal W program or the Clustal V program and therefore allow calculation of % homology between sequences of a multiple alignment and/or calculation of % identity between sequences of a pairwise alignment. The parameters used may be as follows:

Fast pairwise alignment parameters: K-tuple(word) size; 1, window size; 5, gap penalty; 3, number of top diagonals; 5. Scoring method: x percent. Multiple alignment parameters: gap open penalty; 10, gap extension penalty; 0.05. Scoring matrix: BLOSUM

Custal W: Pairwise alignment parameters: 'Slow-Accurate', Gap Penalty: 10, Gap Length: 0.1, Protein Weight Matrix: Gonnet 250, DNA Weight Matrix: IUB. Multiple Alignment Parameters: Gap penalty 10.00, gap length penalty 0.20, Delay Divergent Seqs(%) 30, DNA transition weight 0.50, Protein weight matrix = Gonnet series, DNA weight matrix = IUB.

Clustal V: Pairwise alignment parameters: Ktuple: 1, Gap Penalty: 3, Window: 5, Diagonals: 5; Multiple alignment parameters: Gap penalty 10, gap length penalty 10.

Conservative substitution

As used herein, the term "conservative" amino acid substitutions refers to substitutions made within the same group, and which typically do not substantially affect protein function. By "conservative substitutions" is intended combinations such as Gly, Ala; Val, Ile, Leu; Asp, Glu; Asn, Gln; Ser, Thr; Lys, Arg; and Phe, Tyr. Such variants may be made by techniques well known in the art, such as by site-directed mutagenesis as disclosed in US Patent No 4,302,386 issued 24 November 1981 to Stevens, incorporated herein by reference.

In one embodiment, the Venn diagram of Figure 4 may be used to determine conservative amino acid substitutions: Using Fig. 4., a conservation mutation score (ranging from 0 to 5) may be calculated. A score of 0 is the highest conservation, which, for cysteine, is only assigned for substitution of a cysteine residue with another cysteine residue. For changes from any other amino acid to a cysteine, the score may be 1, 2, 3, 4, 5. A score of 1 is a more conservative substitution that a score of 2, 3, 4 or 5. A score of 5 is assigned to the lowest conservation between a substituted amino acid and the cysteine. The score of 0 to 5 is calculated from Fig. 4 as the number of boundaries (*i.e.* lines) crossed to go from cysteine

zerland, Poster Abstract P38; the poster abstract may be viewed on the world wide web at http://www.micro.biol.ethz.ch/cellwall/). If a yeast other than *S. cerevisiae* is used, disruption of one or more of the genes equivalent to the *PMT* genes of *S. cerevisiae* is also beneficial, *e.g.* in *Pichia pastoris* or *Kluyveromyces lactis*. The sequence of *PMT1* (or any other *PMT* gene) isolated from *S. cerevisiae* may be used for the identification or disruption of genes encoding similar enzymatic activities in other fungal species. The cloning of the *PMT1* homologue of *Kluyveromyces lactis* is described in WO 94/04687.

The yeast may or may not also have a deletion of the *HSP150* and/or *YAP3* genes as taught respectively in WO 95/33833 and WO 95/23857.

The HSA variant may be produced by recombinant expression and secretion. Where the expression system (*i.e.* the host cell) is yeast, such as *Saccharomyces cerevisiae*, suitable promoters for *S. cerevisiae* include those associated with the *PGK1* gene, *GAL1* or *GAL10* genes, *TEF1*, *TEF2*, *PYK1*, *PMA1*, *CYC1*, *PHO5*, *TRP1*, *ADH1*, *ADH2*, the genes for glyceraldehyde-3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, triose phosphate isomerase, phosphoglucose isomerase, glucokinase, α-mating factor pheromone, a-mating factor pheromone, the *PRB1* promoter, the *PRA1* promoter, the *GPD1* promoter, and hybrid promoters involving hybrids of parts of 5' regulatory regions with parts of 5' regulatory regions of other promoters or with upstream activation sites (*e.g.* the promoter of EP-A-258 067).

Suitable transcription termination signals are well known in the art. Where the host cell is eukaryotic, the transcription termination signal is preferably derived from the 3' flanking sequence of a eukaryotic gene, which contains proper signals for transcription termination and polyadenylation. Suitable 3' flanking sequences may, for example, be those of the gene naturally linked to the expression control sequence used, *i.e.* may correspond to the promoter. Alternatively, they may be different. In that case, and where the host is a yeast, preferably *S. cerevisiae*, then the termination signal of the *S. cerevisiae ADH1*, *ADH2*, *CYC1*, or *PGK1* genes are preferred.

