(11) Application No. AU 2003238441 B2

(19) AUSTRALIAN PATENT OFFICE

(54) Title

Protein tag comprising a biotinylation domain and method for increasing solubility and determining folding state

(51)[6] International Patent Classification(s)

G01N 33/68 (2006.01)
C07K 14/47 (2006.01)
C07K 19/00 (2006.01)
C12N 9/00 (2006.01)
C12N 15/00 (2006.01)
C12N 15/62 (2006.01)
C12N 15/67 (2006.01)
C12P 21/02 (2006.01)
G01N 33/68 20060101AFI2005122 0BMJP
C07K 14/47 20060101ALI2005122 0BMJP
C07K 19/00 20060101ALI2005122 0BMJP
C12N 9/00 20060101ALI2005100 8BMEP
C12N 15/00 20060101ALI2005122 0BMJP
C12N 15/62 20060101ALI2005100 8BMEP
C12N 15/67 20060101ALI2005100 8BMEP
C12P 21/02 20060101ALI2005100 8BMEP
PCT/GB03/00362

(21) Application No: 2003238441 (22) Application Date: 2003.01.29

(87) WIPO No: WO03/064656

(30) Priority Data

(31) Number	(32) Date	(33) Country
0202018.8	2002.01.29	GB

(43) Publication Date : 2003.09.02

(43) Publication Journal Date : 2003.09.18

(71) Applicant(s)

Sense Proteomic Limited

(72) Inventor(s)

Hart, Darren James; Samaddar, Mitali; Dyson, Michael Richard; Blackburn, Jonathan Michael

(74) Agent/Attorney

Davies Collison Cave, Level 15 1 Nicholson Street, Melbourne, VIC, 3000

(56) Related Art

US 5932433 A (Schatz) 3 August 1999

WO 1990/014431 A1 (Biotechnology Research and Development Corporation) 29 November 1990

WO 1995/025172 A1 (Universite Louis Pastuer) 21 September 1995

Germino et al, PNAS, 1993, 90:933-937

US 5723584 A (Schatz) 3 March 1998

Zhen et al, Gene, 1996, 173: 147-154

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
7 August 2003 (07.08.2003)

PCT

(10) International Publication Number
WO 03/064656 A1

(51) International Patent Classification[7]: C12N 15/62, 15/70, 15/10

(21) International Application Number: PCT/GB03/00362

(22) International Filing Date: 29 January 2003 (29.01.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0202018.8 29 January 2002 (29.01.2002) GB

(71) Applicant *(for all designated States except US)*: **SENSE PROTEOMIC LIMITED** [GB/GB]; The Babraham Bioincubator, Babraham Hall, Cambridgeshire CB2 4AT (GB).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **SAMADDAR, Mitali** [IN/IN]; Dr. Reddy's Laboratories Ltd, Biotechnology Division, 7-1-27, Ameerpet, Hyderabad 500 016 (IN). **BLACKBURN, Jonathan, Michael** [GB/GB]; 36 Woodlark Road, Cambridge CB3 0HS (GB). **HART, Darren, James** [GB/GB]; Sense Proteomic Limited, The Babraham Bioincubator, Babraham Hall, Cambridgeshire CB2 4AT (GB). **DYSON, Michael, Richard** [GB/GB]; Sense Proteomic Limited, The Babraham Bioincubator, Babraham Hall, Cambridgeshire CB2 4AT (GB).

(74) Agents: **KILBURN & STRODE** et al.; 20 Red Lion Street, London WC1R 4PJ (GB).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 03/064656 A1

(54) Title: PROTEIN TAG COMPRISING A BIOTINYLATION DOMAIN AND METHOD FOR INCREASING SOLUBILITY AND DETERMINING FOLDING STATE

(57) Abstract: The use of a tag moiety comprising a biotinylation domain, such as biotin carboxyl carrier protein (BCCP), as a protein folding marker and protein solubility enhancer in the orientated surface capture of products of heterologously expressed genes is described. Methods for increasing the solubility of proteins and determining the folded state of a protein are also disclosed. The uses and methods of the invention can be carried out in a multiplexed manner on more than one protein in the formation of libraries. In addition the nucleic acid molecule encoding the biotinylation domain of the tag moiety can be used to increase the proportion of clones in a library that express the protein of interest.

PROTEIN TAG COMPRISING A BIOTINYLATION DOMAIN AND METHOD FOR INCREASING SOLUBILITY AND DETERMINING FOLDING STATE

This invention relates to the use of biotin carboxyl carrier protein (BCCP) as a protein folding marker and protein solubility enhancer in the orientated surface capture of products of heterologously expressed genes.

Expression of human proteins in heterologous systems such as bacteria, yeast, insect cells or mammalian cells can result in the production of incorrectly folded proteins resulting in the formation of insoluble aggregates or a low yield of expressed proteins because of the targeting of the unfolded proteins to the proteosome. For all functional protein procedures the production of correctly folded or native proteins is essential and a great deal of work is often performed to optimise the expression of individual proteins. However, many areas of protein biochemistry involve working with libraries or groups of proteins of such a size that optimisation of individual expression and purification conditions for each protein is impractical. Hence, there exists an unmet need in the art for reagents, protocols and methology that facilitate the multiplexing of these processes.

Affinity tags are a convenient method of purification and immobilisation of recombinant proteins. Hexahistidine tags (6 amino acids (aa); Qiagen/Roche), *Escherichia coli* maltose binding protein ("MBP", 300 aa; New England Biolabs) and *Schistosoma japonicum* glutathione-S-transferase (GST, 220 aa; Amersham Pharmacia Biotech/Novagen) are effective, but have the disadvantage that heterologous host proteins interact with the affinity matrices used for purification of fusion proteins. This results in impure protein preparations and an additional clean up step is often required. Additionally, the relatively weak affinity of these proteins for their ligands results in dissociation, or "leaching" of the fusion proteins from surfaces to which they are immobilised. Such reversible interactions are exploited during resin-based purifications on resins in column or batch formats where,

because of the high local concentrations of ligand, dissociated proteins rapidly rebind, yet are rapidly eluted by free ligand. In contrast, immobilisation of proteins to planar surfaces such as microtiter plates or microarrays, for example, biochips, requires that they remain bound and do not leach from the substrate during storage and use. As such, lower affinity tags as used for purification (e.g. MBP, GST and hexahistidine tags) are suboptimal. Frequently, covalent immobilisation strategies are employed such as coupling of purified proteins via surface lysine residues to amine-reactive chemical groups. This is generally accepted to result in reduced activity of the protein.

In contrast to the lower affinity, non-covalent interactions described above, the interaction of biotin with streptavidin, avidin or neutravidin exhibits some of the highest affinities known in biology, with equilibrium dissociation constants of 10^{-15} M (several orders of magnitude higher affinity than the MBP – amylose or GST – glutathione interactions). Whilst still a weaker interaction than covalent coupling, biotinylated proteins bound to a streptavidin-derivatised surface show negligible dissociation. This interaction therefore provides a improved means for tethering proteins to a planar surface for applications such as protein arrays and enzyme-linked immunoassays (ELISAs).

Biotin can be attached chemically to proteins (e.g. using NHS-activated biotin), or via genetically fused protein domains which are biotinylated *in vivo*. The "PinPoint™" vectors from Promega are designed to facilitate the creation of fusions to the biotinylation domain (which is a fragment of the biotin carboxyl carrier protein (BCCP) of methylmalonyl-CoA carboxyl transferase from *Propionibacterium freudenreichii shermanii* [US Patent 5,252,466]). This protein has 40% homology with the *E. coli* BCCP. This system allows the production of BCCP-protein fusions capable of being biotinylated either *in vivo* or *in vitro* by biotin ligase, allowing one to use the highly specific biotin – streptavidin interaction

for surface capture. In addition to the BCCP domain, phage display selected short peptides capable of being biotinylated on a lysine residue have been commercialised by Avidity Inc. [US Patent 5,932,433].

The Inventors herein describe a novel approach whereby BCCP from *E. coli* is fused either N- or C-terminally to a protein partner. In addition to the function of permitting orientated immobilization of the fusion protein to microarray compatible surfaces derivatised with avidin, streptavidin or neutravidin. The Inventors describe new, previously unreported functions of BCCP which greatly facilitate the creation of libraries of solubly expressed folded human, mammalian, fungal, plant or microbial proteins in heterologous systems.

i) N-terminally or C-terminally fused BCCP improves levels of folding of fusion partner

The factors determining the solubility of recombinant proteins are poorly understood and so rational design of solubility and increased expression into recombinant proteins is only possible to a limited extent. However, by fusing well expressed soluble proteins to the N-terminus of a protein, both properties can be greatly improved compared with expression of ORFs alone. Examples include MBP, GST and thioredoxin (Trx, 109 aa; Novagen). A possible mechanism of action is thought to be the recruitment of chaperones to the nascent polypeptide and co-over-expression of chaperones can result in increased yield of soluble protein. Some fusion proteins can then be purified via their fusion protein domain (e.g. amylose resin for MBP or glutathione resin for GST. Although the Trx tag has not been used for protein purification it can both improve the solubility of many target proteins and it appears to catalyse the formation of disulphide bonds in the cytoplasm of *E. coli* trx B mutants.

The Inventors have determined that addition of BCCP to the N-terminus or C-terminus of a protein increases the solubility of the fusion protein and in the case of addition to the N-terminus at least, increases the proportion of clones in a library that express encoded proteins (relative to a library that is not modified to also encode a BCCP tag). Additionally, the BCCP domain is biotinylated *in vivo*. This is particularly useful when attempting to multiplex protein purification for fabrication of protein arrays since the proteins can be simultaneously purified from cellular lysates and immobilised in a single step via the high affinity and specificity exhibited by a streptavidin surface. The Inventors term this simultaneous purification and immobilisation as "surface capture".

ii) N-terminally or C-terminally fused BCCP permits monitoring of folding of fusion partner

Fusion of reporter proteins (with an assayable activity) onto the C-terminus of partner proteins has been previously shown to allow monitoring of the folding of the partner. Notable examples of reporter systems known in the art utilise green fluorescent protein (GFP), chloramphenicol acetyl transferase (CAT), β-galactosidase and the α-complementation of β-galactosidase.

