UNIVERSITY OF SOUTHAMPTON

FACULTY OF MEDICINE, HEALTH AND LIFE SCIENCES

School of Biological Sciences

Structural Studies of Components from the minimal Polyketide and Fatty Acid Synthases from *Streptomyces coelicolor.*

By

Katherine Pottage

Thesis for the degree of Master of Philosophy

November 2007

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF MEDICINE, HEALTH AND LIFE SCIENCES
SCHOOL OF BIOLOGICAL SCIENCES

Master of Philosophy

STRUCTURAL STUDIES OF COMPONENTS FROM THE MINIMAL POLYKETIDE AND FATTY ACID SYNTHASES FROM *STREPTOMYCES COELICOLOR.*

By Katherine Pottage

Polyketides are small biologically active molecules synthesised by large enzyme complexes, termed polyketide synthases (PKS), in response to environmental factors. PKS's operate in a similar repetitive way to fatty acid synthases (FAS). The initial stages of both processes rely on similar enzymes, some of which have been characterised structurally by NMR or x-ray crystallography. However, to date, the structures and interactions in both systems have not been compared. This work looks at the central role played by the acyl carrier proteins (ACP) of PKS and FAS systems in *Streptomyces coelicolor*, and the role played by the ketosynthase/chain length factor (KS/CLF) from the same species. The PKS ACP structure has been published previously allowing comparison with the FAS ACP structure studied here.

Samples of unlabelled, ^{15}N and $^{15}N/^{13}C$ FAS ACP from *Streptomyces coelicolor* were expressed in *E. coli* and purified using anionic exchange chromatography. A range of 2D and 3D double and triple resonance NMR spectra were recorded and analysed in order to use an automated assignment program to assign chemical shift resonances for nitrogen (^{15}N), carbon (^{13}C) and hydrogen (^{1}H) atoms within the protein. NOE data was used to generate a list of distance restraints, which were fed into a simulated annealing protocol to generate an ensemble of three dimensional structures for FAS ACP. Structural analysis has shown that, although the FAS and PKS ACPs from *S. coelicolor* share the same overall tertiary structure, subtle differences are observed. The orientation of helix I differs by ~20° between the two structures whilst the loop region between helices I and II is highly flexible in PKS ACP, but is well ordered within the FAS ACP structure.

The *gris* PKS ACP from *Streptomyces griseus* has been crystallised in conditions containing ammonium sulphate, sodium chloride and tris buffer, and two data sets collected in house at 2.5 Å and 1.8 Å. Data has been processed using the CCP4 suite of programs. Molecular replacement, using a variety of different ACP models, has so far proved unsuccessful. Whilst heavy atom soaks using gadolinium and samarium have failed to produce a heavy atom derivative.

Attempts have been made to crystallise a sample of KS/CLF complex from *Streptomyces coelicolor*, provided by the University of Bristol, using standard methods. This complex is, however, very unstable and precipitates, with a loss of activity, between minutes to hours. Therefore attempts to crystallise the complex with the act ACP are to be attempted in the hope that this will help to stabilise the complex.

Table of Contents

Table of Figures

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Table of Tables

Acknowledgements

I would like to thank my supervisors Dr Alun Coker, Dr Matt Crump, Dr Stuart Findlow and Professor Steve Wood for their help, advice and support throughout my research.

I would also like to thank the BBSRC for their funding of this research.

I am very grateful to the help and advice of everyone in the NMR and crystallography groups. I would especially like to thank Dr Jed Long, Dr Danica Buttler, Dr James Youell and Dr Graham Dunn for their friendship.

A very special thank you to Stu for all his help in proof reading and endless attempts to explain the world of NMR and mechanisms and for his numerous chats on anything and everything.

I would also like to thank Dr Chris Williams of the University of Bristol for his work on the FAS ACP.

Thank you to Lucy for her friendship and support, thanks for trying to keep me sane.

Thanks to Mr C and Mr L for putting me back together and to Gillie for continuously persevering to keep me moving.

A big thank you to all my family for their love and support.

The biggest thank you of all goes to my parent for their continuing love, understanding, care and unending support, without which this thesis would not have happened. They have been amazing during the last 6 very difficult years. Thank you both so much for everything.

And a final thank you to the inventors of seatbelts and air bags, without which I would not have been here to write this thesis.....

I dedicate this thesis to the memory of:

Matt,

Hilda,

Douggie,

Bert,

Ri

and

Pearl

Who died during the making of this thesis.

Abbreviations

Å	Angstrom, unit of measurement equivalent to 0.1nm
ACP	Acyl Carrier Protein
acpS	Acyl Carrier Protein Synthase
act	Actinorhodin
AT	Acetyl Transferase
CLF	Chain Length factor
CO_2	Carbon Dioxide
CoA	Co Enzyme A
C-terminus	Carboxyl terminus
DH	Dehydrase
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
E. coli	*Eschericia coli*
EDTA	Ethylenediaminetetraaacetic acid
ER	Enoyl Reductase
FAS	Fatty Acid Synthase
FFT	Fast Fourier Transform
FID	Free induction decay
FPLC	Fast protein liquid chromatography
HSQC	Heteronuclear single-quantum correlation
IPTG	Isopropyl-β-D-thiogalactosidase
KAS I	β-Ketoacyl acyl carrier protein synthase I
KAS II	β-Ketoacyl acyl carrier protein synthase II
KAS III	β-Ketoacyl acyl carrier protein synthase III
KAS	β-Ketoacyl acyl carrier protein synthase
KR	Ketoreductase
LB	Lauria-Bertani Broth
MCAT/MT	Malonyl CoA:*holo*-Acyl carrier protein Transacylases
mPKS	Minimal Polyketide Synthase
N-terminus	Amino-terminus
NMR	Nuclear Magnetic Resonance
NOESY	Nuclear Overhauser effect spectrometery
ORF	Open Reading Frame
otc	Oxytetracycline from *Streptomyces rimosus*
PKS	Polyketide Synthase
ppm	Parts per million
RF	Radio frequency
Rpm	Revolutions per minute
S. coelicolor	*Streptomyces coelicolor*
SDS PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
S. coelicolor	*Streptomyces coelicolor*
TMS	Tetramethylsilane
TOCSY	Total correlation spectroscopy
UV	Ultraviolet
2D	Two dimensional
3D	Three dimensional
h	Planks constant (Joule-seconds)

$\hbar$	Planks constant (h) divided by 2π
Hz	Frequency (cycles per second)
I	Quantum spin number
J	Internuclear coupling constant (Hz)
λ	Wavelength (m)
MHz	MegaHertz (10^6 Hertz)
$\boldsymbol{I}$	Angular momentum
γ	Gyromagnetic ratio
δ	Chemical shift
μ	Magnetic moment

Chapter 1

Introduction

1 Introduction

Polyketides are a diverse class of secondary metabolites produced by a variety of natural sources including plants, fungi, marine organisms and some bacteria in response to a range of environmental factors (*1*). They are of interest, both scientifically and medically, as they encompass many biologically active compounds including antibiotics (e.g. oxytetracycline (*2-4*)), anti cancer drugs (e.g. dynemycin A (*5*)), cholesterol lowering agents (e.g. lovastatin (*6*)), antiparisitics (e.g. avermectin (*7*)), antifungal agents (e.g. griseofulvin (*8*)) and immunosuppressants (e.g. rapamycin (*9-12*)). The polyketide product lycyclic ether maitotoxin (*13*) (Figure 1-1), at 3422Da, is the largest known polyketide secondary metabolite (*12*). Polyketide products possess a diverse range of structures, examples of which can be seen in Figure 1-2, and are composed of either acyl chains or aromatics or a combination of the two.

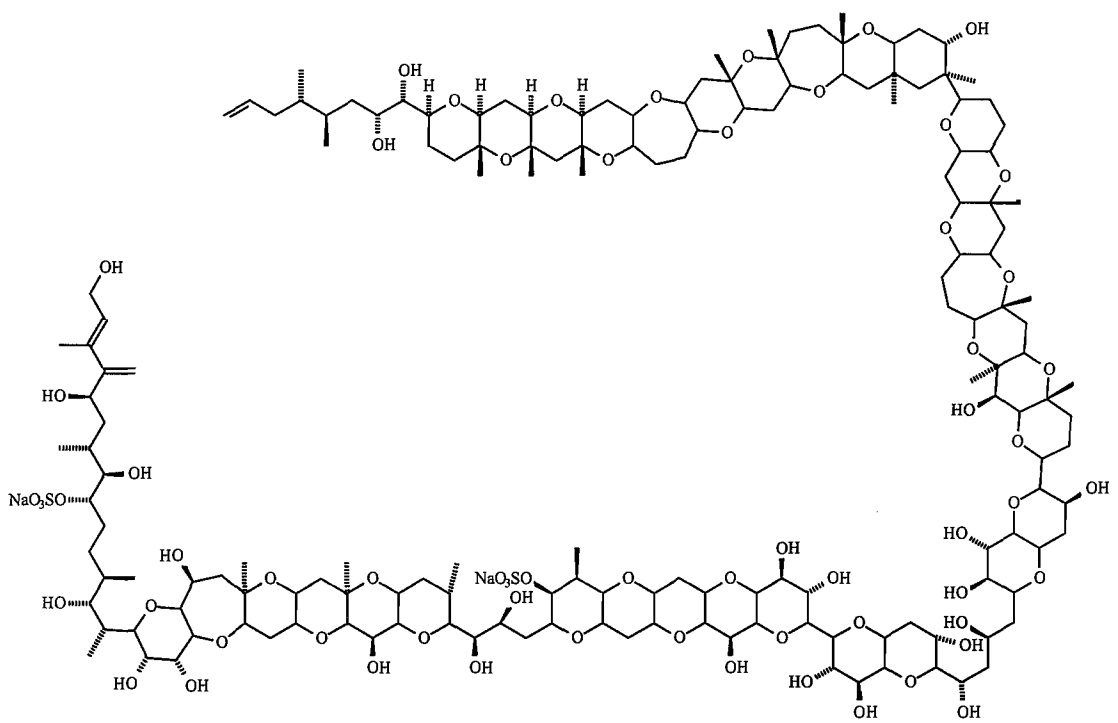


Figure 1-1 The chemical structure of lycyclic ether maitotoxin, a polyketide and the largest secondary metabolite known to man (*13*).

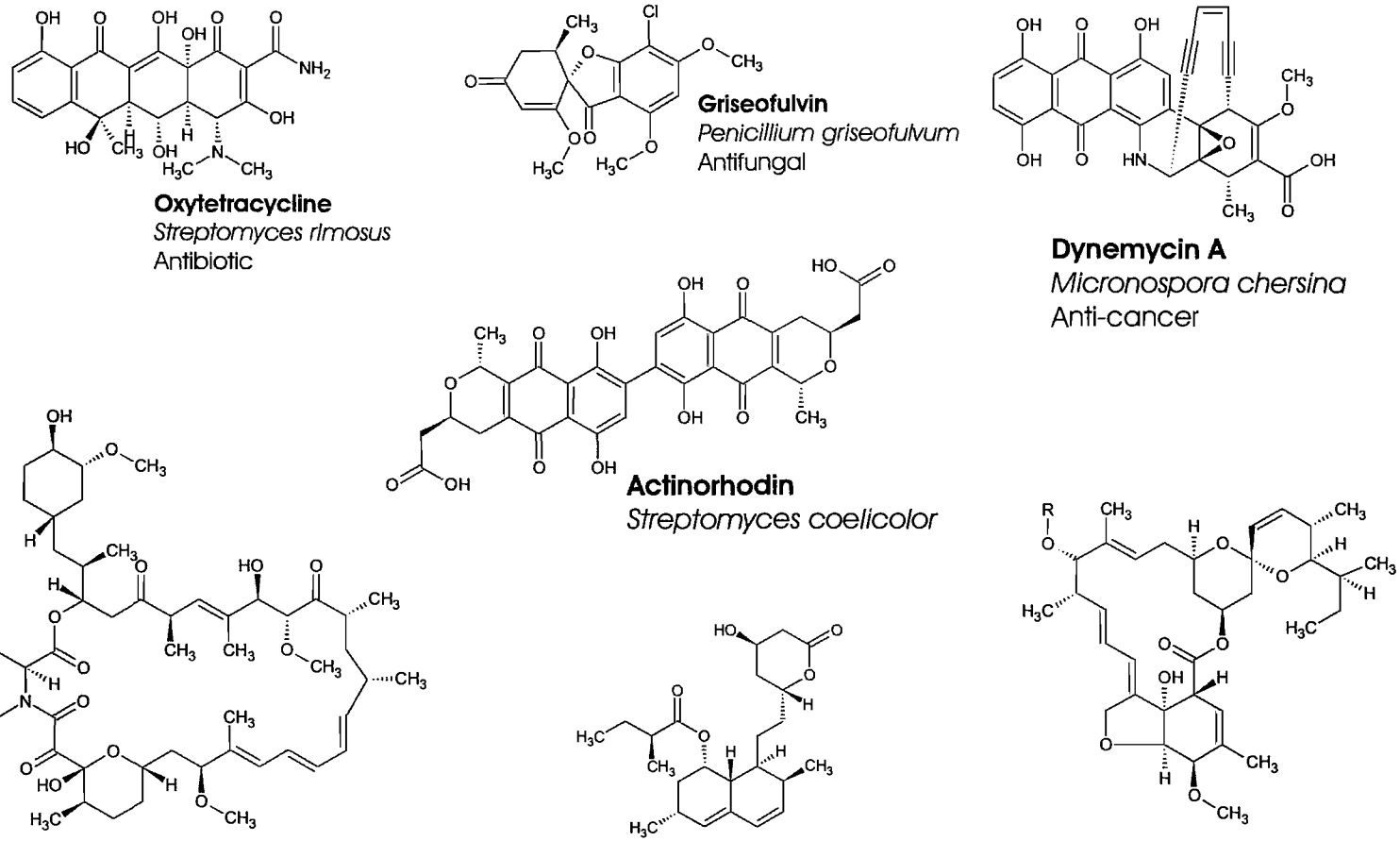


Figure 1-2 Structure diversity of a seven currently identified polyketides, the organisms which synthesis them and their pharmacological uses.

The biosynthetic mechanisms of polyketides and fatty acids are closely related. The majority of polyketide and fatty acid compounds are derived from acetate starter units and malonate extender units (Figure 1-3), although more diverse starter units e.g. propibenzone, cinnamate and amino acids, and more diverse extender units e.g. methylmalonate or ethylmalonate may be utilised. The starter and extender units undergo successive, decarboxylative condensations which are catalysed by multi-enzyme complexes which have a clear structural, functional and evolutional relationship, the polyketide synthases (PKS) and fatty acid synthases (FAS) (*1, 12, 14*).

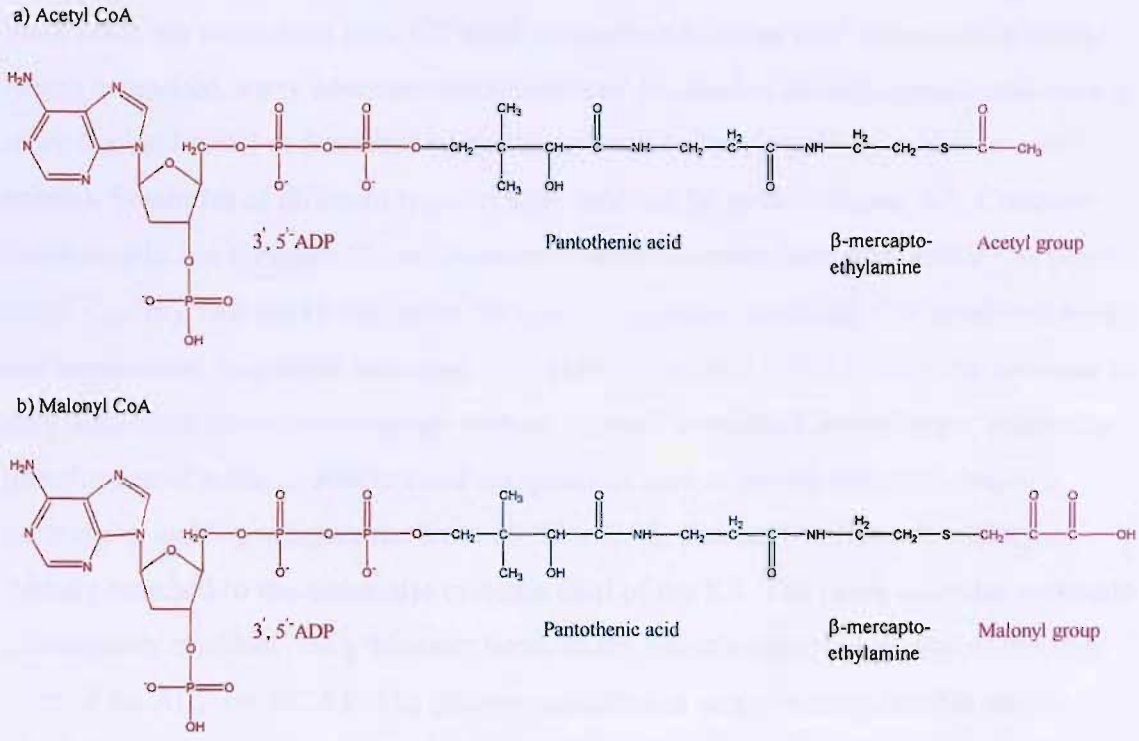


Figure 1-3 Structure of the usual CoA thioesters, acetyl CoA, and, malonyl CoA, Used for polyketide synthesis and fatty acid synthesis. These are converted into either acetyl or malonyl ACP, which are utilised by FAS's and PKS's.

Although individual enzymes are often homologous, both FASs and PKSs show very different associations of there component enzymes. Animal FASs and fungal PKSs and FASs, known as Type I systems, are composed of large multifunctional enzymes with distinct domains controlling the various enzymatic reactions, e.g. 6-methylsalicilic acid synthase (6-MSA) from *Penicillium patulum.* In prokaryotic systems however, the Type II PKS e.g. actinorhodin PKS from *Streptomyces coelicolor* (*act)* and FAS synthases are composed of individual enzymes with discrete functions. Type III systems are only found in polyketide synthases, and are found within the plant kingdom as iterative homodimers.

1.1 Fatty Acid Biosynthesis

Fatty acids are primary metabolites which contain a long hydrocarbon chain and a terminal carboxylate group. Fatty acid synthesis takes place within the cytoplasm where the growing fatty acid is synthesised by the FAS (Figure 1-4) containing the following enzymatic components; acyl carrier protein (ACP), malonyl CoA: acyl transferase (MCAT/MAT), holo-acyl carrier protein synthase (acpS), ketosynthase (KS), ketoreductase (KR), dehydratase (DH), enoyl reductase (ER) and thioesterase (TE). Fatty acids are assembled from C2 units, condensed together until the required chain length is reached. Fatty acids are either saturated (no double bonds), unsaturated (one or more double bonds) or branched (with the fatty acid chain branching at one or more points). Examples of different types of fatty acid can be seen in Figure 1-5. Common chain lengths are between 12 and 20 carbons with saturated fatty acids being C_{12} (lauric acid), C_{14} (myristic acid), C_{16} (palmitic acid), C_{18} (stearic acid) and C_{20} (arachidic acid) and unsaturated fatty acids including $C_{16:1}$ (palmitoleic) and $C_{18:1}$ (oleic). An increase in fatty acid chain length corresponds with an increase in melting temperature, whilst the introduction of a single double bond e.g. palmitic acid to palmitoleic acid causes a decrease in melting temperature from 63 °C to 0 °C. The acyl starter unit moiety is initially attached to the active site cysteine thiol of the KS. The chain extender malonate is covalently attached, via a thioester bond, to the phosphopantetheine arm of the *holo* form of the ACP by MCAT. The phosphopantetheine acts as a long flexible arm to which the growing acyl chain is attached, and which delivers this chain to the various enzymes involved in fatty acid synthesis. The starter unit undergoes a condensation reaction, catalysed by the KS, with an ACP bound malonyl unit. The resulting β-ketoester is then successively reduced (by KR), dehydrated (DH) and finally reduced again (by ER) to produce a saturated acyl chain, two methylene units longer than the starter unit. This process continues in a repetitive fashion until the required chain length is reached (usually C_{16} or C_{18}), when the acyl chain is released from the ACP by the TE (bacteria and animals) or transferred to CoA-SH (yeast) (*15*). An example of this process is shown below in Figure 1-4

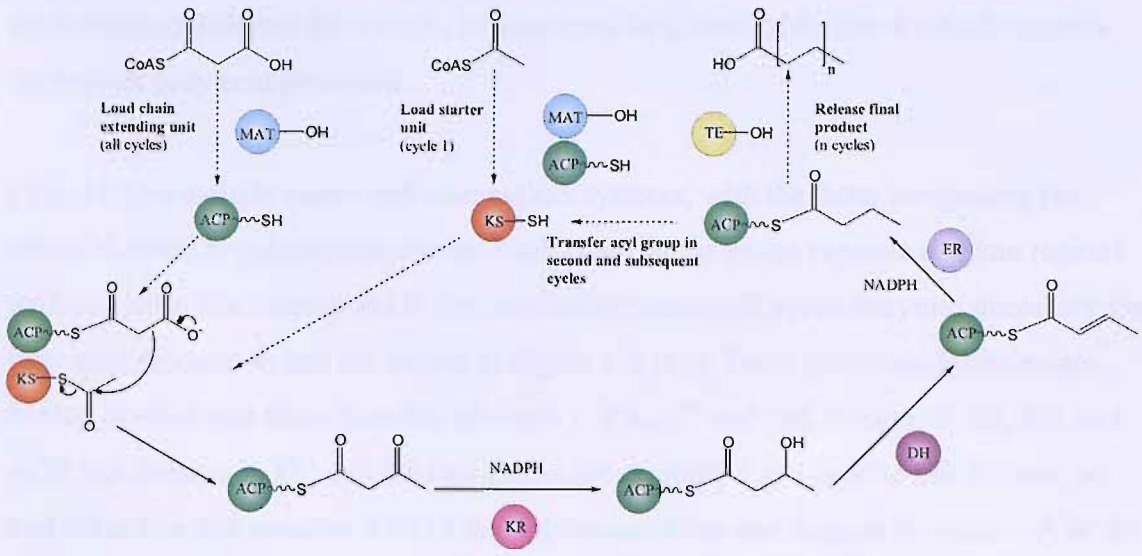


Figure 1-4 Overview of fatty acid synthesis illustrating the roles of different enzymes within the pathway. The acetyl starter unit undergoes a condensation reaction, catalysed by the KS, with an ACP bound malonyl unit. The resulting β-ketoester is then successively reduced (by KR), dehydrated (DH) and finally reduced again (by ER) to produce a saturated acyl chain, two methylene units longer than the starter unit. This process continues in a repetitive fashion until the required chain length is reached, when the acyl chain is released from the ACP by the TE (bacteria and animals) or transferred to CoA-SH (yeast). (DH: Dehydratase, KS: ketosynthase, MAT: malonyl-acyl: ACP transferase, ER: enoyl reductase, TE: thioesterase, KR: ketoreductase, ACP: acyl carrier protein). Adapted from Staunton and Weissman (*1*).

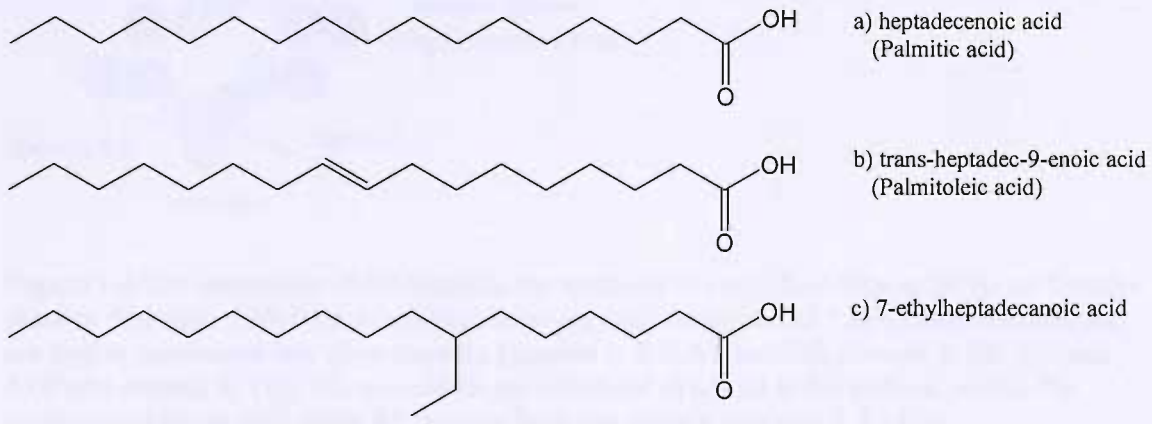


Figure 1-5 Examples of different types of fatty acids showing: a) A saturated fatty acid, palmitic acid, b) An unsaturated fatty acid, palmitoleic acid and c) A branched fatty acid, 7-ethylheptadecanoic acid.

The length of these fatty acids and the degree of unsaturation is dependent upon the temperature at which the bacteria are growing, as membrane fluidity is essential for cell survival. Therefore, at lower temperatures, shorter, unsaturated fatty acids are produced, whereas at higher temperatures longer more saturated moieties are produced. In *E. coli* three KAS are found: KAS I, KAS II and KAS III, each of which is responsible for the formation fatty acid chains of different length and saturation, utilizing different starter units. It is possible that these vary in different intracellular environments to control the

production of different fatty acids, or there may be a form of feedback which controls the type of fatty acid produced.

Type I FASs include yeasts and mammalian systems, with the latter comprising two identical 500kDa polypeptide chains. Each chain forms seven separate enzyme regions with activities that correspond to the previously mentioned seven enzymes necessary for fatty acid production and are shown in Figure 1-6 (*16*). These polypeptide chains are further divided into three domains (domain 1: KS, AT and DH, domain 2: ER, KR and ACP and domain 3: TE) and the two chains are orientated in a head to tail fashion, so that the active site cysteine thiol of the KS domain from one domain is within 2 Å of the phosphopantetheine thiol of the ACP from the other dimer. Mammalian FAS functions iteratively to generate a long chain fatty acid. The fatty acid chain undergoes a number of catalytic cycles in the multidomain enzyme until the correct chain length is reached when the chain is off-loaded by the terminal thioesterase.