It may be beneficial for the promoter and open reading frame of the gene encoding the recombinant protein comprising the sequence of an albumin mutant to be flanked by transcription termination sequences so that the transcription termination sequences are located both upstream and downstream of the promoter and open reading frame, in order to prevent transcriptional read-through into any neighbouring genes, such as 2μm genes, and vice versa.

In one embodiment, the favoured regulatory sequences in yeast, such as *Saccharomyces cerevisiae*, include: a yeast promoter (*e.g.* the *Saccharomyces cerevisiae PRB1*

quence. Accordingly, the term "signal peptide" is understood to mean a presequence which is predominantly hydrophobic in nature and present as an N-terminal sequence of the precursor form of an extracellular protein expressed in yeast. The function of the signal peptide is to allow the expressed protein to be secreted to enter the endoplasmic reticulum. The signal peptide is normally cleaved off in the course of this process. The signal peptide may be heterologous or homologous to the yeast organism producing the protein. Known leader sequences include those from the *S. cerevisiae* acid phosphatase protein (Pho5p) (see EP 366 400), the invertase protein (Suc2p) (see Smith *et al.* (1985) *Science*, **229**, 1219-1224) and heat-shock protein-150 (Hsp150p) (see WO 95/33833). Additionally, leader sequences from the *S. cerevisiae* mating factor alpha-1 protein (MFα-1) and from the human lysozyme and human serum albumin (HSA) protein have been used, the latter having been used especially, although not exclusively, for secreting human albumin. WO 90/01063 discloses a fusion of the MFα-1 and HSA leader sequences (also known as the fusion leader sequence (FL)). In addition, the natural albumin leader sequence may or may not be used to direct secretion of the recombinant protein comprising the sequence of an albumin mutant.

The skilled person will appreciate that any suitable plasmid may be used, such as a centromeric plasmid. The examples provide suitable plasmids (centromeric YCplac33-based vectors) for use to transform yeast host cells of the present invention. Alternatively, any other suitable plasmid may be used, such as a yeast-compatible 2μm-based plasmid.

Plasmids obtained from one yeast type can be maintained in other yeast types (Irie *et al*, 1991, *Gene*, **108**(1), 139-144; Irie *et al*, 1991, *Mol. Gen. Genet.*, **225**(2), 257-265). For example, pSR1 from *Zygosaccharomyces rouxii* can be maintained in *Saccharomyces cerevisiae*. In one embodiment the plasmid may or may not be a 2μm-family plasmid and the host cell will be compatible with the 2μm-family plasmid used (see below for a full description of the following plasmids). For example, where the plasmid is based on pSR1, pSB3 or pSB4 then a suitable yeast cell is *Zygosaccharomyces rouxii*; where the plasmid is based on pSB1 or pSB2 then a suitable yeast cell is *Zygosaccharomyces bailii*; where the plasmid is based on pSM1 then a suitable yeast cell is *Zygosaccharomyces fermentati*; where the plasmid is based on pKD1 then a suitable yeast cell is *Kluyveromyces drosophilarum*; where the plasmid is based on pPM1 then a suitable yeast cell is *Pichia membranaefaciens*; where the plasmid is based on the 2μm plasmid then a suitable yeast cell is C *Saccharomyces cerevisiae* or *Saccharomyces carlsbergensis*. Thus, the plasmid may be based on the 2μm plasmid and the yeast cell may be *Saccharomyces cerevisiae*. A 2μm-family plasmid can be said to be "based on" a naturally occurring plasmid if it comprises one, two or preferably three of the genes *FLP*, *REP1* and *REP2* having sequences derived from that naturally occurring plasmid.

as a combined diagnostic and therapeutic tool to visualize and monitor, for example, dystrophic muscle by magnetic resonance imaging (MRI) and for the delivery of putative therapeutics bound to albumin for effective targeting to dystrophic muscle (Amthor *et al.* (2004) Neuromuscular Disorders 14912: 791-796). Malignant tumours often show an increased uptake and metabolism of albumin. The use of gadolinium-albumin conjugate has also been described for improved imaging of malignant tumours and to determine by MRI tumours sensitive to a therapy with drug-conjugated albumin (Kiessling *et al.* (2002) Investigative Radiology 37(4): 93-198).

Current imaging agents often degrade quickly whilst longer-lasting agents are often toxic. The use of albumin conjugates may be especially useful to increase the half life of imaging agents and would therefore permit imaging over an extended period of time. WO2005/082423 describes the use of serum albumin conjugated to fluorescent substances for imaging.