The Inventors have determined that addition of BCCP to the N-terminus or C-terminus of a protein permits the monitoring of fusion protein folding by measuring the extent of *in vivo* biotinylation. This can be measured by standard blotting procedures, using SDS-PAGE or *in situ* colony lysis and transfer of samples to a membrane, followed by detection of biotinylated proteins using a streptavidin conjugate such as streptavidin-horseradish peroxidase. Importantly, the addition of biotin to the BCCP domain permits purification by surface capture as described above.

Thus in a first aspect the invention provides the use of a tag moiety comprising a biotinylation domain for increasing the solubility of a protein of interest by

attachment of said tag moiety to the N-terminal or C-terminus of said protein of interest.

A tag moiety comprising a biotinylation domain as defined herein is an amino acid sequence comprising a protein or protein domain which is capable of being biotinylated, or to which a biotin group can be attached. In accordance with the first aspect of the invention the tag is highly soluble in the cytoplasm of the host cell in which it is expressed as a tag attached to a protein of interest.

Essentially, the biotinylation domain of the invention is a protein or protein domain having secondary and tertiary structure and which is biotinylated *in vivo* post translationally. Generally the secondary and tertiary structure of the protein or domain is essential for recognition and hence biotinylation by the biotin ligase of the host cell in which expression of the tag is taking place.

Preferably the biotinylation domain of the tag comprises the sequence of *E. coli* BCCP (Biotin Carboxyl Carrier Protein of Acetyl-Coa Carboxylase (ACCB) - Swiss-Prot Database Accession no. P02905), the nucleotide and amino acid sequence of which is:

BCCP domain:
Nucleotide

gcagcagcggaaatcagtggtcacatcgtacgttccccgatggttggtactttcta
ccgcaccccaagcccggacgcaaaagcgttcatcgaagtgggtcagaaagtcaacg
tgggcgataccctgtgcatcgttgaagccatgaaaatgatgaaccagatcgaagcg
gacaaatccggtaccgtgaaagcaattctggtcgaaagtggacaaccggtagaatt
tgacgagccgctggtcgtcatcgagtaa

Amino acid:
AAAEISGHIVRSPMVGTFYRTPSPDAKAFIEVGQKVNVGDTLCIVEAMKMMNQIEA
DKSGTVKAILVESGQPVEFDEPLVVIE-

Alternatively, other sequences encoding BCCP known in the art can be used as the biotinylation domain of the invention, for example other BCCP proteins from the Swiss-Prot database:

BCCA_MYCLE (**P46392**)
Acetyl-/propionyl-coenzyme A carboxylase alpha chain [Includes: Biotin carboxylase (EC 6.3.4.14); Biotin carboxyl carrier protein (BCCP)]. {GENE: BCCA OR ML0726 OR B1308_C1_129} - Mycobacterium leprae

BCCA_MYCTU (**P46401**)
Acetyl-/propionyl-coenzyme A carboxylase alpha chain [Includes: Biotin carboxylase (EC 6.3.4.14); Biotin carboxyl carrier protein (BCCP)]. {GENE: ACCA1 OR BCCA OR RV2501C OR MT2576 OR MTCY07A7.07C} - Mycobacterium tuberculosis

BCCP_ANASP (**Q06881**)
Biotin carboxyl carrier protein of acetyl-CoA carboxylase (BCCP). {GENE: ACCB} - Anabaena sp. (strain PCC 7120)

BCCP_ARATH (**Q42533**)
Biotin carboxyl carrier protein of acetyl-CoA carboxylase, chloroplast precursor (BCCP). {GENE: CAC1 OR BCCP1 OR AT5G16390 OR MQK4.12} - Arabidopsis thaliana (Mouse-ear cress)

BCCP_BACSU (**P49786**)
Biotin carboxyl carrier protein of acetyl-CoA carboxylase (BCCP). {GENE: ACCB OR FABE} - Bacillus subtilis

BCCP_CHLMU (**Q9PKR5**)
Biotin carboxyl carrier protein of acetyl-CoA carboxylase (BCCP). {GENE: ACCB OR TC0399} - Chlamydia muridarum

BCCP_CHLPN (**Q9Z901**)
Biotin carboxyl carrier protein of acetyl-CoA carboxylase (BCCP). {GENE: ACCB OR CPN0183 OR CP0585} - Chlamydia pneumoniae (Chlamydophila pneumoniae)

BCCP_CHLTR (**O84125**)
Biotin carboxyl carrier protein of acetyl-CoA carboxylase (BCCP). {GENE: ACCB or CT123} - Chlamydia trachomatis

BCCP_CYACA (**O19918**)
Biotin carboxyl carrier protein of acetyl-CoA carboxylase (BCCP). {GENE: ACCB} - Cyanidium caldarium [Chloroplast]

BCCP_ECOLI (**P02905**)
Biotin carboxyl carrier protein of acetyl-CoA carboxylase (BCCP). {GENE: ACCB OR FABE OR B3255 OR Z4615 OR ECS4127} - Escherichia coli, Escherichia coli O157:H7

BCCP_HAEIN (**P43874**)
Biotin carboxyl carrier protein of acetyl-CoA carboxylase (BCCP). {GENE: ACCB OR FABE OR H10971} - Haemophilus influenzae

BCCP_LYCES (**P05115**)

Biotin carboxyl carrier protein of acetyl-CoA carboxylase (BCCP) (Fragment). - Lycopersicon esculentum (Tomato)

BCCP_PORPU **(P51283)**
Biotin carboxyl carrier protein of acetyl-CoA carboxylase (BCCP). {GENE: ACCB} - Porphyra purpurea [Chloroplast]

BCCP_PROFR **(P02904)**
Biotin carboxyl carrier protein of methylmalonyl-CoA carboxyl-transferase (Transcarboxylase, 1.3S subunit). - Propionibacterium freudenreichii shermanii

BCCP_PSEAE **(P37799)**
Biotin carboxyl carrier protein of acetyl-CoA carboxylase (BCCP). {GENE: ACCB OR FABE OR PA4847} - Pseudomonas aeruginosa

BCCP_SOYBN **(Q42783)**
Biotin carboxyl carrier protein of acetyl-CoA carboxylase, chloroplast precursor (BCCP). {GENE: ACCB-1} - Glycine max (Soybean)

BCCP_STRMU **(P29337)**
Biotin carboxyl carrier protein (BCCP). - Streptococcus mutans

Also included within the scope of the invention are biotinylation domains encoded by or comprising artificial sequences, for example where one or more amino acids have been altered by conservative substitution. Such sequences can be rationally designed or derived from the sequences of BCCP given above, by methods known in the art. It is essential that these sequences have a secondary and tertiary structure that permits the artificial sequence to be recognised and biotinylated by a biotin ligase enzyme.

In a second aspect, the invention provides the use of a tag moiety comprising a biotinylation domain for determining the folded state of a protein of interest by attachment of said tag moiety to the N-terminus or C-terminus of said protein of interest.

In this second aspect, the tag moiety comprising a biotinylation domain as defined herein is a protein or protein domain which is conditionally biotinylated by a biotinylating enzyme, for example biotin ligase expressed in the host cell in which expression takes place or exogenously applied biotin ligase, for example, used to biotinylate proteins in a cell-free extract. Essentially, the domain can only be

biotinylated through recognition of the folded structure of the domain by the enzyme such that the domain in linear, mis-folded or aggregated, form for example in inclusion bodies, is not biotinylated. The folding of the tag and its subsequent biotinylation is dependent on the correct folding of the protein N-terminal to the C-terminal tag and vice versa.

In a third aspect the invention provides a method of increasing the solubility of a protein of interest when expressed in a host cell comprising the steps of:

a) attaching a first nucleic acid molecule encoding a tag moiety comprising a biotinylation domain to a second nucleic acid molecule encoding said protein of interest to form a construct such that the tag moiety in the expressed product of the combined first and second nucleic acid molecules comprises said tag moiety located at the N-terminus or C-terminus of said protein of interest

b) expressing said construct in a host cell

In a fourth aspect the invention provides a method of determining the folded state of a protein of interest comprising the steps of:

a) attaching a first nucleic acid molecule encoding a tag moiety comprising a biotinylation domain to a second nucleic acid molecule encoding said protein of interest to form a construct such that the tag moiety in the expressed product of the combined first and second nucleic acid molecules comprises is located at the N-terminus or C-terminus of said protein of interest

b) expressing said construct in a host cell under conditions such that only a correctly folded biotinylation domain present in said tag moiety is ligated with biotin

c) determining the folded state of the protein of interest comprising said tag moiety by the presence or absence of a biotin group in the protein expressed from said construct

The uses of the first and second aspect of the invention and the methods of the third and fourth aspects of the invention are preferably carried out in a multiplexed manner on more than one protein of interest. For example, wherein the protein of interest is encoded by nucleic acid molecule which forms part of a library comprising two or more different coding sequences and, optionally, wherein the different coding sequences are modified to contain the tag moiety and expressed in parallel.

Thus in a fifth aspect the invention provides a library of nucleic acid molecules encoding proteins of interest wherein each coding sequence is modifed to incorporate at the N-terminus or C-terminus of the encoded protein a tag moiety comprising a biotinylation domain. Such libraries may be generated using known techniques in the art. Usefully, the library can be generated using the COVET methodology described in WO 01/57198.

Accordingly, in a sixth aspect, the invention provides a library of proteins produced from the methods of the third and fourth aspects of the invention or expressed from the library of the fifth aspect of the invention. Such libraries may be arrayed on a solid substrate, for example through immobilisation to that substrate via, for example, a streptavidin-biotin link via the BCCP tag present on the proteins of the library.

The Inventors have also determined that the addition of DNA encoding a BCCP tag 5' to and in-frame with genes of interest in a library has the effect of significantly increasing the number of encoded proteins of interest which are expressed from that library compared to a library encoding the same proteins, but lacking the BCCP tag encoding sequence. Such relative expression differences between "tagged" and "un-tagged" libraries can be detected or measured qualitatively, for example using western blotting techniques as known in the art.

