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.

Fig. 11 is a table showing mutations (see second column) made to native human serum albumin to generate molecules having a single free thiol group in addition to Cys-34 of native human serum albumin.

Fig. 12 is a map of plasmid pDB3927.

Fig. 13 is a map of plasmid pDB3964.

Fig. 14 is a map of plasmid pBD3936.

Fig. 15 shows SDS-PAGE analysis, and HPLC data (bar chart), showing the expression (µg/ml with standard deviation) of albumin molecules having a free thiol group at Cys-34 of SEQ ID No. 1 and an additional free-thiol at the position indicated below the bar chart.

Fig. 16 is a table showing mutations (see second column) made to native human serum albumin to generate molecules having one or more free thiol groups in addition to Cys-34 of native human serum albumin and/or having Cys-34 removed.

Fig. 17 shows SDS-PAGE analysis, and HPLC data (bar chart), showing the expression (µg/ml with standard deviation of albumin molecules having one or more free thiol groups in addition to Cys-34 of native human serum albumin and/or having Cys-34 removed.

Fig. 18 is a table showing the fermentation yield and relative level of conjugation to albumin molecules comprising one or more free-thiols.

Fig. 19 is a mass spectrogram of a rHA molecule designed to have three free-thiols (Cys-34, A2C and L585C) before treatment with DTNB.

Fig. 20 is a mass spectrogram of a rHA molecule designed to have three free-thiols (Cys-34, A2C and L585C) after treatment with DTNB.

Fig. 21 is a mass spectrogram of a rHA molecule designed to have four free-thiols (Cys-34, D129C, C360S and L585C) before treatment with DTNB.

Fig. 22 is a mass spectrogram of a rHA molecule designed to have four free-thiols (Cys-34, D129C, C360S and L585C) after treatment with DTNB.

Fig. 23 is a mass spectrogram of a rHA molecule designed to have three free-thiols (Cys-34, A2C and a C-terminal free-thiol) before treatment with DTNB.

Fig. 24 is a mass spectrogram of a rHA molecule designed to have three free-thiols (Cys-34, A2C and a C-terminal free-thiol).

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 resi-

due 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 posi-

tion 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 (con-

jugation 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).

Position with Leader Sequence: Refers to the position in the unprocessed form of human serum albumin containing the 24 amino acid secretory leader sequence.

Amino Acid: The standard one letter code for each of the 20 amino acids (*e.g.* A = Ala = Alanine).

% SASA: The solvent accessible surface area calculated for each residue, using the DSSP software described in Kabschand and Sander (1983). Biopolymers 22(12): 2577-637. Each solvent accessible surface area was divided by a standard value for the particular amino acid found in that position and multiplied by 100, thereby obtaining a percentage of the standard value for each residue. The standard solvent accessible surface areas for the 20 different amino acids are defined as (using one-letter codes for the amino acids): A=62, C=92, D=69, E=156, F=123, G=50, H=130, I=84, K=174, L=97, M=103, N=85, P=67, Q=127, R=211, S=64, T=80, V=81, W=126, Y=104).

B-Factor: The crystallographic B-factor value for the C-alpha atom was extracted directly from the PDB file. The B-factor is in column number 11 of the 1ao6 PDB file PDB, (http://www.rcsb.org/pdb/).

SS (Secondary Structure): The secondary structure determined for each residue using the DSSP software Kabsch and Sander (1983). Biopolymers 22(12): 2577-637. If the secondary structure is defined as H (Helix), B (isolated beta bridge) or E (Extended sheet), the residue is marked ’1’, otherwise as ’0’.

Align 1 (Mamm. W): The homology level for an alignment of various mammalian 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 W; six levels of homology are determined with the highest =100%, decreasing in 20% increments, to the lowest =0% (Fig. 2).

Adj. 100%'s (Align 1): The score according to whether the adjacent residue was highly (100%) conserved when HSA is aligned with the mammalian albumins of Fig. 2. A score of 0 indicates the residue is not adjacent to a residue with 100% homology when HSA is aligned with the mammalian albumins of Fig. 2; a score of 1 indicates that the residue is adjacent to one residue with 100% homology when HSA is aligned with mammalian albumins; a score of 2 indicates that the residue is adjacent to two residue with 100% homology when HSA is aligned with mammalian albumins.

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 con-

servation 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 albumin);

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 addi-

tion 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 %SASA may be preferred (Figure 6B, fifth column).

For clarity, for albumins other than the human serum albumin of SEQ ID No. 1, equivalent residues are favoured for mutation. Such equivalent residues may or may not have identical residue numbers to those of SEQ ID No. 1 but would be clearly identifiable by the skilled person, for example with reference to homology alignments with SEQ ID No. 1 and other albumins such as those of Fig. 2 and/or Fig. 3. For example, in Fig. 2, the residue at positions at 160 of the horizontal 'ruler' are equivalent but have differing residue numbers and, sometimes, are differing amino acids, *e.g.* L159 in human, W134 in goat fragment, L151 in Macacque and M159 in mouse. It is preferred that, for an alignment such as Fig. 2 or Fig.3, equivalent residues are within 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 amino acid of the candidate amino acid of SEQ ID No. 1. The 'ruler' above the alignment indicates whether an amino acid is within 1 to 10 amino acids of a candidate amino acid of SEQ ID No. 1.

The selection criteria of Group (A) are more preferable than those of Group (B) which in turn are more preferable than those of Group (C) and in turn are more preferable than those of Group (D).

The method may or may not further comprise determining the receptor binding capacity and/or the conjugation competence of the polypeptide and optionally selecting a polypeptide which does or does not have a receptor binding capacity and/or conjugation competence.

'Preparing' a polypeptide may or may not include expressing the polypeptide in a host cell and/or purifying the polypeptide from the host or host cell media. The method may com-

prise favouring selection of residues meeting one or all of the following criteria:

- residues having high surface accessibility are preferred to those having low surface accessibility;
- conservative mutations from another amino acid to cysteine are preferred over non-conservative mutations;

Alternatively, or in addition, selection criteria as detailed throughout this specification may or may not be used to select residues in the method of the first aspect of the invention.

A second aspect of the invention provides a thio-albumin comprising a polypeptide sequence and/or polypeptide designed and/or produced according to the first aspect of the invention.

Preferably the polypeptide is a recombinant polypeptide. Preferably the polypeptide is an isolated and/or purified polypeptide. Preferably the polypeptide is synthetic and/or does not naturally occur in nature.

Specifically, the invention provides a polypeptide which has an amino acid sequence which is at least 60% identical to residues 1 to 585 of SEQ ID No. 1 or a fragment or fusion thereof, in which:

(a) at a position equivalent to position 34 of SEQ ID No. 1, there is a conjugation competent cysteine residue; and

(b) elsewhere in the polypeptide there is one or more conjugation competent cysteine residues, preferably 2 or more.

In addition the invention provides a conjugation competent polypeptide comprising an amino acid sequence which is at least 60 % identical to residues 1 to 585 of SEQ ID No. 1, or a fragment or fusion thereof, 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 is one or more conjugation competent cysteine residues, preferably 2 or more or 3 or more.

The polypeptide may or may not comprise 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20 conjugation competent cysteine residues.

More specifically, the polypeptide (which may be described in relation to a known albumin sequence such as SEQ ID No. 1) may or may not comprise one or more of:

(a) substitution of a non-cysteine amino acid with a cysteine at a position corresponding to a position equivalent to any of residues L585, D1, A2, D562, A364, A504, E505, T79, E86, D129, D549, A581, D121, E82, S270, Q397 and A578 of SEQ ID No. 1;

(b) insertion of a cysteine at a position adjacent the N- or C- side of an amino acid which may or may not correspond to a position equivalent to any of residues L585, D1, A2, D562, A364, A504, E505, T79, E86, D129, D549, A581, D121, E82, S270, Q397 and A578 of SEQ ID No. 1;

(c) a cysteine with a free thiol group at a position which may or may not correspond to any of C369, C361, C91, C177, C567, C316, C75, C169, C124 and C558 which may or may not be generated by deletion or substitution of C360, C316, C75, C168, C558, C361, C91, C124, C169 and/or C567.

(d) addition of a cysteine to the N- side of the N-terminal residue of an albumin sequence and/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), (c) and (d) 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. 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 the C- or N- terminal amino acid of the albumin, such as to the C-side of L585 in SEQ ID No. 1.

The polypeptide may or may not further comprise a further linker to which a conjugation partner, such as a bioactive compound, may be linked. For example a linker may comprise a primary amine such as a lysine.

One or more conjugation competent cysteines may or may not be created in an albumin by *insertion*; for example by adding one or more additional cysteines without removal of an amino acid residue from the albumin sequence, or by substituting one or more adjacent amino acids with a larger number of residues containing at least one cysteine, thus extending the overall length of the polypeptide. For example, a cysteine residue may be introduced immediately adjacent an albumin residue identified herein. The cysteine residue may be introduced as a single cysteine residue or within a polypeptide. The polypeptide may be from 2 to 50 amino acids long, preferably from 2, 10, 20, 30, or 40 to 10, 20, 30, 40 or 50 amino acids long.

Alternatively, or in addition, the invention includes substitution of one of the cysteine residues in one or more disulphides bond of an albumin with a different amino acid residue, so breaking the disulphide bond to leave an additional free thiol group. For example, a cysteine of one or more of the 17 naturally occurring disulphide bonds of HSA may be substituted

to provide a conjugation-competent cysteine. Such a substitution causes the cysteine which has not been substituted to no longer have a disulphide binding partner and therefore provide a free thiol group. Conjugation competent cysteines may be provided from one or more of the naturally occurring disulphide bonds of an albumin such as 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 or 17 of the naturally occurring disulphide bonds of an albumin such as HSA (*e.g.* SEQ ID No. 1). For example, one cysteine residue which naturally forms a disulphide bond with another cysteine residue may or may not be substituted with a relatively conserved amino acid residue, particularly Ser, Thr, Val or Ala. With reference to SEQ ID No. 1, cysteine residues 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. Cysteine residues preferred for modification (*i.e.* deletion or substitution) may in particular correspond to C360, C316, C75, C168, C558, C361, C91, C124, C169 and/or C567 thus generating a conjugation competent cysteine at one or more of C369, C361, C91, C177, C567, C316, C75, C169, C124 and C558 of SEQ ID No. 1.

In addition, conjugation competent cysteines may or may not be created in albumin by *deletion* of one of the cysteines of a disulphide bond in the protein structure, so breaking the disulphide bond to provide an additional free thiol group.

Alternatively, or in addition, one or more of the cysteine residues present in the albumin molecule, but not involved in disulphide bonding (*e.g.* Cys-34 in the case of SEQ ID No. 1) may or may not be deleted (*i.e.* without substitution with a different amino acid) or may or may not be substituted with a different amino acid, particularly Ser, Thr, Val or Ala.

For a polypeptide comprising two or more conjugation competent cysteine residues, when the polypeptide is folded, the conjugation competent cysteine residues may or may not be relatively evenly distributed over the surface of the folded protein. The term ‘folded’ includes folding of a polypeptide/protein into its natural configuration, for example the most thermodynamically stable folded configuration. An advantage of relatively even distribution is that it allows conjugation of two or more moieties to the thio-albumin without steric hindrance between two or more of the conjugated moieties. This has the advantage of minimising, and optionally eliminating, potential loss of activity due to issues such as steric hindrance between adjacent moieties (conjugation partners) which may be conjugated to the thio-albumin. Such moieties, for example bioactive molecules, may be relatively bulky.

Preferably the two or more conjugation competent cysteines are distributed over the surface of the thio-albumin molecule such that they are spaced as far from each other as possible, for example geometrically possible. Preferably the distance between two or more

conjugation competent cysteines is at least 10, 20, 30, 40, 50, 60, 70, or 80 Angstroms. Preferably each conjugation competent cysteine is at least 10, 20, 30, 40, 50, 60, 70, or 80 Angstroms distant from all other conjugation competent cysteines in the molecule. The distance between two conjugation competent cysteines is preferably a distance which is at least 10, 20, 30, 40, 50, 60, 70, 80, 90, or 95 % and most preferably 100% of the length of the longest axis of the folded albumin molecule, for example as shown in a model of an albumin. For example, the longest axis of SEQ ID No. 1 as shown in protein structure 1AO6 is approximately 85 Angstroms. Therefore, it is preferred that the two or more of the cysteine residues are at least 65, 70, 75 or most preferably 80 Angstroms apart. Most preferably each conjugation competent cysteine residue is at a distance of at least 80, 90, or 95 % and most preferably 100% of the length of the longest axis of the folded albumin molecule.

Preferably the side chains of conjugation competent cysteines are directed away from each other and/or directed so that a moiety conjugated to the cysteine will be directed away from the centre of the albumin structure. This provides the advantage of preventing interactions between the conjugated moieties and/or the albumin moiety itself.

With reference to an amino acid sequence, candidate amino acid residues may be visually inspected using software such as PyMOL (Warren L. DeLano "The PyMOL Molecular Graphics System" DeLano Scientific LLC, San Carlos, CA, USA. http://www.pymol.org). Candidate amino acids may be divided into categories based on their proximity to other members of that group. For example, candidate amino acids may be divided into 2, 3, 4, 5, 6, 7, 8, 9 or 10 categories. It is preferred that combinations of candidate amino acids are selected from different categories. That is, it is preferred that a thio-albumin contains one or fewer mutations from each category.

With reference to SEQ ID No. 1, PyMOL was used to analyse the candidate residues of Figs 5A and 5B in order to identify particularly favoured combinations of cysteine mutations. Such combinations may be used to design a thio-albumin having two or more conjugation competent cysteine residues. Selection Groups 1 and 2 correspond to the selection criteria (A) and (B) (respectively) from Figs 5A and 5B of the selection method described herein. Selection Group 3 corresponds to the residues identified in Fig. 6B. Particularly favoured residues are given in Figure 6A and 6B in which the column headings are the same as those in Fig. 1 with the addition of ‘Selection Group’ and ‘Proximity Group’ as described herein.

The results of the analysis are given in Fig. 10 in which column headings are the same as those used in Fig. 1, and Fig. 6 with the addition of:

Proximity Group: allocation of a proximity group as described herein to descrive subsets of sites within HSA (specifically SEQ ID No. 1).

For example, candidate amino acid residues selected in Selection Group 1 (listed in Figure 5A) were visually inspected using the PyMOL software, and the amino acids selected were divided into categories based on their proximity to other members of Selection Group 1. Five groups were generated (labeled A to E in Figure 10 'proximity group', right hand column), four were generated by visual inspection. Group E contains amino acid residues not visible in 1AO6 which are known to be in the N-terminal region. In addition it was observed that cys-34 is present in the proximity group A.

Similarly amino acid residues selected in Selection Group 2 (listed in figure 5B) were visually inspected using the PyMOL software, and the amino acids selected were divided into categories based on their proximity to other members of Selection Group 2. Five groups were generated (labeled F-J in Figure 10 'proximity group', right hand column)

Similarly, the preferred free cysteine residues selected in Selection Group 3 (listed in Figure 6B), which can be generated by mutations causing disruption of disulphide bonds were visually inspected using the PyMOL software, and the amino acids selected were divided into categories based on their proximity to other members of that group. Four groups were generated (labeled K-N in Figure 10 'proximity group', right hand column). When referring to the residues of selection group 3, the cited residues are the resultant conjugation competent cysteines (*e.g.* Fig. 6B). In order to generate such a conjugation competent cysteine it is 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.
- The above mentioned albumin variants may or may not further comprise a cysteine at Cys34 of SEQ ID No. 1, or at an equivalent potision in another albumin.

