Structural and Functional Studies of Biotin Protein Ligase and its Bacterial Substrate Acetyl-CoA Carboxylase

Cecile M. V. Tron

PhD Thesis

The University of Edinburgh

2008

Declaration

I, Cecile M. V. Tron, hereby certify that this thesis has been composed by myself, and that it is a record of my work, and that it has not been accepted in partial or complete fulfilment of any other degree or professional qualification.

Acknowledgements

I would like to acknowledge Dr. Dominic Campopiano and Prof. Robert Baxter for offering me the opportunity to carry out this PhD. I would like to thank them for their supervision, their guidance and continued support during the course of this work.

Without the following people this work could not have been completed.

Crystallographic studies were carried out in the Structural Biology Group of Prof. Malcom Walkinshaw with the precious help of Dr. Iain McNae (University of Edinburgh).

Prof. Alan Cooper and Margaret Nutley (Biological Microcalorimetry Facility, University of Glasgow).

Prof. Carol Robinson and Laura Lane (University of Cambridge)

I would like to acknowledge Dr. Andrew Cronshaw (University of Edinburgh) and all the members of the lab 229, past and present, and in particular Dave Clarke, Josefin Bartholdson and Ross Stevenson for their help with experiments.

Finally, I would also like to thank my family for supporting me since the beginning of my studies and all the friends I met during my PhD at the University of Edinburgh, especially Maria Manoli and Aileen Mitchell.

Abstract

Biotin protein ligase (BPL, EC: 6.3.4.15) catalyzes the formation of biotinyl-5'-AMP from biotin and ATP, and the succeeding biotinylation of the biotin carboxyl carrier protein (BCCP). The gene encoding the BPL from the hyperthermophyle *Aquifex aeolicus* were overexpressed in *Escherichia coli* and the recombinant protein *Aa*BPL was purified to homogeneity. *Aa*BPL was proven to be catalytically active and to biotinylate specifically the C-terminal biotinyl domain of BCCP (BCCPΔ67). The isolation and characterisation of a chemically cross-linked *Aa*BPL:BCCPΔ67 complex is described here. Isothermal titration calorimetry experiments revealed that in the class I *Aa*BPL, the presence of biotin is not required for ATP binding in the absence of Mg^{2+} ions and the binding of biotin and ATP has been determined to occur *via* a random but cooperative process. The residues involved in substrate binding were further investigated by X-ray crystallography and we report the structures at 2.3 Å resolution of *Aa*BPL in the apo-form and bound to biotin and ATP. This is the first crystal structure of a BPL in complex with biotin and ATP but also of an ATP-bound BPL. The Arg40 residue from the conserved GXGRXG motif is shown to play a key role in stabilizing the two ligands within the active site. The structures of the mutant *Aa*BPL R40G both in the apo- and biotin-bound forms reveal significant conformational differences in the catalytic domain of the enzyme and ITC measurements indicate that binding of ATP to *Aa*BPL R40G is very weak in the absence or presence of biotin. *Aa*BPL R40G remains catalytically active but shows poor substrate specificity.

In a separate study, the assembly states of the full-length *E. coli* BCCP (*Ec*BCCP) and biotin carboxylase (BC) subunits of acetyl-CoA carboxylase (ACC) have been investigated by mass spectrometry in collaboration with Prof. Carol Robinson (University of Cambridge). The *accB* and *accC* genes were overexpressed in *E. coli* and analysis of the purified recombinant *Ec*BCCP and BC proteins indicated that while BC is dimeric, *Ec*BCCP forms decamers and other oligomers in solution. Cloning of the C-terminal fragment BCCP87 and of the BC:*Ec*BCCP complex were also carried out for further understanding of the exact stoichiometry of the BCCP, BC and CT components within the multi-subunit *E. coli* ACC.

Abbreviations

aaRS – Aminoacyl tRNA synthetase

ACC – Acetyl-CoA carboxylase

AMP – Adenosine 5'-monophosphate

ATP – Adenosine 5'-triphosphate

BC – Biotin Carboxylase

BCCP – Biotin carboxyl carrier protein

Bp – Base pair(s)

BPL – Biotin protein ligase

CSD – Charge state distribution

CID – Collision induced dissociation

CoA – Coenzyme A

CT – Carboxyltransfase

Da – Daltons

DNA – Deoxyribonucleic acid

DTT – Dithiothreitol

EDC – 1-ethyl-3(dimethylamino-popyl)-cabodiimide

EDTA – Ethylene diaminetetracetic acid

ESI – Electrospray ionization

FAS – Fatty acid synthase

HCS – Holocarboxylase synthetase

HEPES – N-(2-hydroxyethyl)piperazine-N'-ethanesulfonic acid

IPTG – Isopropyl-1-thio-β-D-galactopyranoside

ITC – Isothermal titration calorimetry

LB – Luria Bertani

LplA – Lipoate protein ligase A

MALDI – Matrix assisted laser desorption-ionization

β-ME – β-mercaptoethanol

MS – Mass spectrometry

nESI – Nanoelectrospray ionization

NHS – N-hydrosuccinimide

ORF – Open reading frame

PAGE – Polyacrylamide gel electrophoresis

PC – Pyruvate carboxylase

PCC – Propionyl-CoA carboxylase

PCR – Polymerase chain reaction

PEP – Phosphoenolpyruvate

Q-ToF – Quadrupole time-of-flight

Rpm – Revolutions per minute

RRM – RNA recognition motif

SDS – Sodium dodecyl sulfate

SH3 – Src-homology 3

TFA – Trifluoro acetic acid

TRIS – Trishydroxymethylaminomethane

Table of Contents

List of Figures

Chapter 4: The *A. aeolicus* Mutant *Aa*BPL R40G

Chapter 5: Mass Spectrometry Analysis of the BCCP and BC subunits of *E. coli* ACC

Chapter 6: Materials & Methods

List of Tables

Chapter 1: Biotinylation – Specificity of an Essential Posttranslational Modification

1.1 The Vitamin Biotin

1.1.1 The Vitamin B_7: Biotin

Vitamins are ubiquitous organic compounds required by all kinds of life that act as co-enzymes in essential metabolic reactions.[1; 2] A common feature of the vitamins is that their biosynthesis does not occur in mammals but does in other organisms. Humans are believed to have lost the ability to biosynthesize these components and have developed instead mechanisms of uptake and absorption.[3] The significance of eating certain foods in order to maintain health was recognized long before vitamins were identified in the nutritional diet and several mysterious and fatal diseases resulting from vitamin deficiency were still prevalent in the past century.[4; 5]

In the late 19th century, Eijkman, who was awarded the 1929 Nobel Prize in Physiology or Medicine with Hopkins for their discovery of several vitamins, and the Russian surgeon Lunin suggested the existence of nutrients other than minerals, fatty-acids, carbohydrates or amino-acids essential in small amounts for the metabolism in living organisms.[6; 7] In 1912, Funk proposed that the diseases beriberi, pellagra, rickets and scurvy resulted from the lack of four different vital nutrients and was first to isolate a water-soluble complex of micronutrients which he proposed to be named "Vitamine", standing for *vital* and *amine*.[8] Those accessory growth factors were divided in 1915 by McCollum and Davis into the fat-soluble (A) and the water-soluble (B) vitamins.[9] Ten years latter, Jansen successfully isolated a small amount of thiamine and in 1933, Williams purified the crystalline compound from rice polishings which allowed the characterisation and the synthesis of the vitamin.[10; 11] Over time, more compounds were identified as vitamins in both the fat-soluble (A, D, E, K) and water-soluble (B_1-B_{12}, C) families, many of whom were latter shown not to

be essential and withdrawn from the list.[3] It is not yet known if all vitamins have been discovered but it has been shown that rats can survive on an almost completely synthetic diet.[12] Because vitamins are required by all organisms, their biosynthesis in microorganisms and their functions remain of considerable interest.[13; 14]

Biotin is the water-soluble B_7 vitamin, also named vitamin H or co-enzyme R, and was first discovered in the laboratory of Karl Folker at Merck in 1935 while searching for the anti-pernicious anemia factor. The vitamin was isolated the same year by Kögl and Tönnis, who purified 1.1 mg of the crystalline compound as its methyl ester form from more than 250 kg of egg yolk.[15-17] The structure of biotin was subsequently deduced by Du Vigneaud *et al.* in 1942.[18; 19] Biotin has the chemical formula $C_{10}H_{16}N_2O_3S$ and contains three chiral centres. Of the eight stereoisomers, only the dextrorotatory D-(+)-biotin is biologically active (Fig. 1.1). The structure reveals an ureido (tetrahydroimidazolidone) group fused to a tetrahydrothiophene ring attached to a valeric acid.

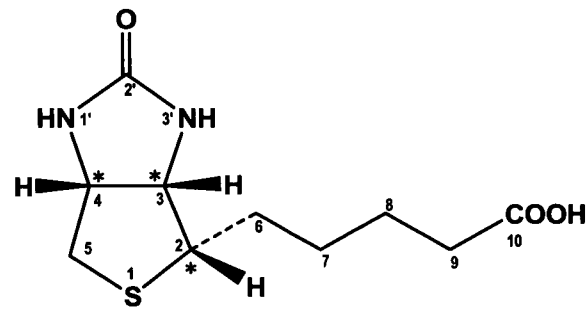


Figure 1.1 - Structure of the biologically active isoform D-(+)-biotin. The atoms are numbered and the (*) indicate the chiral centres of biotin.

Biotin is produced in small quantities by microorganisms and plants while mammals rely on dietary intake to satisfy their biotin requirements (human RDA: 30.0 μg).[20] The vitamin is synthesized from pimeloyl-CoA and the enzymes involved

in the biosynthetic pathway have been studied in details in organisms including *Escherichia coli, Bacillus subtilis* and *Saccharomyces cerevisae.*[21-24] The well characterised *E. coli* biotin biosynthetic operon *bioABFCD* is composed of five closely linked genes which encode respectively the enzymes 7,8-diaminononanoate synthetase (DANS), biotin synthase, 8-amino-7-oxononanoate synthase (AONS), BioC and dethiobiotin synthase (DTBS).[22] AONS catalyses the conversion of pimeloyl-CoA and alanine to 8-amino-7-oxononanoate and the subsequent transfer of the amino group from S-adenosine-L-methionine resulting in 7,8-diaminononanoate is catalysed by DANS. DTBS catalyses the formation of the ureido ring in the presence of carbon dioxide and MgATP and the final step catalysed by biotin synthase consists of the insertion of the sulfur atom.[25; 26]

1.1.2 Biotin-Dependant Carboxylation

Biotin plays key roles in essential carboxylation reactions involved in the metabolism of glucose, fatty-acids and amino-acids. Enzymatic carboxylation reactions within the cell are divided in four categories: the first category is formed by the three enzymes that carboxylate the substrate phosphoenolpyruvate (PEP) leading to the formation of oxaloacetate (PEP carboxylase, PEP carboxykinase and PEP carboxytransphosphorylase) while the second category contains the enzymes that require biotin as a co-factor.[27- 29] The third is represented by the enzyme ribulose-1,5-diphosphate carboxylase, a well characterised carboxylase in photosynthetic plants and microorganisms, and the fourth category constitutes the still poorly understood enzymes responsible for the carboxylation of glutamyl residues in glycoproteins

which require as a co-factor vitamin K, a fascinating coagulant discovered in the late 1920s by Danish scientist Henrik Dam.[30-32]

While carbon dioxide is the natural substrate for the enzymes PEP carboxykinase, PEP carboxytransphosphorylase, ribulose-l,5-disphosphate carboxylase, as well as the vitamin K-dependent carboxylases, all the biotin-dependant enzymes and PEP carboxylase use hydrogen carbonate and ATP as co-substrates.[33] At physiological pH and temperature, the concentration of dissolved carbon dioxide in equilibrium with the atmosphere is only 10 μM, whereas that of bicarbonate ion is 200 μM.[34] Therefore, although it is preferable for an enzyme to use the small proportion of electrophilic carbon dioxide, effective ways of fixing the more abundant bicarbonate occur in biotin-dependant enzymes.

There are six well-characterised biotin-dependant carboxylases which have in common a requirement for bicarbonate and ATP.[27; 28] Pyruvate (α-keto propanoic acid) carboxylase and urea carboxylase which lead to the formation of oxaloacetate and N-carboxyurea respectively. Other biotin enzymes include acetyl-CoA carboxylase, propionyl-CoA carboxylase, methylcrotonyl-CoA carboxylase and geranyl-CoA carboxylase. The specific substrates acetyl-CoA and propionyl-CoA are carboxylated on their α-carbon resulting in the formation of the essential metabolites malonyl-CoA and methylmalonyl-CoA (Fig. 1.2). Carboxylation of β-methylcrotonyl-CoA and geranyl-CoA occur on their γ-carbon which is activated by the double bond between the α- and β- carbons generating β-methylglucanoyl-CoA and isohexenyl-glutaconyl-CoA respectively.[27] In mammals, biotin serves in the carboxylation catalysed by pyruvate carboxylase, acetyl-CoA carboxylase (ACC),

propionyl-CoA carboxylase and β-methylcrotonyl-CoA carboxylase while in bacteria only ACC requires biotin as a co-factor.[35]

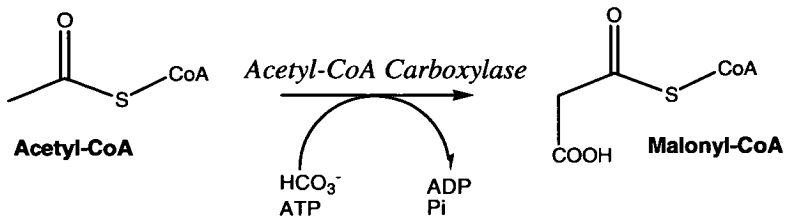


Figure 1.2 - The biotin-dependant carboxylation of acetyl-CoA. The reaction is catalysed by ACC in presence of bicarbonate and ATP and results in the formation of malonyl-CoA.

Biotin acts as a transporter molecule for carbon dioxide when covalently bound to the dependant carboxylases.[27; 28] The first step of the carboxylation reaction involves the ATP-dependent fixation of CO_2 derived from bicarbonate by biotin to form a carboxybiotin complex followed by decarboxylation of carboxybiotin and CO_2 transfer to different acceptor substrates.[33] In 1959, Lynen *et al.* showed that incubation of purified β-methylcrotonyl-CoA carboxylase in presence of ATP and radioactive bicarbonate generated labelled [^{14}C]carboxybiotin and the release of ADP and inorganic phosphate (Pi).[36] While he argued that the structure of biotin suggested that the bicarbonate could reversibly be incorporated at position $N_{3'}$, stabilisation of the labile [^{14}C]carboxybiotin by esterification with diazomethane showed that the resulting dimethyl ester form was identical to synthetic N_1-methoxycarbonylbiotin methylester. Further, limited proteolysis allowed the characterisation of the $N_{1'}$-carboxybiotin.[36] The localization of the carbon dioxide at the sterically less hindered $N_{1'}$ was proven by X-ray analysis of the bis(p-bromoanilide) derivative of carboxybiotin by Bonnemere *et al.* in 1965.[37]

Within the cell, the biotin prosthetic group is covalently linked *via* an amide bond with the valerate carboxyl group and the ε-amino group of specific lysine residues of the biotin dependant-enzymes, resulting in a flexible aliphatic chain, also called a "swinging arm" by some authors.[38; 39] This was first demonstrated by the isolation of N_ε-biotinyl-lysine, or biocytin, from an autolysate of growing yeast.[40] The flexible arm of approximatively 1.6 nm allows the translocation of biotin and carboxylated biotin between the different active sites of the carboxylase.[33; 39-43]

A biotin-dependant carboxylase which does not uses ATP or bicarbonate is the transcarboxylase methylmalonyl-CoA carboxyltransferase.[44; 45] Three decarboxylases, methylmalonyl-CoA decarboxylase, glutaconyl-CoA decarboxylase and oxaloacetate decarboxylase have also been shown to require biotin as a carboxyl carrier. In these three enzymes the CO_2 unit is transferred from either an acyl derivative or keto-ester to the enzyme before being transferred to either an acceptor or released as free CO_2. To date, biotin-dependant decarboxylases have been found only in anaerobic bacteria which employ the free energy of decarboxylation to pump Na^+ across a membrane. The resulting Na^+ gradient is then used for the synthesis of ATP.[46; 47]

1.1.3 The Biological Roles of Biotin

In addition to its role as a carboxyl carrier, the vitamin biotin has been shown to possess other functions within the cell.[48; 49] Recent characterisations of histones covalently bound to biotin suggest roles for the vitamin beyond its attachment to carboxylases and so far, at least seven biotinylation sites have been identified in

human histones.[50] Biotin has also been shown to be involved in the regulation of gene expression and signal transduction. Cell signaling molecules such as biotinyl-5'-AMP, Sp1 and Sp3, nuclear factor –κB and receptor tyrosine kinase depend on biotin supply and more than 2000 biotin-dependant genes have been found in various human tissues.[51; 52] The mechanisms involved in these other functions are still unclear but it seems that biotin may have been selected during evolution to play many different roles in the metabolism and gene expression of higher organisms.[51]

Biotin is only required in trace amounts and thus, deficiency in the nutritional diet is generally rare and mild and can be treated successfully by dietary supplements. Seborrheic dermatitis, a disease characterised by scaly skin changes, may be related to the poor ability of the patients to use biotin.[53] People with type 2 diabetes often have low levels of biotin and it has been suggested that the vitamin may be involved in the synthesis and release of insulin.[54; 55] Deficiency in the vitamin has also been proven teratogenic in animals and to be the cause of several neurological diseases.[56; 57] Biotin deficiency disorders also occur due to genetic lesions. Biotin-responsive multiple carboxylase deficiency (MCD) is a rare autosomal recessive disorder of biotin utilization associated with reduced activity of all four biotin-dependent carboxylases. The MCD disease may be caused by a primary deficiency of holocarboxylase synthetase (HCS), required for the biotin-dependent activation of apo-carboxylases, or of biotinidase, which functions in the recovery of biotin from degraded carboxylases. MCD, which can be reversed successfully by treatment with additional pharmacological dose of biotin, is characterized by metabolic ketoacidosis and organic acidemia accompanied by

developmental delay, hypotonia, and erythematous rashes followed by coma and death if left untreated.[49]

Nowadays, biotin is commonly used in biotechnology due to its extremely high affinity for the proteins avidin (from egg white) and streptavidin (from *Streptomyces avidinii*) which are among the strongest protein-ligand associations characterised ($K_D=10^{-15}$ M and $K_D=10^{-13}$ M respectively).[58; 59] Therefore, biotin is used in a wide range of biochemical applications. The most common remain the purification and the detection of biotin-tagged biomolecules from complex mixtures by avidin / streptavidin chromatography and the use of streptavidin-biotin complexes in western-blots.[35; 60] Biotinylated antibodies are captured by avidin or streptavidin in both the ELISPOT and ELISA techniques for the study of protein-protein interactions and related biochemical phenomena.[61]

1.2 Acetyl-CoA Carboxylase

1.2.1 The Biological Roles of Acetyl-CoA Carboxylase

Acetyl-CoA carboxylase catalyses the first committed and rate-determining step of fatty acid biosynthesis which is the biotin-dependant conversion of acetyl-CoA to malonyl-CoA,.[62; 63] The malonyl-CoA product of the ACC-catalysed reaction is used as a building block in the cyclic stepwise condensation reactions catalysed by fatty acid synthase (FAS) and is the source of the two carbons units employed in the elongation of fatty acids.[64-66] Fatty acids are important constituents of biological membranes, energy storage compounds and messenger substances and, therefore, ACC is essential in many organisms.[67] In addition to fatty acid production, the

malonyl-CoA produced by bacterial ACCs is also the building block for a variety of natural products including the vitamins biotin and lipoic acid as well as polyketide antibiotics.[68-70] Moreover, malonyl-CoA has been shown recently to play a major role in regulating the expression of the transcription factor FapR involved in fatty acid and phospholipid metabolism.[71; 72]

In mammals, two isoforms of acetyl-CoA carboxylase, ACC1 and ACC2, encoded by different genes have been characterized.[73] ACC1 is a cytosolic enzyme which regulates fatty acid biosynthesis in lipogenic tissues, whereas ACC2 is mainly located in the heart and skeletal muscle and is associated with the mitochondrial membrane.[74] ACC2 controls the rate of lipid oxidation as its product malonyl-CoA is a potent inhibitor of carnitine palmitoyltransferase, an enzyme which allows the transfer of long chain acyl-CoA into the mitochondria.[75] Since ACC plays critical roles in the synthesis and regulation of fatty acid biosynthesis in humans, the enzyme has become an attractive target for drug discovery.[76] ACC has been validated for the development of novel therapeutics treatments against obesity, type 2 diabetes, and other manifestations of the metabolic syndrome.[77-79] However, only a few small molecule inhibitor of mammalian ACC are currently known. Animal testing has shown that compound CP-640186 (a Pfizer product) reduces tissue malonyl-CoA levels, inhibits fatty acid biosynthesis, stimulates fatty acid oxidation and reduces body fat and weight (Fig 1.3, A).[79; 80]

Bacterial ACCs also constitute an attractive target for the development of novel antibiotics. Until now, most inhibitors have been directed against the FAS-dependent enzyme complex and very few inhibitors of bacterial ACCs are known.[81; 82] Since the fatty-acid composition of bacteria varies significantly between species,

ACC, which is a broadly conserved enzyme among prokaryotes and, with the exceptions of some actinobacterias, does not possess any other isoform, has become a new antibiotic target area. Pyrrolidinedione derivatives, particularly moiramide B, have been shown recently to inhibit the bacterial ACC while having little effect on eukaryotic enzyme activity (Fig 1.3, B).[83; 84] In plants, ACC is also a proven target for drug action. Two classes of compounds, aryloxyphenoxypropionate (FOP) and cyclohexanedione (DIM), which kill sensitive plants by inhibiting their fatty acid biosynthesis, have been used commercially as herbicides for more than 20 years (Fig 1.3, C). These herbicides are reversible inhibitors of grass plastid ACC and weak inhibitors of mammalian and yeast ACCs [85-87] Soraphen A, a macrocyclic polyketide natural product secreted by the soil-dwelling myxobacterium, *Sorangium cellulosum*, is a potent inhibitor of eukaryotic ACC and has been tested for agricultural applications (Fig 1.3, D).[88]

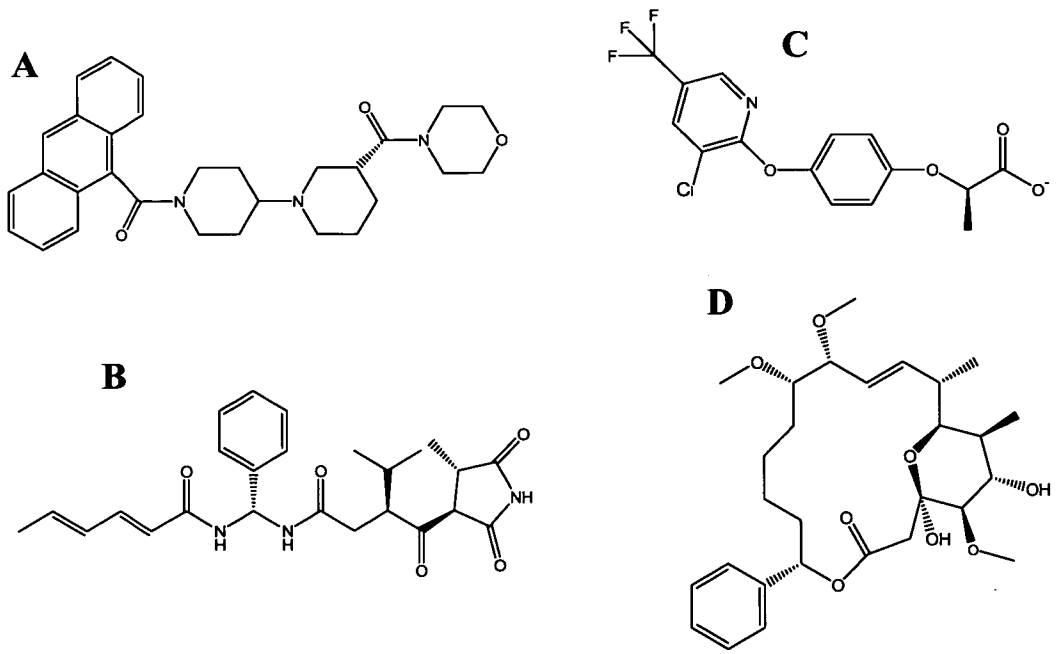


Figure 1.3 - Chemical structures of representative ACC inhibitors. A: CP-640186. **B**: Moiramide B. **C**: Haloxyfop (FOP compound). **D**: Soraphen A.

1.2.2 The Biological Properties of Acetyl-CoA Carboxylase

Two physically distinct types of ACC are found in nature.[64] The first ACCs studied were those from mammals and yeast in which the reaction is catalysed by a single large multifunctional polypeptide forming a homodimer. In contrast, bacterial ACCs are heterodimeric multi-subunit proteins.[63; 89] *E. coli* acetyl-CoA carboxylase constitutes a paradigm of multi-subunit ACC and is constituted by four subunits: the biotin carboxyl carrier protein (BCCP), biotin carboxylase (BC) and carboxyltransferase (CTα and CTβ) (Table 1.1).[90] The predicted oligomeric states of the BC, CT and BCCP subunits would form an enzyme with an overall mass of 200 kDa. Isolation of such complex has not yet been achieved and the exact stoichiometry of the subunits within the ACC complex remains unclear. Until recently, the overall activity of *E. coli* ACC could be measured only when all four subunits were present at high concentrations, although both partial reactions catalysed by the subunits BC and CT α/β can be measured in dilute protein solutions.[91-93] In 2006, Soriano *et al.* established a robust assay which monitors the overall ACC activity by measuring the phosphate production at physiologically relevant concentrations of the four subunits.[94]

Subunit	Encoding gene	MW (kDa)	Solution Form
Biotin Carboxyl Carrier Protein (BCCP)	*accB*	16.7	Tetramer
Biotin Carboxylase (BC)	*accC*	49.4	Dimer
Carboxyltransferase, α subunit (CTα)	*accA*	35.1	Dimer plus dimer of CTβ
Carboxyltransferase, β subunit (CTβ)	*accD*	33.2	Dimer plus dimer of CTα

Table 1.1: The four subunit constituents of *E. coli* ACC.

Although the four subunits of *E. coli* ACC were purified to homogeneity more than 50 years ago, the four genes *accB*, *accC* and *accA/D* of acetyl-CoA carboxylase encoding the different subunits were discovered only in the early 1990s (Table 1.1).[92] The *accB* gene which encodes the biotin carboxyl carrier protein had been described previously by groups working on other projects and Li and Cronan characterized the gene by a reverse genetic approach.[95-97] Cronan subsequently showed that the BC ORF (a*ccC*) was located immediately downstream of *accB*.[98] In the same period, Kondo *et al.* independently isolated the *accA* gene encoding CT-α by reverse genetics leading directly to the identification of the *accD* gene encoding the β-subunit.[99; 100] Citrate, which promotes polymerization of ACC, and long-chains acyl-CoAs, which promote the dissociation of the ACC polymers when bound to the enzyme, together regulate the enzymatic activity of mammalian acetyl-CoA carboxylase.[76]

The identification of the *E. coli acc* genes has greatly facilitated the study of the ACC proteins from different organisms. Until their discovery, a long-standing controversy had remained in the field of plant lipid biosynthesis and contradictory studies reported the presence of multi-functional as well as multi-subunit ACC proteins.[101; 102] It has now been established that the multi-subunit ACCs of plants are expressed in the plastid and are used in *de novo* fatty acid biosynthesis while multi-functional ACC are expressed in the plant cytosol.[103] The multi-functional enzyme is required for the biosynthesis of several secondary metabolites including flavonoids and anthocyanins as well as the formation of long acyl chains which are essential components of plant waxes. CT-β is mostly encoded by the plastid genome while the rest of the subunits are encoded by the nuclear genome and imported into the plastid.

However, the locations of the genes can vary within lineages reflecting the complexity of lipid biosynthesis in plants.[104]

1.2.3 The Biotin-Dependant Carboxylation Catalysed by ACC

In the functional ACC complex, biotin acts as a carboxyl carrier in a two-step carboxylation reaction of acetyl-CoA (Fig. 1.4).[33; 38; 105] The biotin prosthetic group is covalently linked *via* an amide linkage between its carboxyl group and a specific lysine residue on the biotin carboxyl carrier protein BCCP. Biotin carboxylase catalyses the first step of the ACC reaction which is the carboxylation of biotinylated BCCP (holo-BCCP) at the $N_{1'}$ atom of the biotin co-factor in the presence of bicarbonate and ATP. By means of the flexible arm, the biotin and subsequent carboxybiotin groups are presumably able to move between the different active sites of biotin carboxylase and carboxyltransferase.[39]

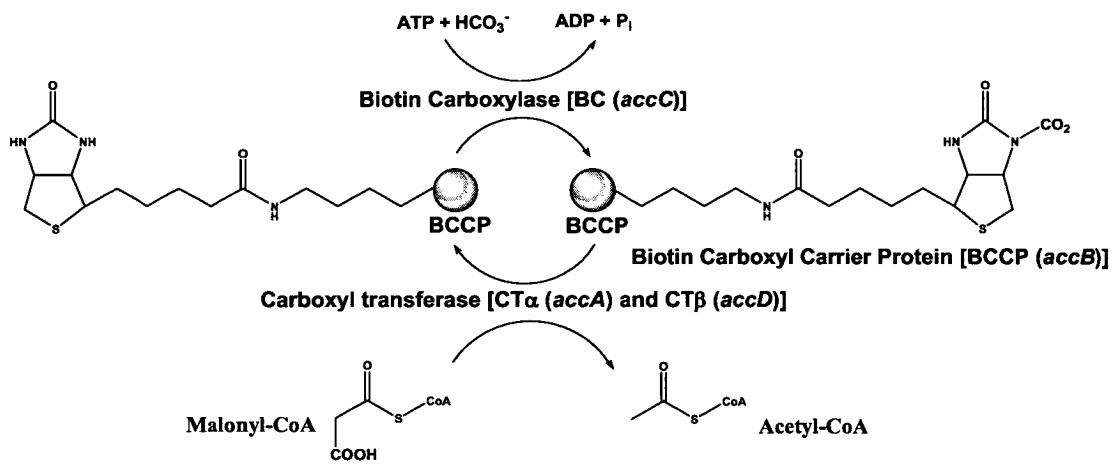


Figure 1.4 - The two-step reaction mechanism of ACC. The biotin prosthetic group is covalently bound to BCCP and acts as a carrier of CO_2 by shuttling the carboxylate moiety between the BC and CT subunits.

The last step of the ACC reaction is catalysed by carboxyltransferase and involves the transfer of the carboxyl group from carboxybiotin to the substrate acetyl-CoA to form malonyl-CoA (Fig. 1.4). Therefore, ACC catalyses the formation of a new carbon-carbon bond and the energy required for this new linkage is derived from the hydrolysis of ATP.[33] This two-step carboxylation reaction mechanism and the structural organization of the ACC multi-enzyme into distinct BCCP, BC and CT components is shared among the other members of the biotin-dependent enzyme family, including pyruvate carboxylase, propionyl-CoA carboxylase, methylcrotonyl-CoA carboxylase, transcarboxylase and glutaconyl-CoA decarboxylase.[45]

1.2.4 The Biotin Carboxyl Carrier Protein BCCP

1.2.4.1 Intact and Truncated Forms of BCCP

The BCCP domain of acetyl-CoA carboxylase was first characterized by Albert and Vagelos who, by treatment of cell-free extract of *E. coli* ACC with avidin, demonstrated that a biotinylated protein, subsequently called biotin carboxyl carrier protein (BCCP), was involved in the ACC catalysed-reaction.[38] However, protein chemistry quickly showed that the initial BCCP preparations were heterogeneous due to cleavage by an unknown protease. Amino-acid sequence analyses indicated that the preparations consisted of a short C-terminal protease-resistant biotinyl domain as well as N-terminal segments of various lengths. Further purification to homogeneity of the full-length 156 amino acids BCCP showed that the protein aggregated and formed large oligomers.[106-108] Subtilisin cleavage was proven to convert the recalcitrant BCCP in a more stable form and consequently most studies have been carried out using an 87 amino acid C-terminal fragment containing the biotinylated

lysine (BCCP87).[108-110] *In vivo* activities with [^{3}H]biotin performed on C-terminal fragments of BCCP expressed as fusion proteins with β-galactosidase showed that only 80 amino acids are necessary for biotinylation and sequence analysis of the truncated BCCPs indicated that the prosthetic group is covalently linked to the lys122 of a highly conserved MKM motif.[97; 111] A high degree of sequence homology is also observed among other biotinyl domains and studies carried out by Cronan *et al.* on the biotin carrier subunit 1.3S of the transcarboxylase from *Propionibacterium shermanii* indicated that only a 75 amino acid sequence is required for biotinylation.[35]

Those results suggest that the N-terminal domain is somehow independent of the rest of the BCCP protein. The N-terminus of BCCP, rich in alanine and proline residues, is thought to act as flexible linker and to interact with the other subunits of ACC.[97] The structure of this upstream linker segment could not be determined due to the high degree of mobility and only the structure of the C-terminal 72 amino acids of the BCCP-like *P. shermanii* 1.3S transcarboxylase subunit has been solved by NMR spectroscopy.[112; 113] Although the biotinyl domain BCCP87 is not physiologically relevant, the short form of BCCP provides a valuable tool for structural analysis and studies on the mechanism of biotinylation of BCCP. Apo-BCCP87 is readily modified by sulfhydryl reagents and is thought to be a disulfide-linked dimer in solution. In contrast, the biotinylated form of BCCP87 (holo-BCCP87) remains monomeric, indicating conformational changes within the biotinyl domain which alter the reactivity of the sole Cys116 upon biotin attachment.[114]

1.2.4.2 Structure of Truncated Apo- and Holo-BCCP

The first structure of the biotinyl domain was determined by X-ray crystallography at a resolution of 1.8 Å by Athappilly *et al.* for the C-terminal 80 amino acids of biotinylated *E. coli* BCCP (BCCP80).[115] Despite low sequence conservation, the biotinyl domain of BCCP shows striking structural similarities with the lipoyl binding domains of pyruvate dehydrogenase.[116] The crystal structure of holo-BCCP80 consists of two sets of antiparallel β-sheets with a 2-fold axis of quasi-symmetry that form a flattened β-barrel with the N- and C-termini close together on one side (Fig. 1.5, b).[115] The biotin cofactor is located at the other end of the molecule and is covalently attached to the side chain of Lys122 exposed on a tight β-turn which connects the two β-sheets. Biotinylated BCCP80 is stabilized by a core of hydrophobic residues in which the reactive Cys116 is buried. The ureido ring of the biotin moiety interacts with a non-symmetric "thumb-like" protrusion segment comprising residues 94-101.

Further heteronuclear multidimensional NMR studies of the two forms, apo- and holo-, of BCCP87 have confirmed the symmetrical nature of BCCP80 (Fig. 1.5, a).[117; 118] The core of apo-BCCP87 is shown to have a less favourable side-chain packing than holo-BCCP87 and this change in dynamics explains the reactivity of Cys116 in the apo-form towards thiol reagents as well as the susceptibility of apo-BCCP87 to limited proteolysis.[114; 119] In the initial crystal structure, Thr94 was reported to form hydrogen bonds with the carbonyl and $N_{1'}$ protons of the biotin uriedo ring. Free biotin acts as a poorer substrate than covalently attached biotin in the BC-catalysed carboxylation reaction suggesting that the biotin $N_{1'}$ atom of holo-BCCP87 is freely accessible.[120] NMR studies have also shown that the rates of

deuterium exchange of the uriedo N_1, and N_3, amide protons of BCCP87 are similar (<5 min) and comparable to those determined for free biotin.[118]

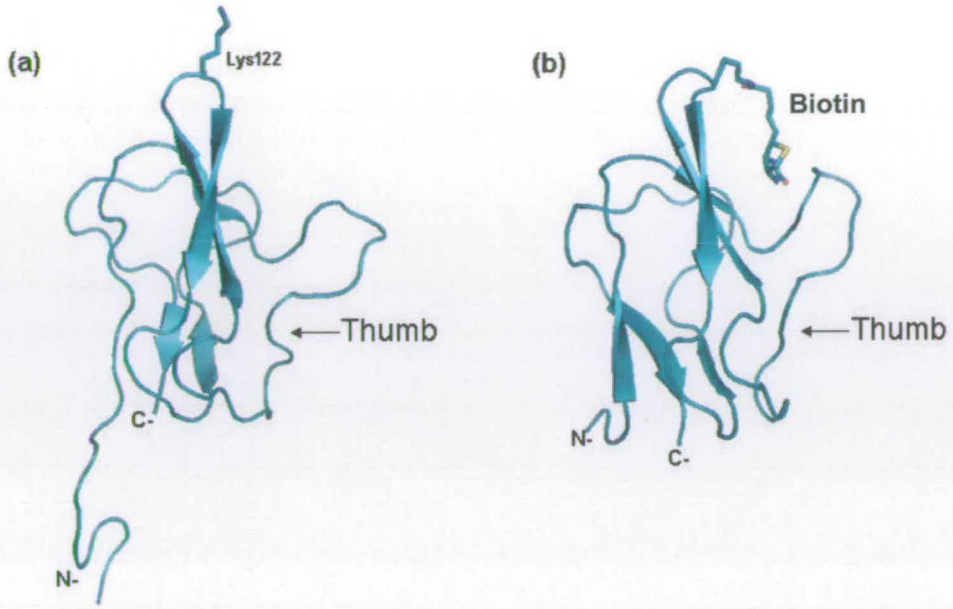


Figure 1.5 - The structures of *E. coli (a)* apo-BCCP87 and *(b)* holo-BCCP80. The NMR structure of *E. coli* apo-BCCP87 and X-ray structure of holo-BCCP80 reveal that the biotinyl-domain consists of a β-barrel with the biotin moiety, drawn in sticks with blue carbons, exposed on a tight β-turn and interacting with the characteristic thumb region. (PBD codes 1A6X and 1BDO).

Although the NMR data argue against strong hydrogen bonds linking the biotin moiety with Thr94, other NMR results indicate that the biotin rings do interact weakly with residues located on the thumb.[118; 119] The amide protons of residues Thr94 and Ser96 show slower solvent exchange rates in the biotinylated form and the proton chemical shifts of the thumb residues differ in apo- and holo-BCCP87. The thumb of holo-BCCP87 is less mobile than the thumb of the apo-form due mainly to biotin–thumb interactions. Analyses of mutants lacking the thumb region indicated an increased sensitivity to proteolysis even in the presence of bound biotin.[121] The structure of holo-BCCP87 is a dynamic picture in which the biotin rings are not

completely immobilized and therefore, although the interactions have not been fully characterized as they cannot be restricted to a specific geometry, virtually, all of the residues from the thumb appear to interact with the relatively flexible biotin moiety.[118; 119]

Despite high sequence similarities between the biotinyl domains of biotin-dependent carboxylases and decarboxylases, the thumb insertion is only found among ACC enzymes of bacteria and plant plastids that synthesize fatty acids.[33] The thumb was proven to be essential for the ACC reaction as expression of two different "thumbless" BCCP mutants failed to complement growth of a temperature-sensitive *E. coli accB* strain.[122] However, thumbless BCCP mutants expressed in *E. coli* were also shown to be efficiently biotinylated and there is no evidence that the protruding region is involved in biotinylation.[60] Therefore, without having an essential role in either protein biotinylation or protein stability, the thumb was suggested to act as a mobile lid for either the biotin carboxylase or carboxyltransferase active sites.[121; 122] Single-site mutations within the thumb region of *E. coli* BCCP have been shown to have strong effects on the functions of BCCP and ACC indicating that the rather mobile protruding segment must have a defined structure to be effective.[122]

1.2.5 Biotin Carboxylase (BC)

1.2.5.1 BC Mechanism

The first half of the ACC reaction is catalysed by biotin carboxylase (BC) and constitutes the most interesting and still incompletely understood mechanistic aspect of the overall ACC reaction. Biotin carboxylase catalyses the carboxylation of the biotin prosthetic group covalently attached to BCCP at the $N_{1'}$ position in a two-step

reaction.[33; 91; 105] During the initial step, BC catalyses the dehydratation of the poor electrophilic HCO_3^- by using ATP. Nucleophilic attack of HCO_3^- on the γ-phosphate of ATP results in the formation of carboxyphosphate and release of ADP (Fig. 1.6).[123; 124] Bicarbonate is the source of CO_2 for all biotin-dependant carboxylase and BC dehydrates the poor electrophilic HCO_3^- by using ATP.[33] The intermediate carboxyphosphate has not been trapped but tracer experiments showed that when labelled [^{18}O]bicarbonate was used as substrate in the propionyl-CoA carboxylase catalysed reaction, two ^{18}O atoms ended up in the new carboxyl group of methylmalonyl-CoA while the third was found in the released inorganic phosphate moiety.[123; 125] Moreover, BC is able to catalyse, in a reverse reaction, the transfer of a phospho group to ADP from carbamoyl phosphate, an analogue of carboxyphosphate (Fig. 1.6).[124] A slow bicarbonate-dependant ATPase activity has also been observed in ACC in absence of biotin and dramatically increased in the presence of biotin or holo-BCCP.[120; 126] Therefore, hydrolysis of ATP has been suggested to be synergistic with biotin binding.[127; 128] It appears that the role of biotin is to mediate the cleavage of the co-substrate ATP in an exergonic reaction.[33]

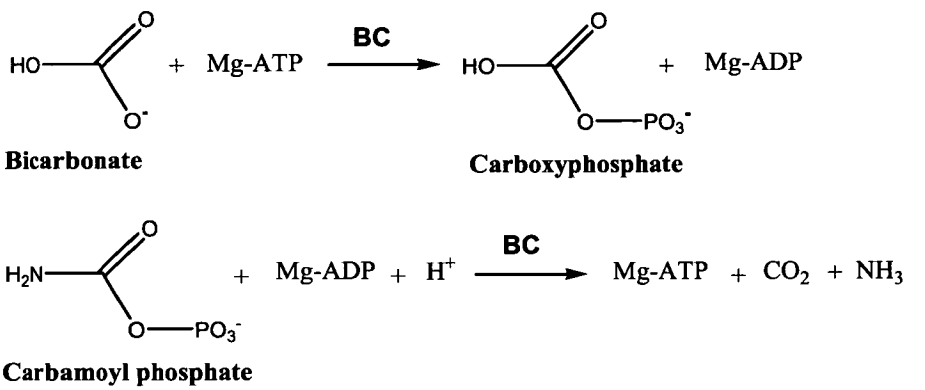


Figure 1.6 - Reactions catalysed by BC. Formation of carboxyphosphate from hydrogen carbonate and ATP and transfer of inorganic phosphate from carbamoyl phosphate to MgADP catalysed by BC.

The second BC-catalysed reaction involves the carboxylation of the biotin co-factor from carboxyphosphate.[33] Several mechanistic pathways have been proposed which seek in different ways to overcome the fact that biotin is a poor nucleophile and bicarbonate a poor electrophile. Each of these pathways accommodates the important finding that one of the three bicarbonate oxygen atoms ends up in the inorganic phosphate product.[123] One possible mechanistic pathway involves the transfer of the γ-phosphate group of ATP onto the ureido oxygen of biotin, leading to the formation of an enol phosphate. Cleavage of the phosphoanhydride bond by attack of HCO_3^- to form carboxyphosphate would simultaneously create a nucleophilic centre at the $N_{1'}$ atom of biotin.[129; 130] However, this mechanism remain in contradiction with the ATPase activity observed in the absence of biotin.[126] Another alternative remains the elimination of Pi from the intermediate carboxyphosphate and the release of the more electrophilic carbon dioxide.[125] The eliminated basic anion PO_4^{3-} could remove the proton from the $N_{1'}$ atom of biotin, generating a biotin ureido anion which reacts with CO_2.[131] However, no bicarbonate-dependent ATP/ADP isotopic exchange in the absence of biotin could be observed.[33]

Analyses carried out with carbamoylphosphate synthetase have shown that the rate of ADP production is equal to the rate of proton release which is consistent with the notion that there is direct attack on carbamoyl phosphate.[132] Other enzymes of the ATP-grasp superfamily such as D-Ala:D-Ala ligase and glutathione synthetase which catalyse similar bicarbonate-dependent ATPase reactions also indicate a model in which there is direct attack on the carbon of the acyl-phosphate intermediate.[90] BC could presumably first catalyses the dehydrogenation of the $N_{1'}$ hydrogen of biotin to form an ureido anion and the simplest mechanism remains the nucleophilic attack of

the $N_{1'}$ atom of biotin onto the carboxyl group of the carboxyphosphate.[133] This mechanism is supported by steady-state kinetics of BC carried out by Tipton and Cleland which showed that acid-base reactions are involved in the tautomerization of biotin which enhances the nucleophilicity of $N_{1'}$ toward the poor electrophilic bicarbonate.[124]

1.2.5.2 BC Structure

The crystal structures of the *E. coli* BC subunit and its mutant Glu288Lys in complex with ATP have been determined by X-ray crystallography.[134; 135] BC shows high structural homology to carbamoylphosphate synthetase, D-Ala:D-Ala ligase and glutathione synthetase and both the structural and mechanistic similarities of these enzymes indicate that they belong to the ATP-grasp fold superfamily.[136] The structure of the free BC subunit is cylindrical and consists of three distinct domains. The N-terminal (A-domain) forms a dinucleotide binding motif consisting of a parallel β-sheet flanked on either side by α-helices (Fig 1.7). The B-domain, formed by two α-helices and three β-strands, extends from the main body of BC and the two domains are connected with an AB-linker region. The C-terminal forms the C-domain and contains an eight-stranded and a smaller four-stranded antiparallel β-sheet and seven α-helices. Together with the AB linker, the A and C domains form the cylindrical core of the enzyme.[134]

The bacterial BC subunit is a dimer in solution and is also in the crystal structure.[92; 134] In the structure, the monomers are associated through their C-domains and the active sites are located away from the dimer interface, at the top of the cylinder between the B and C domains (Fig 1.7). In the mutant BC complexed with

ATP, the adenine ring and the phosphates are stabilised *via* hydrogen bonds with the two domains as well as to the AB-linker.[135] The B-domain acts as a lid to the active site and is shown to undergo large conformational changes upon ATP binding. In the structure of unbound BC, the B-domain is more disordered and positioned away from the core in an open conformation. In the structure of the complex, the domain rotates by approximately 45° in order to cover the active site.[134] Modelling studies of BC have revealed the relative orientation of the two ligands within the active site and show that the γ-phosphate of ATP lies adjacent to the $N_{1'}$ atom of biotin.[135]

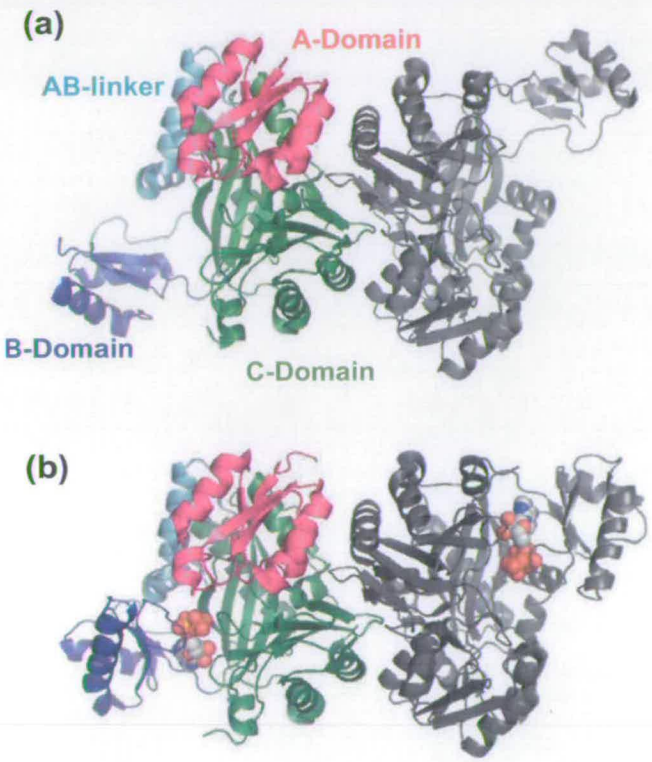


Figure 1.7 - Crystal structures of *(a) E. coli* BC and *(b)* mutant BC E288K complexed with ATP. (a): Apo-biotin carboxyalse (PBD code 1DV1). BC is dimeric and each monomer contains 3 domains. One monomer is shaded grey, while the second monomer is coloured to highlight the 3 distinct domains. The N-terminal (A-domain) is shaded green, the B-domain is shaded dark blue, the AB-linker is shaded cyan, and the C-terminal (C-domain) is shaded pink. **(b):** BC mutant (E288K) in complex with ATP (PBD code 1DV2). ATP is drawn in spheres with white carbons. The B-domain has rotated by approximately 45 ° to cover the ATP-bound active site.

More recently the monomeric structure of the eukaryotic yeast BC domain was solved, both in the unbound state and in complex with the potent natural product inhibitor soraphen A (***See figure 1.3, D***).[137] The overall structures of yeast and *E. coli* BC are similar although, interestingly, the B-domain of yeast biotin carboxylase adopts a closed conformation even in absence of ATP in the active site.[137] While the active site of BC is located at the top of the cylinder formed by the A and C domains, Soraphen A is bound in an allosteric site at the bottom of the cylinder. Structural comparison with the bacterial BC dimer suggests that soraphen A may be located in the interface of the putative dimer of yeast BC.[138]

1.2.6 Carboxyltransferase (CT)

1.2.6.1 CT Mechanism

In contrast to biotin carboxylase BC, little is known about the catalytic mechanism of carboxyltransferase (CT). It could be presumed that, once formed, the carboxybiotin head group could swing to the carboxyltransferase active site where transfer of the carboxyl group to form malonyl-CoA takes place.[90; 91; 105] The recombinant forms of the CTα/β enzymes have been subjected to rigorous steady-state kinetic analysis and the activity of CT was assayed in the non-physiological direction, in which malonyl-CoA reacts with biotin to form acetyl-CoA and carboxybiotin.[139] Initial velocity patterns as well as inhibition studies have revealed an ordered kinetic mechanism of carboxyltransferase with malonyl-CoA binding before biocytin (ε-biotinyl lysine).[140] As is the case for pyruvate carboxylase, carboxylation of the thioester substrates in ACC by the carboxyltransferase subunit requires abstraction of a proton from the methyl group to form an enolate anion

intermediate inducing the subsequent nucleophilic attack on the carbonyl carbon of carboxybiotin (Fig 1.8).[141]

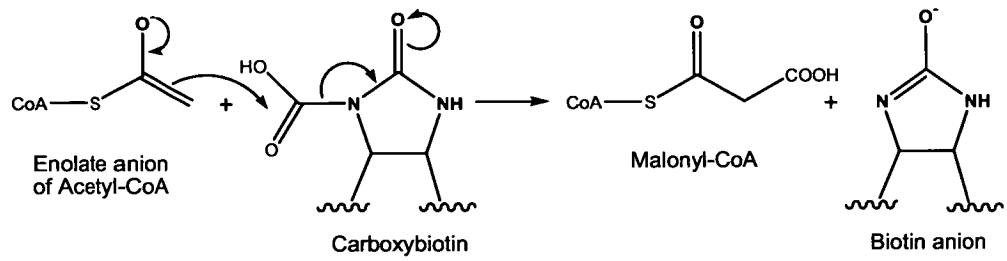


Figure 1.8 – Carboxylation of acetyl-CoA catalysed by CT. Nucleophilic attack of the enolate anion of acetyl-CoA on the carbonyl of carboxybiotin results in the formation of malonyl-CoA.

The chemical mechanism of carboxyltransferase was originally probed by determining the pH rate profile of the reverse reaction.[139] Acid-base catalysis is required in this process because the proton on the $N_{1'}$ atom of biotin must be removed in order to allow the carboxyl transfer from malonyl-CoA. The study demonstrated that a single ionising group on the enzyme, with a pKa of 7.5, must be unprotonated for catalysis to occur. It was postulated that this group was a cysteine residue which acts as the base to remove the proton from the $N_{1'}$ of biotin since CT was found to be susceptible to inactivation by the sulfhydryl-modifying reagent N-ethylmaleimide.[139]

Recent structural evidence and mutational analysis of the active site have revealed that no residue acts as a base in the catalytic mechanism of CT.[87; 142] Therefore, it has been suggested that like in propionyl-CoA carboxylase (PCC), the $N_{1'}$ atom of biotin itself functions as the general base.[33; 87] This is in direct contrast with the crotonase superfamily of enzymes which are characterized by reactions of CoA thioesters in which the enolate intermediate is stabilized by polar interactions

with basic side chain residues and the oxyanion and where an acidic side chain in the enzyme is required for catalysis.[143; 144] Carboxyl transfer from carboxybiotin to acetyl-CoA can occur *via* either a stepwise mechanism in which proton abstraction from the methyl group of acetyl-CoA occurs before carboxyl transfer or *via* a concerted mechanism in which proton removal is concurrent with carboxyl transfer. Studies to distinguish between these two possibilities have not been carried out in ACC, although a stepwise mechanism of carboxyl transfer has been proposed for the CT subunit of pyruvate carboxylase.[145]

1.2.6.2 CT Structure

The crystal structures of the bacterial CT subunits from *E. coli* and *Staphylococcus aureus* were determined last year at 3.2 and 2.0 Å resolution respectively.[142] The overall structures of the two bacterial CT subunits are very similar and have the $\alpha_2\beta_2$ heterotetrameric assembly proposed on the basis of gel filtration and sedimentation equilibrium analysis experiments with the *E. coli* enzyme.[92] The tetramer has a rectangular shape and the dimer of dimers is related by a 2-fold axis running through a central cavity which gradually narrows from an opening of 13 Å diameter to a solvent-inaccessible surface within a 23 Å depth (Fig 1.9, a).[142] The wedge-shaped monomers CTα and CTβ are structurally very homologous and the active site is located at the αβ interface suggesting duplication and divergence of a single ancestral carboxyltransferase subunit gene. CTα and CTβ both possess a spiral core composed of a long and twisted 7-stranded mixed β-sheet flanked by helical regions and orthogonal to a short 2-stranded parallel β-platform as well as a 3-helix bundle located at the edge of the rectangular tetramer.[142] A

surprising and novel feature of the bacterial CT subunits is the presence of a Zn^{2+} binding domain composed of a tetrahedral cluster of cysteine residues that terminate the CTβ dimers.[142]

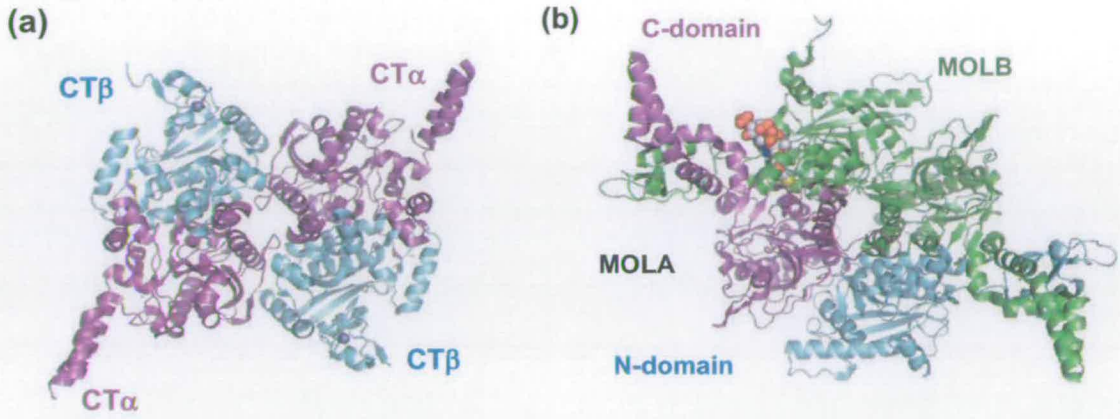


Figure 1.9 - Crystal structures of *(a)* the *S. aureus* CT subunit and *(b)* the yeast CT domain in complex with HSCoA of ACC. (a): The *Sa*CT $\alpha_2\beta_2$ heterotetramer has a rectangular shape. The α and β monomers are coloured in magenta and cyan respectively. **(b):** each monomer of the yeast CT dimer is divided into two distinct subdomains. One CT monomer is shaded in green, while the other monomer is shaded magenta and cyan to highlight the N-terminal (CTβ -cyan) and C-terminal (CTα - magenta) subdomains. The binding pocket is highlighted by the presence of HSCoA drawn in spheres with white carbons (PBD codes 2F9I and 1OD2).

This dimeric structural arrangement as well as the β-β-α superhelix fold is also observed in the structures of the CT domain of yeast ACC solved by Tong *et al.* (Fig. 1.9, b).[87; 146; 147] Yeast CT is a dimer and each monomer contains two subdomains, the N- and C- domains, which are equivalent to the CTβ and CTα subunits of bacterial carboxyltransferases respectively. The two monomers are arranged in a head-to-tail fashion in the dimer, such that the N-domain of one monomer is in contact with the C-domain of the other monomer.[87] The β-β-α superhelix fold of bacterial and yeast carboxyl transferases is characteristic of the crotonase/ClpP superfamily and is found in the structures of the CT subunit of

glutaconyl-CoA decarboxylase and the β subunit of propionyl-CoA carboxylase (PCC).[144; 148; 149]

Comparison of the complexes of the CT domains from yeast ACC with HSCoA and the CTβ subunit from *Streptomyces coelicolor* PCC with propionyl-CoA and biotin have allowed the modelling of the active site of the CT subunit of bacterial ACCs.[142; 149] The acyl-CoA and the biotin binding pocket are both located at the interface between the CTα of one monomer and the CTβ of its dimeric partner. In the structure of yeast CT in complex with HSCoA, most of the ligand is associated with the N-domain while the thiol group of CoA is located at the dimer interface suggesting that the biotin carboxyl approaches the active site from the C-domain of the other monomer.[87] This is supported by the fact that biotin binds predominantly to the alternate domain from the dimeric partner of the acyl-CoA domain in PCC and both ligands interact at the dimer interface.[149]

The structure of the CT domain of yeast ACC has also been determined in complex with two FOP herbicides (haloxyfop and diclofop) and the inhibitor CP-640186 (***See figure 1.3, A & C***).[146; 147] The FOP compounds are bound at the dimer interface in a pocket near the active site region and kinetic studies indicated that haloxyfop is a competitive inhibitor of malonyl-CoA for yeast CT.[87; 150] The inhibitor of mammalian ACC CP-640186 is a moderate inhibitor of the yeast CT domain and more potent against the CT domain of human ACC1.[76] However, the CT domains of yeast ACC and human ACC1 are both dimers in solution and thus, the yeast structure has been suggested as a good model to study the human enzyme. CP-640186 is thought to bind in the putative biotin-binding site and is a non-competitive inhibitor with acyl-CoA substrates for yeast CT.[147; 151] Interestingly, there is little structural

overlap between the bound HSCoA, haloxyfop and CP-640186, suggesting the presence of several distinct regions for inhibitor binding in the active site of the CT domain.