A thio-albumin of this invention may be conjugated to two or more molecules selected from imaging agents, diagnostic agents, therapeutic compounds and contrast agents.

Tumours (and muscle degeneration) show enhanced uptake of albumin (EPR: Enhanced Permeation and Retention). Albumin conjugates may be used for enhanced imaging, and also to assess whether tumours (or other tissues and organs) would be suitable for albumin conjugated drugs.

Bioactive compound

The bioactive compound may be a therapeutic or diagnostic compound. The therapeutic compound may be a chemotherapy drug for use in cancer chemotherapy. It may be cytostatic or cytotoxic; it may be a tumor-inhibiting agent.

The bioactive compound may already contain a free thiol group, *e.g.* a polypeptide containing a Cysteine residue with a free thiol group. Alternatively, the bioactive compound may be modified so as to contain a free thiol group. Thus, the amino acid sequence of a polypeptide may be altered so as to include a Cysteine residue with a free thiol group, or the bioactive compound may be chemically derivatized to include a free thiol group.

The bioactive compound may be a polypeptide (protein), particularly a recombinant protein pharmaceutical. It may be a chemotherapy or radiotherapy drug used to treat cancers and other related diseases.

The free thiol containing albumin mutein of the invention (thio-albumin) can be conjugated *via* the free thiol group, or groups if the albumin mutein of the invention contains more than one free thiol, to at least one bioactive compound by methods known to the art. The

kine hMCP1b, Chemokine hMCP2, Chemokine hMCP3, Chemokine hSDF1b, Chemokine MCP-4, chemokine TECK and TECK variant, Chemokine-like protein IL-8M1 Full-Length and Mature, Chemokine-like protein IL-8M10 Full-Length and Mature, Chemokine-like protein IL-8M3, Chemokine-like protein IL-8M8 Full-Length and Mature, Chemokine-like protein IL-8M9 Full-Length and Mature, Chemokine-like protein PF4-414 Full-Length and Mature, Chemokine-like protein PF4-426 Full-Length and Mature, Chemokine-like protein PF4-M2 Full-Length and Mature, Cholera vaccine, Chondromodulin-like protein, c-kit ligand; SCF; Mast cell growth factor; MGF; Fibrosarcoma-derived stem cell factor, CNTF and fragment thereof (such as CNTFAx15`(Axokine™)), , coagulation factors in both pre and active forms, collagens, Complement C5 Mab, Connective tissue activating protein-III, CTAA16.88 Mab, CTAP-III, CTLA4-Ig, CTLA-8, CXC3, CXC3, CXCR3; CXC chemokine receptor 3, cyanovirin-N, , Darbepoetin, designated exodus, designated huL105_7., DIL-40, Dnase, EDAR, EGF Receptor Mab, ENA-78, Endostatin, Eotaxin, Epithelial neutrophil activating protein-78, EPO receptor; EPOR, erythropoietin (EPO) and EPO mimics, Eutropin , Exodus protein, Factor IX, , Factor VII, , Factor VIII, , Factor X and Factor XIII, FAS Ligand Inhibitory Protein (DcR3), FasL, FasL, FasL, FGF, FGF-12; Fibroblast growth factor homologous factor-1, FGF-15, FGF-16, FGF-18, FGF-3; INT-2, FGF-4; gelonin, HST-1; HBGF-4, FGF-5, FGF-6; Heparin binding secreted transforming factor-2, FGF-8, FGF-9; Glia activating factor, fibrinogen, , flt-1, flt-3 ligand, Follicle stimulating hormone Alpha subunit, Follicle stimulating hormone Beta subunit, Follitropin, Fractalkine, fragment. myofibrillar protein Troponin I, FSH, Galactosidase, Galectin-4, G-CSF, GDF-1, Gene therapy, Glioma-derived growth factor, glucagon, glucagon-like peptides, Glucocerebrosidase, glucose oxidase, Glucosidase, Glycodelin-A; Progesterone-associated endometrial protein, GM-CSF, gonadotropin, Granulocyte chemotactic protein-2 (GCP-2), Granulocyte-macrophage colony stimulating factor, growth hormone, , Growth related oncogene-alpha (GRO-alpha), Growth related oncogene-beta (GRO-beta), Growth related oncogene-gamma (GRO-gamma), hAPO-4; TROY, hCG, Hepatitus B surface Antigen, Hepatitus B Vaccine, HER2 Receptor Mab, hirudin, HIV gp120, HIV gp41, HIV Inhibitor Peptide, HIV Inhibitor Peptide, HIV Inhibitor Peptide, HIV protease inhibiting peptides, HIV-1 protease inhibitors, HPV vaccine, Human 6CKine protein, Human Act-2 protein, Human adipogenesis inhibitory factor, human B cell stimulating factor-2 receptor, Human beta-chemokine H1305 (MCP-2), Human C-C chemokine DGWCC, Human CC chemokine ELC protein, Human CC type chemokine interleukin C, Human CCC3 protein, Human CCF18 chemokine, Human CC-type chemokine protein designated SLC (secondary lymphoid chemokine), Human chemokine beta-8 short forms, Human chemokine C10, Human chemokine CC-2, Human chemokine CC-3, Human chemokine CCR-2, Human chemokine Ckbeta-7, Human chemokine ENA-78, Human chemokine eotaxin, Human chemokine GRO alpha, Human chemokine GROalpha, Human chemokine GRObeta, Human chemokine