Examples of preferred mutations selected from within Selection Group 2 may include the following:

- For 2 amino acid residues selected from Selection Group 2; amino acid residues which occur in proximity groups G + I such as S270 + A581 are preferred. Alternatively, amino acid residues which occur in proximity groups G + H such as S270 + D129 are preferred.
- For 3 amino acid residues selected from Selection Group 2; amino acid residues which occur in proximity groups G + I + F such as S270 + A581 + E82 are preferred. Alternatively, amino acid residues which occur in proximity groups G + I + H such as S270 + A581 + D129 are preferred.
- For 4 amino acid residues selected from Selection Group 2; amino acid residues which occur in proximity groups G + I + F + H such as S270 + A581 + E82 + D129.
- For 5 amino acid residues selected from Selection Group 2; amino acid residues which occur in proximity groups G + I + F + H + J such as S270 + A581 + E82 + D129 + Q397 are preferred. However, D549 is not preferred in combination with mutations

A578, A581. Also, mutations to D549, A578, A581 or are not preferred in combination with mutation of L585 from Selection Group 1.

Examples of preferred site selected from within Selection Group 3 for the conjugation competent free-thiols may include the following:

- For 2 amino acid residues selected from Selection Group 3; amino acid residues which occur in proximity groups M + L are preferred, such as C369 + C177. Similarly, C361 + C124 are also preferred.
- More than two mutations disrupting disulphide bonds are less preferred.
- The above mentioned albumin variants may or may not further comprise a cysteine at Cys34 of SEQ ID No. 1, or at an equivalent potision in another albumin.

Combinations of sites from Selection Groups 1, 2 and 3 can also be made, where sites from Selection Group 1 are typically preferred to sites from Selection Group 2, which are typically preferred to sites selected from Selection Group 3.

Examples of sites from Selection Groups 1 + 2 may include residues from proximity groups C + I, such as A364 + A581. Alternatively, residues from proximity groups A + G + I, such as C34 + S270 + A581, from proximity groups A + H + G + I, such as C34 + D129 + S270 + A581, from proximity groups A + C + I, such as T79 + A364 + A581, or residues from proximity groups C + I + H such as A364 + A581 +D129 are also preferred.

Examples of sites from Selection Groups 1 + 3 may include residues from proximity groups A + L + M, such as C34 + C169 + C316, from proximity groups C + L, such as A364 + 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 natural leader sequence of HSA (amino acids 1 to 24 of SEQ ID No. 102) or an HSA with a fusion leader sequence (amino acids 1 to 24 of SEQ ID No. 49).

A fourth aspect of the invention provides a plasmid comprising the polynucleotide of the third aspect of the invention. The plasmid may be a 2 micron based plasmid such as those described in WO2005/061719, WO2005/061718 and WO2006/067511 (all incorporated herein by reference). The plasmid may exhibit enhanced chaperone activity, for example through over expression of a chaperone, particularly PDI.

A fifth aspect of the invention provides an expression system such as a host cell comprising a polynucleotide according to the third aspect of the invention and/or a plasmid of the fourth aspect of the invention. Preferably the host cell is a mammalian cell such as a human or bovine cell, or a fungal cell such as a yeast cell. Alternatively, the host cell may be a bacterial cell such as a *Bacillus* or *Escherichia coli* or a viral cell such as Baculovirus or a plant cell such as a rice *e.g. Oryza sativa*. Most preferably, the cell is a yeast cell such as a *Saccharomyces* (*e.g. S. cerevisiae*), a *Pichia* or an Aspergillus cell.

A sixth aspect of the invention provides a conjugate which comprises a conjugation

partner, such as a bioactive compound, and a polypeptide according to the invention, wherein the conjugation partner is linked to the polypeptide through a conjugation competent cysteine residue of the polypeptide. The conjugation partner may be a therapeutic, diagnostic or imaging compound such as those mentioned herein. The conjugate may comprise 2 or more, for example 2, 3, 4, 5, 6, 7,8, 9 or 10, conjugation partners which may each be different and/or may be multiple copies of the same compound. Preferably, each conjugation partner is linked to the polypeptide through a conjugation competent cysteine residue of the polypeptide, however conjugation partners may be linked by other means for example by a genetic fusion or covalent bonds to non-cysteine amino acids such as lysine.

A seventh aspect of the invention provides a method of producing a polypeptide of the invention comprising:

(a) culturing a host cell according to the invention under conditions that allow expression of the polypeptide; and

(b) recovering the polypeptide from the host cell and/or from host cell growth medium.

Accordingly, the present invention also provides a method for producing a polypeptide (or protein) of the invention, the method comprising: (a) providing a host cell of the invention comprising a polynucleotide encoding protein product of choice as defined above; and (b) growing the host cell (for example, culturing the host cell in a culture medium); thereby to produce a cell culture or recombinant organism comprising an increased level of the protein product of choice compared to the level of production of the protein product of choice achieved by growing (for example, culturing), under the same conditions, the same host cell that has not been genetically modified to cause over-expression of one or more helper proteins.

The step of growing the host cell may or may not involve allowing a host cell derived from a multicellular organism to be regrown into a multicellular recombinant organism (such as a plant or animal) and, optionally, producing one or more generations of progeny therefrom.

The method may or may not further comprise the step of purifying the thus expressed protein product of choice from the cultured host cell, recombinant organism or culture medium.

The production method may comprise linking a conjugation partner to the polypeptide of the invention through a conjugation competent cysteine residue of the polypeptide. Suitable conjugation methods and conjugation partners are described herein.

An eighth aspect of the invention provides a composition comprising a conjugate ac-

cording to the invention and at least one pharmaceutically acceptable carrier and/or diluent.

A ninth aspect of the invention provides a method for making a pharmaceutical ingredient and/or a pharmaceutical product comprising making a thio-albumin according to the present invention, optionally conjugating a further molecule to the thio-albumin, optionally formulating the resultant conjugate with a pharmaceutically acceptable diluent and/or carrier and optionally preparing the product in unit dosage form.

A tenth aspect of the invention provides use of a polypeptide according to the invention for the production of a thio-albumin-conjugate.

An eleventh aspect of the invention provides use of a conjugate according to the invention and/or produced by a method according to the invention for treatment of disease, treatment of illness and/or diagnosis.

A twelfth aspect of the invention provides a gel comprising one or more albumins according to the invention. The gel may be formed by any suitable method. For example the gel may be formed by incubating an albumin solution, or suspension, at a suitable temperature *e.g.* room temperature (15 to 25 °C, such as 20 °C) or body temperature (36 to 38 °C, preferably 36.9°C). A gel may be used to coat medical devices, such as a stent. A gel may be used in or on a wound dressing. The albumin may be applied to the medical device or wound dressing before or after it has gelled. The albumin may be applied *ex situ* or *in situ* (*e.g.* applied to a medical device or dressing before, after or during its application on to or insertion into a human or animal body).

The polypeptides and/or conjugates of the invention may be used to prepare nanoparticles which may be used, for example, in angiogenic applications, anti-angiogenic applications and to coat a medical device such as a stent. Nanoparticles are effective at targeting, for example to non tight–junctions, and therefore can be useful for targeting tumours such as cancerous tumours. Nanoparticles can also be useful to target antigen in order to provoke an immune response since nanoparticles are particularly susceptible to engulfment and presentation by phagocytes. The invention provides nanoparticles consisting only of thio-albumin according to the invention which may or may not be conjugated to a moiety (conjugation partner). The invention also provides nanoparticles comprising thio-albumin according to the invention, which may or may not be conjugated to a moiety, and one or more other constituents of a nanoparticle which may or may not be albumin related. In a preferred embodiment, a thio-albumin according to the invention comprises at least two conjugation competent cysteine residues located on the surface of the polypeptide. Such a thio-albumin may be used for the preparation of nanoparticles in which one or more conjugation competent cysteine residues may be used in the formation of a nanoparticle and one or more conjuga-

tion competent residue is used for conjugation to a conjugation partner, for example to a bioactive molecule.

The invention relates to all albumins. Whilst preferred residues have been identified in relation to SEQ ID No. 1, the skilled person would be able to identify equivalent residues in other albumin sequences, such as the albumins disclosed in Figures 2 and 3, and understand that mutations of albumins (other than SEQ ID No. 1) at such equivalent residues are part of the invention. Equivalent residues can be identified by, for example, homology alignment with SEQ ID No. 1. A residue in an albumin other than SEQ ID No. 1 may or may not have an identical residue coordinate to its equivalent residue in SEQ ID No. 1. Thus the invention provides thio-albumins based on any albumin sequence, such as the sequences shown in Table 1 and, more preferably, those shown in Fig. 2 and/or 3. 'Based on' includes modification of an albumin sequence to introduce one or more additional free-thiols.

Recombinant albumins can offer advantages over animal-derived albumins by having a higher level of conjugation-competent free thiol groups, and can be manufactured without the risk of contamination with pathogenic prions and viruses. An advantage of a thio-albumin conjugate is that the thio-albumin part may be prepared separately to a conjugation partner. Therefore, one batch of thio-albumin may be used to produce many different thio-albumin conjugates. Also, the individual components of the conjugate can be manufactured by different methods and therefore are not restricted to a single method, such as heterologous protein expression in a host cell such as a yeast. Furthermore, a thio-albumin may comprise multiple conjugation sites and therefore a single thio-albumin may be conjugated to more than one type of conjugation partner (*e.g.* therapeutic agent, diagnostic agent, targeting agent, imaging agent) and/or to multiple copies of one or more types of conjugation partner. The ability to conjugate the thio-albumin to different types of conjugation partners allows the provision of a multi-functional species. The ability to conjugate the thio-albumin to multiple copies of a conjugation partner allows the concentration of molecule to be increased and therefore increase the amount, or volume, of thio-albumin conjugate required for a given purpose relative to a conjugate having only a single copy of the conjugation partner. Advantages of delivering drugs *via* an albumin fusion protein are discussed in Osborn, *et al*. (2002). J Pharmacol Exp Ther 303(2): 540-8. It is expected that delivery of drugs *via* a conjugation of the invention would have similar advantages.

Further details which may or may not be used in accordance with the invention are described below:

Three dimensional (3D) Models

The above disclosure has been made in relation to the albumin model known as 1A06 (Protein Data Bank) which relates to SEQ ID No. 1. Figure 1 gives the amino acid residues for 1AO6.

However, the invention relates to all albumins and their structures. Structures of albumin are available to the skilled person, for example the atomic coordinates for the tertiary structure of human albumin are available at the GenBank DNA database at www.ncbi.nlm.nih.gov.

Structures may be viewed using suitable software such as RasM.1 Chime (Sayle, TIBS 20, 374, 1995). Available albumin coordinates include:

1AO6, **1BM0** (Sugio *et al*. (1999). Protein Eng **12**(6): 439-46), which was among the top 17 requested proteins.

1UOR, He & Carter (1992). Nature **358**(6383): 209-15.

1bj5 ***and*** **1bke**, Curry *et al*. (1998). Nat Struct Biol **5**(9): 827-35.

1e7a,**1e7b, 1e7c,** Bhattacharya *et al*. (2000). J Biol Chem **275**(49): 38731-8.

1e7e, **1e7f**, **1e7g**, **1e7h** and **1e7i**, Bhattacharya *et al*. (2000). J Mol Biol **303**(5): 721-32.

1GNJ, Petitpas *et al*. (2001). J Mol Biol **314**(5): 955-60.

1HA2 and **1H9Z** Petitpas *et al*. (2001). J Biol Chem **276**(25): 22804-9.

Albumin

The albumin used in the invention may be a naturally occurring albumin, an albumin-related protein or a variant thereof such as a natural or engineered variant. Variants include polymorphisms, fragments such as domains and sub-domains, fragments and/or fusion proteins. An albumin, of this invention, may comprise the sequence of an albumin protein obtained from any source. Typically the source is mammalian such as human or bovine. In one preferred embodiment the serum albumin is human serum albumin ("HSA"). The term "human serum albumin" includes a serum albumin having an amino acid sequence naturally occurring in humans, and variants thereof. Preferably the albumin has the amino acid sequence of SEQ ID No. 1 or a variant or fragment thereof, preferably a functional variant or fragment thereof. The HSA coding sequence is obtainable by known methods for isolating cDNA corresponding to human genes, and is also disclosed in, for example, EP 0 073 646 and EP 0 286 424. A fragment or variant may or may not be functional. For example, a fragment or variant may retain the ability to bind to an albumin receptor such as FcRn to at least 10, 20, 30, 40, 50, 60, 70, 80, 90 or 100 % of the ability of the parent albumin (from which the frag-

ment or variant derives) to bind to the receptor. Relative binding ability may be determined by methods known in the art such as surface plasmon resonance studies.

The albumin may be a naturally-occurring polymorphic variant of human albumin or of a human albumin analogue. Generally, variants or fragments of human albumin will have at least 5%, 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, (preferably at least 80%, 90%, 95%, 100%, 105% or more) of human albumin's ligand binding activity (for example FcRN-binding), mole for mole.

The “albumin” may comprise the sequence of bovine serum albumin. The term “bovine serum albumin” includes a serum albumin having an amino acid sequence naturally occurring in cows, for example as taken from Swissprot accession number P02769, and variants thereof as defined herein. The term “bovine serum albumin” also includes fragments of full-length bovine serum albumin or variants thereof, as defined herein.

A number of proteins are known to exist within the albumin family; a non-exclusive list is shown in Table 1, below. The list indicates full-length of sequences including the mature protein and leader sequence (unless indicated otherwise).

The albumin may comprise the sequence of an albumin derived from one of serum albumin from dog (*e.g.* see Swissprot accession number P49822-1), pig (*e.g.* see Swissprot accession number P08835-1), goat (*e.g.* as available from Sigma as product no. A2514 or A4164),), cat (*e.g.* see Swissprot accession number P49064-1), chicken (*e.g.* see Swissprot accession number P19121-1), ovalbumin (*e.g.* chicken ovalbumin) (*e.g.* see Swissprot accession number P01012-1), turkey ovalbumin (*e.g.* see Swissprot accession number O73860-1), donkey (*e.g.* see Swissprot accession number Q5XLE4-1), guinea pig(*e.g.* see Swissprot accession number Q6WDN9-1), hamster (see DeMarco *et al*. (2007). International Journal for Parasitology 37(11): 1201-1208), horse (*e.g.* see Swissprot accession number P35747-1), rhesus monkey (*e.g.* see Swissprot accession number Q28522-1), mouse (*e.g.* see Swissprot accession number P07724-1), pigeon (*e.g.* as defined by Khan *et al*, 2002, *Int. J. Biol. Macromol.*, **30**(3-4),171-8), rabbit (*e.g.* see Swissprot accession number P49065-1), rat (*e.g.* see Swissprot accession number P02770-1) and sheep (*e.g.* see Swissprot accession number P14639-1) and includes variants and fragments thereof as defined herein.

The albumin may comprise the sequence of an albumin such as a serum albumin or an ovalbumin, for example those shown in Table 1, below, and includes variants and fragments thereof as defined herein.