1.2.7 Towards a Working Model of ACC

E. coli multi-subunit ACC has been shown to be a markedly unstable enzyme which readily dissociates into complexes and sub-complexes.[63; 91] Therefore, despite extensive studies on the individual subunits of ACC, their exact stoichiometry and respective arrangement within the ACC complex remain unclear. In multi-subunits and multi-functional enzymes, the metabolites must be efficiently transferred between the active sites through tunnels and channels or through the use of prosthetic groups in order to avoid the energetic penalty associated with their loss to diffusion, degradation, or competing side-reactions.[152] The biotin prosthetic group is attached to a carrier domain and in *E. coli* ACC, the lack of interaction of isolated biotinyl domains BCCPs *in vitro* suggests a model similar to that of another swinging domain enzyme, pyruvate dehydrogenase.[39; 153] The biotinyl and lipoyl binding domains of these enzymes are structurally highly homologous and mutational analyses have shown that modified BCCP were able to switch substrate specificity and be efficiently lipoylated.[116; 154] In lipoyl domains, the sequence rich in proline and alanine of the N-terminal linker region results in a flexible structure and allows the co-factor to move freely between the active sites of pyruvate dehydrogenase in a random fashion.[42; 111]

Deletion of 30 linker residues located adjacent to the biotinyl domain resulted in inactive BCCP species *in vivo* although it was efficiently biotinylated.[155] Expression of this BCCP species failed to restore normal growth and fatty acid synthesis to an *E. coli accB* mutant strain which encodes a temperature-sensitive BCCP (BCCPG133S) that rapidly degrades at 37 °C. Replacement of the deleted BCCP linker with a linker derived from *E. coli* pyruvate dehydrogenase gave a chimeric BCCP species shown to restore activity *in vivo* and this may suggests that these two carrier proteins share a common evolutionary origin.[155] Expression of BCCPs having deletions of various segments of the linker region showed that the inactive BCCPs all lacked an APAAAAA sequence located adjacent to the tightly folded biotinyl domain, whereas deletions that removed only upstream linker sequences had little or no effect on the activity of the proteins.[116; 155]

The full-length BCCP has been shown to be a recalcitrant protein due to its strong tendency to aggregate, however, the exact stoichiometry of the C-terminal biotinyl domain (BCCP87) missing the linker region also remains unclear.[108; 109] Sedimentation equilibrium experiments have indicated that BCCP87 is monomeric even at the high concentrations used in the NMR analyses.[108; 117; 118] Apo-BCCP87 was also shown to form disulfide linked dimers, although, based on the NMR data, the proposed BCCP pairing interactions appear relatively weak.[114] Furthermore, expression of a mutant BCCP lacking the biotinylated lysine residue was able to partially restore growth and fatty-acid synthesis to the temperature-sensitive *E. coli accB* strain.[122] It was proposed that the properly folded, albeit inactive, mutant stabilized the biotinylated temperature-sensitive unfolded BCCP G133S implying

that dimerization of BCCP is required for the ACC reaction and that the biotinyl domains are located close to one another.[122]

The *accB* and *accC* genes encoding BCCP and BC respectively lie in the same operon in the genomes of very diverse bacteria and have a different pattern of regulation than that of the unlinked *accA* and *accD* genes.[98] Co-expression of the *E. coli* BCCP and BC proteins from the *accBC* operon has been previously reported but the stoichiometry of the complex was not determined.[107; 108] Recent studies of the *E. coli* BC:BCCP complex by Cronan *et al.* have revealed that the ratio of BC per BCCP molecule is 1:2.[156] Although the active sites of the *E. coli* BC dimer are located on the monomer faces opposite those used in the dimer interface and are therefore relatively far apart (62 Å), dimerization of BC was initially thought to be required for activity.[134; 135; 157] Hybrid BC dimers in which one monomer was native and the other carried an inactivating mutation were found to be catalytically inactive.[158] Despite recent evidence that monomeric *E. coli* BC mutants retain significant catalytic activity, it is still believed that there is communication between the active sites of the BC dimer.[159] A model has been proposed in which the two subunits of BC are mechanistically linked in a fixed cycle and can not catalyse carboxylation simultaneously.[158] The two subunits alternate catalytic reactions, with one site performing catalysis while the second is releasing product and this requires the two BCCP molecules to be well separated from each other, as each would interact with a different BC active site. The flat and hydrophobic bottom formed by the N- and C- termini has been suggested as a possible interface for the BCCP dimer which would position the two biotin moieties about 60 Å apart, a spacing similar to that of the BC active sites.[122]

The complex BC_2:$BCCP_4$ should be aligned in order to interact with the active sites of the CT component and multi-subunit ACCs may have parallel arrays of active sites. CT is a dimer in solution and in the crystal structure, and it appears likely to remain dimeric in the multi-subunit ACCs.[87; 90; 160] Interestingly, the distance between the two active sites of the CT dimer is about the same as that of the BC dimer suggesting direct interactions with the BC subunits which would allow the positioning of the different active sites in close proximity.[90] The fact that the complex has two BCCPs per BC molecule (BC_2:$BCCP_4$) removes the conflict between the two BCCP dimer models discussed above. Each pair of the BC and CT active sites may be in contact with two BCCP subunits in the ACC complex.[122; 156; 158]

1.3 Biotin Protein Ligase

1.3.1 The Biotinylation Reaction Catalysed by BPL

Biotin protein ligase (BPL, EC: 6.3.4.15) catalyses the covalent attachment of biotin to the specific target lysine residues of the different biotin-dependant enzymes.[44; 161] BPL is also known as holocarboxylase synthetase (HCS, EC 6.3.4.10), which is somehow misleading since some biotin enzymes are decarboxylases and transcarboxylases. The enzymatic biotinylation reaction catalysed by BPL is Mg^{2+} and ATP-dependant and occurs in two steps (Fig 1.10).[162; 163] In the first step, biotin reacts with ATP in presence of Mg^{2+} generating the reactive intermediate biotinyl-5'-AMP with release of Pi. In the second step, the ε-

amino group of the target lysine of BCCP reacts with the activated carboxyl group of biotin to form the amide linkage of holo-BCCP and the AMP moiety is eliminated.

The reaction catalysed by BPL is mechanistically homologous to that catalysed by lipoyl protein ligase and aminoacyl-tRNA synthetases. The three different types of enzymes activate their specific substrate through an adenylate intermediate in presence of ATP before post-translational attachment to their respective acceptor.[164-167] Most organisms contain fewer than five different biotin-dependant enzymes and in *E. coli* and most bacteria, only the BCCP subunit of acetyl-CoA carboxylase is biotinylated.[168] Thus, the biotinylation reaction catalysed by BPL is a remarkably specific post-translational modification as only one of the >4000 protein species in *E. coli* is recognized and biotinylated.[161]

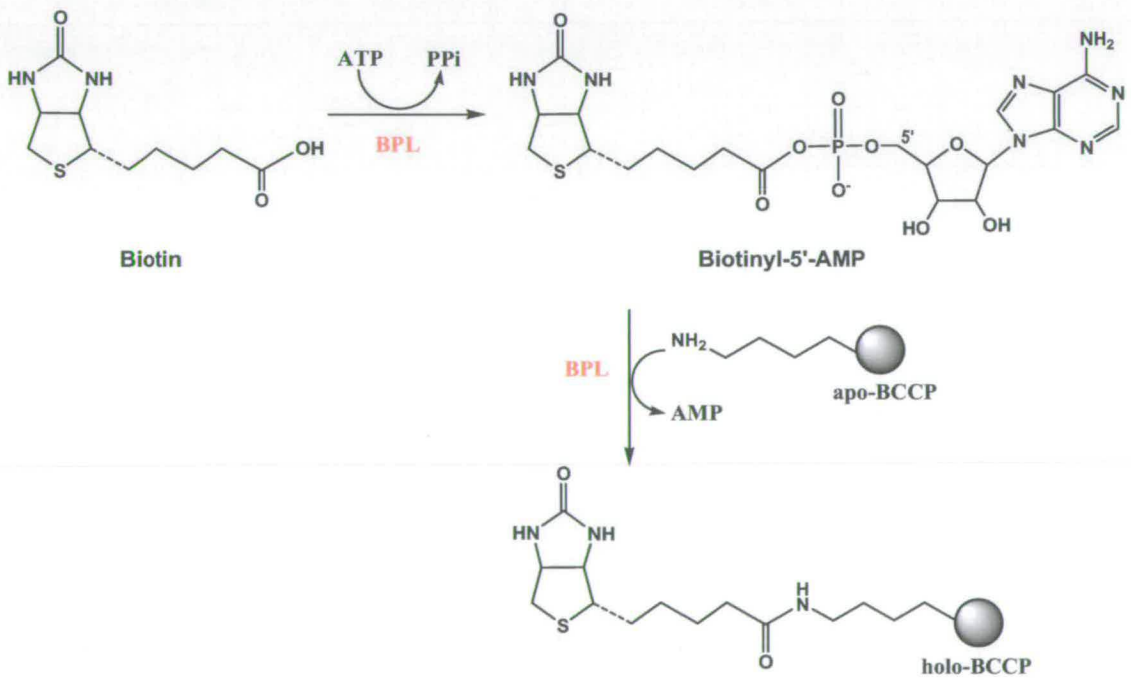


Figure 1.10 - Biotinylation reaction catalysed by Biotin Protein Ligase. The reaction occurs in two-steps with intermediary of biotinyl-5'-AMP.

The functional interactions between BPLs and biotin-dependant enzymes are highly conserved throughout Nature. Biotin-accepting enzymes can be recognised and biotinylated by BPLs derived from widely divergent species and the biotin protein ligase from *E. coli*, BirA, has been shown to biotinylate *in vivo* biotin carriers from other bacteria, yeast and mammals.[35; 169; 170] Genetic analyses in microorganisms and humans have revealed that almost all organism contains a single BPL-encoding gene responsible for processing each biotin-dependant enzyme.[161] Mutational analysis within the *birA* gene of *E. coli* showed that BPL is required for cell growth and mutations within the human HCS gene are known to lead to multiple carboxylase deficiency disease.[49; 171; 172] Only *Buchnera* species, *Borrelia burgdorferi*, *Aeropyrum pernix*, thermoplasmas, and mycoplasmas have neither the BPL nor the biotin biosynthetic genes, which is consistent with the absence of genes for biotin-dependent carboxylases in the genomes of these microorganisms.[173]

BPLs have been classified into four sub-classes according to their size, domain structure and ability to control biotin metabolism (Fig. 1.11).[174] All the BPL/HCS enzymes share a highly conserved central catalytic domain but differ in sequence at their amino termini. Class I and II BPLs are derived from bacteria and both possess a small C-terminal domain of unknown function.[174] BPL enzymes from class II, including *E. coli* BirA and *Bacillus subtilis* BPL, display an additional N-terminal DNA binding domain.[175; 176] Class II BPLs are bifunctional proteins that act as ligases and as transcriptional repressors of biotin biosynthesis.[177] The class I BPL genes from microorganisms such as *Aquifex aeolicus* and *Pyrococcus horikoshii* lack the DNA-binding domain indicating that class I BPLs are monofunctional enzymes that act solely as biotin ligases.[178; 179]

Eukaryotic BPLs are larger than the prokaryotic enzymes and fall into classes III and IV.[174] They often have long N-terminal domains which have little or no sequence similarity with bacterial enzymes. None of the eukaryotic proteins contain sequences that would suggest any DNA-binding activity. The role of the N-termini in class III and class IV BPLs remains unclear but an N-terminally truncated BPL from yeast *Saccharomyces cerevisiae* showed a dramatic reduction in enzyme activity. The N-terminal domain of the yeast enzyme has been speculated to play a role in enzymatic activity and acceptor-substrate recognition.[170]

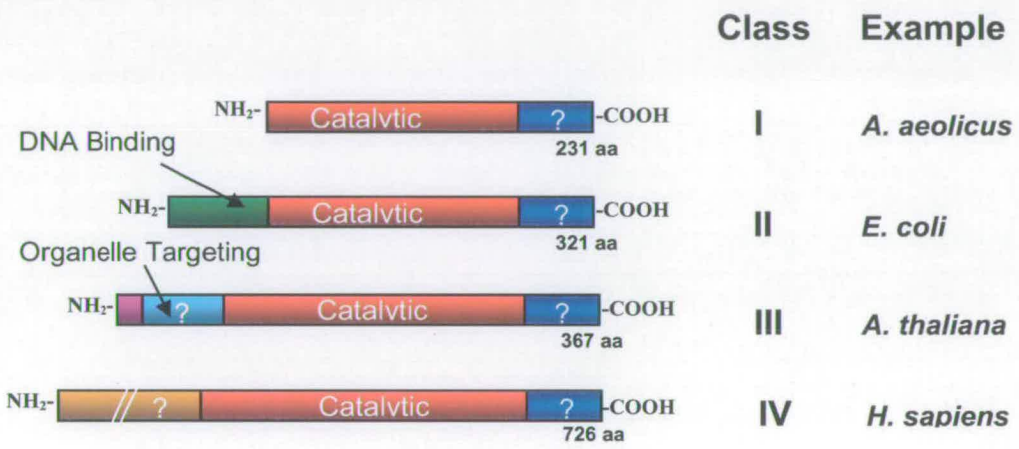


Figure 1.11 - Biotin Protein Ligase classification. BPL enzymes can be divided into four separate classes based on there tertiary structure. Bacterial BPLs are members of classes I and II, whereas classes III and IV are eukaryotic BPLs.

1.3.2 *E. coli* Biotin Protein Ligase: BirA

1.3.2.1 The Bifunctional Nature of BirA

E. coli BPL, BirA, is a 35.3 kDa enzyme that functions both as a biotin ligase and as a repressor of biotin biosynthesis.[176; 180; 181] The biotin regulatory system in *E. coli*, first discovered in the 1970s, constitutes a paradigm of biotin-sensing

processes.[177] The repressor function of BirA is triggered by the formation of biotinyl-5'-AMP which acts as an essential intermediate in the biotin transfer reaction but also serves as a positive allosteric effector for site-specific DNA binding.[182; 183] The intermediate biotinyl-5'-AMP promotes DNA binding by enhancing dimerization of BirA.[177] Thermodynamic and kinetic analysis of DNA binding to BirA have shown that only the BirA:biotinyl-5'-AMP dimer binds to the 40 base pair biotin operator sequence *bioO* located in between *bioA* and *bioB*.[184-186] In this system, the DNA-binding domains of the two BirA:biotinyl-5'-AMP monomers (holo-BirA) interact with the 12 base pair sequences at each terminus of the bidirectional *bioO* operator.[184; 187; 188] Structural and mutational studies of BirA have shown that residues important for catalytic activity are located at the dimer interface.[187-189] Biotinyl-5'-AMP binding has been suggested to promote binding to the *bioO* sequence by increasing the stability of the homodimer interface.[186]

Thermodynamic analysis of this model has indicated that the total enhancement in the free energy for assembly of the protein-DNA complex closely matches the enhancement of dimerization affected by binding of biotinyl-5'-AMP.[186] These results are consistent with the homodimer interface serving as a major mediator for transduction of the allosteric signal. The functional switch in which BirA is depleted of its function as a ligase reflects competing homodimer and heterodimer interactions.[190] BirA can be selectively targeted towards its enzymatic function simply by increasing the kinetic probability of heterodimerisation with apo-BCCP.[191] Initial kinetic measurements showed that at sufficiently high concentrations, apo-BCCP87 preferentially sequesters the BirA:biotinyl-5'-AMP

monomer in a heterodimeric complex, thereby reducing the availability of holo-BirA for homodimerization and *bioO* binding.[191]

However, these protein-protein interactions remain weak, particularly those involved in homodimerization of holo-BirA.[185; 186] In this system, a preequilibrium between the holo-repressor monomer and dimer is assumed to exist. At relatively high repressor concentrations, the dimer rapidly forms a very stable complex with *bioO* characterized by a half-life of approximately 7.5 min which is not affected by the presence of apo-BCCP.[191] The kinetic stability of the final complex ensures effective transcription repression of the biotin operon when biotin requirements are poor resulting in very low concentrations of both free biotin and unligated BirA within the cell.[191] Intracellular biotin levels are intimately linked to gene expression due to the ordered sequential addition of substrates to BirA, as biotin binds prior to ATP during the formation of the co-repressor biotinyl-5'-AMP.[192; 193] However, repression of *bioO* by BirA and biotin biosynthesis are directly controlled by the level of acceptor protein apo-BCCP *in vivo*.

Studies of another class II BPL from the bacterium *Bacillus subtillis* have revealed a similar role for BPL in regulating biotin biosynthesis.[23; 175] *B. subtilis* BPL also functions as a dual-purpose protein in that it acts as a repressor of the *B. subtilis* biotin operon and has ligase activity.[175] Orthologs of *birA* and biotin biosynthetic genes from *E. coli* and *B. subtilis* have been identified in the available bacterial genomes by similarity search.[173] The BPL gene is widely distributed in eubacteria and archaea and is more widespread than the biotin biosynthetic genes among all complete genomes. Comparative genomics suggests that the existence of class II BPL within a genome strictly correlates with the presence of putative BPL-binding

sites upstream of the biotin operon. Interestingly, comparative genomics showed that the microorganisms *Clostridium acetobutylicum*, *Lactococcus lactis*, *Pyrococcus furiosus* and *Pyrococcus abyssii* contain the two BPL paralogues, with and without regulatory domains.[173]

1.3.2.2 The Crystal Structure of BirA

The crystal structures of *E. coli* BirA in the ligand-free form (apo-BirA) was solved at 2.3 Å resolution in 1992 by Wilson *et al.,* followed in 2001 by the structure of the BirA:biotin complex (Fig. 1.12, a).[187; 194] More recently, the X-ray structure of BirA was also solved bound to the biotinyl-5'-AMP analogue biotinol-5'-AMP (Fig. 1.12, b & c).[188; 195] The structure of apo-BirA is monomeric and shows three distinct domains: an N-terminal DNA-binding domain, a central catalytic domain and a small C-terminal domain.[194] The N-terminal domain which is used for *bioO* binding adopts a helix-turn-helix fold characteristic of major DNA-binding proteins.

The central domain of BirA consists of five α-helices and a mixed β-sheet of seven strands. In apo-BirA, four surface loops located in the central domain are disordered and cannot be visualized in the electron density (loops 110–128, 140–146, 193–199 and 212–234).[194] The biotin binding site was determined by exposing the apo-BirA crystal to biocytin (ε-biotinyl lysine) which is a mimic of the final product from the biotin transfer reaction and the structure was refined at 2.8 Å resolution.[194] The biotin moiety of biocytin interacts with parts of three strands from the β-sheet, the α_5-helix and main-chain atoms of residues 114-118.[194] Residues 116-118 are disordered in the structure of apo-BirA and become localized upon biocytin binding, partially burying the ligand in the active site. The loop 116-124 contains the glycine-

rich motif $_{115}$GRGRRG$_{120}$, identified as a consensus sequence (GXGXXG). This glycine-rich sequence has been previously characterized as an ATP and GTP binding motif.[196] Because ATP and biotin must be spatially close to permit the formation of biotinyl-5'-adenylate, in BirA, the motif was initially associated with nucleotide binding.[194]

The C-terminal domain comprises an anti-parallel β-sheet sandwich and adopts a Src-homology 3 (SH3) fold. SH3 domains are protein modules of 50-70 amino acids long found in a variety of proteins involved in signal transduction.[197-199] However, since direct interactions with the substrates have yet to be observed by crystallography, the exact role of the C-terminus in BirA remains unclear.[187; 188] Evidence of protection by biotinyl-5'-AMP against hydroxyl radical cleavage of the BirA backbone at several sites within the C-terminal domain implies a possible role in the enzymatic reaction.[200] Single-site mutation studies on both BirA and the acceptor apo-BCCP87 also suggest that the small SH3-like domain may be involved in substrate recognition and ATP binding.[201]

In contrast to biocytin which interacts only weakly with BirA, the crystal of apo-BirA was destroyed on soaking with free biotin, suggesting that conformational changes had occurred upon binding.[194] However, co-crystallization of BirA with biotin has allowed the characterization of the structure of the BirA:biotin complex at a resolution of 2.4 Å.[187] The main differences between apo-BirA and the BirA:biotin complex are localized in the regions of the surface loops 110–128, 140–146, and 193–199. In the apo-BirA structure, residues within these loops ranged from partially to completely disordered. In contrast, in the BirA:biotin complex, all of the residues within the three loops are visible (Fig. 1.12, a).[187]

The hydrogen-bonding interactions between BirA and biocytin described by Wilson *et al.* are also observed in the structure of the enzyme bound to biotin.[187] However, in the structure of the BirA:biotin complex, the complete loop encompassing residues 110-128 is ordered and forms the biotin binding loop. Unlike the monomeric apo-repressor, the complex of BirA with biotin is dimeric.[187] Although biotin is not the physiological corepressor, the free co-factor has been shown to act as a weak allosteric activator of BirA binding to *bioO*.[186] In the dimer structure, the β-sheets in the central domain of each monomer are arranged side by side forming a single, seamless β-sheet. At the region of contact, residues 188-195 of one subunit form an extended β-sheet strand that is hydrogen-bonded to the same residues in the neighbouring monomer.[187] The two loops 140-146 and 193-199 which are partially disordered in the structure of apo-BirA are located at the dimerization interface. Only the loop composed of residues 212-234 remains disordered in the co-crystal structure.[187]

Functional studies indicated that the dimerization and DNA-binding properties exhibited by BirA in complex with the analogue biotinol-5'-AMP closely resemble those of the functional biotinyl-5'-AMP repressor complex.[195] The X-ray structure revealed that the BirA:biotinol-5'-AMP complex is dimeric and is related to the BirA:biotin dimer (Fig. 1.12, c).[188] However, several conformational differences are observed between the two BirA complexes including the ordering of residues 212-223 which form the adenylate binding loop upon biotinol-5'-AMP binding.[195; 202] Furthermore, the BirA:biotinol-5'-AMP complex forms a much tighter dimer than BirA bound to biotin, consistent with previous equilibrium sedimentation studies. In the structure of BirA bound to biotinol-5'-AMP, the "hinge angle" between the

catalytic domains has bent by 12 °, increasing significantly the buried surface area of the dimer. Concomitant with those changes, the C-terminal domains are shifted and lie closer together resulting in new interactions at the dimer interface.[195] The organization of the structure of BirA upon ligand binding suggests that the allosteric activator function reflects its ability to induce disorder-to-order transitions in multiple segments of BirA. These folding events are prerequisites for tight binding of the holo-BirA dimer to *bioO*.[185-188; 200]

In contrast to biotin binding, there is no preformed binding site for the adenylate in apo-BirA.[188] Only after biotin is bound and the biotin binding loop is ordered, is the full adenylate binding site formed. This is consistent with the sequential substrate addition in BirA with biotin binding prior to ATP. The binding mode for adenylate was unanticipated and although the glycine-rich motif is important for phosphate binding in BirA, it conserves neither the structure nor phosphate interactions characteristic of the canonical GXGXXG nucleotide-binding motif (Fig. 1.12, b).[188; 196] The $_{115}$GRGRRG$_{120}$ motif is located at the turn of the biotin binding loop and in the structure of BirA bound to biotin, the backbone nitrogen of Arg118 interacts with the carboxyl group of biotin.[187] Additionally, the backbone nitrogen and carbonyl of Arg116 interact with the carbonyl and the ureido nitrogen of biotin respectively. In the structure of BirA in complex with biotinol-5'-AMP, Arg118 and the adjacent Arg121 make the critical phosphate binding interactions.[188]

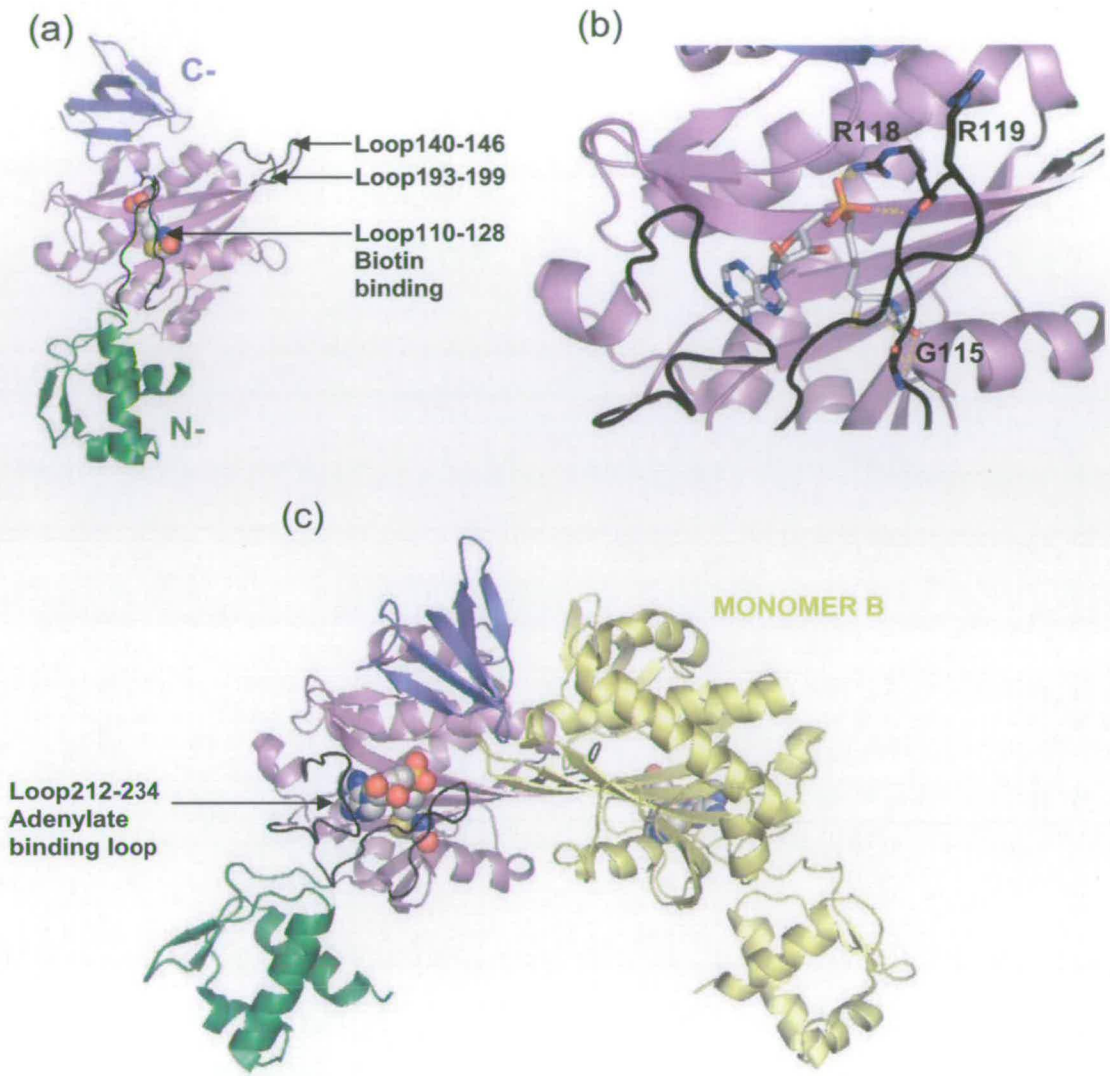


Figure 1.12 - The crystal structure of *E. coli* BirA. The N-terminal, the central, and the C-terminal domains are shaded green, violet and blue slate respectively and the disordered loops are highlighted in black **(a):** Monomeric BirA in complex with biotin. The four disordered loops in the apo-structure of BirA are coloured in black. Biotin is drawn in spheres whith white carbons. **(b):** Close up view of BirA active site with biotinol-5'-AMP bound. The residue Gly115, Arg118 and Arg119 on the biotin binding loop are labelled and drawn in sticks with black carbons. Biotin is drawn in sticks whith white carbons. **(c):** BirA dimer in complex with biotinol-5'-AMP bound at each active site. Monomer B is shaded yellow. The adenylate binding loop coloued blak is ordered and highlighted. Biotinol-5'-AMP is drawn in spheres with white carbons. (PDB codes 1HXD for *(a)* and 2EWN for *(b)* and *(c)*)

1.3.2.3 The Glycine-Rich Motif of BirA

Early biochemical assays had shown the importance of the glycine-rich motif $_{115}$GRGRRG$_{120}$ in BirA catalytic activity, dimerization and repressor function. Hydroxyl radical cleavage of BirA with [Fe-EDTA]$^-$ was performed to determine the ligand-induced changes in the reactivity of the peptide backbone of BirA and the regions most affected upon binding of biotin were shown to be residues 115-120 and the biotin binding loop.[200] The DNA-binding and assembly properties were analysed for the three mutants BirA G115S, R118G and R119W of the glycine motif and revealed that both R118G and R119W mutants were defective in dimerization and *bioO* binding while BirA G115S bound to the biotin operator with the same affinity as wild-type holo-BirA.[189]

Although structural perturbations to apo-BirA induced by the mutations should be minor, isothermal titration calorimetry experiments carried out to determine the thermodynamic profile of ligand binding to the mutants indicated that the structural consequence of amino acid replacement was the disruption of hydrogen-bonding interactions and acquisition of the local disorder which characterizes apo-BirA.[203]. Kinetic analysis of the three BirA mutants indicated that BirA G115S and R118G were defective in binding both biotin and biotinyl-5'-AMP.[204] The crystal structure of BirA shows that the side chain of Arg119 is largely solvent-exposed and therefore, the third mutation R119W did not affect the affinity for the substrates (Fig. 1.12, b).[188] However, the three mutant-biotin complexes were shown to remain competent for ATP binding and subsequent synthesis of biotinyl-5'-AMP.[204].

The dissociation constant of BirA G115S and R118G for biotinyl-5'-AMP are 3000 and 400-fold greater than for wild-type BirA respectively.[204] Biotinyl-5'-AMP is a good biotinylation reagent and the possibility of the mutants acting as a promiscuous biotinylating agents was tested by streptavidin western blotting.[205] Analysis of the extract from BirA G115S, R118G and R119W indicated that, in contrast to wild-type BirA, the two mutants were self-biotinylated *in vivo*. This observation has been suggested to arises as a result of dissociation of the biotinyl-5'-AMP-mutants binary complexes to give free adenylate which reacts with exposed lysine residues. When the R118G protein was expressed in an *E. coli* biotin auxotroph in the presence of [^{14}C]-labeled biotin, the molar ratio of covalently attached biotin to the mutant was 0.4 suggesting significant protein modification during expression.[205] Additionally, the R118G mutant was also shown by western blot to biotinylate a wide range of non-cognate cellular proteins. Surprisingly, the mutant G115S showed no labelling of proteins apart the specific target BCCP. Binding of biotin and biotinyl-5'-AMP to the three mutants were determined with the loss of the increase in intrinsic fluorescence and thus, the high dissociation constants measured for BirA G115S are mainly attributed to a shift in the position of the ligands within the active site.[204; 205]

Western blotting studies indicated that both BSA and RNAse A were readily biotinylated *in vitro* by the R118G mutant in a much greater extent than with wild-type BirA.[205] Because biotinyl-5'-AMP is a readily hydrolysable mixed anhydride, the ability of the mutant to modify distant proteins was studied by analysing the extent of biotinylation of a histidine antibody bound either to the C-terminal of BirA R118G or to a His-tagged acceptor protein. The chemical biotinylation reaction

catalysed by BirA R118G was shown to be more efficient when the antibody was bound to the mutant and thus it was suggested that BirA R118G could be manipulated to generate active biotinyl-5'-AMP and biotinylate target peptides and proteins in a proximity dependant process when studying protein-protein interactions within the cell.[205; 206] Furthermore, the mechanistic importance of this residue is underscored by the observation that mutation of the corresponding arginine by a tryptophan in the conserved $_{505}$GKGRGG$_{510}$ glycine motif of human BPL (R508W) is the most recurrent mutations in the HCS gene. The mutation was shown to alter the affinity of human BPL for biotin and gives rise to multiple carboxylase deficiency syndrome.[172; 207]

1.3.2.4 Structural Homologs

The core of the catalytic domain of BirA can be described as a circular permutation where the N- and C-termini permit folding of the RNA recognition motif (the RRM or ferrodoxin-like fold) which is a mixed αβ fold containing two α-helices packed against a four-stranded, antiparallel β-sheet.[208; 209] The RRM fold of BirA contains several insertions including the biotin and adenylate binding loops and thus, BirA belongs to the circular permutation Cyc2.[188] BPL, lipoate-protein ligase A (LplA) and aminoacyl-tRNA synthetases belong mechanistically to the adenylating-class of enzyme.[165; 167] The similarities between the three families of enzyme extend in some cases to the structure.

Despite low sequence identities, the closest structural homologs to BirA are members of the bacterial LplA families, another example of the Cyc2 permutation of the RRM-like fold.[188; 210] BirA and LplA bind similar substrates in equivalent

positions and catalyse essentially the same chemical reactions suggesting that they are evolutionarily closely related.[188; 210; 211] However, the substrate binding loops of bacterial LplA indicate that the two enzymes have undergone substantial divergence.[212] In LplA, the insertion that corresponds to the biotin binding loop of BirA is a small αβ domain that interacts with the lipoate moiety and is ordered in the apo-enzyme of *E. coli* and *Thermophilus acidophilum*.[210-212] This loop contains residues that are conserved in all lipoate protein ligases and plays an important role for the binding of the substrate and the stabilization of the three-dimensional structure. The loop sequence has been suggested to play a role in substrate specificity by preventing biotin entering the active site of LplA.[211; 212] LplA also has an insertion similar to the adenylate binding loop in BirA, but it is larger and does not undergo a disorder-to-order transition. The absence of a DNA binding domain in the LplA structure may be related to the finding that LplA metabolizes lipoic acid that is exogenously supplied rather than synthesized endogenously in *E. coli*.[167; 210; 213]

BirA also belongs to the same structural family as class II aminoacyl tRNA synthetases (aaRS) which are associated with, Ala, Asn, Asp, Gly, His, Lys, Phe, Pro, Ser and Thr substrate amino acids.[199; 214; 215] Active sites of class II aaRS are built around an antiparallel β-sheet flanked by alpha helices corresponding to a Rossmann fold.[216] Initial comparison of the structures of BirA and seryl-tRNA synthetase revealed striking similarities between their catalytic domains suggesting the possibility of a remote evolutionary relationship.[214] Investigation of the structure of phenylalanyl-tRNA synthetase (PheRS) showed, that although the catalytic domain of the synthetase represents a Cyc1 circular permutation of the RRM-like fold, the DNA-binding domain containing the helix-turn-helix motif and the SH3-

fold barrel-like domain of BirA are conserved.[199] Despite the fact that there is no clear homology between the structure-derived sequence alignment of BirA and PheRS, all three domains of BirA appeared to have counterparts in the tRNA synthetase structure, supporting the concept of a common ancestor for the two enzymes. However, comparisons of the structure of BirA/LplA and class II tRNA synthetases show poor conservation of the adenylate binding sites.[188] Furthermore, ATP and amino-acyl binding has been proven to be random in class II tRNA synthetase which indicate significant mechanistic divergence within the adenylating enzymes family.[217] Thus, it has been suggested that the primary function of this group of enzymes was to scavenge co-factors or amino-acyls raising the question wether the adenylating property appeared before or after selection of their respective substrates.[188]

1.3.3 The BPL:BCCP Complex

Following the binding of biotin and ATP and the subsequent formation of biotinyl-5'-AMP, the BPL ligase must catalyse the transfer of the biotin moiety onto the specific lysine of BCCP.[161] Concomitant with the specificity of the biotinylation reaction of BCCP, and with the complication of the competitive homo- and heterodimerization observed with BirA, a BPL:BCCP complex must be formed to allow nucleophilic attack from the ε-amino group of the correct lysine of BCCP onto the phospho-anhydride bond of biotinyl-5'-AMP.[185; 190; 191] The complex formed between the class I BPL and the C-terminal fragment BCCPΔ67 of the hyperthermophilic bacteria *Aquifex aeolicus* has been characterised by chemical

cross-linking. SDS-PAGE as well as gel-filtration analysis indicated a heterodimer of 1:1 ratio.[178]

The elucidation of the three-dimensional structures of *E. coli* BirA and BCCP87 has allowed Weaver *et al.* to construct a model of the BirA:BCCP complex.[218] Because of the conformational differences observed upon biotin binding to apo-BirA, the BirA:biotin complex was used for modelling the heterologous protein-protein interaction with BCCP.[194] In the proposed model, BCCP is aligned with BirA such that the ε-NH_2 of the target Lys122 is brought close to the carboxyl group of the biotin moiety. The β-strand 122-128 of BCCP can be readily modelled so that it forms an extended parallel β-sheet with the same β-strand 188-195 which is involved in homodimerization of BirA.[218]

The protein interactions between BirA and BCCP described in this system are supported by previous structural and biochemical studies. Heteronuclear NMR spectroscopy studies have shown that the chemical shifts associated with the two regions of the BCCP sequence (residues 119–128, 87–90 and 145) which are the most perturbed in the presence of BirA correspond precisely to the regions of BCCP that have the most extensive contacts with BirA in the model.[219] The BirA:BCCP model is also consistent with the results of mutagenesis studies of BCCP in which the residues of BirA Glu119 (close to the target Lys122 of BCCP) and Glu147 (close to Pro145 of BCCP) were shown to be critical for the biotin transfer reaction.[201] The model interaction between BirA and BCCP presents Val88, Gly89, and Pro145 of BCCP to residues 294-297 of the C-terminal domain of BirA. Thus, the SH3 module appears to make a modest favourable contribution to the binding of BCCP and possibly to its correct orientation on BirA.[201; 218]

1.3.4 *Pyrococcus horikoshii OT3 BPL*

The first structure of a class I BPL from the hyperthermophilic bacterium *Pyrococcus horikoshii* OT3 (*Ph*BPL) was determined at 1.6 Å resolution in 2005 by Bagautdinov *et al.*[179; 220] The structure of *Ph*BPL was solved in its apo-form and in complex with biotin, ADP and biotinyl-5'-AMP (Fig. 1.13).[179] Because *Ph*BPL is mono-functional, it would seem that this is simpler model with which to study the botinylation reaction than BirA. The structure of *Ph*BPL reveals a constitutive homodimer associated through an inter-subunit β-sheet formed by the β_1-strands. The polypeptide chain shares strong structural homologies with BirA and comprises of a large N-terminal catalytic domain and a smaller C-terminal domain. In the structure of apo-*Ph*BPL, only the biotin binding loop is disordered and this becomes structured upon biotin or ADP binding. ADP binds at a site adjacent to the biotin binding pocket and is largely stabilized by the biotin binding loop (Fig. 1.13). Therefore, only minor conformational changes to the structure are induced during the formation of biotinyl-5'-AMP. The conserved Lys111 from the β_7-strand has been proposed to play a role in the formation of biotinyl-5'-AMP by stabilising the pentacoordinate transition state of the α-phosphate before formation of the phosphoanhydride bond of biotinyl-5'-AMP.[179]

In 2006, after initiation of the work described in this thesis, several new BPL structures were deposited in the Protein Data Bank including the structures of *Ph*BPL in complex with ATP by Bagaudtinov *et al.* and the BPL from *Mycobacterium tuberculosis* by Ma *et al.* (PDB entries 1XO1 and 2CGH respectively). In 2007, the structures of the complex formed by *Ph*BPL and biotin and ATP as well as the apo-

form of *A. aeolicus* BPL became available on the PDB (entries 2DTO and 2EAY respectively). Preliminary crystallographic studies of the *P. horikoshii* single and double mutant complexes BPL R48A:BCCPNΔ76 and BPL R48A-K111A:BCCPNΔ76 have also been recently reported but no structure is yet available in the databank.[221]

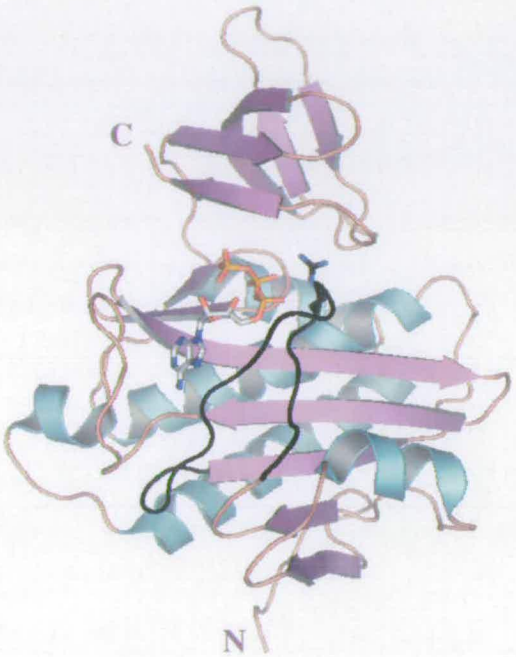


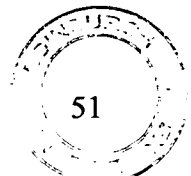
Figure 1.13 – Crystal structure of *Pyrococcus horikoshii* BPL in complex with ATP. The disordered biotin binding loop in the apo-structure is shaded black. The Arg48 (equivalent to the Arg118 of *E. coli* BirA) and ATP are shown in sticks with black and white carbons respectively. (PBD code 1X01)

1.4 Analysis of Non-Covalent Protein Complexes by Mass Spectrometry

The recent introduction of soft ionisation techniques such as electrospray ionisation (ESI) and matrix-assisted laser desorption/ionisation (MALDI) has had a

tremendous impact in the field of protein mass spectrometry. In 2002, Fenn and Tanaka were awarded the Nobel Prize in Chemistry for their role in the development of these two ionisation methods.[222-224] These techniques are increasingly important in proteomics - allowing protein identification and quantification, protein profiling and protein interactions analysis.[225-227] The initial role of proteomic mass spectrometry was to characterise the molecular masses of proteins according to their primary sequences. However, the recent advances in the field have allowed mass spectrometry to become an essential complementary tool in structural biology for the investigation of secondary, tertiary and quaternary structures of large protein complexes.[228-230]

The principle of mass spectrometry MS is to generate ions in the gas phase, separate them according to their *mass-to-charge ratio* (*m/z*) and detect them qualitatively and quantitatively according to their respective *m/z*.[231] In order to study large unstable protein complexes, the generation of gas phase ions for MS analysis must occur directly from the protein in solution in a manner that avoids fragmentation or other undesirable alteration of the target proteins.[223; 232] Although MALDI is also applied to study these assemblies, analyses of the stability of intact protein complexes with ESI and the derived applications to study their overall structure far outnumber analyses using MALDI.[233-235] This is primarily due to the sample-preparation requirements for MALDI studies which typically involve evaporation of solvent in highly acidic conditions that will undoubtedly perturb protein-protein interactions present in solution.[224] Additionally, MALDI mass spectra often yield intense signals for protein aggregates which may be artefacts of the laser desorption/ionisation process.[230; 232]



A typical electrospray setup involves passing a sample through a metallic capillary held at high electrical potential (1-4 kV) making the tip enriched in ions usually in the positive mode.[236] The applied potential causes the charges to gather preferentially at the tip of the capillary with an elongated meniscus forming a "Taylor cone". At the tip of the cone, the stream of solvent is drawn out into highly charged droplets of several micrometers in diameter. This ionisation process takes place at atmospheric pressure and is, therefore, very gentle.[236] Emerging droplets are subsequently guided by a potential and pressure gradient toward the mass analyser. Aided by flows of nebulizing gas, evaporation of solvent from the droplets decreases its radius and since the charge is conserved, at some critical radius (the "Rayleigh limit"), Coulombic forces overcome the surface tension of the liquid and lead to fission of the droplets into even smaller droplets. Subsequent depletion of solvent by further evaporation results in several generations of unequal droplet fission, until ultimately gas phase ions are produced.[236]

A conventional electrospray apparatus uses a spray capillary of approximately 0.5 mm in diameter and requires flow rates of several microliters per minute to maintain the Taylor cone stable.[236] In 1994, Wilm and Mann introduced an important variant of conventional ESI, termed nanoelectrospray (nESI).[237; 238] nESI is typically performed using glass or quartz capillaries which have been pulled to a fine tip (1 μm inner diameter) and given a metallic (usually gold) coating to hold the electric potential in place. The sample is electrosprayed at flow rates in the range of 1-10 nl/min driven primarily by the 0.5-2 kV potential applied to the capillary and an auxiliary backing gas pressure to maintain a steady stream of the solution through the tip.[239] nESI desolvation process has been shown to be a gentler, more reliable,

method of introducing extremely labile protein assemblies into the gas phase intact. It is generally accepted that the smaller initial droplet sizes produced by the nESI source (100 to 500 nm) reduce the number and the energy of the collisions required to desolvate the macromolecules of interest.[232; 238]

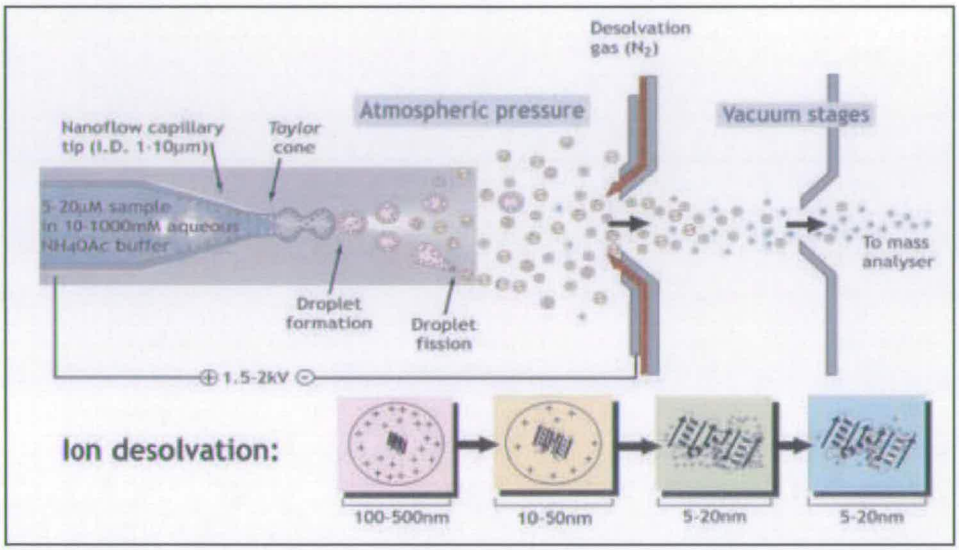


Figure 1.14 - Scheme of the nanoelectrospray ionisation process. (Robinson *et al.*)

Although a wide range of mass analysers has been developed for the analysis of protein molecules, only a small subset of these has been successfully applied to the study of large protein complexes.[232] Typically, large protein ions appear at high *m/z* which considerably limits the use of typical mass spectrometers. Time-of-flight (ToF) analysers have been the preferred tool for the study of large protein complexes The principal advantage of the ToF spectrometer is the theoretically unlimited *m/z* range which is essential to study macromolecular assemblies.[240] Hybrid instruments have been developed using quadrupole (Q-ToF) or/and hexapole filters used in radiofrequency-only mode which transmit and focus the ion beam as well as acting as an interface and are orthogonally coupled to the TOF tube.[241] In tandem MS the

quadrupole and ToF analysers arranged in series are separated by a collision cell and ions of interest can be selected according to their *m/z* ratios in the quadrupole analyser, subsequently activated in the collision cell, with their products being analysed in the ToF analyser.[241] In most commercial instruments the quadrupole mass range is limited to m/z 3,000–4,000 and in order to be able to select ions at high *m/z* ratios, quadrupoles operating at low frequency have been implemented in Q-ToF instruments by several groups.[242-244]

The oligomeric state of protein complexes is often preserved during transfer from solution to the gas phase and thus, mass spectrometry provides an ideal approach to investigate such protein-protein interactions as the measured molecular mass of the oligomer directly reveals the number of subunits in the quarternary structure.[230] The ESI mass spectrometry analysis carried out by Fitzgerald *et al.* in 1996 on the quaternary structure of 4-oxalocrotonate tautomerase is one of the classical studies in which the oligomeric state of a protein was probed.[245] Additionally, mass spectrometry can also provide information on aspects such as the dynamics of assembly, cooperativity, and ligand binding of intact protein complexes in solution phase.[230] Elegant studies revealing the ability of mass spectrometry to investigate cooperativity were carried on the transthyretin system by McCammon *et al.*[246] Their observation of the multifaceted ten-component complex containing six protein subunits (four transthyretin and two retinol binding proteins), two retinol molecules, and two synthetic ligands allowed them to conclude that inhibitor binding does not inhibit association of transthyretin with retinol bound retinol binding protein.[246] Most MS studies show correlation of the structural and thermodynamical properties between the solution phase and gas the phase. However, quarternary

structures may be altered since water molecules cannot longer assist in stabilising the molecular structures in the gas phase.[230]

Activation and dissociation of non-covalent protein complexes in the gas phase can also afford further compositional information.[247] Analysis by MS of a multi-subunit protein and its sub-complexes can generate a three-dimensional interaction map of the different components within the overall complex.[248] The majority of the studies in which a non covalent protein complex is disassembled in the gas phase have employed collision induced dissociation (CID).[232; 249] This is performed using tandem MS where ions of large protein assemblies are mass selected and activated *via* collisions with neutral gas atoms or molecules.[249] Activation occurs as a portion of the kinetic energy of an ion is converted into internal energy during each collision event and this may lead to dissociation if sufficient internal energy is accumulated.[250] The CID technique can be used on homologous protein complexes as well as heterogeneous assemblies, often oligomers of different protein subunits.[230] The nESI mass spectrometric structural analysis by the group of Robinson of intact 70S ribosome complexes consisting of proteins and RNA is one of the first examples in which such a heterogeneous assembly has been characterised.[251] While dissociation studies of protein complexes in the gas phase have mostly used CID, blackbody infrared radiative dissociation (BIRD) and, more recently, electron-capture dissociation (ECD) and surfaceinduced dissociation (SID) have also been employed.[252-254]

Large multi-protein complexes show a tendency to decompose by CID *via* a similar pathway consisting in the expulsion of a small monomeric subunit.[230; 243] Gas phase dissociation of the streptavidin tetramer (S_4) follows the typical monomer-

trimer decomposition pattern with the resulting monomer fragment moiety carrying a disproportionately large fraction of the charge (S^{7+} and S_3^{7+}).[255] A model explaining this anomalous charge distribution observed in the dissociation of numerous large protein complexes suggested that the asymmetric dissociation process is analogous to the fission of charged droplets in the electrospray process in which the smaller offspring droplets are believed to form with much higher *m/z* than the parent droplet.[255] However, in the recent years, gas phase dissociation pathways have been observed in which the loss of a highly charged monomer ion is not the principal decay pathway.[244] In these cases, non-covalently bound tetrameric proteins were found to undergo symmetric dissociation into dimers. Such exceptions provide the opportunity to manipulate the dissociation of protein complexes in the gas phase. As more information on the mechanism of dissociation comes to light, it is becoming clear that considerable structural information can be obtained for protein complexes following such dissociative approaches.[247] Information can be generated regarding the nature of intra- and inter-subunit interactions, on both global and local levels, as well as details of the overall organization of subunits within an oligomer.[249]

1.5 Aims

To understand the chemistry of the biotinylation reaction catalysed by biotin protein ligase, it was important to characterise the protein-protein interactions that control biotin transfer onto the substrate apo-BCCP. Preliminary studies on the class I biotin protein ligase from the hyperthermophilic organism *Aquifex aeolicus*

(*Aa*BPL) carried out at the University of Edinburgh showed that the enzyme was functional and isolation of a chemically cross-linked *Aa*BPL:BCCPΔ67 complex was reported for the first time. However, in the absence of a crystalline complex, site directed mutagenesis of the residues suspected to be involved in the cross-linking appeared necessary in order to gain an insight into the geometry of the structural interactions between *Aa*BPL and BCCPΔ67. Furthermore, the enzyme mechanisms leading to the formation of the intermediate biotinyl-5'-AMP remained unclear in the monofunctional *Aa*BPL and the binding order of the substrates biotin and ATP needed to be explored. Indeed, in contrast to *E coli* BirA, previous limited proteolysis experiments of *Aa*BPL revealed only one single protease-sensitive site adjacent to the predicted substrate-binding pocket. The task here was to determine the crystal structure of *Aa*BPL and further characterise the interactions made with biotin and ATP which govern the extreme specificity of the biotinylation reaction.

Isolation of the *E. coli* multi-subunit ACC has yet not been achieved due to its high tendency to aggregate. Although the components biotin carboxylase and carboxyltransferase have been well characterized, the stoichiometry of the full length *E. coli* biotin carrier *Ec*BCCP remains unknown. Prof. Carol Robinson from the University of Cambridge has been specialised in determining the intact assemblies and the structural organization of large, unstable protein complexes by mass spectrometry. Therefore, in collaboration with her laboratory, analyses of the ACC subunits using mass spectrometry would constitute a novel approach towards the characterization of the assembly state of the full *E. coli* ACC complex.

1.6 References

1. Lieberman, S. & Bruning, N. (1990). *The Real Vitamin and Mineral Book.* Avery Publishing Group Inc, Garden City Park, New York.
2. Smith, S. M., McDonald, A. & Webb, D. (1998). *Complete Book of Vitamins and Minerals*. Publications International Ltd, Lincolnwood, Ill.
3. Leeper, F. J. & Smith, A. G. (2007). Editorial: Vitamins and cofactors-chemistry, biochemistry and biology. *Nat Prod Rep* **24**, 923-926.
4. Carpenter, K. J. (2003). A short history of nutritional science: part 2 (1885-1912). *J Nutr* **133**, 975-84.
5. Carpenter, K. J. (2003). A short history of nutritional science: part 1 (1785-1885). *J Nutr* **133**, 638-45.
6. Eijkman, C. (1929). Nobel lectures: Antineuritic vitamin and beriberi. *The Nobel Prize in Physiology or Medicine.* Elsevier Publishing Company, Amsterdam.
7. Hopkins, F. G. (1929). Nobel lectures: The earlier history of vitamin research. *The Nobel Prize in Physiology or Medicine.* Elsevier Publishing Company, Amsterdam.
8. Funk, C. (1922). *The Vitamines 2nd ed,* Williams and Wilkins Company, Baltimore.
9. Carpenter, K. J. (2003). A short history of nutritional science: part 3 (1912-1944). *J Nutr* **133**, 3023-32.
10. Kinnersley, H. W., O'Brien, J. R. & Peters, R. A. (1935). Crystalline vitamin B(1). *Biochem J* **29**, 701-15.
11. McCollum, E. V., Orent-Keiles, E. & Day, H. (1939). *The Newer Knowledge of Nutrition 5th ed.* The MacMillan Co, New York.
12. Abras, E. & Walser, M. (1982). Growth of rats fed by a continuous intragastric infusion containing amino acids and keto acids. *Am J Clin Nutr* **36**, 154-61.
13. Holliday, G. L., Thornton, J. M., Marquet, A., Smith, A. G., Rebeille, F., Mendel, R., Schubert, H. L., Lawrence, A. D. & Warren, M. J. (2007). Evolution of enzymes and pathways for the biosynthesis of cofactors. *Nat Prod Rep* **24**, 972-87.
14. Rebeille, F., Ravanel, S., Marquet, A., Mendel, R. R., Smith, A. G. & Warren, M. J. (2007). Roles of vitamins B5, B8, B9, B12 and molybdenum cofactor at cellular and organismal levels. *Nat Prod Rep* **24**, 949-62.
15. Dakin, H. D. & West, R. (1935). Observations on the chemical nature of a hematopoietic substance occurring in liver. *J Biol Chem* **109**, 489-517.
16. Olson, R. E. (2001). Karl August Folkers (1906-1997). *J Nutr* **131**, 2227-30.
17. Kögl, F. & Tönnis, B. (1936). Uber das Bios Problem. Darstellung von krystallisiertem Biotin aus Eigelb. *Zeitschr Physiol Chem* **242**, 43-73
18. Melville, D. B., Moyer, A. W., Hofmann, K. & Du Vigneaud, V. (1942). The Structure of Biotin: The Formation of thiophenvaleric acid from biotin. *J Biol Chem* **146**, 487-92.
19. Traub, W. (1956). Crystal Structure of Biotin. *Nature* **147**, 649-50.

20. Eldridge, A. L. (2004). Comparison of 1989 RDAs and DRIs for Water-Soluble Vitamins. *Nutr Today* **39**, 88-93.
21. Eisenberg, M. A. (1973). Biotin: biogenesis, transport, and their regulation. *Adv Enzymol Relat Areas Mol Biol* **38**, 317-72.
22. Otsuka, A. J., Buoncristiani, M. R., Howard, P. K., Flamm, J., Johnson, C., Yamamoto, R., Uchida, K., Cook, C., Ruppert, J. & Matsuzaki, J. (1988). The *Escherichia coli* biotin biosynthetic enzyme sequences predicted from the nucleotide sequence of the bio operon. *J Biol Chem* **263**, 19577-85.
23. Bower, S., Perkins, J. B., Yocum, R. R., Howitt, C. L., Rahaim, P. & Pero, J. (1996). Cloning, sequencing, and characterization of the *Bacillus subtilis* biotin biosynthetic operon. *J Bacteriol* **178**, 4122-30.
24. Phalip, V., Kuhn, I., Lemoine, Y. & Jeltsch, J. M. (1999). Characterization of the biotin biosynthesis pathway in *Saccharomyces cerevisiae* and evidence for a cluster containing BIO5, a novel gene involved in vitamer uptake. *Gene* **232**, 43-51.
25. Schneider, G. & Lindqvist, Y. (2001). Structural enzymology of biotin biosynthesis. *FEBS Lett* **495**, 7-11.
26. Lotierzo, M., Tse Sum Bui, B., Florentin, D., Escalettes, F. & Marquet, A. (2005). Biotin synthase mechanism: an overview. *Biochem Soc Trans* **33**, 820-3.
27. Alberts, A. W. & Vagelos, P. R. (1972). *The Enzymes*. Vol 6, Academic Press, New York.
28. Moss, J. & Lane, M. D. (1971). The biotin-dependent enzymes. *Adv Enzymol Relat Areas Mol Biol* **35**, 321-442.
29. Utter, M. F. & Kolenbrander, H. M. (1972). *The Enzymes*. Vol 6, Academic Press, New York.
30. Berkner, K. L. (2005). The vitamin K-dependent carboxylase. *Annu Rev Nutr* **25**, 127-49.
31. Cranenburg, E. C., Schurgers, L. J. & Vermeer, C. (2007). Vitamin K: the coagulation vitamin that became omnipotent. *Thromb Haemost* **98**, 120-5.
32. Portis, A. R., Jr. & Parry, M. A. (2007). Discoveries in Rubisco (Ribulose 1,5-bisphosphate carboxylase/oxygenase): a historical perspective. *Photosynth Res* **94**, 121-43.
33. Knowles, J. R. (1989). The mechanism of biotin-dependent enzymes. *Annu Rev Biochem* **58**, 195-221.
34. Butler, J. N. (1992). *Carbon Dioxide Equilibria and their Applications. Reading*. Addison-Wesley Publishing Company, Massachusetts.
35. Cronan, J. E., Jr. (1990). Biotination of proteins in vivo. A post-translational modification to label, purify, and study proteins. *J Biol Chem* **265**, 10327-33.
36. Lynen, F., Knappe, J., Lorch, E., Jütting, G. & Ringelmann, E. (1959). Die biochemische Funktion des Biotins. *Angew Chem* **71**, 481-486.
37. Bonnemere, C., Hamilton, J. A., Steinrauf, L. K., Knappe, J. (1965). Structure of the bis-p-bromoanilide of carbon dioxide biotin. *Biochemistry* **4**, 240-245.
38. Alberts, A. W., Nervi, A. M. & Vagelos, P. R. (1969). Acetyl CoA carboxylase, II. Demonstration of biotin-protein and biotin carboxylase subunits. *Proc Natl Acad Sci U S A* **63**, 1319-26.