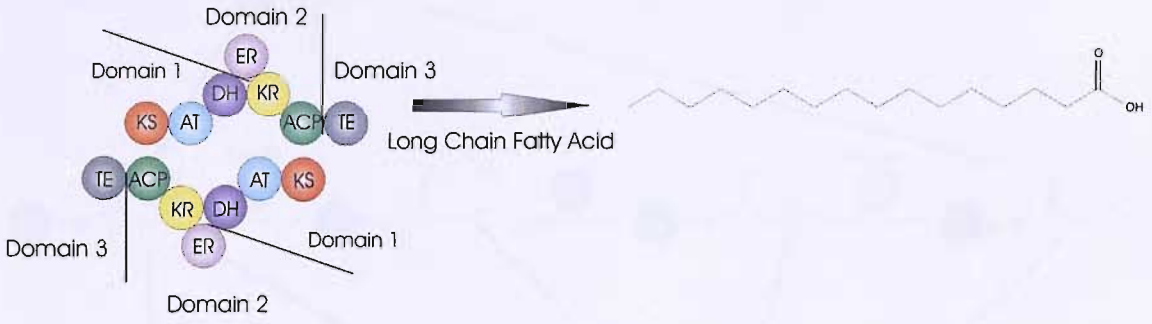


Figure 1-6 The mammalian FAS catalyses the synthesis of long chain fatty acids via an iterative process. The two ~500kDa polypeptide chains are each comprised of 7 enzymatic domains and are further subdivided into three domains (domain 1: KS, AT and DH, domain 2: ER, KR and ACP and domain 3: TE). The two chains are orientated in a head to tail fashion, so that the active site cysteine thiol of the KS domain from one domain is within 2 Å of the phosphopantetheine thiol of the ACP from the other dimer.

1.2 Polyketide Biosynthesis and Polyketide synthases

Polyketide synthases (PKS), like fatty acid synthases, catalyse a series of repetitive decarboxylative condensation reactions, commonly between KS bound acyl thioesters and malonyl-ACP. After each round of condensation some or all of the following reactions may take place: ketoreduction, dehydration and enoyl reduction, depending

upon the final polyketide product (*12*). The enzymes in each individual PKS are encoded for by various open reading frames, comprising some or all of the enzymes associated with polyketide biosynthesis with the makeup and selection of these genes being responsible for the specific polyketide produced (*12*). An overview of polyketide production can be seen below in Figure 1-7. In order to produce a 'classical' polyketide it is thought that the growing acyl chain undergoes n rounds of condensation reactions to produce a polyketide product of 2n carbon atoms, catalysed by the ketosynthase, to produce a classical polyketide. In the 'progressive' pathway a single/series of enzymatic modifications takes place to introduce modifications into the polyketide chain following some/all of the condensation reactions.

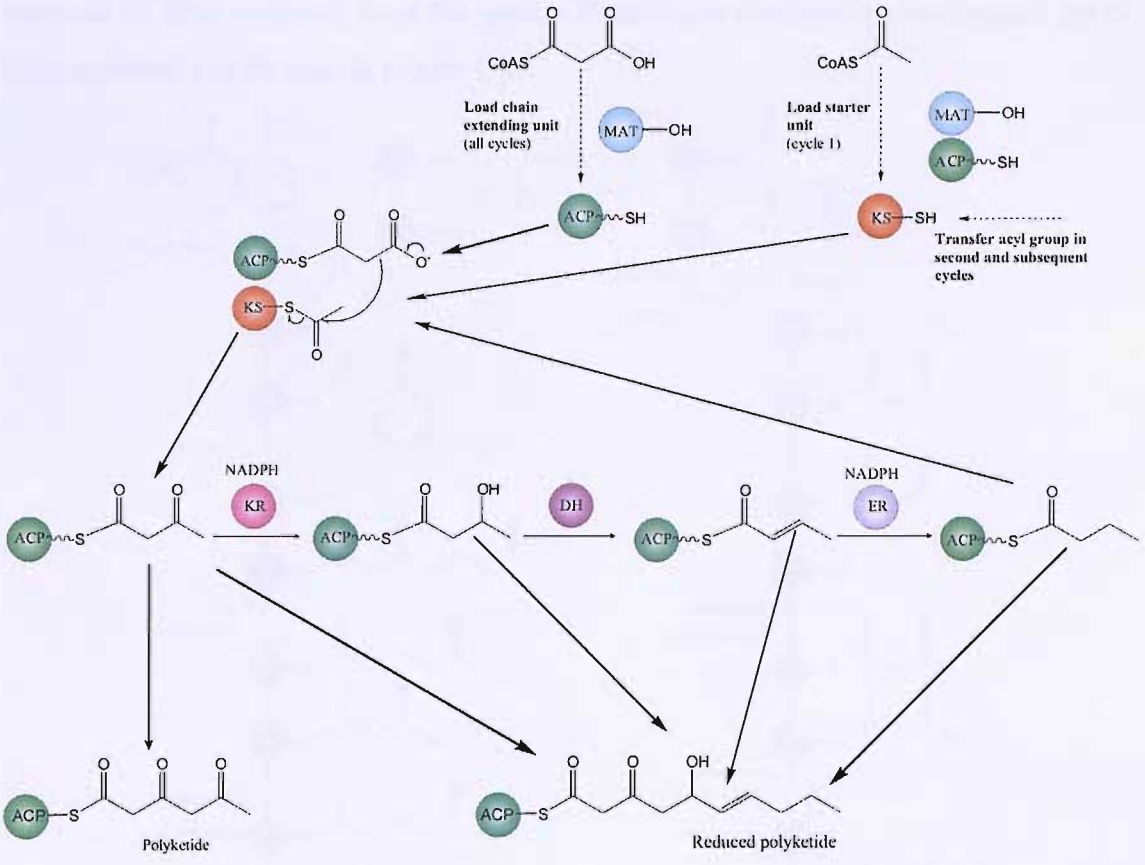


Figure 1-7 The polyketide biosynthetic cycle. MAT: malonyl transferase; ACP: acyl carrier protein; KS: ketosynthase; KR: ketoreductase; DH: dehydratase; ER: enoyl reductase. A series of repetitive decarboxylative condensation reactions take place, commonly between KS bound acyl thioesters and malonyl-ACP. After each round of condensation some or all of the following reactions may take place: ketoreduction, dehydration and enoyl reduction, depending upon the final polyketide product. To produce a 'classical' polyketide there is no modification to the polyketide chain after each round of condensation, whilst during the production of a 'reduced' polyketide various reactions may take place between each round of condensation.

1.3 Polyketide Synthases

Polyketide synthases fall into three categories, Type I, II and III, depending upon their origins, their tertiary/quaternary structures, and their mechanism of action.

1.3.1 Type I Synthase Systems

Type I polyketide synthases are large, multi-functional enzymes with independent catalytic domains which work iteratively. Each individual domain completes just one round of elongation and/or dehydration per extender unit; however the domains are joined in a covalent manner by linker regions. It has been shown that fungal systems fall into this category, with the best studied Type I PKS being 6-methylsalicylic acid synthase (*6-MSA* synthase) from the species *Penicillium patulum,* the mechanism for *6-MSA* synthase can be seen in Figure 1-8.

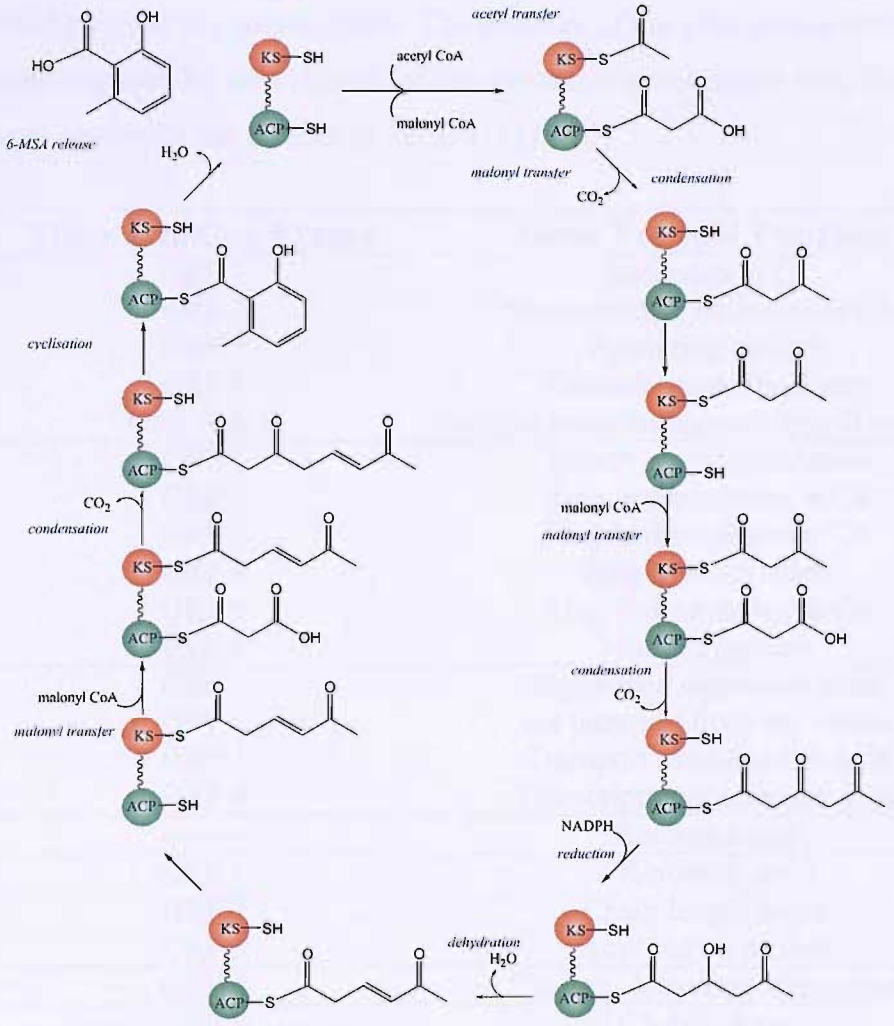


Figure 1-8 Mechanism of *6-MSA* synthesis in *Penicillium patulum.* The multi-functional *6-MSA* synthase converts one molecule of acetyl CoA and 3 molecules of malonyl CoA in order to synthesise one molecule of *6-MSA*. The KS, Mat, DH, KR and ACP domains in the *6*-MSA PKS catalyse three rounds of chain extension, whilst carrying out different levels of reduction during each cycle. After the third cycle the chain undergoes cyclisation, dehydration and enolisation to give the aromatic end product, *6*-MSA. Adapted from Staunton and Weissman (*1*)

This protein has been purified as a homodimer with a subunit mass of 180kDa (*17*) and has been shown to use malonyl-ACP as its starter unit. It has also been shown that *6-MSA* synthase will accept acetyl CoA as well as acetyl-ACP. Type I polyketide synthases are encoded for by a single open reading frame (ORF) within the genomic DNA (*17*).

1.3.2 Type II Synthase Systems

Type II polyketide synthases are found within *Streptomyces sp.* and other filamentous bacteria. Unlike the Type I synthases, the Type II synthases consist of individual enzymes, each with their own catalytic centre. The individual enzymes are presumed to form a multi-enzyme complex, in a non-covalent manner, although thus far this has not been established. The individual enzymes are encoded for by a series of different ORFs that are found grouped in a single cluster. The structure of the gene cluster encoding those proteins required for the synthesis of the polyketide actinorhodin *(act)* from *Streptomyces coelicolor* can be seen in Table 1-1 (*18-25*).

Gene	Open Reading Frame	Gene Product Function
act VI	ORF 1	Reduction at C3
	ORF 2	Stereospecific reduction at C15
	ORF 3	Pyran ring closure
	ORF 4	Controls stereochemistry
	ORF A & B	Stabilise multi-component type II complex
act VA	ORF 1	Export across membrane
	ORF 2	Ring hydroxylation at C8
	ORF 3	Ring hydroxylation at C6
	ORF 4	Ring hydroxylation
	ORF 5	Ring hydroxylation at C6
	ORF 6	Monooxygenase
act II	ORF 1	Resistance repression gene
	ORF 2	*act* transport from mycelium
	ORF 3	Transport associated protein
	ORF 4	Transcriptional activator gene
act III	--------	Ketoreductase
act I	ORF 1	Ketosynthase
	ORF 2	Chain length factor
	ORF 3	Acyl carrier protein
act VII	ORF 4	Cyclase, dehydrase, aromatase
act IV	ORF 5	C5 dehydrase
act VB	ORF 6	Dimerase?

Table 1-1 The *act* genes, open reading frames and their associated protein products from *Streptomyces coelicolor.*

It is interesting to note that not all PKS contain all of the enzymes associated with the synthases. For example the *Streptomyces coelicolor act* gene cluster lacks both ACPS and MCAT encoding genes. Although the *act* ACP has been shown to self malonate it is thought that the fas ACPS from *Streptomyces coelicolor* is able to cross talk to convert the *apo act* ACP into the *holo* form (*12*). A comparison of the known Type II PKS gene clusters is shown in Figure 1-9.

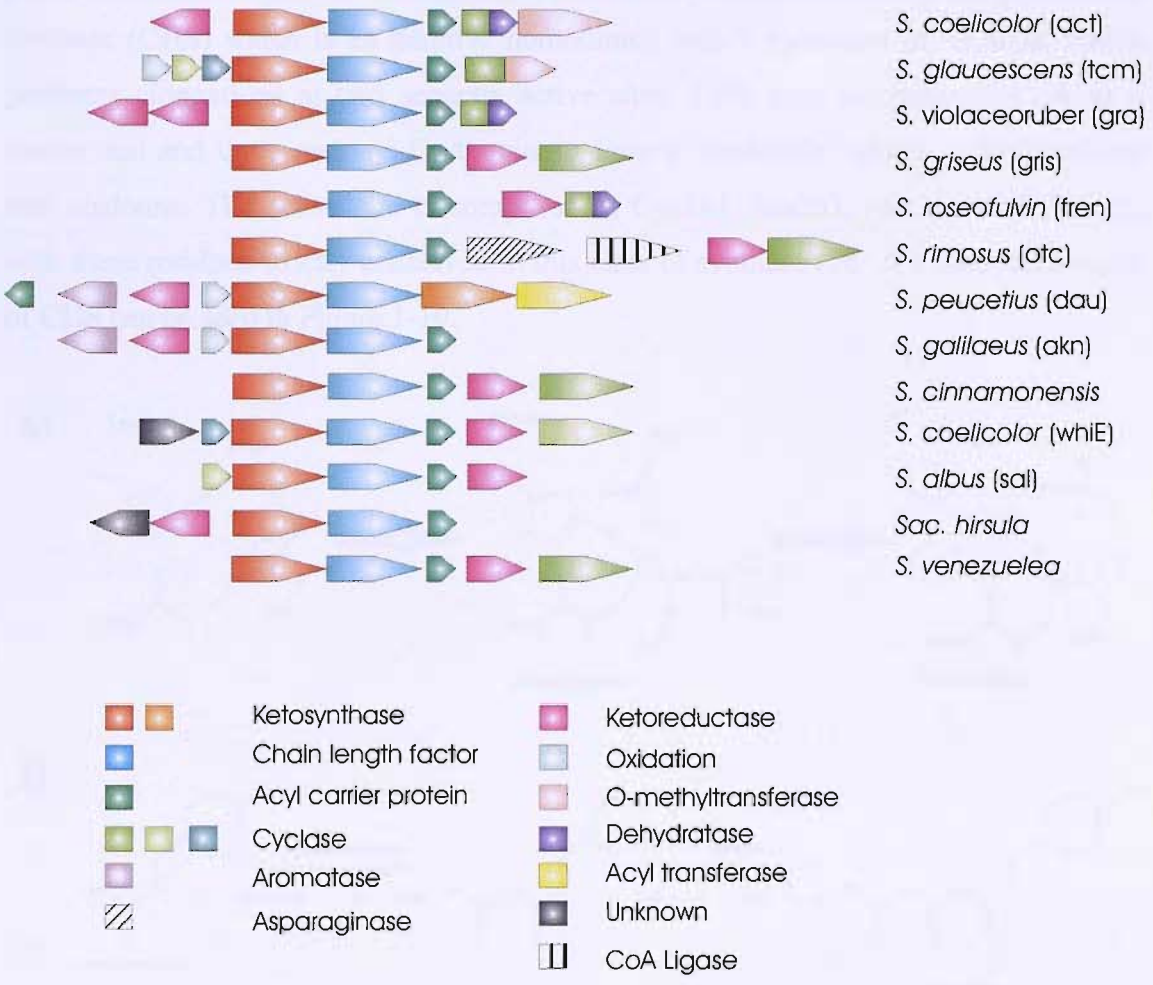


Figure 1-9 Comparison of the organisation of known gene clusters encoding Type II PKSs. The size of each wedge represents the relative size of each gene and are coloured relative to their function. As can be seen, only the gene cluster from *S. peucetius* encodes for an acyl transferase, in line with experimental results that ACP can undergo self malonylation, and that therefore MAT is not required and is therefore not present in the gene clusters. It is thought that the fas MAT from each organism may cross-talk and be responsible for acyl transferase activity in the PKS.

1.3.3 Type III Systems

Type III PKS are found within the plant kingdom and are structurally dissimilar to the Type I and II PKS in that they consist of a homodimer of relatively small molecular weight. They produce a variety of important biological molecules including flavonoids (flower colour pigments) and phytoalexins for defence against pathogens. Type III PKS use coenzyme-A linked thioester substrates rather than ACP bound 4' phosphopantetheine thioesters. The most commonly studied Type III PKS is chalcone synthase (CHS) which is an iterative homodimer, with a monomer of ~42kDa, which performs elongations at two separate active sites. CHS uses *p*-coumaroyl-CoA as a starter unit and three malonyl-CoA units to form a tetraketide, which is then cyclised into chalcone. The active site is composed of Cys164, His303, Asn336 and Phe215, with these residues strictly conserved in this class of synthase (*26, 27*). The mechanism of CHS can be seen in Figure 1-10.

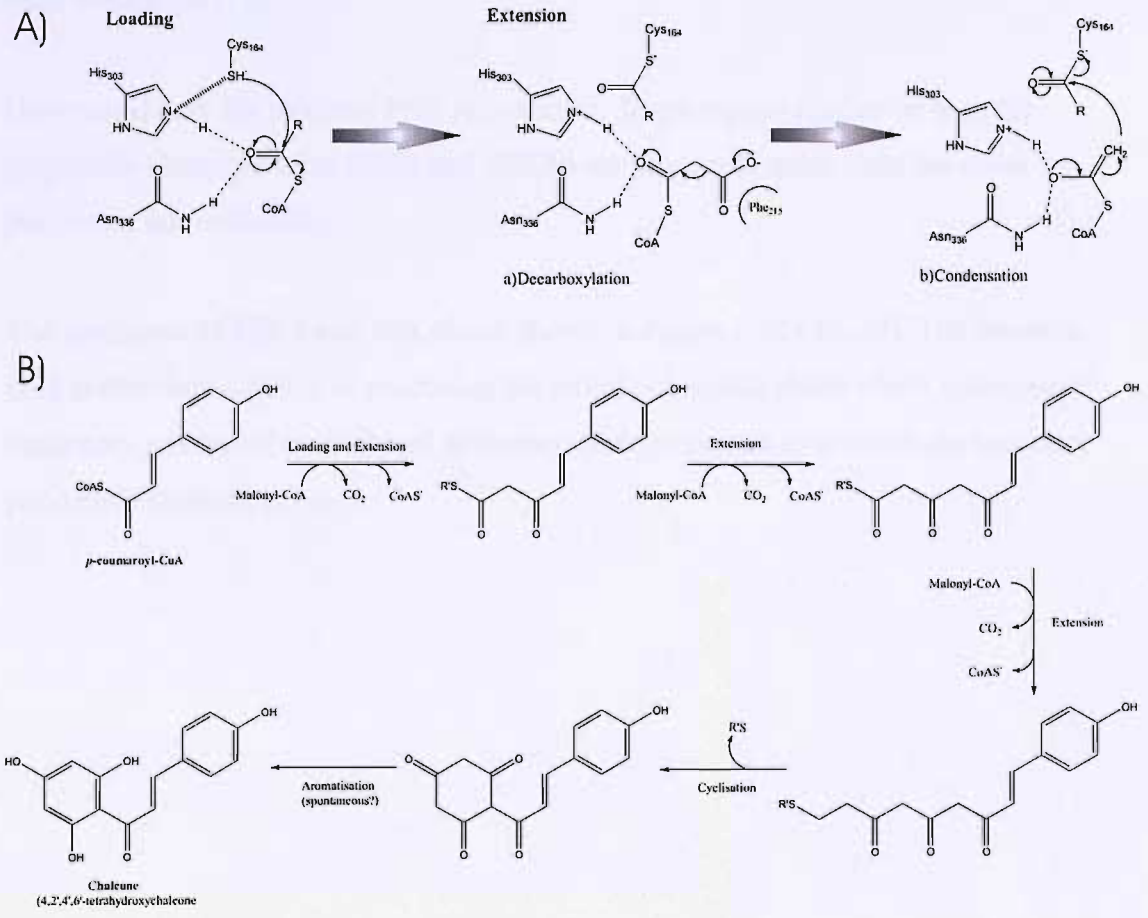


Figure 1-10-Proposed reaction mechanism for chalcone synthase. A) The substrate loading and extension stages catalysed by the CHS homodimer (R is a p-coumaroyl moiety in the loading reaction, a coumaroyl-acetyl group in the decarboxylation reaction and a maroyl-diacetyl group in the elongation reaction). B) Chalcone synthase reaction intermediates. (R' is either CoA or Enzyme Cysteine). (Adapted from Austin *et al.* (*28*))

1.4 Enzyme components of Polyketide and Fatty Acid Synthases

Although ACPS, MCAT, ACP, KS, CLF, KR, ER and DH are generally thought of as being required for initial polyketide and fatty acid synthesis, within individual species e.g. *Streptomyces coelicolor*, it has been shown that a minimal PKS, consisting of a ketosynthase, chain length factor and acyl carrier protein, is all that is required for basic chain assembly (*12*). Figure 1-11 shows the proposed mechanism for the production of actinorhodin (*act)*, an aromatic antimicrobial polyketide, in *Streptomyces coelicolor* (*29*). The *act* minimal PKS is responsible for the decarboxylation of a malonate molecule to form the acetate starter unit, which undergoes seven successive rounds of decarboxylative condensation reactions with further malonate units. Other enzymes such as ketoreductases, aromatases and cyclases are responsible for chemically modifying the growing polyketide chain to produce the required functionality of the final metabolite (*12*).

However if only the minimal PKS is present in *Streptomyces coelicolor* then the polyketide shunt products SEK4 and SEK4b are produced, rather than the usual polyketide actinorhodin.

The structures of SEK4 and SEK4b are shown in Figure 1-12 (*30, 31*). The minimal PKS is therefore capable of producing the initial polyketide chain which subsequently undergoes a series of uncatalysed spontaneous ring closures to produce the incorrect polyketide product. (*29-31*).

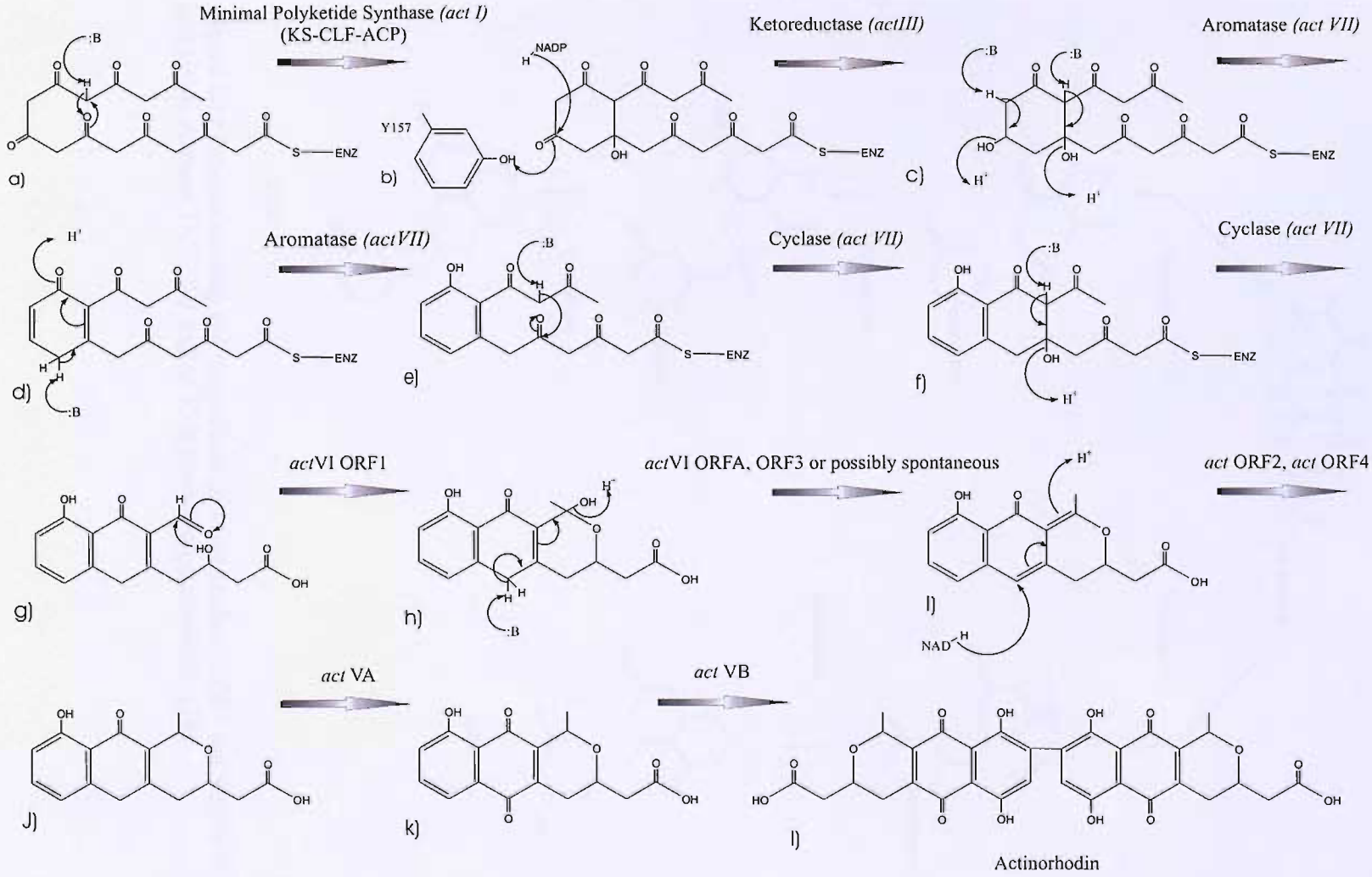


Figure 1-11 Proposed scheme for the biosynthesis of the polyketide Actinorhodin (*act*) in S*treptomyces coelicolor* (*29*). Steps catalysed by enzymes form in the *act* gene clusters are shown, corresponding to those detailed in Table 1-1.