Table 1: Albumins from various species

Protein	Common Name	Species	SwissProt Accession No	Identity to SEQ ID NO: 1 (Clustal V)	Length (aa)
SA	African clawed frog	*Xenopus laevis*	P08759-1	37.3	608
SA	Bovine	*Bos taurus* (SEQ ID No. 94)	P02769-1	76.1	607
SA	Cat	*Felis catus* (SEQ ID No. 95)	P49064-1	82.2	608
SA	Chicken	*Gallus gallus*	P19121-1 (Version 2 text)	46.5	615
SA	Cobra ALB	*Naja kaouthia*	Q91134-1	30.9	614
SA	Dog	*Canis lupus familiaris* (SEQ ID No. 96)	P49822-1	80.3	608
SA	Donkey	*Equus asinus* (SEQ ID No. 97)	Q5XLE4-1	76.6	607
SA	European water frog	*Rana shqiperica*	Q9YGH6-1	31.5	603
SA	Blood fluke	*Schistosoma mansoni*	AAL08579 Q95VB7-1	76.0	608
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

Protein	Common Name	Species	SwissProt Accession No	Identity to SEQ ID NO: 1 (Clustal V)	Length (aa)
SA	Australian Lung-fish	*Neoceratodus forsteri*	P83517-1	22.8	101 (NL)
SA	Macaque (Rhe-sus 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 cuni-culus* (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 macu-latum*	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 mari-nus*	Q91274-1	16.9	1423
SA	Sea lamprey	*Petromyzon mari-nus –AS*	O42279-1	17.4	551
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)

Protein	Common Name	Species	SwissProt Accession No	Identity to SEQ ID NO: 1 (Clustal V)	Length (aa)
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 atleast 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

to the appropriate amino acid. Thus the score for cysteine is 0 as no boundaries are crossed. Likewise, the score of aspartic acid (D) is 3, since 3 boundaries are crossed.

The conservation mutation score (with respect to Fig. 4) for the 20 different amino acids are defined as (using one-letter codes for the amino acids): A=1, C=0, D=3, E=4, F=4, G=2, H=5, I=4, K=4, L=4, M=3, N=2, P=3, Q=3, R=5, S=1, T=1, V=3, W=3, Y=3. With reference to Figs. 1, 5A, 5B, 5C, and 5D, these scores are provided for each of the amino acid residues in the column labelled 'Conserved Mutation to Cysteine'. Using the conservation mutation score residues with a score of 3 or less, *i.e.* aspartic acid methionine, proline, glu-

tamine, valine, tryptophan, tyrosine, glycine, asparagine, alanine, serine and threonine are preferred since they are relatively conserved with cysteine. More preferred are those amino acids with a score of 2 or less *i.e.* glycine, asparagine, alanine, serine, threonine. Most preferred are those with a score of 1, *i.e.* alanine, serine, threonine. Similarly using the conservation mutation score system of Fig. 4, residues with a score of 4 or more, *i.e.* glutamic acid, phenylalanine, isoleucine, lysine, leucine, histidine and arginine are less preferred and may not be preferred at all.

Alternatively, or in addition, "conservative" amino acid substitutions refers to substitutions made within the same group such as within the group of basic amino acids (such as arginine, lysine, histidine), acidic amino acids (such as glutamic acid and aspartic acid), polar amino acids (such as glutamine and asparagine), hydrophobic amino acids (such as leucine, isoleucine, valine), aromatic amino acids (such as phenylalanine, tryptophan, tyrosine) and small amino acids (such as glycine, alanine, serine, threonine, methionine).

For example, a conservative substitution of alanine-2 in SEQ ID No 1 can include glycine or serine. Non-conservative substitutions encompass substitutions of amino acids in one group by amino acids in another group. For example, a non-conservative substitution could include the substitution of a polar amino acid for a hydrophobic amino acid.

Fragment

The term "fragment" as used herein includes any fragment of full-length albumin or a variant thereof, so long as at least one basic property, for example binding activity (type of and specific activity *e.g.* binding to bilirubin), 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. A fragment will typically be at least 50 amino acids long. A fragment may comprise at least one whole sub-domain of albumin. Domains of HSA have been expressed as recombinant proteins (Dockal *et al*., 1999, J. Biol. Chem., 274, 29303-29310), where domain I was defined as consisting of amino acids 1-197, domain II was defined as consisting of amino acids 189-385 and domain III was defined as consisting of amino acids 381-585. Partial overlap of the domains occurs because of the extended α-helix structure (h10-h1) which exists between domains I and II, and between domains II and III (Peters, 1996, op. cit., Table 2-4). HSA also comprises six sub-domains (sub-domains IA, IB, IIA, IIB, IIIA and IIIB). Sub-domain IA comprises amino acids 6-105, sub-domain IB comprises amino acids 120-177, sub-domain IIA comprises amino acids 200-291, sub-domain IIB comprises amino acids 316-369, sub-domain IIIA comprises amino acids 392-491 and sub-domain IIIB comprises amino acids 512-583. A fragment may comprise a whole or part

of one or more domains or sub-domains as defined above, or any combination of those domains and/or sub-domains. A fragment may comprise or consist of at least 50, 60, 70, 75, 80, 85, 90, 95, 96, 97, 98, or 99 % of an albumin or of a domain of an albumin.

Additionally, single or multiple heterologous fusions comprising any of the above; or single or multiple heterologous fusions to albumin, or a variant or fragment of any of these may be used. Such fusions include albumin N-terminal fusions, albumin C-terminal fusions and co-N-terminal and C-terminal albumin fusions as exemplified by WO 01/79271.

Homology

Figures 2 and 3 show alignments of various albumin family proteins with HSA (SEQ ID NO: 1), identified as ‘P02768’. The protein sequences include the albumin leader sequence. These alignments can be used to identify conserved regions and amino acid residues corresponding to those in HSA selected as described above. One or both alignments can also be used to assign a homology score to an amino acid residue in an albumin sequence.

An example of such a procedure is the MegAlign program (version 8.0.2) developed by DNASTAR Inc., part of the Lasergene suite, based on Hein, J.J. (1990) “Unified approach to alignment and phylogenies.” In Methods in Enzymology, Vol. 183: pp. 626-645. Using the Jotun Hein Method and the settings GAP PENALTY = 11, GAP LENGTH PENALTY = 3 for multiple alignments and KTUPLE = 2 for pairwise alignments a series of percentage identity values can be calculated. Alternatively the Clustal V Method and the settings GAP PENALTY = 10, GAP LENGTH 10 for multiple alignments and KTUPLE = 1. GAP PENALTY = 3, WINDOW=5 and DIAGONAL=5 for pairwise alignments a series of percentage identity values can be calculated. Alternatively the Clustal W Method and the settings GAP PENALTY = 10, GAP LENGTH PENALTY = 0.2, DELAY DIVERGENCE =30 DNA transition =0.5 and using GONNET SERIES for Protein Weight matrix and IUB for DNA Weight matrix for multiple alignments, and Slow accurate, GAP PENALTY = 10, GAP LENGTH PENALTY = 0.1, and using GONNET SERIES for Protein Weight matrix and IUB for DNA Weight matrix for pairwise alignments a series of percentage identity values can be calculated. Alternatively the Clustal V method may be used (above).The alignment of two amino acid sequences may also be determined by using the Needle program from the EMBOSS package (http://emboss.org) version 2.8.0. The Needle program implements the global alignment algorithm described in Needleman and Wunsch (1970) "A general method applicable to the search for similarities in the amino acid sequence of two proteins." J. Mol. Biol. 48, 443-453. The substitution matrix used is BLOSUM62, gap opening penalty is 10, and gap extension penalty is 0.5.

Fig. 2 shows an alignment of sixteen mammalian albumin family proteins including HSA (SEQ ID NO: 1, identified in the alignment as P02768) compiled using MegAlign program (version 8.0.2) based on Clustal W. The protein sequences include the albumin leader sequence. Each sequence is labelled with the animal from which it derives and its database accession number.

Fig. 3 shows alignments of thirty three albumin family (both mammalian and non-mammalian) proteins including HSA (SEQ ID NO: 1, identified in the alignment as P02768) compiled using MegAlign program (version 8.0.2) based on Clustal V. The protein sequences include the albumin leader sequence.

Homology may be determined with reference to Fig 2 and/or Fig 3. The degree of identity between a given amino acid sequence and SEQ ID NO: 1 may be calculated as the number of exact matches in an alignment of the two sequences, divided by the length of the shorter of the two sequences. The result may be expressed in percent identity. An exact match occurs when the two sequences have identical amino acid residues in the same positions of the overlap. The length of a sequence is the number of amino acid residues in the sequence.

Conservation of adjacent residues

In addition, individual amino acid residues of HSA can be ranked according to their conservation to the amino acids of other albumin family proteins at the same position. Figure 1, column labelled 'Align 1 (Mamm. W ('mammalian, Clustal W)) provides the homology level for each position of SEQ ID No. 1 as calculated by the alignment of mammalian albumins given in Fig. 2. The homology level score may be calculated. One method to score the homology level, which is used herein, is calculated by using the strength of the histogram provided by MegAlign program (version 8.0.2) (ranging 0 to 100%); six levels of homology are determined with the highest =100%, decreasing in 20% increments to the lowest (0%) and shown by bar height in Figure 2. A score of 100 is the highest conservation and indicates there are no changes at that residue when the sequence from human serum albumin is compared with other mammalian albumin sequence, whereas a score of 0 indicates the lowest level of conservation between the aligned sequences.

Similarly, the homology level score for each amino acid residue in HSA (Figure 1, column labelled 'Align 2 (Var. Sps. V ('various species (*i.e.* mammalian and non-mammalian), Clustal V')) calculated using the strength of the histogram provided by Megalign when various albumins (including non-mammalian albumins) are aligned using Clustal V. A person skilled in the art will appreciate that a range of different albumin sequences and

alignment algorithms may be used to calculate the homology level score.

When using homology scores, such as those described in reference to the alignments of Figs 2 and 3, to identify candidate residues for the present invention, preferred residues include those which are not highly conserved (for example those with a score of less than 40, more preferably less than 20 and most preferably 0) are preferred and those with a higher level of homology (for example those with a score of more than 40, more than 60, more than 80 and most preferably 100) are less preferred.

Each of the amino acid residue in HSA (SEQ ID No. 1) were scored according to whether the adjacent residue was highly (100%) conserved when HSA is aligned with mammalian albumins (Figure 1, column labelled 'Adj. 100%'s (Align 1). This is because if an amino acid is within a highly conserved domain, it may be important to the structure of function of the protein and thus disruption may be undesirable. In Figure 1, a score of 0 indicates the residue is not adjacent to any residue with 100% homology (with reference to the alignment of Fig. 2) when HSA is aligned with mammalian albumins; a score of 1 indicates that the residue is adjacent to one residue with 100% homology when HSA is aligned with mammalian albumins; a score of 2 indicates that the residue is adjacent to two residues with 100% homology when HSA is aligned with mammalian albumins. Residues with a score of 0 or 1 are preferred. Residues with a score of 0 are most preferred.

For example, amino acid residues with a score of 2 (such as valine-7 (V7)) are preferably deselected using the method of the invention since these amino acid residues were assumed to occur in a region of high homology which would be unlikely to accept a mutation to an alternative amino acid. Similarly, phenylalanine-11 (F11) is adjacent to one 100% conserved residue, in a region of conserved residues, and is less preferred to a residue, such as alanine-2 (A2), which has no adjacent 100% conserved residues.

In accordance with the invention, additional information such as preferred sites for insertion, deletions or substitutions may also be obtained by alignment analysis. For example, mouse albumin contains 36 cysteine residues, all the cysteines involved in disulphide bonding (by homology to HSA) are present, as is cysteine-34, however a cysteine residue is present at 579 on mature mouse protein but not other mammalian albumin sequences therefore thio-albumin mutein S579C may be preferred as its lack of homology with other mammalian albumins suggests that it may not be particularly important to the structure and/or function of this albumin.

In addition, using the alignments of various mammalian albumin family and Clustal W (Fig. 2) shows that compared to other mammalian albumins, gerbil albumin has an additional alanine residue between alanine-2 (A2) and histidine-3 (H3), indicating that insertion of a

cysteine residues after residue 2 (*e.g.* A2 of SEQ ID No. 1) and before residue 3 (*e.g.* H3 of SEQ ID No. 1) is preferred.

Compared to other mammalian albumins, guinea pig albumin has a serine residue at cysteine-34 (C34). In examples where deletion of the free thiol group at cysteine-34 is required, a mutation such as C34S may be preferred.

Most mammalian albumin sequences (with the exception of human serum albumin) have a sequence which is less than or equal to 584 amino acids in length (less than or equal to 608 amino acids including leader sequence). Using the alignment in Fig. 2 the additional amino acid residue present on human serum albumin appears to be at the C-terminus without any cognate alignment amino acid residues in the other mammalian serum albumin sequences. Thus a thio-albumin variant containing G584C and a deletion of L585 may be preferred.

A number of albumin sequences (Bovine, Donkey, Goat, Horse, Sheep, Pig)are 583 amino acids in length (607 amino acids including leader sequence). Using the alignment in Fig. 2, it can be seen that these species albumin sequences do not have a residue corresponding R117 (R141 including leader sequence) therefore a thio-albumin containing V116C and a deletion of R117 or a thio-albumin containing a deletion of R117 and P118C may be preferred. In such a thio-albumin the length of the amino acid sequence would be reduced relative to SEQ ID No. 1.

The alignment used in Fig. 2 and the conclusions drawn are in particular with reference to mammalian albumins and Clustal W, the skilled person will appreciate that the teaching applies likewise to other members of the albumin family and to alternative alignment algorithms.

Alignment and identity

The albumin variant may have at least 40 % identity with SEQ ID NO: 1, particularly at least 45%, 50%, 55%, 60%, 65%, 70%, 75 %, at least 80 %, at least 85 %, at least 90 %, at least 95 %, at least 97 %, at least 98 % or at least 99 % identity.

Identity may be calculated using any method, for example those described herein.

Conjugation

The thio-albumin may optionally be fused to one or more conjugation partners for example through a genetic or chemical fusion. For a thio-albumin which comprises a genetic fusion, the fusion may be at the N- or C-terminus or comprise an insertion.

With respect to genetic fusions of albumin, the skilled person will also appreciate

that the open reading frame of any other gene or variant, or part or either, can be utilised as an open reading frame for use with the present invention. For example, the open reading frame may encode a protein comprising any sequence, be it a natural protein (including a zymogen), or a variant, or a fragment (which may, for example, be a domain) of a natural protein; or a totally synthetic protein; or a single or multiple fusion of different proteins (natural or synthetic). Such proteins can be taken, but not exclusively, from the lists provided in WO 01/79258, WO 01/79271, WO 01/79442, WO 01/79443, WO 01/79444 and WO 01/79480, or a variant or fragment thereof; the disclosures of which are incorporated herein by reference. Although these patent applications present the list of proteins in the context of fusion partners for albumin, the present invention is not so limited and, for the purposes of the present invention, any of the proteins listed therein may be presented alone or as fusion partners for albumin or any other protein or fragment or variant of any of the above, as a desired polypeptide. Examples of chemical fusions (also known as conjugations) of albumin are given in Leger *et al*. (2004) Bioorg Med Chem Lett 14(17): 4395-8; and Thibaudeau *et al*. (2005). Bioconjug Chem 16(4): 1000-8.

An advantage of using a genetically or chemically fused albumin is that either or all of the molecules which contribute to the fusion may have improved properties relative to the unfused molecule(s) (Balan *et al*. (2006) Antivir Ther 11(1): 35-45). Albumins and albumin particles are also important for carrying and delivering drugs and prodrugs to their sites of action (Kratz, F. (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 conjugation partner, for example bioactive molecule, from inactivation.

The polypeptide may display modified (*e.g.* reduced) glycosylation, such as, but not limited to reduced N-linked glycosylation or reduced O-linked glycosylation. The N-linked glycosylation pattern of an albumin molecule can be modified by adding/removing amino acid glycosylation consensus sequences such as N-X-S/T, at any or all of the N, X, or S/T position. Albumin polymorphisms exist with N-linked glycosylation. Albumin mutants may have recycling time such that the efficacy of a mutant as a bioactive carrier is improved. Recombinantly expressed proteins can be subject to undesirable post-translational modifications by the producing host cell. For example, the albumin protein sequence of SEQ ID No. 1 does not contain any sites for N-linked glycosylation and has not been reported to be modified, in nature, by O-linked glycosylation. However, it has been found that recombinant human albumin ("rHA") produced in a number of yeast species can be modified by O-linked glycosylation, generally involving mannose. The mannosylated albumin is able to bind to the lectin Concanavalin A. The amount of mannosylated albumin produced by the yeast can be re-

duced by using a yeast strain deficient in one or more of the *PMT* genes (WO 94/04687). The most convenient way of achieving this is to create a yeast which has a defect in its genome such that a reduced level of one of the Pmt proteins is produced. For example, there may be a deletion, insertion or transposition in the coding sequence or the regulatory regions (or in another gene regulating the expression of one of the *PMT* genes) such that little or no Pmt protein is produced. Alternatively, the yeast could be transformed to produce an anti-Pmt agent, such as an anti-Pmt antibody.