39. Perham, R. N. (2000). Swinging arms and swinging domains in multifunctional enzymes: catalytic machines for multistep reactions. *Annu Rev Biochem* **69**, 961-1004.
40. Welch, A. D. & Nichol, C. A. (1952). Water-soluble vitamins concerned with one and two-carbon intermediates. *Annu Rev Biochem* **21**, 633-86.
41. Bruice, T. C. & Hegarty, A. F. (1970). Biotin-bound CO_2 and the mechanism of enzymatic carboxylation reactions. *Proc Natl Acad Sci U S A* **65**, 805-9.
42. Perham, R. N. & Reche, P. A. (1998). Swinging arms in multifunctional enzymes and the specificity of post-translational modification. *Biochem Soc Trans* **26**, 299-303.
43. Lane, M. D. (2004). The biotin connection: Severo Ochoa, Harland Wood, and Feodor Lynen. *J Biol Chem* **279**, 39187-94.
44. Samols, D., Thornton, C. G., Murtif, V. L., Kumar, G. K., Haase, F. C. & Wood, H. G. (1988). Evolutionary conservation among biotin enzymes. *J Biol Chem* **263**, 6461-4.
45. Wood, H. G. & Barden, R. E. (1977). Biotin enzymes. *Annu Rev Biochem* **46**, 385-413.
46. Buckel, W. (2001). Sodium ion-translocating decarboxylases. *Biochim Biophys Acta* **1505**, 15-27.
47. Dimroth, P., Jockel, P. & Schmid, M. (2001). Coupling mechanism of the oxaloacetate decarboxylase $Na(^{+})$ pump. *Biochim Biophys Acta* **1505**, 1-14.
48. Gravel, R. A. & Narang, M. A. (2005). Molecular genetics of biotin metabolism: old vitamin, new science. *J Nutr Biochem* **16**, 428-31.
49. Pacheco-Alvarez, D., Solorzano-Vargas, R. S. & Del Rio, A. L. (2002). Biotin in metabolism and its relationship to human disease. *Arch Med Res* **33**, 439-47.
50. Kothapalli, N., Camporeale, G., Kueh, A., Chew, Y. C., Oommen, A. M., Griffin, J. B. & Zempleni, J. (2005). Biological functions of biotinylated histones. *J Nutr Biochem* **16**, 446-8.
51. Rodriguez-Melendez, R. & Zempleni, J. (2003). Regulation of gene expression by biotin (review). *J Nutr Biochem* **14**, 680-90.
52. Zempleni, J. (2005). Uptake, localization, and noncarboxylase roles of biotin. *Annu Rev Nutr* **25**, 175-96.
53. Gupta, A. K., Bluhm, R., Cooper, E. A., Summerbell, R. C. & Batra, R. (2003). Seborrheic dermatitis. *Dermatol Clin* **21**, 401-12.
54. McCarty, M. F. (2000). Toward practical prevention of type 2 diabetes. *Med Hypotheses* **54**, 786-93.
55. Reddi, A., DeAngelis, B., Frank, O., Lasker, N. & Baker, H. (1988). Biotin supplementation improves glucose and insulin tolerances in genetically diabetic KK mice. *Life Sci* **42**, 1323-30.
56. Ozand, P. T., Gascon, G. G., Al Essa, M., Joshi, S., Al Jishi, E., Bakheet, S., Al Watban, J., Al-Kawi, M. Z. & Dabbagh, O. (1998). Biotin-responsive basal ganglia disease: a novel entity. *Brain* **121 (Pt 7)**, 1267-79.
57. Zempleni, J. & Mock, D. M. (2000). Marginal biotin deficiency is teratogenic. *Proc Soc Exp Biol Med* **223**, 14-21.
58. Hytonen, V. P., Nyholm, T. K., Pentikainen, O. T., Vaarno, J., Porkka, E. J., Nordlund, H. R., Johnson, M. S., Slotte, J. P., Laitinen, O. H. & Kulomaa, M. S. (2004). Chicken avidin-related protein 4/5 shows superior thermal stability

when compared with avidin while retaining high affinity to biotin. *J Biol Chem* **279**, 9337-43.

59. Klumb, L. A., Chu, V. & Stayton, P. S. (1998). Energetic roles of hydrogen bonds at the ureido oxygen binding pocket in the streptavidin-biotin complex. *Biochemistry* **37**, 7657-63.
60. Cronan, J. E., Jr. & Reed, K. E. (2000). Biotinylation of proteins *in vivo*: a useful posttranslational modification for protein analysis. *Methods Enzymol* **326**, 440-58.
61. Chen, I., Howarth, M., Lin, W. & Ting, A. Y. (2005). Site-specific labeling of cell surface proteins with biophysical probes using biotin ligase. *Nat Methods* **2**, 99-104.
62. Gregolin, C., Ryder, E. & Lane, M. D. (1968). Liver acetyl coenzyme A carboxylase. I. Isolation and catalytic properties. *J Biol Chem* **243**, 4227-35.
63. Lane, M. D., Moss, J. & Polakis, S. E. (1974). Acetyl coenzyme A carboxylase. *Curr Top Cell Regul* **8**, 139-95.
64. Wakil, S. J., Stoops, J. K. & Joshi, V. C. (1983). Fatty acid synthesis and its regulation. *Annu Rev Biochem* **52**, 537-79.
65. Smith, S., Witkowski, A. & Joshi, A. K. (2003). Structural and functional organization of the animal fatty acid synthase. *Prog Lipid Res* **42**, 289-317.
66. Maier, T., Jenni, S. & Ban, N. (2006). Architecture of mammalian fatty acid synthase at 4.5 Å resolution. *Science* **311**, 1258-62.
67. Christie, W. W. (1973). *Lipid analysis.* Pergamon Press, Oxford.
68. Sanyal, I., Lee, S. L. & Flint, D. H. (1994). Biosynthesis of pimeloyl-CoA, a biotin precursor in *Escherichia coli*, follows a modified fatty acid synthesis pathway: ^{13}C-labeling studies. *J Am Chem Soc* ***116***, 2637-2638.
69. White, R. H. (1980). Stoichiometry and stereochemistry of deuterium incorporated into fatty acids by cells of *Escherichia coli* grown on [methyl-$^{2}H_3$]acetate. *Biochemistry* **19**, 9-15.
70. Hopwood, D. A. & Sherman, D. H. (1990). Molecular genetics of polyketides and its comparison to fatty acid biosynthesis. *Annu Rev Genet* **24**, 37-66.
71. Schujman, G. E., Paoletti, L., Grossman, A. D. & de Mendoza, D. (2003). FapR, a bacterial transcription factor involved in global regulation of membrane lipid biosynthesis. *Dev Cell* **4**, 663-72.
72. Schujman, G. E., Guerin, M., Buschiazzo, A., Schaeffer, F., Llarrull, L. I., Reh, G., Vila, A. J., Alzari, P. M. & de Mendoza, D. (2006). Structural basis of lipid biosynthesis regulation in Gram-positive bacteria. *Embo J* **25**, 4074-83.
73. Abu-Elheiga, L., Jayakumar, A., Baldini, A., Chirala, S. S. & Wakil, S. J. (1995). Human acetyl-CoA carboxylase: characterization, molecular cloning, and evidence for two isoforms. *Proc Natl Acad Sci U S A* **92**, 4011-5.
74. Munday, M. R. (2002). Regulation of mammalian acetyl-CoA carboxylase. *Biochem Soc Trans* **30**, 1059-64.
75. Abu-Elheiga, L., Almarza-Ortega, D. B., Baldini, A. & Wakil, S. J. (1997). Human acetyl-CoA carboxylase 2. Molecular cloning, characterization, chromosomal mapping, and evidence for two isoforms. *J Biol Chem* **272**, 10669-77.
76. Tong, L. (2005). Acetyl-coenzyme A carboxylase: crucial metabolic enzyme and attractive target for drug discovery. *Cell Mol Life Sci* **62**, 1784-803.

77. Abu-Elheiga, L., Matzuk, M. M., Abo-Hashema, K. A. & Wakil, S. J. (2001). Continuous fatty acid oxidation and reduced fat storage in mice lacking acetyl-CoA carboxylase 2. *Science* **291**, 2613-6.
78. Lenhard, J. M. & Gottschalk, W. K. (2002). Preclinical developments in type 2 diabetes. *Adv Drug Deliv Rev* **54**, 1199-212.
79. Harwood, H. J., Jr. (2004). Acetyl-CoA carboxylase inhibition for the treatment of metabolic syndrome. *Curr Opin Investig Drugs* **5**, 283-9.
80. Shi, Y. & Burn, P. (2004). Lipid metabolic enzymes: emerging drug targets for the treatment of obesity. *Nat Rev Drug Discov* **3**, 695-710.
81. Campbell, J. W. & Cronan, J. E., Jr. (2001). Bacterial fatty acid biosynthesis: targets for antibacterial drug discovery. *Annu Rev Microbiol* **55**, 305-32.
82. Heath, R. J., White, S. W. & Rock, C. O. (2001). Lipid biosynthesis as a target for antibacterial agents. *Prog Lipid Res* **40**, 467-97.
83. Freiberg, C., Brunner, N. A., Schiffer, G., Lampe, T., Pohlmann, J., Brands, M., Raabe, M., Habich, D. & Ziegelbauer, K. (2004). Identification and characterization of the first class of potent bacterial acetyl-CoA carboxylase inhibitors with antibacterial activity. *J Biol Chem* **279**, 26066-73.
84. Pohlmann, J., Lampe, T., Shimada, M., Nell, P. G., Pernerstorfer, J., Svenstrup, N., Brunner, N. A., Schiffer, G. & Freiberg, C. (2005). Pyrrolidinedione derivatives as antibacterial agents with a novel mode of action. *Bioorg Med Chem Lett* **15**, 1189-92.
85. Alban, C., Job, D. & Douce, R. (2000). Biotin metabolism in plants. *Annu Rev Plant Physiol Plant Mol Biol* **51**, 17-47.
86. Seng, T. W., Skillman, T. R., Yang, N. & Hammond, C. (2003). Cyclohexanedione herbicides are inhibitors of rat heart acetyl-CoA carboxylase. *Bioorg Med Chem Lett* **13**, 3237-42.
87. Zhang, H., Yang, Z., Shen, Y. & Tong, L. (2003). Crystal structure of the carboxyltransferase domain of acetyl-coenzyme A carboxylase. *Science* **299**, 2064-7.
88. Gerth, K., Pradella, S., Perlova, O., Beyer, S. & Muller, R. (2003). Myxobacteria: proficient producers of novel natural products with various biological activities-past and future biotechnological aspects with the focus on the genus Sorangium. *J Biotechnol* **106**, 233-53.
89. Alberts, A. W., Bell, R. M. & Vagelos, P. R. (1972). Acyl carrier protein. XV. Studies of -ketoacyl-acyl carrier protein synthetase. *J Biol Chem* **247**, 3190-8.
90. Cronan, J. E., Jr. & Waldrop, G. L. (2002). Multi-subunit acetyl-CoA carboxylases. *Prog Lipid Res* **41**, 407-35.
91. Alberts, A. W. & Vagelos, P. R. (1968). Acetyl CoA carboxylase. I. Requirement for two protein fractions. *Proc Natl Acad Sci U S A* **59**, 561-8.
92. Guchhait, R. B., Polakis, S. E., Dimroth, P., Stoll, E., Moss, J. & Lane, M. D. (1974). Acetyl coenzyme A carboxylase system of *Escherichia coli*. Purification and properties of the biotin carboxylase, carboxyltransferase, and carboxyl carrier protein components. *J Biol Chem* **249**, 6633-45.
93. Davis, M. S., Solbiati, J. & Cronan, J. E., Jr. (2000). Overproduction of acetyl-CoA carboxylase activity increases the rate of fatty acid biosynthesis in *Escherichia coli*. *J Biol Chem* **275**, 28593-8.

94. Soriano, A., Radice, A. D., Herbitter, A. H., Langsdorf, E. F., Stafford, J. M., Chan, S., Wang, S., Liu, Y. H. & Black, T. A. (2006). *Escherichia coli* acetyl-coenzyme A carboxylase: characterization and development of a high-throughput assay. *Anal Biochem* **349**, 268-76.
95. Alix, J. H. (1989). A rapid procedure for cloning genes from lambda libraries by complementation of *E. coli* defective mutants: application to the fabE region of the *E. coli* chromosome. *DNA* **8**, 779-89.
96. Mizuno, T. (1987). Random cloning of bent DNA segments from *Escherichia coli* chromosome and primary characterization of their structures. *Nucleic Acids Res* **15**, 6827-41.
97. Li, S. J. & Cronan, J. E., Jr. (1992). The gene encoding the biotin carboxylase subunit of *Escherichia coli* acetyl-CoA carboxylase. *J Biol Chem* **267**, 855-63.
98. James, E. S. & Cronan, J. E. (2004). Expression of two *Escherichia coli* acetyl-CoA carboxylase subunits is autoregulated. *J Biol Chem* **279**, 2520-7.
99. Kondo, H., Shiratsuchi, K., Yoshimoto, T., Masuda, T., Kitazono, A., Tsuru, D., Anai, M., Sekiguchi, M. & Tanabe, T. (1991). Acetyl-CoA carboxylase from *Escherichia coli*: gene organization and nucleotide sequence of the biotin carboxylase subunit. *Proc Natl Acad Sci U S A* **88**, 9730-3.
100. Li, S. J. & Cronan, J. E., Jr. (1992). The genes encoding the two carboxyltransferase subunits of *Escherichia coli* acetyl-CoA carboxylase. *J Biol Chem* **267**, 16841-7.
101. Choi, J. K., Yu, F., Wurtele, E. S. & Nikolau, B. J. (1995). Molecular cloning and characterization of the cDNA coding for the biotin-containing subunit of the chloroplastic acetyl-coenzyme A carboxylase. *Plant Physiol* **109**, 619-25.
102. Harwood, J. L. (1996). Recent advances in the biosynthesis of plant fatty acids. *Biochim Biophys Acta* **1301**, 7-56.
103. Sasaki, Y. & Nagano, Y. (2004). Plant acetyl-CoA carboxylase: structure, biosynthesis, regulation, and gene manipulation for plant breeding. *Biosci Biotechnol Biochem* **68**, 1175-84.
104. Reith, M. & Munholland, J. (1995). Complete nucleotide sequence of the *Porphyra purpurea* chloroplast genome. *Plant Mol Biol Rep* **13**, 333-5.
105. Polakis, S. E., Guchhait, R. B., Zwergel, E. E., Lane, M. D. & Cooper, T. G. (1974). Acetyl coenzyme A carboxylase system of *Escherichia coli*. Studies on the mechanisms of the biotin carboxylase- and carboxyltransferase-catalyzed reactions. *J Biol Chem* **249**, 6657-67.
106. Nervi, A. M., Alberts, A. W. & Vagelos, P. R. (1971). Acetyl CoA carboxylase. 3. Purification and properties of a biotin carboxyl carrier protein. *Arch Biochem Biophys* **143**, 401-11.
107. Fall, R. R., Nervi, A. M., Alberts, A. W. & Vagelos, P. R. (1971). Acetyl CoA carboxylase: isolation and characterization of native biotin carboxyl carrier protein. *Proc Natl Acad Sci U S A* **68**, 1512-5.
108. Nenortas, E. & Beckett, D. (1996). Purification and characterization of intact and truncated forms of the *Escherichia coli* biotin carboxyl carrier subunit of acetyl-CoA carboxylase. *J Biol Chem* **271**, 7559-67.
109. Fall, R. R. & Vagelos, P. R. (1975). Biotin carboxyl carrier protein from *Escherichia coli*. *Methods Enzymol* **35**, 17-25.

110. Chapman-Smith, A., Turner, D. L., Cronan, J. E., Jr., Morris, T. W. & Wallace, J. C. (1994). Expression, biotinylation and purification of a biotin-domain peptide from the biotin carboxy carrier protein of *Escherichia coli* acetyl-CoA carboxylase. *Biochem J* **302 (Pt 3)**, 881-7.
111. Reche, P., Li, Y. L., Fuller, C., Eichhorn, K. & Perham, R. N. (1998). Selectivity of post-translational modification in biotinylated proteins: the carboxy carrier protein of the acetyl-CoA carboxylase of *Escherichia coli*. *Biochem J* **329 (Pt 3)**, 589-96.
112. Reddy, D. V., Rothemund, S., Shenoy, B. C., Carey, P. R. & Sonnichsen, F. D. (1998). Structural characterization of the entire 1.3S subunit of transcarboxylase from *Propionibacterium shermanii*. *Protein Sci* **7**, 2156-63.
113. Reddy, D. V., Shenoy, B. C., Carey, P. R. & Sonnichsen, F. D. (2000). High resolution solution structure of the 1.3S subunit of transcarboxylase from *Propionibacterium shermanii*. *Biochemistry* **39**, 2509-16.
114. Chapman-Smith, A., Forbes, B. E., Wallace, J. C. & Cronan, J. E., Jr. (1997). Covalent modification of an exposed surface turn alters the global conformation of the biotin carrier domain of *Escherichia coli* acetyl-CoA carboxylase. *J Biol Chem* **272**, 26017-22.
115. Athappilly, F. K. & Hendrickson, W. A. (1995). Structure of the biotinyl domain of acetyl-coenzyme A carboxylase determined by MAD phasing. *Structure* **3**, 1407-19.
116. Reche, P. & Perham, R. N. (1999). Structure and selectivity in post-translational modification: attaching the biotinyl-lysine and lipoyl-lysine swinging arms in multifunctional enzymes. *Embo J* **18**, 2673-82.
117. Yao, X., Wei, D., Soden, C., Jr., Summers, M. F. & Beckett, D. (1997). Structure of the carboxy-terminal fragment of the apo-biotin carboxyl carrier subunit of *Escherichia coli* acetyl-CoA carboxylase. *Biochemistry* **36**, 15089-100.
118. Roberts, E. L., Shu, N., Howard, M. J., Broadhurst, R. W., Chapman-Smith, A., Wallace, J. C., Morris, T., Cronan, J. E., Jr. & Perham, R. N. (1999). Solution structures of apo and holo biotinyl domains from acetyl coenzyme A carboxylase of *Escherichia coli* determined by triple-resonance nuclear magnetic resonance spectroscopy. *Biochemistry* **38**, 5045-53.
119. Yao, X., Soden, C., Jr., Summers, M. F. & Beckett, D. (1999). Comparison of the backbone dynamics of the apo- and holo-carboxy-terminal domain of the biotin carboxyl carrier subunit of *Escherichia coli* acetyl-CoA carboxylase. *Protein Sci* **8**, 307-17.
120. Blanchard, C. Z., Chapman-Smith, A., Wallace, J. C. & Waldrop, G. L. (1999). The biotin domain peptide from the biotin carboxyl carrier protein of *Escherichia coli* acetyl-CoA carboxylase causes a marked increase in the catalytic efficiency of biotin carboxylase and carboxyltransferase relative to free biotin. *J Biol Chem* **274**, 31767-9.
121. Solbiati, J., Chapman-Smith, A. & Cronan, J. E., Jr. (2002). Stabilization of the biotinoyl domain of *Escherichia coli* acetyl-CoA carboxylase by interactions between the attached biotin and the protruding "thumb" structure. *J Biol Chem* **277**, 21604-9.

122. Cronan, J. E., Jr. (2001). The biotinyl domain of *Escherichia coli* acetyl-CoA carboxylase. Evidence that the "thumb" structure is essential and that the domain functions as a dimer. *J Biol Chem* **276**, 37355-64.
123. Ogita, T. & Knowles, J. R. (1988). On the intermediacy of carboxyphosphate in biotin-dependent carboxylations. *Biochemistry* **27**, 8028-33.
124. Tipton, P. A. & Cleland, W. W. (1988). Catalytic mechanism of biotin carboxylase: steady-state kinetic investigations. *Biochemistry* **27**, 4317-25.
125. Kaziro, Y., Hass, L. F., Boyer, P. D. & Ochoa, S. (1962). Mechanism of the propionyl carboxylase reaction. II. Isotopic exchange and tracer experiments. *J Biol Chem* **237**, 1460-8.
126. Climent, I. & Rubio, V. (1986). ATPase activity of biotin carboxylase provides evidence for initial activation of HCO^{3-} by ATP in the carboxylation of biotin. *Arch Biochem Biophys* **251**, 465-70.
127. Blanchard, C. Z., Lee, Y. M., Frantom, P. A. & Waldrop, G. L. (1999). Mutations at four active site residues of biotin carboxylase abolish substrate-induced synergism by biotin. *Biochemistry* **38**, 3393-400.
128. Bruice, T. C. & Benkovic, S. J. (2000). Chemical basis for enzyme catalysis. *Biochemistry* **39**, 6267-74.
129. Kluger, R., Davis, P. P. & Adawadkar, P. D. (1979). Mechanism of Urea Participation in Phosphonate Ester Hydrolysis. Mechanistic and Stereochemical Criteria for Enzymic Formation and Reaction of Phosphorylated Biotin. *J Am Chem Soc* **101**, 5995-6000.
130. King, S. W., Natarajan, R. Bembi, R. & Fife T. H. (1992). Intramolecular ureido and amide group participation in reactions of carbonate diesters. *J Am Chem Soc* **114**, 10715-10721.
131. Tipton, P. A. & Cleland, W. W. (1988). Carbon-13 and deuterium isotope effects on the catalytic reactions of biotin carboxylase. *Biochemistry* **27**, 4325-31.
132. Raushel, F. M., Mullins, L. S. & Gibson, G. E. (1998). A stringent test for the nucleotide switch mechanism of carbamoyl phosphate synthetase. *Biochemistry* **37**, 10272-8.
133. Perrin, C. L. D., T.J. (1987). Proton Exchange in Biotin: A Reinvestigation, with Implications for the Mechanism of CO_2 Transfer. *J Am Chem Soc* **109**, 5163-67.
134. Waldrop, G. L., Rayment, I. & Holden, H. M. (1994). Three-dimensional structure of the biotin carboxylase subunit of acetyl-CoA carboxylase. *Biochemistry* **33**, 10249-56.
135. Thoden, J. B., Blanchard, C. Z., Holden, H. M. & Waldrop, G. L. (2000). Movement of the biotin carboxylase B-domain as a result of ATP binding. *J Biol Chem* **275**, 16183-90.
136. Artymiuk, P. J., Poirrette, A. R., Rice, D. W. & Willett, P. (1996). Biotin carboxylase comes into the fold. *Nat Struct Biol* **3**, 128-32.
137. Shen, Y., Volrath, S. L., Weatherly, S. C., Elich, T. D. & Tong, L. (2004). A mechanism for the potent inhibition of eukaryotic acetyl-coenzyme A carboxylase by soraphen A, a macrocyclic polyketide natural product. *Mol Cell* **16**, 881-91.

138. Weatherly, S. C., Volrath, S. L. & Elich, T. D. (2004). Expression and characterization of recombinant fungal acetyl-CoA carboxylase and isolation of a soraphen-binding domain. *Biochem J* **380**, 105-10.
139. Blanchard, C. Z. & Waldrop, G. L. (1998). Overexpression and kinetic characterization of the carboxyltransferase component of acetyl-CoA carboxylase. *J Biol Chem* **273**, 19140-5.
140. Levert, K. L. & Waldrop, G. L. (2002). A bisubstrate analog inhibitor of the carboxyltransferase component of acetyl-CoA carboxylase. *Biochem Biophys Res Commun* **291**, 1213-7.
141. Attwood, P. V. (1995). The structure and the mechanism of action of pyruvate carboxylase. *Int J Biochem Cell Biol* **27**, 231-49.
142. Bilder, P., Lightle, S., Bainbridge, G., Ohren, J., Finzel, B., Sun, F., Holley, S., Al-Kassim, L., Spessard, C., Melnick, M., Newcomer, M. & Waldrop, G. L. (2006). The structure of the carboxyltransferase component of acetyl-coA carboxylase reveals a zinc-binding motif unique to the bacterial enzyme. *Biochemistry* **45**, 1712-22.
143. Holden, H. M., Benning, M. M., Haller, T. & Gerlt, J. A. (2001). The crotonase superfamily: divergently related enzymes that catalyze different reactions involving acyl coenzyme a thioesters. *Ac Chem Res* **34**, 145-57.
144. Gerlt, J. A. & Babbitt, P. C. (2001). Divergent evolution of enzymatic function: mechanistically diverse superfamilies and functionally distinct suprafamilies. *Annu Rev Biochem* **70**, 209-46.
145. Attwood, P. V. & Cleland, W. W. (1986). Decarboxylation of oxalacetate by pyruvate carboxylase. *Biochemistry* **25**, 8191-6.
146. Zhang, H., Tweel, B. & Tong, L. (2004). Molecular basis for the inhibition of the carboxyltransferase domain of acetyl-coenzyme-A carboxylase by haloxyfop and diclofop. *Proc Natl Acad Sci U S A* **101**, 5910-5.
147. Zhang, H., Tweel, B., Li, J. & Tong, L. (2004). Crystal structure of the carboxyltransferase domain of acetyl-coenzyme A carboxylase in complex with CP-640186. *Structure* **12**, 1683-91.
148. Wendt, K. S., Schall, I., Huber, R., Buckel, W. & Jacob, U. (2003). Crystal structure of the carboxyltransferase subunit of the bacterial sodium ion pump glutaconyl-coenzyme A decarboxylase. *Embo J* **22**, 3493-502.
149. Diacovich, L., Mitchell, D. L., Pham, H., Gago, G., Melgar, M. M., Khosla, C., Gramajo, H. & Tsai, S. C. (2004). Crystal structure of the beta-subunit of acyl-CoA carboxylase: structure-based engineering of substrate specificity. *Biochemistry* **43**, 14027-36.
150. Secor, J. & Cseke, C. (1988). Inhibition of acetyl-CoA carboxylase activity by haloxyfop and tralkoxydim. *Plant Physiol* **86**, 10-12.
151. Harwood, H. J., Jr., Petras, S. F., Shelly, L. D., Zaccaro, L. M., Perry, D. A., Makowski, M. R., Hargrove, D. M., Martin, K. A., Tracey, W. R., Chapman, J. G., Magee, W. P., Dalvie, D. K., Soliman, V. F., Martin, W. H., Mularski, C. J. & Eisenbeis, S. A. (2003). Isozyme-nonselective N-substituted bipiperidylcarboxamide acetyl-CoA carboxylase inhibitors reduce tissue malonyl-CoA concentrations, inhibit fatty acid synthesis, and increase fatty acid oxidation in cultured cells and in experimental animals. *J Biol Chem* **278**, 37099-111.

152. Raushel, F. M., Thoden, J. B. & Holden, H. M. (2003). Enzymes with molecular tunnels. *Acc Chem Res* **36**, 539-48.
153. Brocklehurst, S. M. & Perham, R. N. (1993). Prediction of the three-dimensional structures of the biotinylated domain from yeast pyruvate carboxylase and of the lipoylated H-protein from the pea leaf glycine cleavage system: a new automated method for the prediction of protein tertiary structure. *Protein Sci* **2**, 626-39.
154. Jitrapakdee, S. & Wallace, J. C. (2003). The biotin enzyme family: conserved structural motifs and domain rearrangements. *Curr Protein Pept Sci* **4**, 217-29.
155. Cronan, J. E., Jr. (2002). Interchangeable enzyme modules. Functional replacement of the essential linker of the biotinylated subunit of acetyl-CoA carboxylase with a linker from the lipoylated subunit of pyruvate dehydrogenase. *J Biol Chem* **277**, 22520-7.
156. Choi-Rhee, E. & Cronan, J. E. (2003). The biotin carboxylase-biotin carboxyl carrier protein complex of *Escherichia coli* acetyl-CoA carboxylase. *J Biol Chem* **278**, 30806-12.
157. Guchhait, R. B., Polakis, S. E. & Lane, M. D. (1975). Biotin carboxylase component of acetyl-CoA carboxylase from *Escherichia coli*. *Methods Enzymol* **35**, 25-31.
158. Janiyani, K., Bordelon, T., Waldrop, G. L. & Cronan, J. E., Jr. (2001). Function of *Escherichia coli* biotin carboxylase requires catalytic activity of both subunits of the homodimer. *J Biol Chem* **276**, 29864-70.
159. Shen, Y., Chou, C. Y., Chang, G. G. & Tong, L. (2006). Is dimerization required for the catalytic activity of bacterial biotin carboxylase? *Mol Cell* **22**, 807-18.
160. Guchhait, R. B., Polakis, S. E. & Lane, M. D. (1975). Carboxyltransferase component of acetyl-CoA carboxylase from *Escherichia coli*. *Methods Enzymol* **35**, 32-7.
161. Chapman-Smith, A. & Cronan, J. E., Jr. (1999). The enzymatic biotinylation of proteins: a post-translational modification of exceptional specificity. *Trends Biochem Sci* **24**, 359-63.
162. Lane, M. D., Rominger, K. L., Young, D. L. & Lynen, F. (1964). The enzymatic synthesis of holotranscarboxylase from apotranscarboxylase and (+)-Biotin. Investigation of the reaction mechanism. *J Biol Chem* **239**, 2865-71.
163. Lane, M. D., Young, D. L. & Lynen, F. (1964). The enzymatic synthesis of holotranscarboxylase from apotranscarboxylase and (+)-Biotin. I. purification of the apoenzyme and synthetase; characteristics of the reaction. *J Biol Chem* **239**, 2858-64.
164. Jakubowski, H. & Goldman, E. (1992). Editing of errors in selection of amino acids for protein synthesis. *Microbiol Rev* **56**, 412-29.
165. Schimmel, P. (1987). Aminoacyl tRNA synthetases: general scheme of structure-function relationships in the polypeptides and recognition of transfer RNAs. *Annu Rev Biochem* **56**, 125-58.
166. Green, D. E., Morris, T. W., Green, J., Cronan, J. E., Jr. & Guest, J. R. (1995). Purification and properties of the lipoate protein ligase of *Escherichia coli*. *Biochem J* **309 (Pt 3)**, 853-62.

167. Morris, T. W., Reed, K. E. & Cronan, J. E., Jr. (1995). Lipoic acid metabolism in *Escherichia coli*: the lplA and lipB genes define redundant pathways for ligation of lipoyl groups to apoprotein. *J Bacteriol* **177**, 1-10.
168. Fall, R. R. (1979). Analysis of microbial biotin proteins. *Methods Enzymol* **62**, 390-8.
169. Leon-Del-Rio, A. & Gravel, R. (1994). Sequence requirements for the biotinylation of carboxyl-terminal fragments of human propionyl-CoA carboxylase alpha subunit expressed in *Escherichia coli*. *J Biol Chem* **269**, 22964-22968.
170. Polyak, S. W., Chapman-Smith, A., Brautigan, P. J. & Wallace, J. C. (1999). Biotin protein ligase from *Saccharomyces cerevisiae*. The N-terminal domain is required for complete activity. *J Biol Chem* **274**, 32847-54.
171. Barker, D. F. & Campbell, A. M. (1980). Use of bio-lac fusion strains to study regulation of biotin biosynthesis in *Escherichia coli*. *J Bacteriol* **143**, 789-800.
172. Suzuki, Y., Yang, X., Aoki, Y., Kure, S. & Matsubara, Y. (2005). Mutations in the holocarboxylase synthetase gene HLCS. *Hum Mutat* **26**, 285-90.
173. Rodionov, D. A., Mironov, A. A. & Gelfand, M. S. (2002). Conservation of the biotin regulon and the BirA regulatory signal in Eubacteria and Archaea. *Genome Res* **12**, 1507-16.
174. Mukhopadhyay, B., Purwantini, E., Kreder, C. L. & Wolfe, R. S. (2001). Oxaloacetate synthesis in the methanarchaeon *Methanosarcina barkeri*: pyruvate carboxylase genes and a putative *Escherichia coli*-type bifunctional biotin protein ligase gene (bpl/birA) exhibit a unique organization. *J Bacteriol* **183**, 3804-10.
175. Bower, S., Perkins, J., Yocum, R. R., Serror, P., Sorokin, A., Rahaim, P., Howitt, C. L., Prasad, N., Ehrlich, S. D. & Pero, J. (1995). Cloning and characterization of the *Bacillus subtilis* birA gene encoding a repressor of the biotin operon. *J Bacteriol* **177**, 2572-5.
176. Cronan, J. E., Jr. (1989). The *E. coli* bio operon: transcriptional repression by an essential protein modification enzyme. *Cell* **58**, 427-9.
177. Beckett, D. (2007). Biotin sensing: universal influence of biotin status on transcription. *Annu Rev Genet* **41**, 443-64.
178. Clarke, D. J., Coulson, J., Baillie, R. & Campopiano, D. J. (2003). Biotinylation in the hyperthermophile *Aquifex aeolicus*. *Eur J Biochem* **270**, 1277-87.
179. Bagautdinov, B., Kuroishi, C., Sugahara, M. & Kunishima, N. (2005). Crystal structures of biotin protein ligase from *Pyrococcus horikoshii* OT3 and its complexes: structural basis of biotin activation. *J Mol Biol* **353**, 322-33.
180. Barker, D. F. & Campbell, A. M. (1981). The birA gene of *Escherichia coli* encodes a biotin holoenzyme synthetase. *J Mol Biol* **146**, 451-67.
181. Barker, D. F. & Campbell, A. M. (1981). Genetic and biochemical characterization of the birA gene and its product: evidence for a direct role of biotin holoenzyme synthetase in repression of the biotin operon in *Escherichia coli*. *J Mol Biol* **146**, 469-92.
182. Eisenberg, M. A., Prakash, O. & Hsiung, S. C. (1982). Purification and properties of the biotin repressor. A bifunctional protein. *J Biol Chem* **257**, 15167-73.

183. Prakash, O. & Eisenberg, M. A. (1979). Biotinyl 5'-adenylate: corepressor role in the regulation of the biotin genes of *Escherichia coli* K-12. *Proc Natl Acad Sci U S A* **76**, 5592-5.
184. Otsuka, A. & Abelson, J. (1978). The regulatory region of the biotin operon in *Escherichia coli*. *Nature* **276**, 689 - 694
185. Eisenstein, E. & Beckett, D. (1999). Dimerization of the *Escherichia coli* biotin repressor: corepressor function in protein assembly. *Biochemistry* **38**, 13077-84.
186. Streaker, E. D., Gupta, A. & Beckett, D. (2002). The biotin repressor: thermodynamic coupling of corepressor binding, protein assembly, and sequence-specific DNA binding. *Biochemistry* **41**, 14263-71.
187. Weaver, L. H., Kwon, K., Beckett, D. & Matthews, B. W. (2001). Corepressor-induced organization and assembly of the biotin repressor: a model for allosteric activation of a transcriptional regulator. *Proc Natl Acad Sci U S A* **98**, 6045-50.
188. Wood, Z. A., Weaver, L. H., Brown, P. H., Beckett, D. & Matthews, B. W. (2006). Co-repressor induced order and biotin repressor dimerization: a case for divergent followed by convergent evolution. *J Mol Biol* **357**, 509-23.
189. Kwon, K., Streaker, E. D., Ruparelia, S. & Beckett, D. (2000). Multiple disordered loops function in corepressor-induced dimerization of the biotin repressor. *J Mol Biol* **304**, 821-33.
190. Streaker, E. D. & Beckett, D. (2003). Coupling of protein assembly and DNA binding: biotin repressor dimerization precedes biotin operator binding. *J Mol Biol* **325**, 937-48.
191. Streaker, E. D. & Beckett, D. (2006). The biotin regulatory system: kinetic control of a transcriptional switch. *Biochemistry* **45**, 6417-25.
192. Beckett, D. & Matthews, B. W. (1997). *Escherichia coli* repressor of biotin biosynthesis. *Methods Enzymol* **279**, 362-76.
193. Xu, Y. & Beckett, D. (1997). Biotinyl-5'-adenylate synthesis catalyzed by *Escherichia coli* repressor of biotin biosynthesis. *Methods Enzymol* **279**, 405-21.
194. Wilson, K. P., Shewchuk, L. M., Brennan, R. G., Otsuka, A. J. & Matthews, B. W. (1992). *Escherichia coli* biotin holoenzyme synthetase/bio repressor crystal structure delineates the biotin- and DNA-binding domains. *Proc Natl Acad Sci U S A* **89**, 9257-61.
195. Brown, P. H., Cronan, J. E., Grotli, M. & Beckett, D. (2004). The biotin repressor: modulation of allostery by corepressor analogs. *J Mol Biol* **337**, 857-69.
196. Bellamacina, C. R. (1996). The nicotinamide dinucleotide binding motif: a comparison of nucleotide binding proteins. *Faseb J* **10**, 1257-69.
197. Noble, M. E., Musacchio, A., Saraste, M., Courtneidge, S. A. & Wierenga, R. K. (1993). Crystal structure of the SH3 domain in human Fyn; comparison of the three-dimensional structures of SH3 domains in tyrosine kinases and spectrin. *Embo J* **12**, 2617-24.
198. Sparks, A. B., Rider, J. E., Hoffman, N. G., Fowlkes, D. M., Quillam, L. A. & Kay, B. K. (1996). Distinct ligand preferences of Src homology 3 domains from Src, Yes, Abl, Cortactin, p53bp2, PLCgamma, Crk, and Grb2. *Proc Natl Acad Sci U S A* **93**, 1540-4.

199. Safro, M. & Mosyak, L. (1995). Structural similarities in the noncatalytic domains of phenylalanyl-tRNA and biotin synthetases. *Protein Sci* **4**, 2429-32.
200. Streaker, E. D. & Beckett, D. (1999). Ligand-linked structural changes in the *Escherichia coli* biotin repressor: the significance of surface loops for binding and allostery. *J Mol Biol* **292**, 619-32.
201. Chapman-Smith, A., Mulhern, T. D., Whelan, F., Cronan, J. E., Jr. & Wallace, J. C. (2001). The C-terminal domain of biotin protein ligase from *E. coli* is required for catalytic activity. *Protein Sci* **10**, 2608-17.
202. Naganathan, S. & Beckett, D. (2007). Nucleation of an allosteric response *via* ligand-induced loop folding. *J Mol Biol* **373**, 96-111.
203. Kwon, K., Streaker, E. D. & Beckett, D. (2002). Binding specificity and the ligand dissociation process in the *E. coli* biotin holoenzyme synthetase. *Protein Sci* **11**, 558-70.
204. Kwon, K. & Beckett, D. (2000). Function of a conserved sequence motif in biotin holoenzyme synthetases. *Protein Sci* **9**, 1530-9.
205. Choi-Rhee, E., Schulman, H. & Cronan, J. E. (2004). Promiscuous protein biotinylation by *Escherichia coli* biotin protein ligase. *Protein Sci* **13**, 3043-50.
206. Cronan, J. E. (2005). Targeted and proximity-dependent promiscuous protein biotinylation by a mutant *Escherichia coli* biotin protein ligase. *J Nutr Biochem* **16**, 416-8.
207. Dupuis, L., Leon-Del-Rio, A., Leclerc, D., Campeau, E., Sweetman, L., Saudubray, J. M., Herman, G., Gibson, K. M. & Gravel, R. A. (1996). Clustering of mutations in the biotin-binding region of holocarboxylase synthetase in biotin-responsive multiple carboxylase deficiency. *Hum Mol Genet* **5**, 1011-6.
208. Chen, Y. & Varani, G. (2005). Protein families and RNA recognition. *Febs J* **272**, 2088-97.
209. Delarue, M., Poterszman, A., Nikonov, S., Garber, M., Moras, D. & Thierry, J. C. (1994). Crystal structure of a prokaryotic aspartyl tRNA-synthetase. *Embo J* **13**, 3219-29.
210. Fujiwara, K., Toma, S., Okamura-Ikeda, K., Motokawa, Y., Nakagawa, A. & Taniguchi, H. (2005). Crystal structure of lipoate-protein ligase A from *Escherichia coli*. Determination of the lipoic acid-binding site. *J Biol Chem* **280**, 33645-51.
211. Kim, D. J., Kim, K. H., Lee, H. H., Lee, S. J., Ha, J. Y., Yoon, H. J. & Suh, S. W. (2005). Crystal structure of lipoate-protein ligase A bound with the activated intermediate: insights into interaction with lipoyl domains. *J Biol Chem* **280**, 38081-9.
212. McManus, E., Luisi, B. F. & Perham, R. N. (2006). Structure of a putative lipoate protein ligase from *Thermoplasma acidophilum* and the mechanism of target selection for post-translational modification. *J Mol Biol* **356**, 625-37.
213. Morris, T. W., Reed, K. E. & Cronan, J. E., Jr. (1994). Identification of the gene encoding lipoate-protein ligase A of *Escherichia coli*. Molecular cloning and characterization of the lplA gene and gene product. *J Biol Chem* **269**, 16091-100.

214. Artymiuk, P. J., Rice, D. W., Poirrette, A. R. & Willet, P. (1994). A tale of two synthetases. *Nat Struct Biol* **1**, 758-60.
215. Klipcan, L. & Safro, M. (2004). Amino acid biogenesis, evolution of the genetic code and aminoacyl-tRNA synthetases. *J Theor Biol* **228**, 389-96.
216. Eriani, G., Delarue, M., Poch, O., Gangloff, J. & Moras, D. (1990). Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. *Nature* **347**, 203-6.
217. Pan, F., Lo, K. Y., Pai, S. H. & Lee, H. H. (1982). Kinetic mechanism of threonyl-tRNA synthetase from human placenta. *Int J Pept Protein Res* **20**, 159-66.
218. Weaver, L. H., Kwon, K., Beckett, D. & Matthews, B. W. (2001). Competing protein:protein interactions are proposed to control the biological switch of the *E coli* biotin repressor. *Protein Sci* **10**, 2618-22.
219. Reche, P. A., Howard, M. J., Broadhurst, R. W. & Perham, R. N. (2000). Heteronuclear NMR studies of the specificity of the post-translational modification of biotinyl domains by biotinyl protein ligase. *FEBS Lett* **479**, 93-8.
220. Bagautdinov, B., Kuroishi, C., Sugahara, M. & Kunishima, N. (2005). Purification, crystallization and preliminary crystallographic analysis of the biotin-protein ligase from *Pyrococcus horikoshii* OT3. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **61**, 193-5.
221. Bagautdinov, B., Matsuura, Y., Bagautdinova, S. & Kunishima, N. (2007). Crystallization and preliminary X-ray crystallographic studies of the biotin carboxyl carrier protein and biotin protein ligase complex from *Pyrococcus horikoshii* OT3. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **63**, 334-7.
222. Cho, A. & Normile, D. (2002). Nobel Prize in Chemistry. Mastering macromolecules. *Science* **298**, 527-8.
223. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C. M. (1989). Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**, 64-71.
224. Karas, M. & Hillenkamp, F. (1988). Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* **60**, 2299-301.
225. Aebersold, R. & Mann, M. (2003). Mass spectrometry-based proteomics. *Nature* **422**, 198-207.
226. Hamdan, M. & Righetti, P. G. (2002). Modern strategies for protein quantification in proteome analysis: advantages and limitations. *Mass Spectrom Rev* **21**, 287-302.
227. Mo, W. & Karger, B. L. (2002). Analytical aspects of mass spectrometry and proteomics. *Curr Opin Chem Biol* **6**, 666-75.
228. Loo, J. A. (1997). Studying noncovalent protein complexes by electrospray ionization mass spectrometry. *Mass Spectrom Rev* **16**, 1-23.
229. Hernandez, H. & Robinson, C. V. (2001). Dynamic protein complexes: insights from mass spectrometry. *J Biol Chem* **276**, 46685-8.
230. Heck, A. J. & Van Den Heuvel, R. H. (2004). Investigation of intact protein complexes by mass spectrometry. *Mass Spectrom Rev* **23**, 368-89.
231. Kienitz, H. E. (1968). *in Massenspektrometrie.* Verlag Chemie, Weinheim.

232. Benesch, J. L., Ruotolo, B. T., Simmons, D. A. & Robinson, C. V. (2007). Protein complexes in the gas phase: technology for structural genomics and proteomics. *Chem Rev* **107**, 3544-67.
233. Strupat, K., Sagi, D., Bonisch, H., Schafer, G. & Peter-Katalinic, J. (2000). Oligomerization and substrate binding studies of the adenylate kinase from *Sulfolobus acidocaldarius* by matrix-assisted laser desorption/ionization mass spectrometry. *Analyst* **125**, 563–567.
234. Kiselar, J. G. & Downard, K. M. (2000). Preservation and detection of specific antibody-peptide complexes by matrix-assisted laser desorption ionization mass spectrometry. *J Am Soc Mass Spectrom* **11**, 746-50.
235. Wattenberg, A., Sobott, F., Barth, H. D. & Brutschy, B. (2000). Studying noncovalent protein complexes in aqueous solution with laser desorption mass spectrometry. *Int J Mass Spectrom* **203**, 49–57.
236. Kebarle, P. (2000). A brief overview of the present status of the mechanisms involved in electrospray mass spectrometry. *J Mass Spectrom* **35**, 804-17.
237. Wilm, M. S. & Mann, M. (1994). Electrospray and Taylor-Cone theory, Dole's beam of macromolecules at last? *Int J Mass Spectrom Ion Processes* **136**, 167-180.
238. Wilm, M. & Mann, M. (1996). Analytical properties of the nanoelectrospray ion source. *Anal Chem* **68**, 1-8.
239. El-Faramawy, A., Siu, K. W. & Thomson, B. A. (2005). Efficiency of nano-electrospray ionization. *J Am Soc Mass Spectrom* **16**, 1702-7.
240. Tang, X. J., Brewer, C. F., Saha, S., Chernushevich, I., Ens, W. & Standing, K. G. (1994). Investigation of protein-protein noncovalent interactions in *soybean agglutinin* by electrospray ionization time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom* **8**, 750-4.
241. Morris, H. R., Paxton, T., Dell, A., Langhorne, J., Berg, M., Bordoli, R. S., Hoyes, J. & Bateman, R. H. (1996). High sensitivity collisionally-activated decomposition tandem mass spectrometry on a novel quadrupole/orthogonal-acceleration time-of-flight mass spectrometer. *Rapid Commun Mass Spectrom* **10**, 889-96.
242. Chernushevich, I. V. & Thomson, B. A. (2004). Collisional cooling of large ions in electrospray mass spectrometry. *Anal Chem* **76**, 1754-60.
243. Sobott, F., Hernandez, H., McCammon, M. G., Tito, M. A. & Robinson, C. V. (2002). A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies. *Anal Chem* **74**, 1402-7.
244. van den Heuvel, R. H., van Duijn, E., Mazon, H., Synowsky, S. A., Lorenzen, K., Versluis, C., Brouns, S. J., Langridge, D., van der Oost, J., Hoyes, J. & Heck, A. J. (2006). Improving the performance of a quadrupole time-of-flight instrument for macromolecular mass spectrometry. *Anal Chem* **78**, 7473-83.
245. Fitzgerald, M. C., Chernushevich, I., Standing, K. G., Whitman, C. P. & Kent, S. B. (1996). Probing the oligomeric structure of an enzyme by electrospray ionization time-of-flight mass spectrometry. *Proc Natl Acad Sci U S A* **93**, 6851-6.
246. McCammon, M. G., Scott, D. J., Keetch, C. A., Greene, L. H., Purkey, H. E., Petrassi, H. M., Kelly, J. W. & Robinson, C. V. (2002). Screening transthyretin amyloid fibril inhibitors: characterization of novel multiprotein, multiligand complexes by mass spectrometry. *Structure* **10**, 851-63.

247. Benesch, J. L. & Robinson, C. V. (2006). Mass spectrometry of macromolecular assemblies: preservation and dissociation. *Curr Opin Struct Biol* **16**, 245-51.
248. Hernandez, H., Dziembowski, A., Taverner, T., Seraphin, B. & Robinson, C. V. (2006). Subunit architecture of multimeric complexes isolated directly from cells. *EMBO Rep* **7**, 605-10.
249. Benesch, J. L., Aquilina, J. A., Ruotolo, B. T., Sobott, F. & Robinson, C. V. (2006). Tandem mass spectrometry reveals the quaternary organization of macromolecular assemblies. *Chem Biol* **13**, 597-605.
250. Senko, M. W. & McLafferty, F. W. (1994). Mass spectrometry of macromolecules: has its time now come? *Annu Rev Biophys Biomol Struct* **23**, 763-85.
251. Rostom, A. A., Fucini, P., Benjamin, D. R., Juenemann, R., Nierhaus, K. H., Hartl, F. U., Dobson, C. M. & Robinson, C. V. (2000). Detection and selective dissociation of intact ribosomes in a mass spectrometer. *Proc Natl Acad Sci U S A* **97**, 5185-90.
252. Felitsyn, N., Kitova, E. N. & Klassen, J. S. (2001). Thermal decomposition of a gaseous multiprotein complex studied by blackbody infrared radiative dissociation. Investigating the origin of the asymmetric dissociation behavior. *Anal Chem* **73**, 4647-61.
253. Geels, R. B., van der Vies, S. M., Heck, A. J. & Heeren, R. M. (2006). Electron capture dissociation as structural probe for noncovalent gas-phase protein assemblies. *Anal Chem* **78**, 7191-6.
254. Jones, C. M., Beardsley, R. L., Galhena, A. S., Dagan, S., Cheng, G. & Wysocki, V. H. (2006). Symmetrical gas-phase dissociation of noncovalent protein complexes via surface collisions. *J Am Chem Soc* **128**, 15044-5.
255. Schwartz, B. L., Light-Wahl, K. J. & Smith, R. D. (1994). Observation of noncovalent complexes to the avidin tetramer by electrospray ionization mass spectrometry. *J Am Chem Soc* **5**, 201-204.

Chapter 2: Biotin Protein Ligase from *Aquifex aeolicus* - Characterisation of a *Aa*BPL:BCCP Complex

2.1 Analysis of the *A. aeolicus bpl* and *accB* Genes

Aquifex aeolicus is a hyperthermophilic bacteria found near underwater volcanoes and hot springs which grows optimally at 95°C.[1] The complete genome sequence was successfully mapped by Deckert *et al.* in 1998 and consists of 1512 predicted open reading frames.[1] The *A. aeolicus* genome is only about a third of the length of the *E. coli* genome and 16% of its genes originate from the archaea domain. A BLAST search identified two ORFs of 233 and 154 amino acids with high sequence homologies to *E. coli* BirA (20.9% identity, 35.2% similarity) and BCCP (33.8% identity, 46.9% similarity) respectively.[2] The pairwise sequence alignments for BPL and BCCP from *E. coli* and *A. aeolicus* are shown in figure 2.1 and 2.2.

Initial analysis of *A. aeolicus* BPL (*Aa*BPL) indicated that the ligase belongs to the class I of BPLs and lacks the N-terminal DNA-binding domain characterized in *E. coli* BirA (Fig 2.1).[3] *Aa*BPL is a mono-functional enzyme which acts solely as a ligase. The amino acid sequence of *Aa*BPL shows that the enzyme contains the conserved glycine rich motif $_{37}$GRGRLG$_{42}$. Limited proteolysis experiments of *Aa*BPL previously carried out by Clarke *et al.* have shown that cleavage had occurred at the adjacent Arg43 and Lys44 and was sensibly decreased in the presence of substrates.[2] In *E. coli* BirA, the equivalent residues are contained in the biotin binding loop located in proximity of the binding site.[4]

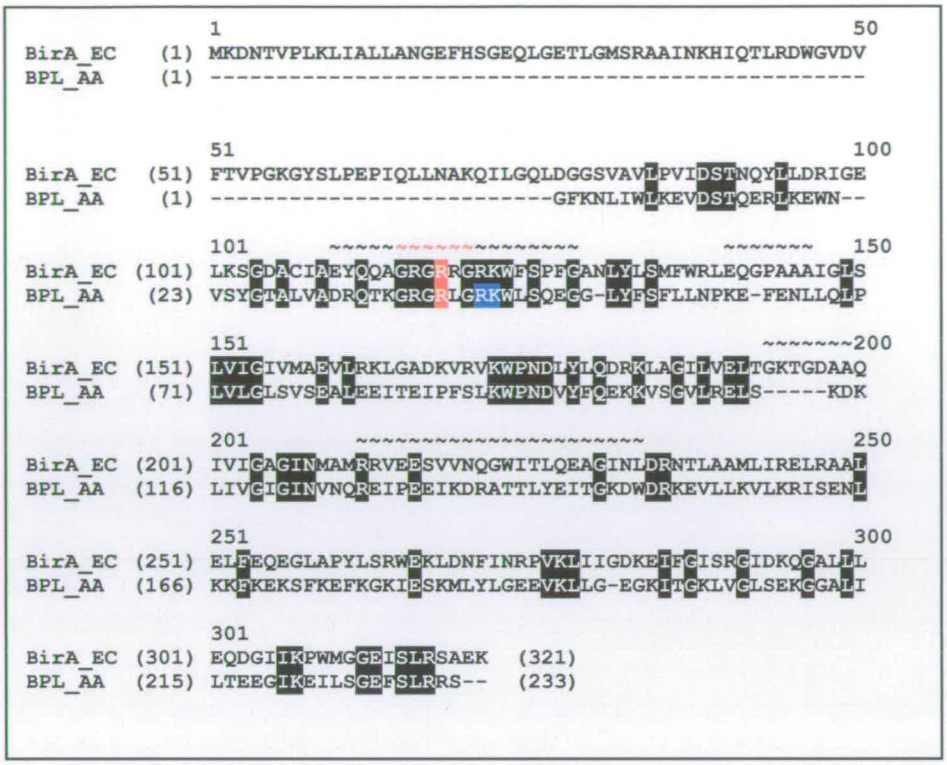


Figure 2.1 - Sequence alignments of *E. coli* BirA and *A. aeolicus Aa*BPL. The pairwise alignment was prepared using CLUSTAL W.[5] The black (~) indicate the disordered surface loops in the apo-structure of *E. coli* BirA and the red (~) indicate the residues forming the glycine-rich motif with the Arg118 of BirA and the equivalent Arg40 of *Aa*BPL highlighted in red. The residues corresponding to the trypsin cleavage sites in *Aa*BPL are indicated in blue.

In *E. coli* BCCP, a sequence rich in proline and alanine residues is found between the amino acids 34 and 101 while the rest of the N-terminus has a large proportion of charged residues and displays little sequence similarity to any other BCCPs (Fig 2.2).[6] In contrast, this pro/ala-rich sequence is absent between residues 28 and 89 of the N-terminal domain from *A. aeolicus* BCCP (BCCP154) which consists mainly of charged residues such as glutamic acids and lysines. In *A. aeolicus*, deletion of 67 residues of the N-terminal domain results in a truncated BCCP containing 87 amino acids (BCCPΔ67).[2] The homology scores between the *A.*

aeolicus BCCPΔ67 and *E. coli* BCCP87 domains are 51.9% identity and 69.6% similarity.[7] Interestingly, although the sequence identities and similarities between the two biotinyl domains are significantly high, the canonical biotinylated lysine (Lys117) of *A. aeolicus* BCCPΔ67 is flanked by aliphatic residues (AVKL motif) instead of the conserved methionines of the *E. coli* AMKM biotinylation motif.[8]

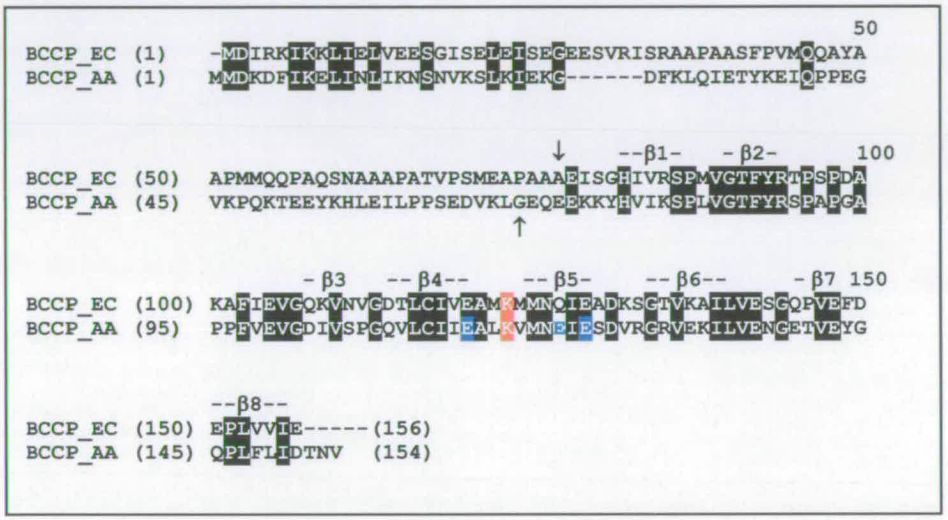


Figure 2.2 - Sequence alignments of *E. coli* and *A. aeolicus* BCCP. The pairwise alignment was prepared using CLUSTAL W.[5] The (↓) shows the start residue of the *E. coli* BCCP87 domain used in previous studies and the (↑) shows the start residue of the *A. aeolicus* BCCPΔ67 domain. The biotinylated lysine residues are highlighted in red. Residues highlighted in blue were suspected cross-linking residues (*See section 2.6*). Secondary structural elements of the BCCP87 domain are indicated.

2.2 Expression and Purification of *A. aeolicus* *Aa*BPL

The *A. aeolicus bpl* gene had been previously cloned into the kanamycin-resistant pET28a expression vector using the restriction enzymes *NcoI* and *BamHI* (Fig. 2.3).[2] The gene encoding *Aa*BPL was expressed in *E. coli* BL21 (DE3)

competent cells (Novagen) at 37 °C and induction was performed with 1 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG). The amino acid sequence of *Aa*BPL has shown that the enzyme consists of a high proportion of positively charged residues and the predicted isoelectric point of 9.1 for *Aa*BPL is unusually high. *Aa*BPL was therefore purified in a single-step by cation exchange chromatography (Fig. 2.4).

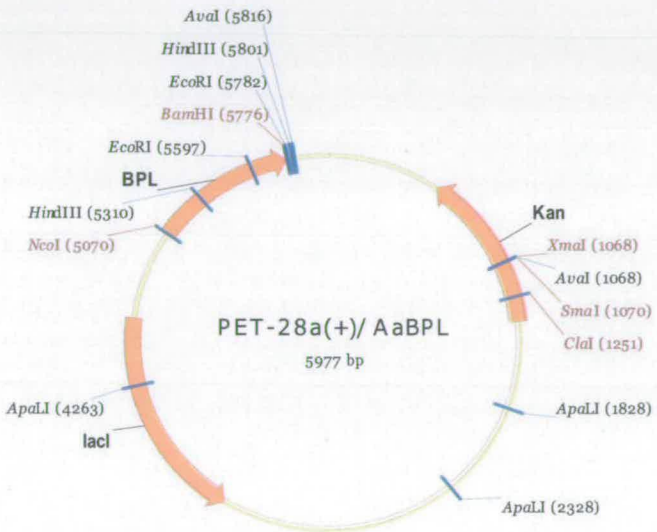


Figure 2.3 - Expression vector pET28a/*Aa*BPL. The *Aquifex aeolicus bpl* gene was cloned using the *NcoI* and *BamHI* restriction sites.

Initially, the crude lysate was incubated at 60 °C and this resulted in precipitation of a large number of *E. coli* proteins which were subsequently removed by ultra-centrifugation. The sample was then dialysed overnight at room temperature against 10 mM HEPES (pH 7.5) as immediate loading of an untreated extract onto a ResourceS column (Amersham) resulted in very poor binding. It is unclear why this step was necessary, but after dialysis the enzyme eluted from the column with 200 mM NaCl with a purity greater than 95%.[2]

Another purification method of *Aa*BPL was established to avoid the heat-shock and dialysis steps for further crystallisation studies of the enzyme. This consisted in directly loading the cell-free extract onto a 10 ml column HR 10/100 containing high resolution cationic 15S beads (GE Healthcare). *Aa*BPL eluted from the column with 200 mM NaCl with a purity of approximately 95% as determined by SDS-PAGE (Fig. 2.4). The protein was then dialysed overnight at 4 °C against 10 mM HEPES (pH 7.5) and the final yield of *Aa*BPL using this method was > 5mg per litre of cell culture. This protein was used for all subsequent activity assays, thermodynamic analyses, cross-linking studies and crystallisation.

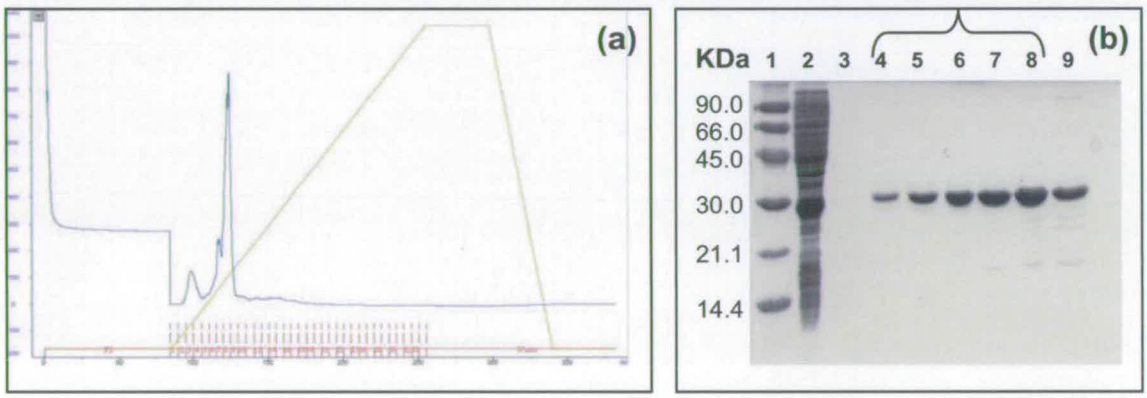


Figure 2.4 - Purification and SDS-PAGE analysis of *Aa*BPL. (a): chromatogram of *Aa*BPL purification on a HR 10/100 column loaded with 15S beads. The protein eluted with 200 mM NaCl. **(b):** protein purification was analysed by SDS-PAGE under reducing conditions. *Lane 1*, low molecular weight marker (Amersham), *lane 2, Aa*BPL flow-through, *lane 3,* wash fraction, *lane 4-9,* fractions 6 to 12 eluted with NaCl (collected fractions are indicated in brackets).