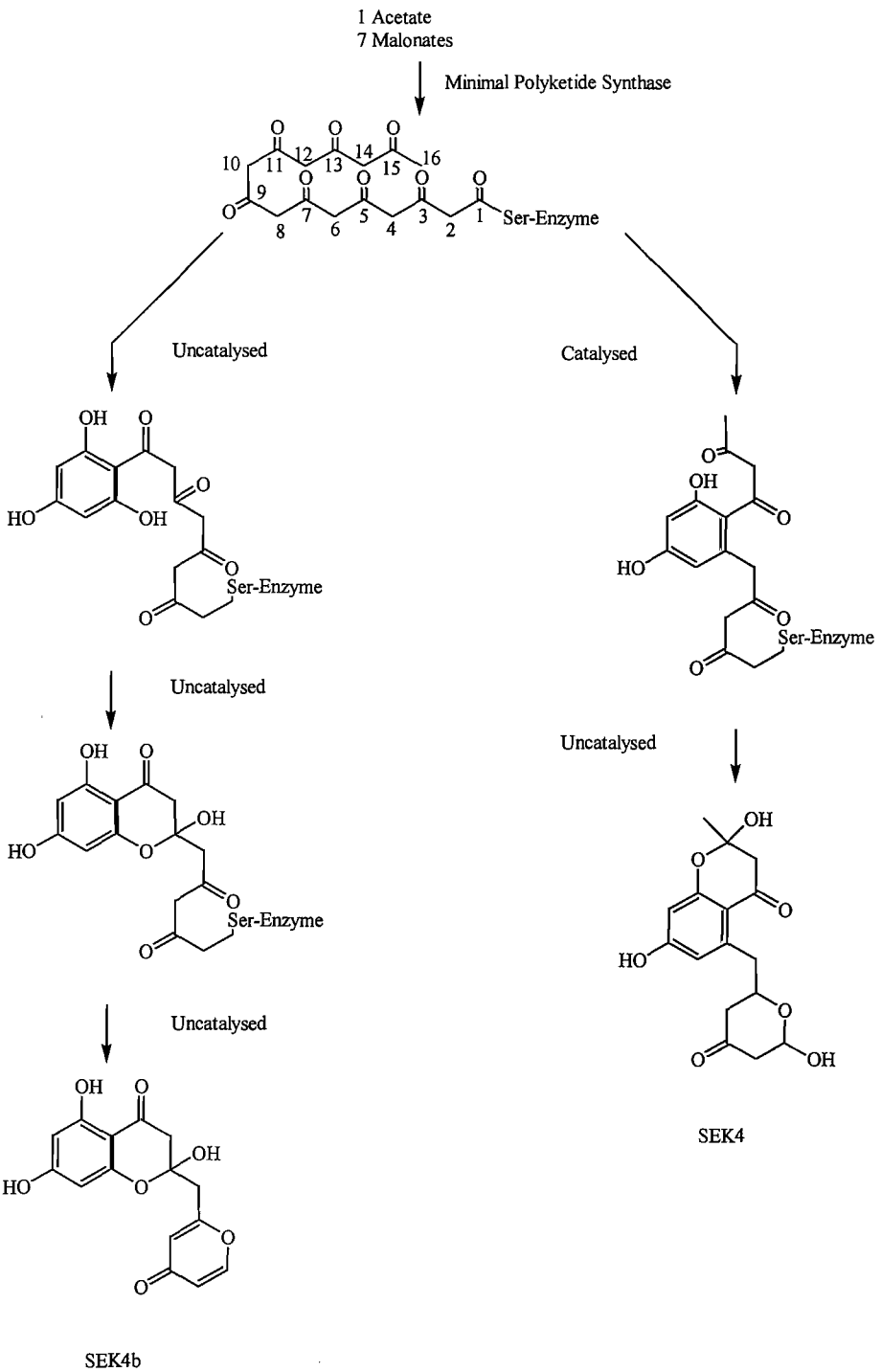


Figure 1-12 Scheme showing the biosynthesis of shunt products SEK4 and SEK4b by minimal polyketide synthase (KS, CLF and ACP) in *Streptomyces coelicolor* (*29*).

1.5 Acyl Carrier Proteins

Acyl carrier proteins (ACP) are essential for the activity of both PKSs and FASs. They are produced in the inactive *apo-* form but converted immediately into the active *holo-* form, resulting in virtually no *apo* ACP being present within the cell at any one time. The *holo-* form has a covalently attached 4′-phosphopantetheine prosthetic group which is derived from Coenzyme A (Figure 1-13). This 4'-phosphopantetheine prosthetic group is covalently attached via a phosphodiester bond to the hydroxyl group of a conserved serine residue on the ACP. The conversion of the *apo-* to the *holo-* form is essential for the ACP functionality, and is an event catalysed by the magnesium dependent enzyme, *holo*-ACP synthase (acpS) (*32-34*). *Holo*-ACP then mediates the transfer of acyl intermediates via covalent attachment to the thiol group of the 4′-phosphopantetheine moiety (*32-38*). In FAS systems a catalytic serine on the MCAT attacks a malonyl-CoA molecule at the thioester bond, before this malonyl-MCAT ester carbonyl is itself attacked by the thiol group of the phosphopantetheine, yielding malonyl-ACP. A histidine residue in the active site of the MCAT is also essential for the enzymes activity. The serine of the ACP that bears the 4'phosphopantetheine arm is 100 % conserved and forms part of a DSL/ESL motif (located at the base of helix II in all solved structures), and is shown in the sequence alignment in Figure 1-14. Work by Keating *et al.* (*39*) has shown that mutation of this conserved serine in *E. coli* FAS ACP resulted in no phosphopantetheinylation *in vitro*, whilst work by Flugel *et al.* (*36*) confirmed this effect *in vivo*.

A malonyl group is then transferred to the free thiol group of the *holo* ACP bound 4'phosphopantetheine either by MCAT (in FAS) (*40*) or by self malonylation or MCAT in polyketide biosynthesis (*30, 31, 41, 42*). Work by Arthur *et al.* (*30, 31*) showed that self-malonylation of type II ACPs occurred *in vitro*, contradicting Dreier *et al.* (*43*) who attributed this phenomenon to low level MCAT contamination of the ACP preparation. Arthur *et al.* synthesised the *apo-act* ACP using solid state peptide methods and refolded the ACP using the GroEL/ES chaperone system from *E. coli*, with negative controls confirming that MCAT was not present in the GroEL/ES preparation. Arthur *et al.* (*31*) demonstrated that *act* ACP can catalyse the transfer of malonate to type II *S. coelicolor* FAS ACP and other type II polyketide ACPs *in vitro*, however the FAS ACP is unable to transfer to the *act* ACP. The proposed mechanism for self-malonylation of *act* ACP in *S. coelicolor* is shown in Figure 1-15 (*31*).

Work by Keating *et al.* (*39*) has shown that over expression of the *apo*-FAS ACP in *E. coli* results in a decrease in growth of the cells when compared to that seen with wild type expression of the ACP, but with no cell lysis. A pool of *apo*-ACP was seen within the cells due to the saturation of the post-transcriptional modification mechanism within the cell. In wild type cells no pool of *apo*-ACP is seen due to the immediate conversion of the *apo* form of ACP to the *holo* form by the acpS suggesting that the pool of *apo*-ACP may be the cause of the growth inhibition. (Jackowski *et al.* (*44*))

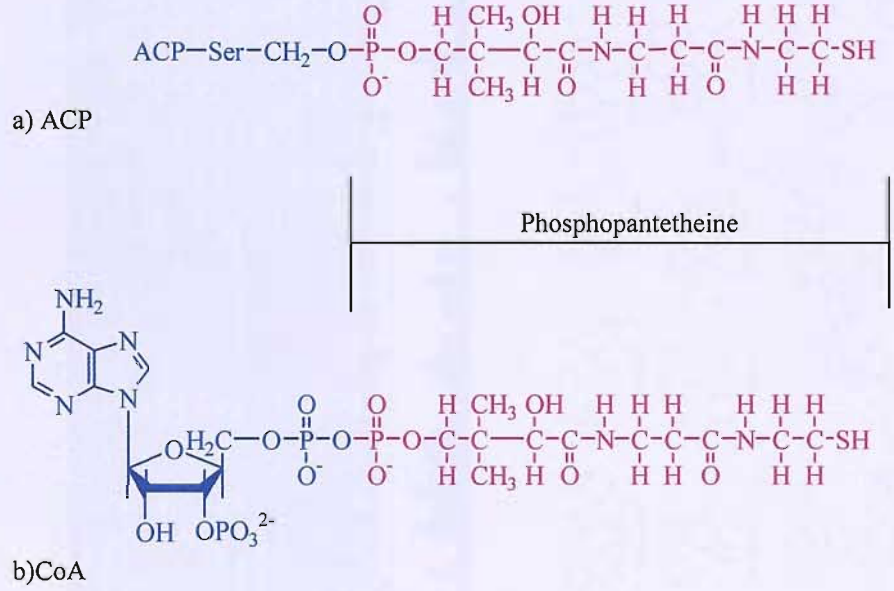


Figure 1-13 Phosphopantetheine moiety, which is vital for ACP and CoA action. The phosphopantetheine moiety is transferred to the conserved serine of the *apo*-ACP by ACPS, a magnesium dependent enzyme, from the CoA source. Malonate/acetate group is added by MCAT/self malonylation to the thiol group of the phosphopantetheine.

STRCM Monensin producer *Streptomyces cinnamonensis*
STRRM Oxytetracycline producer *Streptomyces rimosus*
STRVN Granaticin producer *Streptomyces violaceoruber*
STRCO Actinorhodin producer *Streptomyces coelicolor A3(2)*
STRGA Tetracenomycin C *Streptomyces glaucescens*
STRRS Frenolicin producer *Streptomyces roesofulvus*
STRGS Griseusin producer *Streptomyces griseus*
STRCY Landomycin producer *Streptomyces cyanogenus*
STRCN Curamycin producer *Streptomyces curacoi*
STRFR Urdamycin A producer *Streptomyces fradiae*
STRVN Jadomycin B producer *Streptomyces venezuelae*
STRRB Rubromycin B producer *Streptomyces sp. PGA64*
STRAT Oviedomycin producer *Streptomyces antibioticus*
STRAU Auricin producer *Streptomyces aureofaciens*
STRMT Enterocin producer *Streptomyces maritimus*
STRMH Mithramycin producer *Streptomyces argillaceus*

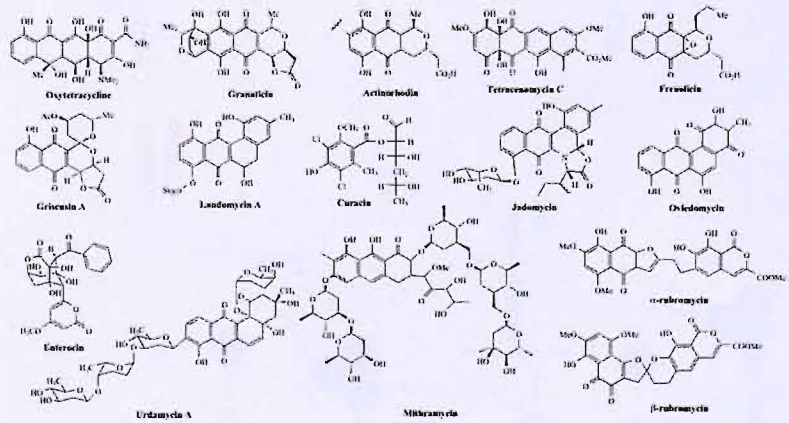


Figure 1-14 Sequence alignments of polyketide Acyl Carrier Proteins from *Streptomyces sp.* showing the conserved DSL/ESL motif to which the phosphopantetheine arm is attached in the *holo-* form of the enzyme. Also shown are the chemical structures of some of the polyketide products. Sequence alignment executed by clustalX (*45*).

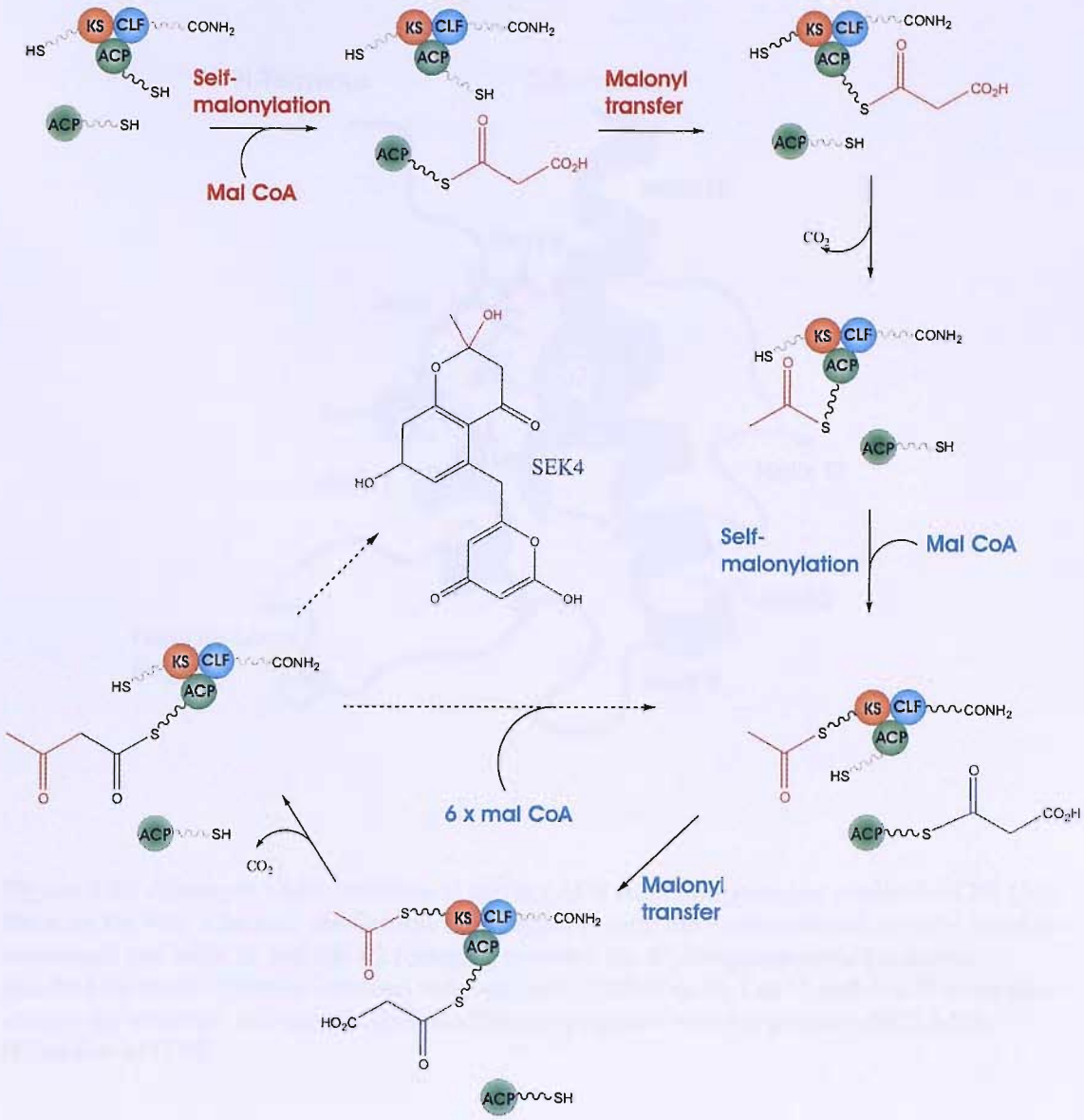


Figure 1-15 Proposed scheme for the Type II minimal polyketide synthase, synthesising SEK4 in the absence of MCAT. The minimal PKS, consisting of ketosynthase (KS-red), acyl carrier protein (ACP-green) and chain length factor (CLF-blue), interacts with free malonyl ACP. The malonate required for SEK4 production is always provided by free malonyl ACP. Adapted from (*31*).

The first three dimensional structure of an ACP, *holo*-FAS ACP from *E. coli*, was solved by Prestegard *et al.* (*46-56*) in 1986 using ^{1}H NMR spectroscopy. The first polyketide *apo*-ACP structure, of actinorhodin (*act*) ACP from *Streptomyces coelicolor,* was solved by Crump *et al.* in 1996 (*29, 57*). This showed that the *act apo* ACP consisted of a four α helical bundle as shown in Figure 1-16. These helices (I, residues 7-16; II, residues 42-53; III, residues 62-67 and IV, residues 72-85 in *act*-ACP) are separated by structured loop regions, with the exception of residues 18-30 (loop 1), joining helix I to helix II, which is highly flexible and poorly defined.

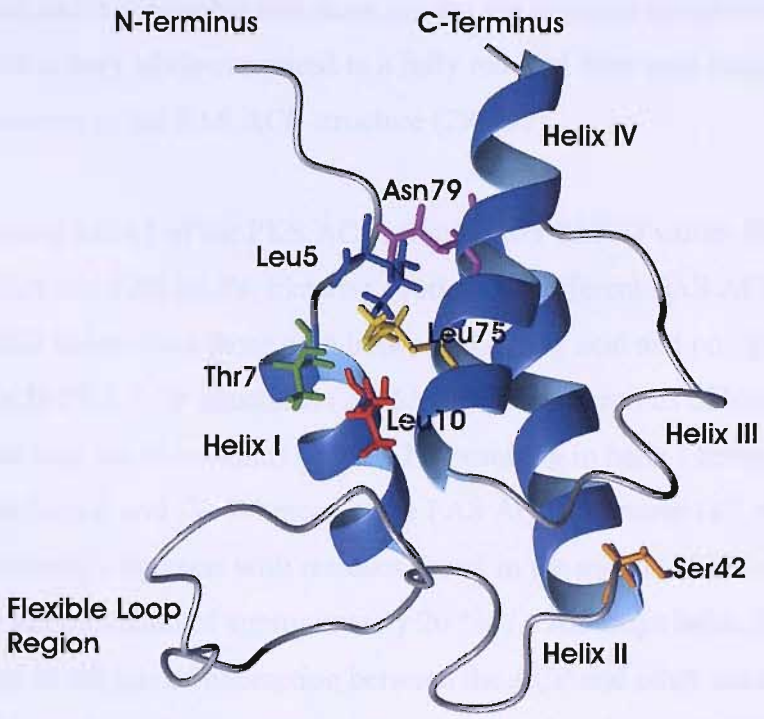


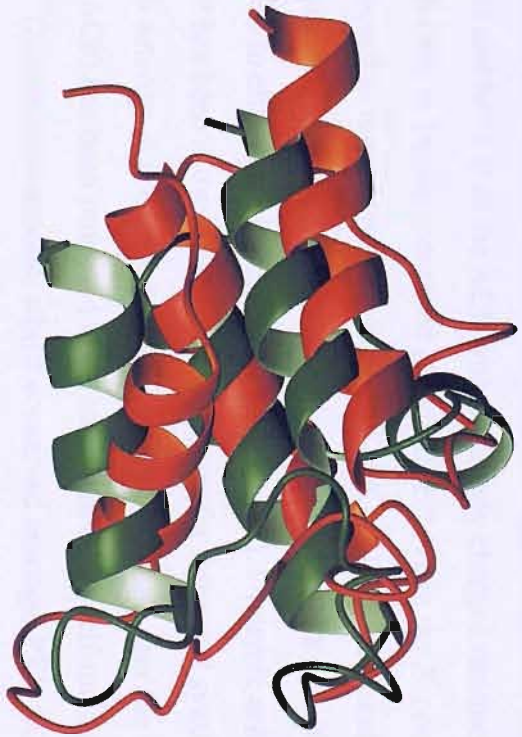
Figure 1-16 Averaged NMR structure of apo *act* ACP from *Streptomyces coelicolor* (*29*) (*58*), showing the four α helices, the flexible loop region joining the c-terminal end of helix I and the n-terminal end helix II, and Ser 42 (orange), to which the 4'phosphopantetheine moiety is attached by acpS. Contacts between residues Leu5, Thr7, Leu10, Leu75 and Asn79 were also seen in the structure and are also ilustrated[Figure prepared with the program MOLMOL (Koradi *et al* (*59*)].

Helices I and IV were shown to run anti-parallel to each other, and showed contacts between residues Leu5, Thr7, Leu10, Leu75 and Asn79. Helices I, II and IV were shown to pack around a hydrophobic cleft with the most structured part of the loop connecting helix I and helix II forming part of the base (*29*). When the apo *act* ACP structure (*29*) is compared to that of the known *E. coli* FAS ACP (*60*) and the apo *otc* ACP from *Streptomyces rimosus* (*61*) the same overall helical bundle structure is observed with variations in the N and C terminal regions and loop regions, as can be seen in the overlays in Figure 1-17. Strong sequence homology is seen within the helical regions whilst more variation being seen within the unstructured loop regions. Although the solved FAS and PKS ACP structures have been shown to exhibit a similar overall fold, differences within the amino acid sequence, between FAS and PKS ACPs are crucial to the observed differences seen in functionality between the two synthases, and this is investigated further during this work. There are a number of hydrophilic residues

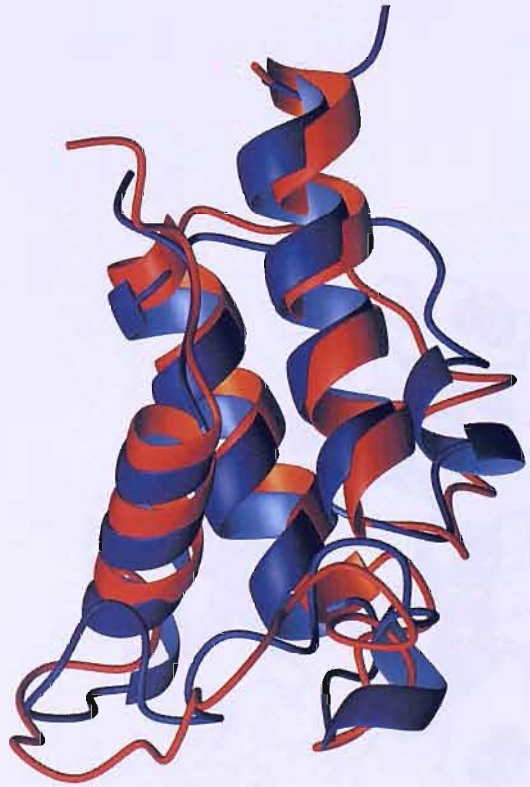
buried within the core of the protein which are not observed in the FAS ACP structure from *E. coli* and it is possible that these groups are required to stabilize the polyketide chain which is very labile compared to a fully reduced fatty acid chain, thus accounting for their absence in the FAS ACP structure (*29*) (*62*).

Superimposing helix I of the PKS ACPs show lower RMSD values than those seen between PKS and FAS ACPs, likewise overlays of different FAS ACP structures show lower RMSD values than those seen between the fatty acid and polyketide synthase ACPs (*61*). In PKS ACP structures (*29, 61, 63*) the N-terminus of helix I shows interactions with the N-terminus of helix IV, resulting in helix I being almost at right angles to helices II and IV. Whereas in the FAS ACP structures (*49, 64-66*) the N-terminus of helix I interacts with residues found in the middle of helix IV, resulting in a difference in orientation of approximately 20 ° (*61*). Although helix II has been proposed to be the site of interaction between the ACP and other enzymes in both FAS and PKS biosynthesis it may well be that the differences in orientation of helix I between the two systems plays an important part in recognition and selectivity. Overlays between FAS ACPs and PKS ACPs result in a higher RMSD value than that seen between PKS ACPs, which indicated subtle structural differences between the ACPs from the two systems, which may account for the interchangeability between PKS ACPs but not between the PKS and FAS ACPs.

The crystal structure of ACP-acpS from *Bacillus subtilis* revealed that the two molecules interacted as a trimer of dimers, as shown in Figure 1-18. Interactions were seen between helix III of the ACP (Leu37 and Met44) and helix I of the acpS, with residues of the ACP protruding into hydrophobic pockets on the acpS. Leu37 is shown to interact with Met18, Phe25, Phe54 and Ile15 of the acpS, and Met44 to interact with Phe25 and aliphatic parts of the side chains of Arg28 and Gln22 of the acpS (*66*). The structure of acpS-CoA has also been solved (*66*) and, by superimposing that structure with that of the ACP-acpS structure, shows that the loop region of the acpS (residues 64-78) shifts by ~ 2 Å which allows for the accommodation of helix IV from the ACP. It is also seen that the dipole of helix II of the ACP is directed to the phosphate of the CoA that is being transferred to the ACP.



a) *act* ACP *Streptomyces coelicolor* (*29*) (red) overlayed with FAS ACP *Bacillus subtilis* (*64*)(green).



b) *act* ACP *Streptomyces coelicolor* (*29*) (red) overlayed with *otc* ACP *Streptomyces rimosus* (*61*) (blue).

Figure 1-17 Backbone overlays of a) *act* ACP *Streptomyces coelicolor* (red) and FAS ACP *Bacillus subtilis* (green) giving an RMSD of 9.994 and b) *act* ACP *Streptomyces coelicolor* (red) and *otc* ACP *Streptomyces rimosus* (blue) giving an RMSD of 4.001. Similarity in size and orientation of the overall 4 helical bundle fold can be seen. The loop regions can be seen to be less well ordered between the different ACPs. The flexible c-terminal tail region of the *otc* ACP has been removed for clarity. [Figure prepared with the program MOLMOL (Koradi *et al* (*59*))].