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 step of "purifying the thus expressed heterologous protein from the cultured host cell or the culture medium" optionally comprises cell immobilization, cell separation and/or cell breakage, but always comprises at least one other purification step different from the step or steps of cell immobilization, separation and/or breakage.

Cell immobilization techniques, such as encasing the cells using calcium alginate beads, are well known in the art. Similarly, cell separation techniques, such as centrifugation, filtration (*e.g.* cross-flow filtration, expanded bed chromatography and the like are well known in the art. Likewise, methods of cell breakage, including beadmilling, sonication, enzymatic exposure and the like are well known in the art.

The at least one other purification step may be any other step suitable for protein purification known in the art. For example purification techniques for the recovery of recombinantly expressed albumin have been disclosed in: WO 92/04367, removal of matrix-derived dye; EP 464 590, removal of yeast-derived colorants; EP 319 067, alkaline precipitation and subsequent application of the albumin to a lipophilic phase; and WO 96/37515, US 5 728 553 and WO 00/44772, which describe complete purification processes; all of which are incorporated herein by reference.

Production of conjugation competent albumin ("thio-albumin")

The thio-albumin or fusions of thio-albumin and another protein or proteins can be prepared by methods know to the art (Sanker, (2004), Genetic Eng. News, 24, 22-28, Schmidt, (2004), Appl. Microbiol. Biotechnol., 65, 363-372) including but not limited to expression in mammalian cell culture (Mason *et al*., (2004), Protein Expr. Purif., 36, 318-326;

Mason *et al*., (2002), Biochemistry, 41, 9448-9454) from cells lines such as CHO (and its variants), NS0, BHK, HEK293, Vero or PERC6 cells by transformation or transient expression; insect cell culture (Lim *et al*., (2004) Biotechnol. Prog., 20, 1192-1197); plant cell culture from such plants as Lemna or *Oryza sativa*; transgenic animals (Dyck *et al*., (2003) Trends in Biotechnology, 21, 394-399); transgenic plants (Ma *et al*., (2003) Nature Reviews Genetics, 4, 794-805); Gram positive and Gram negative bacteria such as *Bacillus* and *Escherichia coli* (Steinlein, and Ikeda, (1993), Enzyme Microb. Technol., 15, 193-199); filamentous fungi including but not restricted to *Aspergillus* spp (EP 238023, US 5,364,770, US 5,578,463, EP184438, EP284603, WO 2000/056900, WO9614413), *Trichoderma* spp and *Fusarium* spp (Navalainen *et al*., (2005), Trends in Biotechnology, 23, 468-473).

The host cell may be any type of cell. The host cell may or may not be an animal (such as mammalian, avian, insect, *etc*.), plant (such as *Oryza sativa*), fungal or bacterial cell. Bacterial and fungal, such as yeast, host cells may or may not be preferred.

Typical prokaryotic vector plasmids are: pUC18, pUC19, pBR322 and pBR329 available from Biorad Laboratories (Richmond, CA, USA); p*Trc*99A, pKK223-3, pKK233-3, pDR540 and pRIT5 available from Pharmacia (Piscataway, NJ, USA); pBS vectors, Phagescript vectors, Bluescript vectors, pNH8A, pNH16A, pNH18A, pNH46A available from Stratagene Cloning Systems (La Jolla, CA 92037, USA).

A typical mammalian cell vector plasmid is pSVL available from Pharmacia (Piscataway, NJ, USA). This vector uses the SV40 late promoter to drive expression of cloned genes, the highest level of expression being found in T antigen-producing cells, such as COS-1 cells. An example of an inducible mammalian expression vector is pMSG, also available from Pharmacia (Piscataway, NJ, USA). This vector uses the glucocorticoid-inducible promoter of the mouse mammary tumour virus long terminal repeat to drive expression of the cloned gene.

Methods well known to those skilled in the art can be used to construct expression vectors containing the coding sequence and, for example appropriate transcriptional or translational controls. One such method involves ligation *via* cohesive ends. Compatible cohesive ends can be generated on the DNA fragment and vector by the action of suitable restriction enzymes. These ends will rapidly anneal through complementary base pairing and remaining nicks can be closed by the action of DNA ligase.

A further method uses synthetic double stranded oligonucleotide linkers and adaptors. DNA fragments with blunt ends are generated by bacteriophage T4 DNA polymerase or *E.coli* DNA polymerase I which remove protruding 3' termini and fill in recessed 3' ends. Synthetic linkers and pieces of blunt-ended double-stranded DNA which contain recognition sequences for defined restriction enzymes, can be ligated to blunt-ended DNA fragments by T4 DNA lig-

ase. They are subsequently digested with appropriate restriction enzymes to create cohesive ends and ligated to an expression vector with compatible termini. Adaptors are also chemically synthesised DNA fragments which contain one blunt end used for ligation but which also possess one preformed cohesive end. Alternatively a DNA fragment or DNA fragments can be ligated together by the action of DNA ligase in the presence or absence of one or more synthetic double stranded oligonucleotides optionally containing cohesive ends.

Synthetic linkers containing a variety of restriction endonuclease sites are commercially available from a number of sources including Sigma-Genosys Ltd, London Road, Pampisford, Cambridge, United Kingdom.

The thio-albumin or fusions of thio-albumin and another protein or proteins may be expressed from a nucleotide sequence, which may or may not contain one or more introns. Additionally the nucleotide sequence may or may not be codon optimised for the host by methods known to the art.

The thio-albumin or fusions of thio-albumin and another protein or proteins can be expressed as variants with reduced N-linked glycosylation. Accordingly, in case of human serum albumin (HSA), it may be particularly advantageous to use a yeast deficient in one or more protein mannosyl transferases involved in O-glycosylation of proteins, for instance by disruption of the gene coding sequence. Recombinantly expressed proteins can be subject to undesirable post-translational modifications by the producing host cell. The mannosylated albumin would be able to bind to the lectin Concanavalin A. The amount of mannosylated albumin produced by the yeast can be reduced by using a yeast strain deficient in one or more of the *PMT* genes (WO 94/04687). The most convenient way of achieving this is to create a yeast which has a defect in its genome such that a reduced level of one of the Pmt proteins is produced. For example, there may or may not be a deletion, insertion or transposition in the coding sequence or the regulatory regions (or in another gene regulating the expression of one of the *PMT* genes) such that little or no Pmt protein is produced. Alternatively, the yeast could be transformed to produce an anti-Pmt agent, such as an anti-Pmt antibody. Alternatively, the yeast could be cultured in the presence of a compound that inhibits the activity of one of the PMT genes (Duffy *et al*, "*Inhibition of protein mannosyltransferase 1 (PMT1) activity in the pathogenic yeast Candida albicans*", International Conference on Molecular Mechanisms of Fungal Cell Wall Biogenesis, 26-31 August 2001, Monte Verita, Switzerland, 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*

promoter), as taught in EP 431 880; and a transcription terminator, preferably the terminator from *Saccharomyces ADH1*, as taught in EP 60 057.

It may be beneficial for the non-coding region to incorporate more than one DNA sequence encoding a translational stop codon, such as UAA, UAG or UGA, in order to minimise translational read-through and thus avoid the production of elongated, non-natural fusion proteins. The translation stop codon UAA is preferred.

In one preferred embodiment, the recombinant protein comprising the sequence of an albumin mutant is secreted. In that case, a sequence encoding a secretion leader sequence may be included in the open reading frame. Thus, a polynucleotide according to the present invention may comprise a sequence that encodes a recombinant protein comprising the sequence of an albumin mutant operably linked to a polynucleotide sequence that encodes a secretion leader sequence. Leader sequences are usually, although not necessarily, located at the N-terminus of the primary translation product of an ORF and are generally, although not necessarily, cleaved off the protein during the secretion process, to yield the “mature” protein. Thus, in one embodiment, the term “operably linked” in the context of leader sequences includes the meaning that the sequence that encodes a recombinant protein comprising the sequence of an albumin mutant is linked, at its 5' end, and in-frame, to the 3' end of a polynucleotide sequence that encodes a secretion leader sequence. Alternatively, the polynucleotide sequence that encodes a secretion leader sequence may be located, in-frame, within the coding sequence of the recombinant protein comprising the sequence of an albumin mutant, or at the 3' end of the coding sequence of the recombinant protein comprising the sequence of an albumin mutant.

Numerous natural or artificial polypeptide leader sequences (also called secretion pre regions and pre/pro regions) have been used or developed for secreting proteins from host cells. Leader sequences direct a nascent protein towards the machinery of the cell that exports proteins from the cell into the surrounding medium or, in some cases, into the periplasmic space.

For production of proteins in eukaryotic species such as the yeasts *Saccharomyces cerevisiae*, *Zygosaccharomyces* species, *Kluyveromyces lactis* and *Pichia pastoris*, a secretion leader sequence may be used. This may comprise a signal (pre) sequence or a prepro leader sequence. Signal sequences are known to be heterogeneous in their amino acid sequence (Nothwehr and Gordon 1990, Bioessays 12, 479-484, or Gierasch 1989, Biochemistry 28, p923-930). In essence, signal sequences are generally N-terminally located, have a basic n-region, a hydrophobic h-region and a polar c-region. As long as this structure is retained the signal sequence will work, irrespective of the amino acid composition. How well

they work, *i.e.* how much mature protein is secreted, depends upon the amino acid sequence. 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 oc-

curring plasmid.

Useful yeast episomal plasmid vectors are pRS403-406 and pRS413-416 and are generally available from Stratagene Cloning Systems (La Jolla, CA 92037, USA), YEp24 (Botstein, D., *et al*. (1979) Gene **8**, 17-24), and YEplac122, YEplac195 and YEplac181 (Gietz, R.D. and Sugino. A. (1988) Gene **74**, 527-534). Other yeast plasmids are described in WO 90/01063 and EP 424 117, as well as the "disintegration vectors of EP-A-286 424 and WO2005061719. Plasmids pRS403, pRS404, pRS405 and pRS406 are Yeast Integrating plasmids (YIps) and incorporate the yeast selectable markers *HIS3, TRP1, LEU2* and *URA3*, as are YIplac204, YIplac211 and YIplac128 (Gietz, R.D. and Sugino. A. (1988) Gene **74**, 527-534). Plasmids pRS413-416 are Yeast Centromere plasmids (YCps) as are YCplac22, YCplac33 and YCplac111 (Gietz, R.D. and Sugino. A. (1988) Gene **74**, 527-534).

Where one or more of the helper (also known as 'chaperone') protein(s) and/or protein product of choice are encoded by a plasmid-borne polynucleotide sequence, the host cell type may be selected for compatibility with the plasmid type being used. Such plasmids are disclosed in WO2005061719. Preferred helper proteins include *PDI1*, *AHA1*, *ATP11*, *CCT2*, *CCT3*, *CCT4*, *CCT5*, *CCT6*, *CCT7*, *CCT8*, *CNS1*, *CPR3*, *CPR6*, *DER1, DER3, DOA4*, *ERO1*, *EUG1*, *ERV2, EPS1, FKB2*, *FMO1*, *HCH1*, *HRD3*, *HSP10*, *HSP12*, *HSP104*, *HSP26*, *HSP30*, *HSP42*, *HSP60*, *HSP78*, *HSP82*, *KAR2*, *JEM1*, *MDJ1*, *MDJ2*, *MPD1*, *MPD2*, *PDI1*, *PFD1*, *ABC1*, *APJ1*, *ATP11*, *ATP12*, *BTT1*, *CDC37*, *CPR7*, *HSC82*, *KAR2*, *LHS1*, *MGE1*, *MRS11*, *NOB1*, *ECM10*, *SCJ1*, *SSA1*, *SSA2*, *SSA3*, *SSA4*, SSB1, SSB2, *SSC1*, *SSE2*, *SIL1*, *SLS1*, *ORM1*, *ORM2*, *PER1*, *PTC2, PSE1, UBC7, UBI4* and *HAC1* or a truncated intronless *HAC1* (Valkonen *et al*. 2003, *Applied Environ. Micro*., **69**, 2065). Such helper proteins are disclosed in WO 2005/061718, WO 2006/067511 and WO 2006/136831.

Plasmids as defined herein may be introduced into a host through standard techniques. With regard to transformation of prokaryotic host cells, see, for example, Cohen *et al* (1972) *Proc. Natl. Acad. Sci. USA* **69**, 2110 and Sambrook *et al* (2001) *Molecular Cloning, A Laboratory Manual*, 3rd Ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. Transformation of yeast cells is described in Sherman *et al* (1986) *Methods In Yeast Genetics, A Laboratory Manual*, Cold Spring Harbor, NY. The method of Beggs (1978) *Nature* **275**, 104-109 is also useful. Methods for the transformation of *S. cerevisiae* are taught generally in EP 251 744, EP 258 067 and WO 90/01063, all of which are incorporated herein by reference. With regard to vertebrate cells, reagents useful in transfecting such cells, for example calcium phosphate and DEAE-dextran or liposome formulations, are available from Stratagene Cloning Systems, or Life Technologies Inc., Gaithersburg, MD 20877, USA.

Electroporation is also useful for transforming cells and is well known in the art for

transforming fungal (including yeast) cell, plant cells, bacterial cells and animal (including vertebrate) cells. Methods for transformation of yeast by electroporation are disclosed in Becker & Guarente (1990) *Methods Enzymol.* **194**, 182.

Generally, the plasmid will transform not all of the hosts and it will therefore be necessary to select for transformed host cells. Thus, a plasmid may comprise a selectable marker, including but not limited to bacterial selectable marker and/or a yeast selectable marker. A typical bacterial selectable marker is the β-lactamase gene although many others are known in the art. Typical yeast selectable marker include *LEU2, TRP1, HIS3, HIS4, URA3, URA5, SFA1, ADE2, MET15, LYS5, LYS2, ILV2, FBA1, PSE1, PDI1* and *PGK1*. Those skilled in the art will appreciate that any gene whose chromosomal deletion or inactivation results in an unviable host, so called essential genes, can be used as a selective marker if a functional gene is provided on the plasmid, as demonstrated for *PGK1* in a *pgk1* yeast strain (Piper and Curran, 1990, *Curr. Genet*. **17**, 119). Suitable essential genes can be found within the Stanford Genome Database (SGD), (http:://db.yeastgenome.org). Any essential gene product (*e.g.* PDI1, PSE1, PGK1 or FBA1) which, when deleted or inactivated, does not result in an auxotrophic (biosynthetic) requirement, can be used as a selectable marker on a plasmid in a host cell that, in the absence of the plasmid, is unable to produce that gene product, to achieve increased plasmid stability without the disadvantage of requiring the cell to be cultured under specific selective conditions. By “auxotrophic (biosynthetic) requirement” we include a deficiency which can be complemented by additions or modifications to the growth medium. Therefore, preferred “essential marker genes” in the context of the present application are those that, when deleted or inactivated in a host cell, result in a deficiency which cannot be complemented by additions or modifications to the growth medium. Additionally, a plasmid may comprise more than one selectable marker.

