

Biotination of Proteins *in Vivo*

A POST-TRANSLATIONAL MODIFICATION TO LABEL, PURIFY, AND STUDY PROTEINS*

(Received for publication, December 19, 1989)

John E. Cronan, Jr.

From the Departments of Microbiology and Biochemistry, University of Illinois, Urbana, Illinois 61801

Post-translational modification of proteins with biotin provides the means to specifically label proteins *in vivo* and to purify proteins from crude cell lysates. The carboxyl-terminal protein segments modified by reaction with biotin ligase are strongly conserved in nature. We have demonstrated that the proteins encoded by translational gene fusions of a number of heterologous proteins to these carboxyl-terminal sequences become biotinated *in vivo*. The minimum size of the protein segment needed to allow biotination of fusion proteins is 75 amino acids. This biotination sequence, although of bacterial origin, functions in *Saccharomyces cerevisiae* as well as in *Escherichia coli*. Fusion proteins are readily labeled with [³H]biotin *in vivo* and the labeling is highly specific due to the scarcity (<5) of biotininated protein species. Biotinated fusion proteins can be readily purified in native form by binding to columns of monomeric avidin followed by elution with buffers containing biotin. Alternatively, proteins can be purified in a denatured form in presence of 1% sodium dodecyl sulfate or 8 M urea. Thus, this technology allows purification by affinity chromatography of any protein to which a biotination sequence can be attached. The ability to specifically label a protein *in vivo* should have utility in studies such as intracellular protein trafficking and cytoskeletal dynamics.

The modification of proteins by covalent attachment of certain coenzymes is found throughout nature. Some of these prosthetic groups, biotin, lipoic acid, and 4'-phosphopantetheine, are found on very few of the protein species present in a given cell. For example, *Escherichia coli* contains only one biotininated protein (Fall, 1979), two lipoated proteins (Perham *et al.*, 1987), and one protein that carries 4'-phosphopantetheine (Rock and Cronan, 1981). The enzymes (ligases) that catalyze the covalent attachment of these molecules to protein thus recognize specific sequences of specific proteins.

Recombinant DNA construction of protein fusions in which the protein segment recognized by a coenzyme ligase was fused to a protein of interest could have distinct advantages over conventional protein fusions (e.g. fusion to β -galactosidase; Silhavy and Beckwith, 1985). Obvious advantages would be (i) the fusion proteins could be specifically labeled by growth of cell cultures in the presence of labeled coenzyme; (ii) if the coenzyme possessed specific binding toward an immobilized ligand, the fusion protein could be readily purified,

perhaps in a native form; and (iii) the protein segment fused to the protein of interest would be much smaller than those commonly used, thus giving less alteration of the chemical and biological properties of the protein of interest. Biotin has distinct advantages over the other coenzyme modification groups, the main advantage being the existence of two proteins, avidin and streptavidin, that strongly and specifically bind biotin (Green, 1975).

Biotin (Fig. 1) is a small coenzyme (vitamin H) synthesized by plants, most bacteria, and some fungi (8) which occurs primarily in a protein-bound state within the cell. Biotin is bound to specific proteins via an amide linkage between the biotin carboxyl group and a unique lysine amino group (Samols *et al.*, 1988), a reaction catalyzed by biotin ligase (Fig. 1). Biotinated proteins play enzymatic roles in several essential metabolic carboxylation and decarboxylation reactions such as the acetyl-CoA carboxylase reaction of fatty acid synthesis (Samols *et al.*, 1988). The function of biotin in these enzymes is to carry CO₂ in an activated form from one enzyme subunit to a second subunit and thus the biotin molecule is mobile and located at the protein surface.

Biotinated proteins are rare in nature. *E. coli* contains only one such protein, the biotin carboxyl carrier protein (BCCP)¹ of acetyl-CoA carboxylase, whereas other bacteria contain one to three biotininated proteins (Fall, 1979). Eucaryotic cells contain a slightly greater number of biotininated protein species, *Saccharomyces cerevisiae* contains four or five biotin proteins depending on growth conditions (Lim *et al.*, 1987), whereas both mammals (Robinson *et al.*, 1983; Chandler and Ballard, 1988) and plants (Nikolau *et al.*, 1985) are reported to contain four biotininated proteins.

The lysine residues of the proteins that become modified with biotin are located within an amino acid sequence usually found adjacent to the carboxyl terminus (Samols *et al.*, 1988) (Fig. 2). The biotininated lysine residue is generally located 34 or 35 residues from the carboxyl terminal amino acid. However, two biotininated proteins, the rat (Bai *et al.*, 1989) and avian (Takai *et al.*, 1988) acetyl-CoA carboxylases, have the coenzyme attached at a site far removed from the carboxyl terminus.

In the best studied of the conventionally biotininated proteins, the 1.3 S subunit of *Propionibacterium shermanii* transcarboxylase (Samols *et al.*, 1988), the biotin moiety is attached to a lysine residue located 34 residues from the carboxyl terminus and shortening of the carboxyl terminus by 2 residues results in a lack of biotination of the protein in *E. coli* (Murtif and Samols, 1987). The sequences surrounding the biotination sites of proteins from diverse biological sources are strongly conserved (Fig. 2). A notable (and probably important) conserved feature are regions upstream of the

* This project was supported by National Institutes of Health Grant AI 15650. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¹ The abbreviations used are: BCCP, biotin carboxyl carrier protein; SDS, sodium dodecyl sulfate.