Imaging agents

Gadolinium, magnetite, manganese, technetium, I125, I131, P32, TI201, Iopamidol, PET-FDG.

Purification tags

The albumin may also be fused to one or more purification tags such as (Ala-Trp-Trp-Pro)$_n$, avidin/streptavidin/Strep-tag, BCCP, B-tag (VP7 protein region of bluetongue virus), calmodulin binding protein (CBP), cellulose binding domains (CBD's), chitin binding domain, chloramphenicol acetyltransferase, c-myc, dihydrofolate reductase (DHFR), FLAG™ peptide (DYKDDDDK), galactose-binding protein, glutathione-S-transferase (GST), green fluorescent protein (GFP), Growth hormone, N-terminus, hemagglutinin influenza virus (HAI), His-patch thioredoxin, His-tag, HSB-tag, KSI, lacZ (β-Galactosidase), maltose binding protein (MBP), NusA, ompT/ompA/pelB/DsbA/DsbC, polyarginine, polyaspartic acid, polycysteine, polyphenyalanine, S-tag, staphylococcal protein A, streptococcal protein G, T4 gp55, T7gene10, T7-tag, thioredoxin, trpE, ubiquitin.

Ligand binding

HSA has ligand binding and esterase activities, as described in "All about Albumin", T. Peters Jr., Academic Press N. Y. The ligand binding properties include binding to anionic and neutral ligands such as long-chain fatty acids, bilirubin and other miscellaneous ligands. The long-chain fatty acids, oleic (C18:1), palmitic (C16:0), linoleic (C18:2), stearic (C18:0), arachidonic (C20:4) and palmitoleic (C16:1) are known to bind HSA.

The polypeptide may include insertions, deletions and substitutions, either conservative or non-conservative, where such changes do not substantially reduce the useful ligand-binding, immunological or receptor binding properties of albumin, for example to FcRN, bilirubin and/or a fatty acid. The polypeptide may have at least 5%, 10%, 15%, 20%, 30%, 40% or 50%, 60%, 70%, at least 80%, 90%, 95%, 100%, 105% or more of human serum albumin's receptor binding activity, mole for mole. The polypeptide may have increased affinity for an albumin receptor.

Ligand binding studies can be performed on HSA and thio-albumins using an isothermal titration calorimetry method that had been suitably qualified for this purpose. Samples can be pre-treated by defatting (Sogami, M. and J. F. Foster (1968). Biochemistry **7**(6): 2172-82, incorporated herein by reference) followed by thiol blocking (Sogami, M., H. A. Petersen, et al. (1969). Biochemistry **8**(1): 49-58, incorporated herein by reference) and subsequent gel permeation chromatography. The binding curves generated for thio-albumins and

HSA with octanoate, for example, may subsequently be compared, and functional similarity established.

Conjugation Methods

The albumin mutein (thio-albumin) of the invention can be covalently linked to one or more conjugation partners such as bioactive compounds by methods known in the art (for example those provided by Pierce, Thermo Fisher Scientific, Rockford, IL, USA). These include, but are not limited to incorporating or engineering a thiol reactive group into or onto the conjugation partner, for example by incorporating or engineering another free thiol present on the conjugation partner; or by incorporating or engineering a pyridyl disulphide group on the conjugation partner; or by incorporating or engineering an iodoacetyl group on the bioactive compound or by incorporating or engineering a maleimide group on the conjugation partner. For example, N-ethylmaleimide (NEM, Pierce), 2-amino-2'-aminoethanethiolsulfonate (Pierce), N-beta-maleimidoprpionic acid (BMPA Pierce), methyl methane thiosulfonate (MMTS, Pierce), fluorescein-5-maleimide (Pierce), 5-iodoacetamido-fluorescein (5-IAF, Pierce) or N-[6-7-amino-4-methylcoumarin-3-acetamido) hexyl]-3'-[2'-pyridyldithio] propionamide (AMCA-HPDP, Pierce).