Thus, in a seventh aspect, the invention provides the use of a nucleic acid molecule encoding a tag moiety comprising a biotinylation domain for increasing the proportion of clones in a library that express the protein of interest encoded by each of said clones at detectable levels, for example as measured by conventional western blotting, by attachment of said nucleic acid molecule encoding said tag 5' to and in-frame with the gene encoding said protein of interest in each of said clones.

Accordingly in an eighth aspect, the invention provides a method of increasing the proportion of clones in a library that express the protein of interest encoded by each of said clones in a host cell at detectable levels, comprising the steps of:

a) attaching a first nucleic acid molecule encoding a tag moiety comprising a biotinylation domain 5' to and in-frame with a second nucleic acid molecule encoding said protein of interest in a clonal member of said library to form a construct such that the tag moiety in the expressed product of the combined first and second nucleic acid molecules comprises said tag moiety located at the N-terminus of said protein of interest

b) expressing said construct in a host cell

Preferred features of each aspect of the invention are as defined for each other aspect, *mutatis mutandis*.

Whilst the tags, methods and libraries of the invention are particularly suited to facilitating parallel expression and purification/immobilisation of proteins encoded by a library of sequences (by a common method of solublisation and purification of the proteins of interest), the invention can also be applied to other methodologies known in the art. For example, an N-terminal or C-terminal tag according to the invention (for example BCCP) can be used to increase both protein expression and solubility in:

- Vaccine production

- Therapeutic protein production
- Antigen production used for the generation of monoclonal or polyclonal antibodies, monoclonal antibody or single chain antibody production
- Enzyme production
- Drug target discovery by mapping cellular protein-protein interactions "the interactome"
- Drug target validation by generation of protein drug targets including, but not exclusively, kinases, phosphatases, cell receptors or proteases for screening, enzyme and / or toxicology studies and any other biochemical analysis.

The invention will now be further described by the following non-limiting examples which refer to the accompanying figures in which:

Figure 1 shows the colony western data using Streptavidin-HRP conjugate as the probe. The clones expressing in-frame GFP-BCCP that fluoresced green are also biotinylated. The bottom row are clones that harbour pMSC301 (no *bccp* gene sequence in the plasmid), and signal obtained is the background signal of endogenous biotinylated AccB. The second row from the bottom are the clones harbouring pMSC302 (overexpressing *accB*). The other negative clones (out of frame fusions or vector re-ligated did not fluoresce green and were not biotinylated).

Figure 2 shows colony western data using Streptavidin-HRP conjugate as the probe. The clones expressing in-frame GST-GFP-BCCP that fluoresced green are also biotinylated. Also shown as biotinylation positive signal is the protein GST-BCCP. The negative control is clones that harbour pMSC301 (no *bccp* gene sequence in the plasmid), and signal obtained is the background signal of endogenous biotinylated AccB. The positive control is the clone harbouring pMSC302 (overexpressing *accB*). The other negative clones (out of frame fusions or vector re-ligated did not fluoresce green and were not biotinylated).

Figure 3 shows western blot analysis of the protein extract from cells expressing GFP-BCCP. The signal obtained at approximately 37 kDa., is the expected *Mr* of GFP-BCCP. Another signal seen at 18 kDa is that of endogenous biotinylated AccB protein, also seen in the GFP-BCCP negative lanes. As expected, the 18 kDa. signal is stronger, when no recombinant biotinylated protein is expressed.

Lanes 1, 2 and 3: Protein extract from clones harbouring pGFP-BCCP, expressing intact GFP-BCCP protein.

Lanes 4, 5 and 6: Protein extract from clones harbouring pMSC301A, B, and C respectively, used as negative control in the experiment.

Figure 4 shows western blot analysis of protein extracts from cells expressing GST-GFP-BCCP, and GST-BCCP. Biotinylated proteins of expected *Mr.* are observed (63 kDa for GST-GFP-BCCP and 37 kDa for GST-BCCP). In all the lanes 18 kDa signal for endogenous AccB is present.

Lanes 1, 2 and 4 are protein extract from cells expressing GST-GFP-BCCP.

Lane 3 is the protein extract from cells expressing GFP-BCCP as a positive control in this expt.

Lanes 5 and 6: Protein extract from clones harbouring pMSC301A, and B as negative controls in the blot.

Lanes 7 and 8: Protein extracts from cells expressing GST-BCCP.

Figure 5 shows a colony western blot using streptavidin-HRP as the probe for biotinylation of BCCP in the fusion protein. All clones that were marked to be fluorescing green when excited at 365 nm wavelength, were also biotinylated (positive signal above the background). The intensities of positive signals varies as

does the green phenotype. Increased sensitivity of detection using streptavidin-HRP conjugate, picked up few additional clones.

Figure 6 shows protein expression results of the human gene set cloned into the Avi-Tag vector pQE82L-GFP-biotin. Single ampicillin resistant colonies were used to inoculate 1 ml of LB media containing 100 µg/ ml ampicillin (LB-Amp) and grown over-night at 37°C with shaking. The next day a 1:100 dilution was made into fresh LB-Amp and cells grown at 37°C until OD600 = 0.6 to 1.0. IPTG was then added to a final concentration of 1 mM and growth continued at 30°C for 4 hours. 10 µl of cell culture was then taken and analysed by 4 – 20% SDS-PAGE Western blot and probed with HRP-conjugated streptavidin. Numbers labeled for each lane refer to the B# in Table 1. The molecular weight markers are: aprotin (7.6 kDa), lysozyme (18.4 kDa), soybean trysin inhibitor (32.5 kda), carbonic anhydrase (45.7 kDa), BSA (78 kDa), B-galactosidase (132 kDa) and myosin (216 kDa).

Figure 7 shows protein expression results of the human gene set cloned into the BCCP expressing vector pMD004. Single ampicillin resistant colonies were used to inoculate 1 ml of LB media containing 100 µg/ ml ampicillin (LB-Amp) and grown over-night at 37°C with shaking. The next day a 1:100 dilution was made into fresh LB-Amp and cells grown at 37°C until OD600 = 0.6 to 1.0. IPTG was then added to a final concentration of 1 mM and growth continued at 30°C for 4 hours. 10 µl of cell culture was then taken and analysed by 4 – 20% SDS-PAGE Western blot and probed with HRP-conjugated streptavidin. Numbers labeled for each lane refer to the B# in Tables 1 and 2. The molecular weight markers are: aprotin (7.6 kDa), lysozyme (18.4 kDa), soybean trysin inhibitor (32.5 kDa), carbonic anhydrase (45.7 kDa), BSA (78 kDa), B-galactosidase (132 kDa) and myosin (216 kDa).

Figure 8 shows plasmid maps of pMD002 and pMD004.

Figure 9 shows a plasmid map of pIFM101A/B/C

Figure 10 shows the cloning site of plasmid pIFM101A

Figure 11 shows the cloning site of plasmid pIFM101B

Figure 12 shows the cloning site of plasmid pIFM101C

EXAMPLES

Example 1: Use of BCCP as a Protein Folding Marker

Methods:

1. **Isolation of Biotin carboxyl carrier protein (C-terminal domain of acetyl-CoA carboxylase) from *E. coli* K 12 strain**

The DNA sequence encoding the entire coding region of acetyl-CoA carboxylase was amplified by PCR from genomic DNA of XL1-Blue (Stratagene) cells, using the following gene specific primers.

accbfor1: 5' GATGGATCCGATATTCGTAAGATTAAAAAACTGATCG 3' with *Bam*HI site at the 5' end.

bccprev1: 5'
GATGAGCTCAAGCTTTTACTCGATGACGACCAGCGGCTCGTC 3'
containing *Sac*I and *Hin*dIII site.

The PCR amplification was carried out using Pwo polymerase (Roche) using standard cycling conditions (94°C 5 min; 94°C 30 sec; 64°C 1 min; 72°C 1min; 30 cycles; 72°C 5 min).

The PCR amplified gene sequence was cloned into the *Bam*HI and *Sac*I site of the *E. coli* expression vector pQE-80 (Qiagen) inframe with the N-terminus hexahistidine tag to form the plasmid pMSC302. The identity of the gene sequence was confirmed by restriction mapping and DNA sequencing. The DNA sequence corresponding to the C-terminal domain of AccB known as biotin carboxyl carrier protein (BCCP) was amplified by PCR using the same reverse primer as above and a new forward primer.

bccpfor1:

5'GATCTGCAGGGCTCCGCAGCAGCGGAAATCAGTGGTCACATCG 3'

containing *Pst*I site for cloning and two extra codons for glycine and serine.

2. Construction of vectors:

The vector pQE-80 was redesigned to delete the DNA sequence for hexahistidine tag, add additional cloning sites (*Not*I and *Sfi*I), and have three different reading frames from the start ATG (pMSC301A/B/C). This was carried out by inverse PCR using the primer sets; pQErev1: 5'P CATAGTTAATTTCTCCTCTTTAATGAATTCTG 3'; pQEfwd1: 5' GCGGCCGCGGCCATTACGGCCGGATCCGCATGCGAGCTCGG TACCCCC 3'; pQEfwd2: 5' G + pQEfwd1; pQEfwd3: 5' GC + pQEfwd1 for A, B, and C reading frames respectively. The PCR was carried out using Pwo polymerase (94°C 2 min; 94°C 30 sec; 63.5°C 1 min; 72°C 6min; 25 cycles; 72°C 10 min).

The *bccp* gene sequence was cloned into the *Pst*I-*Hin*dIII sites of pMSC301 A, B, and C vectors to generate pMSC301A,B,C/BCCP.

The DNA sequence encoding GFPuv (Clontech) was amplified by PCR using the primer set pQEGFPfor1: 5' GGGCCGGTGGCAGCGCGAGTAAAGGAG AAGA ACTTTTCACTGG 3' (with *Sma*I half site and a linker region) and pQEGFPrev1: 5' GATCTGCAGGGTACCGGATCCTTTGTAGAGCTCATCCATGCC 3' (with *Pst*I, *Kpn I* and *Bam HI* sites). The PCR amplified product was cloned into the *Sma*I-*Pst*I sites of pMSC301A, B and C/BCCP in-frame to DNA sequence encoding the N-terminus of BCCP (GFP-BCCP) to generate the vectors pMSC303A, B, and C.

