



folding and function of proteins on arrays

Professor Jonathan Blackburn, Director, Centre for Proteomic & Genomic Research, University of Cape Town, South Africa

> *immobilizing proteins on arrays is quite simple – the challenge is in producing and arraying functional proteins that survive the rigors of immobilisation and retain their biological properties.*

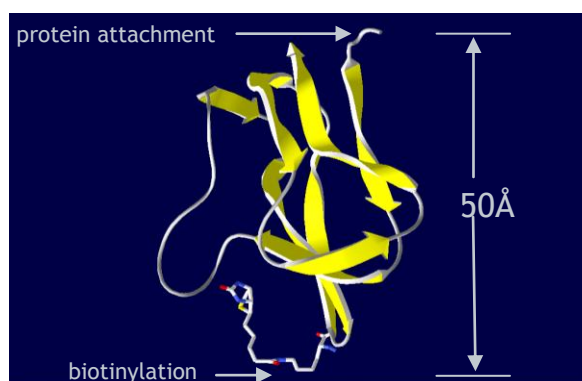
In developing protein function arrays, we have built a technology that allows the immobilisation of a wide variety of proteins whilst taking every step to ensure and maintain their functional state. The technology relies on the use of a proprietary affinity tag, and here we describe some of the important features underlying its development.

benefits of using a fusion tag

Immobilisation of proteins on to any surface is fraught with dangers, particularly if the mode of attachment of the protein to the surface is not tightly controlled. For example, non-specific, physical absorption to a surface through hydrophobic interactions has the potential to be significantly denaturing and to interfere with protein structure and function. At minimum, arrayed proteins will be presented in a heterogeneous manner, leading to a lower specific activity and therefore lower signal-to-noise ratios in assays (1).

By contrast, immobilisation of proteins *via* an affinity tag is a technique employed routinely during the course of affinity purification of proteins prior to functional analysis so there is ample precedent to suggest that, providing the affinity tag itself is tolerated, the process of immobilisation on to a surface *via* the tag does not generally affect protein folding or function. Sense Proteomic therefore attaches each arrayed protein to the surface exclusively *via* an affinity tag, specifically the biotin

carboxyl carrier protein (BCCP) domain of the *E. coli* acetyl CoA carboxylase.



BCCP as a fusion tag

E. coli BCCP is an 80 amino acid domain that folds autonomously into a compact, all beta-strand structure (Figure 1 (2))

Figure 1. Structure of BCCP showing the single biotinylated lysine residue and the N- and C- termini.





and is biotinylated post-translationally *in vivo* on a single, specific lysine residue by the host cell biotin ligase (the bir A product) (3). The *E. coli* BCCP domain is recognised and biotinylated both by the *E. coli* biotin ligase and also by the endogenous biotin ligases found in yeast, insect and mammalian cells (e.g. 4,5,6). Structural studies have shown that the N- and C-termini of BCCP are located in solvent exposed positions on the surface of the protein and are roughly 50Å distal to the biotinylated lysine residue (2). This suggests that it is possible to make in-frame fusions to either terminus of BCCP such that the resultant fusion proteins can become biotinylated *in vivo* on a single specific lysine in the BCCP domain, irrespective of the host cell in which they are expressed.

BCCP is biotinylated – but only when correctly folded

Biotinylation *in vivo* of BCCP does have one constraint; the BCCP domain must be folded into its native 3-dimensional structure in order to become biotinylated. Indeed short peptides derived from BCCP which contain that linear biotinylation epitope are not recognized as substrates by the biotin ligases (3). This property allows the biotinylation state of BCCP to be used as a folding marker.