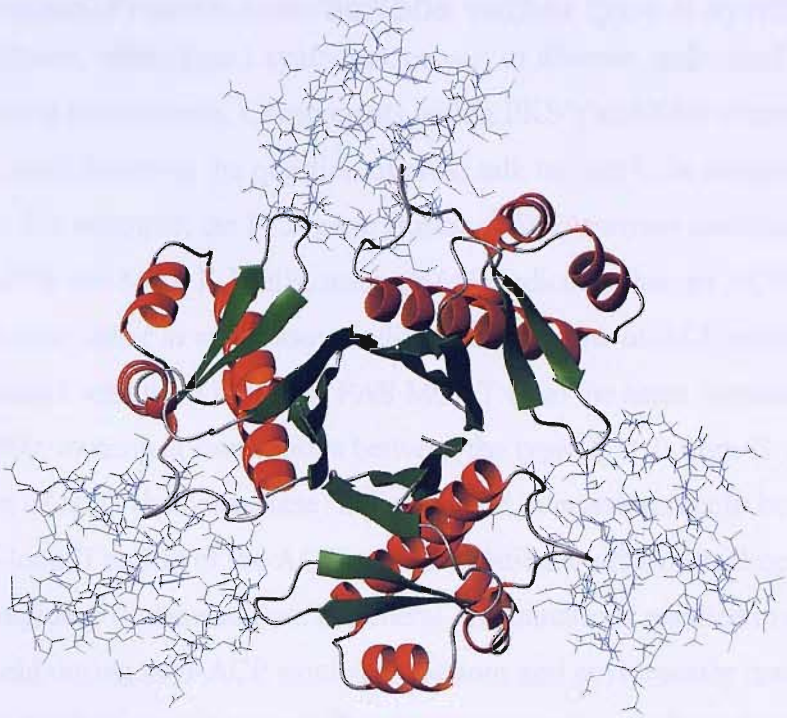


Figure 1-18 Crystal structure of ACP-acpS from *Bacillus subtilis*, showing that the two proteins interact as a trimer of dimers. The acpS is illustrated in a ribbon formation whilst the ACP is shown in a line format. [Figure prepared with the program MOLMOL (Koradi *et al* (*59*)].

The N-terminus of helix II in all solved ACP structures contains a conserved serine residue to which the phosphopantetheine arm is covalently attached. The crystal structure of *Bacillus subtilis* ACP bound to acpS indicates interactions between the DSL motif and helix II of the ACP with positively charged residues on the surface of the acpS (*66*). A highly conserved glutamic acid (41 in the *E. coli* FAS ACP) has been proposed to interact with arginine 249 of KAS III, with mutations of this arginine eliminating activity. Mutations of glutamic acid 47, 48 and 49 on the ACP have been shown to eliminate binding to the KAS III (*67*). Glutamic acid 41 of *Vibro harveyi* FAS ACP has been shown (*68*), to be important for fatty acid synthesis but not essential for acpS activity. A conserved arginine residue is found near the base of helix IV (Arg72 in *act* ACP) has been proposed to be important in the self-malonylation of PKS ACPs (*29, 42*). This arginine residue, whilst conserved amongst PKS ACPs is very poorly conserved amongst FAS ACPs, consistent with their observed inability to self-malonylate.

1.6 Protein-Protein interactions within type II synthases

Type II synthases, unlike type I synthases, consist of discrete, individual enzymes which associate during biosynthesis. Components within PKS's and FAS's interact within their own system, however the question of cross talk has yet to be thoroughly investigated. For example, the PKS gene cluster of *Streptomyces coelicolor* lacks genes encoding ACPS and MCAT. Whilst research (*41*) indicates that *act* ACP is able to self-malonate *in vitro*, other *in vitro* assays indicate that the rate of ACP malonation is greatly increased with the addition of FAS MCAT from the same species. Recent NMR studies by *(69)*, looking at interactions between the type II ACP from *E. coli* and the gene product of fabG (ketoreductase) indicated that interactions occur between the fabG and helix II-loop II region of the ACP, whilst mutation studies on hydrophobic residues within this region (*70*) showed that, in general, the mutations resulted in a decrease in fatty acid yield during acyl-ACP synthase reactions and acyl transfer reactions. From an evolutionary point of view it appears that the two systems may be derived from the same origin due to the presence of conserved regions within individual enzyme types (*1*). Further investigation within this field is required in order to elucidate the discrete differences which prevent interactions between enzymes from different systems.

1.6.1 β-Ketoacyl Acyl Carrier Protein Synthases

β-Ketoacyl Acyl Carrier Protein Synthases (KAS) are condensing enzymes found within both fatty acid synthase and polyketide synthase systems (*71*). They catalyse the formation of carbon-carbon bonds within the growing chain by the process of condensation, usually between acyl groups either in the form of acetyl groups bound to ACP or acetyl CoA and malonyl CoA. Several crystal structures have been solved to high resolution for fatty acid KAS and one structure for the *actinorhodin* polyketide ketosynthase from *Streptomyces coelicolor* (*72*).

In fatty acid synthesis acyl CoA/acyl-ACP donates the acyl group for condensation with malonyl ACP. This is a three step reaction:

1) Initially an acyl group from the *holo*-ACP/acyl CoA is transferred to the active site cysteine of the condensing enzyme to form a thioester.
2) Generation of a carbanion by the decarboxylation of malonyl-ACP followed by carbon-carbon bond formation by nucleophilic attack of the carbanion at the carbonyl carbon atom of the thioester.

This overall process is referred to as a Claisen condensation and is shown in Figure 1-19 (*71*).

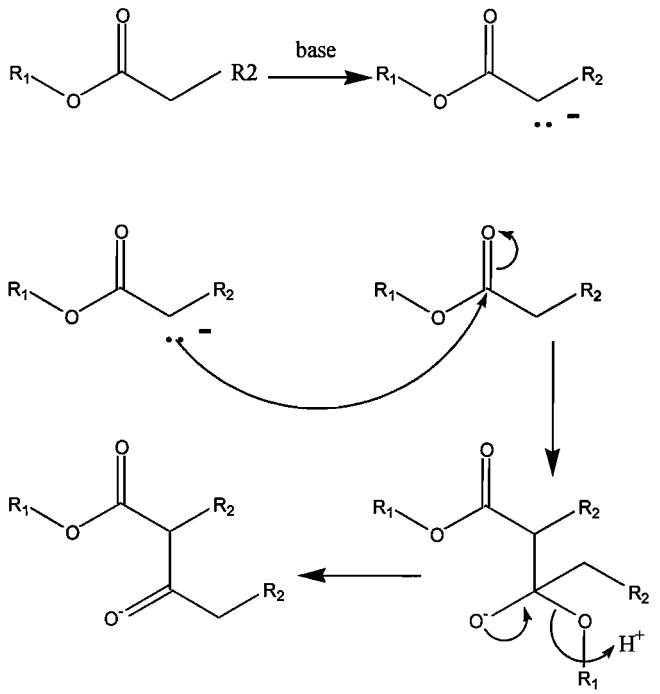


Figure 1-19 The Claisen condensation mechanism. The first step in the Claisen condensation is the production of an enolate from an ester. Condensation of this enolate with another ester forms a tetrahedral intermediate. The carbonyl reforms with the loss of an alcohol group to make a β-keto ester.

Fatty acid synthesis in *E. coli* features three independent KAS enzymes, each involved in acyl chain production, referred to as KAS I, KAS II and KAS III (*73-77*). KAS I, the gene product of fabF, is responsible for a wide range of condensation reactions, including the elongation of C10:1 which KAS II is unable to catalyse. Like KAS I, KAS II (fabB) also undertakes various condensations of different length acyl chains including the elongation of C16:1 (palmoleitic acid) to C18:1 (*cis*-vaccenic acid). It has been shown (*78*) (*79*) that KAS II plays a key role in thermoregulation, as the percentage production of *cis*-vaccenic acid increases with a decrease in environmental temperature. This is important as in various cold temperature species, such as deep sea fish, where *cis*- fatty acids help to keep the cell membranes fluid since the *cis*-isomer does not pack as well in the membranes as the trans-isomer (*71*). KAS III (fabH) has been shown to catalyse the elongation of acetyl-CoA and malonyl-ACP to yield acetoacetyl-ACP, the first step in fatty acid biosynthesis (*71*) (*76*) (*77*). Crystal structures have been determined for the fatty acid ketosynthases KAS I (*80*) KAS II (*71*) and KAS III (*81*) (*82*) from *E. coli*, KAS II from *Synechocystis sp.* (*83*) and KAS III from *Mycobacterium tuberculosis* (*84*), with the details of these structures being shown in Table 1-2. The

overall structures of these enzymes are similar and are close to the five layered α-β-α-β-α fold seen within the biodegradative enzyme yeast thiolase I (*71*).

The KS-CLF structure from the actinorhodin PKS from *Streptomyces coelicolor* was solved by Keatinge-Clay *et al* in 2004 (*72*) and is the first ketosynthase structure from a polyketide synthase. The *S. coelicolor* KS-CLF has the highest sequence similarity to FabF from *E. coli* (37 % [KS] and 28 % [CLF]). Like the FabF, KS catalyses malonyl decarboxylation utilising two His residues at the active site (His 309 and 346 in the act KS) and is acetylated on Cys 169 at the nucleophilic elbow (*72*). The act KS and CLF structures have the same overall fold (αβαβα) as seen in other ketosynthases (*85*). The KS-CLF structure shows highly complementary contacts between the two enzymes along the pseudo two-fold axis, including Tyr 118 (KS) and Phe 116 (CLF), which make tighter interactions than the equivalent in homodimeric ketosynthases. Hydrophobic residues (Met 139, Phe 140, Leu 143 and Val 144) of the KS fill a CLF tunnel and a 6 residue KS insertion (the grasping loop) wraps around helix III of the CLF, possibly to compensate for the binding energy required to build a polyketide tunnel at the heterodimer interface. This tunnel, amphipathic in KS-CLF, whereas hydrophobic in FabF, has been shown, via computational modelling, to allow a KS bound heptaketide to extend into it (*72*).

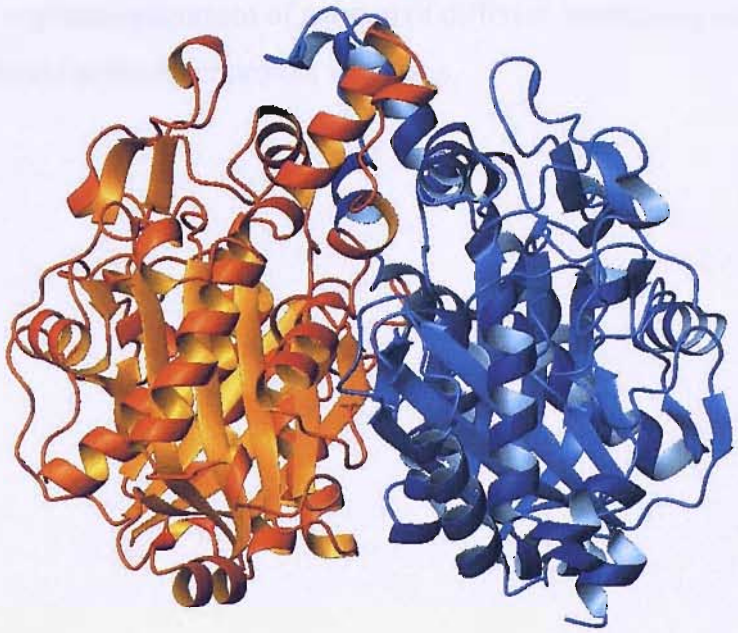


Figure 1-20 Structure of actinorhodin KS-CLF from *Streptomyces coelicolor* showing an overall fold in each subunit similar to that seen in yeast thiolase I and is of an α-β-α-β-α topology. The KS id shown in blue and the CLF is shown in gold. [Figure prepared with the program MOLMOL (Koradi *et al* (*59*)]

Species	KAS	Resolution	Year	PDB code
E. coli	I	2.30 Å	1997	1DD8 (*80*)
E. coli	II	2.40 Å	1997	1KAS (*86*)
Synechocystis Sp.	II	1.54 Å	2000	1E5M (*83*)
E. coli	III	1.80 Å	2000	1EBL (*81*)
E. coli	I	1.85 Å	2000	1EK4 (*87*)
E. coli	I	2.45 Å	2000	1G5X (*67*)
E. coli	III	2.00 Å	2000	1HN9 (*88*)
Mycobacterium tuberculosis	III	2.10 Å	2001	1HZP (*84*)
Streptococcus pneumoniae	II	1.30 Å	2003	1OXO (*89*)
Streptomyces coelicolor	Polyketide synthase KS	2.0 Å	2004	1TQY (*72*)

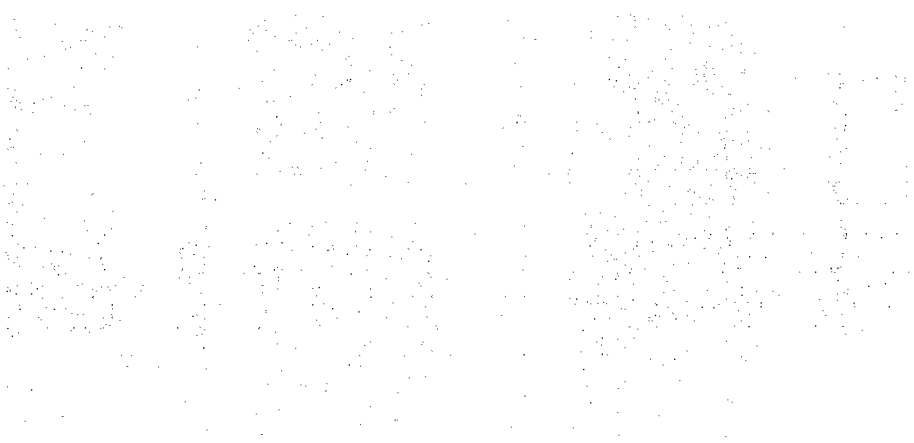
Table 1-2 Details of published KAS structures.

Sequence alignment for different condensing enzymes has identified three different categories of enzymes based on their functionality as is seen in Table 1-3 (*83*). Figure 1-21 shows the sequence alignment of a group of different condensing enzymes, including β-ketoacyl acyl carrier protein synthases.

Property	Group 1	Group 2	Group 3	KS
Carrier of acyl group	ACP. CoA.	ACP.	CoA.	ACP
Identity of elongation substrate	Malonyl-ACP. Malonyl-CoA.	Malonyl-ACP.	Acetyl-CoA.	Malonyl-ACP
Catalytic residues	Cys..His..-..Asn..Phe	Cys..His..Lys..His..Phe	Cys..Asn..-..His..Cys	Cys-His-His
Participation of second subunit to active site	Yes, One residue from loop.	Yes, Two residues from α-helix.	No.	No
Cerulenin sensitive	Yes/No.	Yes.	No.	_____
Enzymes belonging to this group	*E. coli* KAS III. Chalcone Synthase.	*E. coli* KAS I. *E. coli* KAS II. *Synechocystis sp.* KASII.	Biosynthetic and degenerative thiolases.	Polyketide Ketosynthases

Table 1-3 Properties of different classes of condensing enzymes categorised due to sequence alignment (*83*).

A proposed mechanism for the active site of KAS I from *E.coli* can be seen in Figure 1-22.



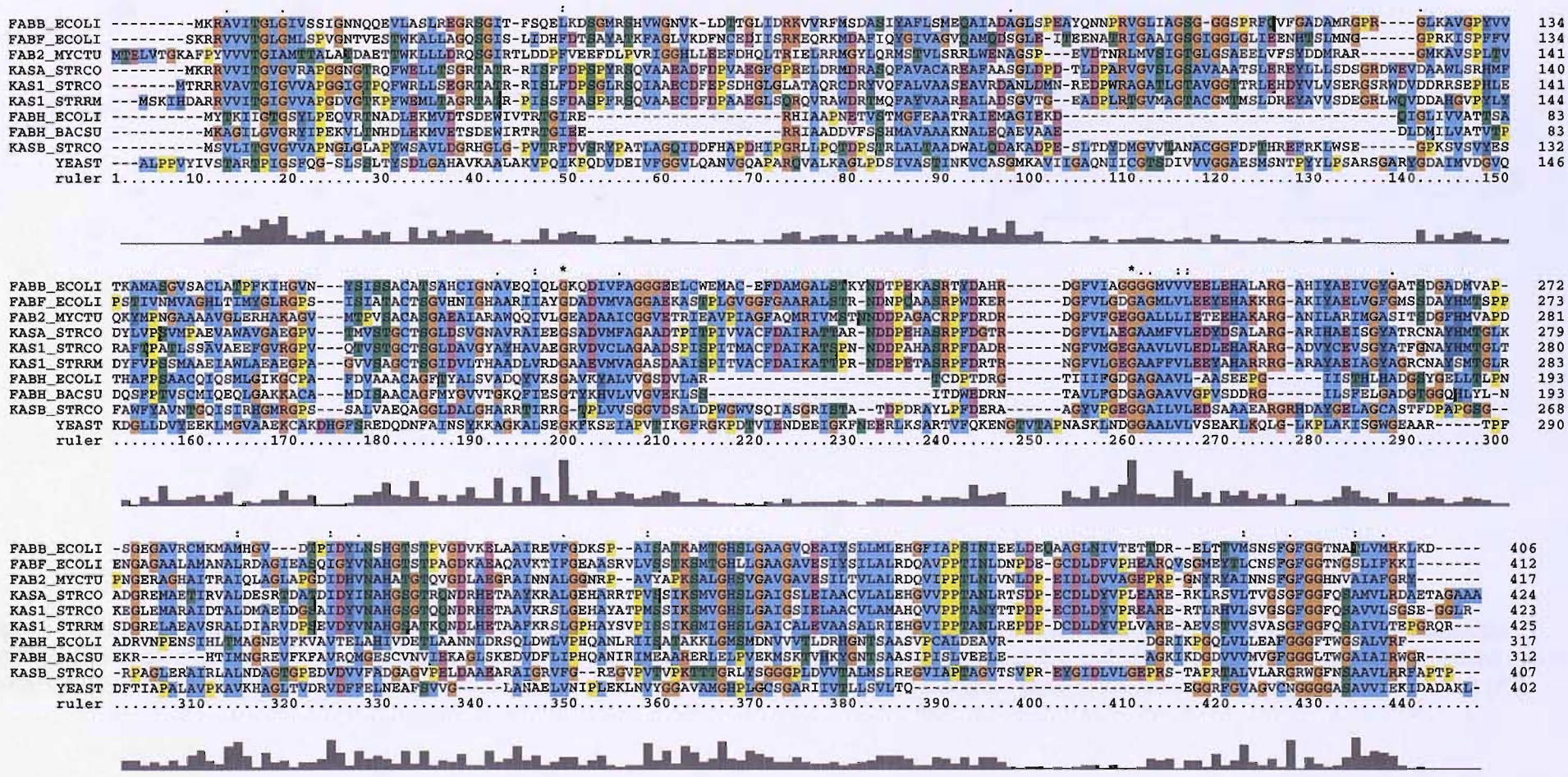


Figure 1-21 Sequence alignment of condensing enzymes, including KASA (KS) and KASB (CLF) from *Streptomyces coelicolor*.

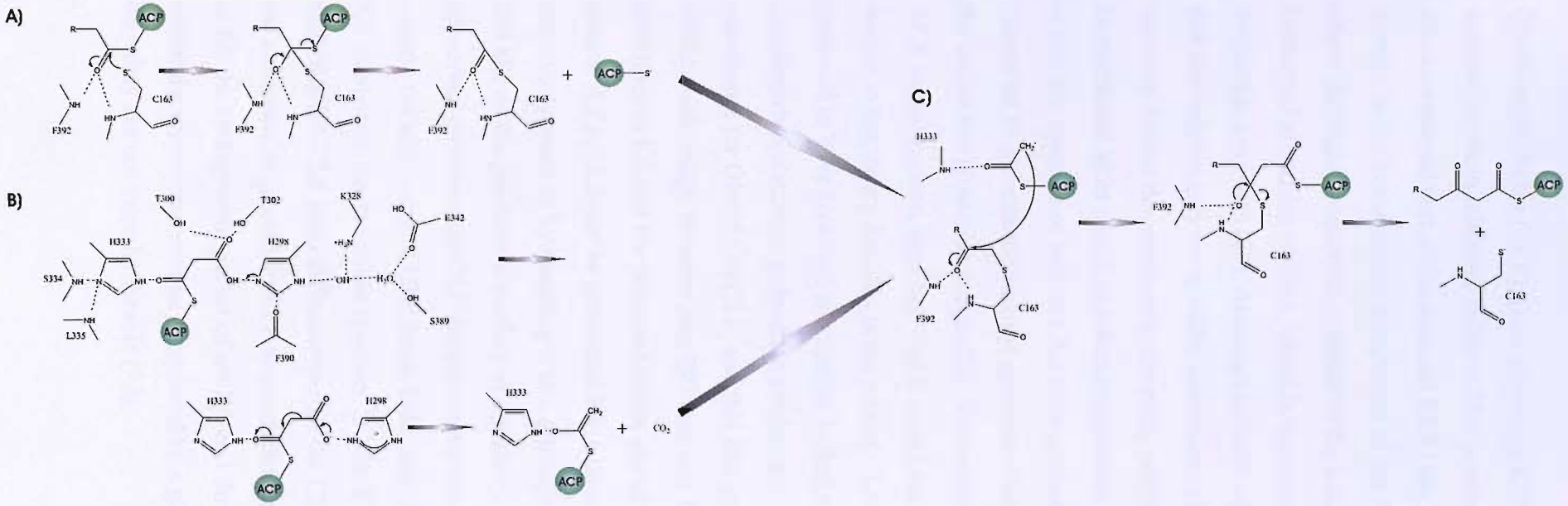


Figure 1-22 Claisen condensation reaction of KAS I from *E. coli*. A) Acyl transfer, via a tetrahedral KAS 1-fatty acid-ACP complex intermediate begins the reaction. B) A malonyl-ACP is bound to the decarboxylase site and the decarboxylation involves Histidines 298 and 333 from the KAS. C) The ACP bound acetyl carbanion attacks the thioester carbon of the bound fatty acid. A second tetrahedral intermediate is formed prior to the releases of 3-oxyacyl-ACP product (*87*).

1.7 Chain Length Factor

Chain length factor (CLF), (also known as KSQ (a previously uncharacterised PKS domain) or chain initiation factor (CIF)) is essential for polyketide production and is also an essential part of the minimal PKS (*90*). The CLF shows a good sequence identity to the ketosynthase component of the PKS, with the exception of the active site where the highly conserved cysteine in the ketosynthase is replaced by a highly conserved glutamine residue, which is important for both decarboxylation and polyketide synthesis (*90*). Although initially called chain length factor, it is now thought that the enzyme may not actually alter/control the chain length (*91, 92*), with other enzymes further downstream in the cycle partially taking on this role. CLF only interacts and takes part in polyketide synthesis in the presence of ketosynthase. If the active site cysteine of animal FAS ketosynthase is reacted with iodoacetamide then it is converted to carboxamidomethyl cysteine which has a similar shape and functionality to the active site glutamine in the CLF. This converts the ketosynthase into a malonyl-ACP decarboxylase, implying that it would act as a chain initiation factor and adds weight to the theory that this is the role of CLF molecule (*90*). CLF was originally proposed to have an active site similar to that of KS, which was thought to carry out the first decarboxylation of polyketide production. The active site Cys 169 of act KS is substituted for Gln 161 in CLF, and this Gln appears to have a structural function, whilst a salt bridge between Asp 297 and Arg 332 in CLF replaces the catalytic histidines in KS and the proposed active site of the CLF is filled by large bulky side-chains. A FAS KS can be converted into a potent decarboxylase by acetylating the reactive cysteine or by mutating it to a glutamine, which inserts into the oxyanion hole and allows the productive binding of a malonyl group (*93*). However the highly conserved glutamine of CLF does not enter the oxyanion hole. The large residues which occupy the active site of CLF leave little room for a malonyl substrate and therefore the KS must carry out the initial reaction. Whilst KS may be interchangeable within mPKS and PKS KS-CLF from different species the CLF entity may not be interchanged (with the exception of granaticin CLF to actinorhodin CLF). The act CLF uses tryptophan 194 to fill the hydrophobic pocket of act KS and this is conserved in the granaticin CLF whilst the tryptophan residue is replaced by a glycine in the most other CLFs, indicating why they are not interchangeable (*72*).

1.8 Reasons for Studying Streptomyces species and their Polyketide Secondary Metabolites

Streptomycetes are members of the bacterial order Actinomycetes and, although prokaryotic, resemble fungi in that they produce mycelia and spores. They are found worldwide in soils and are non-motile, gram-positive bacteria. The largest Streptomycetes genus is *Streptomyces*, which contains more than 500 species. The *Streptomyces sp.* are of great importance to mankind due to the wide range of polyketides that they produce, which are utilised medically and industrially. Actintomycetes make two thirds of all antibiotics from a microbial source, with 80% of these being from *Streptomyces sp.* and with the worldwide market for antibiotics being $30 billion per year it is a financially productive area for research. However a more important reason for studying PKS and FAS systems, both from *Streptomyces* and other sources, is to elucidate the mechanisms of polyketide biosynthesis with the aim of being able to produce new polyketide products to help overcome the growing problem of antibiotic resistance and to provide more effective therapies for other medical conditions such as cancer, high cholesterol and transplant rejections. In addition, by examining differences between PKS and FAS, both within the same organism and between species, it may be possible do develop a range of new drugs which will selectively inhibit bacterial FAS systems without detrimental effects on the infected host.