Liquid chromatography coupled to electrospray mass spectrometry (LC-ESI-MS) analysis of purified *Aa*BPL revealed the presence of a single species with a measured mass of 26,636 ± 2 Da, consistent with the post-translational removal of the N-terminal methionine residue and within experimental error of the predicted

molecular weight for *Aa*BPL (26,634.57 Da). The protein was subjected to rigorous gel filtration analysis on a 120 ml HiLoad 16/60 Superdex75 column. In absence or presence of excess of biotin or/and MgATP, the elution profile of the protein corresponded to a molecular mass of approximately 26,600 Da indicating that *Aa*BPL is monomeric (Fig. 2.5 & 2.6).

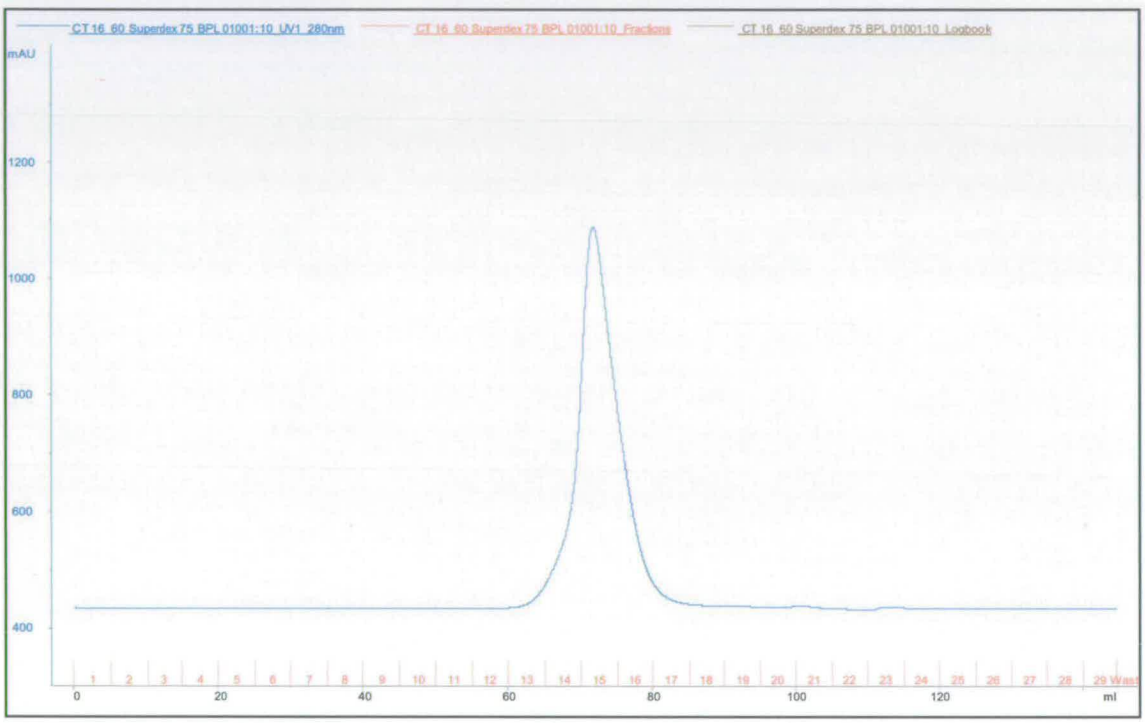


Figure 2.5 - Gel filtration analysis of *Aa*BPL on a HiLoad 16/60 Superdex75 column. The enzyme eluted from the 120 ml column in presence or absence of substrates at 65 ml indicating that *Aa*BPL is monomeric.

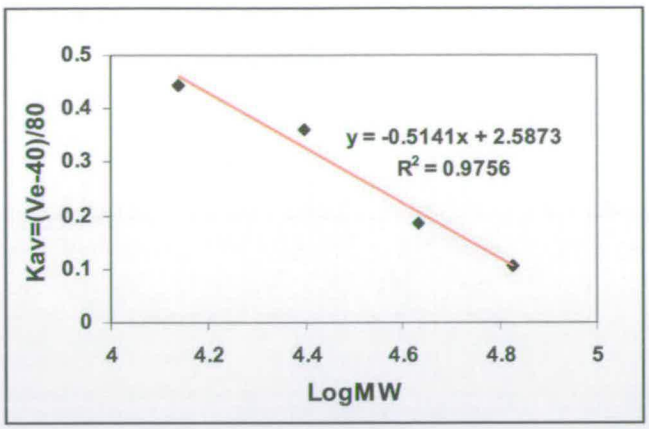


Figure 2.6 - Calibration graph for the HiLoad 16/60 Superdex75 column. Ve corresponds to the elution volume and 40 ml corresponds to the void volume of the 120 ml column.

2.3 The Full-Length and Truncated *A. aeolicus* BCCP

2.3.1 Cloning, Expression and Purification of BCCP154

The full-length *A. aeolicus* BCCP containing 154 amino acids (BCCP154) was initially cloned in a pET6H vector and expressed in *E. coli* BL21(DE3) cells at 37 °C. The resulting N-terminal hexahistidine tagged BCCP154 failed to be purified by nickel affinity chromatography from the cell lysate. The N-terminal sequence which is rich in proline and alanine results in a very flexible structure in the *E. coli* protein. In *A. aeolicus* BCCP154, the conformation of the N-terminal domain appears to have blocked binding of the hexahistidine tag to the nickel column. To avoid complications, BCCP154 was cloned into an ampicillin-resistant pET16b expression vector using the restriction enzymes *NcoI* and *BamHI* (Fig. 2.7).

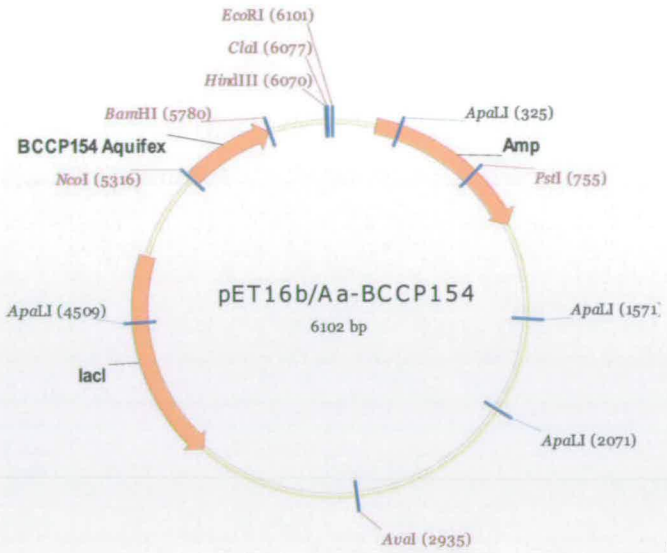


Figure 2.7 - Expression vector pET16b/BCCP154. The *Aquifex accB* gene was cloned using the *NcoI* and *BamHI* restriction sites.

The resulting plasmid pET16b/BCCP154 was expressed in *E. coli* BL21 (DE3) cells at 37 °C and the protein concentrated by two successive anion exchange chromatography steps (Fig 2.8, A & B). A final gel filtration chromatography step yielded BCCP154 which by SDS-PAGE analysis appeared to be > 90% pure (Fig 2.8, C). LC-ESI-MS indicated a molecular weight for BCCP154 of 17,331 ± 6 (Predicted mass: 17,303.88 Da). The mass difference between the experimental and calculated values is due to the poor desolvation of BCCP154 and to the presence of salt adducts. The mass spectrum of the +9 to +15 ion series also revealed that BCCP154 is heterogenous. Streptavidin western blot analysis was carried out with purified BCCP154 and showed that the *A. aeolicus* biotinyl domain was recognized and biotinylated by the host BirA *in vivo* suggesting that it was correctly folded (Fig 2.8, D).

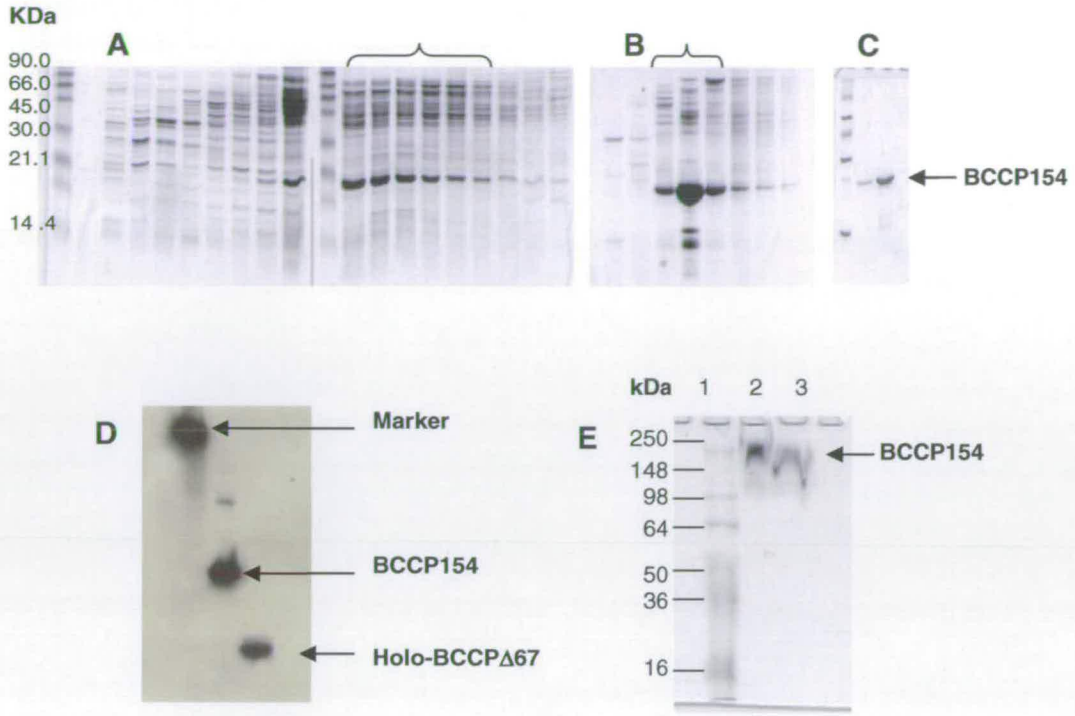


Figure 2.8 – Purification and analysis of BCCP154. A, B, C: purification of BCCP154 was analysed by SDS-PAGE under reducing conditions. The collected fractions are indicated in brackets. *Gel A:* purification on a 55 ml Q-sepharose XL column. *Gel B:* purification on 1 ml MonoQ HR 5/5 column. *Gel C*: purification on 120 ml HiLoad 16/60 Superdex75 column **D:** streptavidin western blot of BCCP154 indicates that the protein is biotinylated by its host *in vivo*. Holo-BCCPΔ67 is used as a standard. **E:** separation of BCCP154 under non-reducing conditions with a Tris-Glycine protein gel (Invitrogen), *lane 1,* pre-stained SeeBlue® Plus2 molecular weight marker (Invitrogen), *lane 2,* non-reducing Tris-Glycine loading buffer, *lane 3,* tricine loading buffer.

The stoichiometry of BCCP154 was analysed by gel filtration on a 320 ml Sephacryl S-200 HR column (Fig 2.9 & 2.10). Although different reducing conditions as well as a wide range of pH and salt concentrations were used, BCCP154 eluted from the column in the void volume and at 120 ml indicating molecular weights greater than 195 kDa. Non-reducing analysis with Tris-Glycine protein gels also showed that BCCP154 formed aggregates corresponding to molecular masses ≥ 200 kDa (Fig 2.8, E).

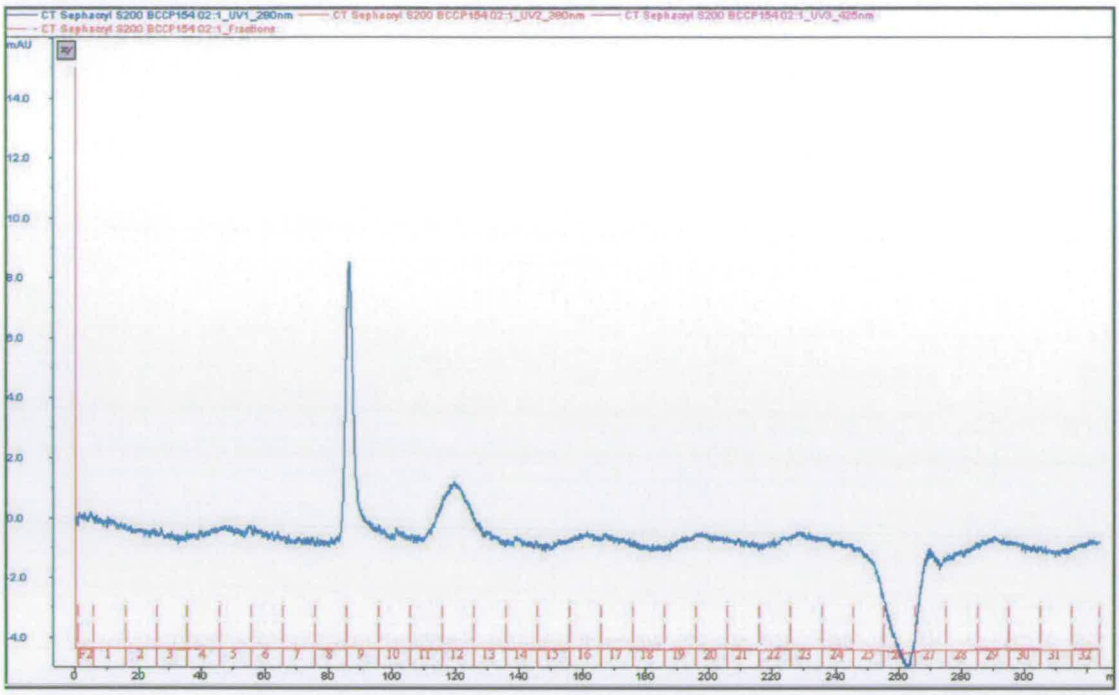


Figure 2.9 - Gel filtration analysis of BCCP154 on a Sephacryl S-200 HR. The protein eluted from the 320 ml column in presence of 2 mM DTT and 150 mM NaCl at 95 ml and 120 ml indicating that BCCP154 forms aggregates of molecular masses ≥ 195 kDa.

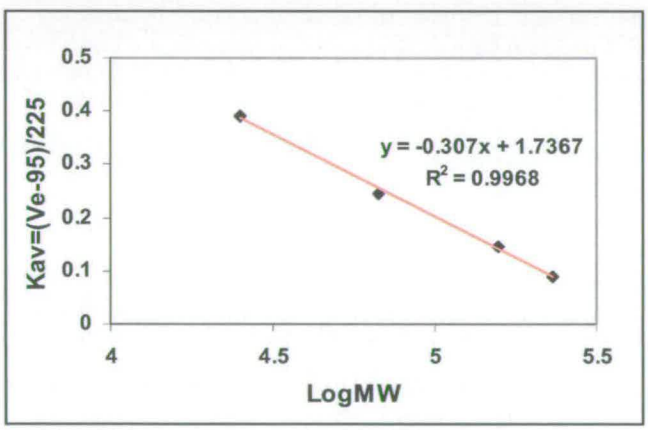


Figure 2.10 - Calibration graph for the Sephacryl S-200 HR column. Ve corresponds to the elution volume and 95 ml corresponds to the void volume of the 320 ml column.

2.3.2 Expression and Purification of BCCPΔ67 and Mutant BCCPΔ67 K117L

The full-length BCCP154 was shown to be difficult to purify, manipulate but also to analyse by mass spectrometry due to its high tendency to aggregate. Previous studies by Clarke *et al.* have been using the truncated domain BCCPΔ67 for activity assays and cross-linking experiments with *Aa*BPL.[2] The *A. aeolicus accB* gene missing the first 201 base pairs which encode the C-terminal 87 amino acids of BCCP154 (BCCPΔ67) was previously cloned into the ampicillin-resistant pET6H vector using the restriction enzymes *NcoI* and *BamHI* (Fig 2.11).[2] BCCPΔ67 is expressed from this construction with an N-terminal six histidine-tag (total length 96 aa) by BL21(DE3) cells.

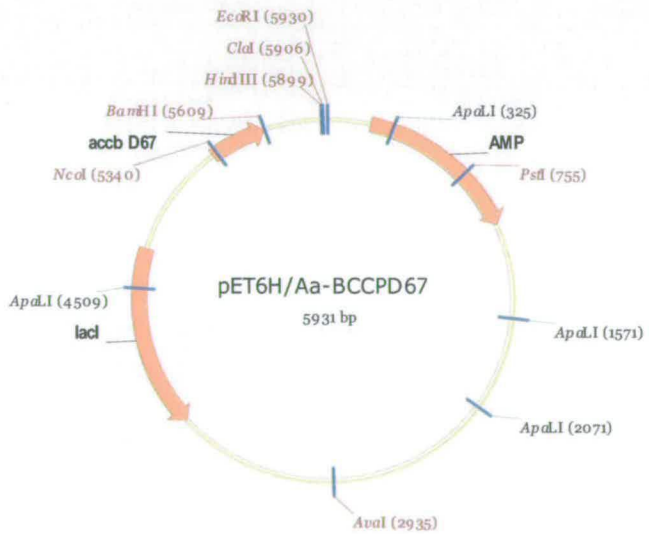


Figure 2.11 - Expression vector pET6H/BCCPΔ67. The *Aquifex accbΔ67* gene was cloned using the *NcoI* and *BamHI* restriction sites. The resulting construct encodes an N-terminal hexahistidine tag.

The hexahistidine-tagged BCCPΔ67 was concentrated from the cell lysate by nickel affinity chromatography on a 5 ml HisTrap™ HP column. His_6-BCCPΔ67 eluted with 100 mM imidazole and was immediately diluted and dialysed against 10 mM HEPES (pH 7.5).[2] SDS-PAGE indicated the protein to have a purity greater than 90% (Fig. 2.12, A). LC-ESI-MS analysis revealed the presence of two distinct species corresponding to the apo- and holo- (biotinylated) forms of BCCPΔ67. The predominant species, of molecular weight 10,739 ± 1 Da, corresponds within experimental error to the predicted mass of the apo-His_6-BCCPΔ67 (10,739.63 Da) while the second species corresponds to the holo-protein with a measured mass of 10,966 ± 1 Da (predicted mass of holo-BCCPΔ67: 10,965.94 Da). The mass difference of 227 ± 2 is equivalent within experimental error to the covalent attachment of one biotin (226.31 Da). The presence of holo-BCCPΔ67 confirmed that the biotinyl domain was recognised and biotinylated by the host BirA.

The apo- and holo-forms of BCCPΔ67 were separated by anion exchange chromatography as previously described for *E. coli* BCCP87.[7] A shallow NaCl gradient was run over 35 column volumes on a 1 ml MonoQ HR 5/5 column and LC-ESI-MS analysis of the isolated protein fractions indicated that apo-BCCPΔ67 eluted at a lower salt concentration (200 mM NaCl) than the biotinylated form (300 mM NaCl). Approximately 80% of the apo-BCCPΔ67 was completely resolved from the holo-form by collecting only the leading fractions of the protein peak (Fig. 2.12, B). The final yield of apo-BCCPΔ67 was approximately 3 mg per litre of cell culture and that of holo-form was less than 1 mg per litre.

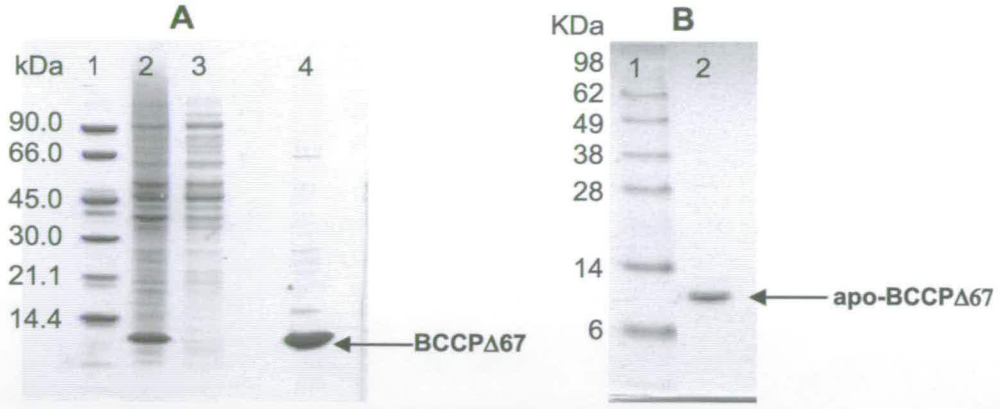


Figure 2.12 - SDS-PAGE analysis of *A. aeolicus* BCCPΔ67 after *(A)* nickel affinity and *(B)* anion exchange chromatography. Protein purification was analysed under reducing conditions. **Gel A:** *lane 1,* low molecular weight marker, *lane 2,* flow-trough, *lane 3,* wash fraction, *lane 4,* BCCPΔ67. **Gel B:** *lane 1,* SeeBlue Plus2 molecular weight marker, *lane 2,* apo-BCCPΔ67.

The hexahistidine-tagged mutant BCCPΔ67 K117L lacking the target biotinylated lysine, previously cloned in the same pET6H vector construction as BCCPΔ67, was expressed and purified by a similar method.[2] As the mutant protein lacks the lysine biotinylation site, purification was simplified by the absence of apo- and holo-forms. The cell lysate was fractionated by nickel affinity chromatography and the bound protein eluted from the HisTrap column with 200 mM imidazole. BCCPΔ67 K117L was further purified by a single anion exchange chromatography step on a MonoQ HR 5/5 column. Analysis of the fractions by SDS-PAGE showed BCCPΔ67 K117L was greater than 95% pure and LC-ESI-MS revealed the presence of a single species with a measured mass of 10,725 ± 1 Da consistent with the predicted molecular weight of the hexahistidine-tagged BCCPΔ67 K117L (10,724.62 Da).

2.4 Activity Assay of *Aa*BPL

In vitro biotinylation assays were carried out to determine if *Aa*BPL was biologically active and if the enzyme could modify apo-BCCPΔ67 in presence of biotin and Mg^{2+}-ATP. The reaction was performed at 65 °C and at pH 7.5 for 1 hour as previously described by Clarke *et al.*[2] Analyses of the reaction mixture over the course of the reaction by LC-ESI-MS indicated the appearance of a new species with a molecular weight of 10,967 ± 1 Da (Fig. 2.13). The mass increase (227 Da) of approximately one biotin is consistent with the formation of holo-BCCPΔ67 and only the singly biotinylated BCCPΔ67 could be identified by electrospray mass spectrometry. *In vitro* biotinylation assays carried out with BCCPΔ67 K117L as the biotin acceptor indicated that the BCCP mutant was not modified by *Aa*BPL, verifying that the target Lys117 is the sole biotinylation target.

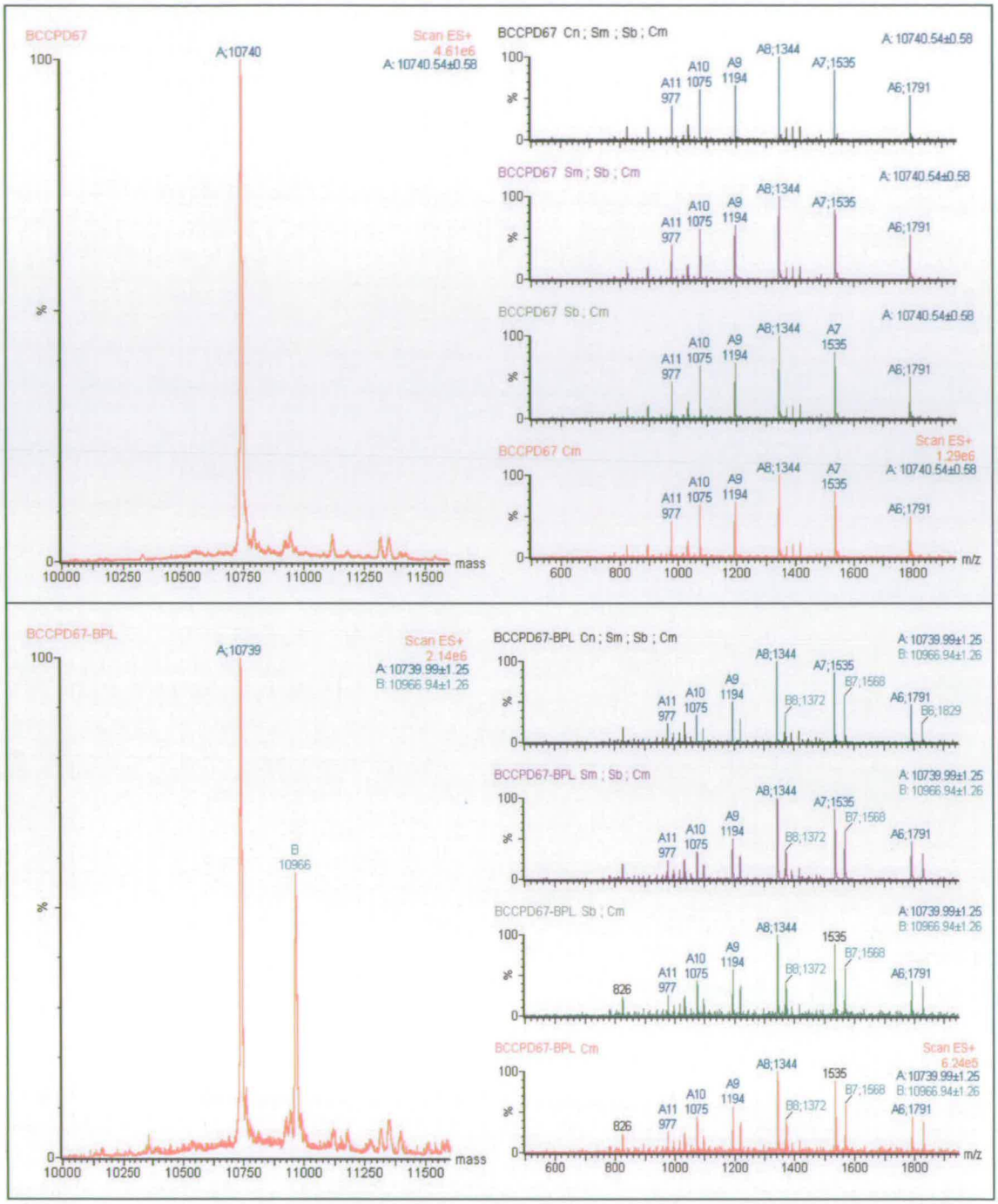


Figure 2.13 – LC-ESI-MS analysis of apo-BCCPΔ67 biotinylated with *Aa*BPL. *The right panels show the charge state distribution after smoothing and the left panel show the charge deconvoluted mass spectra.* The reaction was carried out at 65 °C in presence of MgATP. **A:** apo-BCCPΔ67 after incubation with *Aa*BPL in the absence of biotin. **B:** apo-BCCPΔ67 after incubation with *Aa*BPL for 10 min in the presence of biotin and termination with ice-cold tri-fluoro acetic acid. The mass spectrum indicates the formation of holo-BCCPΔ67.

2.5 Isothermal Titration Calorimetry Analysis of *Aa*BPL

The first step in the enzymatic biotinylation catalyzed by BPLs involves the synthesis of the intermediate biotinyl-5'-AMP from biotin and ATP in the presence of Mg^{2+} ions within the active site and the formation of inorganic pyrophosphate (PPi). In *E. coli* BirA, binding of biotin and ATP has been established to be a sequential ordered addition mechanism. In this system, the binding of biotin precedes that of ATP, and the biotin-induced ordering of the biotin binding loop is required for subsequent nucleotide binding.[9; 10] To determine the thermodynamic parameters for binding of biotin and ATP in the *Aa*BPL system, isothermal titration calorimetry (ITC) was used. Since the formation of biotinyl-5'-AMP is Mg^{2+}- dependent, the binding of the ligands prior to reaction was studied in the presence of 4 mM EDTA.[11] This proved to be a valid simplification since crystallographic studies of *Aa*BPL complexed with biotin and ATP subsequently showed that Mg^{2+} ions are not required for binding of the substrates (***See chapter 3***).

The ITC results shown in figure 2.14 indicate that *Aa*BPL binds both biotin and ATP in exothermic processes. The thermodynamic parameters calculated from these titrations suggest relatively weak protein-ligand interactions for both substrates under these conditions with K_D = 3.5 μM and K_D = 7.2 μM for biotin and ATP respectively (Table 2.1). Biotin and ATP binding to *Aa*BPL are both enthalpy-driven with favourable entropy changes. Since BirA is unable to bind ATP in the absence of biotin, only the dissociation constant of biotin was determined for BirA and this is in the nanomolar range (K_D = 45 nM) indicating a much stronger protein-ligand interaction, some 1000-fold tighter than that for *Aa*BPL and biotin.[4; 12]

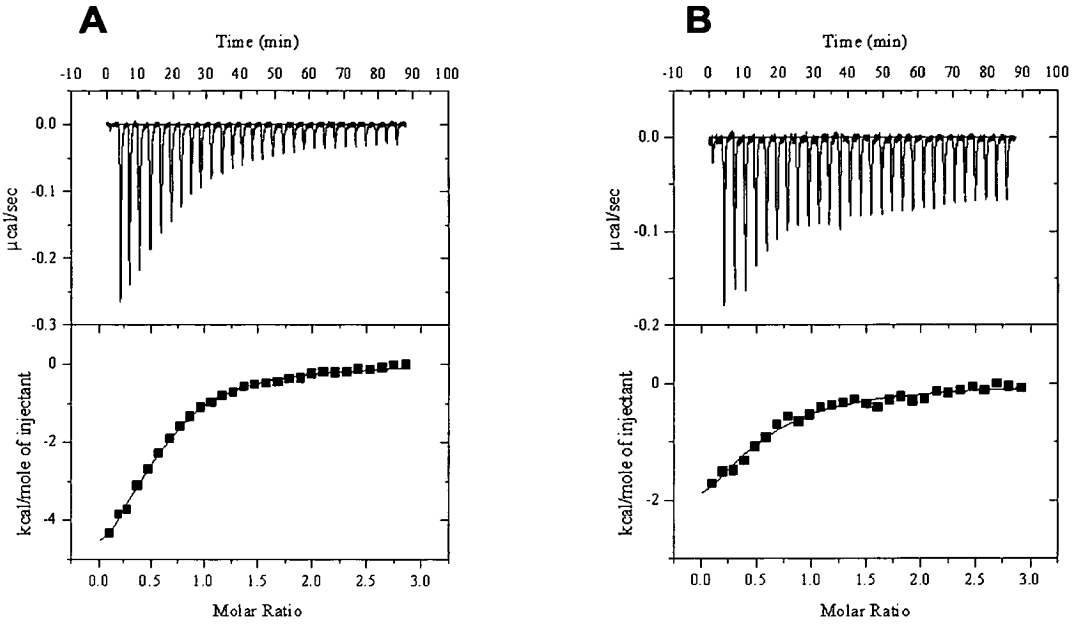


Figure 2.14 - ITC measurements of *Aa*BPL with *(A)* biotin and *(B)* ATP. The upper panels show the raw data of heat changes upon addition of ligand. The lower panels show the processed data corresponding to the heat of each injection plotted against the molar ratio of ligand to enzyme after subtraction of the heat of ligand dilution. All buffers contained 10 mM HEPES pH 7.5 and 4 mM EDTA. . The derived thermodynamic parameters are listed in table I.

	K_D	ΔH	$-T\Delta S^o$	ΔG^o
	μM	$kJ\ mol^{-1}$	$kJ\ mol^{-1}$	$kJ\ mol^{-1}$
(i) *Aa*BPL + biotin	**3.5 ± 0.2**	**-26.8 ± 0.8**	**-4.6 ± 1.3**	**-31.4 ± 0.4**
(ii) *Aa*BPL + ATP	**7.2 ± 1.3**	**-15.1 ± 3.3**	**-14.2 ± 3.8**	**-29.3 ± 3.3**
(iii) *Aa*BPL:biotin + ATP	**4.6 ± 0.5**	**-25.1 ± 2.1**	**-5.4 ± 2.5**	**-30.5 ± 2.1**
(iv) *Aa*BPL:ATP + biotin	**0.7 ± 0.1**	**-43.5 ± 0.8**	**8.4 ± 0.8**	**-35.1 ± 0.8**
BirA + biotin	*(4.5 ±0.2)x10^{-2}*	*-39.3 ± 1.3*	*-2.1 ± 1.3*	*-41.4 ± 0.4*

Table 2.1 - Thermodynamic parameters of biotin and ATP binding to *Aa*BPL. The data are derived from the ITC measurement of *Aa*BPL with biotin and ATP before and after formation of the binary complexes *Aa*BPL:biotin and *Aa*BPL:ATP. The error in ΔH is ± 5% and is mainly due to the differences in enzyme and ligand concentrations. ΔG° is calculated from the binding constant determined by ITC: $\Delta G^\circ = -RT\ \ln K_A = \Delta H - T.\Delta S^\circ$. The thermodynamic parameters of BirA binding to biotin are also reported for comparison.[4; 12]

The small difference (a factor of 2) between the K_D values of *Aa*BPL determined for biotin and ATP suggests that ligand binding to *Aa*BPL occurs *via* a random process. However, binding of ATP to *Aa*BPL could have been an artefact caused by the presence of biotin inside the active site of the enzyme during the ITC experiments. The *Aa*BPL preparation was therefore titrated with streptavidin (Sigma, K_D for biotin = 10^{-13} M) and the occupancy value of biotin shown to be 0.0372, which corresponds to less than 4% of ligand-bound enzyme, indicated that the amount of biotin initially present in the active site of *Aa*BPL could essentially be ignored (Fig 2.15).[11]

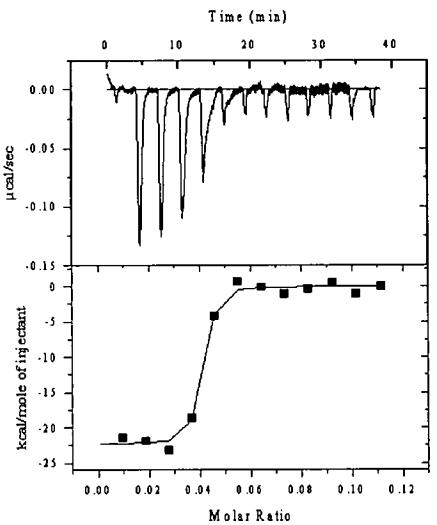


Figure 2.15 – ITC measurements of *Aa*BPL with streptavidin. The titration of 16 µM *Aa*BPL (cell) with 19.5 µM streptavidin (syringe) finished after the initial injection of 1 µl and only four injections of 10 µl. K_D = 1.4 ± 0.6 nM, ΔH = -92.0 kJ mol^{-1}, $-T\Delta S$ = 42.3 kJ mol^{-1}, ΔG^o = -49.8 kJ mol^{-1}, N = 0.0372 ± 5.71 10^{-4}.

For an ordered mechanism to occur, binding of one ligand would be prerequisite for binding of the other but, the ITC results suggest that in contrast to the mechanism of BirA, biotin is not required for ATP binding to *Aa*BPL. The cooperativity of the binding events after formation of the binary complexes

*Aa*BPL:biotin and *Aa*BPL:ATP was further investigated by titration with ATP and biotin, respectively. The ITC results and the thermodynamic parameters reveal that binding of both biotin and ATP to *Aa*BPL is enhanced by prior occupancy by the second substrate (Fig 2.16). The affinity (K_A) of biotin for the *Aa*BPL:ATP binary complex is five times greater than for *Aa*BPL alone, while the K_A of ATP for *Aa*BPL doubles for the *Aa*BPL:biotin complex (Table 2.1).

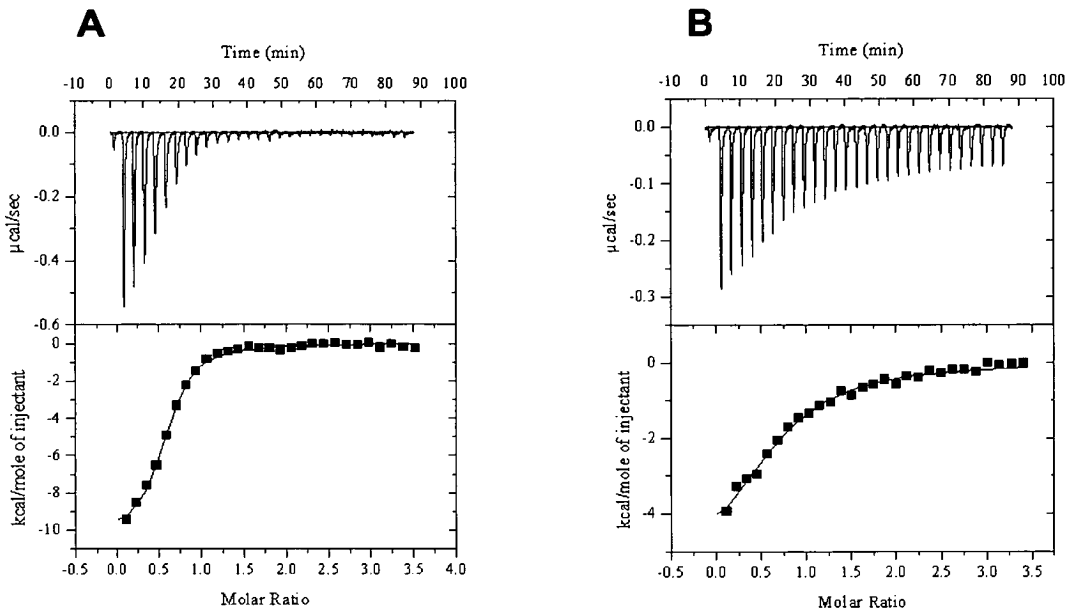


Figure 2.16 - ITC measurements of the binary complexes *(A)* *Aa*BPL:ATP with biotin and *(B)* *Aa*BPL:biotin with ATP. All buffers contained 10 mM HEPES pH 7.5, 4 mM EDTA. The derived thermodynamic parameters are listed in table 2.1.

The enthalpy and entropy changes measured for the two pathways involving the formation of the binary complexes *Aa*BPL:biotin and *Aa*BPL:ATP leading to the formation of the *Aa*BPL:biotin:ATP ternary complex - pathway (i) and (iii) and pathway (ii) and (iv) - correspond within experimental error and indicate cooperativity between biotin and ATP binding (Table 2.1, figure 2.17). It is worth noting that, despite quite large variations in apparent enthalpy and entropy changes

for the different *Aa*BPL binary and ternary complexes, the changes in binding free energies are comparatively much smaller. This appears to be yet another example of the ubiquitous enthalpy-entropy compensation commonly observed in biomolecular systems.[14]

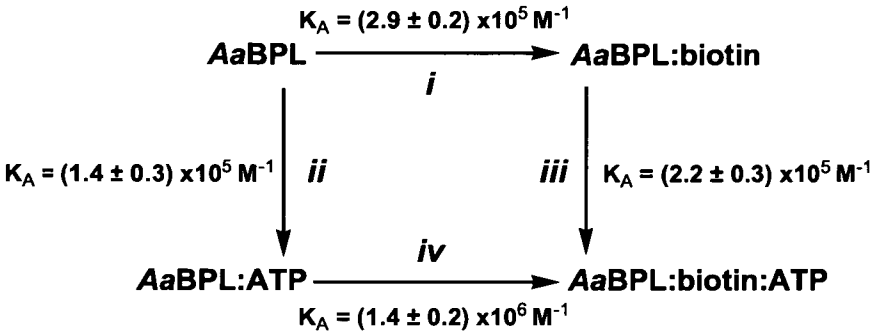


Figure 2.17 - Formation of the ternary complex *Aa*BPL:biotin:ATP. The *Aa*BPL:biotin:ATP complex can be formed *via* the formation of the binary complex *Aa*BPL:biotin (pathway i and iii) or *via* the formation of the binary complex *Aa*BPL:ATP (pathway ii and iv). In *Aa*BPL, all the four reactions for the pathways i to iv take place. K_A = association constant ($1/K_D$)

2.6 The *Aa*BPL:BCCPΔ67 Cross-Linked complex

The specificity of the biotinylation catalyzed by BPLs is well established in many organisms and has been demonstrated in *A. aeolicus* with the sole biotinylation of the target Lys117 on the carrier BCCPΔ67 which was described earlier. However, the nature of protein-protein recognition between the enzyme BPL and the substrate BCCP remain unclear. Clarke *et al.* have previously characterized by SDS-PAGE a chemically cross-linked *Aa*BPL:BCCPΔ67 complex trapped using 1-ethyl-3(dimethylamino-propyl)-cabodiimide (EDC) and subsequently purified by gel

filtration.[2] To increase the yield of *Aa*BPL:BCCPΔ67 cross-linked complex for further biochemical and structural studies aimed at gaining an insight in the residues involved in the formation of the complex, an alternative cross-linking strategy involving the coupling reagent N-hydrosuccinimide (NHS) combined with EDC was investigated. NHS is similar to EDC in that it will activate acidic residues affording succinimidyl esters for further nucleophilic attack by neighbouring amino side chains and subsequent formation of an *Aa*BPL-BCCPΔ67 amide linkage (Fig 2.18).

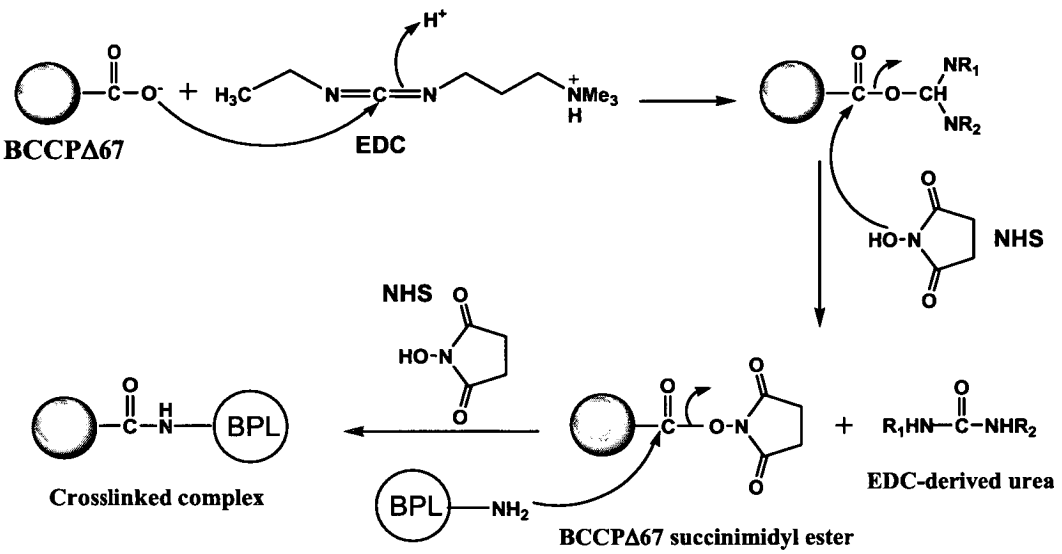


Figure 2.18 –Cross-linked *Aa*BPL:BCCPΔ67 complex with EDC and NHS. Activation of acidic residues on BCCPΔ67 with EDC and NHS leads to the formation of succinimidyl esters that will further react with amino groups of *Aa*BPL.

Chemical cross-linking experiments were carried out by activating *Aa*BPL and apo-BCCPΔ67 at 60 °C with 5 mM EDC and 5 mM NHS before quenching the reagent with β-mercaptoethanol (β-ME). The reaction mixture was incubated for various period of time and termination of the cross-linking was carried out with hydroxylamine to regenerate acidic residues that had been activated but did not

cross-link. The time course of production of cross-linked *Aa*BPL:apo-BCCPΔ67 complex was analysed by SDS-PAGE and electrospray mass spectrometry. EDC and NHS were more reactive than EDC alone and SDS-PAGE indicated the formation of two new species (Fig 2.19, A). The first major band with a molecular weight of approximately 37 kDa corresponds to the *Aa*BPL:apo-BCCPΔ67 complex while the second species has a mass of ~47 kDa and could correspond to a cross-linked complex formed by one *Aa*BPL and two apo-BCCPΔ67 molecules. Holo-BCCPΔ67 and BCCPΔ67 K117L were cross-linked to *Aa*BPL using the same method. Interestingly, the proportion of complex observed by SDS-PAGE was significantly reduced when *Aa*BPL was cross-linked with holo-BCCPΔ67 suggesting that conformational changes in BCCPΔ67 upon biotinylation alter its ability to form the complex.[2] The amount of generated cross-linked species with *Aa*BPL and the BCCPΔ67 K117L mutant was similar to that with apo-BCCPΔ67 indicating that the active Lys117 is not involved in the cross-linking reaction.[2]

Analysis of the cross-linked complexes was carried out by electrospray mass spectrometry. The *Aa*BPL:apo-BCCPΔ67 complex was ionized with difficulty and eluted at the same time as *Aa*BPL during LC-ESI-MS analyses. Direct injection after transfer in ammonium acetate buffer (pH 6.8) using a vivaspin concentrator 20,000 MWCO (Vivasciences) allowed the identification of the complex using Electrospray Maximum Entropy (MassLynx software, Waters Micromass, U. K). The *Aa*BPL:apo-BCCPΔ67 complex was observed at a molecular mass of 37,363 Da suggesting the formation of a single protein-protein amide linkage (predicted mass of non cross-linked and single cross-linked *Aa*BPL:apo-BCCPΔ67 complex: 37,374.4 Da and 37,358.4 Da respectively).

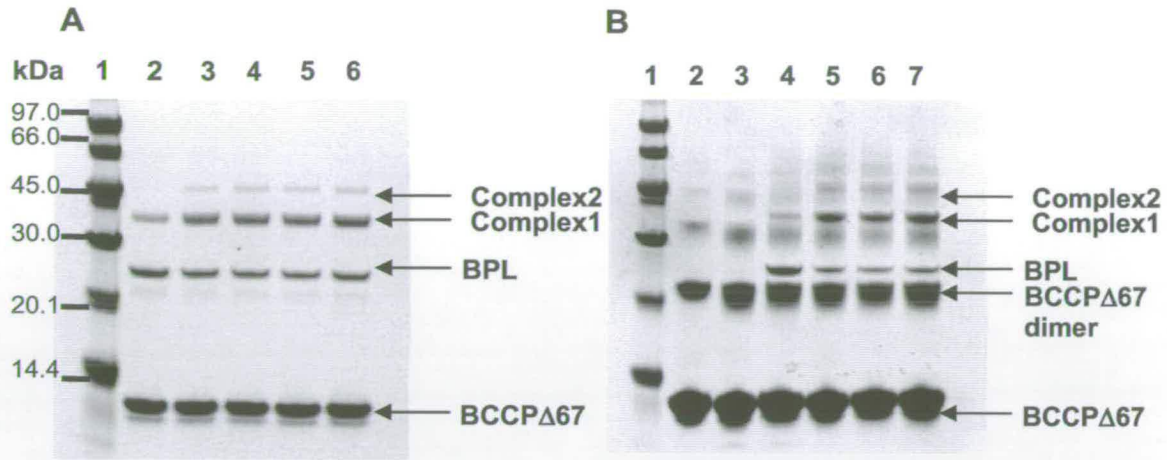


Figure 2.19 - SDS-PAGE of a cross-linked *Aa*BPL:BCCPΔ67 complex. The formation of the cross-linked complex was analysed by SDS-PAGE under reducing conditions. *All gels lane 1:* low molecular weight marker. **A:** activation of *Aa*BPL and apo-BCCPΔ67 together for 5 minutes. *Lanes 2-6:* assay after 1, 5, 15, 30 and 60 minutes respectively. **B:** activation of apo-BCCPΔ67 alone for 2 minutes. *Lane 2, 3:* assay before addition of *Aa*BPL after 1 and 3 min respectively, *lane 3-7:* assay after 1, 5, 15, 30 min.

In a separate experiment aimed at determining which protein possessed the activated carboxylate residue, *Aa*BPL and apo-BCCPΔ67 were activated separately with EDC and NHS and the second protein was added to the reaction mixture after quenching the reagents with β-ME. SDS-PAGE analysis of the reaction after hydroxylamine termination showed that activation of apo-BCCPΔ67 allowed the formation of BCCPΔ67 dimers and other higher aggregates immediately after incubation with EDC and NHS (Fig 2.19, B). However, after addition of *Aa*BPL, apo-BCCPΔ67 appeared to stop aggregating and reacted with *Aa*BPL to generate the cross-linked *Aa*BPL:apo-BCCPΔ67 complex. Holo-BCCPΔ67 did not form aggregates when incubated with EDC and NHS, presumably because the biotinylated form is more structured. Activated *Aa*BPL even for long periods of incubation did

not react with apo-BCCPΔ67 suggesting that the chemical reaction between the two proteins is ordered with an amino side chain of *Aa*BPL reacting with an activated succinimidyl ester located on the surface of BCCPΔ67.

The model of the *E. coli* complex assembled by Weaver *et al.* with the crystal structures of BirA and BCCP87 suggests that the complex is formed by hydrogen bonding of β-strands from each protein, creating an extended β-sheet (Fig 2.20).[15] The 3D model of the *E. coli* BirA:BCCP87 complex showed several pairs of charged residues (Arg or Lys and Asp or Glu) located between ~2 to ~5 Å apart that constitute candidate residues for cross-linking in the *Aa*BPL:BCCPΔ67 complex. In parallel with this work, cross-linking experiments were also carried out by Clarke with four mutated ion pairs on *Aa*BPL and BCCPΔ67, including three conserved in BirA and BCCP87, to identify their role in mediating protein cross-linking (Fig 2.20). The four single-site mutants of the BCCPΔ67 domain were assayed by SDS-PAGE for their ability to cross-link with *Aa*BPL and only the mutant BCCPΔ67 E121A showed significant reduction of generated cross-linked complex. Based on the *E. coli* model of the complex, the Lys19 of *Aa*BPL lies in close proximity to the Glu121 and subsequent mutation of the lysine also showed a considerable reduction of complex formed.

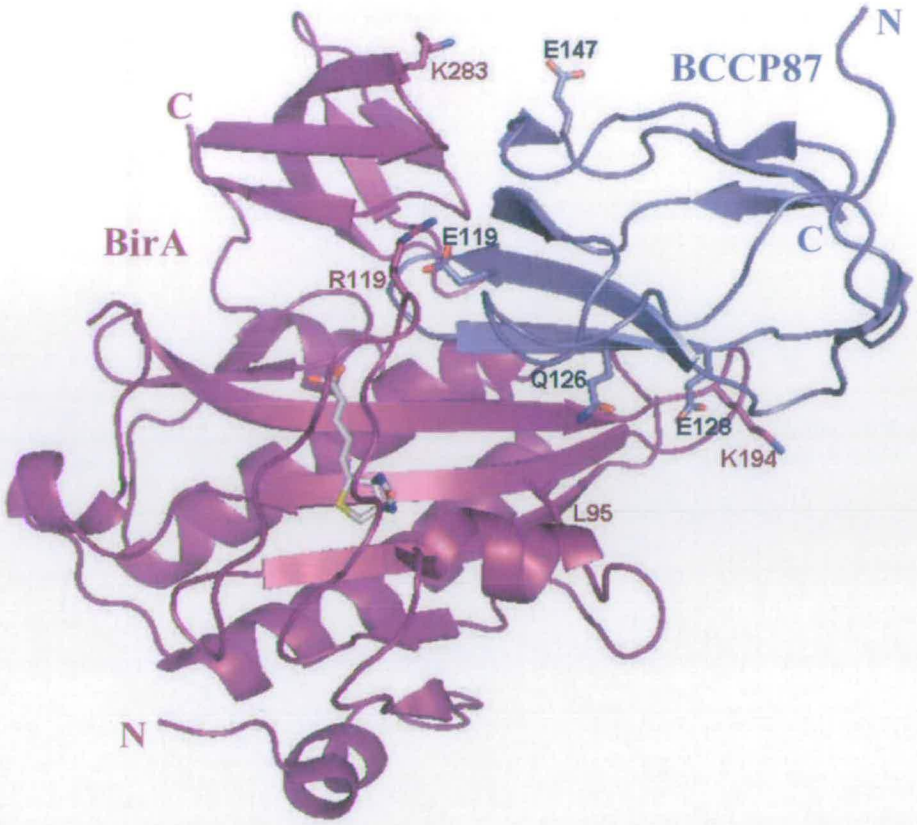


Figure 2.20 – Model of the *E. coli* BirA:BCCP87 complex. The BirA moiety is coloured in violet and the BCCP87 domain in blue. The biotin is drawn in sticks with white carbons. Four amino-acid pairs at the protein:protein interface are drawn in sticks and correspond to potential cross-linking sites. The pair Leu95-Gln126 which corresponds to Lys19 and Glu121 on *Aa*BPL and BCCPΔ67 respectively has been identified as the cross-linked salt bridge in *A. aeolicus*. (PDB entry 1K67)

2.7 Discussion

The recombinant forms of the enzyme *Aa*BPL, the full-length biotin carrier BCCP154 as well as the biotinylation domain BCCPΔ67 from the hyperthermophile *A. aeolicus* have been characterised. *Aa*BPL was shown to be monomeric and the homodimerization interactions found in *E. coli* BirA are not present.[16] Dimerization

in BirA is necessary for regulation of biotin biosynthesis by binding of the N-terminal domains to the bidirectional O/P region of the biotin operon.[17; 18] This dual role is absent in *Aa*BPL which belongs to the class I BPL and lacks the N-terminal DNA-binding domain. *Aa*BPL acts solely as a ligase and consequently remains monomeric even in presence of excess of substrates.

The *E. coli* full-length BCCP has been proven to be a particularly recalcitrant protein due of its strong tendency to aggregate which has been attributed to the presence of the flexible Pro/Ala-rich sequence between amino acids 34 and 101.[19; 20; 21] This Pro/Ala-rich sequence is absent in the N-terminal domain of *A. aeolicus* BCCP154. Recent studies on another full-length BCCP from the thermophilic archaeon *Sulfolobus tokodaii* which does not contain the Pro/Ala-rich sequence of *E. coli* BCCP indicated that the protein exists as a monomer.[22] Thermophilic bacteria are characterized by a higher number of ionic interactions which play a role in the thermostability of the proteins and these interactions could stabilize the N-terminal domains of *S. tokodaii* BCCP and *A. aeolicus* BCCP154.[23; 24] BCCP154 was initially cloned with an N-terminal His-tag and the recombinant protein expressed in *E. coli*. Although the N-terminal domain consists of a significant proportion of charged residues, the His-tagged BCCP154 could not be separated from the lysate by nickel affinity chromatography suggesting that the N-terminus may be unfolded. Thus, BCCP154 was cloned without tag in a pET16b vector and purified by anion exchange chromatography.

The protein was isolated as a mixture of apo- and holo-forms and it seems likely that the *A. aeolicus* biotinyl domain was correctly folded and was a substrate for *E. coli* BirA *in vivo*. However, like its *E. coli* homolog, BCCP154 was shown to

form undecamers and higher oligomers in solution. Moreover, analysis by mass spectrometry was shown to be particularly difficult due to the heterogeneity of purified BCCP154. Consequently, a truncated C-terminal 87 amino acids *A. aeolicus* BCCPΔ67 lacking 67 residues from the N-terminus which was previously shown to be a substrate for biotinylation with *Aa*BPL was expressed and purified in both the apo- and holo-forms.[2] LC-ESI-MS analysis of *in vitro* biotinylation assays indicated that the isolated *Aa*BPL enzyme was active and could biotinylate apo-BCCPΔ67 at elevated temperatures in the presence of biotin. Similar experiments carried out with the mutant BCCPΔ67 K117L, where the target Lys117 was mutated to a leucine, confirmed the extreme specificity of *Aa*BPL for a single lysine.

Isothermal titration calorimetry (ITC) experiments have been carried out with *Aa*BPL to determine the thermodynamic parameters for binding of biotin and ATP and to investigate if the binding events occur *via* the ordered sequential mechanism previously characterized in *E. coli* BirA.[10] The ITC results indicated that *Aa*BPL binds biotin and ATP randomly but substrate binding was also shown to be a cooperative process. This contrasts with BirA, where formation of the BirA:biotin complex is a prerequisite for ATP binding. The ordering of the binding events exhibited by BirA appears necessary in view of its dual role as a repressor and ligase and therefore, recruitment of ATP should be the second step.[25; 26] In contrast, for a monofuctional BPL such as *Aa*BPL, which lacks a regulatory role, the binding order is not crucial.

The kinetic constants of *Aa*BPL for biotin and MgATP have been previously characterized by Clarke *et al.* using steady-state kinetics and are consistent with ATP binding to unliganded *Aa*BPL (Table 2.2).[2] The K_m of *Aa*BPL for biotin was

determined to be 439.6 ± 69.2 nM, a value similar to that reported of *E. coli* BirA for biotin (K_m=300 nM).[27] However, the K_m of *Aa*BPL for MgATP was measured to be 15.14 ± 1.52 μM while the reported K_m value of BirA is much higher (K_m=300 μM).[2; 27] The K_ms of *Aa*BPL for MgATP is similar to that of yeast BPL: K_m=67 ± 11 nM and K_m=20.9 ± 3.0 μM for biotin and ATP respectively (Table 2.2).[28] Yeast BPL belongs to the class IV BPL and pyrophosphate inhibition studies carried out with *S. cerevisiae* BPL have shown that the order of binding is inverted in yeast, with ATP binding prior to biotin.[28] The order of binding of the co-factor and ATP in the adenylate synthesising enzyme family appears to be peculiarly enzyme-specific since in class I and II tRNA synthetases, amino acid and ATP binding has been shown to be a random process.[29; 30]

	K_m Biotin (nM)	K_m MgATP (μM)
*Aa*BPL	439 ± 69	15 ± 2
BirA	300 ± 100	200 ± 30
*Sc*BPL	67 ± 11	21 ± 3

Table 2.2 – Michaelis constants of the *A. aeolicus*, *E. coli* and *S. cerevisae* BPLs.[2; 27]

The second part of the post-translational biotinylation process is thought to be mediated through a network of interactions between BPL and the substrate BCCP.[15] A recent combined mutagenesis/biological selection approach identified two single glutamate residues Glu119 and Glu147 of *E. coli* BCCP87 which appear to interact with BirA. BCCP87 E119K was shown to be inactive as a substrate for BirA, whereas the E147K mutant protein could be biotinylated, albeit poorly.[31] It is presumed that these acidic BCCP87 residues interact with basic BirA counterparts

and the mutant BirA in which Lys277 was substituted with a Glu residue had a significantly higher activity with BCCP87 E119K than wild-type BirA. Those results suggest that ion pair networks are important for the protein recognition of BCCP by BirA in *E. coli.* [27] Ion pair networks are a common feature in heat-resistant proteins and are believed to play important roles in their increased thermal stability.[23] Since both *A. aeolicus Aa*BPL and BCCP contain a large number of charged residues and inhibition of biotinylation was observed at high salt concentrations, specific ionic interactions are presumably involved in the formation of the hyperthermophilic *Aa*BPL:BCCPΔ67 complex.[2]

To investigate the formation of the *Aa*BPL:BCCPΔ67 heterodimer for biochemical and structural studies, EDC and NHS were used as chemical cross-linking reagents. SDS-PAGE analysis of apo-BCCPΔ67 incubated with *Aa*BPL in the presence of EDC/NHS led to the time-dependent appearance of a species with a molecular weight of ~37 kDa which is in agreement with the predicted mass of a 1:1 complex. A species with a mass of ~48 kDa which could correspond to a complex formed by two BCCPs and one *Aa*BPL molecules was also observed by SDS-PAGE. Incubation of apo-BCCPΔ67 alone in the presence of EDC/NHS resulted in the formation of dimers and other higher aggregates stable under the SDS-PAGE reducing conditions which confirmed this hypothesis. Further cross-linking experiments with *Aa*BPL and apo-BCCPΔ67 indicated that the reaction was ordered and involved activated acidic residues on apo-BCCPΔ67 which reacted with an amino side chain on *Aa*BPL. The 3D model of the *E. coli* BirA:BCCP87 complex and the *Aa*BPL/BirA sequence alignments were used to identify potential cross-linked residues in the corresponding *Aa*BPL:BCCPΔ67 complex.[15] Based on SDS-

PAGE results of cross-linking experiments with *Aa*BPL and BCCPΔ67 single mutants, it appears that the salt bridge leading to the formation of a *Aa*BPL-BCCPΔ67 amide bond is due to the two non-conserved residues Lys19 on *Aa*BPL and Glu121 on BCCPΔ67.(Clarke, unpublished results)

2.8 References

1. Deckert, G., Warren, P. V., Gaasterland, T., Young, W. G., Lenox, A. L., Graham, D. E., Overbeek, R., Snead, M. A., Keller, M., Aujay, M., Huber, R., Feldman, R. A., Short, J. M., Olsen, G. J. & Swanson, R. V. (1998). The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. *Nature* **392**, 353-8.
2. Clarke, D. J., Coulson, J., Baillie, R. & Campopiano, D. J. (2003). Biotinylation in the hyperthermophile *Aquifex aeolicus*. *Eur J Biochem* **270**, 1277-87.
3. Mukhopadhyay, B., Purwantini, E., Kreder, C. L. & Wolfe, R. S. (2001). Oxaloacetate synthesis in the methanarchaeon Methanosarcina barkeri: pyruvate carboxylase genes and a putative *Escherichia coli*-type bifunctional biotin protein ligase gene (bpl/birA) exhibit a unique organization. *J Bacteriol* **183**, 3804-10.
4. Kwon, K. & Beckett, D. (2000). Function of a conserved sequence motif in biotin holoenzyme synthetases. *Protein Sci* **9**, 1530-9.
5. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673-80.
6. Nenortas, E. & Beckett, D. (1996). Purification and characterization of intact and truncated forms of the *Escherichia coli* biotin carboxyl carrier subunit of acetyl-CoA carboxylase. *J Biol Chem* **271**, 7559-67.
7. Chapman-Smith, A., Turner, D. L., Cronan, J. E., Jr., Morris, T. W. & Wallace, J. C. (1994). Expression, biotinylation and purification of a biotin-domain peptide from the biotin carboxy carrier protein of *Escherichia coli* acetyl-CoA carboxylase. *Biochem J* **302 (Pt 3)**, 881-7.
8. Reche, P., Li, Y. L., Fuller, C., Eichhorn, K. & Perham, R. N. (1998). Selectivity of post-translational modification in biotinylated proteins: the carboxy carrier protein of the acetyl-CoA carboxylase of *Escherichia coli*. *Biochem J* **329 (Pt 3)**, 589-96.