If the conjugation partner contains at least one thiol group, then the conjugation partner may be cross-linked to the albumin mutein of the invention by methods known to the art such as, but not limited to, oxidation or by the use of cross-linking reagents such as, but not limited to, 1,4-Bis-maleimidibutane (BMB, Pierce); 1,4-Bis-maleimidyl-2,3-dihydroxybutane (BMDB, Pierce); Bis-maleimidohexane (BMH, Pierce), Bis-maleimidoethane (BMOE, Pierce); 1,8-Bis-Maleimidotriethyleneglycol (BM[PEO]3 Pierce); 1,11-Bis-Maleimidotetraethyleneglycol (BM[PEO]4 Pierce); 1,4-Di-[3'-(2'-pyridyldithio)-propionamido]butane (DPDPB, Pierce); dithuio-bis-maleimidoethane (DTME Pierce); 1,6-Hexane-bis-vinylsulfone (HBVS, Pierce) and Tris-[2-maleimimidoethyl]amine (TMEA, Pierce).

If the conjugation partner does not contain a thiol reactive group then it may be modified to incorporate one or more such groups by either chemical modification or genetic engineering by methods know to the art (Chapman, A.P. (2002) Adv. Drug Deliv. Rev., 54 531–545: Humphreys, D.P. *et al.* Protein Engineering, Design & Selection vol. 20 no. 5 pp. 227–234, 2007). While these two references describe methodologies to cross-link PEG to an engineered free thiol within an antibody or antibody fragment, the techniques may be used to cross-link a conjugation partner to an engineered free thiol within the albumin mutein of the invention. Alternatively the Drug Affinity Complex (DAC™) technology developed by ConjuChem Inc. (Montreal, Quebec, Canada, H2X 3Y8) may be used, *e.g.* as described in

described in WO/2005061719 A1. Plasmid pDB2690 (13.018kb) was digested to completion with restriction endonuclease *Not*I and dephosphorylated using calf alkaline intestinal phosphatase and ligated with the 2.992kb *Not*I human serum albumin expression cassette to produce a 16.039 kb plasmid pDB2713 which has the human serum albumin expression cassette orientated in the same direction as the *LEU2* gene (Fig. 9). A person skilled in the art will appreciate that the expression cassette may or may not be codon optimised and that the expression cassette may or may not be cloned in either orientation in the expression vector as part of this invention.

As an alternative to site-directed mutagenesis expression plasmids for thio-albumin (*i.e.* conjugation competent albumin) variants of this invention could be made by subcloning synthesized DNA fragments into plasmid pDB2243 (Fig. 8) prior to cloning into pSAC35 or pDB2690. A method for the construction of a thio-albumin subcloning plasmid containing one extra conjugation competent cysteine (relative to SEQ ID No. 1) is described, by way of example only, below.

The albumin DNA sequence of pDB2243 includes two *Hind*III restriction endonuclease sites.

The synthetic DNA may be modified such that the human serum albumin protein encoding sequence is modified at a selected codon to a cysteine codon, or an existing cysteine codon is deleted or modified to a codon for another amino acid. Alternatively, the coding sequence for the mature thio-albumin may be extended at the 5' or 3' end(s) or insertions made within the polypeptide to add novel sequence(s) coding for cysteine or polypeptides containing one or more cysteine .

Alternatively synthetic DNA may be modified such that the human serum albumin protein encoding sequence is modified at a selected cysteine codon to an alternative codon to create an unpaired cysteine. Alternatively synthetic DNA may be modified such that the human serum albumin protein encoding sequence is modified by substitution of two codons at a specified site to a cysteine codon (the amino acid chain length is reduced). Alternatively synthetic DNA may be modified such that the human serum albumin protein encoding sequence (*e.g.* SEQ ID No. 2 or SEQ ID No. 3 in relation to HSA) is modified by insertion of a cysteine codon at a specified site (the amino acid chain length is increased). Plasmid pDB2243 may be digested to completion with *Hind*III restriction endonuclease and the fragment (approximately 4.383kb) is recovered and dephosphorylated, the synthetic DNA containing the appropriate modification to the human serum albumin encoding sequence may then be cloned to produce the required thio-albumin subcloning plasmid. The thio-albumin subcloning plasmid may then be digested to produce an expression cassette, which may be cloned into a suitable expression plasmid in a similar manner to the construction of pDB2244, pDB2305 or

pDB2244 (Figure 7, WO 0044772A, 'FL': fusion leader sequence). pDB2244 was digested with restriction enzymes *Swa*I and *Hpa*I (both produce blunt ends) and self-ligated to form pDB3927. To create plasmid pDB3964 (Figure 13) restriction enzyme sites were modified in the albumin DNA sequence (SEQ ID No. 2) of pDB3927 without modifying the protein sequence, as outlined below. The resultant DNA sequence is sequence ID No. 4.