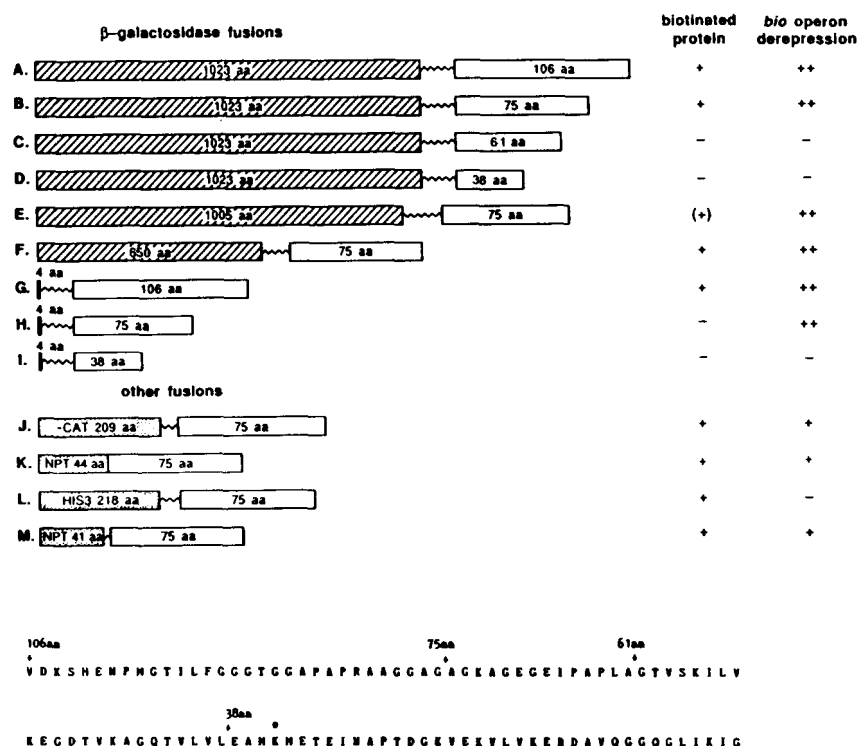


FIG. 3. Properties of the fusion proteins expressed from hybrid genes. All biotin acceptor sequences were derived from the *P. shermanii* 1.3 S gene (lower part of figure). A plus in the biotinated protein column indicates that a fusion protein of the expected size and abundance was detected. A plus in the derepression column indicates that transcription of the biotin operon was increased at least 10-fold in the presence of 41 nM biotin (the minimal concentration giving maximal repression in wild type cells), whereas two plus signs indicate that at least 10-fold derepression was observed at 5 μ M biotin. Fusions A–D encoded active β -galactosidase molecules, and thus derepression was monitored by galactokinase expression using the *bioB-galK* fusion of Ketner and Campbell (1974) rather than the *bioFC-lacZ* fusion used previously (Cronan, 1988). Fusion F produced a very weak by biotin-labeled protein band at the expected migration position. We believe this result is due to proteolytic clipping at the junction between the two protein segments. The junction of the DNA segments is the *EcoRI* site of *lacZ* a site also used in the λ gt11 system where proteolytic clipping at the fusion junction has been observed for a number of fusions (Young and Davis, 1983). At the bottom of the figure the amino acid sequences of the segments of the 1.3 S gene used in these constructions are given. The 106-, 75-, 61-, and 38-amino acid (aa) segments were derived by cutting the structural gene with the restriction enzymes *SalI*, *NarI*, *NaeI*, and *XhoI*, respectively. The *pta1.3t* plasmid carrying the 1.3 S gene was the kind gift of V. Murtif and D. Samols (1987). The polylinkers (sawtooth lines in figure) used to form the fusion junctions were derived the pUR (Rüther and Müller-Hill, 1983) and pUC (Viera and Messing, 1982) plasmids, and the junction sequences can be deduced from the descriptions under "Experimental Procedures." Fusion K contains no polylinker sequences. Fusions J and M are to Tn9 chloramphenicol acetyltransferase (*ScaI* site) and yeast HIS3 (*KpnI* site), respectively, whereas fusions K and M are to the neomycin phosphotransferases of Tn5 (*NarI* site and Tn903 (*ClaI* site), respectively.

biotinated lysine that contain sequences strongly predicted to form unusual secondary structures. The upstream sequences are usually rich in proline and alanine and often contain segments with two proline residues separated by a residue with a small side chain (Gly, Ala, Ser). In one protein, the *Klebsiella pneumoniae* oxalacetate decarboxylase α subunit, this motif is repeated such that over a 22-amino acid stretch, 21 amino acids are either proline or alanine (Schwarz *et al.*, 1988). The biotinated protein segments between the upstream regions and the carboxyl terminus are predicted to be largely α -helical using the standard secondary structural algorithms. Indeed, the profiles of the predicted structures of sequences from a tomato protein and proteins from bacteria, yeast, and humans essentially overlap when aligned at the biotinated lysine residue (data not shown). Thus, these carboxyl-terminal domains seem to possess sufficient common structure to account for their specific recognition by biotin ligase. Indeed, biotin ligases will biotinate the acceptor proteins from very different biological sources (e.g. bacteria *versus* higher eucaryotes) both *in vitro* (McAllister and Coon, 1966) and *in vivo* (see below).

EXPERIMENTAL PROCEDURES²

RESULTS AND DISCUSSION

Production of Biotinated Fusion Proteins—We have fused the DNA sequences encoding the biotin attachment sites of four different biotin proteins to the β -galactosidase structural gene in a manner designed to encode bipartite fusion proteins with amino-terminal β -galactosidase residues and the biotin-acceptor sequence located at the carboxyl terminus (Fig. 3). The four biotin protein segments were those derived from the *P. shermanii* transcarboxylase 1.3 S subunit, the *E. coli* acetyl-CoA carboxylase BCCP subunit, the *K. pneumoniae* oxalacetate α subunit, and a protein sequence encoded by a cDNA from tomato (only data for the first of these segments are given in this paper). The β -galactosidase portions consist of segments ranging from the entire coding sequence (thus en-

² Portions of this paper (including "Experimental Procedures" and Figs. 1, 2, 6, and 7) are presented in miniprint at the end of this paper. Miniprint is easily read with the aid of a standard magnifying glass. Full size photocopies are included in the microfilm edition of the Journal that is available from Waverly Press.

coding an active enzyme), the amino-terminal 65% of the protein (an inactive enzyme), or just the four amino-terminal amino acids. All of the fusion proteins encoded by these chimeric genes became biotinated in *E. coli* (Fig. 4) as long as the biotin ligase recognition sequence remained intact (see below). Indeed, we found that the fusion proteins often successfully competed for biotinylation with the endogenous biotin protein of the host, BCCP, since less [^{3} H]biotin was incorporated into BCCP when a fusion protein was produced (e.g. Fig. 3). It should be noted that Hoffman *et al.* (1987) also detected biotinylation of the tomato cDNA encoded sequence in *E. coli*.

Our most extensive analysis is of the *P. shermanii* 1.3 S subunit sequence where naturally occurring restriction sites enabled us to construct a series of β -galactosidase fusions with various lengths of the carboxyl-terminal 1.3 S segment (Fig. 3). A fusion carrying only the 1.3 S segment from the biotinylated lysine residue to the carboxyl terminus failed to be biotinylated (Fig. 3), indicating that sequences upstream of the biotinylated lysine are required for recognition of the protein by biotin ligase. It seems likely that a required sequence is the Pro-Ser-Pro tripeptide noted above, since protein fusions lacking this segment failed to be biotinylated whereas fusion proteins that included this segment incorporated biotin (Fig. 3). However, there may be more subtle structures in this region that are important for ligase recognition. Of the segments tested, the minimum 1.3 S segment allowing biotinylation of fusion proteins was 75 amino acids. It seems clear that no sequences upstream of this segment are needed for biotinylation since fusion of this segment to several sites within β -galactosidase (including a site only four amino acids from the amino terminus) gave biotinylated proteins as detected by [^{3} H]biotin labeling or by derepression of the biotin biosynthetic operon (see below).