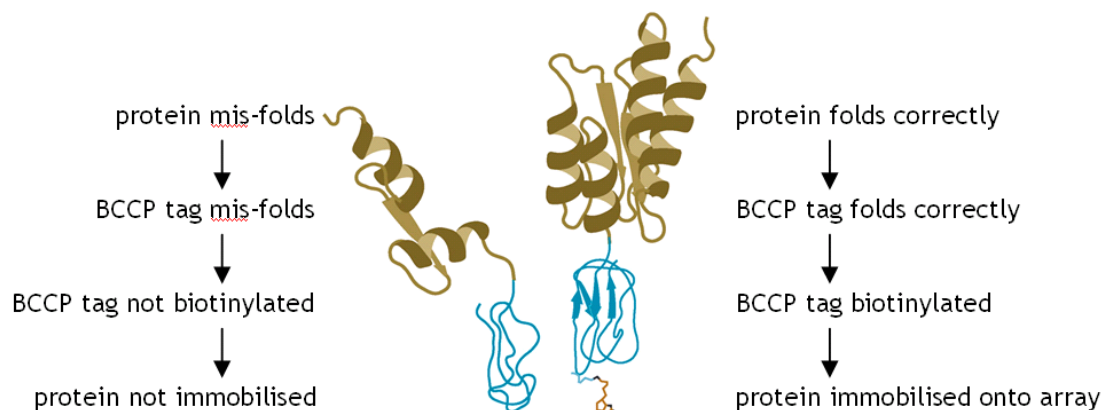
Most studies to date have focused on fusion proteins to BCCP where the fusion partner is an intact domain that is known to fold efficiently in the host cell. In such cases, it is easy to imagine that each domain in the fusion protein will fold autonomously, and that the resultant fusion protein will appear like beads on a string, allowing the BCCP domain to become biotinylated post-folding. However, when eukaryotic proteins are expressed in heterologous hosts, it is not automatic that they will fold correctly. If they do not, there is a strong likelihood that the aberrant folding of the fusion partner will affect the downstream biotinylation of BCCP, particularly if the BCCP domain lies at the C-terminal end of the fusion protein, as it does in Sense Proteomic's recombinant proteins.

> if the expressed protein mis-folds, the correct folding, and therefore biotinylation, of the BCCP tag may be prevented. By contrast if the protein folds correctly then folding of the BCCP tag should be unhindered and biotinylation will result. Consequently it is likely that only correctly folded proteins will be immobilized on to the array surface

The logic behind this expectation is as follows. During translation of large proteins, folding almost certainly begins in a co-translational manner - this is known for eukaryotic systems and is likely to be true in prokaryotic systems, although there are some differences here due to the chaperone-like roles played by heat shock proteins during initial protein synthesis. Co-translational folding of a fusion partner means that in the majority of cases, the N-terminal domain will start to fold first. If this initial folding event results in a properly folded domain, then it is reasonable to suppose that the C-terminal domain will also be able to fold correctly, unhindered by the presence of the N-terminal fusion. However, if the N-terminal domain mis-folds and starts to present aberrant hydrophobic surfaces, there is a strong likelihood that initiation of the normal folding pathway of the C-terminal domain will be affected, leading to aberrant folding of the C-terminal domain as well. If the C-



terminal domain is the BCCP domain, it would be expected that mis-folding of an N-terminal fusion partner will perturb the proper folding of BCCP and therefore prevent downstream biotinylation of BCCP.



In addition to this ‘mis-folding pathway’ mechanism, a second mechanism exists that might prevent biotinylation of a fusion protein in cases where mis-folding of an N-terminal domain occurs but does not affect folding of a C-terminal BCCP domain. Biotinylation of BCCP is a post-translational event that requires BCCP to be physically accessible to the biotin ligase enzyme. However, mis-folded proteins are known to have a strong tendency to aggregate and if a BCCP fusion protein aggregates due to mis-folding of one domain, this will occlude the BCCP domain, thereby physically preventing biotinylation, irrespective of the folded state of the BCCP domain itself.

BCCP acts a marker for correctly folded fusion partners

- > *these theoretical considerations suggest that the ability of a BCCP fusion protein to become biotinylated depends on the correct folding of the fusion partner – in other words, BCCP acts as a useful folding marker. This expectation is borne out by experiment.*

A number of other folding markers are known, of which the most widely used is the green fluorescent protein (GFP) (7). The fluorophore of GFP is formed as an autocatalytic post-translational modification but, as with biotinylation of BCCP, this only occurs if GFP folds correctly and does not aggregate (7). There are many examples in the literature showing that when fusions are made to GFP, only correctly folded, non-aggregated fusion proteins become fluorescent (e.g. 8,9,10). To test the utility of BCCP as a marker of the folded state of upstream fusion partners, we made a series of N-terminal fusions of full-length or truncated glutathione-S-transferase (GST) to both GFP and BCCP,

some of which were expected to fold correctly and some which were not. We initially used colony blots to compare for each construct whether the fusion protein expressed in *E. coli* was fluorescent and whether it was biotinylated. We found a 100% correlation between fluorescence and biotinylation across all constructs in this experiment (data not shown). Western blot analysis of biotinylated clones confirmed that both fluorescence and biotinylation were due to the full length fusion protein, not proteolytic fragments (Figure 2).