The plasmid construct pMSC303B was restricted with *Not*I, the staggered ends were made blunt using the filling in reaction of T4 Polymerase (NEB), restricted with *Sma I* and religated (plasmid designated as pGFP-BCCP).

The vectors pMSC301A/BCCP and pMSC303A were restricted with *Not*I, the overhangs blunted using T4 DNA polymerase, restricted with *Sma*I and were used to clone the DNA fragment encoding GST forming the plasmid constructs pGST-

BCCP and pGST-GFP-BCCP respectively. The DNA sequence encoding GST was amplified by PCR using the primers; GSTfwd01: 5' TCCCCTATACTAGGTTATTGG 3' and GSTrevexoN: 5' GGGCGTCACGA TGAATTCCCGGG 3' and pGEX-2T (Pharmacia) as template.

The *Not*I and *Sfi*I cloning sites of the vectors pMSC303A,B and C were replaced by the *Sfi*I overhang compatible restriction site, *Dra*III to generate the vectors pIFM101A, B, and C. This was carried out by inverse PCR using the primers; DrafwdA: 5' CACTTAGTGGGATCCGCATGCGAGCTCGGTACCCC 3'; DrafwdB: 5' G + DrafwdA; DrafwdC: GA + DrafwdA. The reverse primer used was pQErev1 as described earlier. The PCR conditions used were same as before.

A set of nested deletions recessed at 3' ends of human heart cDNAs (Clontech) were cloned into the *Dra*III-*Sma*I sites of the vectors pIFM101A, B, and C to form the plasmid pX-GFP-BCCP.

The correct DNA sequence of all the constructs used in the study were confirmed by sequencing.

3. Generation of nested deletions (recessed at 3' ends) of human heart cDNAs

The COVET methodology was used to generate the deletion set which is the subject of patent application Nos. GB0020357.0, USSN 60/247995 and WO 01/57198.

In brief, ~100ng template plasmid library (human heart cDNA library in pDNR-LIB from Clontech) was amplified by PCR using vector-specific primers SP5forward: 5'ATGCTCATGAGGCCGGCCGGGAATTC GGCCATTACG GCCGG3' with *Fse*I and *Sfi*I sites, and SP3reverse: 5'GTCTAGAAAGCTTCTCGAGGGCCG3', to optimally incorporate alpha-phosphothioate dTTPs (α-S-dTTP; Amersham). The PCR reaction was carried out using 50pmol each primer, 2.5 units thermostable polymerase (lacking a 3' to 5' exonuclease activity e.g. *Taq* polymerase), a standard buffer and the deoxynucleotide triphosphate mix: 200μM dATP, 200μM dGTP, 200μM dCTP, 100μM dTTP, 100μM α-S-dTTP. The PCR amplified products were purified using QIAquick PCR cleanup kits (Qiagen) and subjected to *Fse*I digestion to produce a 3' nucleotide overhang which protects the 5' end of the dsDNA from

subsequent hydrolysis by exonuclease III (NEB). Exonuclease III digestion was performed using standard conditions and the presence of phosphothioate internucleotide linkages blocked any further hydrolysis. This generated a nested set of sense strand 3' deletions. Mung bean nuclease (New England Biolabs) was used to remove ssDNA from the antisense strand and therefore blunt the dsDNAs in preparation for directional cloning after further digestion with *Sfi*I. These inserts after size fractionation by agarose gel electrophoresis were cloned into the *Dra*III and *Sma*I sites of the vectors pIFM101A, B and C. The ligated products were then used to transform XL1-Blue cells (Stratagene).

4. Expression of the fusion proteins

The *E. coli* strains XL1-Blue or XL10-Gold (stratagene) were used as host cells and were transformed (electroporation or chemical method) using various plasmid constructs. The transformation mixture was plated at an appropriate dilution on a nitrocellulose membrane placed on LB-Agar containing 100 µg/ml carbenicillin. After overnight incubation at 30°C the membranes were transferred onto LB-Agar containing 400 µM IPTG and carbenicillin and incubated for another 4-5 hrs at 30°C. The GFP activity of the clones were assessed by visualizing the clones at 365 nm wavelength of the UV-transilluminator. The membranes were processed for detecting biotinylated BCCP or GFP. For analysing the proteins by western blot the cultures were induced at mid log phase (optical density at 600 nm of 0.5 to 0.6) by adding 400 µM of IPTG to the culture and growth of cells continued for another 3-4 hours at 30°C. At the end of the induction period, cells were harvested, proteins resolved on 10-20 % gradient SDS-gel (Invitrogen), blotted onto nitrocellulose membrane and probed with various antibodies or streptavidin.

5. Detection of biotinylated BCCP

The biotinylation of BCCP was detected by probing with a streptavidin-horseradish peroxidase (HRP) conjugate (Amersham) on colony blots (as described) or on western blots as known in the art.

The clones were either gridded robotically, or the transformation mix was plated, onto nitrocellulose membrane (Amersham) placed on a LB agar plate containing carbenicillin. After overnight incubation at 30°C, the membrane was placed onto a fresh LB agar plate containing carbenicillin and IPTG (400μM). The plate was incubated for another 4-5 hours at 30°C. The colonies on the membrane were subjected to alkaline lysis and the membrane blocked prior to addition of the probe. The membrane is first placed on two sheets of Whatmann 3 paper pre soaked with 0.5 (M) NaOH, 1.5 (M) NaCl for 10 min. The membrane is neutralised by placing on Whatmann 3 sheets soaked with 1 (M) TrisHCl pH 7.5, 1.5 (M) NaCl for 5 min, two times. The membrane is then transferred onto Whatmann 3 sheets wetted in PBS-T (0.1%) containing 1% SDS for 10 mins. The membrane is then washed thoroughly in PBS-T ensuring that all the cell debris has been dislodged. The blot is then ready to be processed in the same manner as a western blot.

The Streptavidin-HRP conjugate was used at a dilution of 1:4000 and the signal was detected by chemiluminescence using the ECL system from Amersham.

6. Detection of GFP activity

The green fluorescence of GFP was visualized by exciting the colonies at 365 nm wavelength using a transilluminator.

7. Detection of GST

An anti-GST monoclonal antibody (Sigma) was used as an immunoprobe to detect expression of GST. The antibody was used at a dilution of (1:3000) and the immunoreactive signal was detected using the ECL system from Amersham.

Results

Absolute correlation of GFP activity and biotinylation of BCCP

Figures 1 and 2 show the colony western data using streptavidin-horseradish peroxidase as the probe. Only the correct in-frame fusion of GST-GFP-BCCP, GST-BCCP and GFP-BCCP gave strong positive signal significantly above the

general background from endogenous biotinylated AccB. Out-of-frame fusions resulting from the cloning strategy used, did not give rise to positive signals. All and only biotinylated fusion proteins (GST-GFP-BCCP and GFP-BCCP) fluoresced green when excited at 365 nm. The fluorescence is indicative of correct folding of the fusion protein and this result demonstrated that correctly folded proteins with BCCP as the C-terminal fusion partner is an active substrate for biotin protein ligase (BPL). Figures 3 and 4 show that the biotinylated proteins are of expected molecular weight, confirming the proteins as intact and unproteolysed.

A more comprehensive study of a group of proteins

Human heart cDNAs were recessed at 3' ends so as to remove the stop codon of the ORFs using controlled Exonuclease III (NEB) digestion. This 3' nested deletion set was then cloned into the vectors pIFM101A,B and C (see Figures 9 to 12). The library of resulting fusions to GFP-BCCP will be either in or out of frame. The in frame fusion proteins when expressed as correctly folded soluble proteins fluoresced green under ultraviolet light at 365 nm (GFP is a visual folding marker) and were also biotinylated. Figure 5 shows a colony western blot probed with streptavidin-horseradish peroxidase conjugate. The positive hits (significantly above the background) are the ones that were marked as green when visualized 365 nm. Only 4 out of 36 were biotinylated but not green visually. This could be due to the fact that the detection method used for biotinylation of BCCP is much more sensitive than visual detection of green fluorescence.

In this experiment many of the fusion proteins would be in-frame to GFP-BCCP but would not fluoresce green as they do not fold properly and are insoluble. The streptavidin-HRP western blot data with a set of complex fusion proteins (figure 5) shows that only when the fusion proteins are correctly folded and soluble, as assessed by green fluorescence of GFP, is the BCCP domain of the fusion protein biotinylated. These observations demonstrate that biotinylation of BCCP in the fusion protein is a folding marker as is the green fluorescence of GFP. Since it is known in the art that GFP is a reliable indicator of correct folding then the results

here demonstrate that biotinylation of BCCP is also a reliable indicator of correct folding.

Example 2: Use of BCCP as a Protein Solubility Enhancer

Materials and Methods

Vectors. The pQE82L-GFP-biotin and pMD004 plasmids (Figure 8) were constructed by standard techniques (T. Maniatis et al (1989) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Press) and both consist of a pQE82L vector (Qiagen) backbone, with a RGS-His tag followed by either the "Avi-Tag" sequence or BCCP protein domain respectively, followed by a multi-cloning site. They encode the $lacI^q$ repressor for tight regulation of the T5 promoter, and when cut with *Sma*I and *Not*I release either the GFP or p53 stuffer fragments to give the vectors ready for gene cloning inserts with a 5'-phosphorylated, blunt end and a 3' – *Not*I sticky end.