1.9 X-Ray Crystallography and Nuclear Magnetic Resonance.

The discovery of X-ray crystallography and its application for solving three dimensional structures was made by Sir William Bragg and his son Sir Lawrence Bragg, for which they received a Nobel Prize in 1915. Protein crystallography was first used successfully in 1934 when J. Bernal and D. Crowfoot discovered that crystals of stomach protease pepsin would yield a diffraction pattern. However it was not until 1960 that the first protein structure, of myoglobin, was published by Kendrew *et al.* (*94*)

Nuclear Magnetic Resonance Spectrometry (NMR) was first demonstrated in 1945 by Purcell *et al* (*95*), with the first protein structure, of proteinase inhibitor II from bull seminal plasma, being published by Wüthrick *et al.* in 1985 (*96*). Unlike crystallography, NMR is able to provide not only structural data, but also information

on dynamics, folding and conformational equilibrium. Another advantage of NMR over crystallography is that samples are studied in solution at near physiological conditions negating the need to grow crystals. However the size of proteins which are conventionally analysed by current NMR techniques is generally limited to approximately 40kDa (although there are exceptions such as the determination of the secondary structure of Malate Synthase G, a 723 residue protein, by Tugarinov *et al.* (*97*)), whereas any protein or protein complex which crystallises may be investigated by crystallography

Therefore both techniques are invaluable in the field of structural biology and allow the scientist a choice of method dependent on the properties of the molecule to be studied and the analysis required.

1.10 X-ray Crystallography

X-ray crystallography is a commonly used technique by which three-dimensional structures of molecules can be determined. Crystals are grown by a process of controlled precipitation from the mother liquor, which slowly drives crystal assembly without causing denaturation of the protein. The hanging-drop vapour-diffusion technique, used during this work, involves suspending a drop of mother liquor containing protein over a pool of mother liquor in a sealed environment. Since the mother liquor in the drop is diluted with the protein solution, over a period of time the excess water diffuses to the main reservoir until equilibrium is reached. If conditions, such as temperature, pH, protein and precipitant concentrations are suitable, then crystals form. The mounting of a single protein crystal, greater than 0.03mm in each dimension, in front of a monochromatic X-ray beam, will give rise to an X-ray diffraction pattern.

X-rays are waves of light (electromagnetic radiation) with wavelengths of between 0.1 and 100 Å. Since their wavelengths are of a similar magnitude to atomic diameters (~1.5 Å), they are able to interact physically with clouds of electrons surrounding atoms. Diffraction patterns occur when X-rays are diffracted by the electrons within a crystal, and from this interaction the three dimensional positions of electron clouds, and hence atoms in space, may be obtained.

X-rays can be generated when high energy electrons collide with an electron cloud surrounding a metal atom and electrons from an inner shell are ejected. When this happens an electron from a higher energy orbital drops down a level, and energy is emitted from the electron in the form of an X-ray photon (*98*). This is called electron cloud bombardment.

A second X-ray source is via electron acceleration. When electrons/positrons, travelling at a velocity near to that of light, are exposed to an accelerating/decelerating force those particles will emit X-ray photons which is what happens at a synchrotron (*98*). In a synchrotron source, electrons or positrons are linearly accelerated and fed into a storage ring where they undergo centripetal acceleration in response to magnetic fields. This process generates X-rays which can be used to record diffraction data.

1.10.1Measurement of Diffraction Data

Crystals are typically flash frozen in a mohair loop, using either liquid ethane or a cryostream. In order to protect the crystal during the freezing process a cryoprotectant is used, often glycerol, to help prevent damage to the crystal. Radical induced damage to the crystal is reduced by freezing, and hence better quality data sets and a larger number of images may be recorded from a single crystal.

For data collection a synchrotron source has several advantages over normal X-ray sources, including quicker data collection (2-3 hours compared to several days for a conventional X-ray source), more intense beams (useful for weakly diffracting crystals), the generation of X-rays with higher frequency (giving higher resolution data) and wavelength-tuneability (allowing selection of multiple wavelengths).

The ability to record diffraction data is a very powerful process. However, since proteins are large molecules, there are far fewer of them present in a crystal of a given size relative to a small molecule crystal. Moreover, proteins are composed of 'light', relatively electron-deficient atoms (hydrogen, oxygen, nitrogen, carbon and sulphur), which interact with X-rays less powerfully than electron-rich heavier atoms. As a consequence of these differences, the overall reinforcable diffraction output from a protein crystal is relatively weak per unit time, whereas from a small, electron rich atom, it is relatively strong. This difference in diffraction power is employed in a

technique called isomorphous replacement, which allows phasing of the net X-rays which contribute the individual spots on diffraction pattern (*99, 100*).

1.10.2 Crystals and Unit cells

A crystal is an ordered structure made up of identical, repeated unit cells. Each unit cell contains one or more molecules in certain orientations. This ordered structure allows for amplification of the diffraction signal. The unit cell of a crystal can take on one of seven crystal systems defined by three cell lengths (a, b and c) and three cell angles (α, β and γ), which describe the shape of the unit cell. These can be seen in Table 1-4.

Crystal system	**Cell length parameters**	**Cell angle parameters**
Cubic	$a = b = c$	$\alpha = \beta = \gamma = 90^\circ$
Tetragonal	$a = b \neq c$	$\alpha = \beta = \gamma = 90^\circ$
Orthorhombic	$a \neq b \neq c$	$\alpha = \beta = \gamma = 90^\circ$
Trigonal (rhombohedral axes) (hexagonal axes)	 $a = b = c$ $a = b \neq c$	 $\alpha = \beta = \gamma$ $\alpha = \beta = 90^\circ, \gamma = 120^\circ$
Hexagonal	$a = b \neq c$	$\alpha = \beta = 90^\circ, \gamma = 120^\circ$
Monoclinic	$a \neq b \neq c$	$\beta \neq \alpha = \gamma = 90^\circ$
Triclinic	$a \neq b \neq c$	$\alpha \neq \beta \neq \gamma$

Table 1-4 The seven crystal systems found within protein crystals. The cell length parameters (a,b,c) and the angle parameters (α,β,γ) vary with no limit on the dimensions other than those needed to satisfy the given statements.

The Bravais lattice of the crystal shows both the crystal system and the centring. Centring describes where certain copies of proteins occur inside the unit cell. It can take on any of four forms: primitive (no centring), C-centring, F-(face) centring and I-(body) centring. The point group (symmetry) of the crystal shows how many rotation axes are present and it also states their nature. The screw axes show how various molecules within the unit cell are related *via* rotation and translation. All these components make up the space group which defines the crystallographic symmetry. For example, the space group $P2_1$ belongs to the primitive (monoclinic) system, and has a single 2_1 screw axis composed of a two-fold rotation about the b axis, and a 1/2- cell translation along the b axis. The asymmetric unit is the smallest unit of the crystal that generates a

complete crystal structure. It requires no symmetry of its own, but is related by symmetry operators to every other asymmetric unit within the unit cell.

The crystal can be divided into a series of imaginary planes (for each set of Miller [hkl] indices) with an interplanar spacing of d, where d values are related to the unit cell parameters. These imaginary planes are useful when we consider Bragg's law (Equation 1-1):

Bragg's law: $n\lambda = 2d_{hkl}\sin\theta$ **Equation 1-1**

This equation shows that when a set of parallel planes with index *hkl* and interplanar spacing d_{hkl} are exposed to X-rays of wavelength λ, if the angle of incidence (θ) is such that Bragg's law is satisfied, then a diffracted beam will emerge from the planes at angle θ, where n is an integer (*98*). When these conditions are met, a diffraction spot appears on the image since X-rays diffracted from separate layers within the crystal interfere constructively. However when $n \neq$ integer destructive interference occurs and no diffraction spots are formed. For each given space group, the spots will always appear in the same relative positions but at different intensities for different molecules. Therefore, from the position of the spots the cell dimensions can be determined.

Each spot on the diffraction pattern is made up of X-rays diffracted by electrons from every atom within the molecule. Each spot is assigned a set of Miller indices [hkl], which are its identification 'tag' in 3D reciprocal space. Reciprocal space is a hypothetical 3 dimensional system which is used to identify and help visualise the diffraction spots in a way which is of ease and practical use to the crystallographer. The central reflection is assigned the indices 0,0,0 and used as a reference, with other reflections assigned whole number indices. This allows for the recording of spot positions and intensities of individual spots. Each net X-ray beam giving rise to a diffraction spot has an associated amplitude, phase and frequency. The crystallographer must determine these parameters for each net X-ray beam and work backwards, in order to reveal the contents of the crystal. The frequency is the same as that of the X-ray beam being used to create the diffraction pattern. The amplitudes of the X-ray beams can be calculated directly from the intensities of the diffraction spot. However, determining the phases for the X-ray beam is a more complex task, which is often one of the most time consuming aspects of crystallography.

The intensities are calculated using specialised computer programs which determine individual spot profiles and then record the intensities of the spots on the diffraction pattern. Once collected, the spot intensities are next scaled together to correct for variations within the crystal and the X-ray beam during data collection. This involves correcting the collected intensities of symmetry-related spots to improve their agreement. After scaling, the intensities are converted to structure factor amplitudes $|F_{hkl}|$ as shown in Equation 1-2.

$$I = |F_{hkl}|.K$$

(K = scale factor) **Equation 1-2**

Once the net X-ray beam amplitudes for each diffraction spot have been calculated, the phases then need to be determined. Isomorphous replacement, multiwavelength anomalous diffraction (MAD) and molecular replacement are the most commonly used procedures employed. Molecular replacement is usually tried first, as it is generally considered to be the easiest and quickest method. Molecular replacement requires obtaining a previously solved structure of a protein with good sequence similarity (typically better than 20%). The phases which relate to this molecule can be applied to the recorded data set to obtain a co-ordinate structure which can then be refined.

Multiwavelength anomalous diffraction phasing can be used to solve the phasing problem in proteins which do not naturally contain a functional heavy atom. These proteins can be expressed in *E. coli* using a medium containing selenomethionine as the only methionine source, to introduce a heavy atom. Data sets are then obtained at different wavelengths to provide sufficient phasing data to enable a structure to be solved. Friedel pair ($I_{hkl} = I_{h\text{-}k\text{-}l}$) intensities at each specific wavelength provide sets of distinct phasing equations which, along with slight changes in intensity seen at different wavelengths, provide sufficient variables to solve a set of linear equations containing phase information (*98*). Isomorphous replacement requires the uniform soaking of heavy atoms such as Hg, Pt and Au into crystals which have been shown to diffract well. Intensities of pairs of reflections are altered by the addition of the heavy atoms and these are used to determine phases (*98*).

Once the phases for the new model have been determined, refinement of the model can begin. This involves putting the model through a series of programs which alter the geometry in order to make the model data best fit the observed data. After the first round of refinement, the model is put into a model building program. Here the structure is checked to ensure that the protein chain fits into the electron density maps (calculated from the target experimental/refined model data before model building). The protein backbone is fitted into the electron density and any regions of unaccounted density are located and side chains, ligands, etc. are positioned within them. The structure output post model building is then run through various iterations of refinement/model building in order to further fit the model to the observed data resulting in the final structure.

1.11 Nuclear Magnetic Resonance

In order to describe the process of protein structure determination by NMR it is necessary to understand the basic principles of the technique.

The NMR experiment was described by Rattle (*101*) as:
an atomic nucleus, with an angular momentum quantum number (I) and an associated magnetic moment (μ), is restricted by space quantisation rules to (2I+1) orientations. Usually these orientations are degenerate, but the application of a magnetic field (Bo) causes the orientations to split into (2I+1) distinct levels, separated by small energy gaps, with a slight excess of nuclei in the lower energy state. The application of electromagnetic radiation of an appropriate frequency stimulates transitions between these states, disturbing the population levels. This excitation, with a net absorption of energy and the subsequent relaxation back to the equilibrium state may be analysed to give information about the molecules of which the nuclei form a part.

Magnetic effects are produced by all currents, and hence the motion of electrons orbiting a nucleus produces a localised magnetic field. This effect is cancelled out in most atoms, due to paired electrons. A few atoms (paramagnetic atoms) have unpaired electrons and act as a small magnet which will align with an externally applied magnetic field (Bo).

The application of a magnetic field to a sample induces currents by altering the orbital motion of the electrons within the sample (the diamagnetic effect), and is common to all materials regardless of whether they are paramagnetic as well. These induced currents produce their own magnetic fields. Diamagnetic effects are much smaller than paramagnetic effects but because they are common to all atoms and molecules they play an important part in NMR.

Magnetic interactions are associated with some form of electromagnetic radiation. The radiation used during a NMR experiment is in the range of 1-1000 MHz and falls into the radio frequency range of the spectrum.

Electromagnetic radiation can interact with atoms and cause transitions between allowed energy states, when the energy of the photon of radiation is equal to the energy gaps between the energy states. The Boltzmann distribution describes the populations at different energy levels, where the higher energy levels have smaller populations. In NMR the distance between the energy levels is small, and as a consequence the ratio between the upper and lower energy states is almost, but not quite, equal.

1.11.1 Angular momentum and spin energy levels

Angular momentum is a vector quantity and is imagined as an arrow along the axis of rotation and with length proportional to its magnitude. The quantisation of the angular momentum applies to both the magnitude and the direction of the vector. The magnitude of the angular momentum is restricted to values related to the angular momentum quantum number j, where j is 0 or a positive integer or half integer. This results in the axis of the spinning system being found in (2j+1) orientations.

Spin is the term used to define the idea of the angular momentum resulting from the rotation of the particle about its own axis. The spin quantum number (I) of the most abundantly used isotopes used during NMR is ½. The isotopes ^{1}H, ^{13}C and ^{13}N may be used during protein structure determination and all have a value of I = ½. All nuclei with a value of $I \neq 0$ have a magnetic moment (μ), which depends upon I and the gyromagnetic ratio (γ) of the specific nucleus (Equation 1-6):

$$\mu = \gamma I \qquad \textbf{Equation 1-3}$$

The natural abundances and other relevant data about ^{1}H, ^{13}C and ^{15}N nuclei which are used in protein determination is given in Table 1-5.

Nucleus	I	Natural Abundance (%)	γ x$10^{7}T^{-1}S^{-1}$	Relative Sensitivity	Resonance frequency at 14.1T (MHz)
^{1}H	½	99.98	26.75	1.0	600
^{13}C	½	1.11	6.73	1.6 x10^{-2}	150.9
^{15}N	½	0.36	-2.71	1.0 x10^{-3}	60.8

Table 1-5 Natural abundance and properties of selected nuclei used in protein NMR spectroscopy. (*102*).

A spinning charged particle may be regarded as a tiny loop current with its own associated magnetic moment. The lowest energy state for this magnet is with the magnetisation lined up along an applied magnetic field. However this magnet is subject to the rules of space quantisation and can occupy only (2I+1) orientations. Because each particle is aligned differently with respect to the magnetic field, each will have a different energy level and the effect of the applied magnetic field will split the particles energy into (2I+1) separate levels (Figure 1-23).

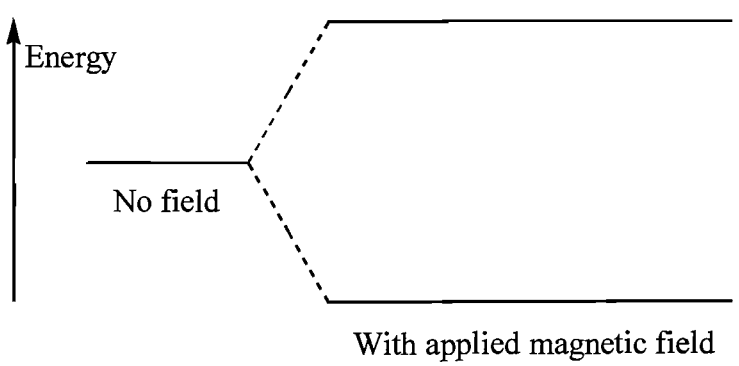


Figure 1-23 Splitting of energy level state of the nucleus by the application of an applied magnetic field.

1.11.2 Precession, excitation and relaxation

The application of a magnetic field to a sample also results in a process known as precession. This precessional effect occurs at a rate known as the Larmor frequency, which is dependent upon the type of nucleus (i.e. the number of protons and neutrons) and the strength of the applied magnetic field B_0. The absorption of an electromagnetic pulse by a population of nuclei within a constant magnetic field results in that population achieving an excited state, whereby the distribution of the nuclei from the Boltzmann equilibrium population is perturbed. Upon removal of this pulse the nuclei return to their pre-equilibrium states by a process know as relaxation. In NMR this relaxation is not spontaneous (due to the very small energy gaps) and is therefore induced by relaxation mechanisms within the sample, which yield motional and conformational information about the sample.

A nucleus possessing a magnetic moment will, when placed into a magnetic field, exhibit space quantisation. For $I = 1/2$ nuclei (^1H, ^{13}C and ^{15}N nuclei used in protein structure determination by NMR) the nucleus has two permitted orientations. The magnetic moment of the nucleus has a component m_z in the same direction as the applied magnetic field (B_0), said to be in the z direction, as the magnetic field is thought of as being along the z-axis in the laboratory frame of reference. The energy difference between the two states is given by $2m_zB_0$ and the frequency of the electromagnetic radiation (v) required to cause a transition between the two states is given by Equation 1-7:

$$v = \gamma B_0/2\pi \text{ (Hz)} \qquad \textbf{Equation 1-4}$$

The Larmor frequency (v_0) of the nucleus is directly proportional to the applied magnetic field and is given by Equation 1-8:

$$v = -\gamma B_0/2\pi \text{ (Hz)} \qquad \textbf{Equation 1-5}$$

By applying a radio frequency pulse (RF pulse) at the Larmor frequency of the nucleus of interest, a rotating magnetic field (B_1) in the xy plane is achieved, and the nucleus will precess around this field. The magnetic moment of a nucleus rotating around B_1 can be split into two components, the z component (in line with the B_0) which is constant, and a component in the xy plane rotating at the Larmor frequency. However in an NMR

sample there is a large population of nuclei of the same type, and at equilibrium the following occurs:

- The two energy states are very nearly equal, the z components therefore nearly cancel out, but the small excess of lower energy spins results in a small net magnetisation aligned along B_o.
- Each nucleus contributes to a rotating magnetic moment in the xy plane. However these are randomly arranged and hence they cancel out, resulting in no net magnetisation in the xy plane.

Therefore, at equilibrium a population of nuclei in an applied field exhibit a small magnetisation M_z in the direction of B_o.

Upon application of a RF pulse at the Larmor frequency a large number of transitions, in both directions, occur reducing the Boltzmann population difference and reducing M_z. The spins are brought into phase with B_1 and precess together. This precession is referred to as phase coherence. Upon removal of the RF pulse (and therefore removal of B_1) the spins no longer precess together and the net magnetisation in the xy plane is no longer zero, but given a value M_{xy}, rotating in the xy plane at the Larmor frequency. The detection of this rotating magnetisation and hence phase coherence is the purpose of an NMR spectrometer. Controlling the phase of B_1 will control the phase of M_{xy}, and therefore by applying B_1 along the x-axis in a rotating frame the direction of M_{xy} will be in the + y direction.

Therefore the application of a RF pulse reduces M_z and introduces M_{xy}, with the resultant net magnetisation (M) tilted away from the z-axis towards the xy-plane. The longer the RF pulse the greater this degree of tilt e.g. a 90^o pulse will result in zero M_z magnetisation and non-zero M_{xy}. The angle which the magnetisation is tilted away from the z-axis is referred to as the flip angle. With the removal of the RF pulse the spin system relaxes to its original state over a period of time. Bloch states that this relaxation occurs exponentially. M_z returns exponentially to its original value by a process of longitudinal relaxation with a characteristic time T_1 (Figure 1-24 a)) . M_{xy} decays exponentially to zero by transverse relaxation with characteristic time T_2 (Figure 1-24 b)).

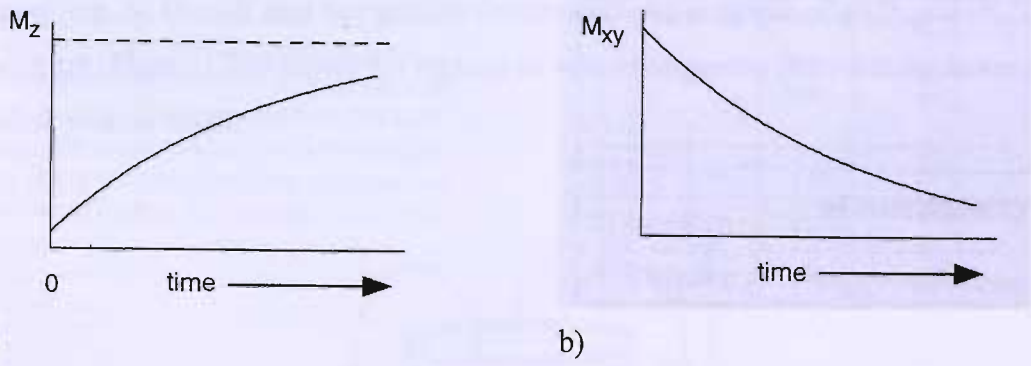


Figure 1-24 a) M_z returns exponentially to its original value by longitudinal relaxation, with characteristic time T_1 b) M_{xy} decays exponentially to zero by transverse relaxation with characteristic time T_2 (*101*).

1.11.3 Chemical shifts

A nucleus within a NMR sample will not be exposed to a magnetic field exactly equal to B_o. This is because of variations in the local magnetic field which it experiences and which is created by the movement of electrons surrounding the nucleus and those nuclei which are in close proximity. This diamagnetic effect produces a small, opposite magnetic field to B_o and therefore B_{loc} (the actual magnetic field to which it is exposed) is slightly smaller than the applied field. This is referred to as 'shielding' of the nucleus by the electrons. The extent of this shielding depends upon the local environment of the nucleus, which in turn depends upon the nature of the molecule. This effect leads to nuclei of the same element, but under different chemical conditions, having slightly different Larmor frequencies. This effect is known as 'chemical shift' and enables the spectrometrist to distinguish between different nuclei of the same chemical element within the molecule of interest. These shifts are measured relative to a standard compound, usually tetramethylsilane (TMS) and are recorded in part per million (ppm) of the applied field. In proteins, ring currents occur in aromatic side chains, where the π system, exhibits a powerful diamagnetic effect (see Figure 1-25). A nucleus at position A in Figure 1-25 would be shielded by the ring current and would experience a B_{loc} smaller than B_o and hence have a lower Larmor frequency, whereas a nucleus at position B would experience some de-shielding and therefore experience a B_{loc} slightly greater than B_o. The presence of paramagnetic ions, such as Fe^{2+}, also gives rise to local magnetic fields, and this phenomenon was used in early protein NMR to probe structures. Hydrogen bonding within the molecule will slightly increase the Larmor frequency of the proton involved in the bond, whilst protons bonded to electronegative

atoms (e.g. N, O) will also be partially de-shielded. An example of a 1D protein NMR spectrum (Figure 1-26) shows the regions in which the proton chemical shifts are found during protein NMR.

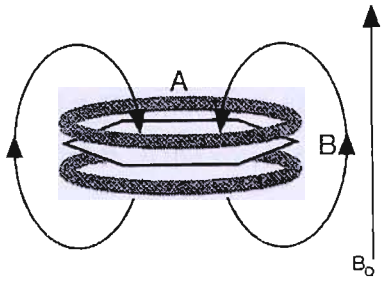


Figure 1-25 Strong diamagnetic effect of an aromatic ring. A nucleus at **A** will be strongly shielded whilst a nucleus at **B** will be desheilded (*101*).

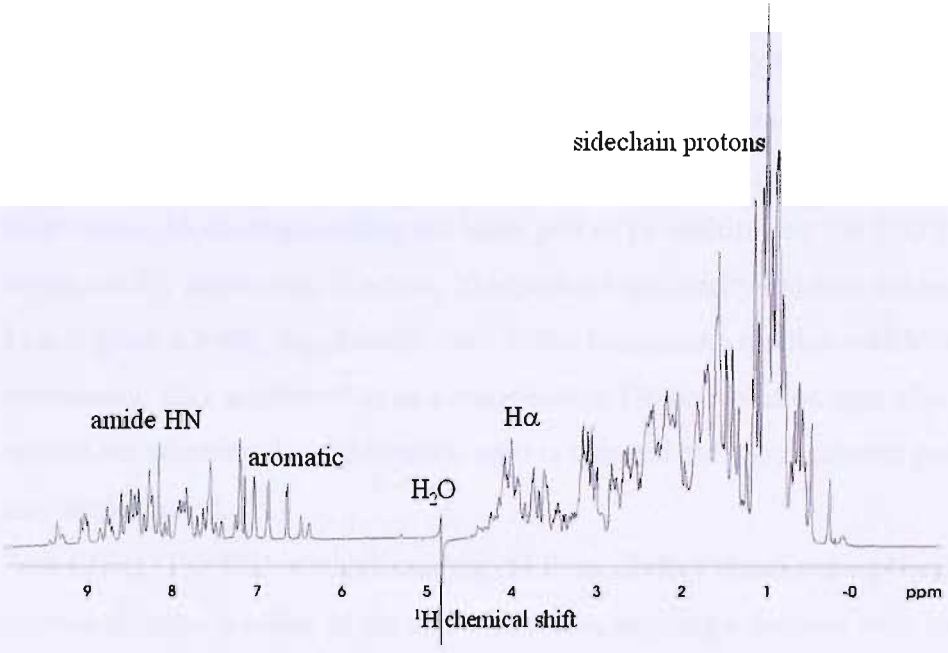


Figure 1-26 1D proton NMR spectrum, indicating the regions in which the chemical shifts for various protons found within a protein are found.

1.11.4 NMR detection and data manipulation

In the NMR spectrometer the net magnetisation of a pulsed sample, which oscillates at a frequency of M_{xy} in a sinusoidal wave decaying to zero due to transverse relaxation, is detected in the xy plane. This signal is referred to as the free induction decay (FID). If a spectrum of a FID of a single nucleus (derived by the process of Fourier Transformation, FT) is plotted a single line, representing the frequency of oscillations is

seen, with the line width depending on the rate of decay of the FID. In practice the FID recorded during a protein NMR experiment will be a combination of many different frequencies and will result in a very complicated waveform.

When recording NMR data it is usual to collect a number of repeat data points at each stage of the experiment in order to give a better signal to noise ratio. Protein NMR is usually carried out on a sample at a concentration of ~1 mM protein for assignment purposes. The protein must be soluble in the solvent at this concentration and at experimental temperatures, without any aggregation, and be stable for a period of time sufficient to record all the required data.