1) Introduced enzyme sites: (basepair positions in brackets refer to positions in SEQ ID NO:4)

	Restriction site	SEQ ID No.		SEQ ID No. 4	Position
a	*Sac*II:	GAGTCAGCTGAAAA (SEQ ID NO:126)	→(to)	GAGTCCGCGGAAAA (SEQ ID NO:127)	(bp 173-178)
b	*Nhe*I/*Bmt*I:	AAGGCTTCGTCTGC (SEQ ID NO:128)	→(to)	AAGGCTAGCTCTGC (SEQ ID NO:129)	(bp 571-576)
c	*Xho*I:	TCTGCTTGAATGTGC (SEQ ID NO:130)	→(to)	TCTGCTCGAGTGTGC (SEQ ID NO:131)	(bp 751-756)
d	*Bam*HI:	GTGGGCAGCAAAT (SEQ ID NO:132)	→(to)	GTGGGATCCAAAT (SEQ ID NO:133)	(bp 751-756)
e	*Sal*I:	GGAAGTCGATGAAA (SEQ ID NO:134)	→(to)	GGAAGTCGACGAAA (SEQ ID NO:135	(bp 1477-1482)

The coding sequence of HSA in pDB3964 is provided as SEQ ID No. 4. DNA synthesis and cloning was used to generate pDB3964 from pDB3927 (DNA2.0 Inc, USA). Synthetic DNA fragments were designed to alter specific amino acid codons within the albumin gene of pDB3964, or with combinations of modifications (see Example 3 below). DNA fragments containing these modifications were synthesised (DNA 2.0 Inc, USA) and cloned into pDB3964 to produce plasmids containing the thio-albumin sequences (Figure 11). These synthetic genes and flanking regions were excised with restriction enzymes *Bst*EII and *Bsr*BI from the plasmids named in Figure 11 for each of the thio-albumin variants and the controls pDB3927, pDB3964 and pDB2244, before purification of the resulting DNA fragments (PCR purification kit, Qiagen). The DNA fragments were used in the yeast transformation procedure described below to allow gap-repair in vivo with linearised pDB3936.

The plasmid pDB3853 (not shown) was constructed from base vector pDB2690 (Ref DB88 / WO2005/061719A1) and the synthetic linker described below. The synthetic linker was constructed from two oligonucleotides (Sigma-Genosys) annealed in distilled water using a temperature gradient from 96°C to room temperature (1 min per 1°C). pDB2690 was digested using *Kpn*I and *Not*I, and purified by gel extraction (Qiagen), before ligation of the annealed linker:

'*Kpn*I (linker) *Bam*HI *Not*I'

5′ - CGCTAGCCTCGAGGTTTAAACGCTAGCGAGCTCGGATCC -3′
(SEQ ID NO: 136)
3′ - CATGGCGATCGGAGCTCCAAATTTGCGATCGCTCGAGCCTAGGCCGG -5′
(SEQ ID NO: 137)

Following the construction of pDB3853, the linker was excised using *Pst*I and *Sca*I (3787bp fragment) before ligation into the gel extracted *Pst*I/*Sca*I cut pSAC35 plasmid (WO 0044772A and WO2005/061719A1)), to form pDB3936 (Figure 14).

pDB3936 was linearised with restriction enzymes *Acc*65I and *Bam*HI before purification of the 9721bp fragment following separation by agarose gel electrophoresis. For the yeast transformation procedure to allow gap-repair in vivo (described below) the concentrations of the linearised pDB3936 and each of the *Bsr*BI-*Bst*EII fragments encoding the thioalbumin coding sequences was calculated and 100ng of each use for each yeast transformation reaction.

Saccharomyces cerevisiae strain BXP10 was used as the expression host throughout (So-low, S. P., J. Sengbusch, et al. (2005). "Heterologous protein production from the inducible MET25 promoter in *Saccharomyces cerevisiae*." Biotechnol Prog 21(2): 617-20.), although alternative expression hosts are also be suitable.