It should be noted that some biotin fusions were degraded by intracellular proteolysis. One such example was a *lacZ*-1.3 S fusion where the fusion joint was the *EcoRI* site located within the *lacZ* gene (fusion E of Fig. 3). Virtually all of this fusion protein was degraded (Fig. 4), probably by clipping at the fusion junction. A similar clipping is often observed in the

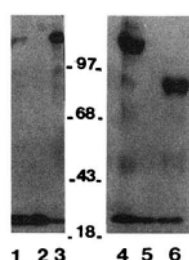


FIG. 4. Biotin labeling of some β -galactosidase fusions. Labeling of some of the fusions of Fig. 3 are shown. *E. coli* strains carrying the hybrid genes on pBR322 or pUC18 derived vectors were grown to early log phase in a medium consisting of minimal medium E (Miller, 1972) with 0.4% glycerol, 0.1% vitamin free casein hydrolysate, 41 nM biotin (1 μ Ci of [^{3} H]biotin/ml), and the antibiotics needed to select plasmid maintenance. The strains were induced for 2 h with 1 mM isopropyl-1-thio- β -D-galactopyranoside (except the culture of lane 3). The cells were collected from 1-ml cultures, lysed in a SDS containing buffer and separated on a 7.5% polyacrylamide gel run in the discontinuous mode in the presence of SDS. The gels were fluorographed using preflashed film. The structures of the fusions are given in Fig. 4. Lane 1, fusion A uninduced (pCY49J); lane 2, no fusion; lane 3, fusion A (pCY49J); lane 4, fusion A (pCY49J); lane 5, fusion E (pCY72) (a faint band was observed at ca 130 kDa upon prolonged exposure); lane 6, fusion F (pCY73). The labeled band at the bottom of all lanes is BCCP, the endogenous biotin protein of *E. coli*.

λ gt11 system which also produces fusions via the *lacZ* *EcoRI* site (Young and Davis, 1983). However, in *E. coli* degradation of biotinized fusions can be distinguished from inactive acceptor sequences by a biological test, the derepression of the biotin operon given by biotinized fusions, but not by fusions inactive in biotin acceptance.

The biological test is based on the fact that biotinoyl-AMP is the corepressor of the biotin (*bio*) operon of *E. coli*. High level expression of a protein with subsequent biotinylation results in consumption of biotinoyl-AMP giving derepression of *bio* operon expression that can be readily assayed by synthesis of β -galactosidase from a *lacZ* gene transcribed from one of the two *bio* operon promoters (Cronan, 1988; for review see Cronan, 1989). Derepression of the biotin operon demonstrates that a given fusion construct is expressed and biotinylated even when no biotinized fusion protein can be detected by [^{3} H]biotin labeling. For example, strains carrying fusion H (Fig. 3) show a 12-fold derepression of *bio* operon transcription, indicating the consumption of biotinoyl-AMP (via protein biotinylation), although no labeled operon protein was observed. The increased transcription indicates that fusion protein H was synthesized, biotinylated (causing derepression), and then rapidly degraded (small proteins are often degraded in *E. coli*). A similar result was seen for fusion E except that a small amount of undegraded biotinized fusion protein was observed.

We have also produced fusions of the 75 amino acid 1.3 S segment with several other proteins (Fig. 3), the neomycin phosphotransferases encoded by transposons Tn5 and Tn903, and the chloramphenicol acetyltransferase encoded by Tn9. In each case a new biotinized protein of the expected molecular weight was obtained. In the case of the acetyltransferase fusion, the resulting protein retained enzymatic activity (data not shown).

As mentioned above, fusions of segments of β -galactosidase with biotinylation sequences from other biological sources were also obtained (data not shown). Note that the *K. pneumoniae* and tomato sequences encode long stretches of unusual protein sequences upstream of the biotinylated lysine residue which should "insulate" the biotin-accepting segment from the remainder of the protein and from β -galactosidase. In the case of the *K. pneumoniae* sequence, the unusual segment is the repetitive Pro-Ala sequence noted above whereas that of the tomato sequence is a stretch of 8 proline residues, a putative proline helix encoded by a poly(dC) sequence introduced by the cDNA cloning protocol.³ It should be noted that the Pro-Ala sequence of the *K. pneumoniae* strongly resembles similar sequences found flanking the sites of lipoic acid attachment of lipoated proteins where the Pro-Ala segments have been shown to be extremely mobile (Radford *et al.*, 1989). The biotinylation of the *K. pneumoniae* and tomato fusions *in vivo* strongly indicates that these protein segments, like the 1.3 S segment, require no sequences upstream of the fusion joint for recognition by the *E. coli* biotin ligase.

A Bacterial Biotin Accepting Sequence also Functions in Yeast—The sequence conservation of biotinylation sites noted above (Fig. 2) coupled with the function of diverse sequences in *E. coli* suggested that a sequence that functioned in *E. coli* might also function in the yeast, *S. cerevisiae*. We chose the HIS3 gene of yeast since this gene is expressed in *E. coli* (due to the adventitious sequences giving promoter and ribosome binding functions) where it complements *E. coli* *hisB* mutants (Struhl and Davis, 1980). A hybrid gene which encoded the HIS3 protein (lacking only the last six amino acids) fused to the minimal *P. shermanii* biotinylation sequence (Fig. 3) was

³ E. Picherch, personal communication.

constructed on YEp24, an *E. coli*-*S. cerevisiae* shuttle vector. The protein encoded by the hybrid gene was biotinylated in both *E. coli* and *S. cerevisiae* (Fig. 5) and had the expected molecular mass (32 kDa) (Fig. 5). In yeast cells the synthesis of this 32-kDa biotinylated protein was regulated as is the normal HIS3 protein (Struhl, 1982) in that its synthesis was derepressed under conditions of histidine limitation. Upon derepression the HIS3 fusion protein became the major biotinylated protein of the yeast cells (Fig. 5). It should be noted that as expected from previous deletion analyses (Struhl and Davis, 1980, 1981) removal of the six terminal amino acids rendered the enzyme inactive in both *E. coli* and yeast. Also, note that synthesis of fusion L in *E. coli* failed to derepress *bio* operon expression (Fig. 3). We attribute this to the low level of fusion protein made (Fig. 5).