Transformed host cells may be cultured for a sufficient time and under appropriate conditions known to those skilled in the art, and in view of the teachings disclosed herein, to permit the expression of the helper protein(s) and the protein product of choice.

The culture medium may be non-selective or place a selective pressure on the maintenance of a plasmid.

Methods for culturing prokaryotic host cells, such as *E. coli*, and eukaryotic host cells, such as mammalian cells are well known in the art. Methods for culturing yeast are generally taught in EP 330 451 and EP 361 991.

The thus produced protein product of choice may be present intracellularly or, if secreted, in the culture medium and/or periplasmic space of the host cell.

Preparation of a polypeptide

The step of “purifying the thus expressed protein product of choice from the cultured host cell, recombinant organism or culture medium” optionally comprises cell immobilisation, cell separation and/or cell breakage, but always comprises at least one other purification step different from the step or steps of cell immobilisation, separation and/or breakage.

Thio-albumin of the invention may be purified from the culture medium by any technique that has been found to be useful for purifying such proteins. Similarly, cell separation techniques, such as centrifugation, filtration (*e.g.* cross-flow filtration, expanded bed chromatography and the like) are well known in the art. Likewise, methods of cell breakage, including beadmilling, sonication, enzymatic exposure and the like are well known in the art.

The “at least one other purification step” may be any other step suitable for protein purification known in the art. For example purification techniques for the recovery of recombinantly expressed albumin have been disclosed in: WO 92/04367, removal of matrix-derived dye; EP 464 590, removal of yeast-derived colorants; EP 319 067, alkaline precipitation and subsequent application of the albumin to a lipophilic phase; and WO 96/37515, US 5 728 553 and WO 00/44772, which describe complete purification processes; all of which are incorporated herein by reference. Suitable methods include ammonium sulphate or ethanol precipitation, acid or solvent extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxyapatite chromatography, lectin chromatography, concentration, dilution, pH adjustment, diafiltration, ultrafiltration, high performance liquid chromatography (“HPLC”), reverse phase HPLC, conductivity adjustment and the like.

The polypeptide may be purified to a commercially or industrially acceptable level of purity. By commercially or industrially acceptable level of purity, we include the provision of the thio-albumin and/or thio-albumin-conjugate in which other material (for example, one or more contaminants) are present at a level of less than 50 %, 40 %, 30 %, 20 %, 10 %, 5 %, 4 %, 3 %, 2 %, 1 %, 0.5 %, 0.1 %, 0.01 %, 0.001 %, 0.0001 %, 0.00001 %, or 0.000001 % and, most preferably at a level of 0 %.

A commercially or industrially acceptable level of purity may be obtained by a relatively crude purification method by which the protein product of choice is put into a form suitable for its intended purpose. A protein preparation that has been purified to a commercially or industrially acceptable level of purity may, in addition to the protein product of choice, also comprise, for example, cell culture components such as host cells or debris derived therefrom. Alternatively, high molecular weight components (such as host cells or debris derived therefrom) may or may not be removed (such as by filtration or centrifugation) to obtain a composition

comprising the protein product of choice and, optionally, a functionally acceptable level of low molecular weight contaminants derived from the cell culture process.

The protein may or may not be purified to achieve a pharmaceutically acceptable level of purity. A protein has a pharmaceutically acceptable level of purity if it is essentially pyrogen free and can be used for its intended purpose and hence be administered in a pharmaceutically efficacious amount without causing medical effects not associated with the activity of the protein.

The thio-albumin and/or thio-albumin-conjugate may be provided at a concentration of at least 10^{-4} g.L^{-1}, 10^{-3} g.L^{-1}, 0.01 g.L^{-1}, 0.02 g.L^{-1}, 0.03 g.L^{-1}, 0.04 g.L^{-1}, 0.05 g.L^{-1}, 0.06 g.L^{-1},0.07 g.L^{-1}, 0.08 g.L^{-1}, 0.09 g.L^{-1}, 0.1 g.L^{-1}, 0.2 g.L^{-1}, 0.3 g.L^{-1}, 0.4 g.L^{-1}, 0.5 g.L^{-1}, 0.6 g.L^{-1}, 0.7 g.L^{-1}, 0.8 g.L^{-1}, 0.9 g.L^{-1}, 1 g.L^{-1}, 2 g.L^{-1}, 3 g.L^{-1}, 4 g.L^{-1}, 5 g.L^{-1}, 6 g.L^{-1}, 7 g.L^{-1}, 8 g.L^{-1}, 9 g.L^{-1}, 10 g.L^{-1}, 15 g.L^{-1}, 20 g.L^{-1}, 25 g.L^{-1}, 30 g.L^{-1}, 40 g.L^{-1},50 g.L^{-1}, 60 g.L^{-1}, 70 g.L^{-1}, 70 g.L^{-1}, 90 g.L^{-1}, 100 g.L^{-1}, 150 g.L^{-1}, 200 g.L^{-1}, 250 g.L^{-1}, 300 g.L^{-1}, 350 g.L^{-1}, 400 g.L^{-1}, 500 g.L^{-1}, 600 g.L^{-1}, 700 g.L^{-1}, 800 g.L^{-1}, 900 g.L^{-1}, 1000 g.L^{-1}.

A method of the present invention may or may not further comprise the step of formulating the purified protein product of choice with a carrier or diluent and optionally presenting the thus formulated protein in a unit dosage form.

Although it is possible for a therapeutically useful protein obtained by a process of the invention to be administered alone, it is preferable to present it as a pharmaceutical formulation, together with one or more acceptable carriers or diluents. The carrier(s) or diluent(s) must be “acceptable” in the sense of being compatible with the desired protein. Typically, the carriers or diluents will be water or saline which will be sterile and pyrogen free.

Alternatively, a method of the present invention may or may not further comprise the step of lyophilising the thus purified protein product of choice.

Formulation of thio-albumin or conjugate

The thio-albumin may be formulated by strategies given in “Protein Formulation and Delivery”, E. J. McNally (Ed.), published by Marcel Dekker Inc. New York 2000 and “Rational Design of Stable Protein Formulations – Theory and Practice”; J. F. Carpenter and M. C. Manning (Ed.) Pharmaceutical Biotechnology Vol 13. Kluwer Academic/Plenum Publishers, New York 2002, Yazdi and Murphy, (1994) Cancer Research 54, 6387-6394, Widera *et al*., (2003) Pharmaceutical Research 20, 1231-1238; Lee *et al*., (2005) Arch. Pharm. Res. 28, 722-729. Examples of formulation methods are as follows:

Method #1: Following purification the free thiol containing albumin mutein of the in-

vention or the conjugate can be stored at 4°C, -20°C or -80°C in 0.01 M – 0.1 M phosphate buffered saline (pH 7.0 - 8.0) containing 0.01 M – 0.2 M NaCl.

Method #2: Following purification the free thiol containing albumin mutein of the invention or the conjugate can be stored at 4°C, -20°C or -80°C in 0.01 M – 0.1 M phosphate buffered saline (pH 7.0 - 8.0) containing 0.01 M – 0.2 M NaCl and containing 10-20mg/L Polysorbate 80.

Method #3: Following purification the free thiol containing albumin mutein of the invention or the conjugate can be stored at 4°C, -20°C or -80°C in 0.01 M – 0.2 M NaCl (pH 7.0 - 8.0).

Method #4: Following purification the free thiol containing albumin mutein of the invention or the conjugate can be stored at 4°C, -20°C or -80°C in 0.01 M – 0.2 M NaCl (pH 7.0 - 8.0) containing 10-20mg/L Polysorbate 80.

Freeze-dried formulations

Method #5: Following purification the free thiol containing albumin mutein of the invention or the conjugate can be dialysed against water, freeze dried and stored at 4°C, -20°C or -80°C.

Method #6: Following purification the free thiol containing albumin mutein of the invention or the conjugate can be dialysed against 0.01 M – 0.2 M NaCl (pH 7.0 - 8.0), freeze dried and stored at 4°C, -20°C or -80°C.

Nanoparticle Formulation

The thio-albumin of the invention (and/or its conjugated form) may be used to produce nanoparticles and/or be entrapped within a nanoparticle or liposome.

The thio-albumin of the invention may be used with and/or in and/or as a nanoparticle and/or liposome. A problem of current conjugation strategies is maintaining both the pharmacological and immunological activity of the conjugation partner, such as a bioactive-targeting ligand conjugate. There is likely to be a maximum number of protein targeting ligand /bioactive moieties (conjugation partners) possible for conjugation to a protein and if this number is exceeded the targeting ligand does not retain its biological activity. Preferably the biological activity of the conjugation partner is not reduced by conjugation to an albumin of the invention.

Liposomes and nanoparticles may be used to entrap bioactive compounds. They provide a mechanism for enhanced delivery of drugs such as bioactive compounds, or uptake by target cells and/or a reduction in the toxicity of the free bioactive to non-target organs

which may result in an increased therapeutic index and/or reduced side effects. In addition, many solvent-based formulations required for the delivery of some bioactive compounds (*e.g.* taxanes) are associated with toxicity which limits the maximum dose which can be given to a patient. Liposome and nanoparticle delivery may also be advantageous for such bioactive compounds, since they would allow larger amounts of the bioactive compound to be delivered whilst avoiding some of the toxicities of solvent-based formulations (Hawkins *et al* (2008) Advanced Drug Delivery Reviews, 60, 8, p876-885).

Methods for attaching targeting ligands to liposomes and nanoparticles are known in the art (reviewed in Nobs *et al* (2004) Journal of Pharmaceutical Sciences Vol 93 p1980-1992) and may be used in accordance with the invention. Attachment methods may be non-covalent or covalent. Covalent reactions appear to be favourable, because covalent linkage is more stable than noncovalent methods. Lipids for the covalent or non-covalent attachment of proteins, peptides, or drugs to the liposome surface are available commercially (for example Avanti Polar Lipids Inc Alabaster, Alabama, USA). There are 3 major classes of functionality: conjugation through disulphide or thioether formation, amide bond formation, or biotin/streptavidin binding, any of these may be used in the invention.

A number of methods relying on covalent coupling ligands to the surface of liposomes *via* thioether bonds have been described, most commonly utilizing the highly efficient reaction of maleimide with thiol groups. Functionalized lipid anchors commonly added to liposomes, and which may be used in or with the invention, include, but are not limited those containing maleimide such as N-[4-(p-maleimidophenyl) butyramide]-PE (N-MPB]-PE) or N-[4-(p-maleimidomethyl) cyclohexane-carboxamide) (MCC-PE) which allow convenient covalent coupling of the targeting moiety *via* a stable thioether bond (Martin & Papahadjopoulos (1982) J. Biol. Chem. 257, 286- 288).

Method #7: Following purification the free thiol containing albumin mutein of the invention or the conjugate can be formulated into nanoparticles prepared according to known procedures for preparing nanoparticles, such as procedures disclosed in WO 2004/071536 A1 and WO 2008/007146 A1, both incorporated herein by reference.

Similarly materials for the formation of nanoparticles, including but are limited to Poly(lactic acid) (PLA), poly(lactic-co-glycolic acid) (PLGA), and COOH-PLA are commercially available and may be functionalized with maleimide or other known chemistries according to known literature for nanoparticle formation. Any of these may be used in or with the invention.

Another convenient way for covalent coupling of ligands to liposomes involves conjugation of two thiols to form a disulphide; however under the reductive conditions in serum

more stable conjugation chemistries involving one free thiol group may be preferred. Chemistries such as (PDP-PE) allow covalent coupling *via* a disulphide bond. Modification of the ligand to introduce a free thiol group or a functionalized linker may be used. An advantage of the thio-albumin of the invention is that no ligand modification is required. However, ligand modification may optionally be used in addition to the invention.

Frequently thiol groups are not present in proteins, or are not present in sufficient amounts or at the desired location. Thus, most cases of covalent coupling of one of more ligands to a liposome *via* thioether or disulphide bonds requires the use of heterobifunctional cross linking agents (described herein with reference to conjugation). Some heterobifunctional cross linking agents (such as SPDP and SATA) require a de-protection step. The thio-albumin of the invention overcomes the requirement for this additional processing.

Alternatively thio-albumin could be conjugated to liposomes or nanoparticles by other chemistries, known to the art. For example, thio-albumin could be attached by an amide bond using a functionalised lipid anchor with either amine or carboxyl functional groups (examples include DSPE-PEG-COOH) which reacts with the primary amine of the ligand. Direct cross linking between primary amines and the surface of liposomes may also be used. The one or more free thiol groups of thio-albumin would then be available for conjugation to another conjugation partner.

Following conjugation, a conjugation partner (*e.g.* bioactive molecule) may show a reduction in its activity (*e.g.* bioactivity). Thio-albumin described in this invention may overcome this problem by providing a conjugate, nanoparticle and/or liposome in which the conjugation partner is located and/or orientated with respect to a thio-albumin such that the conjugation partner retains at least 10, 20, 30, 40, 50, 60, 70, 80, 90 or 100% of its unconjugated activity.

Conjugation partner

The term ‘conjugation partner’ includes bioactive agents, imaging agents, diagnostic agents, contrast agents and therapeutic compounds such as chemotherapeutic drugs and radiopharmaceuticals. A thio-albumin of the invention may be conjugated to one or more conjugation partners.

Imaging agents, diagnostic compounds, contrast agents and therapeutic compounds

The use of diagnostic agents, imaging agents and biological "contrast" agents are well known to the art. A diagnostic agent is any pharmaceutical product used as part of a di-

agnostic test (*i.e.* together with the equipment and procedures that are needed to assess the test result). The diagnostic agent may be used *in vivo, ex vivo* or *in vitro*.

The ability of albumin to accumulate in damaged muscle fibres of dystrophic muscle has been well described. For example, a Gadolinium-DTPA-albumin conjugate may be used 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 bioactive compound includes but is not limited to, peptides, polypeptides or proteins (either natural, recombinant, or synthetic) (Debinski, (2002) Cancer Investigation 20, 801-809, O'Keefe and Draper *et al*., (1985) JBC 260, 932-937, Xia *et al*., (2000) J. Pharmacology Experimental Therapeutics 295, 594-600, Kavimandan *et al*., (2006) Bioconjugate Chem. 17, 1376-1384, Humphries, *et al*., (1994) J. Tissue Culture Methods 16, 239-242, Wenning *et al*., (1998) Biotech. Bioeng. 57, 484-496, Yazdi and Murphy, (1994) Cancer Research 54, 6387-6394, Weaver and Laske (2003) J. Neuro-Oncology 65, 3-13, Widera *et al*., (2003) Pharmaceutical Research 20, 1231-1238, Daniels, T.R. *et al*. (2006) Clinical Immunology 121, 159-176 and the references included therein); therapeutic and diagnostic drugs or compounds (Mishra *et al*., (2006) J. Drug Targeting 14, 45-53, Lim and Shen, (2004) Pharmaceutical Research 21, 1985-1992, Fritzer *et al*., (1996) Biochemical Pharmacology 51, 489-493, Lubgan and Jozwiak (2002) Cell. Mol. Biol. Lett. 7, 98, Daniels, T.R. *et al*. (2006) Clinical Immunology 121, 159-176 and the references included therein); high molecular weight complexes including but not limited to liposomes, viruses and nanoparticles (Mishra *et al*., (2006) J. Drug Targeting 14, 45-53, Daniels, T.R. *et al*. (2006) Clinical Immunology 121, 159-176 and the references included therein); nucleic acids and radionuclides, including DNA, RNA (including siRNA) and their analogs (Lee *et al*., (2005) Arch. Pharm. Res. 28, 722-729, Huang *et al*., (2007) FASEB J. 21, 1117-1125, Daniels, T.R. *et al*. (2006) Clinical Immunology 121, 159-176 and the references included therein) and devices (Humphries, *et al*., (1994) J. Tissue Culture Methods 16, 239-242 and the references included therein). Additionally the entity can itself be modified by methods known to the art.