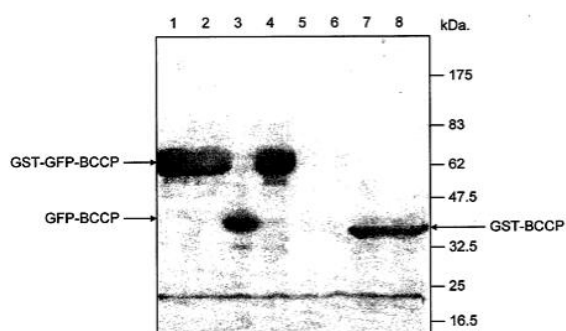


Figure 2. Western blot analysis of GST-GFP-BCCP fusion proteins confirms biotinylation of the full-length proteins.

Lanes 1, 2 and 4 are fluorescent GST-FP-BCCP fusions. Lane 4 is a GFP-BCCP fusion control. Lanes 5 and 6 are negative controls. Lanes 7 and 8 are GST-BCCP fusion controls. The blot was developed using a streptavidin-HRP conjugate. In all lanes, the 18kDa signal from endogenous *E. coli* acetyl CoA carboxylase is just visible. Although detected here under denaturing conditions, the biotin is buried in the native conformation and not available for attachment during spotting of cell lysates onto arrays.

In a second experiment, we took a human cardiac cDNA library and, in a single reaction tube, generated nested deletions from the 3'-end of each clone using Exonuclease III under controlled conditions. We cloned the resultant library of nested deletions into an *E. coli* vector containing a 3'-GFP-BCCP tag and, after inducing expression of the recombinant proteins, we measured the fluorescence and biotinylation of each colony. We observed that, as expected due to the nature of the nested deletion library, only a small subset of colonies were fluorescent but amongst these, we observed a 100% correlation between fluorescence and biotinylation (data not shown).

In a third experiment, we individually cloned a diverse set of 45 human cytosolic proteins (Table 1) as fusions to BCCP-His₆. We expressed each construct in *E. coli* and characterized whether each resultant fusion protein was soluble and whether it was biotinylated. We chose to monitor the solubility of each fusion protein because this is a generally accepted surrogate measure of correct folding in *E. coli* (9,10). By contrast, mis-folded proteins typically aggregate in *E. coli*. We observed that 44 out of 45 clones gave measurable expression, as judged by Western blots on whole cell extracts using an anti-His-tag antibody. Of these 44 clones, 42 expressed a biotinylated protein, as judged by Western blots on whole cell extracts using a streptavidin-HRP conjugate and in each of these 42 cases, we found the biotinylated protein exclusively in the soluble fraction (Table 1). In the two cases where a recombinant protein was expressed but not biotinylated, the recombinant protein could not be detected in the soluble fraction, only in the cell debris, consistent with the prediction that only a properly folded fusion protein would become biotinylated on the BCCP tag.