9. Xu, Y. & Beckett, D. (1997). Biotinyl-5'-adenylate synthesis catalyzed by *Escherichia coli* repressor of biotin biosynthesis. *Methods Enzymol* **279**, 405-21.
10. Wood, Z. A., Weaver, L. H., Brown, P. H., Beckett, D. & Matthews, B. W. (2006). Co-repressor induced order and biotin repressor dimerization: a case for divergent followed by convergent evolution. *J Mol Biol* **357**, 509-23.
11. Perozzo, R., Jelesarov, I., Bosshard, H. R., Folkers, G. & Scapozza, L. (2000). Compulsory order of substrate binding to herpes simplex virus type 1 thymidine kinase. A calorimetric study. *J Biol Chem* **275**, 16139-45.
12. Kwon, K., Streaker, E. D. & Beckett, D. (2002). Binding specificity and the ligand dissociation process in the *E. coli* biotin holoenzyme synthetase. *Protein Sci* **11**, 558-70.
13. Klumb, L. A., Chu, V. & Stayton, P. S. (1998). Energetic roles of hydrogen bonds at the ureido oxygen binding pocket in the streptavidin-biotin complex. *Biochemistry* **37**, 7657-63.
14. Cooper, A., Johnson, C. M., Lakey, J. H. & Nollmann, M. (2001). Heat does not come in different colours: entropy-enthalpy compensation, free energy windows, quantum confinement, pressure perturbation calorimetry, solvation and the multiple causes of heat capacity effects in biomolecular interactions. *Biophys Chem* **93**, 215-30.
15. Weaver, L. H., Kwon, K., Beckett, D. & Matthews, B. W. (2001). Competing protein:protein interactions are proposed to control the biological switch of the *E coli* biotin repressor. *Protein Sci* **10**, 2618-22.
16. Eisenstein, E. & Beckett, D. (1999). Dimerization of the *Escherichia coli* biotin repressor: corepressor function in protein assembly. *Biochemistry* **38**, 13077-84.
17. Streaker, E. D., Gupta, A. & Beckett, D. (2002). The biotin repressor: thermodynamic coupling of corepressor binding, protein assembly, and sequence-specific DNA binding. *Biochemistry* **41**, 14263-71.
18. Streaker, E. D. & Beckett, D. (2003). Coupling of protein assembly and DNA binding: biotin repressor dimerization precedes biotin operator binding. *J Mol Biol* **325**, 937-48.
19. Fall, R. R., Nervi, A. M., Alberts, A. W. & Vagelos, P. R. (1971). Acetyl CoA carboxylase: isolation and characterization of native biotin carboxyl carrier protein. *Proc Natl Acad Sci U S A* **68**, 1512-5.
20. Fall, R. R. & Vagelos, P. R. (1975). Biotin carboxyl carrier protein from *Escherichia coli*. *Methods Enzymol* **35**, 17-25.
21. Perham, R. N. (2000). Swinging arms and swinging domains in multifunctional enzymes: catalytic machines for multistep reactions. *Annu Rev Biochem* **69**, 961-1004.
22. Li, Y. Q., Sueda, S., Kondo, H. & Kawarabayasi, Y. (2006). A unique biotin carboxyl carrier protein in archaeon *Sulfolobus tokodaii*. *FEBS Lett* **580**, 1536-40.
23. Demirjian, D. C., Moris-Varas, F. & Cassidy, C. S. (2001). Enzymes from extremophiles. *Curr Opin Chem Biol* **5**, 144-51.
24. Hough, D. W. & Danson, M. J. (1999). Extremozymes. *Curr Opin Chem Biol* **3**, 39-46.

25. Brown, P. H., Cronan, J. E., Grotli, M. & Beckett, D. (2004). The biotin repressor: modulation of allostery by corepressor analogs. *J Mol Biol* **337**, 857-69.
26. Streaker, E. D. & Beckett, D. (2006). The biotin regulatory system: kinetic control of a transcriptional switch. *Biochemistry* **45**, 6417-25.
27. Chapman-Smith, A., Mulhern, T. D., Whelan, F., Cronan, J. E., Jr. & Wallace, J. C. (2001). The C-terminal domain of biotin protein ligase from *E. coli* is required for catalytic activity. *Protein Sci* **10**, 2608-17.
28. Polyak, S. W., Chapman-Smith, A., Brautigan, P. J. & Wallace, J. C. (1999). Biotin protein ligase from *Saccharomyces cerevisiae*. The N-terminal domain is required for complete activity. *J Biol Chem* **274**, 32847-54.
29. Pan, F., Lo, K. Y., Pai, S. H. & Lee, H. H. (1982). Kinetic mechanism of threonyl-tRNA synthetase from human placenta. *Int J Pept Protein Res* **20**, 159-66.
30. Wang, H. Y. & Pan, F. (1984). Kinetic mechanism of arginyl-tRNA synthetase from human placenta. *Int J Biochem* **16**, 1379-85.
31. Chapman-Smith, A., Morris, T. W., Wallace, J. C. & Cronan, J. E., Jr. (1999). Molecular recognition in a post-translational modification of exceptional specificity. Mutants of the biotinylated domain of acetyl-CoA carboxylase defective in recognition by biotin protein ligase. *J Biol Chem* **274**, 1449-57.

Chapter 3: Structural Analysis of *A. aeolicus* Biotin Protein Ligase

3.1 Crystallization of *A. aeolicus Aa*BPL

The class I *A. aeolicus Aa*BPL studied in this project was initially crystallized prior to the publication at the end of 2005 of the BPL structure from *P. horikoshii*.[1;2] *Ph*BPL also belongs to the class I BPL and possesses significant sequence homology with *Aa*BPL (30.6 % sequence identity and 47.1 % sequence similarity). The sequence homologies displayed by the C-terminal of BirA and *Ph*BPL remain similar to those measured for *Aa*BPL (32.9% similarity for *Ph*BPL and BirA, 20.9% identity, 35.2% similarity for *Aa*BPL and BirA and 22.0% identity). Although the structures of *Ph*BPL in complex with biotin, ADP and biotinyl-5'-AMP shed light on the residues important for catalytic activity, the exact position and orientation of ATP within the active site remained unclear. Mutation of the Arg118 in BirA has been shown to dramatically affect the affinity of the enzyme for the different substrates. In *Ph*BPL, the equivalent conserved Arg48 plays only a minor role in binding biotin.[1] Important sequence differences are displayed within the biotin binding loop glycine-rich motifs of the *P. horikoshii*, *A. aeolicus* and the *E. coli* enzymes (*Ph*BPL: $_{45}$GHGRLN$_{50}$) while the $_{37}$GRGRLG$_{50}$ motif of *Aa*BPL is similar to that of BirA (***See chapter 2, figure 2.1***).[3]

Crystallographic studies of *Aa*BPL were carried out in collaboration with Prof. Malcom Walkinshaw at the University of Edinburgh. The enzyme was crystallized by vapor diffusion and the 24 well plates were stored at 17 °C. An initial screening was performed with the Structure Screen 2 Kit (Molecular Dimension). Forty eight different trial conditions containing sterile filtered reagent combinations incorporating precipitants across a pH range of 4 to 9 were tested with a standard

*Aa*BPL concentration of 6 mg. ml^{-1}. Diamond-shaped microcrystals of 0.05 mm grew in two weeks under three different trial conditions (Fig. 3.1, A). Two crystals from different plates were soaked into an equivalent reservoir solution supplemented with 20% v/v sterile glycerol and frozen in liquid nitrogen. X-ray diffraction measurements at the European Synchrotron Radiation Facility (ESRF) in Grenoble (France) allowed data collection to a resolution of 2.8 Å for both crystals. The sequence similarities displayed by *Aa*BPL and BirA allowed the modeling of an initial structure of *Aa*BPL by molecular replacement using the polyalanine backbone of the *E. coli* enzyme. The preliminary structure of *Aa*BPL was solved by Dr. Iain McNae using the programs MOSFLM and SCALA in the CCP4 suite and the coordinates were refined to a crystallographic residual R_{CRYST} = 31% and R_{FREE} = 41% at 2.8 Å resolution.

To obtain better diffracting crystals, optimization was carried out for the three conditions in which crystals were previously obtained, at varying enzyme and precipitant concentrations, pH, temperature and ratios of enzyme / reservoir solution within the drops. The crystals grew overnight with different shapes and dimensions varying from microcrystals to needles of approximately 1 mm length (Fig. 3.1, B). The size of the needles reached 1.5 mm when crystallization was carried out at 4 °C (data not shown). A large number of crystals were generated in a more or less random process within the different drops of the optimization screens. However, preliminary X-ray diffraction at the home-beam in Edinburgh indicated that the crystals were disordered and the resolution of the data collected ranged from 3.5 to 7 Å. Further X-ray screening of the crystals was performed on the Edinburgh beam and at the ESRF and the Synchrotron Radiation Source (SRS) in Daresbury (U.K) on

crystals grown under more than twenty different conditions. The resolutions of the X-ray diffractions measurements from crystals within a same drop were also found to vary significantly.

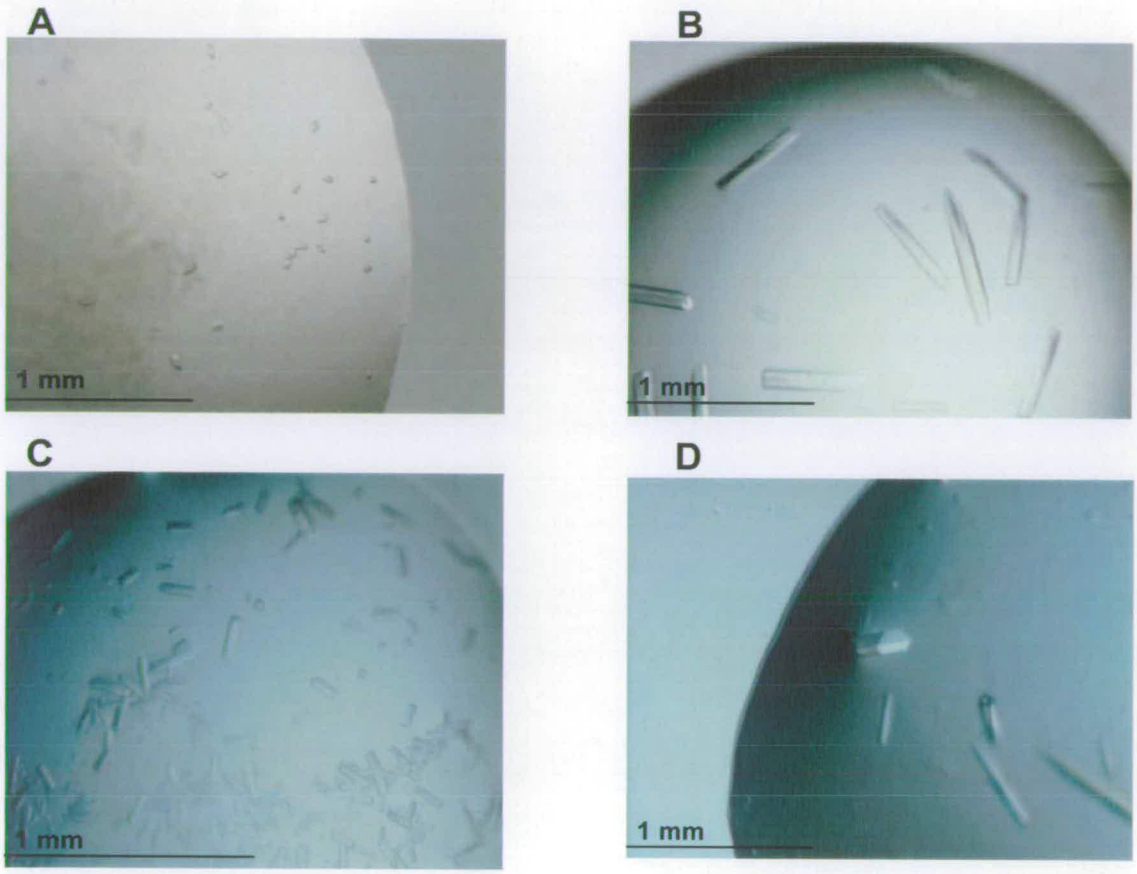


Figure 3.1 – Crystals of apo-*Aa*BPL. The crystals were grown at 17 °C with the following precipitants. **A:** 0.2 M ammonium sulfate, 0.1 M MES pH 6.5, 30 % w/v PEG monoethylether 5,000. **B:** 0.1 M MES pH 7, 12% w/v PEG 20,000. **C:** 0.2 M ammonium sulfate, 0.1 M MES pH 6.5, 20% w/v PEG 5,000 monoethylether; **D:** 0.2 M ammonium sulphate, 0.1 M MES pH 6.5, 14% w/v PEG 5,000 monoethylether.

The conditions in which the crystals of *Aa*BPL diffracted to the greatest resolution were chosen for further optimization. *Aa*BPL was crystallized at a concentration of 5 $mg.ml^{-1}$ and the drops contained 2 µl enzyme / 1 µl reservoir solution. The wells contained 0.1 M MES, 0.2 M ammonium sulfate, 20% w/v PEG

5,000 monoethyl ether with a final pH range of 6 to 7. The crystals grew non-reproducibly with a rectangular shape in few hours to a size of approximately 0.4 mm (Fig. 3.1, C). X-ray diffraction of the crystals indicated resolutions lower than 2.7 Å. The immediate and promiscuous crystallization of *Aa*BPL under such conditions appeared to contribute to the significant disorder observed within the crystals generated. To limit the nucleation sites due to impurities, the enzyme was centrifuged for 30 min at 4 °C prior to the crystallization setup. Lower PEG and enzyme (4.5 $mg.ml^{-1}$) concentrations were also found to slow down the nucleation process and led to the formation of crystals which diffracted to a resolution of 2.3 Å (Fig. 3.1, D). Data for the apo-crystals of *Aa*BPL were collected at BM14 at the ESRF Grenoble and processed by Dr. Iain W. McNae. The structure of *Aa*BPL in the unbound form (apo-*Aa*BPL) was solved by molecular replacement using the newly released structure of apo-*Ph*BPL and the coordinates were refined to a crystallographic residual R_{CRYST} = 21.39% and R_{FREE} = 28.38% at 2.3 Å resolution (Table 3.1).

Many attempts were carried out to co-crystallize biotin with *Aa*BPL under similar conditions as used for the apo-enzyme at different biotin concentrations ranging across 200 μM to 2 mM (pH 12). Co-crystals with biotin grew overnight with a different hexagonal shape. However, these hexagonal crystals were disordered and destroyed within three days. Rectangular crystals which grew more slowly were chosen for X-ray diffraction at the synchrotron but the resolution of the data proved to be lower than 3 Å.

Data collection and Processing.	**Apo**	**ATP, biotin**	**R40G, biotin**
Space group	$P2_1$	$P2_12_12_1$	$P2_12_12_1$
Unit cell parameters *a* (Å) *b* (Å) *c* (Å) α (deg) β (deg) γ (deg)	 56.513, 61.336, 73.308, 90.0, 90.50, 90.0	 41.225 81.695 143.278 90 90 90	 41.2513 79.9277 140.1733 90 90 90
Resolution (Å) [high shell]	31.47-2.30 [2.42-2.30]	39.62-2.30 [2.42-2.30]	28.84-2.55
Observations [high shell]	92480 [8941]	76866 [11290]	63143 [8673]
Unique observations [high shell]	20745 [2066]	21993 [3166]	13851 [1869]
Multiplicity [high shell]	4.5 [4.3]	3.5 [3.6]	4.6 [4.6]
Completeness (%) [high shell]	92.4 [64.6]	98.9 [99.0]	88.1 [83.9]
Mean I/σ(I) [high shell]	22.4 [5.0]	10.4 [2.5]	12.6 [2.2]
R merge (%) [high shell]	3.5 [30.8]	9.7 [42.3]	9.7 [63.9]
Refinement			
Total atoms	3648	4071	3903
Solvent atoms	87	181	131
Resolution range (Å)	31.47-2.30	39.62-2.30	28.84-2.55
Rcryst (%)	21.39	20.86	20.00
Rfree (%)	28.38	28.92	29.31
R.M.S.D from ideal			
Bond lengths (Å)	0.005	0.006	0.007
Bond angles (deg)	0.695	0.902	0.932
Ramachandran plot statistics			
Residues in most favoured regions (%)	86.8	87.0	84
Residues in additionally allowed regions (%)	11.1	11.0	13.3
Residues in generously allowed regions (%)	1.6	0.5	1.5
Residues in disallowed regions (%)	0.5	1.5	1.2

Table 3.1 - X-ray Data collection, Processing and Refinement Statistics.

Similarly, *Aa*BPL was co-crystallized with 1 mM biotin and 5 mM ATP in the absence of $MgCl_2$. Crystals grew within two to three days and these appeared more ordered than the *Aa*BPL:biotin co-crystals. X-ray diffraction at SRS Daresbury station 10.1 allowed data collection to a resolution of 2.3 Å. The structure was solved by molecular replacement using the structures of apo-*Aa*BPL and *Ph*BPL and the coordinates were refined to a crystallographic residual R_{CRYST} = 20.86% and R_{FREE} = 28.92% (Table 3.1).

3.2 Crystal Structure of *Aa*BPL

*Aa*BPL displays linear sequence similarities with *E. coli* BirA and *Ph*BPL and analysis of the crystal structures from the three different enzymes reveals that these similarities are also conserved in their overall folds and secondary structures (Fig. 3.2 & 3.3).[1; 4; 5] In the structure of apo-*Aa*BPL, the first 186 residues of the 26 kDa protein folds into a large N-terminal catalytic domain which adopts the conserved RRM fold containing seven β-strands and five α-helices (Fig. 3.2).[6] In the catalytic domain, the two N-terminal parallel β-strands are inaccessible to solvent while the remainder of the mixed β-sheet has one face which is largely solvent-exposed and forms the active site (strands β_3, β_6 and β_7). The Lys19 which was shown to interact with apo-BCCPΔ67 by chemical cross-linking with EDC/NHS is located on the α_1-helix which constitutes the bottom of the catalytic pocket. The smaller C-terminal domain (47 aa) contains a mixed β-sheet of five strands and forms a β-barrel similar to a SH3-module.

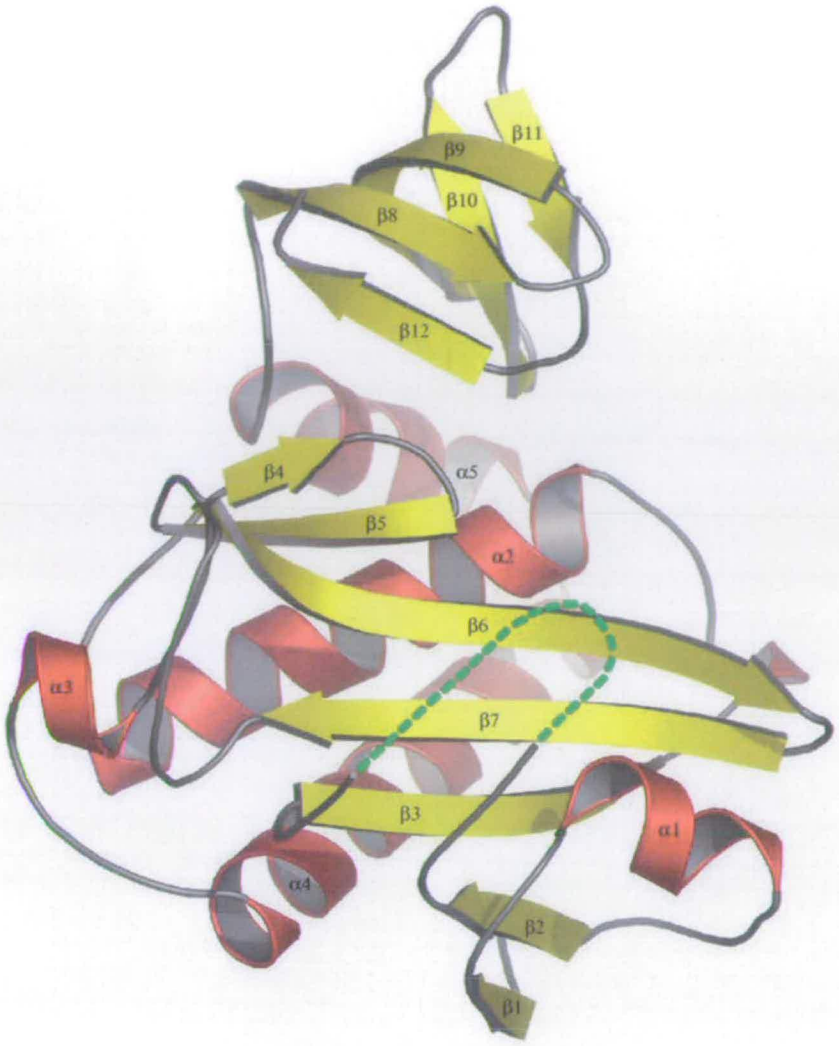


Figure 3.2 – Monomer structure of apo-*Aa*BPL. Secondary structure elements α-helices (coloured red) and β-strands (coloured yellow) are numbered according to their position in the protein sequence. The disordered loop (residues 37 to 47) is represented by a green dashed line.

The biotin binding loop, encompassing residues 37 to 47 (loop β_2-β_3) containing the $_{37}$GRGRLG$_{42}$ motif, is disordered in the structure of apo-*Aa*BPL. The equivalent adenylate-binding loop, which only becomes organized upon ATP binding in BirA, is ordered in the structure of apo-*Aa*BPL and the ATP binding site is consequently more structured (Fig. 3.2).[4; 7] This observation supports our previous

ITC binding studies which showed that in contrast to apo-BirA, apo-*Aa*BPL is capable of binding ATP in the absence of biotin (***See section 2.5***). Additionally, compared to other BPL structures, the surface loop connecting the β_6 and β_7 strands of the active site is shifted towards the catalytic pocket in *Aa*BPL (Fig. 3.3).

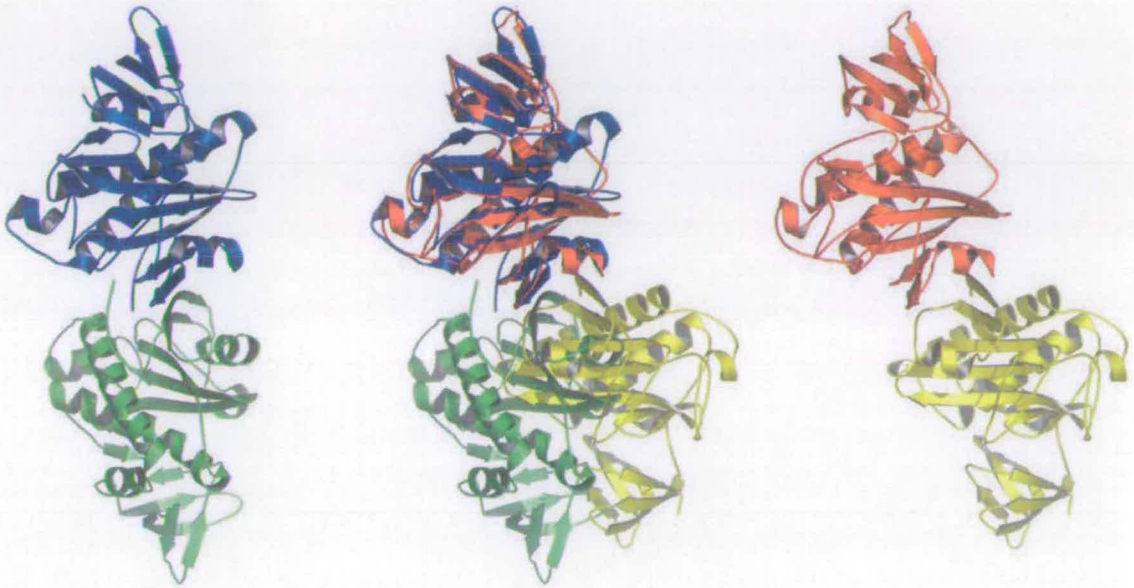


Figure 3.3 - Dimer comparison of *Ph*BPL and *Aa*BPL. ***Left image:*** *Ph*BPL dimer, monomers A and B from the crystal asymmetric unit are coloured blue and green respectively. ***Right image:*** *Aa*BPL monomers A and B are coloured red and yellow respectively. ***Central image:*** overlay of chains A from *Ph*BPL and *Aa*BPL dimers in the crystal asymmetric unit. Colouring is as for individual images.

Unlike *Aa*BPL, *Ph*BPL has been shown to be dimeric in solution by dynamic light-scattering experiment and this dimer is also observed in the crystal form (***See section 2.2***).[1] There are no structural similarities between the homodimers formed by *Ph*BPL and BirA. While in the crystal structure of dimeric BirA the β-strands of the catalytic domains form a seamless β-sheet interface, the structure of *Ph*BPL indicates that the two subunits of the homodimer are associated *via* a β-sheet formed by the β_1-

strands of each monomer (Fig. 3.3).[1; 4] The crystals of apo-*Aa*BPL are monoclinic and belong to the space group $P2_1$, while the crystals of *Aa*BPL in complex with biotin and ATP belong to the orthorhombic system and possess a $P2_12_12_1$ symmetry (Table 3.1). Similarly to the *P. horikoshii* structure, the asymmetric unit of *Aa*BPL consists of two monomers related by a pseudo-two fold axis.[1] However, in the structure of apo-*Aa*BPL, the β_1-strands are located in close proximity but do not form the inter-subunit β-sheet observed in *Ph*BPL (Fig. 3.3). Thus, in contrast to the *P. horikoshii* enzyme, *Aa*BPL does not constitute a functional homodimer.

3.3 The *Aa*BPL:Biotin:ATP Complex

3.3.1 The Biotin and ATP Binding Sites of *Aa*BPL

In the structure of *Aa*BPL in complex with biotin and ATP (*Aa*BPL:biotin:ATP), the two ligands are fully ordered as are residues 37-47 which form the biotin binding loop (Fig. 3.4 & 3.5, a). The peptide segment stabilizes the biotin and ATP moieties inside the active site with strong hydrophobic and hydrophilic interactions. Biotin is shown to interact with the biotin binding loop as well as the solvent exposed β_6 and β_7 strands and α_1-helix. The tetrahydrothiophene ring of the biotin is accommodated in the interior of the binding pocket formed by residues Gly119 and Gly121 of the β_7-strand. The nitrogen atoms of the *ureido* group hydrogen-bond with the side chain of Gln34 and the backbone carbonyl of Arg38 from the biotin binding loop and with the side chain of Thr14 from the α_1-helix. In addition, the *ureido* carbonyl forms hydrogen-bonds with the hydroxyl group of

Ser13 from the β_1-α_1 loop and the backbone NH of Arg38 (Fig. 3.5, b). The aliphatic tail of the biotin is held by a sandwich of two hydrophobic walls formed by the residues Gly106, Val107, and Leu108 from the β_6-strand on one side, and on the other side, the residues Gly37, Gly39, Trp45, and Leu46 from the biotin binding loop.

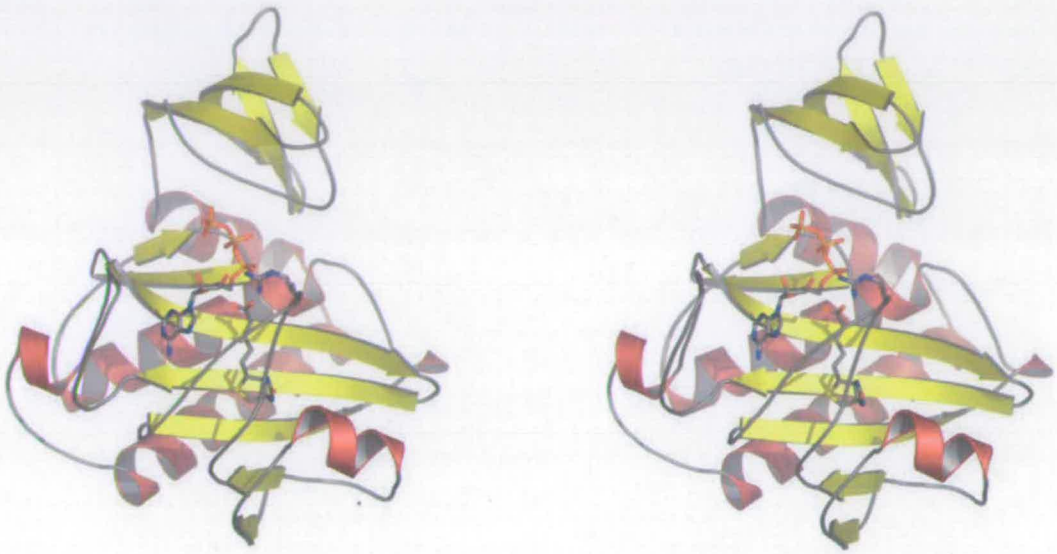


Figure 3.4 – Stereo view of *Aa*BPL bound to biotin and ATP. Cartoon representation of *Aa*BPL monomer A; the ligands ATP, biotin and the Arg40 residue are represented as sticks with grey carbons.

The biotin carboxylate is adjacent to the turn of the biotin binding loop and is strongly stabilised by hydrogen-bonds with the side chain of the Arg40 from the $_{37}$GRGRLG$_{50}$ motif (Fig. 3.4 & Fig. 3.5). Additionally, the carboxyl group interacts weakly with the Arg40 backbone NH and the side chain of Lys103 from the β_6-strand. The biotin carboxylate lies toward the outside of the active site, near the ATP binding pocket. We find that the oxygen atoms of the carboxyl group of the biotin

and the α-phosphate of ATP are located in close proximity (approximately 3.5 Å; figure 3.5, b). Therefore, the nucleophilic attack of the carboxylate anion at the α-phosphate of ATP, resulting in the formation of the phosphoanhydride bond of biotinyl-5'-AMP, induces only minor conformational changes to the active site. Comparison of the structures of the *Aa*BPL:biotin:ATP and *Ph*BPL:biotinyl-5'-AMP complexes shows that the orientation of the biotin and the adenine ring and sugar moiety of ATP within the active sites are similar in the two complexes.[1] In the structure of *Ph*BPL bound to biotinyl-5'-AMP, the intermediate adopts a U-shaped conformation allowing the phosphoanhydride bond to protrude from the adenylate binding site. This conformation exposes it for subsequent reaction with the lysine of BCCP.

Polar and electrostatic interactions with residues Asn123, Ile133 and Arg136 from the adenylate-binding loop and residues Trp45, Leu46 and Ser47 from the biotin binding loop contribute to the correct positioning of the adenine ring of ATP in the active site. The α-phosphate of ATP is stabilised *via* a network of hydrogen-bonding interactions with the side chains of Arg40, Arg43 and Lys103. In addition to these interactions, the crystal packing allows the formation of hydrogen-bonds between the oxygen atoms of the α-phosphate of ATP and the side chain of Lys192 from the second monomer of the asymmetric unit (Fig. 3.5, b).

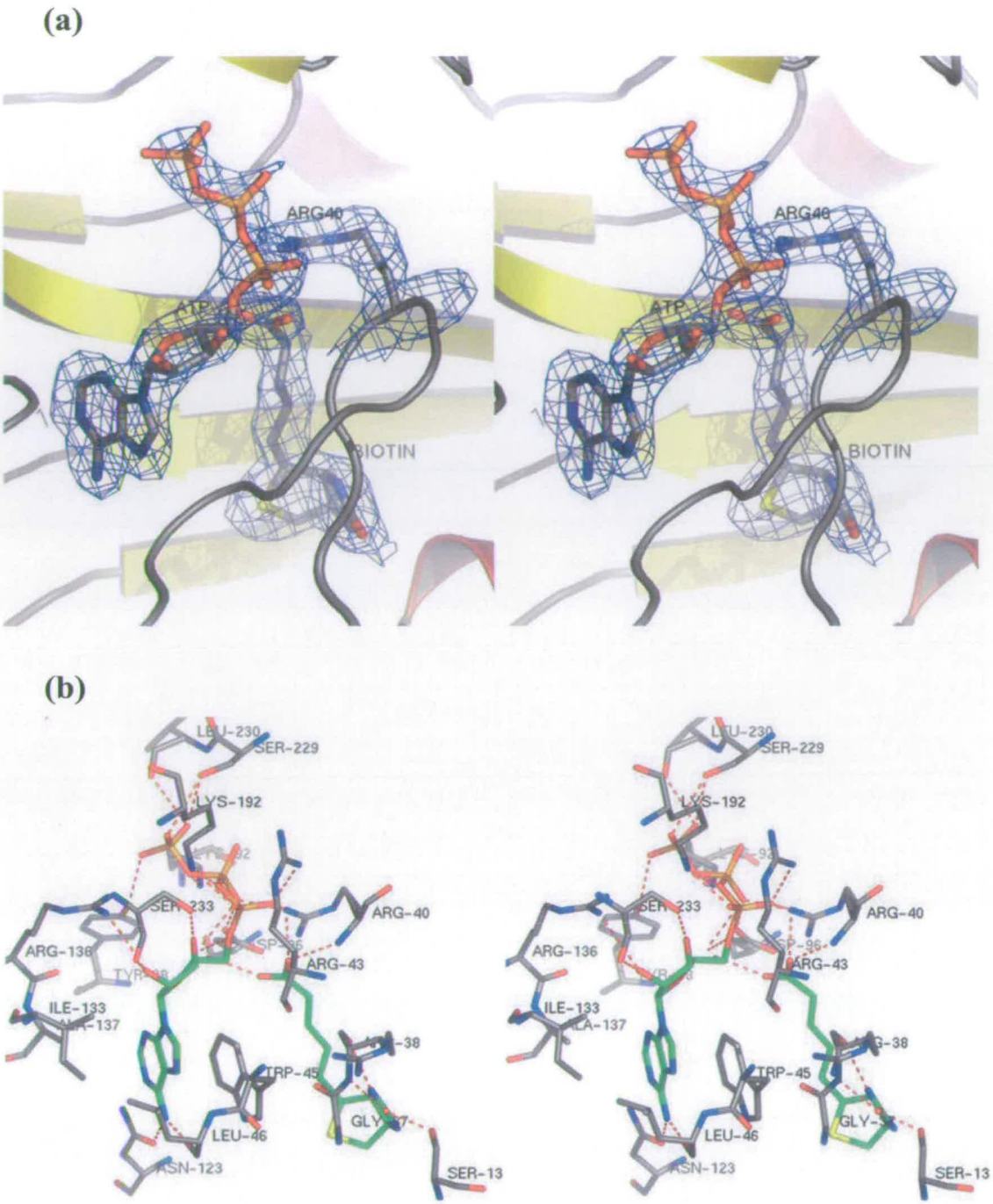


Figure 3.5 –Stereo views of the active site of the *Aa*BPL:biotin:ATP complex. (a) Active site of *Aa*BPL monomer A with the electron density shown around the ligand biotin and ATP and the Arg40 (Blue coloured chickenwire). Final $2|F_o|-|F_c|$ map is contoured at 1σ. **(b)** Binding interactions of ATP and biotin within the *Aa*BPL monomer A active site. ATP and biotin are coloured with green carbons, H-bonds are shown as dashed red lines. Lys192 is from the B monomer.

The structure of the *Aa*BPL:biotin:ATP complex reveals that the β- and γ-phosphates are solvent-exposed and protrude from the adenylate binding site delineated by the β_5 and β_6 strands on one side and the turns of the biotin and adenylate binding loops on the other, and point into the inter-subunit cavity formed by the N- and C-terminal domains (Fig. 3.4). In the crystal, the β-phosphate is stabilised by the hydrophilic environment formed by Asp96 from the β_5-strand and the side chains of Arg40, Lys92 and Lys103. The γ-phosphate interacts weakly with the hydroxyl group of Tyr98 and the side chain of Lys92 from the β_4 and β_5 strands respectively. Interestingly, the structure of the biotin-ATP complex reveals that the hydroxyl group of Ser229 from the β_{12}-strand located on the C-terminal domain also form hydrogen-bonds with the oxygen atoms of the γ-phosphate (Fig. 3.5, b). Although Mg^{2+} can not be visualised, the terminal diphosphate is in a position which allows binding of divalent metal ions. This is presumably a prerequisite for elimination of the pyrophosphate and formation of biotinyl-5'-AMP.

3.3.2 Capture of the Biotin-ATP Intermediate Complex of *Aa*BPL: The Role of the Arginine 40.

The orientation of the side chain of Arg40 and the interactions made by this conserved residue in the structure of the *Aa*BPL:biotin:ATP complex are of particular note. These differ significantly from those of the equivalent Arg118 in BirA and Arg48 in *Ph*BPL.[1; 4; 7] In the biotin-ATP ternary *Aa*BPL complex, Arg40 plays an essential role in stabilising the biotin and ATP ligands within the active site

(Fig. 3.5). The structure of the *Aa*BPL complex reveals that the side chain of Arg40 interacts with the carboxyl group of the biotin and forms direct hydrogen-bonds with the oxygens of the α- and β-phosphates of ATP. The backbone NH of the Arg40 residue also interacts with the carboxyl group of the biotin. This orientation of the conserved arginine side chain has not been observed in other ligand-bound BPL structures.

In the structure of the BirA:biotin complex, the backbone NH of Arg118 forms hydrogen-bonds with the carboxyl group of the biotin and in the complex with biotinol-5'-AMP, the backbone NH interacts with the oxygen atoms of the phosphate group of the analogue.[4; 7] Thus, in the BirA complexes, the side chain of the conserved Arg118 appears to play mainly a structural role by closing the active site and sequestering the ligands *via* weak hydrophilic interactions.

The structure of *Ph*BPL bound to biotin and ATP was recently solved by Bagautdinov *et al.* and the coordinates of the complex have been deposited in the protein databank (PDB entry: 2DTO). In the directly comparable structure of the *Ph*BPL:biotin:ATP complex, the equivalent Arg48 stabilises the α-phosphate of ATP by strong hydrogen-bonding. While in the structure of *Aa*BPL:biotin:ATP, the side chain of the Arg40 adopts a bent conformation in order to interact with the biotin carboxylate, in the *Ph*BPL:biotin:ATP complex, the side chain of the Arg48 points towards the γ-phosphate and has a similar orientation with respect to the triphosphate chain (Fig. 3.6). The conformation of the Arg48 allows the side chain to form a network of weak hydrogen-bonds with the oxygen atoms from the three phosphate groups of ATP. However, neither the backbone NH nor the side chain of the Arg48 interacts with the carboxyl group of the biotin in this structure.

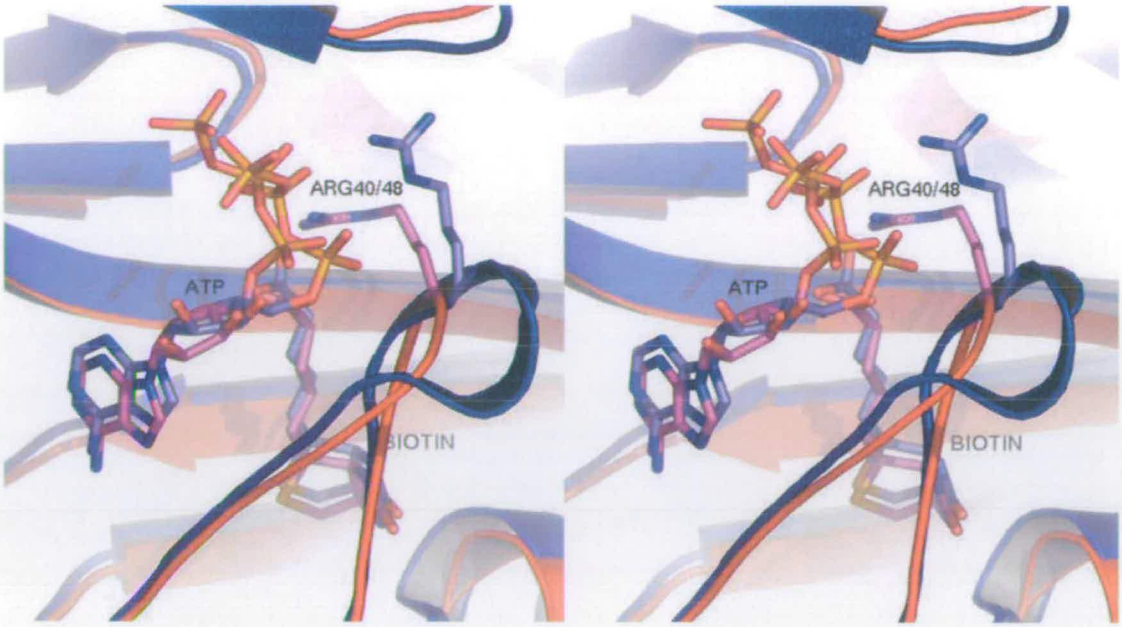


Figure 3.6 - Stereo view of the overlay of the *Aa*BPL and *Ph*BPL biotin-ATP complexes binding sites. *Aa*BPL colouring: red cartoon, ATP, biotin and Arg40 shown as sticks with pink carbons. *Ph*BPL colouring: blue cartoon, ATP, biotin and Arg48 shown as sticks with light blue carbons. The equivalent arginines and ATP are seen to be significantly different in their positions.

In the *P. horikoshii* enzyme, Lys111 from the β_6-strand has been proposed to be involved in the formation of biotinyl-5'-AMP by stabilising the α-phosphate upon formation of the phosphoanhydride bond. In the structure of the *Aa*BPL:biotin:ATP complex, the equivalent conserved Lys103 adopts a similar conformation but here forms weak interactions with the biotin carboxylate and the α-phosphate of ATP which is shifted towards the Arg40.[1] The phosphate chain of bound ATP adopts a slightly different orientation in the active sites of *Aa*BPL and *Ph*BPL while the position of biotin is similar in the two complexes. The main structural differences observed between the *Aa*BPL and *Ph*BPL biotin-ATP complexes are located on the biotin binding loops (Fig. 3.6). The two loops adopt different conformations which

are a result of the sequence differences between their glycine-rich motifs. In comparison to the *Aa*BPL:biotin:ATP structure, in the *Ph*BPL complex, the loop is displaced and the side chain of the Asn50 is solvent-exposed and interacts weakly with His46.

3.4 Discussion

The crystal structures of *Aa*BPL have been determined in the apo-form and in complex with biotin and ATP. The geometry of this complex is revealing and seems similar to that of the intermediate state which is expected to immediately precede biotinyl-5'-AMP formation. In the structure of the *Aa*BPL:biotin:ATP complex, the bulk of both the biotin and the AMP moiety of ATP are buried and the biotin carboxylate and the α-phosphate are located in close proximity to each other at the entrance of the active site. The binding of Mg^{2+} and subsequent nucleophilic attack of the carboxylate anion on the α-phosphate, leading to the formation of the phosphoanhydride bond of the product, would induce only minor changes to the conformation of the active site.

*Aa*BPL displays high structural similarities with BirA and the *P. horikoshii* BPL enzyme.[1; 4] Similarly to *Ph*BPL, the structure of apo-*Aa*BPL reveals only one disordered surface loop.[1] In *Aa*BPL, the peptide segment encompassing residues 37-47 which forms the biotin binding loop becomes organized upon ligand binding and largely encloses the biotin and ATP moieties. The loop appears to act as a partial lid to the active site and protects the ligands from the solvent. On one side of the loop,

Gly37, Gly39 and the aliphatic chain of Arg38 stabilize the hydrophobic tail of the biotin. On the other side, Trp45, Leu46 and Ser47 play important roles in the correct positioning of the adenine ring of ATP in the active site. The Arg40 from the glycine-rich motif $_{37}$GRGRLG$_{42}$ is located at the turn of the loop. The side chain of the conserved arginine interacts with the biotin carboxyl group and together with the Arg42 forms a network of hydrogen-bonds with the α-phosphate of ATP. Arg40 appears to play an important role in positioning the biotin carboxylate in proximity to the α-phosphate and in the neutralisation of the negative charges of the phosphate chain of ATP.

In the structure of the *Aa*BPL:biotin:ATP complex, the β and γ phosphates of ATP are solvent-exposed and pointing towards the cavity formed by the N- and C-terminal domains. The inter-subunit domain can readily accommodate the Mg^{2+} ion required for stabilization and elimination of the diphosphate chain. The oxygen atoms of the ATP γ-phosphate are shown to form long range hydrogen-bonds with the backbone NH of residues Ser229 and Leu230 located on the β_{12} strand of the C-terminal domain and this remains the first direct demonstration that interactions between the C-terminus and ATP occur. Mutagenesis analyses previously carried out on BirA support the interactions between ATP and the C-terminus observed in the *Aa*BPL structure. Single-site mutations within the C-terminal domain of BirA revealed that residues located at some distance from the active site directly affected Mg^{2+}-ATP binding.[8] For example, the affinity of the mutant BirA R317E for MgATP was determined to be 25-fold lower than wild-type BirA. In the structure of the *Aa*BPL:biotin:ATP complex, this conserved arginine, equivalent to the Arg231, is adjacent to Leu230 and the side chain is located in the vicinity of the γ-phosphate.

In the model of the *E. coli* BirA:BCCP87 complex, the turn of the loop of BCCP87 containing the biotinylated Lys122 and the conserved AMKM biotinylation motif is located at the entrance of the inter-subunit cavity formed by the central and C-teminal domains of BirA (***See chapter 2, figure 2.20***).[9] The side chain of Lys122 from BCCP87 is adjacent to the Arg118 of BirA, while the Met121 and Met123 interacts respectively with the corresponding β_5 and β_{12} strands of *Aa*BPL. In the *E. coli* complex, the interior of the inter-subunit cavity of BirA is free from side chain interactions with BCCP87. Therefore, the presence of the terminal diphosphate and Mg^{2+} ion within the cavity would not interfere with the formation of a BPL:BCCP complex. In *Aa*BPL, the Lys19 shown to interact with Glu121 of apo-BCCPΔ67 by chemical cross-linking with EDC and NHS is located on the α_1-helix and is solvent-exposed. Hydrophilic interactions formed by Lys19 and Glu121 of BCCPΔ67 would protect the bottom of the interface of the *Aa*BPL:BCCPΔ67 complex. The catalytic pocket and the ligands appear to be enclosed in the core of the complex which limits loss of biotinyl-5'-AMP by diffusion into the media ensuring that the adenylate is specifically transferred to the target lysine of BCCP.

Additionally, mutation of Lys 277 in BirA, which is spatially in proximity of the Arg317, to a glutamic acid residue was demonstrated to restore the activity of the mutant apo-BCCP87 E119K which was previously shown to be poorly biotinylated by the wild-type enzyme.[8] On the basis of this result, the C-terminal domain of BirA, which had been ascribed no biochemical function, has been proposed to play a role in protein-protein recognition and in mediating the formation of a BirA:BCCP complex. In the structures of the BPLs, the C-terminal domain adopts a Src-homology 3-like fold (SH3) which was also found in the structural homolog phenylalanine tRNA

synthetase (PheRS).[10; 11] The SH3 fold consists of two antiparallel β-sheets, packed against each other at an approximate right angle. Eukaryotic Src homology regions 2 and 3 domains are conserved protein modules of 50-70 amino acids long found in a variety of proteins with important roles in signal transduction pathways.[12; 13] These non-catalytic domains are conserved among a series of cytoplasmic signaling proteins regulated by receptor protein-tyrosine kinases, including phospholipase C-gamma, Ras GTPase (guanosine triphosphatase)-activating protein, and Src-like tyrosine kinases, whereas the SH3 domains of *Aa*BPL, BirA, *Ph*BPL and PheRS are essential constituents of prokaryotic ligases and synthetases. In the structure of *Thermus thermophilus* PheRS, the enzyme undergoes disorder-to-order transition upon aminoacyl and ATP binding similar to that observed in *Aa*BPL and BirA.[14-17] In presence of phenylanyl-AMP, PheRS adopts a closed conformation and the SH3 module is shifted towards the active site. Those concerted conformational changes have been shown to be necessary for the correct order of substrate binding in the overall phenylalanylation reaction.[17] Although the SH3 C-terminal domain of *Aa*BPL appears to undergo only minor conformational changes upon ligand binding, those differences may be sufficient to play a role in the formation of the *Aa*BPL:BCCP complex.

Interestingly, the orientations of the substrates and the triphosphate chain of ATP characterized within the active site of *Aa*BPL are not conserved in the structurally homologous class II tRNA synthetases. In the structures of class II aaRSs, ATP adopts a U-shaped conformation and the β and γ phosphates are solvent-exposed.[18] Investigation of the structures of *T. thermophilus* glycyl-tRNA synthetase bound to ATP and glycyl-AMP showed that the ATP phosphate chain, located at the

entrance of the active site, is bent away from the binding site of the glycyl moiety.[18] The resultant aminoacyl-adenylate adopts an extended conformation which has also been characterized in PheRS complexed with phenylalanyl-AMP.[17] Analyses of the structures of *Staphylococcus aureus* threonyl-tRNA synthetase bound to ATP and Zn^{2+} and to an adenylate analog (threonyl-AMS) as well as the MgATP and alanine complexes of *A. aeolicus* alanyl-tRNA synthetase suggest that this binding mode is common in class II tRNA synthetases.[19; 20]

The U-shaped conformation of ATP has also been demonstrated in the structures of *Thermoplasma acidophilum* LplA (*Ta*LplA).[21] In the complex of *Ta*LplA bound to MgATP, the phosphate chain of the ATP moiety points inward towards the bifurcated lipoyl-AMP binding pocket which is only accessible through a tunnel-like entrance.[21] The β-phosphate occupies roughly the same position as the carboxyl group of lipoic acid inside the cavity. The structure of *Ta*LplA reveals a peptide segment similar to the biotin binding loop which contains a conserved sequence motif $_{71}$RR*XX*GGG*X*V(F/Y)HD$_{82}$. The bent conformation of the triphosphate group is maintained through strong interactions with the side chain of the conserved Arg72 inside the lipoic acid binding pocket.[22; 23] The structure of *Ta*LplA bound to lipoic acid suggests that the Arg72 stabilizes the negative charge of the co-factor carboxylate during ATP binding and formation of the phosphoanhydride bond.[22; 23] If ATP were to bind prior to lipoic acid, the β and γ-phosphate groups of ATP would have to be released from the binding pocket when lipoic acid enters the binding site. The environment surrounding the diphosphate chain in the *Ta*LplA:ATP complex is highly hydrophobic, and thus it would preferentially bind lipoic acid.[21] In *Ta*LplA, the intermediate lipoyl-AMP has been

shown to adopt a bent conformation identical to that described in the *Ph*BPL:biotinyl-5'-AMP and BirA:biotinol-5'-AMP complexes.[1; 7]

3.5 References

1. Bagautdinov, B., Kuroishi, C., Sugahara, M. & Kunishima, N. (2005). Crystal structures of biotin protein ligase from *Pyrococcus horikoshii* OT3 and its complexes: structural basis of biotin activation. *J Mol Biol* **353**, 322-33.
2. Bagautdinov, B., Kuroishi, C., Sugahara, M. & Kunishima, N. (2005). Purification, crystallization and preliminary crystallographic analysis of the biotin-protein ligase from *Pyrococcus horikoshii* OT3. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **61**, 193-5.
3. Clarke, D. J., Coulson, J., Baillie, R. & Campopiano, D. J. (2003). Biotinylation in the hyperthermophile *Aquifex aeolicus*. *Eur J Biochem* **270**, 1277-87.
4. Weaver, L. H., Kwon, K., Beckett, D. & Matthews, B. W. (2001). Corepressor-induced organization and assembly of the biotin repressor: a model for allosteric activation of a transcriptional regulator. *Proc Natl Acad Sci U S A* **98**, 6045-50.
5. Wilson, K. P., Shewchuk, L. M., Brennan, R. G., Otsuka, A. J. & Matthews, B. W. (1992). *Escherichia coli* biotin holoenzyme synthetase/bio repressor crystal structure delineates the biotin- and DNA-binding domains. *Proc Natl Acad Sci U S A* **89**, 9257-61.
6. Chen, Y. & Varani, G. (2005). Protein families and RNA recognition. *Febs J* **272**, 2088-97.
7. Wood, Z. A., Weaver, L. H., Brown, P. H., Beckett, D. & Matthews, B. W. (2006). Co-repressor induced order and biotin repressor dimerization: a case for divergent followed by convergent evolution. *J Mol Biol* **357**, 509-23.
8. Chapman-Smith, A., Mulhern, T. D., Whelan, F., Cronan, J. E., Jr. & Wallace, J. C. (2001). The C-terminal domain of biotin protein ligase from *E. coli* is required for catalytic activity. *Protein Sci* **10**, 2608-17.
9. Weaver, L. H., Kwon, K., Beckett, D. & Matthews, B. W. (2001). Competing protein:protein interactions are proposed to control the biological switch of the *E coli* biotin repressor. *Protein Sci* **10**, 2618-22.
10. Safro, M. & Mosyak, L. (1995). Structural similarities in the noncatalytic domains of phenylalanyl-tRNA and biotin synthetases. *Protein Sci* **4**, 2429-32.
11. Noble, M. E., Musacchio, A., Saraste, M., Courtneidge, S. A. & Wierenga, R. K. (1993). Crystal structure of the SH3 domain in human Fyn; comparison of the three-dimensional structures of SH3 domains in tyrosine kinases and spectrin. *Embo J* **12**, 2617-24.

12. Mongiovi, A. M., Romano, P. R., Panni, S., Mendoza, M., Wong, W. T., Musacchio, A., Cesareni, G. & Di Fiore, P. P. (1999). A novel peptide-SH3 interaction. *Embo J* **18**, 5300-9.
13. Sparks, A. B., Rider, J. E., Hoffman, N. G., Fowlkes, D. M., Quillam, L. A. & Kay, B. K. (1996). Distinct ligand preferences of Src homology 3 domains from Src, Yes, Abl, Cortactin, p53bp2, PLCgamma, Crk, and Grb2. *Proc Natl Acad Sci U S A* **93**, 1540-4.
14. Mosyak, L., Reshetnikova, L., Goldgur, Y., Delarue, M. & Safro, M. G. (1995). Structure of phenylalanyl-tRNA synthetase from *Thermus thermophilus*. *Nat Struct Biol* **2**, 537-47.
15. Goldgur, Y., Mosyak, L., Reshetnikova, L., Ankilova, V., Lavrik, O., Khodyreva, S. & Safro, M. (1997). The crystal structure of phenylalanyl-tRNA synthetase from *thermus thermophilus* complexed with cognate tRNAPhe. *Structure* **5**, 59-68.
16. Reshetnikova, L., Moor, N., Lavrik, O. & Vassylyev, D. G. (1999). Crystal structures of phenylalanyl-tRNA synthetase complexed with phenylalanine and a phenylalanyl-adenylate analogue. *J Mol Biol* **287**, 555-68.
17. Moor, N., Kotik-Kogan, O., Tworowski, D., Sukhanova, M. & Safro, M. (2006). The crystal structure of the ternary complex of phenylalanyl-tRNA synthetase with tRNAPhe and a phenylalanyl-adenylate analogue reveals a conformational switch of the CCA end. *Biochemistry* **45**, 10572-83.
18. Arnez, J. G., Dock-Bregeon, A. C. & Moras, D. (1999). Glycyl-tRNA synthetase uses a negatively charged pit for specific recognition and activation of glycine. *J Mol Biol* **286**, 1449-59.
19. Swairjo, M. A. & Schimmel, P. R. (2005). Breaking sieve for steric exclusion of a noncognate amino acid from active site of a tRNA synthetase. *Proc Natl Acad Sci U S A* **102**, 988-93.
20. Torres-Larios, A., Sankaranarayanan, R., Rees, B., Dock-Bregeon, A. C. & Moras, D. (2003). Conformational movements and cooperativity upon amino acid, ATP and tRNA binding in threonyl-tRNA synthetase. *J Mol Biol* **331**, 201-11.
21. Kim, D. J., Kim, K. H., Lee, H. H., Lee, S. J., Ha, J. Y., Yoon, H. J. & Suh, S. W. (2005). Crystal structure of lipoate-protein ligase A bound with the activated intermediate: insights into interaction with lipoyl domains. *J Biol Chem* **280**, 38081-9.
22. Fujiwara, K., Toma, S., Okamura-Ikeda, K., Motokawa, Y., Nakagawa, A. & Taniguchi, H. (2005). Crystal structure of lipoate-protein ligase A from *Escherichia coli*. Determination of the lipoic acid-binding site. *J Biol Chem* **280**, 33645-51.
23. McManus, E., Luisi, B. F. & Perham, R. N. (2006). Structure of a putative lipoate protein ligase from *Thermoplasma acidophilum* and the mechanism of target selection for post-translational modification. *J Mol Biol* **356**, 625-37.

Chapter 4: The *A. aeolicus* Mutant *Aa*BPL R40G

4.1 Cloning, Expression and Purification of Mutant *Aa*BPL R40G

The structure of *Aa*BPL in complex with biotin and ATP has allowed characterization of the residues important for substrate binding. It is now clear that Arg40 plays a key role in stabilizing and positioning the biotin and ATP moiety within the active site. The equivalent Arg118 in *E. coli* BirA is important for substrate specificity and the BirA R118G mutant was recently demonstrated to biotinylate promiscuously *in vivo* and *in vitro* non-cognate proteins.[1] On the basis of the fact that BirA R118G binds biotinyl-5'-AMP less tightly than wild-type BirA (K_D = 20 nM and K_D = 45 pM respectively), it has been suggested that this could be a direct consequence of non-specific solution reactions of free biotinyl-5'-AMP with exposed lysine residues.[1; 2] However, the mechanism involved in this promiscuous biotinylation is not clearly understood and despite the crystallographic data obtained for the different BirA complexes, the role of the conserved arginine from the glycine motif in the normal catalytic process remains unclear.

Similarly to the *E. coli* mutant BirA R118G, the Arg40 residue of *Aa*BPL was mutated to a glycine to explore the functional role of the interactions made by the side chain of the conserved arginine residue. The *A. aeolicus bpl R40G* gene was produced by site directed mutagenesis and the PCR product was cloned into a pGEM-T easy vector (Promega). The mutation was confirmed by DNA sequencing and the mutated *A. aeolicus bpl R40G* gene was subsequently cloned into the pET28a expression vector using the restriction enzymes *NcoI* and *BamHI* (Fig. 4.1).