Once the FID has been recorded, the data may be manipulated in a number of ways to improve the quality of the data before FT is carried out:

- Apodisation. To reduce the time taken for the experiment the acquisition of the FID is usually halted before it has decayed to zero. Apodisation manipulates the final part of the FID, by a function, to obtain a gentle decay towards zero.
- Exponential manipulation. The earliest part of the FID contains the best signal to noise ratio (before the signal decays too greatly). The FID may therefore be transformed by de-emphasising the latter part or by multiplying the FID by an exponentially decreasing function. This process artificially reduces the apparent T_2 and gives a better signal/noise ratio whilst broadening the line width of the resonances. This is referred to as a convolution. De-convolution may also be carried out whereby the signal/noise ratio is reduced but much sharper peaks may be produced.
- Zero filling. The FID is stored in a digital form. Before transforming the data-set a series of zeros is added to the end of the data, forming a data-set with many more points and providing a sharper final spectrum.
- Phase correction. The phase relationship between the frequency component of the FID and the applied field B_1 may vary throughout the spectrum and is not automatically corrected for during the FT. The phase may therefore be corrected using 2 phase corrections; Zero-order, which corrects across the whole spectrum or first-order, which varies linearly from one end of the spectrum to the other.

The majority of NMR experiments used during this work utilise proteins dissolved in a water based buffer. This does however pose a serious problem due to the much larger

concentration of protons within the water than the sample (protons in H_2O 110 M compared to 1-2 mM in the protein sample!(*102*)). Hence the water peak in the spectrum could be 100,000 time more intense than those derived from the sample. Early NMR experiments used a ~ 100 % D_2O solution but this also presents problems as many protons within a protein quickly exchange with the deuterium resulting in the loss of essential data from the spectrum. A variety of techniques have been developed to suppress the water peak. A variance on the WATERGATE method (*103*) was routinely used during this work. A non-selective RF pulse is applied to the sample followed by a strong gradient pulse which de-phases both solvent and solute magnetisation. A selective water 90 ° pulse is followed by a non-selective 180° pulse. A further 90 ° water pulse effectively leaves the water coherence where it is but inverts the coherence order of the solute magnetisation. A second gradient pulse de-phases the solvent magnetisation further but refocuses the solute magnetisation (*102*). This results in the near eradication of the water peak, and hence a significant increase in relative intensity of the solute resonances.

1.11.5 One Dimensional NMR

Whilst 1D NMR is very useful to determine structural data on small molecules, when dealing with protein spectra several problems arise, the main one being extensive overlapping of resonances. Unlike small molecules, proteins contain a much larger number of nuclear resonances which need to be assigned for structural studies. In a 1D ^1H protein spectrum the majority (if not all) of the proton chemical shifts will be overlapped preventing assignment (Figure 1-26). In order to overcome this problem it is necessary to record data sets in 2D and/or 3D, using a variety of different pulse sequences, in order for full assignment and structure determination to be carried out.

1.11.6 Two Dimensional NMR

The basic 2D pulse sequence, shown in Figure 1-27, contains four main components. During the preparation phase the spin systems are set up for the experiment, e.g. they are allowed to relax to equilibrium, and phase coherence is induced. The phase coherence is allowed to evolve for time t_1 during the evolution phase. During the mixing

period the phase coherence is transferred, using a pulse/series of pulses at the beginning of the mixing period and ended using one or more pulses at the end of the mixing period. The FID is then detected and recorded during the acquisition phase (*101*).

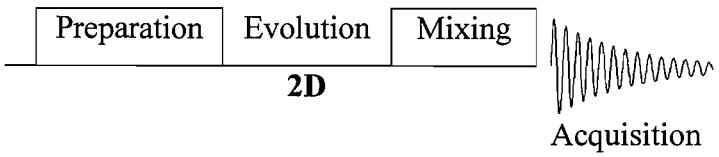


Figure 1-27 Two dimensional NMR experimental setup (*102*). During the preparation phase the spin systems allowed to relax to equilibrium, and phase coherence is induced. The phase coherence is allowed to evolve for time t_1 during the evolution phase. During the mixing phase the phase coherence is transferred, using a pulse/series of pulses at the beginning of the mixing period and ended using one or more pulses at the end of the mixing period. The FID is then detected and recorded during the acquisition phase.

The phase coherence of one group of spins within a molecule, complete with its phase information, may be passed onto another group. This second group does not need to be of the same element and therefore the phase coherence of a group of protons, sharing a Larmor frequency, found within a protein may be passed onto a group of other nuclei, either ^{13}C or ^{15}N, which have been incorporated into the molecule by a process known as isotopic labelling. This coherence transfer may occur even though the Larmor frequencies for the two groups of nuclei are not equal. This transfer is only possible if the spins of interest are coupled in some way, either by spin-spin coupling (scalar), where the nuclei of interest are joined/separated by a few covalent bonds, or between spins which, although not connected through bonds, are close in space (< 5 Å apart) and may continuously exchange their nuclear magnetisation through direct interactions between their dipole moments (dipolar couplings).

If a sample contains magnetically active nuclei, all with the same Larmor frequency (ν_1), then only a single peak will be seen on a 1D NMR spectrum. The application of a 90° RF pulse along the x-axis will tip the net magnetisation (M_z) into the +y direction, where it will begin to precess in the xy plane at the Larmor frequency. After time t_1 the M_{xy} will have passed through an angle of $2\pi\nu_1 t_1$. If the T_2 is ignored for a moment, there are two components of this magnetisation, $M_{xy}\cos(2\pi\nu_1 t_1)$ along the y` axis and $M_{xy}\sin(2\pi\nu_1 t_1)$ along the x axis. The application of a second 90° pulse, along the x` axis, results in the y` component being tilted into the –z axis, where it will decay to zero,

without contributing to the detected magnetisation. The x` component is unaffected and continues to precess, with the signal detected depending on v_1 and t_1. This means that the initial size of the signal detected after the 90° pulse is a label of the frequency of v_1 during time t_1. This signal is collected for a set acquisition time, t_2. The experiment is then rerun a number of times (2^n), incrementally increasing the value of t_1, with each FID digitised to produce a series of data points which are fed into a matrix where each row is one FID. If each FID is Fourier Transformed a series of 1D spectra are produced, representing a function varying with time t_2, and hence each spectrum is a plot of intensity against frequency (v_2). The size of the peak varies with t_1 depending upon M_{xy} $\sin(2\pi v_1 t_1)$ and therefore each column of data in the matrix represents a sinusoidal variation of frequency v_1. If this is plotted then it will appear in the form of another FID decaying at a rate of T_2. Therefore if FT is carried out on these column data a plot of frequency versus time is obtained. A 2D plot of the whole matrix will reveal a single peak at point v_2, v_1.

In this very simple case the plot will represent the measure of the Larmor frequency twice over. However if the nuclei are precessing at two different Larmor frequencies (v_1 and v_2) during time (t_1 and t_2) which may arise if the second 90° pulse caused changes in the spin system resulting in an altered precession frequency, or if it caused phase coherence which had been associated with one group of nuclei precessing at v_1 at time t_1 to be transferred to a second group of nuclei precessing at v_2, this would provide the detected FID at t_2. The resonances from these two sets of nuclei will appear in the 2D plot at v_2, v_1. The FID is collected at frequency v_2 but is also labelled with v_1. This is the basis of all 2D experiments although many different pulse sequences are available to produce a variety of 2D experiments, each providing different information.

1.11.7 TOCSY

In TOtal Correlation SpectroscopY (TOCSY) the mixing time is composed of a sequence of 180° pulses with a fast repetition rate share the phase coherence throughout the entire system of coupled spins, known as the spin system. At the end of the mixing time the FID is collected and the FT spectrum contains cross-peaks connecting every nucleus of the type excited within the spin system. The longer the mixing time the more detail of the spin system is seen with optimum mixing time revealing detail of the whole spin system. Correlated Sectroscopy (COSY) spectra only reveals details of two nuclei

separated by 3 bonds. The TOCSY is used in conjunction with the NOESY experiment, discussed below, to assign the chemical shifts of all the nuclei of interest within the molecule.

1.11.8 Nuclear Overhauser Effect Spectroscopy

Nuclear Overhauser Effect Spectroscopy (NOESY) relies upon nuclear interactions through space and provides information about nuclei close in space (<5 Å), an invaluable tool to the spectroscopist for analysing 3D protein structures. The effect is based on cross-relaxation, whereby the magnetisation is transferred through dipolar couplings between nuclei rather than by scalar couplings as in the TOCSY (*101*).

The NOESY pulse sequence (Figure 1-28) uses a third 90 ° pulse sequence to tip the z component of the magnetisation (tipped into the z plane by the second 90 ° pulse) back into the xy plane after a mixing time, tm.

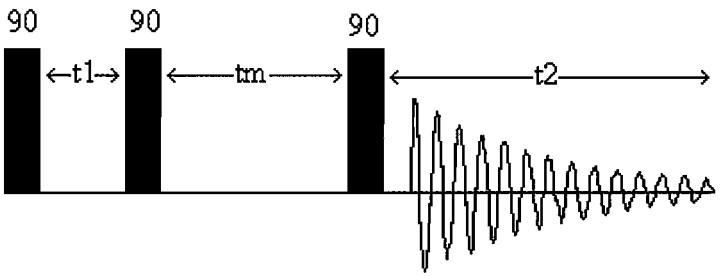


Figure 1-28 Example of a NOESY pulse sequence.

If the phase coherence of a group of spins, precessing at one frequency during t_1, were to migrate by cross-relaxation during t_m, then it would have a different frequency during t_2. This results in a cross-peak in the two dimensional NOESY, revealing the chemical shifts of the two sites. Therefore, if there is an NOE between two sets of spins, a cross-peak is produced using the NOESY pulse sequence. The NOESY experiment is designed to record only magnetisation that is longitudinal during the mixing time (*101*).

The NOESY experiment shows the strong distance dependence of cross-relaxation and this phenomenon is of great importance for three dimensional protein structure determination. Strong NOE cross-peaks are seen between nuclei of less than 2.5 Å and weaker NOEs between 2.5 Å and 3.5 Å. By extending the mixing time cross-peaks may be seen between nuclei up to 5 Å apart. For structure determination the NOE distance

restraints are grouped as either strong (1.8-2.5 Å), intermediate (1.8-3.5 Å) or weak (1.8-5.0 Å) (*101*).

1.11.9 Isotopic Labelling

To carry out most 3D NMR experiments and some 2D experiments it is necessary to incorporate magnetically active nuclei other than ^{1}H into the protein. To do this recombinant proteins are expressed in a minimal medium containing a ^{13}C and/or ^{15}N labelled chemical as the only carbon/nitrogen source. The expressed protein should then be uniformly labelled using these isotopes, enabling their analysis by NMR.

^{1}H-^{15}N HSQC

^{1}H-^{15}N-HSQC shows the correlation between proton and ^{15}N chemical shifts for the protein backbone amides. Every amino acid residue (except proline) has a unique H^{N}-^{15}N pair on its backbone. HSQC spectra correlate the proton and nitrogen chemical shifts and may be used as a reference, allowing for easier interpretation of more complex 3D spectra and use for dynamic and interaction studies. Initially RF pulses are applied in the proton Larmor frequency, the magnetisation is transferred to the labelled amide nitrogen via an INEPT (Insensitive Nuclei Enhanced by Polarisation Transfer) sequence and transferred back again immediately. Over the course of the experiment the time the magnetisation is allowed to remain on the nitrogen is incremented to allow for procession of the magnetisation on the nitrogen before being transferred back to the amide proton. FT of the 1D FID allows the 1D spectra produced to be converted into a second frequency dimension, giving rise to the HSQC spectrum.

1.11.10 Three Dimensional NMR

Although 2D NMR can be used to resolve larger molecular structures than 1D NMR there are still many problems with overlapping resonances. By separating the spectral data into a third dimension these overlaps can be further reduced. The 3D pulse sequence is made up of a combination of two 2D pulse sequences (Figure 1-29).

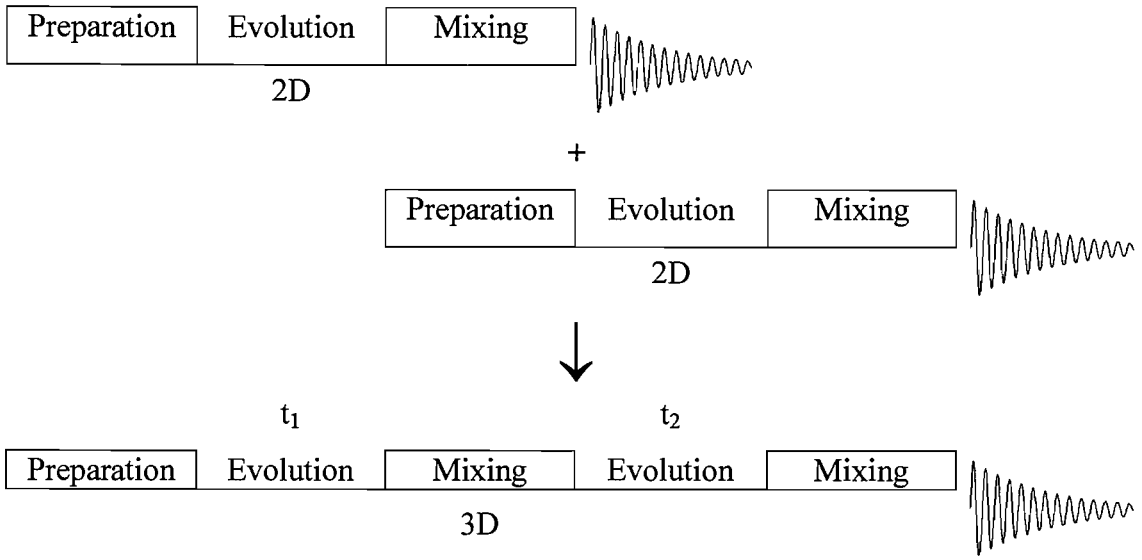


Figure 1-29 Generation of a three dimensional NMR experiment from a combination of two 2D experiments. A second evolution and mixing time are added to the pulse sequence in order to allow a third dimension of data to be recorded (*102*).

During a 3D ^{15}N HSQC- NOESY experiment all the protons within the sample are excited using a RF pulse and after a frequency labelling time (t_1) a second proton labelling pulse is followed by a mixing time as for a 2D NOESY. During the mixing time magnetisation is transferred via the Overhauser effect between protons which are close neighbours in space. A nitrogen frequency pulse transfers the magnetisation to the amide nitrogen and after a second labelling time t_2, during which the magnetisation is labelled with the ^{15}N frequency, a second nitrogen pulse transfers the magnetisation onto the amide proton. The magnetisation is the collected during t_3. The data collected is at the chemical shift of the amide protons, but is also labelled with the chemical shifts of the protons within close proximity and the nitrogen shifts. Three Fourier transforms will result in a stack of 2D NOESY spectra showing neighbours of the amide protons, with each layer of the stack distinguished by a ^{15}N chemical shift. A combination of ^{15}N HSQC-NOESY and ^{15}N HSQC-TOCSY spectra allows for sequential assignment of the protein by determining the chemical shifts of the proton and nitrogen nuclei within the protein (*101*). Samples labelled with ^{13}C and ^{15}N may be used to determine the chemical shifts of the carbon atoms found within the protein and a selection of complementary doubly labelled 3D experiments may be used to carry out sequential assignment and clarify any ambiguities arising from the ^{15}N labelled data-set and determining further NOE restraints.

1.11.11 Three Dimensional Protein Structure Generation

In order to solve a 3D protein NMR structure it is necessary to assign all the chemical shifts within the protein. This may be done using a variety of 2D and 3D spectra. Each residue has a characteristic spin system and the protons within this system need to be identified before the structure may be determined.

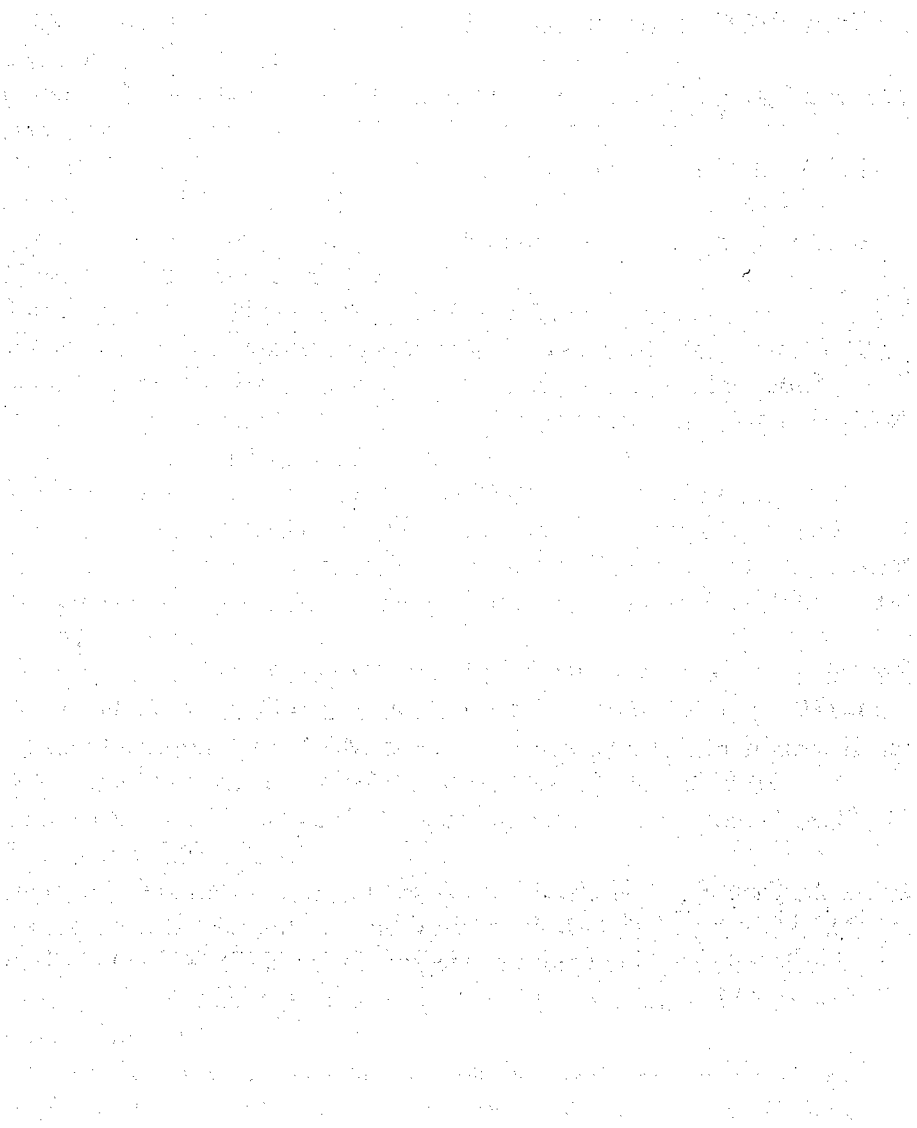
Using pairs of spectra, e.g. 3D TOCSY and 3D NOESY spectra sequential assignment is carried out. The TOCSY spectrum contains the spin systems for each residue on each 'slice' of the spectra corresponding with the ^{15}N chemical shift. By using the NOESY to look at cross peaks between the amide protons (and sometimes NOEs between the side chains) it is possible to 'walk along' the protein sequence, identifying connectivitys, and thus identifying not only the residue type but the actual residue within the protein sequence. The side chain resonances can then be assigned, due to their characteristic chemical shifts.

Once the chemical shift determination is complete the NOESY spectrum is used to identify connectivities through space. These link two protons which are close in space but not necessarily in the protein sequence. The intensity of the NOE is proportional to the distance separating the two protons and these distances are used during structure refinement as restraints for the folding of a structure-less polypeptide chain, of the same sequence as the protein, into a correctly folded 3D protein structure.

1.12 Aims of this research

The aim of this research is to elucidate the three dimensional structures of protein components from the early stages of the fatty acid and polyketide synthase pathways. A multi-disciplinary approach will be employed using a range of protein expression and purification methods and protein NMR and X-ray crystallographic techniques in order to solve the three dimensional protein structures. Targets include the FAS ACP from *Streptomyces coelicolor*, with the aim of comparing the structure with that of the polyketide ACP involved in actinorhodin production (Act ACP) from the same species. Since the two ACPs exist within the same organism but are involved in different pathways the structural differences which allow for the observed specificity will be examined. By solving the structure of *Gris* ACP from *Streptomyces griseus*, further

structural comparisons can take place. Analysis of the differences, and similarities, between ACPs from PKS systems within different species may be undertaken. The three dimensional structure of the KS/CLF complex from *Streptomyces coelicolor* will be the first PKS KS/CLF structure to be solved and, again, will allow comparisons with FAS system KS structures. Once these structures have been solved they may subsequently be used to identify the sites of interactions between the enzymes involved in both FA and PK synthesis, using a range of interaction titrations by NMR and/or co-crystallisation trials.



1.13 References

1. Staunton, J., and Weissman, K. J. (2001) *Nat. Prod. Rep. 18*, 380-416.
2. Finlay, A. C., Hobby, G. L., P'an, S. Y., Regna, P. P., Routien, J. B., Seeley, D. B., Shull, G. M., Sobin, B. A., Solomons, I. A., Vinson, J. W., and Kane, J. H. (1950) *Science 111*, 85.
3. Hochstein, F. A., Stevens, C. R., Conover, L. H., Regna, P. F., Pasternack, R., Gordon, P. N., Pilgrim, F. J., Brunings, K. J., and Woodward, R. B. (1953) *J. Am. Chem. Soc. 75*, 5455-5475.
4. Silver, L., and Bostian, K. (1990) *Eur. J. Clin. Microbiol. Infect. Dis. 9*, 455-461.
5. Konishi M, O. H., Matsumoto K, Tsuno T, Kamei H, Miyaki T, Oki T, Kawaguchi H, VanDuyne GD, Clardy J. (1989) *J Antibiot (Tokyo). 42*, 1449-1452.
6. Alberts, A. W. (1980) *Proc Natl Acad Sci U S A 77*, 3957-3961.
7. Burg, R. W., Miller, B. M., Baker, E. E., Birnbaum, J., Currie, S. A., Hartman, R., Kong, Y.-L., Monaghan, R. L., Olson, G., and Putter, I. (1979) *Antimicrob. Agents Chemother. 15*, 361–367.
8. Grove, J. F., MacMillan, J., Mulholland, T. P. C., and Rogers, M. A. T. (1952) *Structure. J. Chem. Soc.*, 3977-3987.
9. Vezina, C., Kudelski, A., and Sehgal, S. N. (1975) *J.Antibiot. 28*, 721-726.
10. Durous, J., and Suffness, M. (1981) *Cancer Treat. Rev. 8*, 63-69.
11. Calne, R. Y., Collier, D. S., Lim, S., Polard, S.G., Samaan, A., White, D. J. G., and Thiru, S. (1989) *Lancet II*, 227.
12. Simpson, T. J. (1995) *Chemistry & Industry*, 407-411.
13. Murata M, N. H., Iwashita T, Matsunga S, Sasaki M, Yokoyama A, Yasumoto T. (1993) *Journal of the American Chemical Society 115*, 2060-2062.
14. Hopwood, D. A., and Sherman, D. H. (1990) *Annu. Rev. Genet. 24*, 37-66.
15. Wakil, S. J. (1981) *Biochemistry 28*, 4523.
16. Sherman, D. H., and Smith, J. L. (2006) *ACS Chemical Biology 1*, 505-509.
17. Spencer, J. B., and Jordan, P. (1992) *Biochemical Journal 228*, 869.
18. Taguchi, T., Itou, K., Ebizuka, Y., Malpartida, F., Hopwood, D. A., Surti, C. M., Booker-Milburn, K. I., Stephenson, G. R., and Ichinose, K. (2000) *J. Antibiot. 53*, 144-52.
19. Bystrykh, L. V., Fernandez-Moreno, M. A., Herrema, J. K., Malpartida, F., Hopwood, D. A., and Dijkhuizen, L. (1996) *J. Bacteriol. 178*, 2238-44.
20. Fernandez-Moreno, M. A., Martinez, E., Caballero, J. L., Ichinose, K., Hopwood, D. A., and Malpartida, F. (1994) *J. Biol. Chem. 269*, 24854-63.
21. Caballero, J. L., Martinez, E., Malpartida, F., and Hopwood, D. A. (1991) *Mol. Gen. Genet. 230*, 401-12.
22. Sciara, G., Kendrew, S. G., Miele, A. E., Marsh, N. G., Federici, L., Malatesta, F., Schimperna, G., Savino, C., and Vallone, B. (2003) *Embo. J. 22*, 205-15.
23. Hesketh A, C. K. (2003) *J Ind Microbiol Biotechnol. 30*, 523-529.
24. Ichinose, K., Taguchi, T., Bedford, D. J., Ebizuka, Y., and Hopwood, D. A. (2001) *J. Bacteriol. 183*, 3247-50.
25. Ichinose, K., Surti, C., Taguchi, T., Malpartida, F., Booker-Milburn, K. I., Stephenson, G. R., Ebizuka, Y., and Hopwood, D. A. (1999) *Bioorg. Med. Chem. Lett. 9*, 395-400.