Gene Insert Production. Human protein domains were chosen and the corresponding genes were PCR amplified from cDNA libraries. The 5'-phosphorylated forward primers consist of the first 24 bp at the beginning of the relevant sequence, starting with a full codon. Some of the forward primers are longer to incorporate a G or C at the 3' end. The reverse primers consist of the last 24 bp of the relevant sequence (longer if necessary to incorporate a G or C at the 5' end) which is then appended to the beginning of the reverse primer template (TGATAGAAGAGCGGCCGC). The final reverse primer would be the reverse complement of this. This primer results in the stop codon of all the fusions being defined and followed by a *Not*I site for cloning into the N-terminal tagging vector described above. Two cDNA templates were combined at a final concentration of 10ng/μl. These were a) human heart cDNA plasmid library (Life Technologies) & b) HeLa cell cDNA plasmid library (Invitrogen). All primers were reconstituted in distilled water to 100pmols/μl. A master mix was prepared (without primers) from: Template (10ng), PWO polymerase buffer with magnesium sulphate (1x final concentration), dNTPs (5mM final conc.), PWO polymerase (2.5 units), dimethyl

sulfoxide (10% final conc.) and distilled water to a final volime of 48 µl per reaction. The master mix was aliquoted into 96 well PCR plates (Eppendorf) and 1 µl of each primer added on ice. Conditions were as follows: 94 for 3 mins then 94 for 30 secs, 59 for 30 secs, 72 for 2 mins (32 cycles) and finally 72 for 7 mins. Products were checked on 2% agarose gels/TBE and purified using Qiaquick PCR purification columns (Qiagen). Clean dsDNA was digested with *Not*1 in a standard digestion mixture and cleaned again.

Hoescht 33258 assay. To quantify the dsDNA in preparation for cloning a low range standard curve of an unrelated, clean PCR product in 1:1000 Hoescht dye (stock 1mg/ml)/1xTNE (Tris 10mM, EDTA 1mM, NaCl 0.2 M pH 7.4) was set up at 80, 40, 20, 10, 5, 2.5, 1.25, 0 ng/100 µl. 1 µl of each experimental PCR product was added to 99 µl of 1:1000 Hoescht/TNE, mixed in clear bottomed, black sided 96 well microtiter plates (Corning) and fluorescence read at 365/465nm. The standard curve was plotted and dsDNA content of each 'insert preparation' calculated as ng/µl

Cloning the inserts into pQE82L-GFP-biotin or pMD004. Inserts were ligated to the vector prep with an approximate molar ratio of 3:1 (insert:vector). Ligations were carried out in a 96-well PCR plate with the rapid DNA ligation kit (Roche). The ligations (2 µl of each) were used to transform 30 µl of XL1-Gold Supercompetent cells (Stratagene), according to the protocol, in a thin wall 96-well PCR plate. After heat shock, the transformations were added to 300 µl of pre-warmed SOC medium in a 96-well deep well block and shaken at 37°C for 45 minutes. 200 µl of each was plated and incubated at 37°C overnight. Approximately 0.02 pmoles of vector was used for each ligation. Ampicillin resistant clones were analysed by colony PCR to check for correct insert size and positive clones taken forward for expression screening.

Protein Expression. Single ampicillin resistant colonies were used to inoculate 1 ml of LB media containing 100 µg/ ml ampicillin (LB-Amp) and grown over-night at 37°C with shaking. The next day a 1:100 dilution was made into fresh LB-Amp

and cells grown at 37°C until OD600 = 0.6 to 1.0. IPTG was then added to a final concentration of 1 mM and growth continued at 30°C for 4 hours. 10 µl of cell culture was then taken and analysed by 4 – 20% SDS-PAGE Western blot as described and probed with HRP-conjugated streptavidin.

Results and Discussion

To prove that the BCCP domain can aid protein folding, a defined set of 49 human proteins were cloned into the *Sma* I / *Not* I sites of two different vectors: pQE82L-GFP-biotin or pMD004 (Figure 8). Protein expression from these constructs resulted in proteins being expressed with either a short (19 aa) N-terminal peptide tag (consisting of a hexa-histidine sequence followed by the "Avi-Tag" sequence (www.avidity.com; US Patent 5,932,433) for pQE82L-GFP-biotin or as fusions to the C-terminus of the *E. coli* BCCP protein (pMD004). A significantly higher success rate for the production of soluble protein was observed when the proteins were expressed as fusions with the BCCP protein (see Figures 6 and 7), as summarized in Table 1. For example when fused to the BCCP domain 98 % of proteins were expressed solubly compared with when expressed in the absence of the BCCP domain only 48 % of clones gave observable expression of which 81 % were soluble. The observation that a greater overall number of clones expressed from the pMD004 vector compared with the expression from the pQE82L-GFP-biotin in unlikely to be explained by the "N-end rule" where the amino acids at the N-terminus can be crucial in determining targeting to the proteosome for degradation (Rao H, Uhlmann F, Nasmyth K, Varshavsky A. (2001) *Nature,* **410,** 955-9), since in both constructs the N-terminal 12 amino acids are identical. More likely an explanation is that the constructs expressed with an-N-terminal BCCP domain aid protein folding of the downstream proteins, preventing the targeting of the mis-folded proteins to the proteosome. This is also supported by the observation that more proteins expressed in a soluble manner when expressed downstream of BCCP compared with expression from the pQE82L-GFP-biotin vector. The mechanism by which BCCP aids the folding of down-stream protein

domains could be either by recruitment of chaperones or by increasing the overall solubility of the fusion protein.

The results presented here strongly indicate that the BCCP domain can increase the overall number of clones expressing soluble protein when expressed as an N-terminal fusion to the target protein. In addition the result indicate that the BCCP domain can increase the solubility of a protein of interest. The tight correlation observed between biotinylation and solubility of expressed fusions demonstrates that biotinylation of BCCP acts as a folding marker when fused to the N-terminus of a protein of interest. In addition, the ability of the BCCP protein to be biotinylated provides a highly specific means to capture the protein on a streptavidin surface.

Table 1. Protein Expression Summary. Proteins were chosen and corresponding gene inserts were cloned into the pQE-GFP-biotin (vector 1) or the BCCP pMD004 (vector 2) resulting in fusions to the C-terminus of either a hexa-histidine-Avi-Tag peptide or a hexa-histidine-BCCP protein. Only inserts cloned into both vectors are compared in terms of protein expression. Key to table: [1]Internal coding number. [2]Protein database accession number (www.oca.ebi.ac.uk). [3]DNA gene length in base-pairs. [4]. Protein size when expressed as a fusion with BCCP in amino acids (aa). [5]. Protein size when expressed as a fusion with BCCP in kilodalton (kda). [6] Region of ORF cloned (aa). C – cloned but no expression; H – expressing hexa-histidine positive protein in a SDS-PAGE Western blot; B – expressing biotin positive protein in a SDS-PAGE Western blot; S – expressing soluble protein.

Table 1.

Gene	B #[1]	PDB[2]	Insert Length bp[3]	Fusion aa[4]	Fusion Kda[5]	Part Cloned[6]	Expression Vector 1	Expression Vector 2
Ac.Fib. Gr. Factor	1	2AXM	408	241	31.3	1-136/136 orf	C.H.B.S.	C.H.B.S.
Alc. Dehyd.	2	1DEH	1143	486	63.2	1-370/374 orf	C.	C.H.B.S.
Ad. Kinase	3	1BX4	1044	453	58.9	22-362/362 orf	C.H.B.S.	C.H.B.S.
Ald. Red	4	1AZ1	960	425	55.3	2-315/315 orf	C.H.B.S.	C.H.B.S.
Bar-to-Autoint.	5	2EZZ	285	200	26.0	1-89/89 orf	C.H.B.S.	C.H.B.S.
Bleo. Hyd.	6	1CB5	1380	565	73.5	1-454/455 orf	C.H.B.S.	C.H.B.S.
Bone Morph. P2	7	3BMP	19B	171	22.2	291-396/396orf	C.H.B.S.	C.H.B.S.
Carb. Anhyd. II	9	1A42	798	371	48.2	371 / 371 orf	C.	C.H.B.S.
Cyclin-dep Kin 2	11	1F5Q	912	409	53.2	1-298/298orf	C.H.B.S.	C.H.B.S.
C-Raf1	12	1GUA	246	187	24.3	56-131/648orf	C.H.	C.H.B.S.
3-Meth. DNA Glyc.	14	1BNK	663	326	42.4	80-294/298orf	C.	C.H.B.S.
DNA Pase β	15	1BPX	1010	442	57.4	4-334/334orf	C.H.	C.H.B.S.
Gr. F. Rec-bid. P2	17	1CJ1	306	207	26.9	57-152/217orf	C.H.B.S.	C.H.B.S.
Hck Kinase	19	3HCK	336	217	28.2	140-245/526orf	C.H.B.S.	C.H.B.S.
C-Jun Proto-Onc	20	1FOSJ	189	168	21.8	255-322/340orf	C.H.B.S.	C.H.B.S.
Urac.-DNA Glyc.	21	4SKN	678	331	43.0	85-304/304orf	C.	C.
Quin. Red.	22	2QR2	711	342	44.5	1-230/230orf	C.	C.H.B.S.
GSTP1	23	9GSS	652	322	41.9	1-209/209orf	C.	C.H.B.
Orn. Aminotr.	25	2CAN	1224	513	66.7	238-439/439orf	C.H.B.S.	C.H.B.S.
Angiogenin	26	1AWZ	369	228	29.6	25-147/147orf	C.	C.H.B.S.
Prot. Disulf. Isom.	28	1MEK	378	231	30.0	18-137/508orf	C.	C.H.B.S.
Glyc-Inh. Factor	29	1GIF	363	226	29.4	1-114/114orf	C.H.B.S.	C.H.B.S.
Fk506-Bind. Prot	30	1NSG	325	213	27.7	1-107/107orf	C.	C.H.B.S.
Annexin I	34	1BO9	237	184	23.9	40-112/345orf	C.H.	C.H.B.S.
Cyclophillin A	36	1BCK	495	270	35.1	1-164/164orf	C.H.B.S.	C.H.B.S
Ser.-Thr. Phos. B-B	41	1AUIB	507	274	35.6	2-170/170orf	C.	C.H.B.S.
Transcr. Factor IIb	42	1TFB	633	316	41.1	112-316/316orf	C.	C.H.B.S.
S-Admeth. Decarb.	47	1JEN	800	372	48.3	69-329/334orf	C.	C.H.B.S.