Detection and Purification of Biotinylated Fusion Proteins—In most of the experiments above, biotinylated fusion proteins were detected by *in vivo* labeling of *E. coli* with [³H]biotin. This labeling is almost as effective in wild type cells as in biotin auxotrophs due to the effective repression of biotin synthesis by exogenous biotin (Barker and Campbell, 1980, 1981). Labeling with radioactive biotin was also applicable to *S. cerevisiae*, a natural biotin auxotroph. We expect that since mammalian tissue culture cells also require biotin (Dakshinamurti *et al.*, 1985), efficient labeling of these cells should also occur. In addition to *in vivo* labeling, a very sophisticated

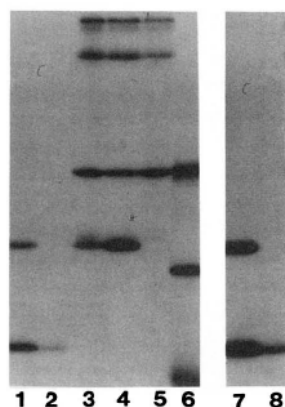


FIG. 5. Biotinylation of a HIS3-1.3S fusion protein in *E. coli* and *S. cerevisiae*. The hybrid gene carried by the yeast-*E. coli* shuttle vector YEp24 was maintained in *E. coli* by selection for antibiotic resistance and in *S. cerevisiae* by selection for uracil independence. The yeast strain used carried a deletion of the chromosomal URA3 and HIS3 loci and had a nonsense lesion in the LYS2 locus. The *E. coli* strains were labeled and prepared for electrophoresis as described in Fig. 4. The yeast strains were grown to early stationary phase in a minimal medium supplemented with glucose (0.6%), lysine 30 μ g/ml, biotin 20 nM (5 μ Ci/ml), and histidine HCl at either 50 μ g/ml (lane 3) or 2.5 μ g/ml (lane 4). The lower histidine concentration results in derepression of HIS3 expression (Struhl, 1982). The yeast cells were disrupted in a French pressure cell, and insoluble debris was removed by centrifugation. The soluble proteins were collected by trichloroacetic acid precipitation, washed free of acid, solubilized in an SDS-containing buffer, and electrophoresed on a 12.5% polyacrylamide gel. The gel was fluorographed as in Fig. 4. Lane 1, *E. coli* DH5 α carrying pCY106 encoding fusion L (Fig. 3); lane 2, *E. coli* DH5 α carrying pCY105, a YEp24 derivative with an intact HIS3 gene; lanes 3 and 4 are *S. cerevisiae* strain CTY182 carrying pCY106; lane 5, yeast strain CTY182 carrying pCY105; and lane 6, ¹⁴C-labeled molecular mass standards (Bethesda Research Laboratories). The standard proteins shown are ovalbumin, 43 kDa; carbonic anhydrase, 29 kDa; and β -lactoglobulin, 18.4 kDa. Lanes 7 and 8 are a longer exposure of lanes 1 and 2. In lane 5, the labeled bands in order of increasing mobility are acetyl-CoA carboxylase (205 kDa), pyruvate carboxylase (130 kDa), and an unknown protein of 44 kDa also observed by Lim *et al.* (1987). The band in lanes 1, 2, 7, and 8 is BCCP (22 kDa).

technology to detect biotinylated macromolecules using streptavidin or avidin has been developed (*e.g.* Buckland, 1986; Wilchek and Bayer, 1988, 1989). The extremely specific and high affinity (K_D 10^{-15} M) binding of biotin by avidin and streptavidin results in specific detection systems of very high sensitivity.

Binding to avidin or streptavidin can also be used to purify biotinylated proteins. However, the extremely tight binding of biotin and biotinylated proteins to avidin or streptavidin cannot be reversed by elution with biotin and thus elution of bound proteins requires much harsher methods. Swack *et al.* (1978) showed that biotinylated proteins present in crude mixtures of SDS solubilized proteins can be quantitatively bound to columns of immobilized avidin, washed free of contaminating proteins, and eluted by boiling the column matrix in SDS. We have used this method, modified by use of streptavidin bound to agarose via an eleven carbon "arm" and omission of urea (to prevent cyanation of proteins) to purify several biotinylated protein to homogeneity in a single step (data not shown). The resulting purified proteins lacked native structure, but were suitable for protein chemistry and/or antibody production. The ability of avidin to retain biotin binding in the presence of SDS or urea (Green, 1975) should be of great utility in purification of membrane proteins. Another approach (now under investigation) is to construct a fusion protein containing an engineered site for protease cleavage at the fusion junction. This is the approach of Nagai and Thorngensen (1987) where the protease cleavage site (*e.g.* for the blood coagulation protease factor Xa) is one of several specific tetrapeptide sequences rarely found in proteins. Such a fusion protein could be bound to immobilized avidin or streptavidin, washed free of contaminating proteins, and simultaneously eluted in native form (and freed of the biotinylated segment) by treatment of the column matrix with the designated protease. Avidin and streptavidin are extremely insensitive to proteases in general, and since the amino acid sequences of both proteins are known, selection of proteases unable to cleave these proteins is readily accomplished.

Purification of Biotinylated Proteins without Denaturation—Avidin can be modified to decrease the affinity for biotin, thus permitting elution of biotinylated proteins by competition with free biotin. Such columns are readily and reproducibly prepared by conversion of avidin from its normal tetrameric form to a monomeric form. These "monomer avidin" columns are obtained by treatment of columns of avidin bound to agarose with guanidine solutions (*e.g.* Green, 1975; Henrickson, *et al.*, 1979; Kohanski and Lane, 1985). This treatment denatures the protein resulting in release of those subunits not covalently bound to the column matrix. Renaturation of the avidin monomers remaining bound to the matrix restores biotin binding ability, but most sites are of much lower affinity (K_D $\sim 10^{-6}$ – 10^{-7} M) than the tetrameric species. Any remaining high affinity sites can be blocked with biotin thus giving columns from which bound biotinylated proteins can be quantitatively eluted by conventional nondenaturing buffers containing biotin. Columns of monomer avidin have routinely been used to purify biotinylated enzymes and we have purified several biotinylated β -galactosidase fusion proteins (Figs. 6 and 7) from crude extracts with almost quantitative recovery of enzyme activity. Binding of β -galactosidase activity to a monomer avidin column required both the biotinylation sequence and unoccupied biotin binding sites (Fig. 6). The only non- β -galactosidase protein eluted with biotin was the endogenous biotin protein of *E. coli*, BCCP (Fig. 7). The columns are readily regenerated (Kohanski and Lane, 1985).