26. Jez. J.M. & Noel, J. P. (2000) *J. Biol. Chem. 275*, 39640-39646.
27. Jez, J. M., Austin, M. B., Ferrer, J., Bowman, M. E., Schroder, J., and Noel, J. P. (2000) *Chem. Biol. 7*, 919-30.
28. Austin, M. B., and Noel, J. P. (2003) *Nat. Prod. Rep. 20*, 79-110.
29. Crump, M. P., Crosby, J., Dempsey, C. E., Parkinson, J. A., Murray, M., Hopwood, D. A., and Simpson, T. J. (1997) *Biochemistry 36*, 6000-8.
30. Arthur, C., Szafranska, A., Evans, S. E., Findlow, I. S., Burston, S. G., Owen, I. C. L., Simpson, T. J., Crosby, J. and Crump, M. P. (2005) *Biochemistry 44*, 15414-15421.
31. Arthur, C., Szafranska, A., Long, J., Mills, J., Cox, R., Findlow, I. S., Simpson, T. J., Crump, M. P. and Crosby, J. (2006) *Chem. Biol. 13*, 587-596.
32. Rock, C. O., and Cronan, J. E. (1996) *Biochim. Biophys. Acta. 1302*, 1-16.
33. Smith, S. (1994) *Faseb. J. 8*, 1248-59.
34. Elovson, J., and Vagelos, P. R. (1968) *J. Biol. Chem. 243*, 3603-11.
35. Lambalot, R. H., and Walsh, C. T. (1995) *J. Biol. Chem. 270*, 24658-61.
36. Flugel, R. S., Hwangbo, Y., Lambalot, R. H., Cronan, J. E., Jr., and Walsh, C. T. (2000) *J. Biol. Chem. 275*, 959-68.
37. Lambalot, R. H., Gehring, A. M., Flugel, R. S., Zuber, P., LaCelle, M., Marahiel, M. A., Reid, R., Khosla, C., and Walsh, C. T. (1996) *Chem. Biol. 3*, 923-36.
38. Majerus, P. W., Alberts, A. W., and Vagelos, P. R. (1965) *J. Biol. Chem. 240*, 4723-6.
39. Keating, D. H., Carey, M. R., and Cronan, J. E., Jr. (1995) *J. Biol. Chem. 270*, 22229-35.
40. Revill, W. P., Bibb, M. J., and Hopwood, D. A. (1995) *J. Bacteriol. 177*, 3946-52.
41. Matharu, A. L., Cox, R. J., Crosby, J., Byrom, K. J., and Simpson, T. J. (1998) *Chem. Biol. 5*, 699-711.
42. Hitchman, T. S., Crosby, J., Byrom, K. J., Cox, R. J., and Simpson, T. J. (1998) *Chem. Biol. 5*, 35-47.
43. Dreier, J., Li, Q., and Khosla, C. (2001) *Biochemistry 40*, 12407-11.
44. Jackowski, S., and Rock, C. O. (1983) *J. Biol. Chem. 258*, 15186-91.
45. Thompson, J. D., Gibson,T.J., Plewniak,F., Jeanmougin,F. and Higgins,D.G. (1997) *Nucleic Acids Research 24*, 4876-4882.
46. Holak, T. A., and Prestegard, J. H. (1986) *Biochemistry 25*, 5766-74.
47. Holak, T. A., Frederick, A. F., and Prestegard, J. H. (1987) *J. Biol. Chem. 262*, 3685-9.
48. Holak, T. A., Kearsley, S. K., Kim, Y., and Prestegard, J. H. (1988) *Biochemistry 27*, 6135-42.
49. Holak, T. A., Nilges, M., Prestegard, J. H., Gronenborn, A. M., and Clore, G. M. (1988) *Eur. J. Biochem. 175*, 9-15.
50. Andrec, M., Hill, R. B., and Prestegard, J. H. (1995) *Protein Sci. 4*, 983-93.
51. Jones, P. J., Cioffi, E. A., and Prestegard, J. H. (1987) *J. Biol. Chem. 262*, 8963-5.
52. Jones, P. J., Holak, T. A., and Prestegard, J. H. (1987) *Biochemistry 26*, 3493-500.
53. Kim, Y., and Prestegard, J. H. (1989) *Biochemistry 28*, 8792-7.
54. Kim, Y., and Prestegard, J. H. (1990) *Proteins 8*, 377-85.
55. Mayo, K. H., Tyrell, P. M., and Prestegard, J. H. (1983) *Biochemistry 22*, 4485-93.
56. Mayo, K. H., and Prestegard, J. H. (1985) *Biochemistry 24*, 7834-8.
57. Crump, M. P., Crosby, J., Dempsey, C. E., Murray, M., Hopwood, D. A., and Simpson, T. J. (1996) *FEBS Lett. 391*, 302-6.
58. Kraulis, P. (1991) *Journal of Applied Crystallography 24*, 946-950.

59. Koradi, R., Billeter, M., and Wüthrich, K. (1996) *J. Mol. Graphics. 14*, 51-55.
60. Holak, T. A., Engstrom, A., Kraulis, P.J., Lindenberg, G., Bennich, H., Jones, T.A. & Gronenborn, A.M. (1988) *Biochemistry 27*, 7620-7629.
61. Findlow SC, W. C., Simpson TJ, Crosby J, Crump MP. (2003) *Biochemistry 42*, 8423-33.
62. Johnson, N. J. (2001).
63. Li, Q., Khosla, C., Puglisi, J.D. and Liu, C.W. (2003) *Biochemistry 42*, 4648-4657.
64. Xu, G. Y., Tam, A., Lin, L., Hixon, J., Fritz, C. C., and Powers, R. (2001) *Structure (Camb) 9*, 277-87.
65. Wong, H. C., Liu, G., Zhang, Y. M., Rock, C. O., and Zheng, J. (2002) *J. Biol. Chem. 277*, 15874-80.
66. Parris, K. D., Lin, L., Tam, A., Mathew, R., Hixon, J., Stahl, M., Fritz, C. C., Seehra, J., and Somers, W. S. (2000) *Structure Fold. Des. 8*, 883-95.
67. Zhang, Y. M., Rao, M. S., Heath, R. J., Price, A. C., Olson, A. J., Rock, C. O., and White, S. W. (2001) *J. Biol. Chem. 276*, 8231-8.
68. Gong, H., and Byers, D. M. (2003) *Biochem. Biophys. Res. Commun. 302*, 35-40.
69. Zhang YM, W. B., Zheng J, Rock CO. (2003) *J Biol Chem. 278*, 52935-43.
70. Worsham, L. M., Earls, L., Jolly, C., Langston, K. G., Trent, M. S., and Ernst-Fonberg, M. L. (2003) *Biochemistry 42*, 167-76.
71. Huag, W., Jia, J., Edwards, P., Dehesh, K., Schneider, G. & Lindqvist, Y. (1998) *Embo. J. 17*, 1183-1191.
72. Keatinge-Clay, A. T., Maltby, D. A., Medzihradsky, K. F., Khosla, C. and Stroud, R. M. (2004) *Nature Structural & Molecular Biology 11*, 888-893.
73. Garwin, J. L., Klages, A. L., and Cronan, J. E., Jr. (1980) *J. Biol. Chem. 255*, 11949-56.
74. de Mendoza, D., Klages Ulrich, A., and Cronan, J. E., Jr. (1983) *J. Biol. Chem. 258*, 2098-101.
75. Kauppinen, S., Siggaard-Andersen, M., and von Wettstein-Knowles, P. (1988) *Carlsberg. Res. Commun. 53*, 357-70.
76. Clough, R. C., Matthis, A. L., Barnum, S. R., and Jaworski, J. G. (1992) *J. Biol. Chem. 267*, 20992-8.
77. Tsay, J. T., Oh, W., Larson, T. J., Jackowski, S., and Rock, C. O. (1992) *J. Biol. Chem. 267*, 6807-14.
78. Garwin, J. L., Klages, A. L., and Cronan, J. E., Jr. (1980) *J. Biol. Chem. 255*, 3263-5.
79. Ohashi, Y., and Okuyama, H. (1986) *Biochim. Biophys. Acta. 876*, 146-153.
80. Olsen, J. G., Kadziola, A., von Wettstein-Knowles, P., Siggaard-Andersen, M., Lindquist, Y., and Larsen, S. (1999) *FEBS Lett. 460*, 46-52.
81. Davies, C., Heath, R. J., White, S. W., and Rock, C. O. (2000) *Structure Fold. Des. 8*, 185-95.
82. Qiu, X., Janson, C. A., Konstantinidis, A. K., Nwagwu, S., Silverman, C., Smith, W. W., Khandekar, S., Lonsdale, J., and Abdel-Meguid, S. S. (1999) *J Biol Chem 274*, 36465-71.
83. Moche, M., Dehesh, K., Edwards, P., and Lindqvist, Y. (2001) *J. Mol. Biol. 305*, 491-503.
84. Scarsdale, J. N., Kazanina, G., He, X., Reynolds, K. A., and Wright, H. T. (2001) *J. Biol. Chem. 276*, 20516-22.
85. Mathieu, M., Zeelen, J., Pauptit, R.A., Erdmann, R., Kanau, W-H. & Wierenga, R.K. (1994) *Structure 2*, 797-808.

86. Huang, W., Jia, J., Edwards, P., Dehesh, K., Schneider, G., and Lindqvist, Y. (1998) *Embo. J. 17*, 1183-91.
87. Olsen, J. G., Kadziola, A., von Wettstein-Knowles, P., Siggaard-Andersen, M., and Larsen, S. (2001) *Structure (Camb) 9*, 233-43.
88. Qiu, X., Janson, C. A., Konstantinidis, A. K., Nwagwu, S., Silverman, C., Smith, W. W., Khandekar, S. K., Lonsdale, J., Abdel-Meguid, S. S. (1999) *J.Biol.Chem. 274*, 36465-36471.
89. Price, A. C., Choi, K. H., Heath, R. J., Li, Z., White, S. W., and Rock, C. O. (2001) *J. Biol. Chem. 276*, 6551-9.
90. Bisang, C., Long, P. F., Cortes, J., Westcott, J., Crosby, J., Matharu, A. L., Cox, R. J., Simpson, T. J., Staunton, J., and Leadlay, P. F. (1999) *Nature 401*, 502-5.
91. Shen, B., Summers, R.G., Wendt-Pienkowski, E. & Hutchinson, C.R. (1995) *J. Am. Chem. Soc. 117*, 6811-6821.
92. Kramer, P. J., Zawada, R.J.X., McDaniel, R., Hutchinson, C.R., Hopwood, D.A. & Khosla, C. (1997) *J. Am. Chem. Soc. 119.*
93. Witkowski, A., Joshi, A. K., Lindqvist, Y., and Smith, S. (1999) *Biochemistry 38*, 11643-50.
94. Kendrew, J. C., and Perutz, M. F. (1957) *Annu. Rev. Biochem. 26.*
95. Purcell, E. M., Torrey, H. C. & Pound, R. V. (1946) *Phys. Rev. 60*, 37-38.
96. Williamson, M. P., Havel, T.F. and Wuthrick, K. (1985) *J. Mol. Biol. 182*, 295-315.
97. Tugarinov, V., Muhandiram, R., Ayed, A., and Kay, L. E. (2002) *J. Am. Chem. Soc. 124*, 10025-10035.
98. Rhodes, G. (2000) *Crystallography made crystal clear.*
99. Drenth, J. (1994) *Principles of Protein X-ray Crystallography*, Springer-Verlag, New York.
100. Giacovazzo, C. (1992) *Fundamentals of Crystallography*, Oxford University Press, NEW YORK.
101. Rattle, H. (1995) *An NMR Primer for Life Scientists.*, Partnership Press., Fareham, UK.
102. Cavanagh, J., Fairbrother, W.J., Palmer III, A.G. and Skelton, N.J. (1996) *Protein NMR spectroscopy. Principles and practice.*, Academic Press, London.
103. Piotto, M., Saudek, V. and Sklenar, V. (1992) *J. Biomol. NMR 2*, 611-665.

Chapter 2

Materials and Methods

2 Materials and Methods

2.1 Chemicals and Reagents

All reagents and chemicals were obtained from Sigma-Aldrich Chemical Company LTD, Dorset, England or Fisher Scientific UK LTD, Leicestershire, England unless otherwise stated below.

DIFCO Bactotryptone
NOVAGEN Competent cells
Melford Laboratories, Ipswich, England Isopropyl-β-D-thiogalactosidase (IPTG), Ampicillin sodium
PROMEGA Restriction enzymes, Plasmids, DNA purification kits
Molecular Dimensions Crystal screens MD1-01 and MD1-02, 24 well Linbro plates
Invitrogen SDS-PAGE molecular weight markers
JenaBioscience JBScreen Mixed

2.2 Microbiological Techniques

2.2.1 Sterilisation

All media, Eppendorfs, tips and flasks used during the course of this work were sterilised using a British Steriliser Autoclave at 120 °C, 15 psi for 20 minutes and once autoclaved, tips and Eppendorfs were dried at ~60 °C for 24 hours. Any solutions containing heat labile substances were sterilised by filtration through 0.2 μm Millipore filters prior to addition to growth media.

2.2.2 Media

Recombinant ACP samples were grown and expressed in Lauria Broth (LB) (*1*). Labelled ACP samples were grown initially in LB and transferred to M9 (*1*) minimal media for expression. KS/CLF was grown and expressed in Super YEME media (*2*). Recipes for each of these three medias are shown below:

Lauria Broth

Bactotryptone 10 g/L	Glycerol 2 mL/L	$MgSO_4.7H_2O$ 0.25 g/L
NaCl 10 g/L	Yeast extract 5 g/L	

The media was adjusted to pH 7.5 using either concentrated HCl or 10 M NaOH and then autoclaved.

M9 Minimal Media

Na_2HPO_4 6 g/L	KH_2PO_4 3 g/L	$MgSO_4.7H_2O$ 0.5 g/L
NaCl 0.5 g/L	NH_4Cl 1 g/L	$CaCl_2.2H_2O$ 15 mg/L

Adjusted to pH 7.4 using either concentrated HCl or 10 M NaOH and then autoclaved. Post autoclaving, 2 g/L glucose was added as a sterile filtered aqueous solution. When ^{15}N and/or ^{13}C labelling proteins then $^{15}NH_4Cl$ and ^{13}C Glucose were used respectively, substitute for their unlabelled equivalents.

Sterile filtered antibiotic (ampicillin) was added prior to inoculation to a final concentration of 100 μg/mL.

Agar plates were prepared by the addition of 1.5 g of agar to 100 mL of required media prior to being autoclaved. Once the media cooled to ~40 °C the required antibiotic was added as above, and the solution poured into sterile plastic Petri dishes to a depth of ~5 mm. Once set these plates were stored at 4 °C for up to one month before use.

2.2.3 Bacterial strains

The following bacterial strains were used as hosts for the growth of the different proteins used during the course of this work. Competent cell lines were purchased from Novagen and stored at –70 °C and are shown in Table 2-1.

Strain	Genotype	Protein expressed using this strain
BL21 DE3	f⁻ *ompT* *hsdS*$_B$ (r^-_B m^-_B) *gal dcm* (DE3)	FAS ACP *gris* ACP

Table 2-1 Competent cell lines used during this work, which were obtained from Novagen.

2.3 DNA Techniques

2.3.1 Plasmids

apo FAS ACP from *Streptomyces coelicolor* was expressed using pET11c plasmids obtained from PROMEGA (*3*).

2.3.2 Agarose Gel Electrophoresis

Agarose gels were used to separate insert and plasmid DNA following enzymatic digests. Gels of between 1 % and 3 % w/v were used depending upon the size of the plasmid and/or the DNA insert. Gels were prepared using 50 mL 1xTAE buffer (50x TAE buffer stock: 242 g Tris base, 57.1 mL glacial acetic acid, 100 mL 0.5 M EDTA, pH 8.0 made up to 1 L) plus the required percentage of agarose. The solution was boiled using a microwave, and ethidium bromide added to a final concentration of 0.3 μg/mL, once the solution had cooled to ~50 °C. Gels were then poured and once set they were run horizontally, submerged in a tank containing 1x TAE buffer. DNA samples were prepared in water with the addition of Promega 6x blue/orange gel loading buffer, and ~5 ng of DNA loaded onto the gel. 1 kb and 100 bp DNA ladders, obtained from Promega, were run on each gel. Gels were run at 100 V until full separation was obtained. Fluorescence bands were visualized using an UV transluminator.

2.3.3 Extraction of DNA from Agarose Gel

DNA inserts were extracted from agarose gels using Bio 101 Gene Clean Kits, leaving the DNA in solution in water.

2.3.4 Restriction Digests

Plasmid DNA for FAS ACP from *Streptomyces coelicolor* was digested following PROMEGA protocols using BamHI and NdeII restriction enzymes from PROMEGA, and PROMEGA restriction buffers D and E, in order to remove the DNA insert and/or to open the plasmid in order to allow ligation to take place.

2.3.5 Plasmid DNA Preparation

Plasmids containing a DNA insert were replicated in *E. coli* strains grown in LB media, 10 mL of which was centrifuged and the plasmid DNA extracted from the pellet using Wizard®Plus SV Minipreps DNA purification system from PROMEGA. Following restriction digests the DNA insert was removed and re-ligated into a new plasmid of choice as described in section 2.3.6.

2.3.6 DNA Ligations

Plasmid DNA was digested with PROMEGA restriction enzymes which complement those sites found within the required DNA insert. The plasmid and insert were then mixed together with a PROMEGA ligase and left at 4 °C over night, following PROMEGA methods, for ligation to complete.

2.3.7 Transformation of Plasmid DNA into Bacterial Strains

20 μL of competent cells were placed on ice and treated with ~2 μL of plasmid (~2-20 ng). This was left on ice for 30 minutes before being heat-shocked at 42 °C for 30 seconds. The cells were placed back on ice for approximately five minutes and 80 μL of SOC media added (supplied by Novagen). The resulting culture was then spread onto agar plates containing the required antibiotic. The plates were then incubated at 37 °C over night. The next day two colonies per plate were selected and streaked onto new agar plates for further use.

2.4 Protein Techniques

2.4.1 Polyacrylamide Gel Electrophoresis

Sample preparation

Protein Samples

Protein samples were boiled for 5 minutes with an equal volume of gel loading buffer (560 mM Tris.HCl pH 6.8, 100 mM β-mercaptoethanol, 2 % SDS, 0.1 % bromophenol and 10 % glycerol).

Bacterial samples

1mL aliquots of bacterial culture were spun at 13,000 rpm for 5 minutes. The resulting pellet was resuspended in 20 μL water, 20 μL 1x gel loading buffer and 1.5 μL β-mercaptoethanol, boiled for 5 minutes and spun at 13,000 rpm for 15 minutes to remove cell debris. 20 μL of supernatant was then loaded onto the gel.

Samples should ideally be loaded at a concentration of 20 ng/well.

Gel preparation

Gels were poured by hand using a Bio Rad Mini-Protean II Electrophoresis cell and were prepared as shown in Table 2-2 (*4*).

Reagent	Stacking Gel	Running Gel
Acrylamide	0.66 mL	3.33 mL
dH_2O	2.90 mL	2.03 mL
Glycerol	n/a	1.30 mL
3 M Tris, 0.3 % SDS pH 8.45	1.24 mL	3.33 mL
TEMED	10 µL	10 µL
100 mg/ml Ammonium persulphate	25 µL	50 µL

Table 2-2 Composition of stacking and running gel components used in SDS PAGE analysis.

3.3 mL running gel was poured and topped with a layer of water until set. The water was removed, stacking gel poured onto the running gel and the comb fitted leaving 1 cm of gel at the base of the comb. Once the gel was set the samples were loaded into the wells (2-30 µL depending on sample concentration) and the gel run, with separate cathode and anode buffers, at 150 V for 45 minutes (cathode buffer: 100 mM tris, 100 mM tricine, 0.1 % SDS pH 8.2, anode buffer: 200 mM tris pH 8.9). Invitrogen MultiMark®Multi-Coloured Standards were run on all gels. Once run, gels were stained for half an hour with rocking (stain: 0.25 g coomasie brilliant blue R250, 45 mL methanol, 45 mL water and 10 mL glacial acetic acid). Gels were destained completely overnight (destain: 45 mL methanol, 45 mL water and 10 mL glacial acetic acid) and then photographed.

2.4.2 Protein Concentration Assays

Protein concentrations were determined by A280 assay.

A280

The absorbance of a protein solution can be measured at 280 nm and the concentration calculated from its extinction coefficient (ε), which can be roughly determined from its protein sequence or calculated experimentally. $A280 = \varepsilon.c.l$ where c is the protein concentration and l is the path length of the sample.

2.4.3 Protein Purification Protocols

All protein samples were purified using a Pharmacia Biotech FLPC system and Pharmacia Biotech columns. All buffers were filtered through a 0.45µM filters as were protein solutions prior to loading onto the columns.

2.4.4 Acyl Carrier Protein Expression and Purification

ACP expression

E. coli strains, transformed with the relevant ACP encoding plasmid, were grown in LB media containing 100 μg/mL ampicillin, at 200 rpm and 37 °C, to an absorbance of ~0.7 at 600 nm. For unlabeled samples expression was then induced using a final concentration of 1 mM IPTG at 37 °C for three hours. In the case of labelled samples the LB culture was spun at 8,000 rpm for 20 minutes and the pellet resuspended in labelled M9 media containing 100 μg/mL ampicillin and the relevant isotopic label (^{15}N ammonium phosphate and/or ^{13}C glucose). This was then induced using a final concentration of 1 mM IPTG and grown at 37 °C for 18 hours. Once expression was complete the culture was spun at 8,000 rpm for 20 minutes and the resulting pellet resuspended in either 50 mM phosphate or tris buffer at pH 8 and frozen at –20 °C.

ACP purification

Frozen pellets containing the expressed ACPs were defrosted and then flash frozen using a dry-ice and acetone bath. Once refrozen the sample was slowly defrosted by the addition, with stirring, of 10 mM sodium phosphate buffer pH 8. The resulting cloudy solution was centrifuged at 8,000 rpm for 20 minutes. The supernatant then underwent a 60 % ammonium sulphate cut (390 g/L) before being centrifuged again at 8,000 rpm for 20 minutes. A 90 % cut was then carried out on the supernatant with the addition of a further 219 g/L ammonium sulphate, the solution adjusted to pH 4 and the sample left at 4 °C overnight. The following day the sample was spun at 18,000 rpm for 20 minutes and the pellet resuspended in 50 mM sodium phosphate buffer pH 8. This protein solution was then passed down a Sephadex G25 superfine column attached to an Amersham Pharmacia Biotech FPLC to remove salt. The protein fraction obtained from this column was then applied to a 26/10 Q Sepharose high performance anionic exchange column, before elution using a sodium chloride gradient of 0-1 M applied over 400 mL.

2.4.5 ACP Protein Characterisation and Purity Assessment

SDS PAGE was performed on all protein peaks detected by FPLC, and those fractions which contained the required protein were pooled and freeze-dried. A final desalt was carried out using a PD10 Sephadex G25M column before samples were taken for characterisation by mass spectrometry to confirm the molecular weight and percentage label incorporation where necessary.

2.4.6 KS/CLF

Protein was expressed, purified and provided by Dr Tom Nichols at the University of Bristol.

2.4.7 Minimal PKS assay

50 μM *holo act* ACP was reduced overnight at 30 °C with 10 μM DTT. Minimal PKS assays were carried out at 30 °C for two hours under the following conditions:

1 μM KS/CLF

50 μM reduced ACP

1 mM DTT

50 mM EDTA

100 mM sodium phosphate buffer, pH 7.3, 10 % glycerol (assay buffer)

1mM malonyl-CoA

After the reaction 100 μL assay buffer and 100 mg NaH_2PO_4 were added to the sample and vortex mixed. 400 μL ethylacetate was added and the sample vortex mixed and spun at 13,000 rpm for 2 minutes. The organic layer was removed to a clean Eppendorf and a further 400 μL of ethylacetate added to the aqueous layer. The sample was again vortex mixed and spun at 13,000 rpm for 2 minutes. The organic layers were combined and the sample dried under nitrogen. The resulting pellet was resuspended in 100 μL HPLC grade methanol and applied to a C18 column. The sample was run as described in Table 2-3, with the production of SEK4 (~20-22 min) indicating active KS/CLF.

Time (min)	%B
0	0
5	0
35	75
37	100
40	100
45	0
50	0

Table 2-3 Parameters for mPKS assay using C18 column. (A=H_2O, 0.04 % TFA, B=Acetonitrile, 0.05 % TFA).

2.5 Crystallographic Techniques

2.5.1 Protein Preparation

All crystallisation trials were carried out using buffered protein solutions as discussed in the relevant chapters.

2.5.2 Crystal Screen Preparations

24-well Linbro plates were greased around the top of each well using Dow Corning high vacuum grease. Glass cover slides were coated using dimethyldichlorosilane solution and dried before use. All crystal screen stocks were filtered using 0.2 μm filters to prevent bacterial/fungal growth, and all buffers were prepared as 1 M stocks at the required pH as with commercially available solutions from Molecular Dimensions. Salt and precipitant components were similarly prepared as concentrated stock solutions to allow for dilution within the wells to the required concentration with filtered distilled water.

2.5.3 Crystal Screens

Initial crystal screens were carried out using Molecular Dimension MD1-01 and MD1-02 screens and Jena Bioscience screens and hanging drop methodology, Figure 2-1. During these trials 1mL of mother liquor was used in the wells (700 μL for Jena Bioscience screens) and typically drops contained 2 μL protein solution and 2 μL mother liquor, with trays being kept at room temperature (~17 °C) in the dark. Wells yielding crystals reported from these conditions were refined using a combination of both grid screens and full factorial screens, using solutions prepared in house, in order to improve and refine the crystals produced. Examples of how grid and full factorial screens were laid out are shown below in Figure 2-2.

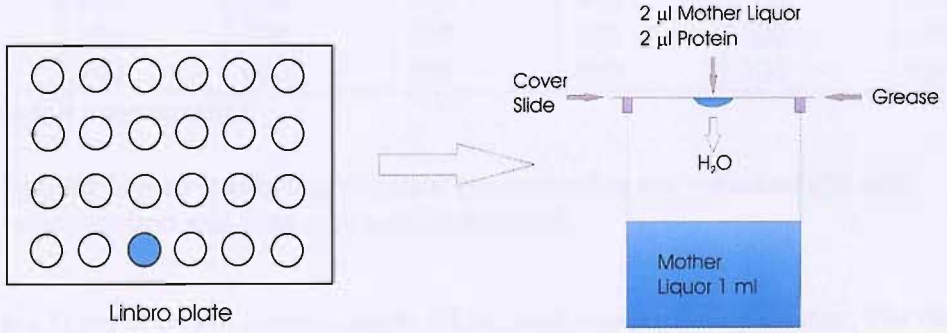


Figure 2-1 Hanging drop method used to crystallise purified protein. A drop of protein solution and mother liquor is placed on a cover slip which is sealed over the well containing the mother liquor. Water from the drop diffuses into the well solution, concentrating protein within the drop, and causes crystalisation of the protein if the conditions are favourable.

Precipitant concentration	pH X (µL)		pH Y (µL)		Drop composition
Concentration A (µL)	100 300 100 500	100 300 500	100 300 100 500	100 300 500	2 µL protein 2 µL mother liquor
	100 300 100 500	100 300 500	100 300 100 500	100 300 500	2 µL protein 4 µL mother liquor
Concentration B (µL)	100 400 100 400	100 400 500	100 400 100 400	100 400 500	2 µL protein 2 µL mother liquor
	100 400 100 400	100 400 500	100 400 100 400	100 400 500	2 µL protein 4 µL mother liquor
	+ Salt	- Salt	+ Salt	- Salt	

a) Full factorial screen- pH, precipitant concentration, salt concentration and protein concentration/drop size may all be varied together in this screen.