Therapeutic compounds

4-1BB ligand, 5-helix, , A human C-C chemokine, A human L105 chemokine, A human L105 chemokine designated huL105_3., A monokine induced by gamma-interferon (MIG), A partial CXCR4B protein, A platelet basic protein (PBP), α1-antitrypsin, □□ACRP-30 Homologue; Complement Component C1q C, Adenoid-expressed chemokine (ADEC), aFGF; FGF-1, AGF, AGF Protein, albumin, an etoposide, angiostatin, Anthrax vaccine, Antibodies specific for collapsin, antistasin, Anti-TGF beta family antibodies, antithrombin III, APM-1; ACRP-30; Famoxin, apo-lipoprotein species, , Arylsulfatase B, b57 Protein, BCMA, Beta-thromboglobulin protein (beta-TG), bFGF; FGF2, Blood coagulation factors, BMP Pro-

cessing Enzyme Furin, BMP-10, BMP-12, BMP-15, BMP-17, BMP-18, BMP-2B, BMP-4, BMP-5, BMP-6, BMP-9, Bone Morphogenic Protein-2, calcitonin, Calpain-10a, Calpain-10b, Calpain-10c, Cancer Vaccine, Carboxypeptidase, C-C chemokine, MCP2, CCR5 variant, CCR7, CCR7, CD11a Mab, CD137; 4-1BB Receptor Protein, CD20 Mab, CD27, CD27L, CD30, CD30 ligand, CD33 immunotoxin, CD40, CD40L, CD52 Mab, Cerebus Protein, Chemokine Eotaxin., Chemokine hIL-8, Chemokine hMCP1, Chemokine hMCP1a, Chemokine 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 recep-

tor, 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 HCC-1, Human chemokine HCC-1, Human chemokine I-309, Human chemokine IP-10, Human chemokine L105_3, Human chemokine L105_7, Human chemokine MIG, Human chemokine MIG-beta protein, Human chemokine MIP-1alpha, Human chemokine MIP1beta, Human chemokine MIP-3alpha, Human chemokine MIP-3beta, Human chemokine PF4, Human chemokine protein 331D5, Human chemokine protein 61164, Human chemokine receptor CXCR3, Human chemokine SDF1alpha, Human chemokine SDF1beta, Human chemokine ZSIG-35, Human Chr19Kine protein, Human CKbeta-9, Human CKbeta-9, Human CX3C 111 amino acid chemokine, Human DNAX interleukin-40, Human DVic-1 C-C chemokine, Human EDIRF I protein sequence, Human EDIRF II protein sequence, Human eosinocyte CC type chemokine eotaxin, Human eosinophil-expressed chemokine (EEC), Human fast twitch skeletal muscle troponin C, Human fast twitch skeletal muscle troponin I, Human fast twitch skeletal muscle Troponin subunit C, Human fast twitch skeletal muscle Troponin subunit I Protein, Human fast twitch skeletal muscle Troponin subunit T, Human fast twitch skeletal muscle troponin T, Human foetal spleen expressed chemokine, FSEC, Human GM-CSF receptor, Human gro-alpha chemokine, Human gro-beta chemokine, Human gro-gamma chemokine, Human IL-16 protein, Human IL-1RD10 protein sequence, Human IL-1RD9, Human IL-5 receptor alpha chain, Human IL-6 receptor, Human IL-8 receptor protein hIL8RA, Human IL-8 receptor protein hIL8RB, Human IL-9 receptor protein, Human IL-9 receptor protein variant #3, Human IL-9 receptor protein variant fragment, Human IL-9 receptor protein variant fragment#3, Human interleukin 1 delta, Human Interleukin 10, Human Interleukin 10, Human interleukin 18, Human interleukin 18 derivatives, Human interleukin-1 beta precursor, Human interleukin-1 beta precursor., Human interleukin-1 receptor accessory protein, Human interleukin-1 receptor antagonist beta, Human interleukin-1 type-3 receptor, Human Interleukin-10 (precursor), Human Interleukin-10 (precursor), Human interleukin-11 receptor, Human interleukin-12 40 kD subunit, Human interleukin-12 beta-1 receptor, Human interleukin-12 beta-2 receptor, Human Interleukin-12 p35 protein, Human Interleukin-12 p40 protein, Human interleukin-12 receptor, Human interleukin-13 alpha receptor, Human interleukin-13 beta receptor, Human interleukin-15, Human interleukin-15 receptor from clone P1, Human interleukin-17 receptor, Human interleukin-18 protein (IL-18), Human

interleukin-3, human interleukin-3 receptor, Human interleukin-3 variant, Human interleukin-4 receptor, Human interleukin-5, Human interleukin-6, Human interleukin-7, Human interleukin-7., Human interleukin-8 (IL-8), Human intracellular IL-1 receptor antagonist, Human IP-10 and HIV-1 gp120 hypervariable region fusion protein, Human IP-10 and human Muc-1 core epitope (VNT) fusion protein, human liver and activation regulated chemokine (LARC), Human Lkn-1 Full-Length and Mature protein, Human mammary associated chemokine (MACK) protein Full-Length and Mature, Human mature chemokine Ckbeta-7, Human mature gro-alpha, Human mature gro-gamma polypeptide used to treat sepsis, Human MCP-3 and human Muc-1 core epitope (VNT) fusion protein, Human MI10 protein, Human MI1A protein, Human monocyte chemoattractant factor hMCP-1, Human monocyte chemoattractant factor hMCP-3, Human monocyte chemotactic proprotein (MCPP) sequence, Human neurotactin chemokine like domain, Human non-ELR CXC chemokine H174, Human non-ELR CXC chemokine IP10, Human non-ELR CXC chemokine Mig, Human PAI-1 mutants, Human protein with IL-16 activity, Human protein with IL-16 activity, Human secondary lymphoid chemokine (SLC), Human SISD protein, Human STCP-1, Human stromal cell-derived chemokine, SDF-1, Human T cell mixed lymphocyte reaction expressed chemokine (TMEC), Human thymus and activation regulated cytokine (TARC), Human thymus expressed, Human TNF-alpha, Human TNF-alpha, Human TNF-beta (LT-alpha), Human type CC chemokine eotaxin 3 protein sequence, Human type II interleukin-1 receptor, Human wild-type interleukin-4 (hIL-4) protein, Human ZCHEMO-8 protein, Humanized Anti-VEGF Antibodies, and fragments thereof, Humanized Anti-VEGF Antibodies, and fragments thereof, Hyaluronidase, ICE 10 kD subunit., ICE 20 kD subunit., ICE 22 kD subunit., Iduronate-2-sulfatase, Iduronidase, IL-1 alpha, IL-1 beta, IL-1 inhibitor (IL-1i)., IL-1 mature, IL-10 receptor, IL-11, IL-11, IL-12 p40 subunit., IL-13, IL-14, IL-15, IL-15 receptor, IL-17, IL-17 receptor, Il-17 receptor, Il-17 receptor, IL-19, IL-1i fragments, IL1-receptor antagonist, , IL-21 (TIF), IL-3 containing fusion protein., IL-3 mutant proteins, IL-3 variants, IL-3 variants, IL-4, IL-4 mutein, IL-4 mutein Y124G, IL-4 mutein Y124X, IL-4 muteins, Il-5 receptor, IL-6, Il-6 receptor, IL-7 receptor clone, IL-8 receptor, IL-9 mature protein variant (Met117 version), immunoglobulins or immunoglobulin-based molecules or fragment of either (*e.g.* a Small Modular ImmunoPharmaceutical™ ("SMIP") or dAb, Fab' fragments, F(ab')2, scAb, scFv or scFv fragment), including but not limited to plasminogen, Influenza Vaccine, Inhibin alpha, Inhibin beta, insulin, insulin-like growth factor, Integrin Mab, inter-alpha trypsin inhibitor, inter-alpha trypsin inhibitor, Interferon gamma-inducible protein (IP-10), interferons (such as interferon alpha species and sub-species, interferon beta species and sub-species, interferon gamma species and sub-species), interferons (such as interferon alpha species and sub-species, interferon beta species and sub-species, interferon gamma species and sub-species), Interleukin 6, Interleukin

8 (IL-8) receptor, Interleukin 8 receptor B, Interleukin-1alpha, Interleukin-2 receptor associated protein p43, interleukin-3, interleukin-4 muteins, Interleukin-8 (IL-8) protein., interleukin-9, Interleukin-9 (IL-9) mature protein (Thr117 version), interleukins (such as IL10, IL11 and IL2), interleukins (such as IL10, IL11 and IL2), Japanese encephalitis vaccine, Kalikrein Inhibitor, Keratinocyte growth factor, Kunitz domain protein (such as aprotinin, amyloid precursor protein and those described in WO 03/066824, with or without albumin fusions), Kunitz domain protein (such as aprotinin, amyloid precursor protein and those described in WO 03/066824, with or without albumin fusions), LACI, lactoferrin, Latent TGF-beta binding protein II, leptin, Liver expressed chemokine-1 (LVEC-1), Liver expressed chemokine-2 (LVEC-2), LT-alpha, LT-beta, Luteinization Hormone, Lyme Vaccine, Lymphotactin, Macrophage derived chemokine analogue MDC (n+1), Macrophage derived chemokine analogue MDC-eyfy, Macrophage derived chemokine analogue MDC-yl, Macrophage derived chemokine, MDC, Macrophage-derived chemokine (MDC), Maspin; Protease Inhibitor 5, MCP-1 receptor, MCP-1a, MCP-1b, MCP-3, MCP-4 receptor, M-CSF, , Melanoma inhibiting protein, Membrane-bound proteins, Met117 human interleukin 9, MIP-3 alpha, MIP-3 beta, MIP-Gamma, MIRAP, Modified Rantes, monoclonal antibody, MP52, Mutant Interleukin 6 S176R, myofibrillar contractile protein Troponin I, Natriuretic Peptide, Nerve Growth Factor-beta, Nerve Growth Factor-beta2, Neuropilin-1, Neuropilin-2, Neurotactin, Neurotrophin-3, Neurotrophin-4, Neurotrophin-4a, Neurotrophin-4b, Neurotrophin-4c, Neurotrophin-4d, Neutrophil activating peptide-2 (NAP-2), NOGO-66 Receptor, NOGO-A, NOGO-B, NOGO-C, Novel beta-chemokine designated PTEC, N-terminal modified chemokine GroHEK/hSDF-1alpha, N-terminal modified chemokine GroHEK/hSDF-1beta., N-terminal modified chemokine met-hSDF-1 alpha, N-terminal modified chemokine met-hSDF-1 beta, OPGL, Osteogenic Protein-1; OP-1; BMP-7, Osteogenic Protein-2, OX40; ACT-4, OX40L, Oxytocin (Neurophysin I), parathyroid hormone, Patched, Patched-2, PDGF-D, Pertussis toxoid, Pituitary expressed chemokine (PGEC), Placental Growth Factor, Placental Growth Factor-2, Plasminogen Activator Inhibitor-1; PAI-1, Plasminogen Activator Inhibitor-2; PAI-2, Plasminogen Activator Inhibitor-2; PAI-2, Platelet derived growth factor, Platelet derived growth factor Bv-sis, Platelet derived growth factor precursor A, Platelet derived growth factor precursor B, Platelet Mab, platelet-derived endothelial cell growth factor (PD-ECGF), Platelet-Derived Growth Factor A chain, Platelet-Derived Growth Factor B chain, polypeptide used to treat sepsis, Preproapolipoprotein “milano” variant, Preproapolipoprotein “paris” variant, pre-thrombin, Primate CC chemokine "ILINCK", Primate CXC chemokine "IBICK", proinsulin, Prolactin, Prolactin2, prosaptide, Protease inhibitor peptides, Protein C, , Protein S, pro-thrombin, prourokinase, RANTES, RANTES 8-68, RANTES 9-68, RANTES peptide, RANTES receptor, Recombinant interleukin-16, Resistin, restrictocin, Retroviral protease inhibitors, ricin, Rota-

virus Vaccine, RSV Mab, saporin, sarcin, Secreted and Transmembrane polypeptides, Secreted and Transmembrane polypeptides, serum cholinesterase, serum protein (such as a blood clotting factor), Soluble BMP Receptor Kinase Protein-3, Soluble VEGF Receptor, Stem Cell Inhibitory Factor, Straphylococcus Vaccine, Stromal Derived Factor-1 alpha, Stromal Derived Factor-1 beta, Substance P (tachykinin), T1249 peptide, T20 peptide, T4 Endonuclease, TACI, Tarc, TGF-beta 1, TGF-beta 2, Thr117 human interleukin 9, thrombin, , thrombopoietin, Thrombopoietin derivative1, Thrombopoietin derivative2, Thrombopoietin derivative3, Thrombopoietin derivative4, Thrombopoietin derivative5, Thrombopoietin derivative6, Thrombopoietin derivative7, Thymus expressed chemokine (TECK), Thyroid stimulating Hormone, tick anticoagulant peptide,, Tim-1 protein, TNF-alpha precursor, TNF-R, TNF-RII; TNF p75 Receptor; Death Receptor, tPA, transferrin, transforming growth factor beta, , Troponin peptides, Truncated monocyte chemotactic protein 2 (6-76), Truncated monocyte chemotactic protein 2 (6-76), Truncated RANTES protein (3-68), tumour necrosis factor, Urate Oxidase, urokinase, Vasopressin (Neurophysin II), VEGF R-3; flt-4, VEGF Receptor; KDR; flk-1, VEGF-110, VEGF-121, VEGF-138, VEGF-145, VEGF-162, VEGF-165, VEGF-182, VEGF-189, VEGF-206, VEGF-D, VEGF-E; VEGF-X, von Willebrand's factor, , Wild type monocyte chemotactic protein 2, Wild type monocyte chemotactic protein 2, ZTGF-beta 9.