Procathepsin B	49	3PBH	948	421	54.7	19-333/339orf	C.	C.H.B.S.
Rhoa	51	1CXZ	561	202	38.0	1-181/183orf	C.	C.H.B.S.
Acid Phosphotase 1A	51A	P24666	471	257	28.0	1-157/157orf	C.	C.H.B.S.
Pax-6	53	6PAX	417	244	31.7	4-138/422orf	C.	C.H.B.S.
Phosfyr. Phostp	55	5PNT	492	269	35.0	1-157/157orf	C.H.B.S.	C.H.B.S.
Thyroid Hormone BP	57A	Q14894	942	314	45	1-314/314orf	C.	C.H.B.S.
Hsp86	58A	-	684	333	43.3	8-235/731orf	C.H.B.S.	C.H.B.S.
Hsp40	59	1HDJ	231	182	23.7	1-76/340orf	C.	C.H.B.S.
NKxB52	61	1A3Q	891	402	52.3	37-327/898orf	C.H.B.S.	C.H.B.S.
Fruc.-Biaph. Ald.*	64	1DOS	1055	470	61.1	1-358/358orf	C.H.B.S.	C.H.B.S.
Fadd	65	1E3Y	312	209	27.2	93-192/208orf	C.	C.H.B.S.
Transcr. Factor Max	66	1HLO	285	200	25.0	4-92/168orf	C.	C.H.B.S.
IL-6	67	2IL6	515	276.7	36.0	47-212/212orf	C.H.	C.H.B.S.
Hyp.-Guan. Phtrbr.	71	1NSF	650	325	42.3	4-217/217orf	C.	C.H.B.S.
Glyoxylase II	78	1QH5	198	371	48.2	1-260/260orf	C.	C.H.B.
Smtp-1a	80	1AM9	268	191	24.8	319-398/1147orf	C.	C.H.B.S.

Throughout this specification and the claims which follow, unless the context requires otherwise, the word "comprise", and variations such as "comprises" and "comprising", will be understood to imply the inclusion of a stated integer or step or group of integers or steps but not the exclusion of any integer or step or group of integers or steps.

The reference in this specification to any prior publication (or information derived from it), or to any matter which is known, is not, and should not be taken as acknowledgement or admission or any form of suggestion that that prior to publication (or information derived from it) or known matter forms part of the common general knowledge in the field of endeavour to which this specification relates.

2003238441 11 Sep 2008

THE CLAIMS DEFINING THE INVENTION ARE AS FOLLOWS:

1. Use of a tag moiety comprising a biotinylation domain for increasing the solubility of a protein of interest by attachment of said tag moiety to the N-terminus or C-terminus of said protein of interest.

2. Use of a tag moiety comprising a biotinylation domain for determining the folded state of a protein of interest by attachment of said moiety to the N-terminus or C-terminus of said protein of interest.

3. Use of a nucleic acid molecule encoding a tag moiety comprising a biotinylation domain for increasing the proportion of clones in a library that express the protein of interest encoded by each of said clones at detectable levels by attachment of said nucleic acid molecule encoding said tag 5' to and in-frame with the gene encoding said protein of interest in each of said clones.

4. A method of increasing the solubility of a protein of interest when expressed in a host cell characterised by:

a) attaching a first nucleic acid molecule encoding a tag moiety comprising a biotinylation domain to a second nucleic acid molecule encoding said protein of interest to form a construct such that the tag moiety in the expressed product of the combined first and second nucleic acid molecules is located at the N-terminus or C-terminus of said protein of interest; and

b) expressing said construct in a host cell

5. A method of determining the folded state of a protein of interest characterised by:

a) attaching a first nucleic acid molecule encoding a tag moiety comprising a biotinylation domain to a second nucleic acid molecule encoding said protein of interest to form a construct such that the tag moiety in the expressed product of the combined first and second nucleic acid molecules is located at the N-terminus or C-terminus of said protein of interest; and

b) expressing said construct in a host cell under conditions such that only a correctly folded biotinylation domain present in said tag moiety is ligated with biotin; and

2003238441 11 Sep 2008

c) determining the folded state of the protein of interest comprising said tag moiety by the presence or absence of a biotin group in the protein expressed from said construct

6. A method of increasing the proportion of clones in a library that express the protein of interest encoded by each of said clones in ahost cell at detectable levels, characterised by:

a) attaching a first nucleic acid molecule encoding a tag moiety comprising a biotinylation domain to a second nucleic acid molecule encoding said protein of interest to form a construct such that the tag moiety in the expressed product of the combined first and second nucleic acid molecules is located at the N-terminus or C-terminus of said protein of interest; and

b) expressing said construct in a host cell

7. The use as claimed in claim 1 or claim 2, characterised in that said protein of interest is encoded by a nucleic acid molecule which forms part of a library comprising two or more different coding sequences.

8. The use as claimed in claim 7, characterised in that said different coding sequences are modified to contain said tag moiety and expressed in parallel

9. A method a claimed in claim 4 or claim 5, characterised in that said protein of interest is encoded by a nucleic acid molecule which forms part of a library comprising two or more different coding sequences.

10. A method as claimed in claim 9, characterised in that said different coding sequences are modified to contain said moiety and expressed in parallel.

11. A library of folded proteins in which each protein has an N- or C-terminal tag moiety comprising a biotinylation domain that is enzymatically biotinylated.

12. A library as claimed in claim 11, wherein said proteins are immobilised on a solid substrate by means of said biotinylated domain to form an array.

2003238441 11 Sep 2008

13. A library as claimed in claim 12, characterised in that said solid substrate comprises a streptavidin surface.

14. A library as claimed in any of claims 11-13, a use as claimed in any one of claims 1-3,7, or 8, or a method as claimed in any one of claims 1-4, 9 or 10, characterised in that said biotinylation domain is Biotin Carboxyl Carrier Protein (BCCP).

15. A library, a use or method as claimed in claim 14, wherein said BCCP is *E.coli* BCCP.

16. A method of making a library of folded proteins as claimed in any one of claims 11-15, characterised by:

a) providing a library of two or more different coding sequences, each coding sequence comprising a respective second nucleic acid molecule encoding a protein of interest;

b) modifying said different coding sequences by attaching a first nucleic acid molecule, which first nucleic acid molecule encodes a tag moiety comprising a biotinylation domain 5'- or 3'- to and in-frame with the second nucleic acid molecule of each member of said library to form a construct such that the expressed products of the combined first and second nucleic acid molecules each comprise said tag moiety located at the N- or C-terminus of each protein of interest; and

c) expressing said construct in parallel in a host cell under conditions such that the biotinylation domain present in the tag moiety of each expressed construct is ligated *in vivo* with biotin when correctly folded.

17. A method as claimed in claim 16, further comprising lysing said host cell and simultaneously immobilising said expressed constructs from the cellular lysates on a solid support by means of said biotinylated domains to form an array.

18. A method as claimed in claim 17, wherein said solid substrate comprises a streptavidin surface.

2003238441 11 Sep 2008

19. A library of nucleic acid molecules encoding a library of folded proteins as claimed in claims 11 to 15, characterised in that each coding sequence is modified to incorporate a tag moiety comprising a biotinylation domain at the N-terminus or C-terminus of the encoded protein.

20. A use as claimed in claim 1, claim 2 or claim 3, a method as claimed in claim 4, claim 5 or claim 6, a library as claimed in claim 11, substantially or herein before described with reference to the Examples and the accompanying drawings.

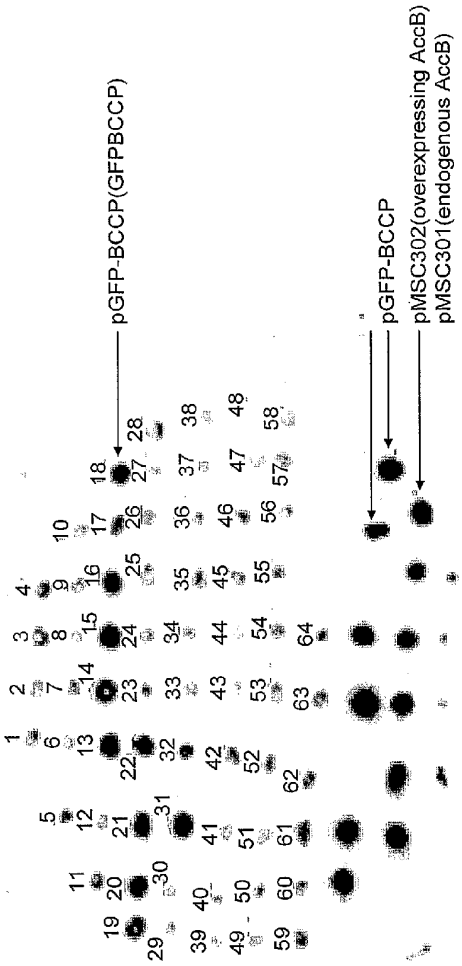

pGFP-BCCP(GFPBCCP)
pGFP-BCCP
pMSC302(overexpressing AccB)
pMSC301(endogenous AccB)