It should also be noted that biotin is an effective hapten

(Berger, 1979) and polyclonal and monoclonal antibodies against biotin are commercially available. Such antibodies should be a useful alternative to avidin and streptavidin for detection and purification of biotinylated fusion proteins.

Efficiency of Fusion Protein Biotinylation—The production of a biotin accepting fusion protein could be toxic to cellular metabolism by decreasing biotinylation of endogenous biotin proteins essential to cellular metabolism. Toxicity could be due to depletion of intracellular biotin or to titration of the available biotin ligase activity or both. The potential problem of biotin depletion can be readily overcome by providing high concentrations of biotin in the growth medium. The biotin transport systems of *E. coli* (Barker and Campbell, 1980), *S. cerevisiae* (Rogers and Lichstein, 1969), and mammalian tissue culture cells (Dakshinamurti *et al.*, 1985) are able to transport biotin at sufficiently high rates to preclude biotin depletion. There also is evidence that biotin at high concentrations can enter *E. coli* by diffusion (Barker and Campbell, 1980).

Titration of biotin ligase activity is more problematical. High level and prolonged expression of a biotinylated fusion proteins can result in deficient biotinylation of endogenous biotin proteins. In *E. coli*, the only endogenous biotinylated protein is the BCCP subunit of acetyl-CoA carboxylase, an enzyme catalyzing an essential step in fatty acid synthesis. High level expression of some fusion proteins resulted in decreased biotinylation of BCCP (Fig. 4), resulting in growth inhibition (data not shown). However, the gene (*birA*) encoding *E. coli* biotin ligase has been cloned (Barker and Campbell, 1981) and multicopy plasmids carrying the *birA* gene are available. Such plasmids overproduce biotin ligase and overcome the growth inhibitory effects of fusion protein production while increasing the yields of biotinylated fusion proteins. Depending on the relative levels of fusion protein and ligase expression quantitative biotinylation of fusion proteins can be obtained in *E. coli*. Buoncristiani and Otsuka (1988) have shown that *E. coli* biotin ligase can be overproduced by >600-fold without deleterious effects on cellular growth. Using a similar plasmid (pCY100, see "Experimental Procedures") we have obtained quantitative biotinylation of very highly expressed ($\sim 3 \times 10^4$ molecules/cell) fusion proteins. The *S. cerevisiae* ligase gene has not yet been cloned although ligase-deficient mutants (Mishina *et al.*, 1980) should allow cloning by genetic complementation. However, in this yeast (unlike *E. coli*) the endogenous biotin proteins are not essential to growth when the products of the two metabolically essential biotin proteins (acetyl-CoA carboxylase and pyruvate carboxylase) are provided in the growth medium (Mishina *et al.*, 1980). Addition of these products, fatty acids and oxalacetate (provided as aspartate), should reduce or eliminate toxic effects of fusion protein biotinylation on metabolism. The *E. coli* ligase, a polypeptide of 35.5 KDa, could also be expressed in these heterologous systems in order to provide increased ligase levels.

CONCLUSIONS

Biotinylated protein fusions provide several distinct advantages over conventional fusions (*e.g.* to β -galactosidase). These include (i) the ability to efficiently purify the fusion protein in either native or denatured form; (ii) specific, sensitive, and facile detection by labeling with radioactive biotin or by indirect detection by avidin or streptavidin using well developed techniques done with commercially available reagents; (iii) fusion of a much smaller protein segment to the protein of interest. This last advantage should decrease the problems seen with the much larger protein segments (*e.g.* β -

galactosidase) commonly used to produce fusions. Large protein segments can dominate the immune response during antibody production such that the epitopes of the protein of interest are seldom recognized. The size of the fusion partner can also preclude use of the fusion to study intracellular trafficking and localization of the protein of interest. For example, β -galactosidase protein fusions are unable to transverse the inner membrane of *E. coli* even when fused to sequences able to promote the export of other proteins (Lee *et al.*, 1989). It seems possible that biotinylated proteins can be secreted through biological membranes. Kurzychaka *et al.* (1988) have shown that proteins biotinylated *in vitro* by an indirect method can be secreted into the lumen of isolated microsomal vesicles. Moreover, most of the biotinylated protein species of mammalian cells are mitochondrial, whereas the ligase activity has the distribution of a cytosolic protein (Cohen *et al.*, 1985). These data suggest these proteins may enter mitochondria as the biotinylated species. Thus, biotinylated fusion proteins may be useful in studying problems of protein trafficking and localization. We anticipate that fusion proteins biotinylated in *E. coli* and yeast should also be biotinylated when expressed in cells of higher eucaryotes, but this has not yet been tested.

It should be noted that the present monomer avidin columns should only be a first step in the technology to purify fusion proteins in a native conformation. The streptavidin crystal structure has recently been solved (Weber *et al.*, 1989; Hendrickson *et al.*, 1989). These structures account for the decreased biotin binding affinity of avidin upon dissociation from the tetrameric form to the monomeric form. One of the four tryptophan residues forming the hydrophobic biotin binding site of a given subunit is derived from a neighboring (cydad-related) subunit. Thus, monomerization removes this residue from the biotin binding site giving a lower affinity. The streptavidin gene has been cloned and sequenced (Agaraña *et al.*, 1986) and expressed in *E. coli* (Sano and Cantor, 1990). Site-directed mutagenesis guided by the crystal structure should produce tetrameric streptavidin molecules with the affinity of the monomer (or in principle any given affinity). Such tetrameric molecules should be more stable than monomers to proteases and denaturants and provide a superior column-bound ligand. It should be noted that a reverse type of fusion strategy, fusion of streptavidin sequences to the protein of interest, is also possible. However, this seems unlikely to be an advantage over the present strategy, since the residues comprising the biotin binding site of streptavidin is spread over at least 100 residues of primary sequence (Weber *et al.*, 1989), and binding of streptavidin to intracellular biotin and/or metabolically essential biotinylated proteins should disrupt cellular metabolism. Indeed, expression of the streptavidin gene in *E. coli* has recently been shown to be lethal to the host (Sano and Cantor, 1990).

Acknowledgments—I thank the persons mentioned in the text for their kind gifts of the DNA segments and microbial strains used in this work. I thank Kelly Reed, S.-J. Li, and Tim Morris of this laboratory for useful discussions and unpublished information.