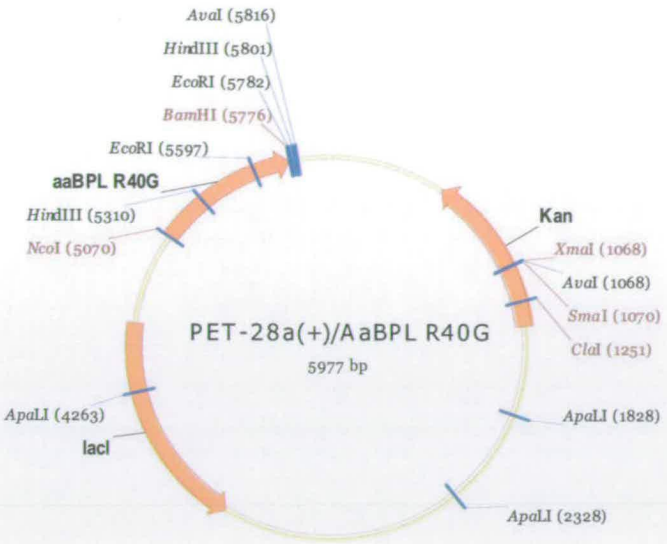


Figure 4.1 - Expression vector pET28a/*Aa*BPL R40G. The mutated *Aquifex bpl R40G* gene was cloned using the *NcoI* and *BamHI* restriction sites.

The resulting construct was then transformed into *E. coli* BL21 (DE3) cells at 37 °C and induction was performed with 1 mM IPTG. The expressed protein was purified by cation exchange chromatography in an identical manner to wild-type *Aa*BPL. The mutant enzyme eluted with 150 mM NaCl on the HR 10/100 15S column with a purity of more than 95% as determined by SDS-PAGE (Fig. 4.2). The resulting yield of pure *Aa*BPL R40G was greater than that of the wild-type enzyme (~ 10 mg / litre of cell culture). Gel filtration analysis of *Aa*BPL R40G carried out on a HiLoad 16/60 Superdex75 column indicated that like the wild-type enzyme, the mutant R40G is monomeric in the apo-form and in the presence of substrates.

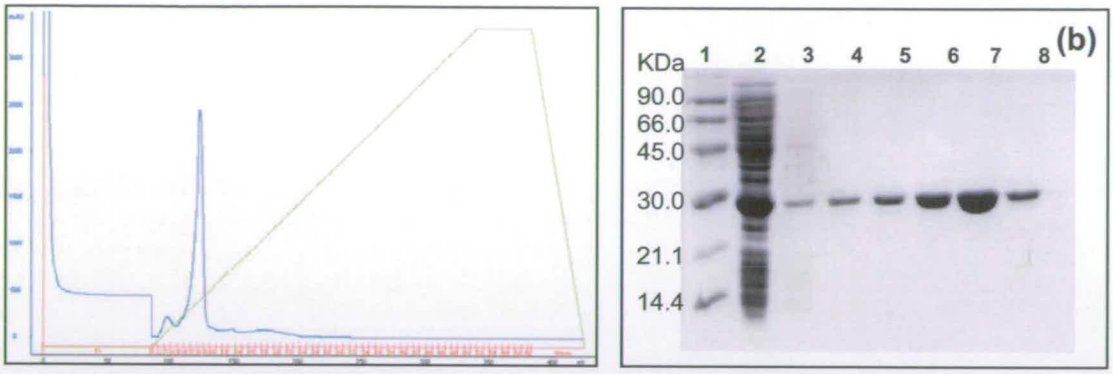


Figure 4.2 - Purification and SDS-PAGE analysis *Aa*BPL R40G. (a): Chromatogram of *Aa*BPL R40G purification on a HR 10/100 15S column. The protein eluted with 150 mM NaCl. **(b):** Protein purification was analysed by SDS-PAGE under reducing conditions. *Lane 1*, low molecular weight marker (Amersham), *lane 2, Aa*BPL R40G flow-through, *lane 3,* wash fraction, *lane 4-8,* eluted fractions 5 to 9.

Analysis by LC-ESI-MS confirmed the purity of the protein as a single species with a measured mass of 26,542 ± 3 Da, which is in agreement with the predicted molecular weight of 26,535.43 Da. Western blotting analyses with streptavidin showed that purified *Aa*BPL R40G was not biotinylated even after incubation with biotin at 60 °C. However, addition of ATP and Mg^{2+} to the incubation mixture resulted in hyper-biotinylation of the mutant indicating that *Aa*BPL R40G remained capable of catalyzing the formation of biotinyl-5'-AMP and had modified itself (Fig. 4.3). In contrast, previous streptavidin western blotting studies had shown that the equivalent *E. coli* mutant BirA R118G was self-biotinylated *in vivo.* The fact that *Aa*BPL R40G remained unbiotinylated when expressed in *E. coli* could be attributed to the temperature dependence of the biotinylation reaction in *A. aeolicus*.[1] Activity assays performed by Clarke *et al.* with native *Aa*BPL at varying temperatures indicated that the optimal activity at 70 °C

decreased by roughly 50% for every 10 °C drop in temperature with almost no activity detected at 37 °C.[3] Increasing the temperature above 70 °C resulted in enzyme precipitation with a dramatic loss in activity.

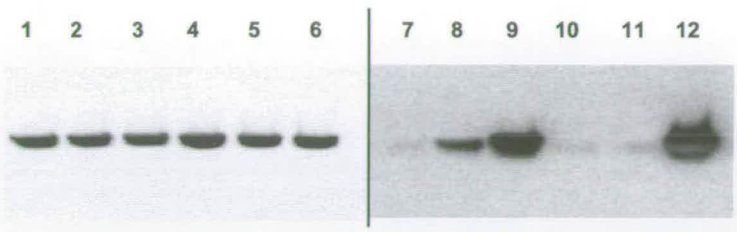


Figure 4.3 – Streptavidin western blot analysis of wild-type *Aa*BPL and mutant *Aa*BPL R40G. *Left panel (lane 1-6): SDS-PAGE; right panel (lane 7-12): streptavidin western blot. Lane 1-3 & 7-9, Aa*BPL; *lane 4-6 & 10-12, Aa*BPL R40G; *lane 2, 4, 8, 11,* incubation in presence of biotin at 60 °C; *lane 3, 6, 9, 12,* incubation in presence of biotin, ATP, $MgCl_2$ at 65 °C.

For comparison by mass spectrometry, *E. coli* BirA R118G was also expressed in an ampicillin-resistant pBTac vector in *E. coli* JM109 competent cells. Induction of the *birA R118G* gene was carried out with 0.4 mM IPTG at 30 °C and the resulting C-terminal hexahistidine-tagged soluble protein was purified by nickel affinity chromatography. LC-ESI-MS analysis of BirA R118G indicated that the purified enzyme mixture was highly heterogeneous (Fig. 4.4). A number of different species with different elution times on the C5 column could be identified with molecular masses ranging from 36,078 ± 11 Da (predicted molecular weight of His_6-BirA R118G = 36,133.02 Da) to approximately 37,500 Da which corresponds to the addition of up to seven biotin molecules. ESI-MS analysis by direct infusion revealed a series of species with increasing molecular weight and intensities confirming that, in contrast to *Aa*BPL R40G, BirA R118G was biotinylated *in vivo*.

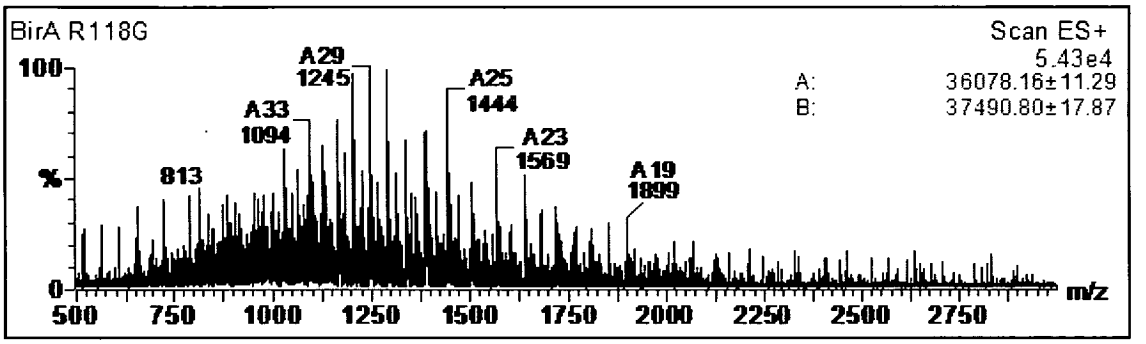


Figure 4.4 – LC-ESI mass spectrum of BirA R118G. The mixture is heterogenous indicating the presence of non-modified and of biotinylated species (predicted mass of BirA R118G = 36,133.02 Da).

4.2 Crystal Structure of *Aa*BPL R40G in Complex with Biotin

4.2.1 Crystallization of *Aa*BPL R40G

The mutant *Aa*BPL R40G was initially crystallized under similar conditions as the native enzyme and optimization was carried out by varying the PEG concentrations across a pH range of 5.5 to 6.5. Crystallization of the mutant was performed at a concentration of 6 mg. ml^{-1} and the mutation was shown to have a dramatic impact on the growth of the crystals. The rapid and extensive nucleation observed with native *Aa*BPL was significantly reduced and it took 4 days to two weeks for the crystals to grow reproducibly with a rectangular shape to a size of approximately 0.5 mm (Fig. 4.5). X-ray diffraction at BM14 at the ESRF Grenoble allowed data for the crystals of apo-*Aa*BPL R40G to be collected to a resolution of 2.3 Å. The structure of the apo-R40G mutant indicated that the mutated biotin binding loop was disordered and thus the mutation itself could not be visualised in

the electron density. The apo-structures of both the native and mutant enzymes are consequently equivalent.

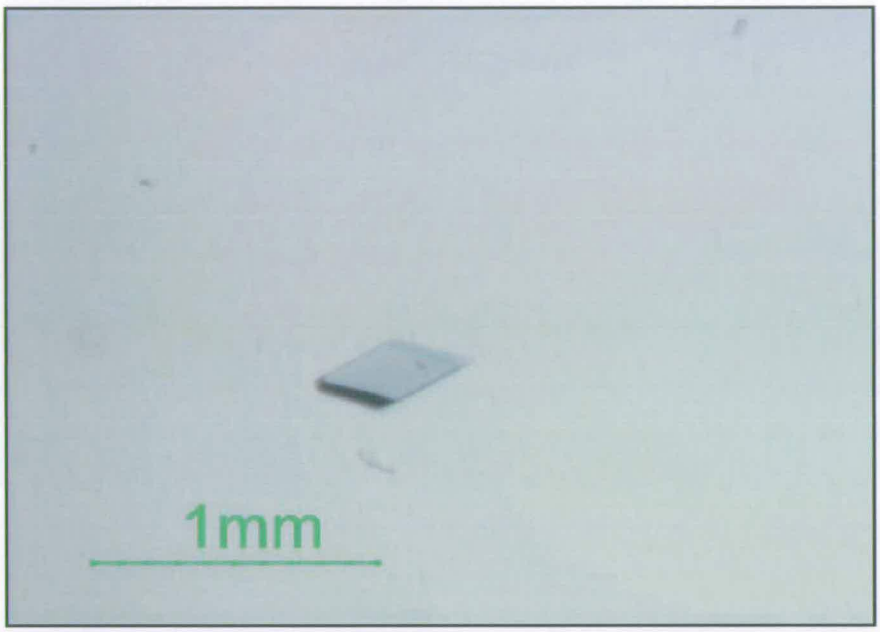


Fig.4.5 –Crystal of mutant *Aa*BPL R40G. The mutation R40G was found to have a dramatic impact on the nucleation process.

*Aa*BPL R40G was also co-crystallized with biotin (2 mM) and ATP (5 mM). The structure of the mutant co-crystallized with ATP indicated no electron density for the nucleotide or for the biotin binding loop suggesting that the mutation might have affected ATP binding. Using a single co-crystal with biotin which diffracted to 2.55 Å at the BM14 at the ESRF Grenoble, the structure of the mutant bound to the ligand was solved. This revealed a clear electron density for the R40G containing loop and for the bound biotin. The crystals of the *Aa*BPL R40G:biotin complex are orthorhombic and belong to the space-group $P2_12_12_1$. The coordinates of this mutant structure were refined to a crystallographic residual R_{CRYST} = 20.00% and R_{FREE} =

29.31% at 2.55 Å resolution (***See chapter 3, table 3.1***). Co-crystallization with both biotin and ATP gave very fragile crystals with different shapes and sizes and X-ray diffraction measurements indicated that they were disordered. Optimization of the co-crystals screens at different pH and temperature was unsuccessful.

4.2.2 Crystal Structure of *Aa*BPL R40G in complex with biotin

In the crystal structure of the mutant *Aa*BPL R40G determined in complex with biotin, the clear density for the R40G containing loop and biotin indicates that biotin binding stabilises the loop (Fig. 4.6, a). The main conformational differences observed between the structures of the mutant and the wild-type enzymes are located within the active site. The biotin binding mode of the *Aa*BPL R40G:biotin complex closely resembles that of the ATP-biotin complex of the wild type enzyme. While the *ureido* group and thiophene ring show near identical interactions with the wall formed by the β_6 and β_7 strands in both structure, in the mutant the carboxylate group of biotin is shifted towards the side chains of Lys103 and Asp96 on the β_6 and β_7 strands (Fig. 4.6).

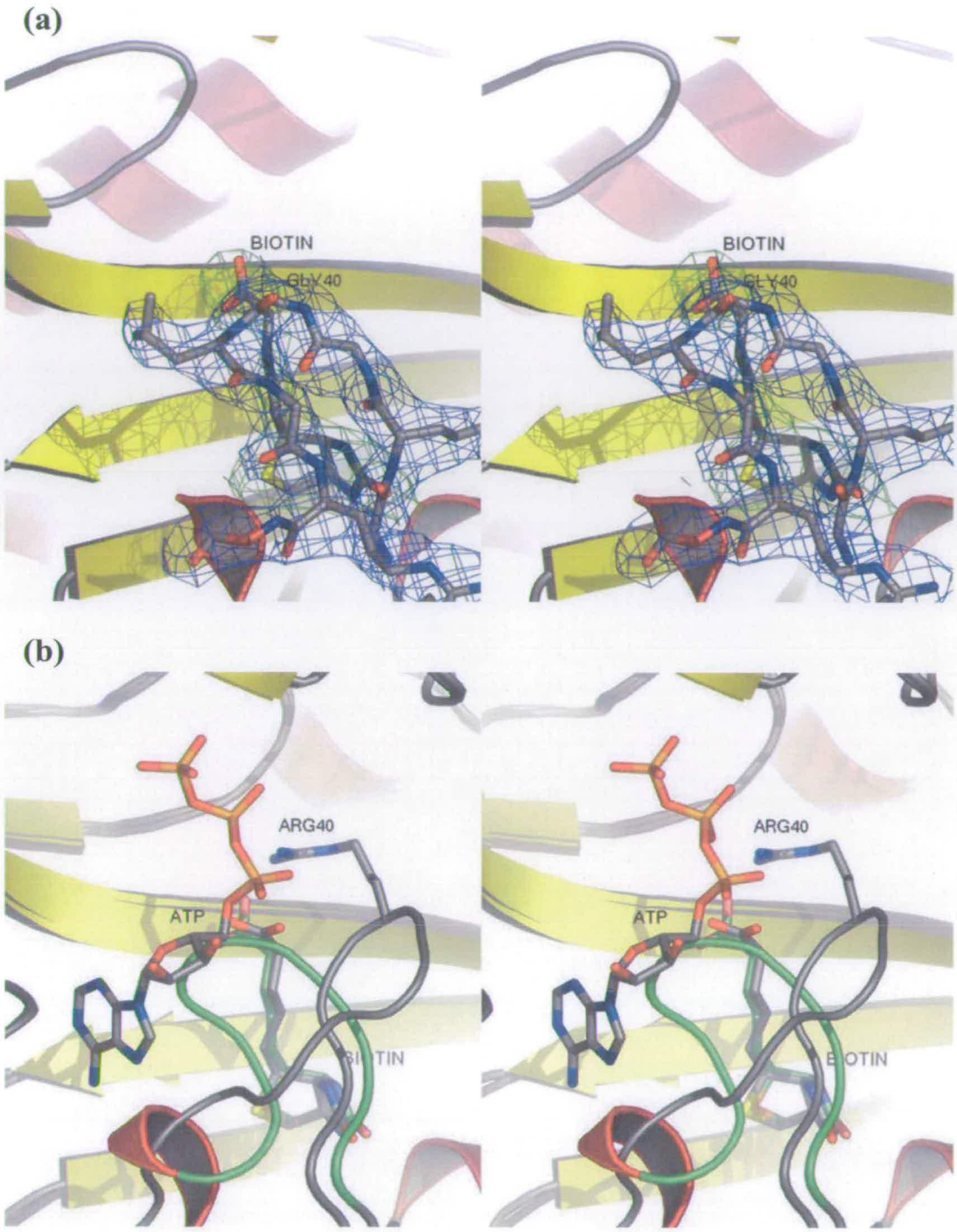


Figure 4.6 - Stereo views of the mutant *Aa*BPL R40G:biotin complex binding site. (a) Electron density is shown around the R40G containing loop (blue coloured chickenwire) and the biotin ligand (green coloured chickenwire). Final $2|F_o|-|F_c|$ map contoured at 1σ is shown. **(b)** Comparison of the loop positions between the mutant *Aa*BPL R40G:biotin complex and native *Aa*BPL:biotin:ATP complex. ATP, biotin and Arg40 are shown as sticks in the native structure with grey coloured carbons. The mutant loop and complexed biotin are coloured green. The loop position of *Aa*BPL R40G can be clearly seen to preclude ATP binding.

The biotin binding loop (residues 37-47) containing the R40G mutation adopts a dramatically different conformation to the one observed in the *Aa*BPL:biotin:ATP wild-type complex structure (Fig. 4.6, b). As a result, the interactions with the glycine-rich motif $_{37}$GRGGLG$_{42}$ have changed significantly and the ATP binding site is now blocked by side chain interactions. In the structure of the mutant, the side chain of Arg38 adopts a different orientation and points towards the oxygen atoms of the *ureido* group. The carboxylate group of the biotin interacts *via* hydrogen-bonds with the backbone NH of the mutated Gly40, which occupies the binding site of the α-phosphate of ATP. The backbone and side chain of Leu41 have shifted towards the active site and now blocks ribose sugar binding.

The strong hydrogen bond formed between the backbone NH of the Gly42 and the backbone carbonyl of the Gly39 stabilises the turn of the loop formed by the aliphatic mutated GRGGLG motif. While in the structure of the native complex the side chain of the Arg43 interacts with the α-phosphate of ATP, in the structure of the mutant complex the side chain of this residue points towards the solvent. The loop is bent at the location of the Arg43 and the side chain of the Trp45 interferes with the positioning of the adenine ring of ATP. Furthermore, the side chain of Leu46, which contributes to the stabilisation of ATP in the native *Aa*BPL complex, interacts with the aliphatic chain of the biotin in the structure of the mutant complex and hinders the binding of the adenine ring of ATP in the active site.

4.3 Binding Studies of *Aa*BPL R40G

To investigate the role of Arg40 in the ligand binding process, isothermal titration calorimetry experiments were also carried out with *Aa*BPL R40G. Titrations of the mutant under the same experimental conditions as used for wild-type *Aa*BPL indicated that *Aa*BPL R40G binds biotin in an exothermic process (Fig. 4.7, A). The affinity for biotin was determined to be only 2.5-fold lower than that observed for the native enzyme suggesting that, in contrast to the *E. coli* mutant BirA R118G, the R to G mutation has little effect on biotin binding in *A*aBPL (Table 4.1).

Thus, while the affinity of BirA for biotin is approximately a hundred-fold greater than that of *Aa*BPL, the dissociation constants for the biotin complexes of the two mutants BirA R118G and *Aa*BPL R40G are similar (K_D = 1.8 μM and K_D = 8.3 μM respectively).[2] Biotin was also shown to bind to *Aa*BPL R40G with a K_D = 15.4 μM when the mutant was titrated in the presence of ATP (Fig. 4.7, B). This small difference, a factor of *ca* 1.9, between the two measured dissociation constants for *A*aBPL R40G indicates that the presence of ATP does not significantly alter biotin binding. The thermodynamic parameters derived from the titrations shows that both apparent enthalpy and entropy changes are favourable for biotin binding to the mutant and remain similar in the presence of ATP.

In contrast to the biotin data, titrations of the mutant enzyme with ATP showed only small signals, principally due to heat of dilution, indicating that binding of ATP to *Aa*BPL R40G is negligible (K_D > 200 μM) under these conditions (Fig. 4.7, C). Thus, the side chain of the Arg40 appears to be necessary for ATP binding in

the absence of Mg^{2+}. Titration of biotin-bound mutant with ATP also showed very weak binding (K_D > 200 μM; figure 4.7, D).

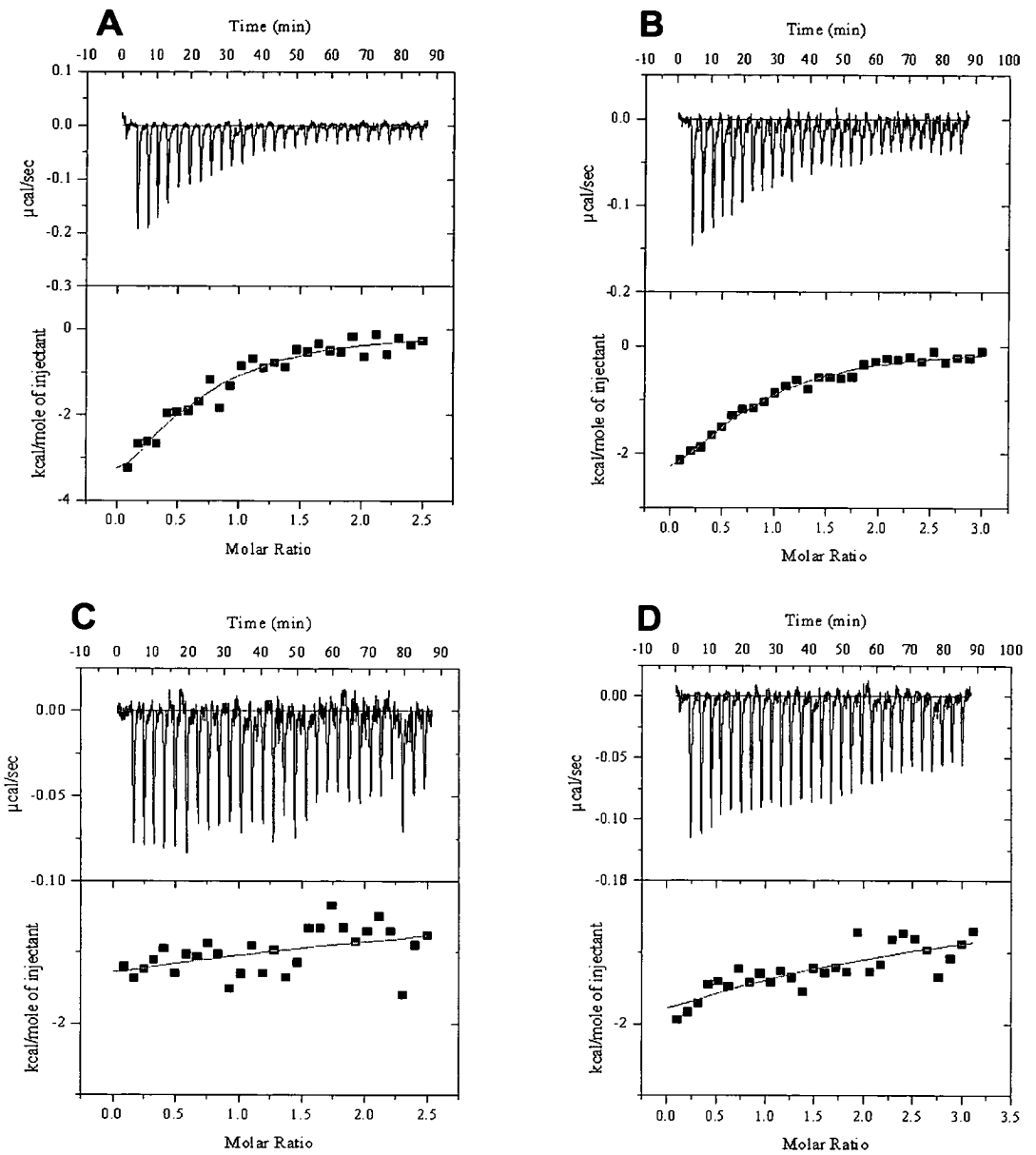


Fig.4.7 - Isothermal titration calorimetry data of *Aa*BPL R40G. Titration ***(A)*** with biotin ***(B)*** in presence of ATP and ***(C)*** with ATP ***(D)*** in presence of the binary complex *Aa*BPL R40G:biotin. All buffers contained 10 mM HEPES pH 7.5, 4 mM EDTA

	K_D	ΔH	$-T\Delta S^o$	ΔG^o
	μM	$kJ\ mol^{-1}$	$kJ\ mol^{-1}$	$kJ\ mol^{-1}$
*Aa*BPL R40G + biotin	**8.3 ± 2.1**	**-23.4 ± 6.3**	**-5.4 ± 15.9**	**-28.9 ± 9.7**
*Aa*BPLR40G:biotin + ATP	**>200**	***Nd***	***Nd***	***Nd***
*Aa*BPL R40G + ATP	**>200**	***Nd***	***Nd***	***Nd***
*Aa*BPL R40G:ATP + biotin	**15.4 ± 3.7**	**-24.7 ± 10.0**	**-2.5 ± 18.8**	**-27.2 ± 8.8**
BirA R118G + biotin	*1.8 ± 0.4*	*-39.3 ± 1.3*	*-7.1 ± 1.3*	*-32.2 ± 0.4*

Table 4.1: Thermodynamic parameters of biotin and ATP binding to *Aa*BPL R40G. The data are derived from the ITC measurements of *Aa*BPL R40G with biotin and ATP in the absence and in the presence of the second ligand. The error in ΔH is ± 5% and is mainly due to the differences in enzyme and ligand concentrations. ΔG° is calculated from the binding determined by ITC: $\Delta G^\circ = -RT\ \ln K_A = \Delta H - T.\Delta S^\circ$. The thermodynamic parameters of *E. coli* BirA R118G binding to biotin are also reported for comparison.[2; 4]

4.4 Substrate Biotinylation with *Aa*BPL R40G

It has been shown that the C-terminal domain of *A. aeolicus* BCCP (BCCPΔ67) is a substrate for biotinylation with *Aa*BPL. Streptavidin western blot studies indicated that incubation of apo-BCCPΔ67 with *Aa*BPL or with the mutant *Aa*BPL R40G in presence of biotin, ATP and $MgCl_2$ for 20 min at 65 °C led to the formation of biotinylated-BCCPΔ67 (Fig. 4.8). On the basis of the western blots of the reactions, the R40G mutant appears to be a poorer catalyst than the wild type enzyme (Fig. 4.8, lanes 9 & 10). Western blot analysis also revealed that the mutant substrate BCCPΔ67 K117L, which lacks the target Lys117, is not biotinylated by *Aa*BPL or *Aa*BPL R40G showing that none of the remaining four lysine residues

were derivatized (Fig. 4.8, lanes 12-14). This suggests a significant degree of target specificity.

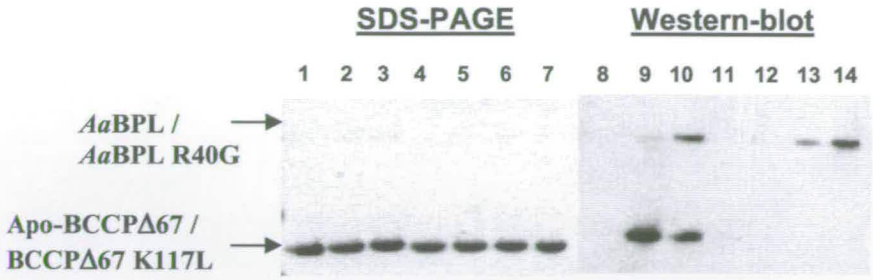


Figure 4.8 – Streptavidin western blot of apo-BCCPΔ67 and BCCPΔ67 K117L biotinylated with *Aa*BPL and *Aa*BPL R40G. Left panel (lane 1-7): SDS-PAGE; right panel (lane 8-14): western blot. *Lane 1 & 8,* apo-BCCPΔ67; *lane 2 & 9,* apo-BCCPΔ67 incubated with *Aa*BPL in presence of biotin, ATP, $MgCl_2$ at 65 °C for 20 min; *lane 3 & 10,* apo-BCCPΔ67 incubated with *Aa*BPL R40G in the same conditions; *lane 4 & 11*, BCCPΔ67 K117L; *lane 5 & 12*, BCCPΔ67 K117L incubated with *Aa*BPL in presence of biotin, ATP, $MgCl_2$ at 65 °C; *lane 6 & 13*, BCCPΔ67 K117L incubated with *Aa*BPL R40G in the same conditions; *lane 7 & 14*, BCCPΔ67 K117L incubated with *Aa*BPL R40G in presence of 100 μM biotin, ATP, $MgCl_2$ at 65 °C.

Biotinylation of apo-BCCPΔ67 with *Aa*BPL R40G was also monitored by mass spectrometry (Fig. 4.9). LC-ESI-MS analysis of the biotinylated BCCPΔ67 from the R40G catalysed reaction revealed the formation of only one species with a measured mass of 10,967 ± 2 Da which is in agreement with the predicted molecular weight of holo-BCCPΔ67 (10,965.94 Da). The increased mass of approximately one biotin (+227 Da) confirmed the formation of only single-biotinylated BCCPΔ67 species.

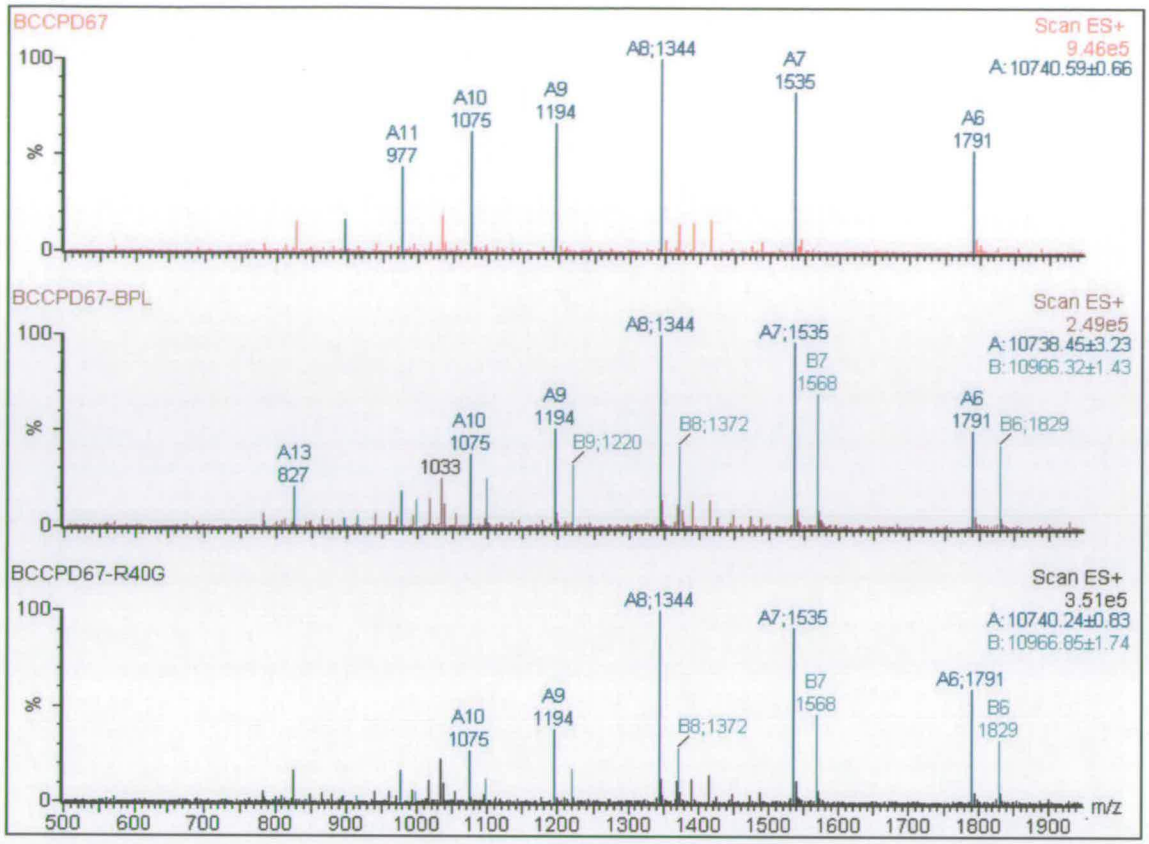


Figure 4.9 - LC-ESI mass spectra of apo-BCCPΔ67 biotinylated with *Aa*BPL and *Aa*BPL R40G. The mass spectra show the +6 to +11 charge states and the charge deconvoluted masses. **(a):** apo-BCCPΔ67 incubated in presence of biotin, ATP, $MgCl_2$ at 65°C for 15 min. **(b):** apo-BCCPΔ67 incubated with *Aa*BPL in the same conditions. **(c):** apo-BCCPΔ67 incubated with *Aa*BPL R40G in the same conditions. Predicted mass for apo-BCCPΔ67: 10,739.63 Da; predicted mass for holo-BCCPΔ67: 10,965.94 Da

However, while the wild-type *Aa*BPL was shown to be stable during the biotinylation assays, the mutant *Aa*BPL R40G was highly modified in presence of biotin and MgATP at a temperature above 45 °C. Western blots of these reactions revealed that, in contrast to the wild-type enzyme, the mutant *Aa*BPL R40G self-biotinylates (Fig. 4.8). This was confirmed by matrix assisted laser desorption-ionization mass spectrometry (MALDI-MS) analysis of the R40G species which

showed mass increases after incubation ranging from ~200 to ~2000 Da indicating that the mutant, which has 29 lysine residues, undergoes multiple biotinylation (Fig. 4.10). The extent of auto-biotinylation observed by MALDI-MS was directly affected by the concentration of biotin and the incubation time as well as temperature. Furthermore, the intensity of the *Aa*BPL R40G blot, given that the protein concentrations of R40G and BCCPΔ67 were 10/1 in this experiment, supports the multiple nature of R40G self-biotinylation (Fig. 4.8, lane 10).

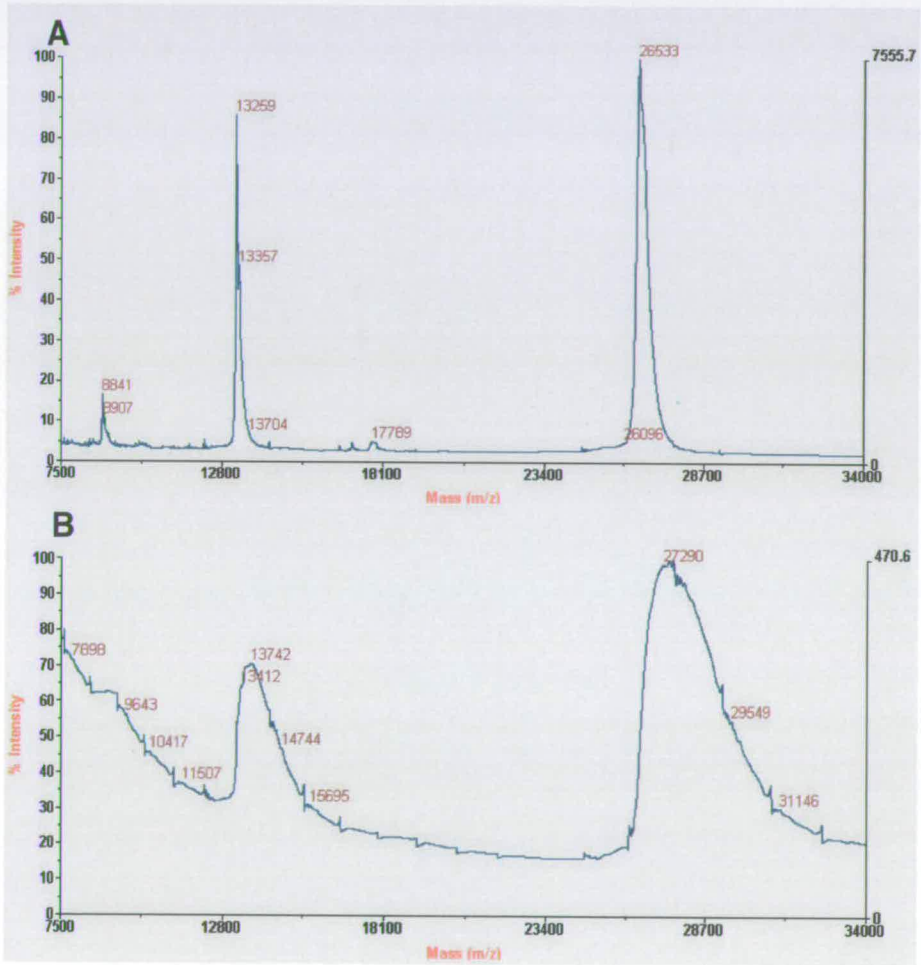


Figure 4.10 – MALDI mass spectra of self-biotinylated *Aa*BPL R40G. A: *Aa*BPL R40G, predicted mass = 26,535.43 Da. **B:** *Aa*BPL R40G incubated with 100 µM biotin, ATP, $MgCl_2$ at 65 °C for 15 min. The observed molecular weight of 27,290 Da corresponds to a mass increase of 757 Da (approximately 3 biotin molecules).

The fact that BCCPΔ67 and BCCPΔ67 K117L are not multiply biotinylated is supported by chemical biotinylation studies with free biotinyl-5'-AMP. Chemical biotinylation of *E. coli* BCCP87 with synthetic biotinyl-5'-AMP has been previously demonstrated and MALDI-MS analysis of the digested fragments of biotinylated BCCP87 indicated that only the target Lys122 had been modified.[5] The five lysines of BCCP are all relatively solvent-accessible and therefore, the specific biotinylation of BCCP87 with free adenylate has been claimed to be due to the intrinsic reactivity of the target lysine to nucleophilic attack. The residue has been suggested to have a lower pKa although it has not, apparently, been measured.[5] The interactions of the *ureido* ring of the biotin moiety covalently attached to Lys122 with residues of the "thumb" region present in both *E. coli* and *A. aeolicus* BCCP may contribute to the stability of the BCCP:biotin complex.[3; 5; 6]

To determine whether *Aa*BPL and *Aa*BPL R40G could carry out non-specific biotinylation, the enzymes were incubated with biotin, ATP and $MgCl_2$ in the presence of bovine serum albumin (Pierce) at 65 °C. Western blot analysis of the products revealed that, while biotinylation of BSA did not occur in the presence of substrates alone or at low concentration of wild-type *Aa*BPL, the mutant *Aa*BPL R40G biotinylated itself and BSA in a manner similar to BirA R118G (Fig. 4.11).[1] However at high enzyme concentrations, the wild-type *Aa*BPL also carried out self-biotinylation and non-specific biotinylation of BSA, albeit less effectively. The fact that *Aa*BPL R40G appears to self-biotinylate preferentially rather than derivatise the substrates BSA or BCCPΔ67 is likely to be a consequence of the elevated temperature of the reaction with thermostable enzymes (***See figures 4.8 and 4.11***). Since chemical biotinylation with biotinyl-5'-AMP is proximity dependent and the

lifetime of AMP mixed anhydrides in aqueous solution (typically $t_{1/2} \sim 1$ min at pH 7, 25 °C) will be significantly shorter at 65 °C, the amount of reagent surviving to derivatise a non-cognate acceptor will be significantly less.[1; 7]

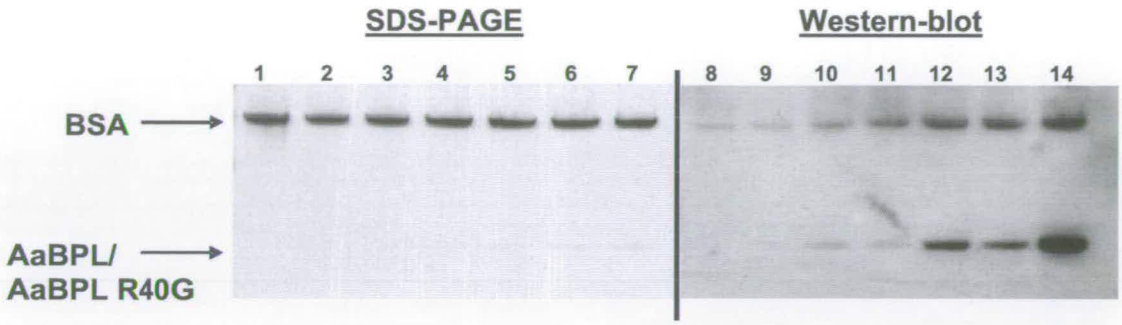


Figure 4.11 - Streptavidin western blot of BSA biotinylated with *Aa*BPL and *Aa*BPL R40G. *Left panel (lane 1-7): SDS-PAGE; right panel (lane 8-14): streptavidin western blot.* BSA (2 μM) incubated in presence of biotin, ATP, $MgCl_2$ at 65 °C for 20 min; *lane 1 & 8,* BSA; *lane 2 & 9*, BSA with 50 nM *Aa*BPL; *lane 3 & 10,* BSA with 50 nM *Aa*BPL R40G; *lane 4 & 11,* BSA with 200 nM *Aa*BPL; *lane 5 & 12*, BSA with 200 nM *Aa*BPL R40G; *lane 6 & 13*, BSA with 500 nM *Aa*BPL; *lane 7 & 14*, BSA with 500 nM *Aa*BPL R40G.

4.5 Discussion

The conserved Arg40 of *Aa*BPL has been mutated to a glycine and the resulting mutant enzyme *Aa*BPL R40G characterized. The crystal structures of *Aa*BPL R40G in both the apo- and biotin-bound forms have been determined. The mutated loop is ordered upon biotin binding but is shown to adopt a dramatically different conformation to that observed in the wild-type *Aa*BPL:biotin:ATP complex. In the structure of the mutant in complex with biotin, the R40G containing loop has shifted towards the active site in order for the backbone NH of Gly40 to stabilize the

biotin carboxylate. The collapse of the biotin binding loop results in side chain interactions from residues located on the loop, which block the ATP binding pocket.

To investigate the effects of the R40G mutation on substrate binding in solution, ITC measurements were carried out with *Aa*BPL R40G to determine the thermodynamic parameters for binding biotin and ATP. The ITC results revealed that the affinity for biotin is relatively unaffected by the R40G mutation. In contrast, the thermodynamic data indicate that binding of ATP to *Aa*BPL R40G is very weak in the absence or presence of biotin (K_D > 200 μM) and the free enthalpy and entropy changes measured during the ATP titrations are too high to be considered. The R40G mutant appears to have lost its ability to sequester ATP within the catalytic pocket. Moreover, in the structure of the mutant co-crystallized with ATP, the biotin binding loop and the nucleotide could not be visualized in the electron density. Thus, the side chain of the Arg40 is essential in order for the active site binding loop to adopt the correct conformation for substrate binding and stabilization of the ATP moiety inside the binding pocket.

The equivalent *E. coli* mutant BirA R118G was recently shown to biotinylate itself as well as a number of cellular proteins *in vivo* but also to modify non-cognate proteins *in vitro*.[1] The promiscuous biotinylation observed with BirA R118G was suggested to arise as a result of dissociation of the BirA R118G:biotinyl-5'-AMP complex to give free adenylate which reacts with available amino groups. Interestingly, purified *Aa*BPL R40G was not biotinylated *in vivo* but was shown to biotinylate extensively itself by streptavidin western blot and mass spectrometry when incubated in presence of biotin and MgATP at temperatures above 50 °C. Although structural and thermodynamic studies of *Aa*BPL R40G have shown that the

mutation directly affects ATP binding and the subsequent formation of the *Aa*BPL R40G:biotin:ATP ternary complex, the R40G mutant remains catalytically active. In the presence of high concentrations of ATP, its reaction with bound biotin, which is correctly positioned for attack, can be envisaged as a result of conformational flux of the active site. The resulting *Aa*BPL R40G:biotinyl-5'-AMP binary complex dissociates prematurely and the reactive adenylate is released from the active site of the mutant.

*Aa*BPL R40G was also shown to biotinylate the non-cognate protein BSA which suggests diffusion of biotinyl-5'-AMP in the media. However, the extent of self-biotinylation are significantly greater that the non-specific biotinylation of BSA. Biotinyl-5'-AMP is a mixed anhydride that readily hydrolyses in solution at neutral pH and elevated temperatures and *Aa*BPL R40G has an unusually high proportion of lysine residues (12.5% frequency). The reactive biotinyl-5'-AMP released from the active site of *Aa*BPL R40G after its formation will therefore preferably biotinylate one of the mutant's own solvent exposed lysine side chains located in the vicinity of the biotin binding pocket.[1]

Because of the high reactivity of anhydrides towards primary amines, similar modifications have also been characterised with aminoacyl-adenylates in the native *E. coli* methionyl-tRNA synthetase (MetRS) and valyl-tRNA synthetase (ValRS). Analysis of the modified peptide fragments of MetRS by MALDI-MS indicated that the lysines covalently attached to [^{14}C]methionine are distributed at the surface of the binding site regions of the three substrates, methionine, ATP, and tRNA.[8] Several different aminoacyl groups have been used to modify ValRS and determine the residues involved in substrate binding.[9] The activity of the enzymes MetRS and

ValRS were shown to be directly affected by these modifications being decreased significantly upon methionylation and valylation respectively.[8; 9] Only the putative editing site of ValRS was modified with [^{14}C]Thr and [^{14}C]Met and the enzyme activity stayed unchanged.[9]

Substrate biotinylation with *Aa*BPL R40G was also carried out with apo-BCCPΔ67 and BCCPΔ67 K117L. The mutant was shown to biotinylate only the target Lys117 in a manner similar to the wild-type enzyme. The specific biotinylation of BCCPΔ67 is supported by biotinylation studies with *E. coli* BCCP87 which showed that only the conserved lysine was modified in presence of free biotinyl-5'-AMP, albeit less effectively than in the enzymatic reaction catalysed by BirA.[5] The unusual reactivity of the target lysine of BCCP has been attributed to intrinsic properties of the residue which, together with the self-biotinylation of the mutant, appears to prevent most of the chemical biotinylation with non-specific acceptors generated upon dissociation of the *Aa*BPL R40G:biotinyl-5'-AMP complex. The extent of self-biotinylation observed with the *A. aeolicus* mutant remains very high however and chemical biotinylation was observed at temperatures below 4 °C when the concentration of mutant and biotin was adjusted. Thus, using *Aa*BPL R40G in order to generate active biotinyl-5'-AMP and biotinylate target peptides and proteins in a proximity dependant process could be envisaged.[10]

This study sheds more light on the effects of mutation within the biotin binding loop on substrate specificity of the bacterial BPLs and provides some insight into the molecular basis of human multiple carboxylase deficiency syndrome (MCD). MCD is a rare autosomal recessive disease characterized by a combination of severe biochemical and clinical disorders which can lead to coma and death if not treated

with a pharmaceutical dose of biotin.[11] This disease is linked to mutations in the human holocarboxylase synthetase enzyme (HCS), a class IV BPL, and is characterized by a decrease in activity of all four human biotin-dependent carboxylases (acetyl-CoA carboxylase, pyruvate carboxylase, propionyl-CoA carboxylase and β-methylcrotonyl-CoA carboxylase) which, in turn, affects gluconeogenesis, fatty acid metabolism and amino acid catabolism.[11]

It has been demonstrated that from the 726 amino acid sequence, the minimum functional HCS protein is retained in the C-terminal 349 amino acids which show a striking sequence similarity with the biotin binding domain of bacterial BPLs.[12] Further deletion of the N-terminus of human HCS to map the 276 amino acids of *Aa*BPL resulted in 33.3 % sequence similarity and 19.2 % sequence identity between the human and *A. aeolicus* enzymes (c.f. a 21.7 % sequence identity with the catalytic and C-terminal domains of BirA). The most recurrent HCS mutations leading to MCD is R508W located in the $_{505}$GKGRGG$_{510}$ motif, which maps to Arg40 of *Aa*BPL.[13; 14] The HCS R508W mutant expressed in a cellular environment has been shown to have low activity in the cells but to remain responsive to biotin *in vitro* and the missense R508W mutation (Exon11, 1522C→T) can be treated clinically with additional doses of the vitamin.[14; 15] Adjacent mutations G518E, V550M and D571M have been shown to affect biotin binding and catalytic activity while other recurrent mutations leading to MCD have also been characterized in the C-terminal as well as the N-terminal domain of the HCS enzyme.[15] It seems likely that further studies on the bacterial proteins may uncover details of enzyme action relevant to human biochemistry.[13; 14; 16]

4.6 References

1. Choi-Rhee, E., Schulman, H. & Cronan, J. E. (2004). Promiscuous protein biotinylation by *Escherichia coli* biotin protein ligase. *Protein Sci* **13**, 3043-50.
2. Kwon, K. & Beckett, D. (2000). Function of a conserved sequence motif in biotin holoenzyme synthetases. *Protein Sci* **9**, 1530-9.
3. Clarke, D. J., Coulson, J., Baillie, R. & Campopiano, D. J. (2003). Biotinylation in the hyperthermophile *Aquifex aeolicus*. *Eur J Biochem* **270**, 1277-87.
4. Kwon, K., Streaker, E. D. & Beckett, D. (2002). Binding specificity and the ligand dissociation process in the *E. coli* biotin holoenzyme synthetase. *Protein Sci* **11**, 558-70.
5. Streaker, E. D. & Beckett, D. (2006). Nonenzymatic biotinylation of a biotin carboxyl carrier protein: unusual reactivity of the physiological target lysine. *Protein Sci* **15**, 1928-35.
6. Solbiati, J., Chapman-Smith, A. & Cronan, J. E., Jr. (2002). Stabilization of the biotinoyl domain of *Escherichia coli* acetyl-CoA carboxylase by interactions between the attached biotin and the protruding "thumb" structure. *J Biol Chem* **277**, 21604-9.
7. Pavela-Vrancic, M., Dieckmann, R., Dohren, H. V. & Kleinkauf, H. (1999). Editing of non-cognate aminoacyl adenylates by peptide synthetases. *Biochem J* **342 Pt 3**, 715-9.
8. Gillet, S., Hountondji, C., Schmitter, J. M. & Blanquet, S. (1997). Covalent methionylation of *Escherichia coli* methionyl-tRNA synthethase: identification of the labeled amino acid residues by matrix-assisted laser desorption-ionization mass spectrometry. *Protein Sci* **6**, 2426-35.
9. Hountondji, C., Lazennec, C., Beauvallet, C., Dessen, P., Pernollet, J. C., Plateau, P. & Blanquet, S. (2002). Crucial role of conserved lysine 277 in the fidelity of tRNA aminoacylation by *Escherichia coli* valyl-tRNA synthetase. *Biochemistry* **41**, 14856-65.
10. Cronan, J. E. (2005). Targeted and proximity-dependent promiscuous protein biotinylation by a mutant *Escherichia coli* biotin protein ligase. *J Nutr Biochem* **16**, 416-8.
11. Wolf, B. (1995). *In The Metabolic and Molecular Bases of Inherited Disease, 2* (Scriver, C., Beaudet, A., Sly, W. & Valle, D., Eds.). Vol 7, McGraw-Hill, New York.
12. Campeau, E. & Gravel, R. A. (2001). Expression in *Escherichia coli* of N- and C-terminally deleted human holocarboxylase synthetase. Influence of the N-terminus on biotinylation and identification of a minimum functional protein. *J Biol Chem* **276**, 12310-6.
13. Dupuis, L., Leon-Del-Rio, A., Leclerc, D., Campeau, E., Sweetman, L., Saudubray, J. M., Herman, G., Gibson, K. M. & Gravel, R. A. (1996). Clustering of mutations in the biotin-binding region of holocarboxylase synthetase in biotin-responsive multiple carboxylase deficiency. *Hum Mol Genet* **5**, 1011-6.

14. Suzuki, Y., Yang, X., Aoki, Y., Kure, S. & Matsubara, Y. (2005). Mutations in the holocarboxylase synthetase gene HLCS. *Hum Mutat* **26**, 285-90.
15. Dupuis, L., Campeau, E., Leclerc, D. & Gravel, R. A. (1999). Mechanism of biotin responsiveness in biotin-responsive multiple carboxylase deficiency. *Mol Genet Metab* **66**, 80-90.
16. Morrone, A., Malvagia, S., Donati, M. A., Funghini, S., Ciani, F., Pela, I., Boneh, A., Peters, H., Pasquini, E. & Zammarchi, E. (2002). Clinical findings and biochemical and molecular analysis of four patients with holocarboxylase synthetase deficiency. *Am J Med Genet* **111**, 10-8.

Chapter 5: Mass Spectrometry Analysis of the BCCP and BC Subunits of *E. coli* ACC

5.1 Stoichiometry of the *E. coli* Full-Length BCCP (*Ec*BCCP)

5.1.1 Purification and Characterization of *Ec*BCCP

Studies with the full-length *E. coli* biotin carboxyl carrier protein (*Ec*BCCP) have been limited by the presence of the conserved N-terminal proline and alanine-rich linker sequence between residues 34 and 101 (***See chapter II, figure 2.2***).[1-4] This N-terminal linker segment located outside the biotinyl domain is though to be highly flexible and extremely sensitive to proteolytic cleavage and most studies have been carried out with truncated C-terminal fragments of *Ec*BCCP.[2; 5-7] The full-length biotin carrier was shown to be a better substrate for the partial ACC-catalysed reactions than the truncated biotinyl domain BCCP87 and deletion of 30 amino acids of the N-terminal linker residues resulted in a *Ec*BCCP mutant which could be biotinylated but was defective in function *in vivo*.[4; 7] Therefore, the N-terminus has been suggested to be important not only for dimerisation of *Ec*BCCP but also to interact with the other subunits of the ACC complex.[8] *Ec*BCCP was previously purified and found to be a heterogeneous mixture of dimers plus higher order aggregates but the exact assembly state of *Ec*BCCP remains unknown [1; 7; 9] Since the subunits of *E. coli* ACC are required in a defined stoichiometry, the aim of this study is to shed light on the exact assembly of *E. coli* ACC. In collaboration with Prof. Carol Robinson from the University of Cambridge, analysis of the oligomeric states of *Ec*BCCP and BC and of the interactions formed by the biotin carrier and the BC subunit within the *E. coli* ACC complex have been carried out using non denaturing mass spectrometry.

In the recent study by Soriano *et al.*, the production of large amounts of recombinant forms of the *E. coli* ACC subunits (*Ec*BCCP, BC and CT) is reported.[10]

The *Ec*BCCP gene, *accB*, had been previously cloned into the pET29a vector with a C-terminal hexahistidine-tag by using the restrictions sites *NdeI* and *XhoI* and was a kind gift from Dr. Aileen Soriano at Schering-Plough. The plasmid pET29a/*Ec*BCCP was expressed in *E. coli* BL21 (DE3) cells and induction was carried out with 1 mM IPTG for 5 hours at 37 °C. Because biotin biosynthesis is strictly regulated by BirA and apo-BCCP in *E. coli*, the pET29a/*Ec*BCCP vector was transformed in presence of a second plasmid, pBirAcm Avidity (AviTag™), which overproduces BirA in order to increase the efficiency of biotinylation of overexpressed *Ec*BCCP (Fig. 5.1).

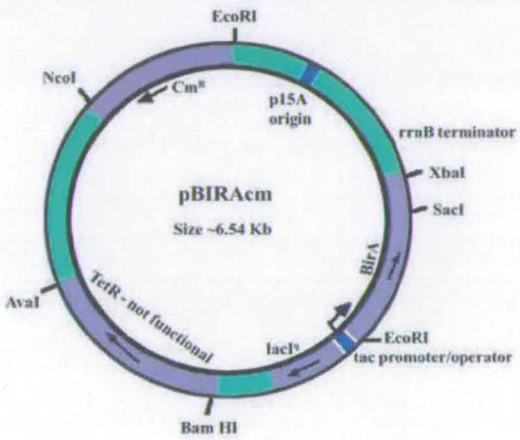


Figure 5.1 - Expression vectors pBirAcm. pBirAcm is a low copy number plasmid which is commercially available (Avidity).

The cell-free extract was ultracentrifugated for 1 hour at 30,000 rpm prior to purification by Ni-NTA agarose chromatography (Qiagen) at 4 °C. Elution with 170 mM followed by 250 mM imidazole of the soluble His_6-*Ec*BCCP protein was monitored using the Bradford reagent (Fig. 5.2, a). *Ec*BCCP was further purified by anion exchange chromatography on a 1 ml MonoQ HR 5/5 column and the protein which eluted at 420 mM NaCl was judged to be 90% pure as deduced from SDS-PAGE analysis (Fig. 5.2, b & c). Purified *Ec*BCCP was extensively dialyzed in the

presence of 20% glycerol, β-ME (20 mM) or DTT (1 mM) and 300 mM NaCl at pH 8, and the final yield of the protein was >10 mg per litre of cell culture.

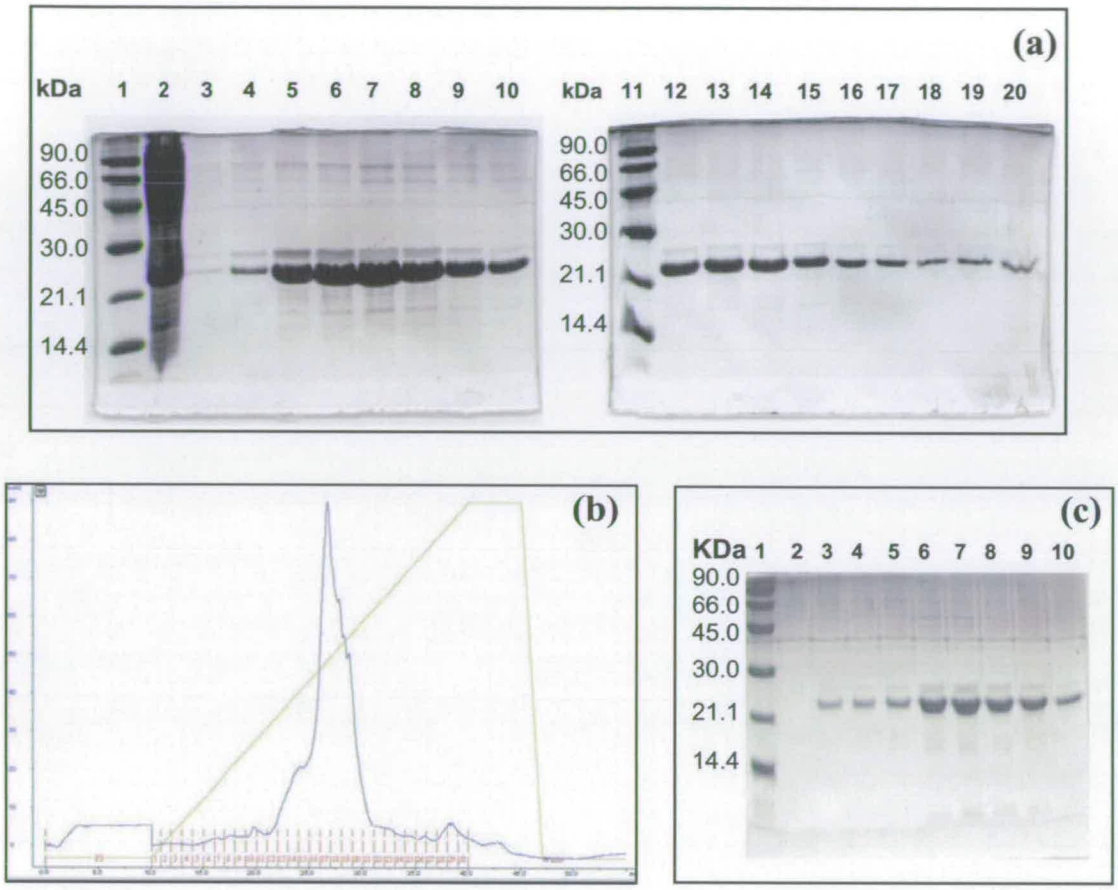


Figure 5.2 – *Ec*BCCP purification *(a)* by nickel-NTA affinity chromatography and *(b & c)* anion exchange chromatography. (a): SDS-PAGE under reducing conditions of His_6-tagged *Ec*BCCP purification with Ni-NTA agarose; *lane 1*, low molecular weight marker (Amersham), *lane 2,* flow-through, *lanes 1-8,* fractions eluted with 170 mM imidazole, *lanes 9-18*, fractions eluted with 250 mM imidazole. **(b):** chromatogram of *Ec*BCCP purification on a 1 ml MonoQ HR 5/5 column. The protein eluted at 420 mM NaCl. **(c):** Protein purification on the MonoQ HR 5/5 column was analysed by SDS-PAGE under reducing conditions. *Lane 1*, low molecular weight marker, *lanes 2,* flow-through, *lanes 3-10,* fractions 13 to 21 eluted with NaCl.