Chemotherapy drugs

13-cis-Retinoic Acid, 2-CdA, 2-Chlorodeoxyadenosine, 5-Azacitidine, 5-Fluorouracil, 5-FU, 6-Mercaptopurine, 6-MP, 6-TG, 6-Thioguanine, A, Abraxane, Accutane®, Actinomycin-D, Adriamycin®, Adrucil®, Agrylin®, Ala-Cort®, Aldesleukin, Alemtuzumab, ALIMTA, Alitretinoin, Alkaban-AQ®, Alkeran®, All-transretinoic Acid, Alpha Interferon, Altretamine, Amethopterin, Amifostine, Aminoglutethimide, Anagrelide, Anandron®, Anastrozole, Arabinosylcytosine, Ara-C, Aranesp®, Aredia®, Arimidex®, Aromasin®, Arranon®, Arsenic Trioxide, Asparaginase, ATRA, Avastin®, Azacitidine, BCG, BCNU, Bevacizumab, Bexarotene, BEXXAR®, Bicalutamide, BiCNU, Blenoxane®, Bleomycin, Bortezomib, Busulfan, Busulfex®, C225 , Calcium Leucovorin, Campath®, Camptosar®, Camptothecin-11, Capecitabine, Carac ™, Carboplatin, Carmustine, Carmustine Wafer, Casodex®, CC-5013, CCNU, CDDP, CeeNU, Cerubidine®, Cetuximab, Chlorambucil, Cisplatin, Citrovorum Factor, Cladribine, Cortisone, Cosmegen®, CPT-11, Cyclophosphamide, Cytadren®, Cytarabine, Cytarabine Liposomal, Cytosar-U®, Cytoxan®, Dacarbazine, Dacogen, Dactinomycin, Darbepoetin Alfa, Dasatinib, Daunomycin, Daunorubicin, Daunorubicin Hydrochloride, Daunorubicin Liposomal, DaunoXome®, Decadron, Decitabine, Delta-Cortef®, Deltasone®, Denileukin diftitox, DepoCyt ™, Dexamethasone, Dexamethasone acetate , Dexamethasone Sodium Phosphate , Dexasone, Dexrazoxane, DHAD, DIC, Diodex, Docetaxel, Doxil®, Doxorubicin, Doxorubicin

liposomal, Droxia ™, DTIC, DTIC-Dome®, Duralone®, Efudex®, Eligard ™, Ellence ™, Eloxatin ™, Elspar®, Emcyt®, Epirubicin, Epoetin alfa, Erbitux ™, Erlotinib, Erwinia L-asparaginase, Estramustine, Ethyol , Etopophos®, Etoposide, Etoposide Phosphate, Eulexin®, Evista®, Exemestane, Fareston®, Faslodex®, Femara®, Filgrastim, Floxuridine, Fludara®, Fludarabine, Fluoroplex®, Fluorouracil, Fluorouracil (cream), Fluoxymesterone, Flutamide, Folinic Acid, FUDR®, Fulvestrant, G-CSF, Gefitinib, Gemcitabine, Gemtuzumab ozogamicin, Gemzar®, Gleevec ™, Gliadel® Wafer, GM-CSF, Goserelin, Granulocyte - Colony Stimulating Factor, Granulocyte Macrophage Colony Stimulating Factor, Halotestin®, Herceptin®, Hexadrol, Hexalen®, Hexamethylmelamine, HMM, Hycamtin®, Hydrea®, Hydrocort Acetate®, Hydrocortisone, Hydrocortisone Sodium Phosphate, Hydrocortisone Sodium Succinate, Hydrocortone Phosphate, Hydroxyurea, Ibritumomab, Ibritumomab Tiuxetan , Idamycin®, Idarubicin, Ifex®, IFN-alpha , Ifosfamide, IL-11 , IL-2 , Imatinib mesylate, Imidazole Carboxamide , Interferon alfa, Interferon Alfa-2b (PEG Conjugate), Interleukin - 2, Interleukin-11, Intron A® (interferon alfa-2b) , Iressa®, Irinotecan, Isotretinoin, Kidrolase®, Lanacort®, Lapatinib, L-asparaginase, LCR, Lenalidomide, Letrozole, Leucovorin, Leukeran, Leukine ™, Leuprolide, Leurocristine, Leustatin ™, Liposomal Ara-C, Liquid Pred®, Lomustine, L-PAM, L-Sarcolysin, Lupron®, Lupron Depot®, M, Matulane®, Maxidex, Mechlorethamine, Mechlorethamine Hydrochloride, Medralone®, Medrol®, Megace®, Megestrol, Megestrol Acetate, Melphalan, Mercaptopurine, Mesna, Mesnex ™, Methotrexate, Methotrexate Sodium, Methylprednisolone, Meticorten®, Mitomycin, Mitomycin-C, Mitoxantrone, M-Prednisol®, MTC, MTX, Mustargen®, Mustine , Mutamycin®, Myleran®, Mylocel ™, Mylotarg®, Navelbine®, Nelarabine, Neosar®, Neulasta ™, Neumega®, Neupogen®, Nexavar®, Nilandron®, Nilutamide, Nipent®, Nitrogen Mustard, Novaldex®, Novantrone®, Octreotide, Octreotide acetate, Oncospar®, Oncovin®, Ontak®, Onxal ™, Oprevelkin, Orapred®, Orasone®, Oxaliplatin, Paclitaxel, Paclitaxel Protein-bound, Pamidronate, Panitumumab, Panretin®, Paraplatin®, Pediapred®, PEG Interferon, Pegaspargase, Pegfilgrastim, PEG-INTRON ™, PEG-L-asparaginase, PEMETREXED, Pentostatin, Phenylalanine Mustard, Platinol®, Platinol-AQ®, Prednisolone, Prednisone, Prelone®, Procarbazine, PROCRIT®, Proleukin®, Prolifeprospan 20 with Carmustine Implant, Purinethol®, R, Raloxifene, Revlimid®, Rheumatrex®, Rituxan®, Rituximab, Roferon-A® (Interferon Alfa-2a) , Rubex®, Rubidomycin hydrochloride, Sandostatin®, Sandostatin LAR®, Sargramostim, Solu-Cortef®, Solu-Medrol®, Sorafenib, SPRYCEL ™, STI-571, Streptozocin, SU11248, Sunitinib, Sutent®, Tamoxifen, Tarceva®, Targretin®, Taxol®, Taxotere®, Temodar®, Temozolomide, Teniposide, TESPA, Thalidomide, Thalomid®, TheraCys®, Thioguanine, Thioguanine Tabloid®, Thiophosphoamide, Thioplex®, Thiotepa, TICE®, Toposar®, Topotecan, Toremifene, Tositumomab, Trastuzumab, Tretinoin, Trexall ™, Trisenox®, TSPA, TYKERB®, VCR, Vectibix ™, Velban®, Velcade®, VePesid®,

Vesanoid®, Viadur ™, Vidaza®, Vinblastine, Vinblastine Sulfate, Vincasar Pfs®, Vincristine, Vinorelbine, Vinorelbine tartrate, VLB, VM-26, Vorinostat, VP-16, Vumon®, Xeloda®, Zanosar®, Zevalin ™, Zinecard®, Zoladex®, Zoledronic acid, Zolinza, Zometa®.

Radiopharmaceuticals

Carbon-11, Carbon-14, Chromium-51, Cobalt-57, Cobalt-58, Erbium-169, Fluorine-18, Gallium-67, Gold-198, Indium-111, Indium-113m, Iodine-123, Iodine-125, Iodine-131, Iron-59, Krypton-81m, Nitrogen-13, Oxygen-15, Phosphorous-32, Rhenium-186, Rubidium-82, Samarium-153, Selenium-75, Strontium-89, Technetium-99m, Thallium-201, Tritium, Xenon-127, Xenon-133, Yttrium-90.

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, bili-

rubin 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-iodoacetamidofluorescein (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 WO200069902. There are three parts of each DAC™ construct: 1) the drug component (the portion responsible for biologic activity); 2) a linker attached to the drug component, and 3) a reactive chemistry group at the opposite end of the linker, usually a soft electrophile selective for thiols; a maleimide is the most useful embodiment. Other applicable conjugation methods are described in WO2007/071068 incorporated herein by reference.

If the conjugation partner does not contain a thiol reactive group but does contain one or more amino groups then it may be modified to incorporate one or more thiol reactive groups by chemical modification by methods known to the art such as the use of cross-linking reagents such as, but not limited to, N-5-azido-2-nitrobenzoyloxysuccinimide (AMAS, Pierce), N-[beta-maleimidopropyloxy] succinimide ester (BMPS, Pierce), N-eta-maleimidocaproic acid (EMCA, Pierce), N-[eta-maleimidocaproyloxy]succinimide ester (EMCS, Pierce), N-[eta-maleimidocaproyloxy]sulfosuccinimide ester (sulfo-EMCS, Pierce), N-[gamma-maleimidobutyryloxy]succinimide ester (GMBS, Pierce), N-[gamma-maleimidobutyryloxy]sulfosuccinimide ester (sulfo-GMBS, Pierce), N-kappa-maleimidoundecanoic acid (KMUA, Pierce), N-[kappa -maleimidoundecanoic acid]hydrazide (KMUH, Pierce), N-[kappa -maleimidoundecanoyloxy]sulfosuccinimide ester (sulfo-KMUS, Pierce), m-maleimidobenzoyl-N-hydroxysuccinimide (MBS, Pierce), m-maleimidobenzoyl-N-hydroxysulfosuccinimide ester (sulfo-MBS, Pierce), N-succinimidyl S-acetylthio-acetate (SATA, Pierce), N-succinimidyl S-acetylthiopropionate (SATP, Pierce), succinimidyl 3-[bromoacetamido]propionate (SBAP, Pierce), N-succinimidyl iodoacetate (SIA, Pierce), N-succinimidyl[4-iodoacetyl]aminobenzoate (SIAB, Pierce), sulfosuccinimidyl[4-iodoacetyl]aminobenzoate (sulfo-SIAB, Pierce), succinimidyl [4-[N-maleimidomethyl]cyclohexane-1-carboxylate (SMCC, Pierce), sulfosuccinimidyl [4-[N-maleimidomethyl]cyclohexane-1-carboxylate (sulfo-SMCC, Pierce), succinimidyl-[4-[N-maleimidomethyl]cyclohexane-1-carboxy-[6-amidocaproate (LC-SMCC, Pierce), 4-succinimidyloxycarbonyl-methyl-alpha[2-pyridyldithio]toluene (SMPT, Pierce), sulfosuccinimidyl-6-[alpha-methyl-alpha□2-pyridyldithio)toluamido]hexanoate (sulfo-LC-

SMPT, Pierce), succinimidyl 4-[p-maleimidophenyl]-butyrate (SMPB, Pierce), sulfosuccinimidyl 4-[p-maleimidophenyl]-butyrate (sulfo-SMPB, Pierce), succinimidyl-6-[(beta-maleimidopropionamido)hexanoate](SMPH, Pierce), N-succinimidyl 3-[2-pyridyldithio]propionate (SPDP, Pierce), succinimidyl [3-(2-pyridyldithio)propionamido]hexanoate (LC-SPDP, Pierce), sulfosuccinimidyl [3'-(2-pyridyldithio)propionamido]hexanoate (sulfo-LC-SPDP, Pierce) and N-succinimidyl-[4-vinylsulfonyl]benzoate (SVSB Pierce). It may be advantageous to block certain amine residue as described by Kavimandan *et al*., (2006) Bioconjugate Chem. 17, 1376-1384.

If the conjugation partner does not contain a thiol reactive group but does contain one or more carbonyl (oxidised carbohydrate) groups then it can be modified to incorporate one or more thiol reactive groups by chemical modification by methods known to the art such as the use of cross-linking reagents such as, but not limited to, N-[eta-maleimidocaproic acid]hydrazide(EMCH, Pierce), 4-[N-maleimidomethyl]cyclohexane-1carboxylhydrazide•HCl•1/2 dioxane (M2C2H, Pierce), 3-maleimidophenyl boronic acid (MPBH, Pierce) and 3-[2-pyridyldithio]propionyl hydrazide (PDPH, Pierce).

If the conjugation partner does not contain a thiol reactive group but does contain one or more hydroxyl groups then it may be modified to incorporate one or more thiol reactive groups by chemical modification by methods known to the art such as the use of cross-linking reagents such as, but not limited to, N-[p-maleimidophenyl]isocyanate (PMPI, Pierce).

Conjugation competence of albumin variant

The conjugation competence of polypeptides of the invention may be tested by fluorescent labelling and cellular uptake, as described by McGraw *et al*., (1987), *The Journal of Cell Biology*, 105, 207-214 and Presley *et al*., (1993), *The Journal of Cell Biology*, 122, 1231-1241. Other methods of testing conjugation competence include conjugating the albumin to another molecule such as HRP. Subsequently, the mass of the resultant conjugate and/or the activity of the conjugated compound may be assayed, for example by mass spectrometry or by enzyme assay.

Microorganism.

A host strain suitable for use in the present invention includes an *hsp150*-deficient version of DXY1, disclosed in S. M. Kerry-Williams *et al*. (1998) Yeast 14:161-169. WO 95/33833 teaches the skilled person how to prepare *hsp150*-deficient yeast. This host strain may be referred to as 'Strain 1'.

All documents cited are incorporated by reference in their entirety.

The invention is described by way of example only with reference to the following examples:

EXAMPLES

EXAMPLE 1: CONSTRUCTION OF ALBUMIN MUTEIN EXPRESSION PLASMIDS

The HSA coding sequence is obtainable by known methods for isolating cDNA corresponding to human genes, and is also disclosed in, for example, EP 0 073 646 and EP 0 286 424. Expression plasmids for albumin variants of this invention can be constructed in a similar way to pDB2244 described in WO 00/44772 or pDB2305 described in WO/2006/013859 for expression of human serum albumin from *S. cerevisiae*. Plasmid pDB2305 contains the HSA sequence codon-optimised for expression in *S. cerevisiae.* Alternative codon optimisation methods may be used for the particular host organism selected for thio-albumin production. Expression plasmids for albumin variants of this invention can also be constructed in a similar way to those described in WO 2005/061719 A1 for improved expression of human serum albumin from *S. cerevisiae*.

Thio-albumin muteins can be made following modification of plasmid pDB2244 (Fig. 7) or pDB2305 by site directed mutagenesis. Overlapping mutagenic oligonucleotide sequences can be used to modify the codon of the selected residue(s) to any DNA sequence which encodes a cysteine residue (TGT or TGC) using the procedures indicated by a commercially available kit (such as Stratagene's Quikchange™ Kit). Alternatively, synthetic DNA fragments can be manufactured containing the desired modifications to the polynucleotide sequence.

Construction of a thio-albumin mutant expression plasmids

Subcloning plasmids which may be used to create plasmid pDB2244 (Fig. 7) are plasmid pDB2243 (Fig. 8) (described in WO 00/44772) and pSAC35 (described in EP 286424). Plasmids pDB2243 and pDB2244 contain the native HSA gene. A skilled person will appreciate that the expression cassette may or may not be codon optimised; methods for constructing expression plasmids containing HSA codon optimised for expression in *S. cerevisiae* are described in WO/2006/013859. The native nucleotide sequence encoding HSA is provided in SEQ ID No. 2. A HSA nucleotide sequence codon-optimised for expression in *S. cerevisiae* is provided as SEQ ID No. 3.

Plasmid pDB2243 (6.203kb) was digested to completion using restriction endonucle-

ases *Not*I to release the 2.992 kb human serum albumin expression cassette.

Plasmid pSAC35 is derivative of pSAC3 by Chinery and Hinchliffe (1989) *Curr. Genet.* **16** , 21-25, and in EP 286424. Plasmid pSAC35 (11.037 kb) 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 14.037 kb pDB2244 which has the human serum albumin expression cassette orientated in the same direction as the *LEU2* gene (Fig. 7). 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.

Alternatively plasmid pDB2690 may be used. The construction of plasmid pDB2690 is 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 pro-

tein 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 pDB2713.

Those skilled in the art will appreciate that expression cassette for thio-albumin variants with additional modifications to the albumin protein sequence could be produced using a similar method to that described for the construction of a thio-albumin subcloning plasmid containing one extra conjugation competent cysteine (relative to SEQ ID No. 1).

A *S. cerevisiae* strain, *e.g.* Strain 1, may be transformed to leucine prototrophy with pDB2244 (WO 00/44772), or pDB2305 (WO/2006/013859) for expression of human serum albumin or the appropriate thio-albumin expression plasmids. Yeast may be transformed using a modified lithium acetate method (Sigma yeast transformation kit, YEAST-1, protocol 2; Ito *et al*, 1983, J. Bacteriol., 153, 16; Elble, 1992, Biotechniques, 13, 18). Transformants may be selected on BMMD-agar plates, and subsequently patched out on BMMD-agar plates. The composition of BMMD is described by Sleep *et al*., 2002, Yeast, 18, 403. Cryopreserved stocks may be prepared in 20% (w/v) trehalose from 10mL BMMD shake flask cultures (24 hours, 30ºC, 200rpm).

EXAMPLE 2: EXPRESSION OF ALBUMIN MUTEINS WITH SINGLE AMINO ACID CHANGES COMPARED TO HSA

Thio-albumin variants with single amino acid changes were selected from Tables 5A, 5B and 6A. These variants were identified as the preferred mutations according to the methods described above. Details of each variant are given in Figure 11, which provides a Construct Reference (e.g. TA1 for rHA A2C), the name of the plasmid encoding each thio-albumin variant expression construct and flanking sequences required for in vivo recombina-

tion by gap-repair, and the number given to a cryopreserved yeast stock (the yeast stock number) producing each thio-albumin variant. Details of the mutant codons compared to SEQ ID No. 2 are also provided, as are the SEQ ID numbers for each thio-albumin variant (DNA and protein).