Human protein	SwissProt identifier	Size of fusion protein (inc. BCCP domain) kDa	Fragment of open reading frame cloned (residue numbers)	Protein expression result
Acidic Fibroblast Growth Factor	P05230	26.7	16-155/155	H.B.S.
Alcohol dehydrogenase 1B	P00325	50.2	1-370/374	H.B.S.
Adenosine Kinase	P55263	49.1	22-362/362	H.B.S.
Aldose Reductase	P15121	46.5	2-315/315	H.B.S.
Barrier-To-Autointegration Factor	O60558	21.0	1-89/89	H.B.S.
Bleomycin Hydrolase	Q13867	63.4	1-454/455	H.B.S.
Bone Morphogenetic Protein 2	P12643	22.9	291-396/396	H.B.S.
Carbonic Anhydrase II	P00918	40.0	1-259/259	H.B.S.
Cyclin Dependent Kinase 2	P24941	44.8	1-298/298	H.B.S.
RAF proto-oncogene serine/threonine-protein kinase	P04049	19.5	56-131/648	H.B.S.
3-Methyladenine DNA Glycosylase	P29372	34.6	80-294/298	H.B.S.
DNA Polymerase Beta	P06746	48.6	4-334/334	H.B.S.
Growth Factor Receptor-Bound Protein 2	P62993	22.1	57-152/217	H.B.S.
Tyrosine-protein kinase HCK (SH2 domain)	P08631	22.9	140-245/525	H.B.S.
C-Jun Proto-Oncogene	P05412	18.3	255-322/331	H.B.S.
Uracil-DNA Glycosylase	P13051	35.3	94-313/313	C.
Quinone Reductase Type 2	P16083	36.7	1-230/230	H.B.S.
Glutathione S-Transferase P	P09211	34.2	1-209/209	H.
Ornithine Aminotransferase	P04181	33.3	238-439/439	H.B.S.
Angiogenin precursor (Ribonuclease 5)	P03950	24.8	25-147/147	H.B.S.
Prolyl 4-Hydroxylase Beta Subunit (Thioredoxin domain 1)	P07237	24.4	18-137/508	H.B.S.
Macrophage migration inhibitory factor	P14174	23.3	1-114/114	H.B.S.
Fk506-Binding Protein 1A	P62942	22.7	1-107/107	H.B.S.
Annexin A1 (domain 1)	P04083	19.1	40-112/345	H.B.S.
Peptidyl-prolyl cis-trans isomerase A (Cyclophilin A)	P62937	28.8	1-164/164	H.B.S.
Serinethreonine Phosphatase 2b (Calcineurin subunit B)	P63099	30.0	2-169/169	H.B.S.
Transcription initiation factor IIB	Q00403	33.5	112-316/316	H.B.S.
S-adenosylmethionine decarboxylase proenzyme	P17707	40.9	69-329/334	H.B.S.
Cathepsin B precursor	P07858	46.1	19-333/339	H.B.S.
Transforming protein RhoA	P61586	31.3	1-181/193	H.B.S.
LMW phosphotyrosine protein phosphatase	P24666	28.8	1-157/157	H.B.S.
Paired box protein Pax-6 (paired domain)	P26367	25.6	4-136/422	H.B.S.
NADP-regulated thyroid-hormone binding protein	Q14894	44.7	1-314/314	H.B.S.
Heat shock protein 90-alpha (N-Terminal domain)	P07900	37.3	8-235/731	H.B.S.
Heat shock protein 40 (N-Terminal domain)	P25685	19.4	1-76/339	H.B.S.
Nuclear factor NF-kappa-B p52 Rel homology domain	Q00653	42.2	37-327/343	H.B.S.
Fructose-bisphosphate aldolase class II	P05062	50.2	1-363/363	H.B.S.
FAS-associating death domain (FADD)	Q13158	22.1	93-192/208	H.B.S.
Myc-associated factor X (basic helix-loop-helix domain)	P61244	21.1	4-92/160	H.B.S.
Interleukin-6 (IL-6)	P05231	29.5	47-212/212	H.B.S.
Hypoxanthine-guanine phosphoribosyltransferase	P00492	35.0	4-217/217	H.B.S.
Hydroxyacylglutathione hydrolase (Glyoxalase II)	Q16775	39.8	1-260/260	H.
Sterol regulatory element-binding protein 1 (Dna Binding Domain)	P36956	20.2	319-398/1147	H.B.S.
Nuclear factor NF-kappa-B p50 Rel homology domain	P19838	45.8	43-353/367	H.B.S.
Nuclear factor of activated T-cells (NFAT) (DNA binding domain)	Q13469	42.7	399-678/925	H.B.S.