Initial mass spectrometry analysis was carried out to confirm the mass of the purified His_6-tagged *Ec*BCCP. LC-ESI-MS revealed the presence of apo- and

biotinylated forms of *Ec*BCCP with measured masses of 17,754 ± 3 Da and 17,981 ± 4 Da (Predicted masses = 17,751.40 Da and 17,977.71 Da for apo- and holo-*Ec*BCCP respectively). The molecular weight of *Ec*BCCP determined by SDS-PAGE is approximately 6 kDa greater than the mass determined by electrospray mass spectrometry. This anomalous migration has been previously observed and has been attributed to the high Pro/Ala-rich sequence of the full-length *Ec*BCCP.[7; 10] The stoichiometry of *Ec*BCCP was initially studied by gel filtration. Analysis of *Ec*BCCP on a Sephacryl S-200 HR column (320 ml) was carried out at protein concentrations of 1 μM, 5μM and 20μM under reducing conditions and in presence of glycerol. The protein eluted in the void volume indicating a molecular weight above 200 kDa (Fig. 5.3, ***see figure 2.10***). Concomitant with the mass of the holo-*Ec*BCCP monomer, this result suggests that the protein forms decamers and other higher aggregates.

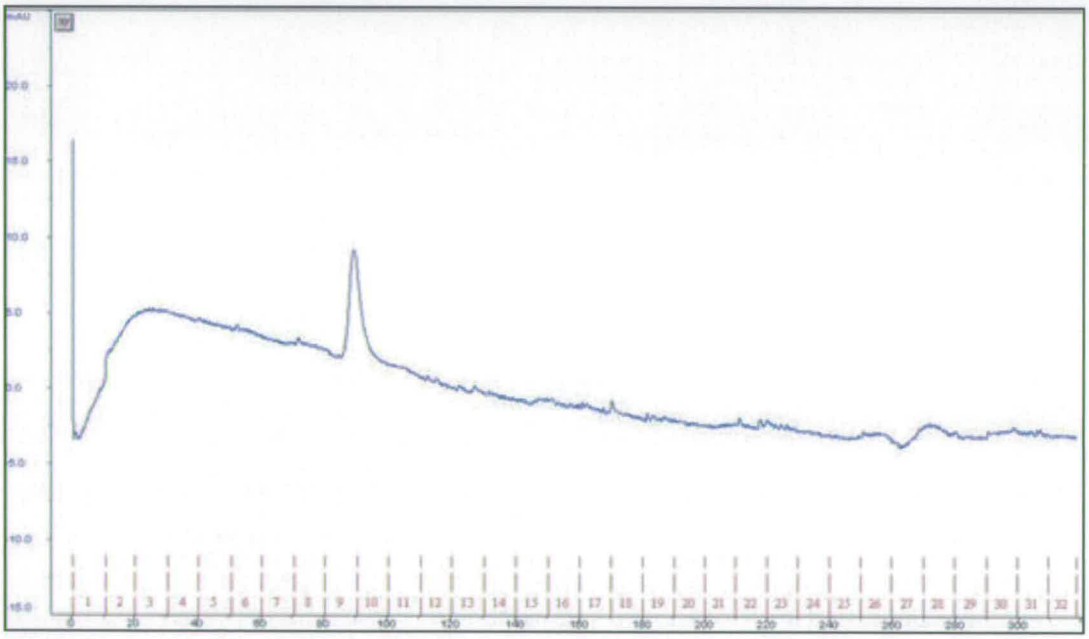


Figure 5.3 - Gel filtration analysis of *Ec*BCCP. The protein eluted from the 320 ml Sephacryl S-200 HR column at 95 ml in the void volume indicating the formation of oligomers of mass > 200 kDa. The calibration graph for the Sephacryl S-200 HR column is shown in chapter 2, figure 2.10.

The stoichiometry of *Ec*BCCP was also analysed by non-reducing protein gels. Tricine and Tris-Glycine gels ran under non-denaturing conditions with their respective buffers indicated clearly that the protein formed a dimer (Fig. 5.4). The assembly states of the full-length *Ec*BCCP and BCCP87 (un-tagged) were previously analysed by equilibrium sedimentation.[7] While the results indicated that *Ec*BCCP forms a number of large aggregates, BCCP87 remained monomeric. Further studies of BCCP87 indicated that the apo-form consisted of disulfide-linked dimer and it is tempting to assume that the non-reducing gels reveal the presence of a full-length *Ec*BCCP dimer bonded *via* the equivalent Cys116 residue.[11]

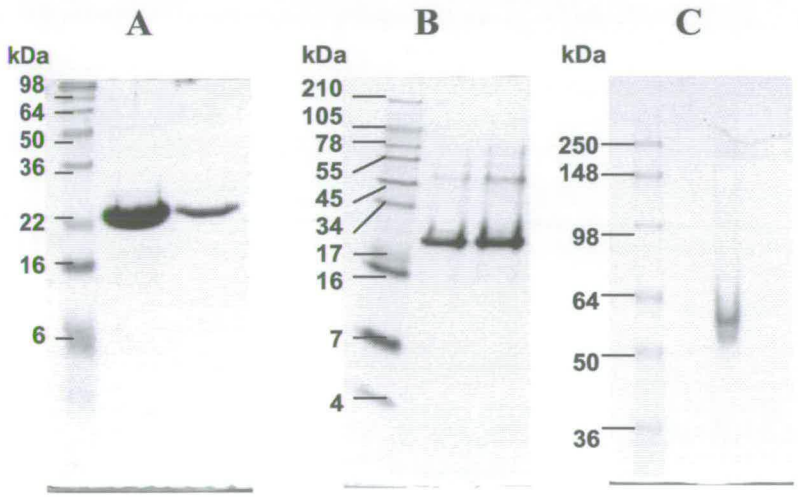


Figure 5.4 – Protein gels of *Ec*BCCP. Gels A, B, C: the pre-stained SeeBlue® Plus2 molecular weight marker is used. **A:** *Ec*BCCP ran on SDS-PAGE under reducing conditions. **B:** *Ec*BCCP separated on a Tricine gel with a Tricine non-reducing loading and running buffer. **C:** *Ec*BCCP separated on a Tris-Glycine gel with a Tris-Glycine non-reducing loading and running buffer. (Invitrogen)

To determine if the biotinylation affected the oligomeric state of the biotin carrier in solution, *Ec*BCCP was further purified by monomeric avidin chromatography. The unbound protein was washed with phosphate buffer while the biotinylated *Ec*BCCP eluted with 2 mM biotin. Interestingly, the wash fractions

indicated that the protein eluted slowly from the column and were found by LC-ESI-MS to consist of a mixture of apo- and holo- forms (Fig. 5.5, A). *Ec*BCCP was shown to elute rapidly in presence of biotin. However, despite a significant increase in the proportion of biotinylated *Ec*BCCP, LC-ESI-MS analysis revealed the presence of apo-protein which suggests that the *Ec*BCCP oligomers are formed by a mixture of apo- and holo-*Ec*BCCP monomers (Fig. 5.5, B).

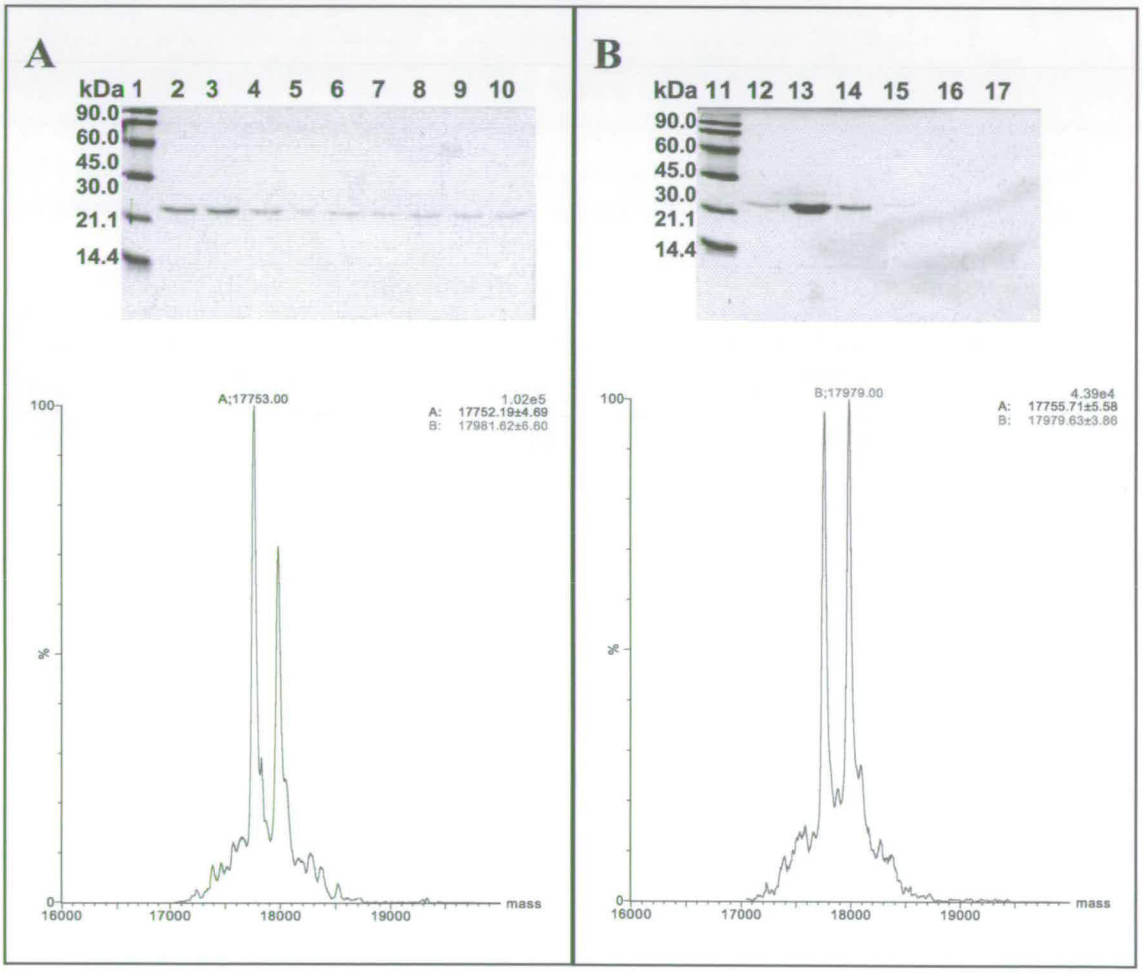


Figure 5.5 - Purification of *Ec*BCCP by monomeric avidin chromatography. A: analysis of the fractions washed with phosphate buffer. SDS-PAGE: *lane 1,* low molecular weight marker, *lanes 2-10,* fractions 1 to 9; LC-ESI-MS analysis: charge deconvoluted mass spectrum of apo- and holo-*Ec*BCCP in the wash fraction 1. **B:** analysis of the bound fractions eluted with 2 mM biotin. SDS-PAGE: *lane 11,* low molecular weight marker, *lane 12,* wash fraction 10, *lane 13-17,* fractions 1 to 5 eluted biotin; LC-ESI-MS: charge deconvoluted mass spectrum of apo- and holo-*Ec*BCCP of the bound fraction 1.

5.1.2 Analysis by Mass Spectrometry of the Assembly State of *Ec*BCCP

The structure of the N-terminal domain of *Ec*BCCP could not be determined due to its high degree of mobility and there have been no reports of successful crystallisation of an isolated full-length bacterial BCCP subunit. The interesting oligomerisation properties of *Ec*BCCP encouraged us to carry out further studies on this upstream linker segment by mass spectrometry (MS). *Ec*BCCP was analysed by Laura Lane from the laboratory of Prof. Carol Robinson (University of Cambridge, UK) on a quadrupole time-of-flight (Q-ToF) mass spectrometer using nanoelectrospray ionization (nESI) in the positive mode. nESI results in the generation of intact, multiply charged ions of the form $[M + nH]n+$. While initial analysis by LC-ESI-MS on a standard analyser generated ion series of mass over charge (*m/z*) in a range across 500 to 2000 *m/z*, the instrumentation in Cambridge enables the study of ions of up to 30,000 *m/z* under non-dissociative conditions. This would allow the identification of ions from large stable *Ec*BCCP oligomers with relatively low charge states.[12]

*Ec*BCCP was buffer exchanged into 20 mM ammonium acetate (pH 6.8) at a concentration of 1 $mg.ml^{-1}$ and 2 µl of the protein were injected into the borosilicate glass capillary and electrosprayed at a flow rate of approximately 5 $nl.min^{-1}$ driven by a potential ranging from 1.2 kV to 1.8 kV with a backing gas pressure to ensure a stable flow. Preliminary analysis by nESI-MS on a modified Q-ToF II instrument (Waters Micromass, UK) indicated a series of low intensity peaks below 2,500 *m/z*, several of which can be attributed to monomeric *Ec*BCCP proteins.[13] The mass

spectrum shows two ion envelopes (A and B) with ion peaks equivalent to the +7, +8 and +9 charge states (Fig. 5.6). These two resolvable charge state distributions (CSD) series have masses of 17,924 ± 2 Da and 18,145 ± 1 Da respectively. The mass difference between the two species (221 ± 3 Da) corresponds to the covalent attachment of one biotin (226.31 Da). The discrepancies observed between the experimental values and the predicted molecular weights of monomeric apo- and holo-*Ec*BCCP (predicted mass: 17,751.40 Da and 17,975.71 Da respectively) can be attributed to the conditions of ionization resulting in poor desolvation of *Ec*BCCP as well as the addition of salt adducts NH_4^+ (+18 Da).

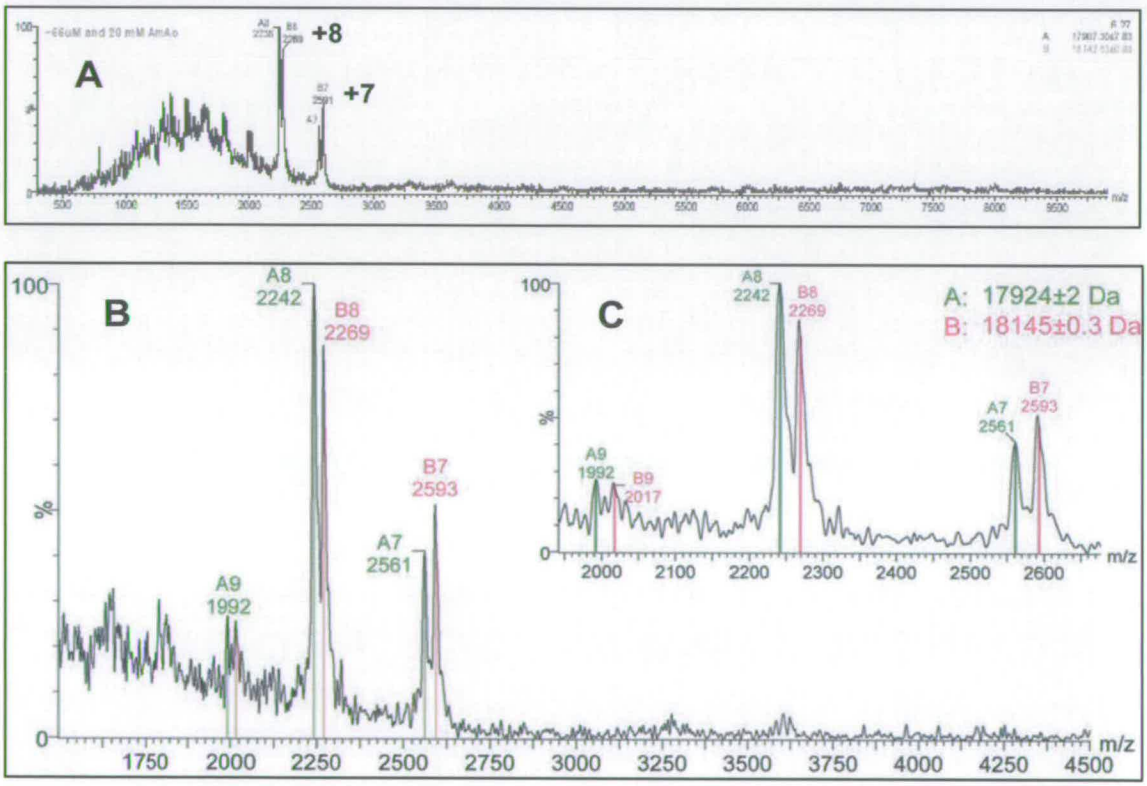


Figure 5.6 – nESI mass spectra of *Ec*BCCP analysed on a modified Q-ToF II instrument. Conditions: ISV 1.5 kV, SC 100 V, EC 0 V, CE 10. Pirani pressure, penning pressure and ToF penning pressure (mbar): 5.7e-3, 7.3e-5 and 9.1e-7. **A:** mass spectrum in the 400 to 9000 *m/z* range. **B and C:** enlargements of the two CSD of +7 to +9 charge states with measured masses of 17,924 ± 2 Da and 18,145 ± 1 which correspond to apo- and holo-*Ec*BCCP respectively.

To obtain a more accurate mass of monomeric *Ec*BCCP from a wider charge state distribution, *Ec*BCCP was denatured and was buffer exchanged into water. Analysis was carried out by nESI-MS on a Q-star XL instrument (Concord, ON, Canada) and gave two extended CSD in the 500 to 2000 *m/z* range (Fig. 5.7).[14] In this case, the two ion series (A and F) were observed for the +8 to +19 charge states and could be resolved with masses equivalent to 17,759 ± 23 Da and 17,969 ± 1 Da which correspond to the apo- and holo-form of *Ec*BCCP with a mass difference of 210 ± 24 Da. The calculated errors between the experimental values of the two apo- and holo- species and the predicted masses of *Ec*BCCP correspond to less than 0.5%. In contrast to the analysis carried out in ammonium acetate buffer, the +7 charge state of monomeric *Ec*BCCP in the *m/z* > 2000 is not apparent on this mass spectrum. The high charge states of the two CSD also suggests unfolding of the monomeric *Ec*BCCP species in absence of salt. Structural and biochemical studies of the truncated biotinyl domain of BCCP87 have shown that biotinylation increases the side-chain packing within the core of the protein and reduces its sensitivity towards proteolysis.[11; 15] It is thus worth noting that the ratio of holo:apo *Ec*BCCP decreases proportionally with the number of charges on the ion peaks. In the case of the low charge states (+8 to +15), the proportion of holo-*Ec*BCCP is greater than the extent of apo-form (Fig. 5.7).

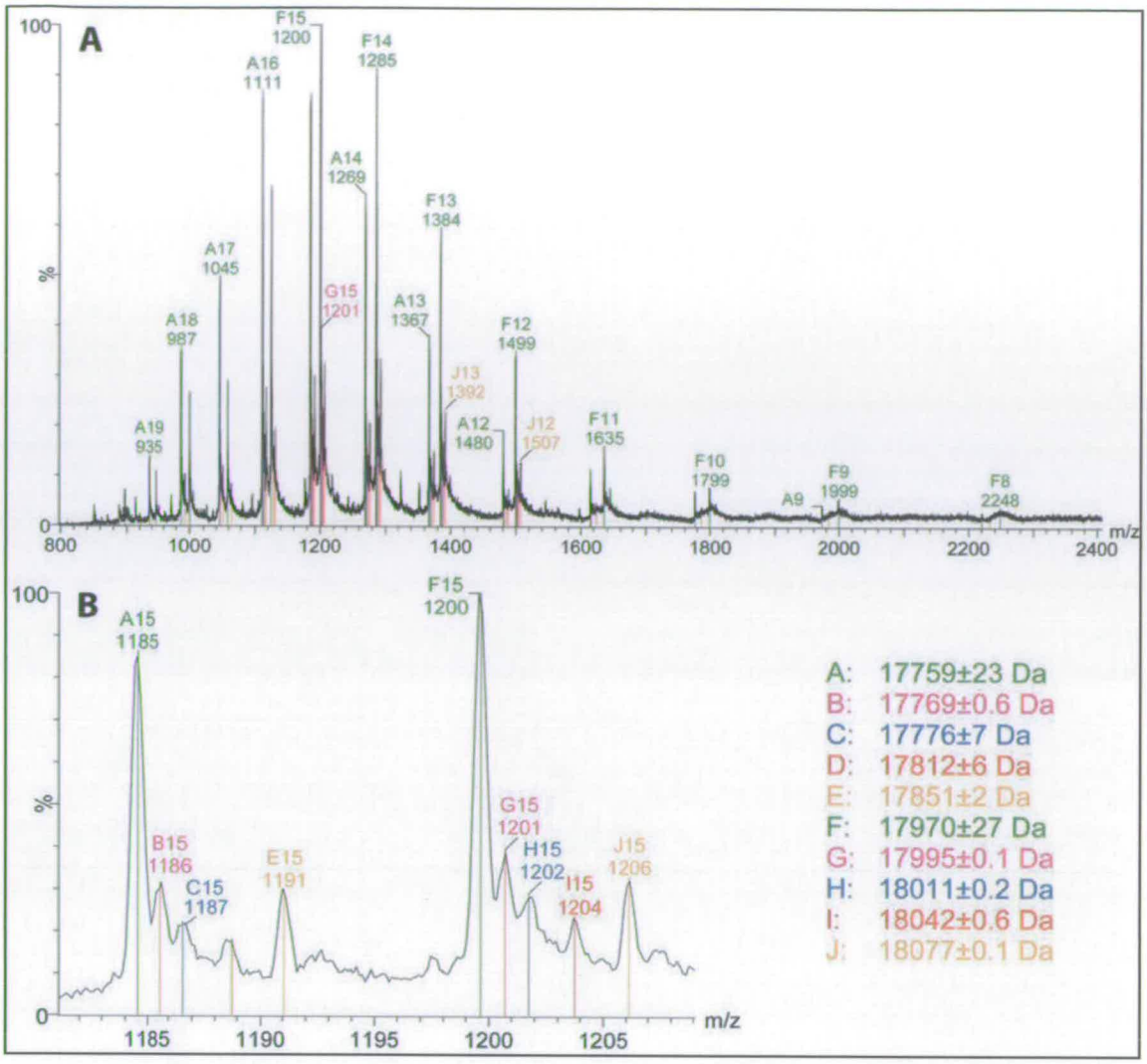


Figure 5.7 – nESI mass spectra of *Ec*BCCP analysed in water on a Q-star XL instrument. Conditions: ISV 1.5 kV, DP 100, FP, 150, DP2 15, Q0 20, CAD 3, IRD 6, IRW 5. **A:** The two 8-19+ CSD (A and F) corresponding to apo- and holo-*Ec*BCCP are well resolved. The high charge state of the ion series suggests that the protein is unfolded. **B:** enlargement of the +15 charge state of apo- and holo-*Ec*BCCP

Optimization consisted principally in decreasing the concentration of *Ec*BCCP to a range across 5 μM to 10 μM. The reduction in protein concentration reduces the likelihood that there is more than one protein species in a single droplet, hence, decreasing the appearance of ESI artefacts.[12] The mass spectrum obtained

after analysis on a Q-star XL instrument indicates that the sample is heterogeneous (Fig. 5.8, A). Two well resolved CSD in the low *m/z* (A and C, +13 to +20 charge states) with measured masses of 17,762 ± 10 Da and 17,986 ± 9 Da correspond within experimental error to the molecular weights of apo- and holo-*Ec*BCCP monomers and the mass difference between the two species is 224 ± 19 Da (Fig. 5.8, B).

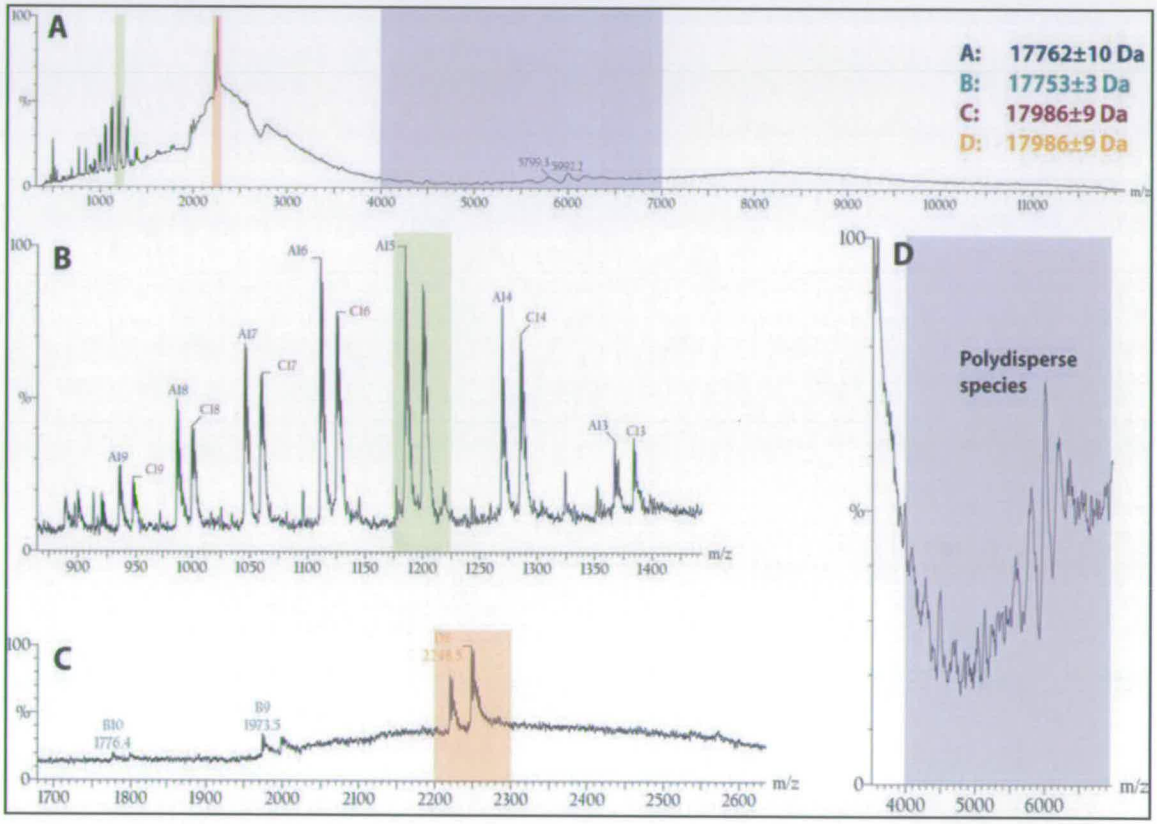


Figure 5.8 – nESI mass spectra of *Ec*BCCP analysed on a Q-Star XL instrument after optimization. Conditions: ISV 1.8 kV, DP 100, FP 200, DP2 25, Q0 20, CAD 3, IRD 6, IRW 5 **A:** The mass spectrum indicates the presence of two different conformations in the low *m/z* (coloured green and red) and the presence of polydisperse *Ec*BCCP species in the high *m/z* (coloured blue). **B:** enlargement of the mass spectrum in the 800 to 1500 *m/z* range showing the ion series A and C (charge states +13 to +19) corresponding to apo- and holo-*Ec*BCCP. **C:** enlargement of the mass spectrum in the 1700 to 2700 *m/z* range showing the ion series B and D (charge states + 8 to + 10). **D**: polydisperse *Ec*BCCP species observed at approximately 6000 *m/z*.

The two ion series (B and D) for the charge states +8 to +10 in the higher *m/z* range which were previously observed are better resolved and the measured masses of 17,753 ± 3 Da and 17,986 ± 9 Da correspond within experimental error to the predicted molecular weights of monomeric apo- and holo-*Ec*BCCP (Fig. 5.8, C; ***see figure 5.6***). Optimisation of the 1750 to 3000 *m/z* range allowed the characterization of the two ion series of +6 to +10 charge states corresponding to molecular weights of 17,754 ± 1 Da and 17,982 ± 3 Da (Fig. 5.9). The mass difference between the two species is equivalent to 228 ± 4 Da and the error between the experimental values and the predicted masses of monomeric apo- and holo-*Ec*BCCP is less than 0.5%. The presence of two different CSD pairs (+6 - +10 and +13 - +20 charge states) equivalent to monomeric *Ec*BCCP indicate a different conformation for the protein which is more compact for the low charge states ions.

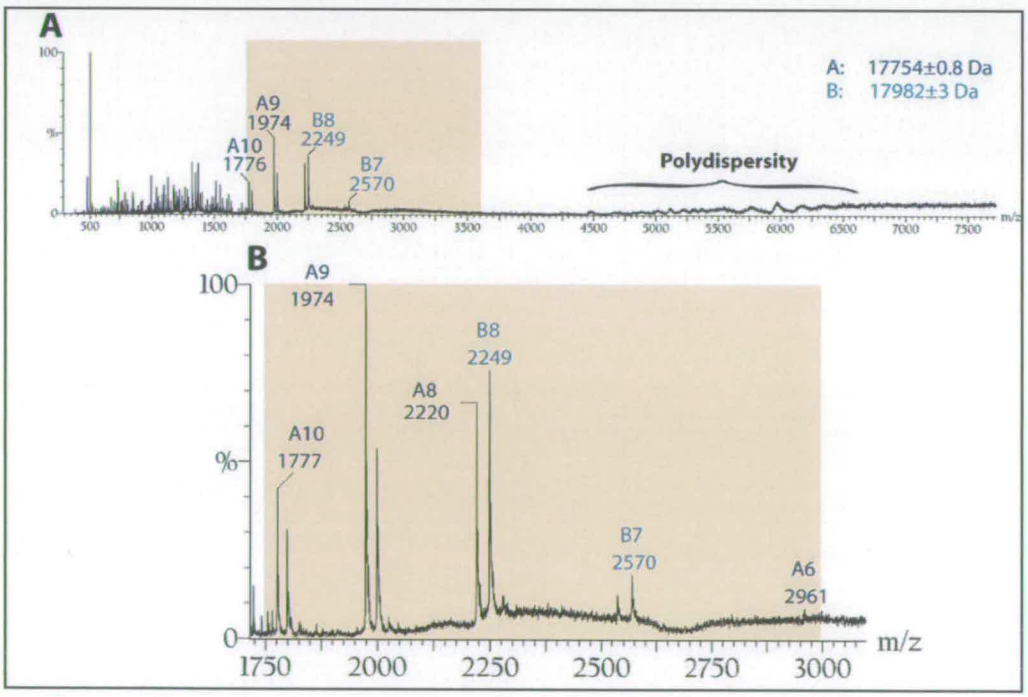


Figure 5.9 – nESI-MS analysis of *Ec*BCCP in the 1,750 to 3,000 *m/z* range. Conditions: ISV 1.2 kV, DP 180, FP 15, DP2 15, Q0 20, CAD 3, IRD 6, IRW 5.The different charge series suggest that *Ec*BCCP adopts a more compact conformation in the high *m/z*.

The polydisperse peaks observed at approximately 6,000 *m/z* appears occasionally to be resolved (Fig. 5.8, D). MS-MS of one of these peaks indicated an average oligomeric state of ten. Although other oligomeric species appear to be present in this *m/z* region, their intensities and stoichiometries are often not apparent in the mass spectrum and after smoothing (Fig. 5.10). Further MS-MS analysis of these polydisperse species may reveal useful stoichiometric information.

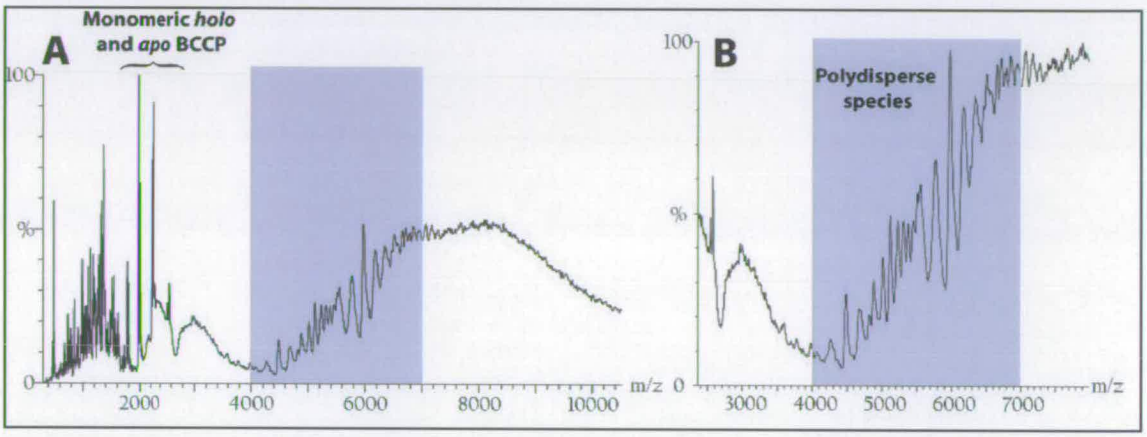


Figure 5.10 – nESI-MS analysis of *Ec*BCCP in the the 4,000 to 7,000 *m/z* range. Conditions: ISV 1.2 kV, DP 180, FP 15, DP2 15, Q0 20, CAD 3, IRD 6, IRW 5. **A and B:** Polydisperse BCCP peaks can be observed

5.2 Assembly state of *E. coli* Biotin Carboxylase (BC)

5.2.1 General Description of *E. coli* BC

The *E. coli* biotin carboxylase gene (*accC*) encodes a protein of 449 residues which has a molecular weight of 49,387.20 Da.[16; 17] The *accC* gene was previously cloned with an N-terminal histidine-tag into the ampicillin-resistant pET16b vector by using the restrictions sites *NdeI* and *XhoI* and was a kind gift from Dr. Aileen

Soriano from Schering-Plough. The resulting sequence of *E. coli* His_{10}-BC shown in figure 5.11 consists of 471 amino acids and has a formula weight of 51,908.94 Da.

```
  1    GHHHHHHHHH HSSGHIEGRH MLDKIVIANR GEIALRILRA CKELGIKTVA
 51    VHSSADRDLK HVLLADETVC IGPAPSVKSY LNIPAIISAA EITGAVAIHP
101    GYGFLSENAN FAEQVERSGF IFIGPKAETI RLMGDKVSAI AAMKKAGVPC
151    VPGSDGPLGD DMDKNRAIAK RIGYPVIIKA SGGGGGRGMR VVRGDAELAQ
201    SISMTRAEAK AAFSNDMVYM EKYLENPRHV EIQVLADGQG NAIYLAERDC
251    SMQRRHQKVV EEAPAPGITP ELRRYIGERC AKACVDIGYR GAGTFEFLFE
301    NGEFYFIEMN TRIQVEHPVT EMITGVDLIK EQLRIAAGQP LSIKQEEVHV
351    RGHAVECRIN AEDPNTFLPS PGKITRFHAP GGFGVRWESH IYAGYTVPPY
401    YDSMIGKLIC YGENRDVAIA RMKNALQELI IDGIKTNVDL QIRIMNDENF
451    QHGGTNIHYL EKKLGLQEKS D
```

Figure 5.11 - Amino acid sequence of His_{10}-BC. The *accC* gene was previously cloned in the pET16b vector using the restrictions sites *NdeI* and *XhoI*. BC contains an N-terminal decahistidine tag and consists of 471 amino acids.

5.2.2 Purification and Characterization of BC

The plasmid pET16b/BC was used to transform *E. coli* BL21 (DE3) cells and BC was induced with 1 mM IPTG for 5 hours at 37 °C. Purification of BC was carried out by nickel affinity chromatography on a 5 ml HisTrap™ HP column. The protein eluted with 250 mM imidazole and was extensively dialyzed in presence of 0.1 mM EDTA (Fig 5.12). The isolated BC was further purified by gel filtration and eluted from the Sephacryl S-200 HR column (320 ml) at a volume of 120 ml which corresponds approximately to a mass of 101 kDa (Fig. 5.13, ***see figure 2.10***; predicted molecular weight of BC dimer = 103.8 kDa). A portion of the protein eluted in the void volume but was not kept for further MS analysis. After the gel filtration purification step, BC was judged to have a purity greater than 95% as determined by SDS-PAGE and the yield of the protein was > 10 mg per litre of cell

culture (Fig 5.13, b). Initial mass spectrometry measurements carried out by LC-ESI-MS indicated a single species with a measured mass of 51,906.04 ± 27.75 which corresponds within experimental error to the predicted molecular weight of monomeric BC.

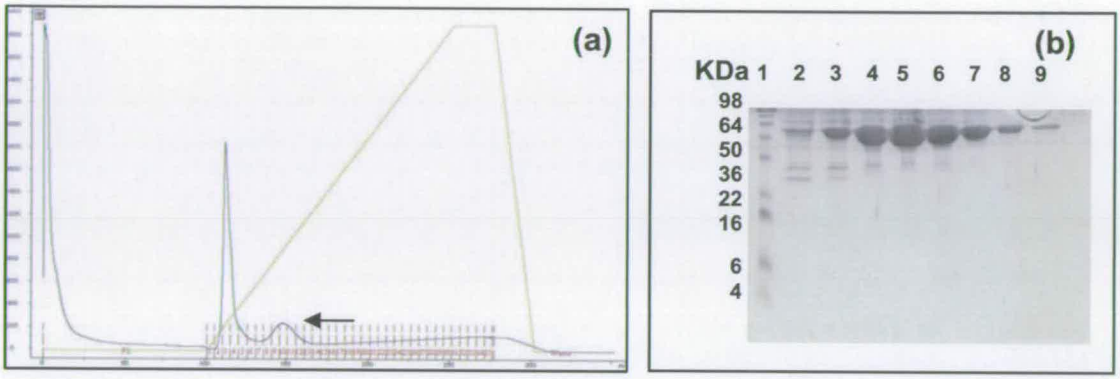


Figure 5.12 - Purification and SDS-PAGE analysis of *E. coli* BC (a): Chromatogram of BC purification on a HisTrap™ HP nickel column. The peak corresponding to BC is indicated by an arrow **(b):** Protein purification was analysed by SDS-PAGE under reducing conditions. *Lane 1*, SeeBlue Plus2 molecular weight marker, *lanes 2-9,* eluted fractions 6 to 13 (collected fractions 8-13).

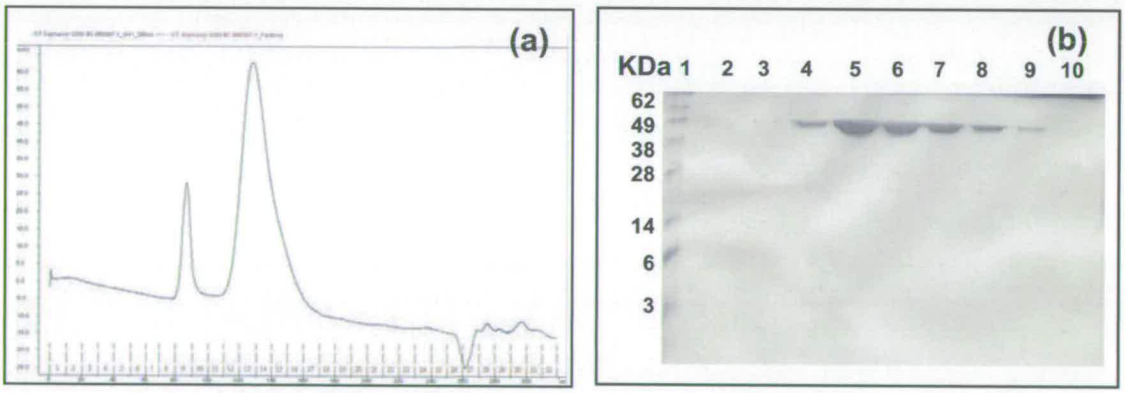


Figure 5.13 - Gel filtration analysis of BC. (a): purification of BC on a Sephacryl S-200 HR column. The enzyme eluted at 120 ml indicating that BC is dimeric. A portion eluted at 85 ml in the void volume. The calibration graph for the column is shown in chapter 2, figure 2.10. **(b):** protein purification was analysed by SDS-PAGE under reducing conditions. *Lane 1*, SeeBlue Plus2 molecular weight marker, *lanes 2-10,* eluted fractions 11 to 19 (collected fractions 13-17).

5.2.3 Mass Spectrometry Analysis of BC

BC was nanoelectrosprayed out of 20 mM ammonium acetate buffer pH 6.8 at a concentration of 2 $mg.ml^{-1}$ and similarly to *Ec*BCCP, the ions generated in positive mode were analysed on a modified Q-ToF II spectrometer (Waters Micromass, UK). Two different CSD observed in the mass spectrum were resolved indicating the presence of monomeric and dimeric BC (Fig. 5.14). The first charge state distribution (A) corresponds to a mass of 52,251 ± 44 which is close to the predicted mass of monomeric BC (51,908.94 Da). The second CSD (B) corresponds to dimeric BC with a measured mass of 104,587 ± 15 Da (predicted mass of BC dimer = 103,817.88 Da). Poor desolvation and addition of salt adducts may contribute to the discrepancy observed between the experimental value and the predicted molecular weight of monomeric and dimeric BC. This phenomenon is commonly observed when studying protein ions because increasing the activation of the ions leads to disruption of noncovalent interactions before full desolvation is reached.[18]

The results of this incomplete desolvation are broad peaks and mass measurements that are several percent larger than those predicted by the amino acid sequence. Therefore, despite the resolving power capabilities of the Q-ToF spectrometer, the peak width achieved for a protein is determined primarily by the sample homogeneity but also by the electrospray process and subsequent ion desolvation. However, it has been shown that the widths of the mass spectrum peaks are directly correlated with the excess mass observed due to adducts for large ions of known molecular weight.[19] On the spectrum A, the peak widths for the +14 charge states of the ion series A corresponds to approximately 700 Da, while the peak

widths for the +21 charge states of the ion series B corresponds to approximately 1000 Da (Fig. 5.14). Those values are higher than the difference measured between the experimental and the predicted masses of the BC monomer and dimer.

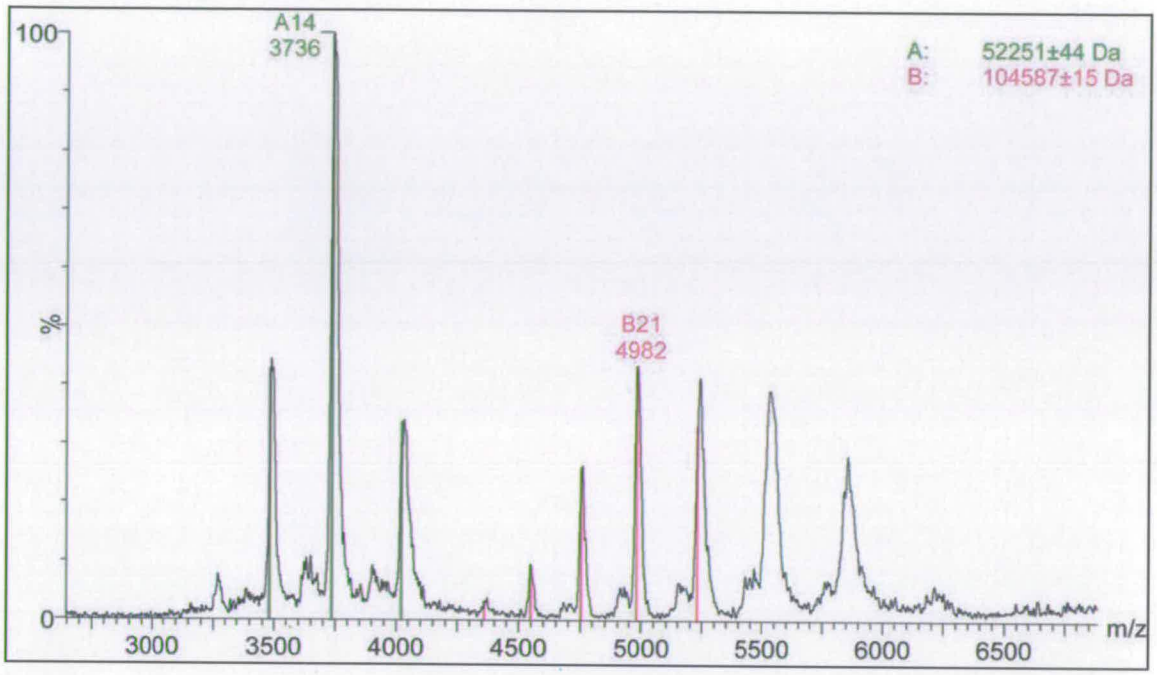


Figure 5.14 – nESI mass spectrum of BC analysed on a modified Q-ToF II instrument. Cap 1.7 kV, SC 150, EC 0 V, CE 10.Pirani pressure, penning pressure and ToF penning pressure (mbar): 7.70.e-4, 2.88.e-4 and 1.42.e-6. Two charge series A and B corresponding to monomeric and dimeric BC respectively are observed.

5.3 Discussion

Studies of the assembly states of the *Ec*BCCP and BC subunits of *E. coli* ACC were carried out on a modified Q-ToF II and a Q-star XL mass spectrometer in the laboratory of Prof. Carol Robinson at the University of Cambridge.[13; 14] Monomeric *Ec*BCCP was initially analysed in water by nESI-MS on a Q-star XL

instrument in the positive mode. The derived mass spectra revealed the presence of charge series which correspond to monomeric forms of apo- and holo-*Ec*BCCP. The +9 to +19 charge state series is well resolved and the sample appears relatively homogenous. However, the high charge states of the ion peaks in the two series indicate that the protein is unfolded. In solution phase, the energy that assists a protein in folding to a specific conformer can be attributed to the driving force to form the hydrophobic core while polar residues are oriented towards the solution and form a hydrophilic surface.[20] This model is consistent with the structure of the BCCP87 domain which forms a flattened β-barrel stabilized by a core of hydrophobic residues and, therefore, the monomeric form of *Ec*BCCP could be stable in solution.

The full-length *Ec*BCCP was similarly analysed in an ammonium acetate buffer (pH 6.8) by nESI-MS on a modified Q-ToF II instrument.[13] The mass spectrum revealed charge state distribution series of +7 to +9 charge states corresponding to apo- and holo-*Ec*BCCP monomers. The charge differences of the ion peaks in the series obtained after nESI-MS analysis in water and in ammonium acetate buffer suggest that while the water condition appeared to be denaturing, the presence of salt seems to be required for the protein to maintain a compact structure. This observation is well described by the charged-residue model (CRM) of the highly charged microdroplets in the ionization source.[21] In the CRM model, complete desolvation of ions occur via successive loss of all solvent molecules. The protons of the ultimate droplet are then transferred onto the protein molecules and thus, the number of charges is directly related to the space that a molecule covers in the

microdroplets. Compact proteins exhibit lower charge state under identical nESI conditions than their unfolded conformations.[21]

Further analysis of *Ec*BCCP in water but at lower protein concentrations (approximately 5 μM to 10 μM instead of 55 μM) confirmed the presence of a different, more compact conformation of monomeric *Ec*BCCP in the high *m/z* region (+6 to +10 charge states). Interestingly, the extent of biotinylated *Ec*BCCP is also higher in this *m/z* range. In the published X-ray and nmr structures of the biotinyl domains of BCCP, flexible regions (either not well resolved or disordered), are the N-terminus and the protruding thumb segment.[22-24] In the nmr structure of holo-BCCP87, the core of the protein is shown to have a more favourable side-chain packing than in the apo-structure and the interactions with the biotin moiety stabilizes the residues located on the flexible thumb region.[15; 24] To investigate the role of the N-terminus in the ionization process, BCCP87 was cloned with the C-terminal His-tag for future studies (Fig. 5.15). Although the N-terminal domain consists principally of aliphatic residues, characterization of both the full-length BCCP and BCCP87 by nESI-MS under similar conditions may reveal the significance of the N-terminus interactions with the core of the protein.

Polydisperse *Ec*BCCP species in the range from 4,000 to 10,000 *m/z* were also observed on the mass spectrum after optimization and suggested the presence of at least *Ec*BCCP decamers. The sample is highly heterogeneous in this *m/z* region and attempts to analyse these species by MS-MS were carried out. In tandem MS, single species of interest can be selected and activated according to their *m/z* ratios in the first analyser (quadrupole) and analysis to determine their stoichiometry is performed in the ToF analyser at the end.[25] However, the ion peaks corresponding to

*Ec*BCCP decamers are not always evident and MS-MS analysis was difficult to perform. Previous studies of *Ec*BCCP have demonstrated the formation of lower and higher assembly states than decamers.[7] Furthermore, purification of *Ec*BCCP carried out by monomeric avidin chromatography revealed that the *Ec*BCCP oligomers consist of a mixture of apo- and holo-forms which could not be separated using affinity chromatography. The formation of *Ec*BCCP oligomers appears to occur in a more or less random process and therefore, the proportion of biotinylated monomers can vary significantly in the *Ec*BCCP complexes which in turn, contributes to the heterogeneity of the ions generated from decamers and other oligomers in this *m/z* region.

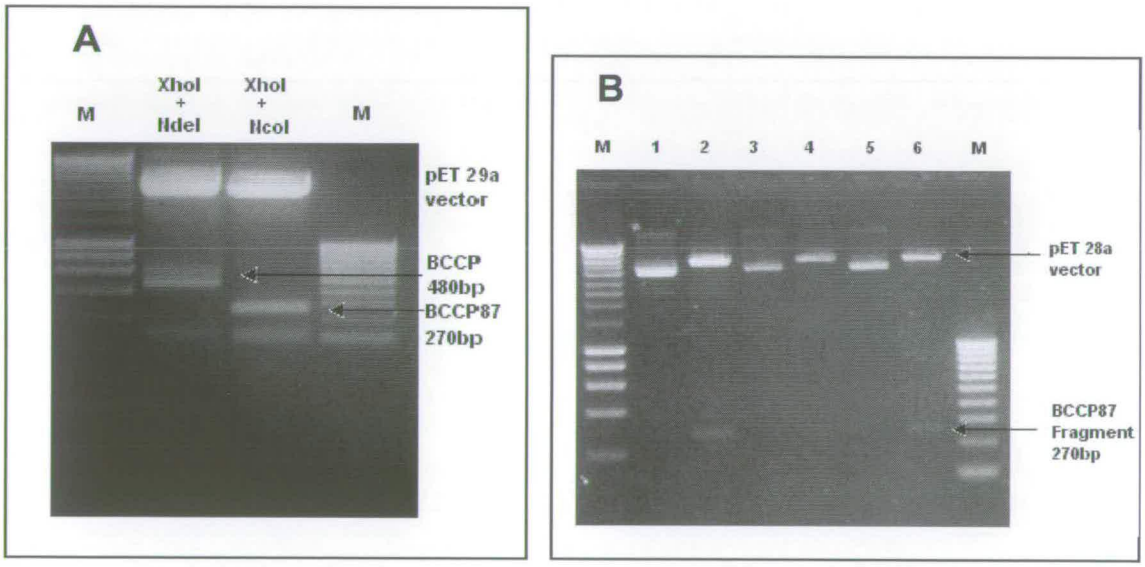


Figure 5.15 - DNA gels of *Ec*BCCP and BCCP87. A: PCR amplification of the *accB* gene followed by digestion with *NdeI* / *XhoI* and *NcoI* / *XhoI* yielded fragments of the size corresponding to *Ec*BCCP and BCCP87 respectively **B:** restriction digest of the pET28a/BCCP87 vector obtained after DNA sequencing of BCCP87 and ligation into a pGEM vector (Promega) followed by cloning into a pET28a vector (Novagen) using the restriction sites *NcoI* and *XhoI*. *Lanes 1, 3 and 5*, undigested BCCP87, *lanes 2, 4 and 6*, BCCP87 digested with *NcoI* / *XhoI*

In *E. coli*, the *Ec*BCCP and BC genes (*accB* and *accC* respectively) form an operon with the *accB* gene located upstream of *accC* and this arrangement is strikingly conserved among bacterial genomes.[17; 26; 27] Although transcription of the biotin biosynthetic operon (*bio*ABFCD) in *E. coli* is regulated by BirA as well as the presence of apo-*Ec*BCCP protein, overproduction of *accC* was recently shown to disrupt regulation of biotin and fatty acid biosynthesis.[28; 29] It is thought that the *accB* and *accC* genes are co-transcribed to ensure stringent regulation of the ratio of the two proteins and the subsequent level of biotin biosynthesis.[28] Thus, to investigate the assembly state of *Ec*BCCP by MS, analysis of BC and of the BC:*Ec*BCCP sub-complex appears necessary.

BC has been shown to be a dimer in solution and this dimeric state assembly is conserved in the crystal structure.[30; 31] The recombinant form of *E. coli* BC was purified to homogeneity and subsequent analysis by nESI-MS on a Q-ToF II instrument confirmed the dimeric nature of BC. The *E. coli* BC and *Ec*BCCP form an unstable complex with a stoichiometry that has been the subject of long-term debate. Preliminary studies in *E. coli* together with the isolation and characterization of the $BCCP_2$:BC_2:$CT\alpha_2$:$CT\beta_2$ multi-subunit ACC from *Pseudomonas citronellolis* suggested that a *Ec*$BCCP_2$:BC_2 complex might be present *in vivo*.[2; 9; 32] A new model of the BC:*Ec*BCCP complex was recently reported by Choi-Rhee and Cronan. Their analysis by SDS-PAGE of a [^{35}S]methionine-labelled BC:*Ec*BCCP complex indicated the ratio of BC per *Ec*BCCP molecule to be 1:2.[33] In the proposed ACC model, the BC_2:$BCCP_4$ complex is thought to bind the $\alpha_2\beta_2$ CT heterodimer in order to form the active ACC complex (BC_2:BCCP4:$CT\alpha_2\beta_2$) which would have a combined mass of ~304 kDa. With the aim to investigate in the future the ratio of BC

and *Ec*BCCP molecules within the BC:*Ec*BCCP sub-complex by nESI-MS analysis, the *accBC* operon was cloned from genomic *E. coli* DNA in a pGEM vector (Promega) to generate a similar clone to that used by Choi-Rhee and Cronan (data not shown).

The structure of the complete biotin-dependant enzyme pyruvate carboxylase from *Rhizobium etli* (*Re*PC) was recently determined at 2.0 Å resolution and revealed the structure of a full-length BCCP for the first time (Fig. 5.16, a).[34] *Re*PC consists of an N-terminal BC, a central CT, and a C-terminal BCCP domain. There are two monomers in the asymmetric unit and BCCP can be visualised only in one monomer. The structure of *Re*PC shows that each monomer contains a central domain which has not been previously described and appears to serve as a mediator by connecting the BC to the CT domain and the BCCP to CT domain.[34] PC is a tetramer in solution and the *Re*PC crystal lattice also contains a tetramer composed of a dimer of dimers (Fig 5.16, b).[34; 35] The BC active site is located at ~65 Å from the CT active site from an opposing polypeptide chain, with the BCCP domain positioned equidistant between the two. Therefore, based on distances and orientation of the different active sites of *Re*PC, the BCCP domain is predicted to swing between the BC and CT components from different polypeptide chains.[34] The structure also reveals that the global position of the BCCP is restrained by the arrangement of the remaining *Re*PC enzyme. The 34 Å proline/alanine-rich linker connecting BCCP to the allosteric domain is highly flexible and facilitates the moves of biotinylated BCCP between the active sites.[3; 36; 37]

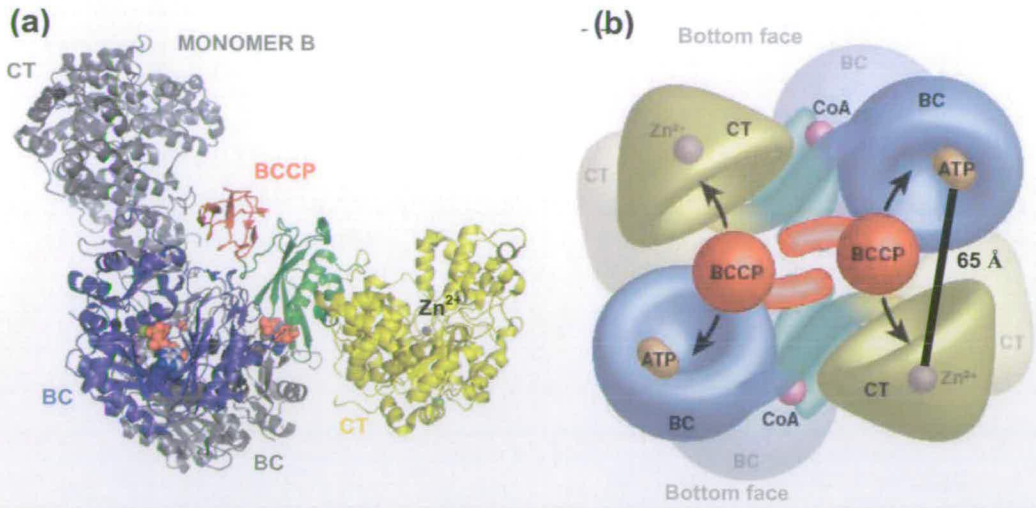


Figure 5.16 – Crystal structure of the *R. etli* PC complex. (a) The structure of the *Re*PC dimer. The BC, CT, BCCP, and allosteric domains are colored blue, yellow, red, and green, respectively. The active sites are highlighted by the presence of ligand drawn in spheres: CoA in the linker domain, the ATP analog adenosine-5'-diphosphate monothiophosphate SAP complexed with Mg^{2+} in the BC active site, and Zn^{2+} in the CT active site. (PBD code 2QF7) **(b)** Model of the *Re*PC tetramer showing the movement of the BCCP domain between neighbouring active sites on opposing polypeptide chains.[34]

More recently, Tong *et al.* also reported the crystal structures at 2.8 Å resolution of the full-length PC from *S. aureus* (*Sa*PC) and the C-terminal region missing the BC domain of human PC.[38] The two structures show a PC tetramer and reveal a previously uncharacterized domain suggested to mediate teramerization. In *Sa*PC, the biotinylated BCCP domain of monomer 4 is located in the active site of the CT domain of monomer 3 and a similar conformation for the BCCP domain is observed in the structure of human PC. In both structures, the biotin moiety is projected away from the BCCP and is located near the pyruvate substrate bound in the CT active site. Therefore, this conformation may represent that of BCCP during catalysis in the carboxyltransfer reaction and is the first observation of BCCP in a catalytically competent conformation.[38] Those PC structures provide the first glimpse

of the interactions made by BCCP with BC and CT. They provide an important insight into the functionality of biotin-dependent enzymes but also reveal the complexity of the assembly states of such protein complexes. A clone that expresses the complete CT enzyme of *E. coli* ACC was generously supplied by Aileen Soriano and attempts to characterize the full ACC complex by MS analysis will continue in the future.

5.4 References

1. Fall, R. R., Nervi, A. M., Alberts, A. W. & Vagelos, P. R. (1971). Acetyl CoA carboxylase: isolation and characterization of native biotin carboxyl carrier protein. *Proc Natl Acad Sci U S A* **68**, 1512-5.
2. Fall, R. R. & Vagelos, P. R. (1975). Biotin carboxyl carrier protein from *Escherichia coli*. *Methods Enzymol* **35**, 17-25.
3. Perham, R. N. (2000). Swinging arms and swinging domains in multifunctional enzymes: catalytic machines for multistep reactions. *Annu Rev Biochem* **69**, 961-1004.
4. Cronan, J. E., Jr. (2002). Interchangeable enzyme modules. Functional replacement of the essential linker of the biotinylated subunit of acetyl-CoA carboxylase with a linker from the lipoylated subunit of pyruvate dehydrogenase. *J Biol Chem* **277**, 22520-7.
5. Sutton, M. R., Fall, R. R., Nervi, A. M., Alberts, A. W., Vagelos, P. R. & Bradshaw, R. A. (1977). Amino acid sequence of *Escherichia coli* biotin carboxyl carrier protein (9100). *J Biol Chem* **252**, 3934-40.
6. Chapman-Smith, A., Turner, D. L., Cronan, J. E., Jr., Morris, T. W. & Wallace, J. C. (1994). Expression, biotinylation and purification of a biotin-domain peptide from the biotin carboxy carrier protein of *Escherichia coli* acetyl-CoA carboxylase. *Biochem J* **302 (Pt 3)**, 881-7.
7. Nenortas, E. & Beckett, D. (1996). Purification and characterization of intact and truncated forms of the *Escherichia coli* biotin carboxyl carrier subunit of acetyl-CoA carboxylase. *J Biol Chem* **271**, 7559-67.
8. Cronan, J. E., Jr. & Waldrop, G. L. (2002). Multi-subunit acetyl-CoA carboxylases. *Prog Lipid Res* **41**, 407-35.
9. Fall, R. R. & Vagelos, P. R. (1972). Acetyl coenzyme A carboxylase. Molecular forms and subunit composition of biotin carboxyl carrier protein. *J Biol Chem* **247**, 8005-15.