REFERENCES

- Agaraña, C. E., Kuntz, I. D., Birkin, S., Axel, R., and Cantor, C. R. (1986) *Nucleic Acids Res.* **14**, 871-882
- Bai, D.-H., Moon, T.-W., Lopez-Casillas, Andrews, P. C. and Kim, K.-H. (1989) *Eur. J. Biochem.* **182**, 239-245
- Barker, D. F., and Campbell, A. M. (1980) *J. Bacteriol.* **143**, 789-800
- Barker, D. F., and Campbell, A. M. (1981) *J. Mol. Biol.* **146**, 469-492
- Berg, D. E., Weiss, A. and Crossland, L. (1980) *J. Bacteriol.* **142**, 439-446

- Berger, M. (1979) *Methods Enzymol.* **62**, 319-326
- Botstein, D., Falco, S. C., Stewart, S. E., Brennan, M., Scherer, S., Stinchcomb, D. T., Struhl, K., and Davis, R. W. (1979) *Gene (Amst.)* **8**, 17-24
- Buckland, R. M. (1986) *Nature* **320**, 557-558
- Buonocristiani, M. R., and Otsuka, A. J. (1988) *J. Biol. Chem.* **263**, 1013-1016
- Chandler, C. S., and Ballard, F. J. (1988) *Biochem. J.* **251**, 749-755
- Cohen, W. D., Thomas, M., and Stack, M. (1985) *Ann. N. Y. Acad. Sci.* **447**, 393-395
- Cronan, J. E., Jr. (1988) *J. Biol. Chem.* **263**, 10332-10336
- Cronan, J. E., Jr. (1989) *Cell* **58**, 427-429
- Dakshinamurti, K., Chalifour, L., and Bhullar, R. P. (1985) *Ann. N. Y. Acad. Sci.* **447**, 38-54
- Fall, R. R. (1979) *Methods Enzymol.* **62**, 390-398
- Green, N. M. (1975) *Adv. Protein Res.* **29**, 85-133
- Hendrickson, W. A., Pähler, A., Smith, J. L., Satow, Y., Merritt, E. A., and Phizackerley, R. P. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86**, 2190-2194
- Henrickson, K. P., Allen, S. H. G., and Maloy, W. L. (1979) *Anal. Biochem.* **94**, 366-370
- Hoffman, N. E., Pichersky, E., and Cashmore, A. R. (1987) *Nucleic Acid Res.* **15**, 3928
- Ketner, G., and Campbell, A. (1974) *Proc. Natl. Acad. Sci. U. S. A.* **71**, 2698-2702
- Kohanski, R. A. and Lane, M. D. (1985) *Ann. N. Y. Acad. Sci.* **447**, 373-385
- Kurzchaka, T. V., Wiedman, M., Bretes, H., Zimmerman, W., Bauschke, E., and Rapport, T. A. (1988) *Eur. J. Biochem.* **172**, 663-668
- Lee, C., Li, P., Inouye, H., Brickman, E. R. and Beckwith, J. (1989) *J. Bacteriol.* **171**, 4609-4616
- Lim, P., Rhode, M., Morris, C. P., and Wallace, J. C. (1987) *Arch. Biochem. Biophys.* **258**, 219-264
- Lim, F., Morris, C. P., Occhiodoro, F., and Wallace, J. C. (1988) *J. Biol. Chem.* **263**, 11493-11497
- Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York
- McAllister, H. C., and Coon, M. J. (1966) *J. Biol. Chem.* **241**, 2855-2861
- Mead, D. A., Szczesma-Skorupa, E., and Kemper, B. (1986) *Protein Eng.* **1**, 67-74
- Miller, J. H. (1972) *Experiments in Molecular Genetics*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York
- Mishina, M., Roggenkamp, R., and Schweizer, E. (1980) *Eur. J. Biochem.* **111**, 79-87
- Murtif, V. L., and Samols, D. (1987) *J. Biol. Chem.* **262**, 11813-11816
- Nagai, K., and Thorgensen, H. C. (1987) *Methods Enzymol.* **153**, 461-481
- Nikolau, B. J., Wurtele, E. S., and Stumpf, P. K. (1985) *Anal. Biochem.* **149**, 488-493
- Perham, R. N., Packman, L. C., and Radford, S. E. (1987) *Biochem. Soc. Symp.* **54**, 67-81
- Radford, J. E., Lane, E. D., Perham, R. N., Martin, S. R., and Appella, E. (1989) *J. Biol. Chem.* **264**, 767-775
- Robinson, B. H., Oei, J., Saunders, M., and Gravel, R. (1983) *J. Biol. Chem.* **258**, 6660-6664
- Rock, C. O., and Cronan, J. E., Jr. (1981) *Methods Enzymol.* **71**, 341-351
- Rogers, T. O., and Lichstein, H. C. (1969) *J. Bacteriol.* **100**, 557-564
- Rüther, U., and Müller-Hill, B. (1983) *EMBO J.* **2**, 1791-1794
- Rüther, U., Koenen, M., Otto, K., and Müller-Hill, B. (1981) *Nucleic Acids Res.* **9**, 4087-4098
- Sano, T., and Canto, C. R. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87**, 142-146
- Samols, D., Thornton, C. G., Murtif, V. L., Kumar, G. K., Haase, F. C., and Wood, H. G. (1988) *J. Biol. Chem.* **263**, 6461-6464
- Schwarz, E., Oesterheld, D., Reinke, H., Beyreuther, K., and Dimroth, P. (1988) *J. Biol. Chem.* **263**, 9640-9645
- Sherman, F., Fink, G. R., and Hicks, J. (1983) *Methods in Yeast Genetics*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York
- Silhavy, T. J., and Beckwith, J. R. (1985) *Microbiol. Rev.* **49**, 398-418
- Struhl, K. (1982) *Nature* **300**, 284-287
- Struhl, K., and Davis, R. W. (1980) *J. Mol. Biol.* **136**, 309-332
- Struhl, K., and Davis, R. W. (1981) *J. Mol. Biol.* **152**, 553-568
- Swack, J. A., Zander, G. L., and Utter, M. F. (1978) *Anal. Biochem.* **87**, 114-126
- Takai, T., Yokoyama, C., Wada, K., and Tanabe, T. (1988) *J. Biol. Chem.* **263**, 2651-2657
- Takeshita, S., Sato, M., Toba, M., Marahashi, W., and Hashimoto-Gotoh, T. (1987) *Gene (Amst.)* **61**, 63-74
- Viera, J., and Messing, J. (1982) *Gene (Amst.)* **19**, 219-268
- Weber, P. C., Ohlendorf, D. H., Wendolaski, J. J., and Salemme, F. R. (1989) *Science* **234**, 85-88
- Wilchek, M., and Bayer, E. A. (1988) *Anal. Biochem.* **171**, 1-32
- Wilchek, M., and Bayer, E. A. (1989) *Trends Biochem. Sci.* **14**, 408-411
- Wilchek, M., and Bayer, E. A. (1990) *Methods Enzymol.* **184**, in press
- Young, R. A., and Davis, R. W. (1983) *Proc. Natl. Acad. Sci. U. S. A.* **80**, 1194-1198

SUPPLEMENTARY MATERIAL TO:

Biotinylation of Proteins in Vivo: A Post-Translational Modification to Label, Purify, and Study Proteins by John E. Cronan, Jr.