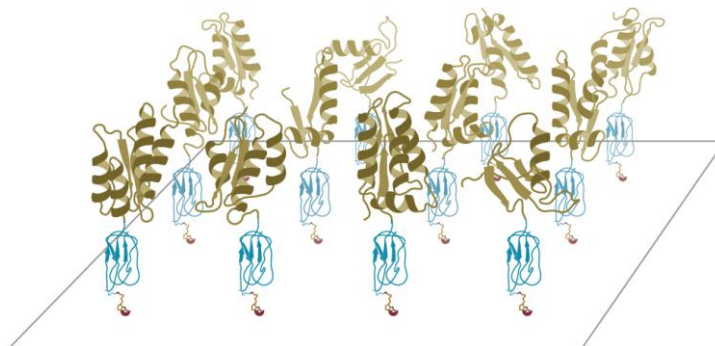
Table 1. A prototype set of human proteins expressed in E. coli as fusions to BCCP.

The protein fragment actually cloned as well as the full length ORF are given for each clone. The results of the protein expression studies are described as follows: C - cloned but no expression; H - expression confirmed by Western blot analysis of whole cell extracts using an anti-His tag antibody; B - biotinylation of expressed protein confirmed by Western blot analysis of whole cell extracts using a streptavidin-HRP conjugate; S - expressed, biotinylated protein is folded, as judged by Western blot analysis of soluble protein fractionations using a streptavidin-HRP conjugate.

Although exceptions might exist, the striking correlation we observed across the different libraries of fusion proteins described above is in direct support of the theoretical predictions, so it seems reasonable that our observation will generally hold true and that BCCP does act as a marker of the folded state of N-terminal fusion partners.

summary

Proteins are by nature diverse in size, structure and function. This means that immobilising them on a solid support such that they retain their structure and biological properties presents a challenge. Conventional methods have the disadvantage that the attachment is random, which means that not only may the protein be effectively denatured



by being pulled down onto the surface, but also the active part of the protein may be orientated so that it is not available for interaction. Tagging proteins with BCCP offers several advantages in array fabrication. It is biotinylated during natural processing, but only if it is correctly folded. This helps to ensure that only correctly folded, and therefore functional, proteins are immobilized on the array. Also BCCP presents the protein at a distance of some 50Å from the slide surface so assay reagents have good access to active sites on the immobilised proteins. Such benefits are essential to ensuring that data from Sense Proteomic protein function arrays are reliable and meaningful.

references

1. Cha, T., Gio, A. & Zhu, X-Y. (2005) Enzymatic activity on a chip: The critical role of protein orientation. *Proteomics* **5**, 416-419.
2. Athappilly, F.K. & Hendrickson, W.A. (1995). Structure of the biotinyl domain of acetyl-CoA carboxylase determined by MAD phasing. *Structure* **3**, 1407-1419.
3. Chapman-Smith, A. & Cronan, J.E. (1999). Molecular biology of biotin attachment to proteins. *J.Nutr.* **129**, 477S-484S.
4. Bertliner, E., Mahtani, H.K., Karki, S., Chu, L.F., Cronan, J.E., Gelles, J. (1994). Microtubule movement by a biotinylated kinesin bound to streptavidin-coated surface. *J. Biol. Chem.* **269**, 8610-8615.
5. Lerner, C.G. & Saiki, A.Y. (1996). Scintillation proximity assay for human DNA topoisomerase I using recombinant biotinyl-fusion protein produced in baculovirus-infected insect cells. *Anal. Biochem.* **240**, 185-196.
6. Parrott, M.B. & Barry, M.A. (2000). Metabolic biotinylation of recombinant proteins in mammalian cells and in mice. *Mol. Ther.* **1**, 96-104.
7. Waldo, G. S., Standish, B. M., Berendzen, J. & Terwilliger, T. C. (1999). Rapid protein-folding assay using green fluorescent protein. *Nature Biotechnology* **17**, 691-695.
8. Kawasaki, M. & Inagaki, F. (2001). Random PCR-based screening for soluble domains using GFP. *BBRC* **280**, 842-844.
9. Pedelacq, J.D. et al (2002). Engineering soluble protein for structural genomics. *Nature Biotechnology* **20**, 927-932.
10. Waldo, G.S. (2003). Genetic screens and directed evolution for protein solubility. *Curr. Opin. Chem. Biol.* **7**, 33-38.