To modify the amino acids in non-human serum albumins, the equivalent positions to a particular position in HSA may be determined from an alignment including human serum albumin (SEQ ID No. 1) such as Figures 2 and 3. The skilled person is familiar with alignments and can readily determine whether or not an amino acid in a sequence is equivalent to an amino acid in another sequence. For example, the position of the amino acid in the non-human albumin is not necessarily the same relative to the N-terminal end of HSA. For example, from Figure 2 position 239 of HSA is an alanine residue, whereas the corresponding residue of the bovine sequence is serine-238. Similarly, valine-479 of HSA corresponds to leucine-478 of sheep albumin. The plasmid pDB3927 (Figure 12) was constructed from plasmid 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 con-

taining 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 thio-albumin 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 fragments for gap-repair, as described above. The plate was mixed by swirling of the plate while flat on a benchtop. 300µl of PEG/LiAc (40% w/v PEG 3350, 100mM lithium acetate) was added to each well and was mixed again. The plate was incubated at 30°C with shaking at 200 rpm for 1 hour before transfer to static incubation at 42°C for 30min. After 1min incubation on ice, the plate was centrifuged (2000xg, 5min, room temperature) followed by removal of the supernatant and resuspension of the pellet in 200µl 1M sorbitol. The full volume was inoculated onto BMMD agar plates with CSM-Leu nutritional supplement (MP Biomedicals, Bio 101) and incubated for 4 days at 30°C.

Single colony transformants were picked and patched onto fresh BMMD agar plates for short term storage. These patches were grown at 30°C and cells then inoculated into 10mL BMMD shake flask cultures and cryopreserved as described earlier. 10µl of yeast stock was inoculated into a 48-well plate containing 0.5mL BMMD per well. Growth of cultures in microtitre plates was achieved in a humidity chamber which was a sealed Perspex box containing wet paper towels to provide ~100% humidity and evaporative loss below 0.25% over 5 days under growth conditions. The plates were incubated in the shaking humidity chamber (30°C, 200rpm,) for 5 days at 30°C. The 48-well plate was centrifuged to pellet cells (2000xg, 10 min, room temperature) and the supernatant was harvested.

The concentration of the thio-albumin variants in the culture supernatants was deter-

mined by Gel Permeation High Pressure Liquid Chromatography (GP-HPLC). Protein concentrations were determined using a LC2010 HPLC system (Shimadzu) equipped with UV detection under Shimadzu VP7.3 client server software control. Injections of 25µL were made onto a 7.8mm internal diameter x 300mm length TSK G3000SWXL column (Tosoh Bioscience), with a 6.0mm internal diameter x 40mm length TSK SW guard column (Tosoh Bioscience). Samples were chromatographed in 25mM sodium phosphate, 100mM sodium sulphate, 0.05% (w/v) sodium azide, pH 7.0 at 1mL.min^{-1}, with a run time of 15 minutes. Samples were quantified by UV detection at 280nm, by peak height, relative to a recombinant human albumin standard of known concentration (10mg/mL).

A non-reducing SDS-PAGE analysis and the expression titres (by GP-HPLC) for each of the thio-albumin variants with single mutations are compared against controls in Figure 15. It is evident that all of the thio-albumin variants have been successfully secreted from *S. cerevisiae* BXP10. Preferred mutations have high expression titres and show a sharp Coomassie stained band equivalent to rHA controls by non-reducing SDS-PAGE analysis.

EXAMPLE 3: EXPRESSION OF ADDITIONAL THIO-ALBUMIN VARIANTS

Figure 16 describes an additional selection of thio-albumin variants with two or more free-thiol groups. Mutations shown to be expressed in Example 2 above were combined to generate sequences designed to have multiple free-thiol groups available for conjugation. This selection includes thio-albumin variants designed to have up to five free-thiol groups, thio-albumin variants designed to have free-thiol groups from within one Selection Group or from more than one Selection Group, thio-albumin variants designed to have free-thiol groups with and without the naturally occurring free-thiol at C34 of HSA, thio-albumin variants designed to have free-thiol groups from a range of Proximity Groups, and thio-albumin variants designed to have free-thiol groups derived from insertions, extensions, additions and/or deletions. It represents a sub-set of thio-albumins with multiple conjugation competent cysteine residues. The details of these thio-albumin variants, the plasmids encoding them and the SEQ ID No. for their DNA and protein sequences are described in Figure 16, in a similar manner to those of Figure 11 for the thio-albumin variants with single modifications. Methods for plasmid construction and expression from *S. cerevisiae* BXP10 are similar to those described above.

Figure 17 shows a non-reducing SDS-PAGE analysis and the expression titres (by GP-HPLC) for each of these additional thio-albumin variants compared against an rHA control (pDB3927 coding sequnce). Again, it is evident that all of the thio-albumin variants have been successfully secreted from *S. cerevisiae* BXP10. It therefore confirms that the selection criteria allow suitable thio-albumin variants to be generated and therefore indicates that there

is no problem with undersirable mis-folding or aggregation. Again, preferred combinations of mutations have high expression titres and show a sharp Coomassie stained band equivalent to rHA controls by non-reducing SDS-PAGE analysis.

EXAMPLE 4: PRODUCTION, PURIFICATION AND CONJUGATION OF THIO-ALBUMIN VARIANTS

Five cryopreserved yeast stocks (9116, 9118, 9124, 9125 and 9130; Figure 11) each in 1mL aliquots were inoculated into shake flasks containing 100mL BMMS growth medium (yeast nitrogen base without amino acids and $(NH_4)_2SO_4$, Difco 1.7g/L; citric acid monohydrate 6.09g/L; anhydrous Na_2HPO_4 20.16g/L; $(NH_4)_2SO_4$ 5.0g/L; pH6.5±0.2; sucrose added to 20g/L). Cells were transferred from the shake flask to the fermenter (10L working volume, Sartorius Biostat C 10-3 fermenter) when the concentration of cells in the shake flask has reached 0.8–1.2g/L achieving a cell inocula concentration of ≥10mg/L (greater than or equal to 10mg/L) in the fermenter.

The thio-albumin variants proteins were produced by axenic culture of each of the five yeast strains in high cell density (HCD) fed-batch fermentation. The aim of the fermentation was to achieve maximum biomass and productivity by controlling feed rate addition so that formation of byproducts such as ethanol and acetate were avoided. Further details of the fermentation process are described in WO96/37515. The temperature and pH were controlled at 30°C and pH5.5 respectively. Culture supernatant was harvested by centrifugation using a Sorvall RC 3C centrifuge (DuPont) and frozen for storage, before being thawed for subsequent purification. Final product concentrations were determined by GP HPLC using a LC2010 HPLC system (Shimadzu) equipped with UV detection under Shimadzu VP7.3 client server software control as described above. Figure 18 provides the yields of each thio-albumin variant (in g/L culture supernatant) and shows that high product titres of greater that 1g/L culture supernatant were obtained in all cases.

A single step chromatography procedure was used to prepare material suitable for mass spectrometry. This purification step used a column (bed volume approximately 200µL) packed with AlbuPure™ matrix (ProMetic BioSciences Ltd, Cambridge UK or Novozymes Biophama UK Ltd.). This was equilibrated with 50mM sodium phosphate, pH5.3, and loaded with neat culture supernatants, at approximately pH5.5-6.5, to approximately 40mg protein/mL matrix. The column was washed with approximately 3 column volumes each of 50mM sodium phosphate, pH5.3, and 50mM ammonium acetate, pH8.0, respectively. Bound protein was eluted using approximately 5 column volumes of 50mM ammonium acetate, 10mM octanoate, pH7.0. The flow rate for the load step was 137µL/min, while the wash and

elution steps were performed by means of centrifugal force, using a Heraeus Multifuge 3 centrifuge at 300 rpm. Final concentrations were in the range 1.8-4.0 mg/mL and samples were approximately 2mL volume. Free thiol determination was performed immediately after sample elution by following the procedure described below.

The number of free thiols on a protein can be determined spectrophotometrically using Ellman's reagent. Ellman's reagent (5'5'-dithio-bis(2-nitronenzoic acid) (DTNB)) is an aromatic disulphide which reacts with thiol groups to form a mixed disulphide of the protein and one mole of 5-thio-2-nitrobenzoic acid (TNB) (per mole of protein sulfhydryl group). This reaction also results in a yellow colour from free TNB being released in solution. Alternatively the number of free thiols on a protein can be determined using mass spectrometric analysis of protein sample treated with DTNB reagent. 5-thio-2-nitrobenzoic acid (TNB) has a molecular weight of 199 Da, thus an increase in mass of 197 Da (TNB minus H_2 lost during disulphide bond formation with the free thiol group on the test protein) indicates the presence of one free thiol group on the protein sample.

700 µL of the test protein sample was added to 100 µl Buffer 2 (4 mg/mL DTNB and 500 mM Sodium Phosphate, pH 7.0) and 900 µL Buffer 1 (0.1M Tris-HCl, 100 mM EDTA, pH8.0). The preparation was allowed to mix for 25 minutes at ambient temperature (21-25 °C) followed by filtration through a low molecular mass cut-off filter (Vivaspin 2 – 10000 MWCO Sartorius Stedim Germany). The filter was washed with two volumes of 0.1% Trifluoroacetic acid (TFA) and the sample was resuspended in 1 ml of 0.1% TFA. TNB labelled and unlabelled samples were prepared for mass spectrometric analysis by desalting/concentrating using Solid Phase Extaction (SPE). SPE columns were prepared by first wetting with 1mL of 70% Acetonitrile (ACN Fisher) / 0.1% TFA and then equilibrating ready for loading with 0.1% TFA. 1mL of sample was loaded on the equilibrated SPE columns allowing time for the protein to bind. The bound protein and SPE columns were then washed three times in 1mL of 0.1% Formic acid (Merck). Finally the bound protein was eluted into pre-washed 1mL microfuge tube with 0.5mL 70% ACN/0.1%FA.

For Time-of-Flight mass spectrometry 30µL of sample was introduced into a hybrid quadru-pole time-of flight mass spectrometer (QqOaTOF, Applied Biosystems, QSTAR-XL®), equipped with an IonSprayTM source in positive ion mode, using flow injection analysis (FIA). The only instrument parameter that is actively tuned is the Decoupling Potential (DP), typically set to 250 V. Typically 2 minutes of sample scans are averaged. For protein analysis the TOF analyser is calibrated against protonated molecular ions of equine myoglobin (Sigma) and resolution is typically >14,000. Instrument control and data acquisition and processing were performed using AnalystTM QS v1.1 software (Applied Biosystems).

The results of the above analysis of the purified thio-albumin samples are described below. On addition of DTNB all samples quickly turned yellow as expected due to the presence of numerous free thiols. When the samples were visually compared to an equivalent sample of rHA, containing a single free thiol, the colour change observed for the thio-albumin samples was significantly more intense, strongly indicating the presence of multiple free thiols on each thio-albumin molecule. Results are summarised in Figure 18, with increasing colour intensity increasing denoted by increased number of "+".

The thio-albumin variants produced at higher fermentation yields were preferred for analysis by the mass spectroscopy method described above. Therefore, the recombinant proteins rHA (A2C, L585C) (total of 3 free thiols), rHA (D129C, C360S, L585C) (total of 4 free thiols), and A2C rHA-Cys (total of 3 free thiols) were analysed by ESI TOF (electrospray ionsation time of flight) mass spectrometry pre- and post-DTNB treatment to determine the numbers of free thiols present on each molecule.

When rHA (A2C, L585C) was analysed pre-DTNB treatment (Figure 19) the major deconvoluted peaks observed were at 66633Da and 66807Da which corresponds to 172Da and 346Da above the expected mass of 66461Da. These modifications were likely to be due to species, of ~172Da present in the growth media cross linking to the free thiols in rHA (A2C, L585C). Post DTNB treatment mass spectrometric analysis (Figure 20) resulted in a major a deconvoluted peak at 67428Da which was 376Da above the expected mass for the protein with 3 free thiols. This extra mass is most likely to be due to an extra TNB linked to a free thiol and a further 179Da, this strongly suggests the presence of a 4 free thiols and a possible a further thiol blocked with a species of ~179Da. Hence, the rHA (A2C, L558C) thio-albumin variant is particularly surprising in that it provides more than the expected number of reactive groups available for conjugation. Also present is a series of peaks ~396Da apart which are due to excess DTNB still present at the time of ionisation causing DTNB adduct formation with the labelled rHA (A2C, L558C) molecule. This adduct formation is known to occur in the presence of excess DTNB.

When rHA (D129C, C360S, L585C) (total of 4 free thiols) was analysed pre-DTNB treatment (Figure 21) the major deconvoluted peaks observed were at 66575Da and 66747Da which corresponds to 172Da and 344Da above the expected mass of 66403Da. These modifications were likely to be due to molecules of ~172Da present in the growth media cross linking to the free thiols in rHA (D129C, C360S, L585C). Post DTNB treatment mass spectrometric analysis (Figure 22) resulted in a major a deconvoluted peak at 67564Da which was 373Da above the expected mass for the protein with 4 free thiols. This extra mass is most likely to be due to an extra TNB linked to a free thiol and a further 176Da, this

strongly suggests the presence of a 5 free thiols and a possible a further thiol blocked with a species of ~176Da. Hence, the rHA (D129C, C360S, L585C) thio-albumin variant is particularly surprising in that it has provided more than the expected number of groups available for conjugation. Also present are a series of peaks ~396Da apart these are due to excess DTNB still present at the time of ionisation causing DTNB adduct formation with the labelled rHA D129C, C360S, L585C. This adduct formation is known to occur in the presence of excess DTNB.

Finally when A2C rHA-Cys (total of 3 free thiols) was analysed pre-DTNB treatment (Figure 23) the major deconvoluted peaks observed were at 66747Da and 66919Da which corresponds to 172Da and 344Da above the expected mass of 66574Da. However also present was a smaller peak corresponding to the expected unmodified mass at 66575Da. This mass spectra indicates the blocking of some free thiols while a proportion of the molecule is present containing the expected 3 free thiols. Post DTNB treatment mass spectrometric analysis (Figure 24) resulted in a major a deconvoluted peak at 67142Da which was 23Da below the expected mass of 67165Da of the protein labelled with 3 TNB molecules, this was likely due to the presence of two TNB molecules and a 175Da modification, suggesting the presence of 3 thiols, one of which was blocked by the unknown ~175Da species. However on closer inspection of Figure 24 the presence of secondary peaks 23Da above each species can be seen. These secondary peaks correspond to small shoulders in the raw data (data not shown) which are likely to indicate the presence of the molecule modified with 3 TNB molecules, indicating 3 free thiols. The other major species present are due to excess DTNB still present at the time of ionisation causing DTNB adduct formation with the labelled rHA D129C, C360S, L585C. This adduct formation is known to occur in the presence of excess DTNB.

In conclusion, a range of thio-albumin variants have been produced with three or more conjugation competent cysteine residues. The conjugation competent cysteines can be in regions of that may or may not have secondary structure, and/or may or may not be generated from natural disulphide bonds, and/or may or may not be additional cysteines residues (such as cysteine residues extending from the natural C-terminus of HSA).

EXAMPLE 5: FORMATION OF GELS FROM THIO-ALBUMIN VARIANTS

Samples of TA35 (*i.e.* A2C, A364, D562 in addition to naturally occurring C34) and TA33 (*i.e.* A2C, L585C in addition to naturally occurring C34) were incubated at room-temperature for 24 hours and both formed gels.

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.