FIG. 1

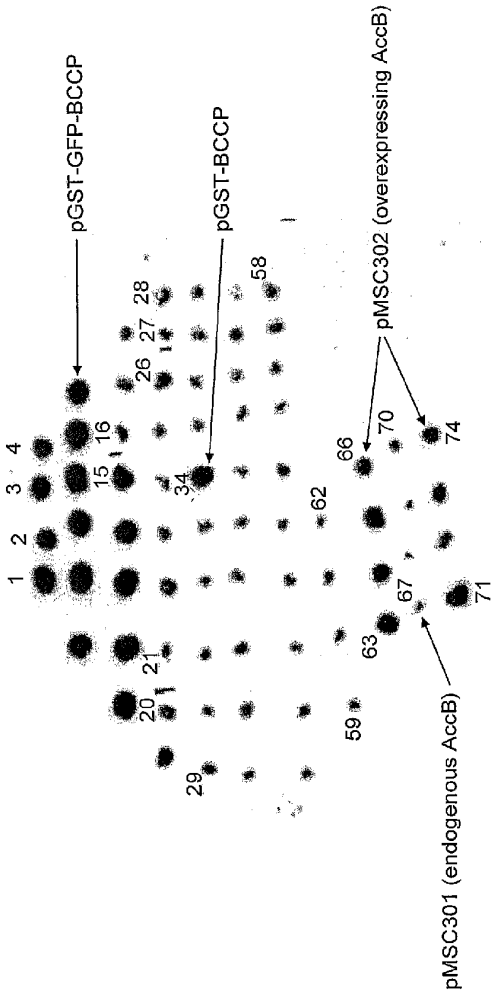
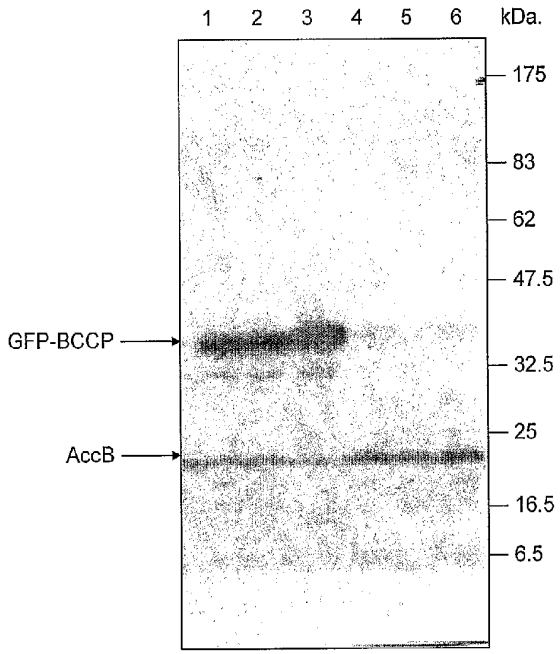

pGST-GFP-BCCP
pGST-BCCP
pMSC302 (overexpressing AccB)
pMSC301 (endogenous AccB)
1
2
3
4
15
16
20
21
26
27
28
29
34
58
59
62
63
66
67
70
71
74

FIG. 2



Probe: Strep.-HRP conjugate

FIG. 3

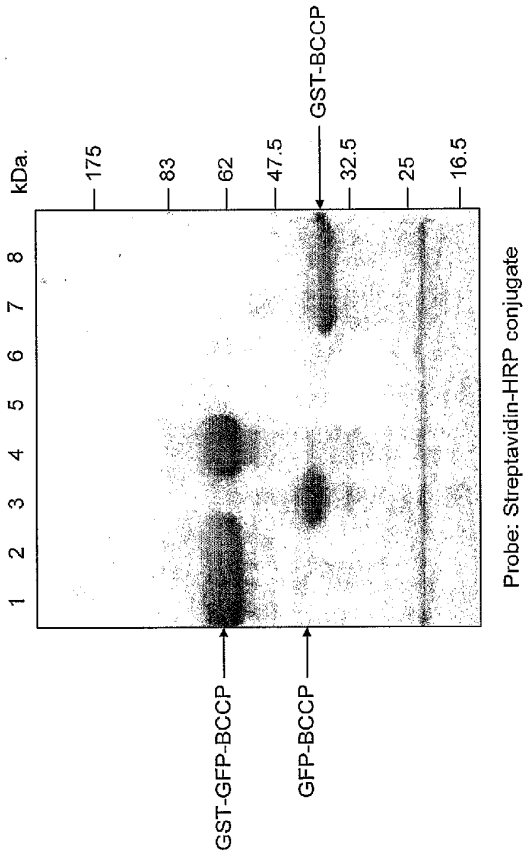

kDa.
175
83
62
47.5
32.5
25
16.5
GST-BCCP
1 2 3 4 5 6 7 8
GST-GFP-BCCP
GFP-BCCP
Probe: Streptavidin-HRP conjugate

FIG. 4

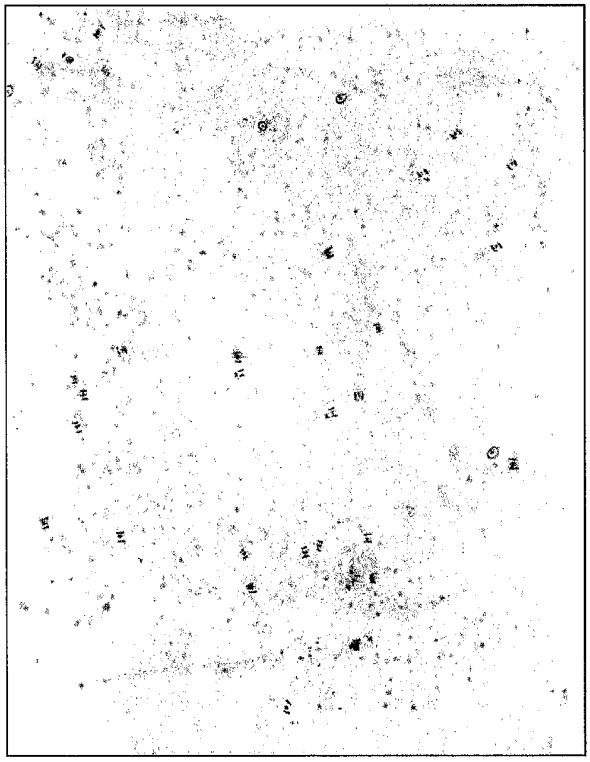
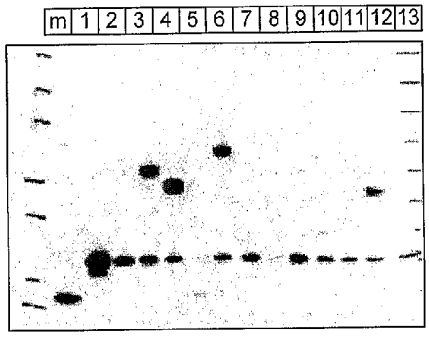
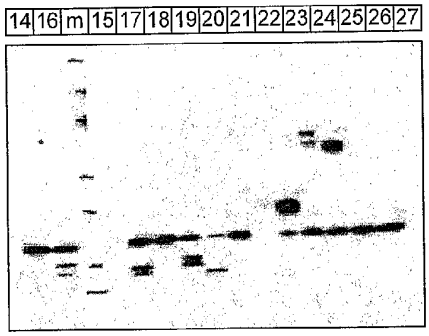


FIG. 5


m 1 2 3 4 5 6 7 8 9 10 11 12 13


14 16 m 15 17 18 19 20 21 22 23 24 25 26 27

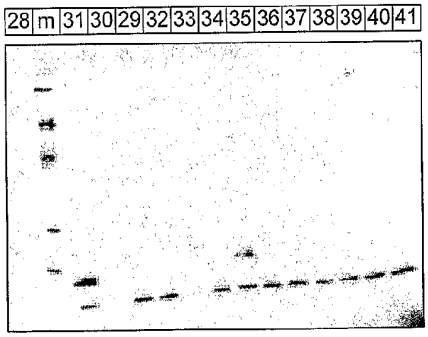

28 m 31 30 29 32 33 34 35 36 37 38 39 40 41

FIG. 6

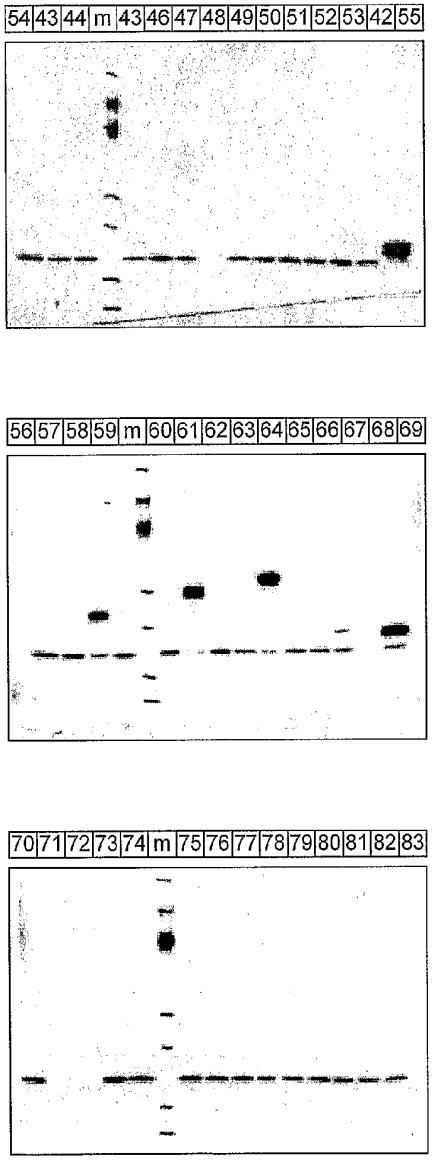

54 43 44 m 43 46 47 48 49 50 51 52 53 42 55
56 57 58 59 m 60 61 62 63 64 65 66 67 68 69
70 71 72 73 74 m 75 76 77 78 79 80 81 82 83