Cryopreserved stocks of *S. cerevisiae* BXP10 were prepared from 10mL YEPPD (1% w/v yeast extract, 2% w/v plant peptone, 2% w/v dextrose)) shake flask cultures (grown for 24 hours, 30°C, 200rpm) mixed with an equal volume of 40% w/v sterile trehalose solution and dispensed in 1mL aliquots for storage at -80°C. 10mL BMMD, YEPPD and LB (1% w/v bacteriological tryptone, 0.5% w/v yeast extract, 0.5% w/v NaCl) shake flasks were inoculated with 100µL cryopreserved yeast stock and incubated for four days at 30°C, 200rpm as above before being observed microscopically to confirm they were axenic.

Frozen competent *S. cerevisiae* BXP10 cells were prepared by inoculating 100µL cryopreserved yeast stock into 10mL YEPPD which were incubated for two days at 30°C, 200 rpm, before being used to inoculate 300mL YEPPD to an OD600 = 0.3. The cells were incubated as above for approximately 4 hours or until a doubling of OD_{600} had been achieved. The cells were harvested by centrifugation (3000xg, 5min, room temperature) before resuspension in 120mL distilled water followed by a further centrifugation step. The pellet was resuspended in 3mL TE/LiAc (10mM Tris, 1mM EDTA, pH7; 500mM lithium acetate) and glycerol added to a final concentration of 15% (v/v), before storage in aliquots at -80°C.

S. cerevisiae BXP10 cells were transformed to leucine prototrophy using a modified lithium acetate method (Elble, R. "A simple and efficient procedure for transformation of yeasts." Biotechniques 13.1 (1992): 18-20. Ito, H., et al. "Transformation of intact yeast cells treated with alkali cations." J.Bacteriol. 153.1 (1983): 163-68.). 50µl of thawed competent cells were aliquoted into a 48-well microtitre plate (Nunc) before the addition of DNA frag-

CLAIMS

1) A conjugation competent polypeptide comprising up to and including five conjugation competent cysteine residues,

a) wherein said polypeptide comprises an albumin sequence having at least 96% sequence identity to SEQ ID NO: 1; and

b) wherein said poly peptide is a mutant comprising a Cysteine at or immediately following an amino acid residue in the albumin sequence which corresponds to the second, third, fourth or fifth residue relative to the N- or C-terminus of the albumin sequence; and/or

c) wherein the polypeptide is a mutant comprising a Cysteine at or adjacent to an amino acid residue in the albumin sequence which:

i) has a solvent accessibility of at least 80%, is not located within secondary structure, has a B-factor of a C-alpha atom of at least 50 as defined by x-ray crystallography and has no polymorphism at this position that is known to cause thermal instability; or

ii) has a solvent accessibility of at least 80%, is located within secondary structure, has a B-factor of a C-alpha atom of at least 30 as defined by x-ray crystallography and has no polymorphism at this position is known to cause thermal instability; and

d) optionally, wherein the polypeptide comprises additional alterations to the albumin sequence where each alteration is an amino acid deletion, substitution, or insertion;

with the proviso that, where there are two or fewer conjugation competent cysteine residues within the polypeptide, such polypeptide does not include a cysteine at the position corresponding to position 1 of SEQ ID NO: 1.

2) A polypeptide according to claim 1 comprising conjugation competent cysteine residues at positions equivalent to positions 2 and 585 of SEQ ID NO: 1.

3) A polypeptide according to claim 1 comprising four or more conjugation competent cysteine residues.

4) A polypeptide according to claim 3 comprising: conjugation competent cysteine residues at positions equivalent to positions 2 and 585 of SEQ ID NO: 1; and a conjugation competent cysteine residue at one or both of positions equivalent to positions 364 and 562 of SEQ ID NO: 1.

5) A polypeptide according to claim 1 in which the polypeptide comprises one or more of:

a) substitution of an amino acid, other than cysteine, with a cysteine at a position corresponding to a position equivalent to any of residues L585, A2, D562, A364, A504, E505, E86, D129, D549, A581, D121, E82, Q397and A578 of SEQ ID No. 1;

b) insertion of a cysteine at a position adjacent the N- or C- side of an amino acid corresponding to a position equivalent to any of residues L585, A2, D562, A364, A504, E505, E86, D129, D549, A581, D121, E82, Q397 and A578 of SEQ ID No. 1; and

c) addition of a cysteine to the N- side of the N-terminal residue of an albumin sequence or to the C- side of the C-terminal residue of an albumin sequence; such that the net result of the substitution, deletion, addition or insertion events of a), b) and c) is that the number of conjugation competent cysteine residues of the polypeptide sequence is increased relative to the polypeptide prior to the substitution, insertion, deletion and addition events.