EXPERIMENTAL PROCEDURES

Bacterial and Yeast Strains and Media

The *E. coli* K-12 strains generally used were F'11recA and BMH71-18 (Rüther and Müller-Hill, 1983) the former for screening for an intact *lacZ* gene and the latter for α -complementation. Strains DH5a (Bethesda Research Laboratories) and F'M15 *recA* Rüther *et al.* (1981) were also used for α -complementation screening. Strain RD100 (*hisB333*) was used for *hisB* complementation (Struhl and Davis, 1980). The yeast strain used was CTY182 (MATa *ura3-52. shis3-200. lys2-80*) from V. Bankaitis, University of Illinois.

The bacterial media were described previously (Cronan, 1988). The yeast defined minimal medium with the inositol selection medium of Sherman *et al.* (1983) with biotin was omitted. Uracil, lysine, and biotin were added at 20, 30, or 1 μ g/ml, respectively. Histidine-HCl was added at either 2.5 or 50 μ g/ml. γ -[^{35} S] Biotin was purchased from Amersham.

Plasmid Constructions

Plasmids pUR288 and 289 (Rüther and Müller-Hill, 1983) were used in construction of that 8-galactosidase fusion encoding 7650 residues of 8-galactosidase. Plasmids pUC8 (Viera and Messing, 1982) and pTZ18R (Mead *et al.*, 1986) were used to construct fusions containing only a few 8-galactosidase residues. Plasmid ptacl.3t (Murtif and Samols, 1987) was the source of the 1.35 gene sequences. The plasmids used for construction of the fusions to the *Tn2* Cat gene, the *Tn923* Kan gene and the *Tn5* Kan gene were pHS397, (Takeshita, *et al.*, 1987) pUC4K (Viera and Messing, 1982), and pCI5, respectively. pCY5 was constructed by *Bam*HI and *Hind*III digestion of a λ Tn5 phage (Berg *et al.*, 1975) followed by filling of the ends with DNA polymerase I and ligation into the *Sma*I site of pUC8 (the *kan*^R gene is transcribed in the same direction as *lacZ*). pMBR10 (Buonocristiani and Otsuka, 1988) was the gift of A. Otsuka. The *HIS3* gene came from plasmid pWJ79, a pBR322 derivative carrying the original

yeast *Bam*HI *HIS3* fragment of Struhl and Davis (1980), that was the gift of T. Davis, University of Washington, as was YEp24 (Botstein *et al.*, 1979).

Plasmid pCY46, was constructed by ligation of the *Sma*I-*Bam*HI fragment of ptacl.3t to pBR328 cut with the same enzymes. Plasmid pCY49J was constructed by *Sma*I-*Pst*I digestion of pCY46 and pUR288 followed by ligation and transformation of F11recA, and selection for *Amp*^R and *Cm*^R with screening for *lac*⁺. pCY56 was constructed by ligating *Eco*RV cleaved pCY49J to *Hind*III digested pUC4K with selection for *Amp*^R, *Kan*^R, and *Cm*^R. The *Kan*^R determinant was oriented opposite to that of the 1.3t segment. An *Amp*^S, *Kan*^S derivative of pCY56 was constructed by digestion of pCY56 with *Sca*I followed by recircularization with ligase. pCY66 and pCY68 were constructed by ligation of *Nar*I digested pCY56 with *Acl*I digested pTZ18R or pUC8, respectively. Ligation was followed by selection for *Amp*^R *Kan*^R followed by screening for *Cm*^S and *lac*⁺. pCY72 resulted from digestion of pCY66 with *Eco*RI and *Xba*I followed by ligation and selection for *Kan*^R *Amp*^R with screening for *lac*⁺. pCY73 and pCY74 was constructed as was pCY72 except the enzymes used with *Sma*I were *Sat*I or *Xba*I, respectively. pCY84 was constructed by digestion of pCY49J with *Sma*I and *Xba*I followed by recircularization. pCY90 was constructed by digestion of pUR288 with *Bam*HI. The resulting *Bam*HI ends were filled with DNA polymerase I and ligated to *Nar*I digested pCY68. pCY94 resulted from ligation of *Sma*I-*Sph*I digested pCY66 to pHS397 digested with *Sca*I and *Sph*I followed by selection for *Cm*^R and screening for loss of α -complementation. pCY105 was constructed by insertion of the *Bam*HI fragment of pWJ79 into the *Bam*HI site of YEp24 such that the orientation of *HIS3* was opposite that of *URA3*. pCY106 was constructed by *Xba*I-*Sph*I digestion of both pCY105 and pCY66 followed by ligation, transformation and selection for *Amp*^R, *Kan*^R plasmids that failed to complement a *hisB* strain. pCY177 was constructed by *Pst*I digestion of pCY66 followed by recircularization to eliminate the *Kan*^R element) to give pCY115. pCY115 and pUC4K were digested with *Eco*RI and ligated to restore *Kan*^R (pCY116). The net result of these manipulations was to move the *Kan*^R cassette from a position downstream of

the 1,35 coding segment to a position upstream of (and with the same orientation as) the 1,35 sequences. The *Kan^R* and 1,35 sequences were then fused by digestion with *Cla*I and *Sma*I followed by filling in the resulting *Cla*I ends and recircularization with ligase to give pCY117.

pCY118 was constructed by digestion of pCY5 and pUC8 with *Nci*I and *Bgl*II followed by ligation and selection for *Amp^R Kan^R*. pCY119 was the result of digestion of pUC8 and ptacl.3t with *Hind*III and *Sma*I (the *Sma*I digestion of ptacl.3t was a partial digest) followed by ligation. pCY120 was constructed by *Xho*I-*Hind*III digestion of ptacl.3t followed by ligation to *Sma*I-*Hind*III cleaved pUC8 followed by digestion with *Sma*I and selection for *Amp^R* with screening for loss of α -complementation. pCY100 was constructed by ligation of the *Bam*HI-*Sma*I fragment of pMBR10 to the large *Bam*HI-*Eco*RV fragment of pACYC184 (Maniatis *et al.*, 1982). pCY100 was shown to give superrepression of *bio* operon expression (Barker and Campbell, 1981).