10. Soriano, A., Radice, A. D., Herbitter, A. H., Langsdorf, E. F., Stafford, J. M., Chan, S., Wang, S., Liu, Y. H. & Black, T. A. (2006). *Escherichia coli* acetyl-coenzyme A carboxylase: characterization and development of a high-throughput assay. *Anal Biochem* **349**, 268-76.
11. Chapman-Smith, A., Forbes, B. E., Wallace, J. C. & Cronan, J. E., Jr. (1997). Covalent modification of an exposed surface turn alters the global conformation of the biotin carrier domain of *Escherichia coli* acetyl-CoA carboxylase. *J Biol Chem* **272**, 26017-22.
12. Benesch, J. L., Ruotolo, B. T., Simmons, D. A. & Robinson, C. V. (2007). Protein complexes in the gas phase: technology for structural genomics and proteomics. *Chem Rev* **107**, 3544-67.
13. Sobott, F., Hernandez, H., McCammon, M. G., Tito, M. A. & Robinson, C. V. (2002). A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies. *Anal Chem* **74**, 1402-7.
14. Chernushevich, I. V. & Thomson, B. A. (2004). Collisional cooling of large ions in electrospray mass spectrometry. *Anal Chem* **76**, 1754-60.
15. Yao, X., Soden, C., Jr., Summers, M. F. & Beckett, D. (1999). Comparison of the backbone dynamics of the apo- and holo-carboxy-terminal domain of the biotin carboxyl carrier subunit of *Escherichia coli* acetyl-CoA carboxylase. *Protein Sci* **8**, 307-17.
16. Kondo, H., Shiratsuchi, K., Yoshimoto, T., Masuda, T., Kitazono, A., Tsuru, D., Anai, M., Sekiguchi, M. & Tanabe, T. (1991). Acetyl-CoA carboxylase from *Escherichia coli*: gene organization and nucleotide sequence of the biotin carboxylase subunit. *Proc Natl Acad Sci U S A* **88**, 9730-3.
17. Li, S. J. & Cronan, J. E., Jr. (1992). The gene encoding the biotin carboxylase subunit of *Escherichia coli* acetyl-CoA carboxylase. *J Biol Chem* **267**, 855-63.
18. Sobott, F. & Robinson, C. V. (2004). Characterising electrosprayed biomolecules using tandem-MS - the noncovalent GroEL chaperonin assembly. *Int J Mass Spectrom* **236**, 25-32.
19. McKay, A. R., Ruotolo, B. T., Ilag, L. L. & Robinson, C. V. (2006). Mass measurements of increased accuracy resolve heterogeneous populations of intact ribosomes. *J Am Chem Soc* **128**, 11433-42.
20. Baldwin, R. L. (2007). Energetics of protein folding. *J Mol Biol* **371**, 283-301.
21. Felitsyn, N., Peschke, M. & Kebarle, P. (2002). Origin and number of charges observed on multiply-protonated native proteins produced by ESI. *Int J Mass Spectrom* **219**, 39-62.
22. Athappilly, F. K. & Hendrickson, W. A. (1995). Structure of the biotinyl domain of acetyl-coenzyme A carboxylase determined by MAD phasing. *Structure* **3**, 1407-19.
23. Roberts, E. L., Shu, N., Howard, M. J., Broadhurst, R. W., Chapman-Smith, A., Wallace, J. C., Morris, T., Cronan, J. E., Jr. & Perham, R. N. (1999). Solution structures of apo and holo biotinyl domains from acetyl coenzyme A carboxylase of *Escherichia coli* determined by triple-resonance nuclear magnetic resonance spectroscopy. *Biochemistry* **38**, 5045-53.
24. Yao, X., Wei, D., Soden, C., Jr., Summers, M. F. & Beckett, D. (1997). Structure of the carboxy-terminal fragment of the apo-biotin carboxyl carrier

subunit of *Escherichia coli* acetyl-CoA carboxylase. *Biochemistry* **36**, 15089-100.

25. Aquilina, J. A., Benesch, J. L., Bateman, O. A., Slingsby, C. & Robinson, C. V. (2003). Polydispersity of a mammalian chaperone: mass spectrometry reveals the population of oligomers in alphaB-crystallin. *Proc Natl Acad Sci USA* **100**, 10611-6.
26. Li, S. J. & Cronan, J. E., Jr. (1993). Growth rate regulation of *Escherichia coli* acetyl coenzyme A carboxylase, which catalyzes the first committed step of lipid biosynthesis. *J Bacteriol* **175**, 332-40.
27. James, E. S. & Cronan, J. E. (2004). Expression of two *Escherichia coli* acetyl-CoA carboxylase subunits is autoregulated. *J Biol Chem* **279**, 2520-7.
28. Abdel-Hamid, A. M. & Cronan, J. E. (2007). Coordinate expression of the acetyl coenzyme A carboxylase genes, *accB* and *accC*, is necessary for normal regulation of biotin synthesis in *Escherichia coli*. *J Bacteriol* **189**, 369-76.
29. Streaker, E. D. & Beckett, D. (2006). The biotin regulatory system: kinetic control of a transcriptional switch. *Biochemistry* **45**, 6417-25.
30. Guchhait, R. B., Polakis, S. E. & Lane, M. D. (1975). Biotin carboxylase component of acetyl-CoA carboxylase from *Escherichia coli*. *Methods Enzymol* **35**, 25-31.
31. Thoden, J. B., Blanchard, C. Z., Holden, H. M. & Waldrop, G. L. (2000). Movement of the biotin carboxylase B-domain as a result of ATP binding. *J Biol Chem* **275**, 16183-90.
32. Fall, R. R. (1976). Stabilization of an acetyl-coenzyme A carboxylase complex from *Pseudomonas citronellolis*. *Biochim Biophys Acta* **450**, 475-80.
33. Choi-Rhee, E. & Cronan, J. E. (2003). The biotin carboxylase-biotin carboxyl carrier protein complex of *Escherichia coli* acetyl-CoA carboxylase. *J Biol Chem* **278**, 30806-12.
34. St Maurice, M., Reinhardt, L., Surinya, K. H., Attwood, P. V., Wallace, J. C., Cleland, W. W. & Rayment, I. (2007). Domain architecture of pyruvate carboxylase, a biotin-dependent multifunctional enzyme. *Science* **317**, 1076-9.
35. Attwood, P. V., Johannssen, W., Chapman-Smith, A. & Wallace, J. C. (1993). The existence of multiple tetrameric conformers of chicken liver pyruvate carboxylase and their roles in dilution inactivation. *Biochem J* **290**, 583-90.
36. Perham, R. N. (1991). Domains, Motifs, and Linkers in 2-Oxo Acid Dehydrogenase Multienzyme Complexes: A Paradigm in the Design of a Multifunctional Protein. *Biochemistry* **30**, 8501-12.
37. Reddy, D. V., Rothemund, S., Shenoy, B. C., Carey, P. R. & Sonnichsen, F. D. (1998). Structural characterization of the entire 1.3S subunit of transcarboxylase from *Propionibacterium shermanii*. *Protein Sci* **7**, 2156-63.
38. Xiang, S. & Tong, L. (2008). Crystal structures of human and *Staphylococcus aureus* pyruvate carboxylase and molecular insights into the carboxyltransfer reaction. *Nat Struct Mol Biol*.

Chapter 6: Materials and Methods

6.1 General Materials

6.1.1 General Reagents

Chemicals and solvents were of the appropriate quality and were purchased from GE Healthcare, Amersham Bioscience, Promega, Sigma-Aldrich, Pierce, Qiagen, Invitrogen, Bio-Rad, Vivascience, Roche and New England Biolabs unless otherwise stated. D-biotin was obtained from sigma.

Aquifex aeolicus plasmids pET28a/BPL, pET6H/BCCP154 pET6H/BCCPΔ67 and pET6H/BCCPΔ67 K117L were obtained from David Clarke (University of Edinburgh). The pBtac/BirA R118G plasmid was a kind gift from John E. Cronan Jr (University of Illinois). The plasmids expressing the histidine-tagged *E. coli* BCCP and BC were generously supplied by Aileen Soriano (Schering-Plough-Hialeah).

6.1.2 Media and Solutions.

Sterilisation of media – All media were autoclaved at 121 °C for 20 minutes at 15 psi prior to use.

Luria Bertani (LB): tryptone (10 $g.l^{-1}$), yeast extract (5 $g.l^{-1}$), sodium chloride (10 $g.l^{-1}$); pH adjusted to 7.5 with sodium hydroxide.

SOC: tryptone (20 $g.l^{-1}$), yeast extract (5 $g.l^{-1}$), sodium chloride (0.5 $g.l^{-1}$), magnesium sulfate (5 $g.l^{-1}l$), glucose (3.2 $g.l^{-1}$); pH adjusted to 7.5 with sodium hydroxide.

Agar plates: bacto-agar (15 $g.l^{-1}$) was added to the LB media. The specific antibiotic was added prior to making the plates.

X-Gal plates: S-Gal™/LB Agar Blend from Sigma-Aldrich containing tryptone (10 $g.l^{-1}$), yeast extract (5 $g.l^{-1}$), sodium chloride (10 $g.l^{-1}$), agar (12 $g.l^{-1}$), S-Gal (0.3 $g.l^{-1}$), Ferric ammoniumcitrate (0.5 $g.l^{-1}$), IPTG (0.03 $g.l^{-1}$) was dissolved in the appropriate amount of deionised water according to the manufacturer instructions. The specific antibiotic was added prior to making the plates.

DNA-running buffer: TAE buffer – 40 mM Tris-HCl, 20 mM Acetic acid, 1 mM EDTA (pH 8.3).

6.1.3 Purification Buffers

Purification of *A. aeolicus* *Aa*BPL, *Aa*BPL R40G, BCCP154, apo- and holo-BCCPΔ67 and BCCPΔ67 K117L

- Buffer A (binding buffer): 10 mM HEPES (pH 7.5).
- Buffer B (elution buffer): 10 mM HEPES, 1 M NaCl (pH 7.5).
- Buffer C (gel filtration buffer): 10 mM HEPES, 150 mM NaCl (pH 7.5).
- Buffer D (binding buffer): 20 mM Tris-HCl, 0.5 M NaCl, 5 mM imidazole (pH 7.5).
- Buffer E (elution buffer): 20 mM Tris-HCl, 0.5 M NaCl, 100 mM imidazole (pH 7.5).

Purification of *E. coli* BirA R118G

- Buffer F (binding buffer): 50 mM Tris-HCl, 300 mM NaCl, 5 mM imidazole (pH 7.8).
- Buffer G (elution buffer): 50 mM Tris-HCl, 300 mM NaCl, 100 mM imidazole (pH 7.8).

Purification of *E. coli* BCCP

- Buffer H (lysis buffer): 50 mM HEPES, 300 mM NaCl, 10% glycerol, 20 mM 2-mercaptoethanol (pH 8.0).
- Buffer I (wash buffer I): 50 mM HEPES, 300 mM NaCl, 10% glycerol, 20 mM 2-mercaptoethanol, 20 mM imidazole (pH 8.0).
- Buffer J: (wash buffer II): 50 mM HEPES, 300 mM NaCl, 10% glycerol, 20 mM 2-mercaptoethanol, 50 mM imidazole (pH 8.0).
- Buffer K (elution buffer I): 50 mM HEPES, 300 mM NaCl, 10% glycerol, 20 mM 2-mercaptoethanol, 170 mM imidazole (pH 8.0).
- Buffer L (elution buffer II): 50 mM HEPES, 300 mM NaCl, 10% glycerol, 20 mM 2-mercaptoethanol, 250 mM imidazole (pH 8.0).
- Buffer M (binding buffer): 50 mM HEPES, 50 mM Nacl, 10% glycerol, 1.0 mM DTT (pH 7.5)
- Buffer N (elution buffer): 50 mM HEPES, 1 M NaCl, 10% glycerol, 1.0 mM DTT (pH 7.5)
- Buffer O (gel filtration buffer): 50 mM HEPES, 150 mM Nacl, 10% glycerol, 1.0 mM DTT (pH 7.5)

Purification of *E. coli* BC

- Buffer P (binding buffer): 20mM HEPES, 150 mM NaCl, 5 mM imidazole, 1 mM DTT, 0.1 mM EDTA, 10% glycerol (pH 7.5).
- Buffer Q (elution buffer): 20 mM HEPES, 150mM NaCl, 1 M imidazole, 1 mM DTT, 0.1 mM EDTA, 10% glycerol (pH 7.5).
- Buffer R (gel filtration buffer): 20 mM HEPES, 150 mM NaCl, 1 mM DTT, 0.1 mM EDTA, 10% glycerol (pH 7.5).

6.2 Molecular Biology

6.2.1 Bacterial Cell Lines

Strain	Genotype	Applications
K-12 MG1655	F⁻ λ *ilvG rfb*50 *rph*1	DNA isolation.
Top 10™ (Competent)	F⁻ *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) ϕ80*lacZ*ΔM15 Δ*lac*X74 *deo*R *rec*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K λ⁻ *rps*L (Str^R) *nup*G	Cloning.
DH5α™ (Competent)	F⁻ ϕ80*lacZ*ΔM15 Δ(*lacZYA-argF*) U169 *deo*R *rec*A1 *end*A1 *hsd*R17 (r_{k-}, m_{k+}) *pho*A *sup*E44Λ⁻ *thi*-1 *gyr*A96 *rel*A1	Cloning.
JM109 (Competent)	*end*A1 *rec*A1 *gyr*A96 *thi hsd*R17 (r_{k-}, m_{k+}) *rel*A1 *sup*E44, Δ(*lac-proAB*), [F′ *tra*D36 *proAB*, *laq*IqZΔM15]	Protein expression. Transformation of DNA ligations.
BL21(DE3) (Competent)	F⁻ *omp*T hsdS_B (r_B^- m_B^-) *gal dcm* (DE3)	Protein expression.

6.2.2 Oligonucleotide primers

The following oligonucleotide primers were used this study. Restriction and mutation sites are underlined.

Name	Sequence (5'-3')
*Aa*BPL R40G For	GGAAGGGGAGGACT**CGA**AGGAAGTGGCTC
*Aa*BPL R40G Rev	GAGCCACTTCCTTC**CGA**GTCCTCCCCTTCC
AccBC For	GCTACAATCAC**TCTAGA**CAAA
AccBC Rev	CCTTATGGGGGTT**CTCGAG**CC
pGEM For	GAATACTCAAGCTATGC
pGEM Rev	AGTGAATTGTAATACGACT

6.2.3 DNA Manipulation

6.2.3.1 Purification of Plasmid DNA

Plasmid DNA was prepared using QIAprep® Spin Miniprep Kit (Qiagen) following the manufacturer's instructions.

6.2.3.2 Isolation of *E. coli* Genomic DNA

The cells from the *E. coli* K-12 MG1655 strain were spread on an agar plate and left at room temperature until sufficient colonies were obtained. Multiple colonies were resuspended in 20 µl lysis buffer containing 0.25% w/v sodium dodecyl sulphate (SDS) and 0.05 M NaOH. The mixture was incubated for 15 min at

95 °C and centrifuged at 11,000 x *g* for 20 sec. Distilled water was added (180 ml) and the cell debris was harvested by centrifugation for 5 min.

6.2.3.3 Transformation *E. coli* Competent Cells with Recombinant DNA

Competent cells were transformed according to the manufacturers' instructions. DNA (up to 40 ng) was added to an aliquot of competent cells and gently mixed. This was left on ice for 30 minutes before the cells were heat shocked (42 °C, 40 sec). The cells were then grown in 250 µl SOC medium at 37 °C for 1 hr. Finally, the cells were spread to dryness on selective agar plates and incubated at 37 °C overnight.

6.2.3.4 Electrophoresis of DNA

The required amount of agarose was added to TAE buffer (typically 1.2 g / 100 ml) and heated at 100 °C until dissolved. The solution was allowed to cool to 55 °C and ethidium bromide was added to a final concentration of 0.5 $\mu g.ml^{-1}$. The gel was then poured into the casting mould and allowed to set at room temperature. Loading dye (Promega) was added to the sample prior to loading and the DNA migrated using a potential difference of 100 V for an adequate time to achieve separation. The gels were viewed and photographed under UV-light. HyperLadder I and IV (Bioline) were used for quantitation of the DNA.

6.2.3.5 Digestion of DNA with Restriction Endonucleases

The required amount of DNA (0.5 - 1 µg) was treated with the appropriate amount of endonuclease and buffer and incubated for at least 3 hours at 37 °C before

analysis by electrophoresis on agarose. The restriction enzyme *EcoR*I was used as a control for the digestion of pGEM-T easy plasmids.

6.2.3.6 Gel-Extraction of DNA

DNA was purified from agarose using QIAquick® Gel Extraction Kit (Qiagen) following the manufacturer's instructions.

6.2.3.7 Direct Cloning of PCR Products

All PCR products were cloned into the pGEM-T easy vector (Promega) using the manufacturer's guidelines. 2x rapid T4 DNA ligase buffer (2 µl), pGEM-T easy vector (1 µl), PCR product (3 µl) and T4 DNA ligase from Promega (1 µl) were gently mixed and incubated for one hour at room temperature. 2 µl of the reaction mixture was used to transform JM109 competent cells.

6.2.3.8 Cloning into Plasmid Vectors

The DNA fragment cut with suitable restriction enzymes (8 µl), the host vector cut with suitable restriction enzymes (2 µl), 2x Quick ligation buffer (2 µl), and Quick T4 DNA ligase from New England Biolabs (1 µl) were gently mixed and incubated for 10 min at room temperature. The reaction mixture (3 µl) was used to transform JM109 competent cells.

6.2.3.9 Storage of Bacterial Stocks

LB medium containing the appropriate antibiotics were used for the short term storage of *E. coli*. Colonies of bacteria were stored on inverted agar plates at 4

°C for up to 28 days. For long term storage, strains were frozen (-80 °C) in LB media containing 20% glycerol. Purified plasmids DNA were stored at -20 °C for up to one year.

6.2.4 Polymerase chain reactions

Polymerase chain reactions (PCRs) were performed using a Perkin Elmer 480 thermal cycler and the reactions were overlayed with ~30 µl oil. PCR sequencing reactions were performed on an Eppendorf Mastercycler Personal which has a hot-top assembly and does not require the addition of oil.

6.2.4.1 Amplification of DNA

A typical amplification PCR contained 2 Ready to Go PCR™ beads (Amersham Biosciences), DNA template (2 µl), primer-forward 10 µM (5 µl,), primer-reverse 10 µM (5 µl), and distilled water (final volume of 50 µl). The cycling parameters of the PCR are listed bellow. The PCR product was then subjected to agarose gel electrophoresis and the required band was excised and the DNA purified as described in section 6.2.3.6.

PCR steps	Temperature (°C)	Time (min)	Number of cycles
Initial denaturation	95	2	1
Denaturation	95	1	25
Annealing	55	1	25
Extension	72	2	25
Termination	72	10	1

6.2.4.2 PCR Megaprimer Mutagenesis

Mutations were performed by using the Stratagene site-directed mutagenesis kit. Each reaction contained plasmid DNA (5 µl), 10× reaction buffer (5 µl), primer-forward 10 µM (1 µl), primer-reverse 10 µM (1 µl), dNTP mix (1 µl), distilled water (final volume 50 µl) and Pfu DNA polymerase 2.5 u/µl (1 µl). The cycling parameters of the PCR are listed bellow

.

PCR steps	Temperature (°C)	Time	Number of cycles
Initial denaturation	95	2 min	1
Denaturation	95	50 sec	18
Annealing	60	50 sec	18
Extension	68	1 min / 1000 bp	18
Termination	68	7 min	1

6.2.4.3 DNA sequencing

The sequencing reactions were performed using the BigDye® Terminator v3.1 Cycle sequencing kit (PE Applied Biosystems). The PCR mixture contained DNA template ~5 pmol (5 µl), 5× reaction buffer (2 µl), primer 10 µM (1 µl) and Big Dye 3.1 (2µl). Sequencing of the 5'-end of the DNA template was carried out with the primer pGEM-forward while the sequencing of the 3'-end was carried out with the primer pGEM-reverse. The cycling parameters of the PCR are listed bellow.

PCR steps	Temperature (°C)	Time	Number of cycles
Denaturation	95	30 sec	30
Annealing	60	20 sec	30
Extension	68	4 min	30

Automated DNA sequencing was performed on an ABI prism 377 DNA sequencer using the Sanger dideoxy chain termination method. Sequence data were analysed using Contig Express within the Vector NTI Advance™ V9 software package.

6.3 Protein Expression and Purification

6.3.1 Polyacrylamide Gel Electrophoresis (PAGE)

SDS-PAGE was used to analyse proteins on the basis of their molecular mass with a Tris-Glycine or "Laemmli" discontinuous buffering system.[1] Alternatively, proteins were analysed on precast 12% Bis-Tris Nu-PAGE gels (Invitrogen)

according to manufacturer's instructions. Proteins were analysed in their native form by non-denaturing gel electrophoresis using Novex® 12% Tricine or 12% Novex® Tris-Glycine precast gels (Invitrogen). Gels were visualised using Comassie blue.

6.3.2 Expression and Purification of *Aa*BPL and Mutant *Aa*BPL R40G

6.3.2.1 Large Scale Expression of *Aa*BPL and *Aa*BPL R40G

*Aa*BPL was expressed and purified as described previously by Clarke *et al* with the following modifications.[2] Overexpression of *A. aeolicus* *Aa*BPL and mutant *Aa*BPL R40G was achieved by transforming *E. coli* BL21(DE3) (Novagen) cells with the vectors pET28a/*Aa*BPL and pET28a/*Aa*BPL R40G respectively. A single colony was added to 250 ml LB supplemented with kanamycin (30 $\mu g.ml^{-1}$) and grown overnight at 37 °C and 200 rpm. This seed culture was then used to inoculate 4 litres of fresh LB medium and grown under similar conditions to OD_{600} = 1.0 before induction with isopropyl-1-thio-β-D-galactopyranoside (ITPG) to a final concentration of 1.0 mM. After a further 3 hours, the cells were harvested by centrifugation (5,000 x *g* for 15 minutes at 4 °C), washed with buffer A, and stored at -20 °C.

6.3.2.2 Purification of *Aa*BPL and *Aa*BPL R40G

Cells overexpressing *Aa*BPL were resuspended in buffer A (5 ml per gram of wet cell paste) with one tablet of Complete™ Proteinase Inhibitor Cocktail (Roche) and disrupted by sonication (15 pulses of 30 seconds at 30 second intervals) at 4 °C.

The cell debris was removed by centrifugation two times for 15 minutes at 27,000 x *g* and at 4 °C in presence of 1 µl DNase 10 u/µl (Roche). The supernatant was filtered through a 0.45 µM membrane before being loaded on a cation exchange 10 ml Tricorn HR 10/100 column containing 15S beads (GE Healthcare) pre-equilibrated with buffer A. The column was then washed with 5 column volumes of buffer A and the protein was eluted with a linear salt gradient (0-100% buffer B) over 20 column volumes at 4 °C. Fractions containing *Aa*BPL were run on either SDS-PAGE or Nu-PAGE gels and those fractions judged to be 95% pure were dialysed overnight against buffer A at 4 °C. Protein concentration was determined with the Bradford protein assay (Bio-Rad) and by measuring the absorbance at 280 nm using the molar extinction coefficient 34,850 M^{-1} cm^{-1} which was calculated using the Vector NTI Advance™ V9 software. *Aa*BPL was concentrated to 1 $mg.ml^{-1}$ using a Vivaspin concentrator 20,000 MWCO (Vivascience) and stored in buffer A containing 20% glycerol (v/v) at –20°C. The mutant *Aa*BPL R40G mutant was purified in a similar manner.

6.3.3 Expression and Purification of *A. aeolicus* Full-Length BCCP: BCCP154

6.3.3.1 Large Scale Expression of BCCP154

Overexpression of *A. aeolicus* His_6-Tagged BCCP154 and untagged BCCP154 was carried out by transforming *E. coli* BL21(DE3) cells with the pET6H/BCCP154 and pET16b/BCCP154 vectors respectively. A single colony was added to 250 ml LB supplemented with ampicillin (100 $\mu g.ml^{-1}$) and grown

overnight at 37 °C and 200 rpm. This seed culture was then used to inoculate 4 litres of fresh LB medium and grown to OD_{600} = 0.6 before induction with ITPG (1.0 mM final concentration) for 3 hours at 37 °C and 200 rpm. The cells were then harvested by centrifugation (5,000 x *g* for 15 minutes at 4 °C), washed with buffer A and stored at -20 °C.

6.3.3.2 Purification of N-terminal His_6-Tagged BCCP154

Initial purification of *A. aeolicus* full length BCCP154 carried out with the N-terminal His_6-Tagged protein by nickel affinity chromatography was shown to be unsuccessful. Cells overexpressing His_6-BCCP154 were resuspended in binding buffer D (5 ml per gram of wet cell paste) with one tablet of Complete™ Proteinase Inhibitor Cocktail and disrupted by sonication (15 pulses of 30 seconds at 30 second intervals) at 4 °C. The cell debris was removed by centrifugation at 27,000 x *g* for 20 minutes at 4 °C in presence of 1 µl DNase 10 u/µl. The supernatant was filtered through a 0.45 µm membrane prior to chromatography at 4 °C. The cell lysate was loaded onto a 5 ml HisTrap™ HP column (GE Healthcare) previously equilibrated with binding buffer D. The column was then washed with 5 column volumes of buffer D before elution using a linear gradient of imidazole (0-100% buffer E). Fractions were analysed by SDS-PAGE and the His-tagged protein which did not bind to the nickel column, was shown to be in the flow-through fraction with the rest of the unbound material.

6.3.3.3 Purification of Untagged BCCP154

After being cloned into a pET16b vector, *A. aeolicus* BCCP154 was purified by cation exchange chromatography. Cells overexpressing BCCP154 were resuspended in binding buffer A (5 ml per gram of wet cell paste) with one tablet of Complete™ Proteinase Inhibitor Cocktail and disrupted by sonication (15 pulses of 30 seconds at 30 second intervals) at 4 °C. The cell debris was removed by centrifugation at 27,000 x *g* for 20 minutes at 4 °C in presence of 1 µl DNase 10 u/µl. The supernatant was filtered through a 0.45 µm membrane prior to chromatography at 4 °C. The cell lysate was first loaded on a 55 ml Q-Sepharose XL (Amersham) previously equilibrated with binding buffer A. The column was then washed with 5 column volumes of buffer A and the bound material eluted using a linear salt gradient of 0-100% buffer B over 15 column volumes. The fractions were analysed by SDS-PAGE and those containing a significant proportion of protein BCCP154 were dialyzed overnight against buffer A at 4 °C. The protein mixture was filtered through a 0.45 µm membrane before being loaded on a 1 ml MonoQ HR 5/5 (Amersham) column equilibrated with buffer A. The bound material eluted with a salt gradient (0-100% buffer B) over 20 column volumes and the fractions analysed be SDS-PAGE which contained BCCP154 were dialyzed overnight against buffer A at 4 °C. The protein was concentrated to a volume of 4 ml using a Vivaspin concentrator 10,000 MWCO and, after filtration through a 0.45 µm membrane, was loaded on a 120 ml HiLoad 16/60 Superdex75 gel filtration column (Amersham) previously equilibrated with buffer C. The bound material eluted with buffer C at a flowrate of 1ml.min^{-1} and SDS-PAGE analysis of the fractions indicated BCCP154 to be 95% pure. BCCP154 was dialysed overnight against buffer A at 4 °C. Due to the

low proportion of aromatic residues in BCCP154, protein concentration was determined using the BCA™ protein assay kit (Pierce) and by measuring the absorbance at 280 nm (molar extinction coefficient 6,520 M^{-1} cm^{-1}). BCCP154 was concentrated to 1 $mg.ml^{-1}$ using a 10,000 MWCO and stored in buffer A containing 20% glycerol (v/v) at –20°C.

6.3.4 Expression and purification of *A. aeolicus* BCCPΔ67 and BCCPΔ67 K117L

6.3.4.1 Large scale expression of BCCPΔ67 and BCCPΔ67 K117L

The pET6H/BCCPΔ67 and pET6H/BCCPΔ67 K117L vectors were used to transform *E. coli* BL21(DE3). A single colony was added to 250ml LB supplemented with ampicillin (100 $\mu g.ml^{-1}$) and grown overnight at 37 °C and 200 rpm. This seed culture was then used to inoculate 4 litres of fresh growth medium and grown under similar conditions to OD_{600} = 1.0 before being induced with ITPG (1.0 mM final concentration). After a further 3 hours, the cells were harvested by centrifugation (5,000 x *g* for 15 minutes at 4 °C) and stored at -20 °C.

6.3.4.2 Purification of BCCPΔ67 and BCCPΔ67 K117L

BCCPΔ67 and mutant BCCPΔ67 K117L were expressed and purified as described previously by Clarke *et al.*[2] Cells overexpressing His_6-BCCPΔ67 were resuspended in binding buffer D (5 ml per gram of wet cell paste) with one tablet of Complete™ Proteinase Inhibitor Cocktail and disrupted by sonication (15 pulses of 30 seconds at 30 second intervals) at 4 °C. The cell debris was removed by

centrifugation at 27,000 x *g* for 20 minutes at 4 °C, after which the supernatant was filtered through a 0.45 μm membrane prior to chromatography. The cell lysate was loaded onto a 5 ml HisTrap™ HP column (GE Healthcare) previously equilibrated with binding buffer D. The column was then washed with 5 column volumes of buffer D before the bound material was eluted using a linear gradient of imidazole (0-100% buffer E) over 20 column volumes at 20 °C. Fractions were analysed by SDS-PAGE and those containing BCCPΔ67 were pooled and dialysed overnight against 4 litres of buffer A at 20 °C. The mutant BCCPΔ67 K117L was purified in a similar manner.

Apo- and holo-BCCPΔ67 were separated by applying the BCCPΔ67-containing fractions eluted from the nickel column onto a 1 ml MonoQ HR 5/5 column (Amersham) pre-equilibrated with buffer A The column was then washed with 10 column volumes of buffer A before the protein eluted with a linear salt gradient (0-100% buffer B) over 30 column volumes at 20 °C. The elution was monitored by SDS-PAGE and fractions containing apo- and holo-BCCPΔ67 were screened by LC-ESI-MS. The concentration of apo-BCCPΔ67, holo-BCCPΔ67 and mutant BCCPΔ67 K117L was determined by using the BCA™ protein assay kit and by measuring the absorbance at 280 nm (molar extinction coefficient 3,960 M^{-1} cm^{-1}). The separated apo- and holo-proteins and the K117L mutant BCCPΔ67 were concentrated to 1 $mg.ml^{-1}$ using a Vivaspin concentrator 3,000 MWCO and stored in buffer A containing 20% glycerol (v/v) at -20°C.

6.3.5 Expression and Purification of *E. coli* Mutant BirA R118G

6.3.5.1 Large Scale Expression of BirA R118G

The *E. coli* BirA R118G mutant was expressed according to the method described by Ting.[3] The pBTac vector containing the gene for the His-tagged *E. coli* mutant BirA R118G was used to transform JM109 cells (Novagen). A single colony was added to 10 ml LB medium supplemented with ampicillin (100 $\mu g.ml^{-1}$) and grown overnight at 37°C and 200 rpm. This seed culture was used to inoculate 250 ml LB supplemented with ampicillin (100 $\mu g.ml^{-1}$) and grown overnight under similar conditions. The overnight growth was then added to 4 litres of fresh LB medium and grown at 37 °C and 200 rpm to OD_{600}=0.9. Enzyme expression was induced with IPTG to a final concentration of 0.4 mM at 30 °C and 200 rpm. After a further 3 hours, the cells were harvested by centrifugation (5,000 x *g* for 15 minutes at 4 °C), washed with buffer F in presence of one tablet of Complete™ Proteinase Inhibitor Cocktail and stored at -20 °C.

6.3.5.2 Purification of BirA R118G

The cells overexpressing His_6-BirA R118G were resuspended in buffer F (5 ml per gram of cell paste) and disrupted by sonication at 4 °C (three 10-second pulses with 1 minute interval between each pulse). The cell debris was removed by centrifugation (27,000 x *g*, 20 min, 4 °C) and the supernatant kept at 4 °C, was filtered through a 0.45 μm membrane. The His_6-tagged soluble protein was purified at 20°C from the lysate using a 5ml HisTrap™ HP nickel column (GE Healthcare) pre-equilibrated with buffer F. The column was washed with 10 column volumes of binding buffer before bound material was eluted using a linear gradient of 0-100% of

buffer G over 25 column volumes. The fractions were analysed by SDS-PAGE and those fractions, judged to be 95% pure, were dialysed overnight against phosphate buffer saline (pH 7.4) at 4 °C. The concentration of BirA R118G was determined using the Bio-Rad protein assay kit and by measuring the absorbance at 280 nm (molar extinction coefficient 47,630 M^{-1} cm^{-1}). The purified BirA R118G was concentrated to 1 $mg.ml^{-1}$ using a Vivaspin concentrator 20,000 MWCO and stored in phosphate buffer saline containing 20% glycerol (v/v) at -80°C.

6.3.6 Expression and Purification of *E. coli* BCCP (*Ec*BCCP)

6.3.6.1 Large Scale Expression of *Ec*BCCP

Overexpression of the full length *E. coli* *Ec*BCCP was achieved by transforming *E. coli* BL21(DE3) (Novagen) cells with the vectors pET29a/*Ec*BCCP and pBirAcm Avidity (AviTag™). A single colony was added to 250 ml LB supplemented with kanamycin (200 $\mu g.ml^{-1}$) and chloramphenicol (20 $\mu g.ml^{-1}$) and grown overnight at 37 °C and 200 rpm. This seed culture was then used to inoculate 4 litres of fresh LB medium supplemented with kanamycin (200 $\mu g.ml^{-1}$), chloramphenicol (20 $\mu g.ml^{-1}$) and biotin (30 $\mu g.ml^{-1}$) and grown at 37 °C and 200 rpm to OD_{600} = 0.7. Overexpression was induced with ITPG to a final concentration of 1.0 mM for 5 hours. The cells were then harvested by centrifugation (5,000 x *g* for 15 minutes at 4 °C), washed with buffer H, and stored at -20 °C.

6.3.6.2 Purification of *Ec*BCCP

The C-terminal His_6-tagged BCCP was first purified with nickel affinity followed by anion exchange chromatography. Cells overexpressing *Ec*BCCP were resuspended in buffer H (5 ml per gram of wet cell paste) with one tablet of Complete™ Proteinase Inhibitor Cocktail. The resuspended cell paste was shaked at 37 °C for 30 min and the cells were then disrupted by sonication (5 times 1 min at 30 seconds interval) at 4 °C. The cell debris was removed by centrifugation two times for 15 minutes at 27,000 x *g* and at 4 °C in presence of 1 µl DNase 10 u/µl (Roche). The supernatant was then further clarified by ultracentrifugation at 120,000 x *g* and at 4 °C for 1 hour. 2 ml of Ni-NTA agarose (Qiagen) was washed with 10 ml of distilled water and equilibrated with 3 times 10 ml of buffer H. The Ni-NTA agarose was centrifuged at 1,000 x *g* and at 4 °C for 1 min and was then slowly mixed in a beaker with the lysate and left to set for 2 hours at 4 °C. The mixture was then loaded on a 5 ml disposable column (Pierce) washed with buffer I. The presence of protein in the washes was monitored by using the Bradford reagent from the Bio-Rad protein assay kit (1 drop of solution / 1 ml of Bradford reagent). The slurry was then washed with buffer J until no protein could be detected with the Bradford method. The slurry was further purified with 5 column volumes of buffer J before elution of the protein with buffer K (170 mM of imidazole) over 10 column volumes. The protein was collected in 1 ml fractions and was further eluted with buffer L (250 mM imidazole) until no more protein could be detected with the Bradford method. The flow-through and the 17 fractions were analysed by SDS-PAGE before being extensively dialysed against buffer M at 4 °C.

The protein was then filtered through a 0.45 μm membrane before being loaded on a 1 ml MonoQ HR 5/5 column (Amersham) pre-equilibrated with binding buffer M. Elution of the protein was performed with a linear gradient of salt 0-1 M NaCl (buffer N) over 20 column volumes. The fractions were analysed by SDS-PAGE and judged to be 90% pure and the protein was then extensively dialysed against buffer O at 4 °C. Protein concentration was determined by measuring the absorbance at 280 nm (molar extinction coefficient 2,680 M^{-1} cm^{-1}) and the fractions were concentrated to 4 ml using a Vivaspin concentrator 3,000 MWCO. After filtering through a 0.45 μM membrane, *Ec*BCCP was loaded on a gel filtration 320 ml Sephacryl S-200 HR column (Amersham) pre-equilibrated with buffer O. The protein eluted with buffer O at a flowrate of 1ml.min^{-1} and SDS-PAGE analysis of the fractions showed *Ec*BCCP to be 95% pure. *Ec*BCCP was dialysed overnight against 4 L buffer O in presence of 20 % glycerol at 4 °C. Protein concentration was determined with the Bio-Rad protein assay and by measuring the absorbance at 280 nm. *Ec*BCCP was concentrated to 1 mg.ml^{-1} using a Vivaspin concentrator 3,000 MWCO and stored in buffer O (20% glycerol v/v) at -20°C.

6.3.7 Expression and Purification of *E. coli* BC

6.3.7.1 Large Scale Expression of BC

The plasmid pET16b/BC was used to transform *E. coli* BL21(DE3). A single colony was added to 250 ml LB supplemented with ampicillin (1 μg.ml^{-1}) and grown overnight at 37 °C and 200 rpm. This seed culture was then used to inoculate 3 litres of fresh LB medium and grown subject to the same conditions to OD_{600} = 0.6 before being induced with ITPG (1.0 mM final concentration). The overexpression was

allowed to proceed for 5 hours, and the cells were harvested by centrifugation (5,000 x *g*, 15 min, 4 °C), washed with buffer R and stored at -20 °C.

6.3.7.2 Purification of BC

Cells overexpressing BC were resuspended in buffer P (5 ml per gram of wet cell paste) with one tablet of Complete™ Proteinase Inhibitor Cocktail (Roche) and disrupted by sonication (5 times 1 min at 30 seconds intervals) at 4 °C. The cell debris was removed by centrifugation two times for 15 minutes at 27,000 x *g* and at 4 °C in presence of 1 µl DNase 10 u/µl (Roche). The supernatant was filtered through a 0.45 µM membrane before being loaded on a HisTrap™ HP column (GE Healthcare) previously equilibrated with buffer P. The column was then washed with 5 column volumes of buffer P and the BC protein eluted with a linear gradient of imidazole (0-100% buffer Q) over 20 column volumes at 4 °C. Fractions were analysed by SDS-PAGE and those containing BC were pooled and dialysed overnight against 4 litres of buffer R at 4 °C. Protein concentration was determined by measuring the absorbance at 280nm using (molar extinction coefficient 25,850 M^{-1} cm^{-1}) and the fractions were concentrated to 4 ml using a Vivaspin concentrator 20,000 MWCO. After filtering through a 0.45 µM membrane, BC was loaded on a gel filtration 320 ml Sephacryl S-200 HR column previously equilibrated with buffer R. The protein eluted with buffer R at a flowrate of 1ml.min^{-1} and SDS-PAGE analysis of the fractions indicated BC to be 95% pure. BC was dialysed overnight against 4 L buffer R at 4 °C. Protein concentration was determined with the Bio-Rad protein assay and by measuring the absorbance at 280 nm. BC was concentrated to 2 mg.ml^{-1} and stored in buffer R containing 20% glycerol (v/v) at –20°C.

6.4 Protein Chemistry

6.4.1 Protein Characterisation

6.4.1.1 Liquid chromatography-mass spectrometry (LC-ESI-MS)

LC-ESI-MS was performed on a MicroMass Platform II quadrupole mass spectrometer equipped with an electrospray ion source. The spectrometer cone voltage was ramped from 40 to 70 V and the source temperature set to 140 °C. Protein samples were separated on a Jupiter C5 reverse phase column (5 μm, 250 x 4.6 mm, Phenomenex) with a Waters HPLC 2690 directly connected to the spectrometer. Proteins were eluted from the column with a 5-95% acetonitrile (containing 0.01% TFA) gradient at a flow rate of 0.1 $ml.min^{-1}$. The total ion count in the range 500-2000 m/z was scanned at 0.1 s intervals. The scans were accumulated, spectra combined and the molecular mass determined by the MaxEnt and Transform algorithms of the Mass Lynx software (Micromass, U. K.).

6.4.1.2 Nano-electrospray mass spectrometry (nESI-MS)

Nanoelectrospray MS data were acquired on a Q-ToF II instrument (Waters, Manchester, UK) and on a MDS Sciex Q-star XL (Concord, ON, Canada) in positive mode, under conditions optimised for the transmission and efficient desolvation of high-mass non-covalent complexes. Instrument parameters and pressures throughout the instruments were adjusted to enhance these processes.[4] Samples were introduced into the instruments via nanoflow-electrospray capillaries. These capillaries were prepared using a micropipette puller (Flaming/Brown P-97, Sutter Instruments, Novato, CA, USA) and glass capillary tubes of 1 mm outer-diameter and 0.78 mm inner-diameter (Harvard Apparatus, Holliston, MA, USA). A thin layer of gold

coating, to make the capillaries electrically conductive, was applied using a SEM sputter-coater (Polaron, Newhaven, UK). The pulled end of the capillaries were clipped under a stereomicroscope resulting in an inner tip diameter of 2-5 μm. Capillaries were loaded, using GEloader tips (Eppendorf), with 2 μl of aqueous protein solution. Proteins were sprayed from these capillaries with the aid of a backing pressure to initiate and maintain a steady spray. Prior to data collection the mass spectrometer was calibrated using caesium iodide (100 mg ml^{-1}). Data were acquired and processed with MassLynx software 4.0 (Waters, Manchester, UK). All spectra are shown with minimal smoothing and without background subtraction.

The Q-TOF II mass spectrometer is equipped with a Z-spray source in which the direction of the electrospray is orthogonal to the sample cone orifice. A radio frequency (RF) is applied to the first hexapole to direct and collimate the whole ion beam into the quadrupole analyser. The ion path is further focused by a second hexapole, before ions reach the ToF analyser which is orthogonal to the ion beam and where ion separation is achieved. A reflectron is used in the ToF analyser to reduce the kinetic energy spread of ions which would diminish spectral resolution. Ions are finally detected using a Microchannel Plate Detector (MCP) detector. Here, the incident ion signal results in an electron emission from a channel wall in the MCP which is then accelerated by a potential applied across the plate. The current which reaches the rear of the plate is then transformed into a mass spectrum using a time-to-digital converter.

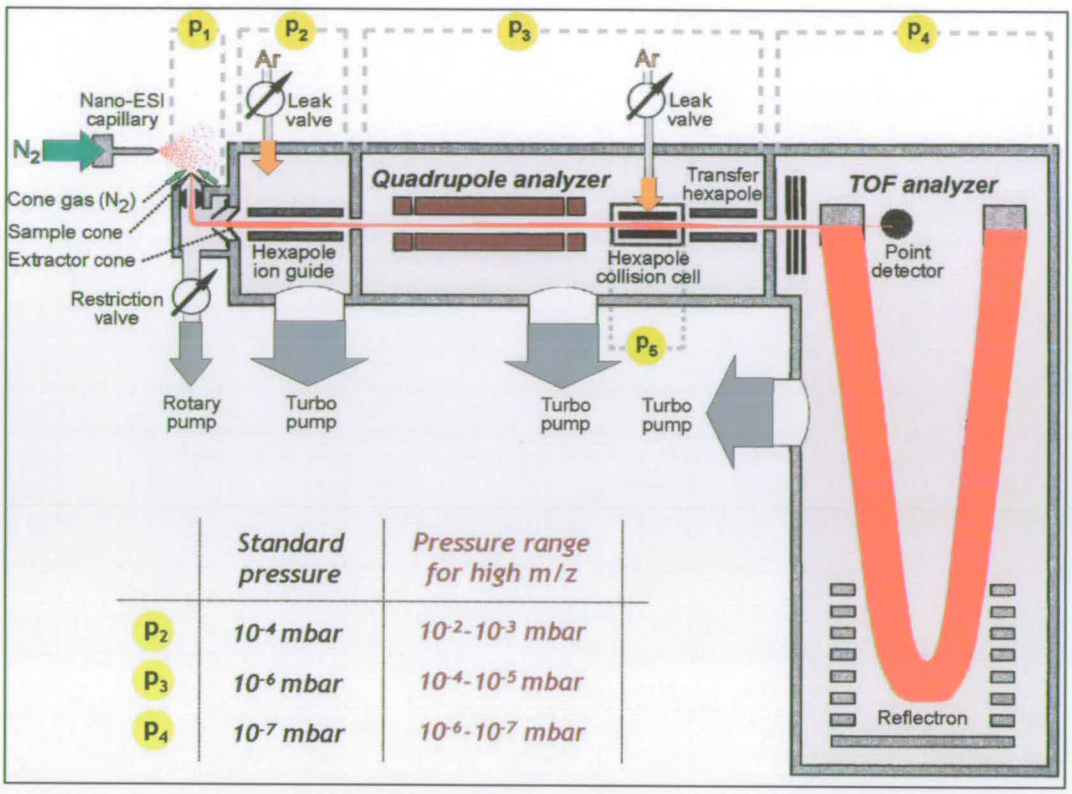


Figure 6.1 - Schematic of a Q-ToF II mass spectrometer.

In the MDS Sciex Q-star spectrometer, the pre-ToF ion path consists of three quadrupoles (ion guide Q0, mass analyser RO1, collision cell RO2). RO1 is the mass-filtering device, whereas Q0 and RO2 are operated in the RF-mode only and function as ion guide devices. In contrast with the Q-ToF II, the nanospray set-up of the Q-star spectrometer has a liner trajectory.[5]

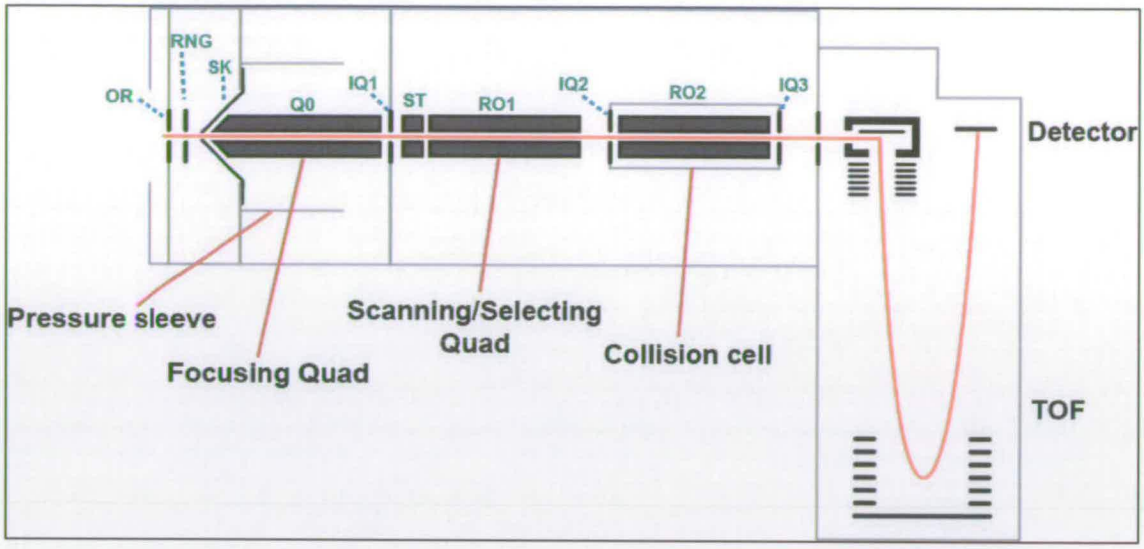


Figure 6.2 - Schematic of the MDS Sciex Q-star XL.

6.4.1.3 Streptavidin Western Blot

Pure proteins were separated by SDS-PAGE on two identical gels. One gel was kept for control and analysis while the proteins from the second gel were transferred on a 0.2 μm pore sized Hybond-ECL nitrocellulose membrane (Amersham) using a Trans blot Semi-Dry Transfer cell. The membrane was incubated overnight in the biotin-free SuperBlock solution at 4 °C. Excess blocking solution was then removed by washing with phosphate buffer containing 0.1% Tween (Sigma) four times before being incubated in SuperBlock solution containing 1/50,000 Streptavidin HRP for one hour at room temperature. The membrane was washed and dried before immunolabelled proteins were detected using the ECL enhanced chemiluminescence detection system (Amersham) in accordance with the manufacturer's protocol. Proteins were visualised on a BioMax XAR film (Kodak).

6.4.2 Isothermal Titration Calorimetry

Isothermal titration calorimetry (ITC) was carried out in collaboration with Pr. Alan Cooper and the Glasgow Biological Microcalorimetry Facility. The ITC measurements were performed on a VP-ITC calorimeter (Microcal Inc Northampton, USA) at 25 °C. *Aa*BPL and *Aa*BPL R40G were exhaustively dialyzed against the buffer system containing 10 mM HEPES, 4 mM EDTA, pH 7.5. The final dialysis buffer was used to make up the ligand solution as well as for instrument equilibration and baseline controls. The concentrations of the wild-type *Aa*BPL and mutant *Aa*BPL R40G were determined by measuring the absorption at 280 nm (15.4 μM *Aa*BPL and 18 μM *Aa*BPL R40G) and the concentrations of ligands were approximately 10 times higher (0.200 mM biotin, 0.204 mM ATP).

The reactants *Aa*BPL and *Aa*BPL R40G were placed in the 2 ml sample chamber and the substrates biotin and ATP in the syringe. A typical ITC measurement consisted of a first control injection of 1 μl followed by 29 successive injections of 10 μl for 20s with a 3 min interval between each injection. *Aa*BPL and *Aa*BPL R40G were titrated with the first ligand biotin or ATP which was left in the cell in order to titrate the binary complexes with the second ligand. The concentrations of the proteins were adjusted in function of the volume added during the first titration. During the titration of *Aa*BPL with streptavidin (Sigma, 0.198 mM in the syringe), the first 1 μl injection was followed by only 12 injections of 10 μl. Control experiments in which the ligands were directly injected in the buffer were performed in order to evaluate the heat contributions due to coupled protonation events upon binding. The observed heat effects were identical to the heat signals

after complete saturation of the proteins. Therefore, the non-specific background was usually estimated by averaging the small heats at the end of the *Aa*BPL and *Aa*BPL R40G titrations.

Raw data were collected, corrected for ligand heats dilution, and the peaks generated integrated using the ORIGIN software (Microcal Inc) by plotting the values in microcalories against the molar ratio of injectant to reactant within the cell. Data were fitted using the one single-site binding model. From the dissociation constant K_D and the reaction enthalpy value ΔH, the change in free Gibbs energy (ΔG^o) and entropy change (ΔS^o) can be calculated using the equation $\Delta G^o = -RT \ln (1/K_D) = \Delta H - T\Delta S^o$ where R is the universal gas constant and T the absolute temperature.

6.4.3 *In vitro* Biotinylation Assays

6.4.3.1 Biotinylation of apo-BCCPΔ67 and BCCPΔ67 K117L with *Aa*BPL and *Aa*BPL R40G

In vitro biotinylation assays of the apo-form of BCCPΔ67 and mutant BCCPΔ67 K117L were carried out with *Aa*BPL or *Aa*BPL R40G and were analysed by LC-ESI-MS and streptavidin Western blot. All biotinylation experiments, except otherwise stated, were carried out in 10 mM HEPES at pH 7.5. For LC-ESI-MS analysis, the reaction mixture contained 100 μM biotin, 1 mM ATP, 2 mM $MgCl_2$ and 40 μM substrate apo-BCCPΔ67 or BCCPΔ67 K117L. The reaction was initiated by the addition of purified *Aa*BPL or *Aa*BPL R40G to a final concentration of 1 μM, and incubated at 65 °C for 30 min. Aliquots of the incubation mixture were taken at

5 min and 10 min intervals and terminated by the addition of ice-cold trifluoroacetic acid (final concentration 25% w/v) and incubation on ice for 30 min. For the streptavidin Western blot studies, the reaction contained 5 μM biotin, 100 μM ATP, 200 μM $MgCl_2$, 2 μM of substrate apo-BCCPΔ67 or BCCPΔ67 K117L and was incubated with 200 nM *Aa*BPL or *Aa*BPL R40G at 65 °C for 15 min. The reactions were terminated by addition of SDS sample buffer prior to SDS-PAGE and western blotting.

6.4.3.2 Biotinylation of BSA with *Aa*BPL and *Aa*BPL R40G

Biotinylation of Bovine Serum Albumin (Pierce) was assayed with wild-type *Aa*BPL and mutant *Aa*BPL R40G at 3 different enzyme concentrations (50 nM, 200 nM, and 500 nM). The reaction contained 2 μM BSA, 5 μM biotin, 1 mM ATP, 2 mM $MgCl_2$ and was initiated by the addition of wild-type or mutant *Aa*BPL. Because BSA aggregates at elevated temperatures, the reaction was incubated at 60 °C for 15 min and was terminated by addition of SDS sample buffer prior to western blotting analysis.

6.4.4 Chemical Crosslinking of *A. aeolicus Aa*BPL and BCCPΔ67

*Aa*BPL was cross-linked chemically with apo-BCCPΔ67, holo-BCCPΔ67 and BCCPΔ67 K117L using the coupling reagents 1-ethyl-3(dimethylamino-propyl)-carbodiimide (EDC) and N-hydrosuccinimide (NHS). All cross-linking experiments, except otherwise stated, were carried out with 30 μM *Aa*BPL and 50 μM of substrate BCCPΔ67 and in presence of 5 mM EDC and 5 mM NHS in a MES buffer pH 6.0.

The reaction was incubated at 60 °C and the formation of activated succidinimidyl ester was quenched after 7 minutes with 10 mM β-mercaptoethanol. The reaction mixture was set at 60°C for 2 hours and aliquots were withdrawn at different time intervals. Termination was carried out using 10 mM hydroxylamine to regenerate acidic residues that had been activated but did not crosslink. Analysis of the time-dependent formation of cross-linked complex was carried out by SDS-PAGE. ESI-MS was used to characterize the molecular mass of the cross-linked species. Prior to ESI-MS analysis, the *Aa*BPL:BCCPΔ67 complex was transferred into 20 mM ammonium acetate pH 6.8 using Vivaspin concentrators of 10,000 and 20,000 MWCO (Vivascience).

To investigate further the cross-linking pathway, apo-BCCPΔ67 was incubated alone for 3 minutes with EDC and NHS at 60 °C and the reaction was quenched with 10 mM β-mercaptoethanol before addition of the enzyme *Aa*BPL. In a similar manner, *Aa*BPL was activated alone for 15 minutes before quenching and addition of apo-BCCPΔ67. The formation of complex was analysed by SDS-PAGE.

6.5 Crystallography

6.5.1 Crystallization of *Aa*BPL and Mutant *Aa*BPL R40G

Crystallization of wild-type *Aa*BPL and mutant *Aa*BPL R40G was carried out in collaboration with the Structural Biochemistry Group of Pr. Malcom Walkinshaw (University of Edinburgh). The purified *Aa*BPL protein was concentrated to 6 mg-ml^{-1} using a Vivaspin concentrator 20,000 MWCO

(Vivascience). The enzyme concentration was verified using the Bio-Rad protein assay kit and by measuring the absorbance at 280 nm (molar extinction coefficient 34,850 M^{-1} cm^{-1}). *Aa*BPL was crystallized by the hanging drop vapour diffusion method on XRL 24 well crystallization plate (Molecular Dimension Ltd) at 17 °C. Initial crystal trials were set up using the Crystal Screen Kit - Structure Screen 2 (Molecular Dimensions). The drops contained 1.3 µl of *Aa*BPL and 1.3 µl of reservoir solution. The crystals obtained in two different screening conditions (condition 26: 0.2 M ammonium sulfate, 0.1 M MES pH 6.5, 30 % w/v PEG monoethylether 5,000 and condition 28: 0.1 M HEPES pH 7.5, 20 % w/v PEG 10,000) were soaked in their respective reservoir solution supplemented with 20% sterile glycerol (v/v) before being frozen in liquid nitrogen and analysed by X-ray diffraction at the European Synchrotron Radiation Facility (ESRF) in Grenoble (France).

During the optimization set up, each crystallization drop contained 2 µl of protein at a concentration of 5 $mg.ml^{-1}$ and 1 µl of the reservoir solution. The well buffer contained 0.1 M MES, 0.2 M ammonium sulfate over a pH ranging from 6 to 7 and with a concentration of PEG 5,000 mono-ethyl ether varying from 5 % to 25% w/v. *Aa*BPL was centrifuged for 30 min at 11,000 x *g* and at 4 °C prior to crystallization. The crystals were soaked in a cryoprotectant solution (Non-drying immersion oil Type B Formula code 1248, Cargille Laboratories Inc) before being frozen in liquid nitrogen and screened by X-ray diffraction at the ESRF and at the Synchrotron Radiation Source (SRS) in Daresbury (U.K). Crystallization of the *Aa*BPL:biotin:ATP complex was achieved by adding biotin (2 mM, 1 M NaOH) and ATP (5 mM, pH 7) to the enzyme prior to centrifugation and the *Aa*BPL

complex was kept at 4 °C during the crystallization set up. The mutant *Aa*BPL R40G was also crystallized by hanging drop vapour diffusion at 17 °C at a concentration of 6 mg-ml^{-1}. The mutant crystallized in similar conditions with wild-type *Aa*BPL (0.2 M ammonium sulphate, 0.1 M HEPES pH 6.2 10 % w/v PEG 5,000 mono-ethyl ether) and co-crystals with biotin were obtained in presence of 2 mM biotin.

6.5.2 Data collection and Structure Analysis

Data collection and processing were carried out by Dr. Iain McNae. Data were collected at station BM14 ESRF Grenoble and station 10.1 SRS Daresbury. Data were processed using the programs MOSFLM and SCALA as part of three CCP4 suite of programs and the structure solved using the program PHASER using the structure from *Pyrococcus Horikoshii* (pdb-1WPY).[6] [7] Initial refinement was performed using the program REFMAC and finished using the program phenix.refine as part of the PHENIX package.[8] [9] Manual refinement was performed using the program COOT.[10]

6.6 References

1. Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-5.
2. Clarke, D. J., Coulson, J., Baillie, R. & Campopiano, D. J. (2003). Biotinylation in the hyperthermophile *Aquifex aeolicus*. *Eur J Biochem* **270**, 1277-87.

3. Chen, I., Howarth, M., Lin, W. & Ting, A. Y. (2005). Site-specific labeling of cell surface proteins with biophysical probes using biotin ligase. *Nat Methods* **2**, 99-104.
4. Sobott, F., Hernandez, H., McCammon, M. G., Tito, M. A. & Robinson, C. V. (2002). A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies. *Anal Chem* **74**, 1402-7.
5. Chernushevich, I. V. & Thomson, B. A. (2004). Collisional cooling of large ions in electrospray mass spectrometry. *Anal Chem* **76**, 1754-60.
6. (1994). Collaborative Computational Project Number 4, The CCP4 suite: programs for protein crystallography. *Acta Crystallog* **D50**, 760-763.
7. McCoy, A. J., Grosse-Kuntsleve, R. W., Storoni, L. C. & Read, R. J. (2005). Likelihood enhanced fast translation functions. *Acta Crystallog* **D61**, 458–464.
8. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallogr* **D53**, 240-255.
9. Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K. & Terwilliger, T. C. (2002). PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr* **D58**, 1948-1954.
10. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr* **D60**, 2126-2132