FIG. 6 CONT'D

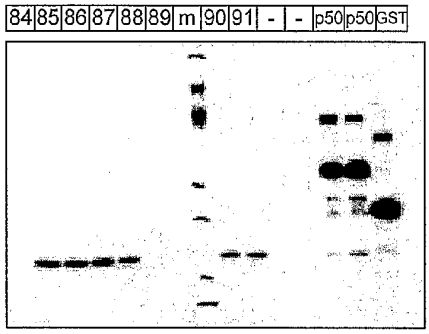


FIG. 6 CONT'D

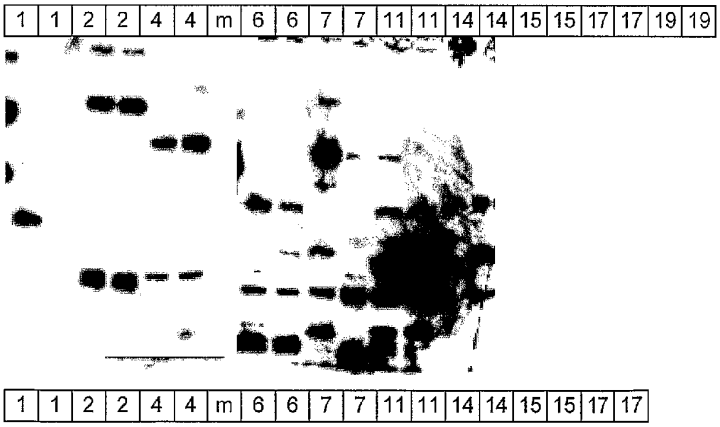
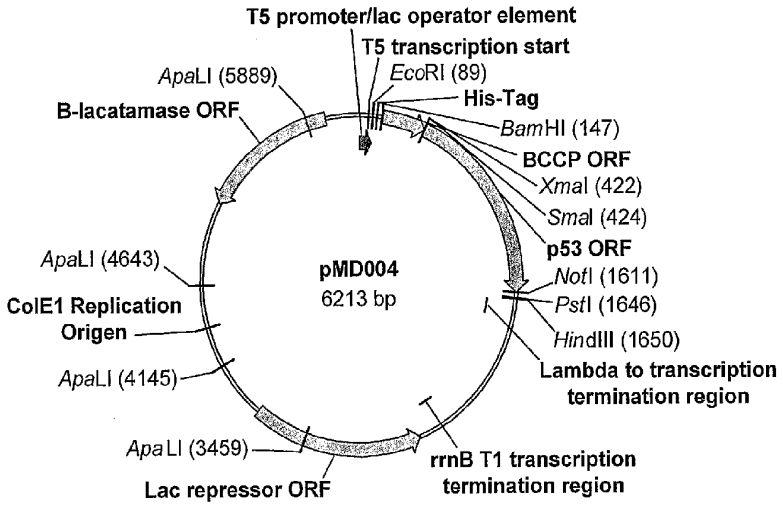


FIG. 7


T5 promoter/lac operator element
T5 transcription start
ApaLI (5889)
EcoRI (89)
B-lacatamase ORF
His-Tag
BamHI (147)
BCCP ORF
XmaI (422)
SmaI (424)
p53 ORF
ApaLI (4643)
pMD004
6213 bp
NotI (1611)
ColE1 Replication
Origen
PstI (1646)
HindIII (1650)
ApaLI (4145)
Lambda to transcription
termination region
ApaLI (3459)
rrnB T1 transcription
termination region
Lac repressor ORF

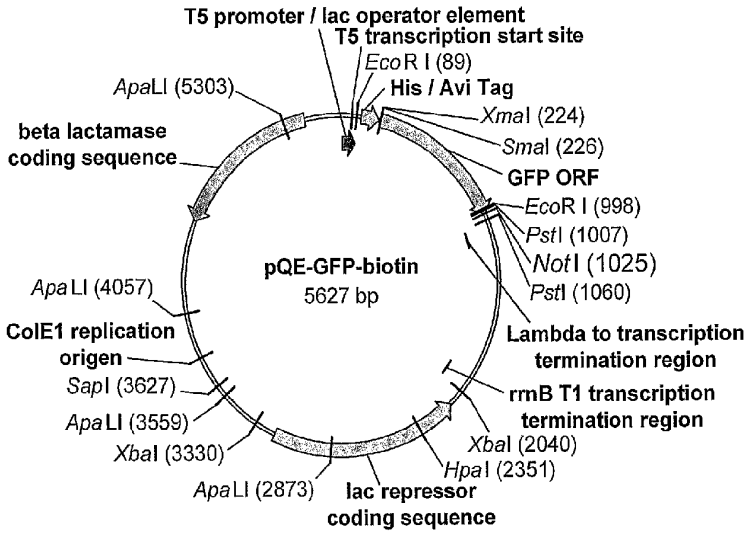

T5 promoter / lac operator element
T5 transcription start site
EcoR I (89)
ApaLI (5303)
His / Avi Tag
XmaI (224)
beta lactamase
coding sequence
SmaI (226)
GFP ORF
EcoR I (998)
PstI (1007)
pQE-GFP-biotin
5627 bp
NotI (1025)
ApaLI (4057)
PstI (1060)
ColE1 replication
origen
Lambda to transcription
termination region
SapI (3627)
rrnB T1 transcription
termination region
ApaLI (3559)
XbaI (2040)
XbaI (3330)
HpaI (2351)
ApaLI (2873)
lac repressor
coding sequence

FIG. 8

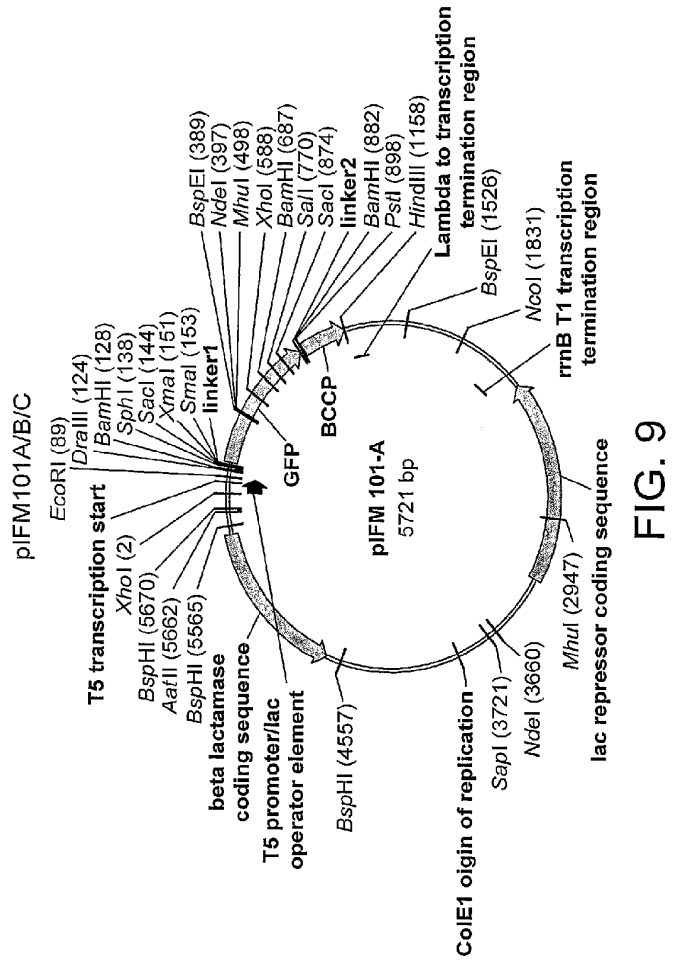

pIFM101A/B/C
EcoRI (89)
DraIII (124)
BamHI (128)
SphI (138)
SacI (144)
XmaI (151)
SmaI (153)
linker1
BspEI (389)
NdeI (397)
MhuI (498)
XhoI (588)
BamHI (687)
SalI (770)
SacI (874)
linker2
BamHI (882)
PstI (898)
HindIII (1158)
Lambda to transcription termination region
BspEI (1526)
NcoI (1831)
rrnB T1 transcription termination region
MhuI (2947)
lac repressor coding sequence
NdeI (3660)
SapI (3721)
ColE1 oigin of replication
BspHI (4557)
T5 promoter/lac operator element
beta lactamase coding sequence
BspHI (5565)
AatII (5662)
BspHI (5670)
XhoI (2)
T5 transcription start
GFP
BCCP
pIFM 101-A
5721 bp

FIG. 9

pIFM101A

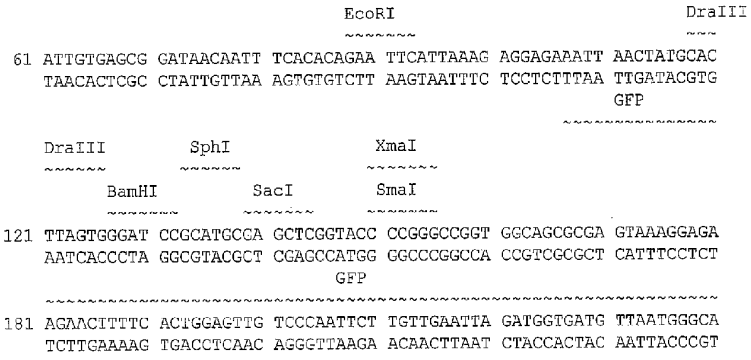


FIG. 10

pIFM101B

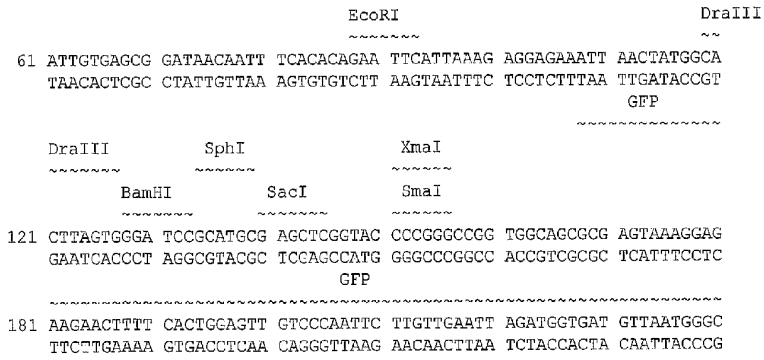


FIG. 11

pIFM101C

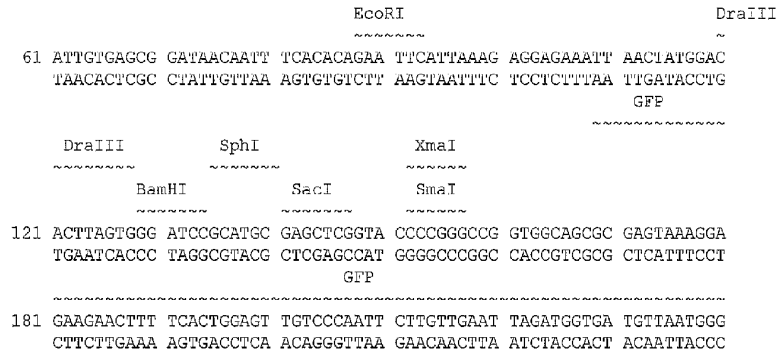


FIG. 12