All plasmid constructs were confirmed by detailed restriction mapping. In several cases, the fusion junction was within a new (hence diagnostic) restriction site. Other junctions were confirmed by DNA sequencing, by production of a biotinylated fusion protein or by biotin operon derepression. Plasmid constructions and transformation employed conventional procedures (Maniatis *et al.*, 1982; Sherman *et al.*, 1983).

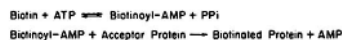
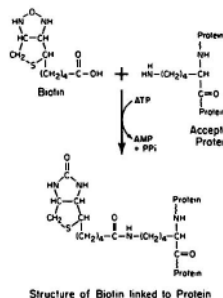


Fig. 1. Biotin, biotinylated proteins, and the biotin ligase reaction.

Biotin is attached to acceptor proteins by biotin ligase in a two step reaction given in the lower part of the figure. Biotin is first converted to biotinoyl-AMP which then reacts with the ϵ amino group of a specific lysine residue of the acceptor protein. The biotinoyl lysine moiety is called biocytin.



Fig. 2. Sequences of the carboxyl termini of several biotinylated proteins.

The sequences are aligned at the lysine residue which becomes biotinylated (arrow). The sequences shown are ECRCCP, the *Escherichia coli* biotin carboxyl carrier protein (BCCP, a subunit of acetyl-CoA carboxylase); PS 1,35, the 1,35 subunit of *Propionibacterium shermanii* transcarboxylase; YPYC, *Saccharomyces cerevisiae* pyruvate carboxylase; HPTC, human pyruvate carboxylase; and Tom, a c-DNA sequence from tomato. The protein corresponding to the tomato sequence is unknown, the segment was identified by its encoded biotin acceptor activity and homology to the *P. shermanii* sequence (Hoffman *et al.*, 1987). The boxed residues are those residues which are conserved among the proteins. Shaded boxes are residues of very similar properties (the branched chain amino acids or the amidated acids). Much less homology is found elsewhere in these proteins. A more extensive comparison is given by Samols *et al.* (1988) and the original references are found therein except for yeast pyruvate carboxylase (Lim *et al.*, 1988).

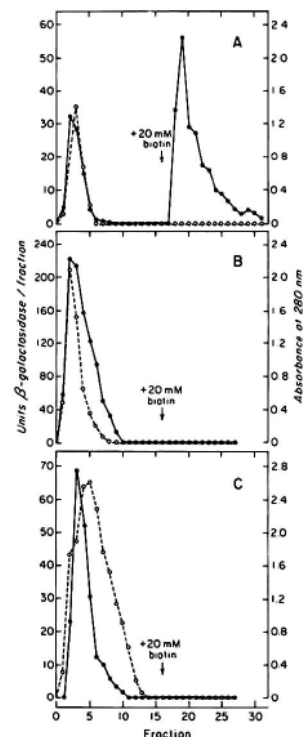


Fig. 6. Purification of a β -galactosidase fusion protein by monomer avidin chromatography.

One hundred ml cultures of *E. coli* Fli reek carrying either the fusion plasmid pCY74 (encoding fusion B, Fig. 3) or the β -galactosidase vector pUR288 were grown to early exponential phase in a broth medium, induced with 1 mM IPTG for 3 h and harvested. These strains also carried pBA11 (Barker and Campbell, 1981) a compatible plasmid that overproduces biotin ligase ≈ 10 -fold. The cells were harvested and disrupted in Z buffer (Miller, 1972). The resulting lysate was centrifuged at $48,000 \times g$ for 1 h and the supernatants (≈ 2 mg protein) were applied to 0.5 ml columns of monomer avidin-Sepharose prepared as described by Hendrickson *et al.*, (1979) (exchangeable biotin binding capacity of 35 nmol/ml gel). The columns were eluted with Z buffer or Z buffer containing 20 mM biotin (arrows). Fractions of ≈ 250 μ l were collected and assayed for β -galactosidase (Miller, 1972) and protein (absorbance at 280 nm of a twenty-fold dilution). Panel A, fusion B encoded by pCY74; panel B, wild type β -galactosidase encoded by pUR288; panel C, as in panel A except the column was saturated with 20 mM biotin before application of protein.

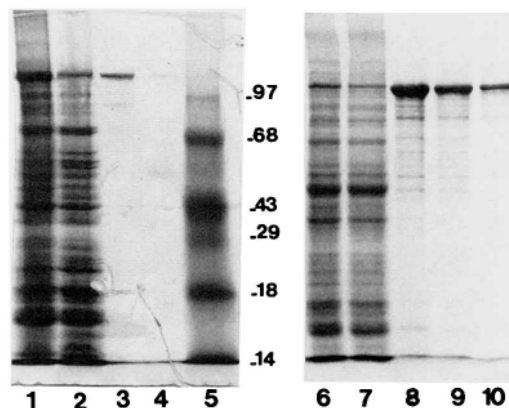


Fig. 7. SDS gel electrophoresis of monomer avidin-Sepharose fractions.

The results of two β -galactosidase purifications done as described in Fig. 6, panel 1. Lanes 1-5 result from chromatography of ≈ 20 mg (total extract protein) of a fusion B extract (Fig. 3) on a 2.5 ml monomer avidin-Sepharose column. The gels were of 8% acrylamide and were stained with Coomassie Blue R. Lane 1 is a sample of the original supernatant applied to the column. Lane 2 is the unbound protein (flow-through) whereas lane 3 is a sample of the eluate obtained with 20 mM biotin. Lane 4 is a sample identical to that of lane 3 except that the sample was treated with a monoclonal anti- β -galactosidase (Promega) followed by absorption with protein A-agarose and centrifugation. The resulting supernatant was loaded on the gel. The faint bands visible in the lower half of the gel are unabsorbed immunoglobulin chains.

Lanes 6-10 are the results of chromatography of ≈ 5 mg (total extract protein) of a fusion B (Fig. 3) preparation on a 1.5 ml monomer avidin-Sepharose column. Lanes 6 and 7 are samples of the supernatant loaded and unbound protein, respectively whereas lanes 8, 9, and 10 are samples of the peak and tailing regions of the 20 mM biotin eluate, respectively. The host strains carried pCY74 plus either of two *bio* (biotin ligase overproducing) plasmids. The strain used in lanes 1-5 carried plasmid pCY100 which overproduces the *bio* protein ≈ 100 -fold whereas the strain used in lanes 6-10 carried pBA11 which overproduces ligase activity ≈ 10 -fold (Barker and Campbell, 1981). The minor bands in lanes 8-10 were removed by absorption with anti- β -galactosidase (data not shown) except the band of greatest mobility which is a protease cleaved form of BCCP (Fall, 1979).