6) The polypeptide according to claim 1, wherein at least one cysteine residue has been substituted with a different amino acid residue.

7) A polypeptide according to claim 1 in which one or more of the conjugation competent cysteines is conjugated to a bioactive molecule.

8) A polypeptide according to claim 1 in which the polypeptide has at least 98% sequence identity to SEQ ID NO: 1.

9) A polypeptide according to claim 1 in which the polypeptide further comprises a further linker to which a bioactive compound may be linked.

10) A polypeptide according to claim 1 in which the polypeptide has at least 96% sequence identity to SEQ ID NO: 1 and comprises conjugation competent cysteine residues located at: (a) A2+L585, (b) A2+A364+D562+L585C, (c) A2 and adjacent the C-side of the C-terminus of the albumin (d) T79+A364; (f) T79+D562+A364; (g) D562+A364+D1; (h) T79+D562+A364+A504; (i) T79+D562+A364+L585; (j) T79+D562+A364+D1; (k) T79+D562+A364+L585+D1; (1) E86+D562+A364+A504+A2; (m) S270+A581; (n) S270+D129; (o) S270+A581+E82; (p) S270+A581+D129; (q) S270+A581+E82+D129; (r) S270+A581+E82 +D129+Q397; (s) C369+C177; (t) A364+A581; (u) T79+A364+A581; (v) A364+A581+D129; (w) A364+C177; (x) D562+C369; (y) D129+C369; (z) A581+C369; or (ai) D562+D129+C369.

11) The polypeptide according to claim 1 in which the polypeptide has at least 97% sequence identity to SEQ ID NO: 1.

12) The polypeptide according to claim 1 in which the polypeptide has at least 98% sequence identity to SEQ ID NO: 1.

13) The polypeptide according to claim 1 in which the polypeptide has at least 99% sequence identity to SEQ ID NO: 1.

14) A conjugate which comprises a bioactive compound and a polypeptide according to claim 1, wherein the bioactive compound is linked to the polypeptide through a conjugation competent cysteine residue of the polypeptide.

15) A conjugate which comprises a bioactive compound and a polypeptide, wherein the polypeptide comprises an amino acid sequence having at least 98% sequence identity to residues 1 to 585 of SEQ ID NO: 1 in which: a) at a position equivalent to position 34 of SEQ ID NO: 1, there is a conjugation competent cysteine residue; and b) within those residues corresponding to positions 1-585 of SEQ ID NO: 1 there are two or more conjugation competent cysteine residues, wherein the bioactive compound is linked to the polypeptide through a conjugation competent cysteine residue of the polypeptide.

16) A conjugate which comprises a bioactive compound and a polypeptide, wherein the polypeptide comprises an amino acid sequence having at least 98% sequence identity to residues 1 to 585 of SEQ ID NO: 1 in which: a) at a position equivalent to position 34 of SEQ ID NO: 1, there is not a conjugation competent cysteine residue; and b) elsewhere in the polypeptide there are three or more conjugation competent cysteine residues, wherein the bioactive compound is linked to the polypeptide through a conjugation competent cysteine residue of the polypeptide.

17) A conjugation competent polypeptide comprising an amino acid sequence which has at least 98% sequence identity to residues 1 to 585 of SEQ ID NO: 1 in which: a) at a position equivalent to position 34 of SEQ ID NO: 1, there is a conjugation competent cysteine residue; and b) within those residues corresponding to positions 1-585 of SEQ ID NO: 1 there are two or more conjugation competent cysteine residues.

18) A conjugation competent polypeptide comprising an amino acid sequence which is at least 98% identical to residues 1 to 585 of SEQ ID NO: 1 in which: a) at a position equivalent to position 34 of SEQ ID NO: 1, there is not a conjugation competent cysteine residue; and b) elsewhere in the polypeptide there are three or more conjugation competent cysteine residues.

19) A polynucleotide which encodes the polypeptide of claim 1.

20) A plasmid comprising the polynucleotide of claim 19.

21) A host cell comprising a polynucleotide of claim 19.

22) The host cell of claim 21, which is a yeast cell.

23) A method of producing a polypeptide of claim 1, comprising: culturing the host cell of claim 21 under conditions that allow expression of the polypeptide; and recovering the polypeptide from the host cell and/or from host cell growth medium.