	Increasing precipitant concentration (µL) →					
Increasing pH (µL) ↓	100 100 100 700	100 200 100 600	100 300 100 500	100 400 100 400	100 500 100 300	100 600 100 200
	100 100 100 700	100 200 100 600	100 300 100 500	100 400 100 400	100 500 100 300	100 600 100 200
	100 100 100 700	100 200 100 600	100 300 100 500	100 400 100 400	100 500 100 300	100 600 100 200
	100 100 100 700	100 200 100 600	100 300 100 500	100 400 100 400	100 500 100 300	100 600 100 200

Constant salt concentration

b) Grid Screen- pH and precipitant concentration are varied whilst salt concentration and drop size remain constant.

Figure 2-2 Types of crystal screens used to refine conditions producing crystals. The full factorial screen (a) varies precipitant, protein and salt concentration and pH and drop size/protein concentration. The grid screen (b) varies precipitant concentration and pH whilst retaining the same drop size and salt and protein concentration. Precipitant volume is shown in green, pH buffer volume shown in red, salt volume shown in black and added water volume shown in blue.

2.5.4 Data Collection.

Diffraction data were collected either at the ESRF synchrotron (Grenoble, France) or using the in-house Enraf-Nonius CuKα-rotating anode generator/MAR scanner data collection system. Crystals were either wet mounted in a capillary tube at room temperature, or cryo-cooled in each case to 100 K, with the addition of cryoprotectant, using an Oxford Cryosystem cooler.

2.5.5 Data Processing

The following series of programs were used to interpret the diffraction data in this work:

IPDISP
Displays diffraction patterns obtained from crystallographic data collection
⇩
MOSFLM
Auto-indexing and integration of spot intensities
⇩
SORTMTZ
Sorts MTZ reflection files into ascending/descending h,k,l values
⇩
SCALA
Scales together and merges multiple observations of reflections
⇩
HKLVIEW
Displays zones of reciprocal space as pseudo-precession images
⇩
TRUNCATE
Obtains structure factor amplitudes for the processed data
⇩

MOLREP*	BEAST*	AMORE*	EPMR* (*5*)
Automated molecular replacement program	Automated molecular replacement program	Automated molecular replacement program	6 D Molecular replacement

* Provides estimates for initial phase information.

The programs IPDISP, MOSFLM (*6, 7*), SORTMTZ, SCALA (*8*), HKLVIEW, TRUNCATE (*9*), MOLREP (*10*) , MTZ2VARIOUS, BEAST (*11-13*) and AMORE (*14-18*) are all part of the CCP4 (*19*) suite.

2.5.6 Programs used during X-ray crystallography protein structure determination.

IPDISP

The diffraction images were first viewed using IPDISP, an image display program, which provides a reliable initial estimate for the beam centre of the images belonging to a given set of diffraction data.

MOSFLM

Once the beam centre and resolution had been determined, MOSFLM was used to process the data. This program performs a series of processes in order to index and integrate the spot intensities as shown below:

Auto-indexing

⇩

Preliminary refinement of cell parameters and crystal orientation

(matrix optimisation)

⇩

Positional refinement

⇩

Generation of standard profiles and profile-fitting

⇩

Integration of spot intensities

In order to auto-index, the program picks approximately 100-500 of the most intense spots on one image and a reduced cell is generated from a vector histogram and a distortion index table is produced. This distortion index table is a measure of how much the other Bravis lattices must be distorted to fit the reduced P1 cell. This distortion index table gives a variety of different possible cell dimensions along with possible space groups. One of these is then selected on the basis of (low) distortion penalty and (high) symmetry considerations. It is then used in the processing of the whole data set. The cell and space group is chosen with caution, and may be confirmed later by processing in a lower symmetry space group and analysing the pattern of data using HKLVIEW.

It is important to gain reliable values for the crystal orientation and unit cell dimensions. These must be refined before processing the entire data set. Three different sets of images (five in each) are used for this initial refinement. These sets were 30°, 60° and 90° from the starting data collection angle. Whilst refining the data it is important to watch both the RMS residual and the weighted residual. The RMS residual should ideally be 0.04-0.06 mm and shows how close the spot predictions are to the actual positions of the spots in the diffraction data. The weighted residual should be <1.0. Larger values of these two parameters suggest errors in crystal orientation or cell parameters.

After this interactive refinement process MOSFLM is run automatically. This involves performing positional refinement, generation of standard profiles and profile-fitting, and spot intensity integration for the entire data set.

During positional refinement, MOSFLM varies the detector, beam and crystal parameters in such a way as to refine the fit of the predicted positions to the actual diffraction spots. This process continues with each new image in the data set, and so the refinement improves. The crystal, detector and beam parameters are all refined simultaneously with the predicted spot positions.

Each diffraction pattern is then split into a grid of squares and a standard profile for the spots in each of these sections is created. For a given spot, its standard profile is then adjusted to a unique shape depending on its physical distance from each of its neighbouring grid boxes. The standard profile from a neighbouring grid box, which is close to the spot in question, will have a strong influence on the spot's individual shape, whereas that from a more distant grid box will only have a weak influence.

Once all the spots have been profile-fitted, i.e. their individual shapes determined, each spot is integrated, which is a process whereby the overall intensity is effectively broken down into very fine units which are collected together to give the overall spot intensity. MOSFLM applies three corrections to the spot intensities, the Lorentz correction, the Polarisation correction and the Absorption correction.

The Lorentz correction corrects for the time each individual set of planes within the crystal are found in the beam in a diffracting position, and therefore giving rise to spots.

The Polarisation correction adjusts the intensities to correct for changes brought about by strongly polarised synchrotron X-ray beam.

The Absorption correction takes into account X-ray absorption effects on the intensities, which are related to the variation in thickness of the crystal as it is rotated.

The reflection (MTZ) file output from MOSFLM was then input into SORTMTZ. This sorts the reflections in the MTZ file into numerical order using the Miller indices (h then k then l), making it easier for the subsequent programs to process the data.

SCALA

The output from SORTMTZ is then input into SCALA (*8*), a scaling program. Scaling is required to obtain consistency between collected intensities for symmetry related spots. The program aims to get the differences from the mean intensity for each set of Miller indices, as close as possible to one another. Variations between the intensities recorded for symmetry related spots can arise for a number of different reasons. These include variations in the intensity of the X-ray beam between one block of recorded data and the next, damage to the crystal lattice from free radicals formed when the X-ray beam hits the crystal (reduced when the crystal is frozen) and errors in measurement and changes to the thickness of the crystal as it is rotated within the X-ray beam. In perfect conditions the intensities for each set of symmetry related spots (Freidel pairs and Bijvoet pairs) should be identical. Once the data is scaled, each set of symmetry-related spots gives rise to an associated single spot. This has an intensity derived from the average value of its corresponding set of symmetry-related spots. As such, a unique list of reflections is generated. The $R_{sym}(I)$ which is calculated by SCALA is used to compare the intensity of symmetry related reflections and is found using equation 2-1:

$$R_{sym}(I) = \frac{\Sigma_{hkl}\Sigma_i|I_i(h\ k\ l)-I(h\ k\ l)|}{\Sigma_{hkl}\Sigma_i I_i(h\ k\ l)} \qquad \textbf{Equation 2-1}$$

i-observations of a given reflection

I(h k l)- average intensity of the i observations

HKLVIEW

HKLVIEW is a display program which allows the intensities of Miller indices to be examined and compared and systematic absences observed. It is used to identify space groups for ambiguous data sets, and allows identifications of screw axis.

TRUNCATE

The program TRUNCATE is used to convert collected spot intensities (I) into structure factor amplitudes ($|F_{hkl}|$). These are important crystallographic values which are combined with phases (once known) to produce structure factors (F_{hkl}), one for each diffraction spot. The $|F_{hkl}|$ is calculated by TRUNCATE using Equation 2-2:

$$I = |F_{hkl}|^2 \times K \qquad \textbf{Equation 2-2}$$

K = scale factor derived from the Wilson plot.

TRUNCATE will also give an estimate of the number of objects to be found within an asymmetric unit and the solvent content of the unit cell using Equation 2-3.

$$S = 1 - \frac{N \times Vp \times Mr}{N_A \times V} \qquad \textbf{Equation 2-3}$$

Where:

S = Solvent content of unit cell (%)

N = Number of protein molecules in unit cell

V_p = Partial specific volume of protein molecule (ml/g)

M_r = Molecular weight of protein molecule (g/mole)

N_A = Avogadro's Number (number of protein molecules in one mole)

V = Total volume of unit cell (ml) (from X-ray data)

After processing in TRUNCATE, the log file contains a Wilson plot, a plot of average intensity vs. resolution, which, ideally should be linear, at high resolution, beyond 3 Å.

2.5.7 Molecular Replacement

AmoRe

Using the AmoRe package a large number of rotation and translation function solutions may be calculated in a fast, automatic way. Rotation functions are performed in reciprocal space. Two preliminary programs (SORTING and TABLING) are run prior to the rotation function, with SORTING packing and sorting the data into AmoRe format, whilst TABLING adjusts the model co-ordinates to the optimal position and prepares a table of continuous Fourier coefficients then being used in the structure factor calculations by interpolation (*20*). The program ROTING calculates the structure factor amplitudes for the search model in a P1 box (the size depending on the radius of integration, resolution of the data and dimensions of the search model). It then computes spherical harmonic coefficients for the search model and target Patterson distributions. Finally it calculates the rotation function. The translation search is carried out using the program TRAING, using four different fast Fourier Functions. The program allows automated examination of multiple rotation function peaks in the translation search, with both the R-factor and correlation coefficient being used to judge the correctness of each translation function.

MOLREP

MOLREP can perform all stages of molecular replacement automatically using X-ray data from the target crystal and search model. For each trial position/orientation structure factors are calculated by FFT of the electron density rather than by interpolation from a table of structure factors calculated on a fine grid. The rotation and translation functions performed by MOLREP are overlapping functions between the target and calculated model Pattersons. The Crowther fast rotation function (*21*) is used and is evaluated using FFT. The noise level is reduced due to the fact MOLREP computes the rotation function for three different orientations prior to averaging them out. The translation search algorithm (*22*) is run using FFT in reciprocal space. The R-factor and correlation coefficient between model and target structure factor amplitudes are used to determine the correctness of each solution.

BEAST

The program BEAST (*12*) carries out maximum likelihood molecular replacement for both the rotation and translation functions. The program may use a single or a variety of model structures, for which it calculates statistically weighted averaged structure factors. Beast is able to carry out both the rotation and translation functions in the same processing. An RMS value is applied to each model by the user, depending on the sequence identity between the model and the target. NMR structures may be used as models but generally require higher RMS values due to the way the NMR structure is derived.

EPMR

The molecular replacement program EPMR (*5*) (evolutionary programming for molecular replacement) was also utilised as a means to a possible solution for the phasing problem. EPMR utilises a six dimensional search using an evolutionary optimisation algorithm. The program takes the model data and performs FFT just once to calculate structure factors with the model at the origin of a P1 cell. The indices for each observed reflection are transferred into the molecular lattice and rotated according to the specified current rotation for the search model, the structure factors at the non-integral indices is calculated using linear interpolation into the table of P1 structure factors. Phase shifts corresponding to the current translation of the search model are applied and the contribution of all symmetry-related molecules are summed. A number of different runs were carried out using a population size of 300 over 50 generations (*5*). Initially the model is placed at 300 random positions within the unit cell and the fit of the model structure factors to the observed structure factors is calculated for each member of this population. Randomly selected sub-sets of this population are then compared in a stochastic tournament to find the best solution within that set. Each solution is then taken to the next generation where, in order to restore the population size, offspring are produced by introducing small random variables into the parameters comprising a parent solution. This process of evolutionary searching gradually focuses the search into the most promising regions of the search space (*5*). Figure 2-3 shows a flowchart showing the evolutionary search process.

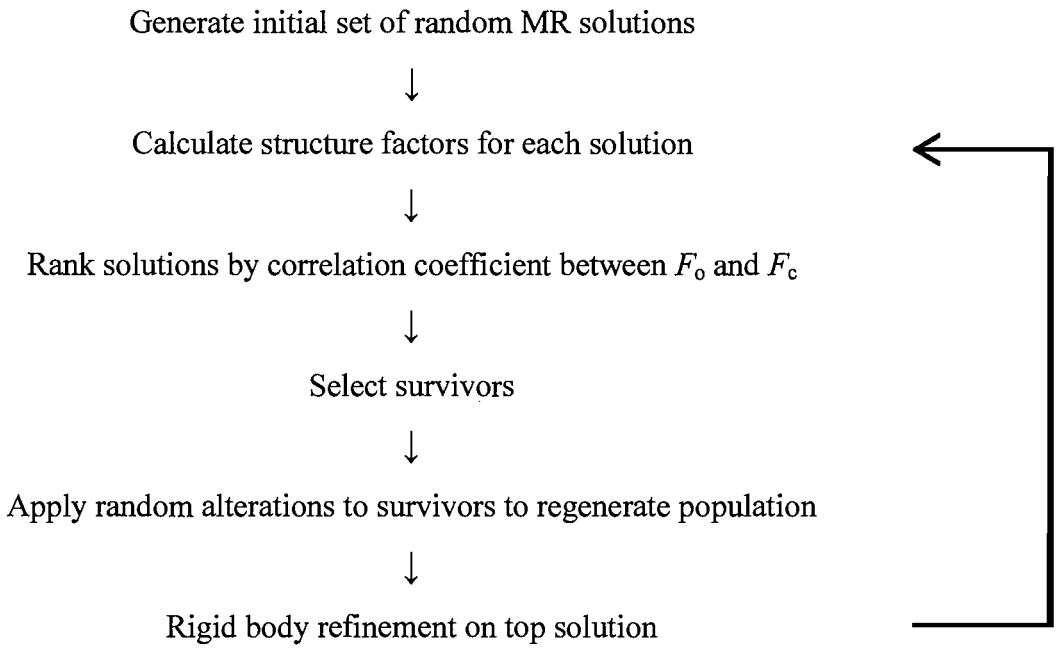


Figure 2-3 Flowchart showing the processing cycles undertaken during molecular replacement by evolutionary search. Taken from Kissinger *et al.* (*5*).

2.6 NMR Techniques

All experiments were carried out on a Varian *INOVA* 600 MHz spectrometer at 25 °C (unless otherwise stated). ^{15}N labelled protein was used to record ^{15}N HSQC-NOESY, ^{15}N HSQC-TOCSY, HSQC and HNHA and HNHB experiments. CBCACONNH, HNCA, HNCO, HNCOHB, HNCOCA, HNCACB, HCCHTOCSY and CN NOESY data were collected on a sample of ^{13}C,^{15}N-protein. Unlabelled protein samples, in 100 % $^{2}H_2O$, were used to determine aromatic chemical shifts and NOE data for aromatic side chains. ACP samples were prepared at 1 mM protein, 20 mM phosphate buffer pH 6.3 to a final volume of 600 µL with 5% D2O present.

Spectra were processed using nmrPipe (*23*) and phased in nmrDraw. Peak picking was carried out using PIPP (*24*). Unassigned HN, N, CO, CA and CB chemical shifts, selected from ^{1}H-^{15}N HSQC, HNCA, HNCACB, HNCO, HNCOCA and CBCACONNH spectra, for *apo* FAS ACP were fed into AUTOASSIGN (*25*), which sequentially assigned the back bone of the protein.

2.6.1 Automated assignment of backbone atom (HN, N, CO, CA and CB) chemical shifts using AUTOASSIGN

Traditionally a combination of NOESY and TOCSY spectra would be used to sequentially assign the protein backbone. Each amino acid type has a characteristic 'spin system' within a specific spectrum. These may be used to assign the residue type and subsequently locate them in sequence. The TOCSY would be used to determine the residue type whilst the NOESY would be used to identify connectivities between two sequential residues and hence allows the spectroscopist to 'walk' through the primary amino acid sequence assigning chemical shifts to the backbone and side chain atoms in the protein.

This sequential assignment technique is, however, a major bottleneck when attempting to solve a protein structure by NMR. Many spin systems are not unique to one amino acid type, which can introduce ambiguities to the assignment process. Problems resulting from weak signals for some resonances and/or lack of NOE data further complicate assignment. Recently, programs have been developed which are able to utilise three dimensional data sets recorded on $^{15}N/^{13}C$ labelled proteins and assign the backbone resonances of a protein in a matter of seconds. During the course of this work the program AUTOASSIGN (*26*) was utilised to assign the HN, N, CO, CA and CB chemical shifts for *apo* FAS ACP. AUTOASSIGN reads in the chemical shifts of unassigned peaks that have been selected from $^{1}H-^{15}N$ HSQC, HNCA, HNCACB, HNCO, HNCOCA and CBCACONNH spectra. The programme then looks for connectivities between the data sets and assigns the resonances that it identifies.

^{15}N-HN resonance frequencies of cross peaks from the above experiments are used to cross reference the spectra to ^{15}N-HN resonances in the reference HSQC. Generic spin system resonance frequencies for ^{13}C and ^{1}H, derived from the triple resonance experiments, are recorded as a ladder of individual shifts, e.g. CA ladder, CO ladder. AUTOASSIGN then carries out rounds of constraint based matching, where the strongest matches are initially recorded, with the constraints being relaxed progressively during subsequent analysis to allow for identification of weaker spin systems. Figure 2-4 shows a schematic overview of AUTOASSIGN's default execution system. In the first stage ('Make strongest matches') unambiguously inferred chemical shifts are designated on ladders, with only completely specified ladders participation in constraint

based matching. 'Allow designate shifts' includes refinement of incompletely specified ladders and allows the matching of remaining, unmatched ladders. 'Extend assigned segments' uses established assignments to give the specification of incompletely assigned ladders and extends currently assigned fragments of the sequence. 'Match weaker spin systems' sets aside the weakest generic spin systems for data sets containing extraneous spin systems. 'Finish assignments' is then used to evaluate the designated chemical shifts and sequential assignments, makes refinements and corrections where necessary and concludes with a final round of constraint based matching (*26*).

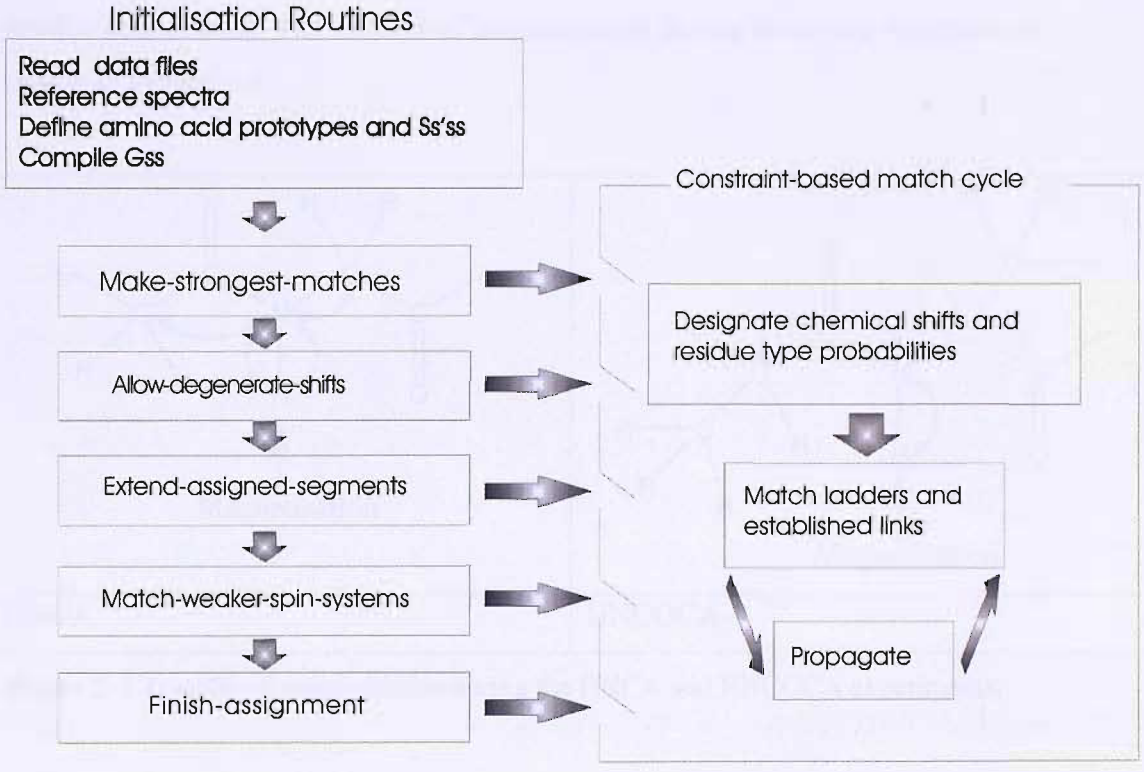


Figure 2-4 Schematic overview of AUTOASSIGN's default execution sequence. AUTOASSIGN reads in chemical shifts of unassigned peaks that have been selected from ^{1}H-^{15}N HSQC, HNCA, HNCACB, HNCO, HNCOCA and CBCACONNH spectra, looks for connectivities between the data sets and assigns the HN, NH, Co, CA and CB resonances that it identifies.

2.6.2 Spectra utilised by AUTOASSIGN

HNCA/HNCOCA

In an HNCA experiment magnetisation is created at the amide proton (H), and is transferred to the attached nitrogen atom (N) and allowed to evolve. This constitutes the

first spectral dimension. The magnetisation is then transferred to the CA nucleus to evolve, which is measured as the second dimension. Afterwards, the magnetisation is transferred back the same way to the amide proton and this is measured as the third (direct) dimension. In each step magnetisation is transferred via strong couplings between the nuclei. The coupling, between the nitrogen atom and the α carbon of the preceding amino acid (^2J= 7 Hz), is only marginally smaller than the coupling to the directly attached CA atom (^1J = 11 Hz). Thus, the nitrogen atom of a given amino acid is correlated with both CA i, and i-1. It is therefore theoretically possible to sequentially assign the protein backbone exclusively using an ideal HNCA spectrum. The HNCOCA correlates the backbone amide resonance for residue i with the CA of the preceding residue. It can be used in conjunction with the HNCA to clarify intra-residue and inter-residue connections. The transfer of magnetisation during these two experiments is shown in Figure 2-5

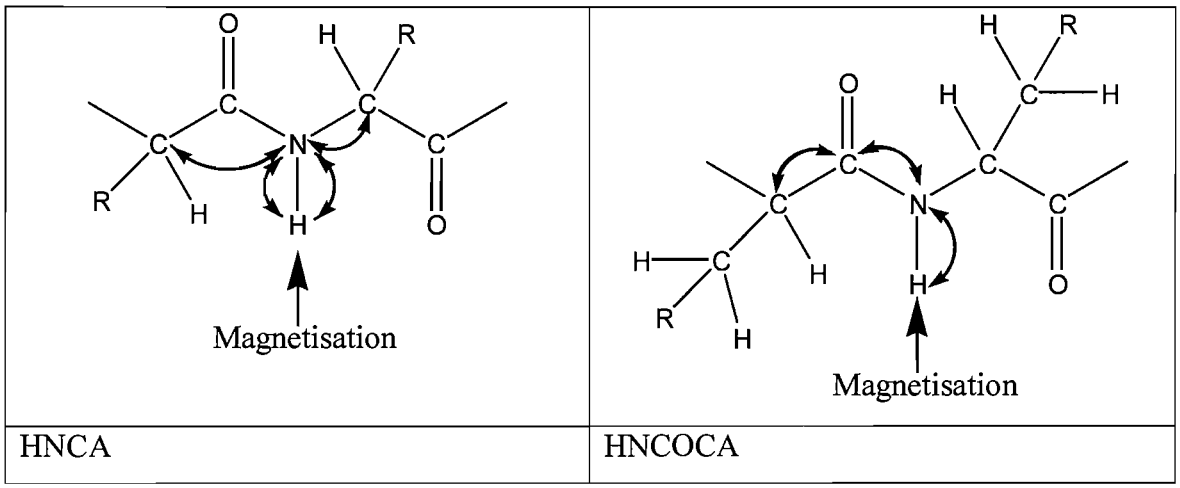
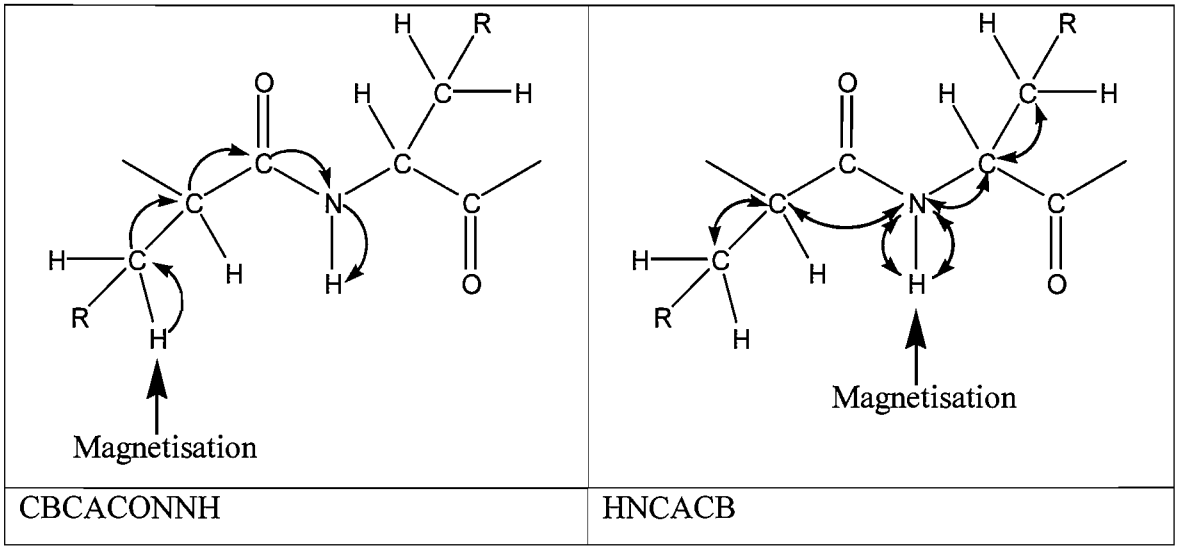


Figure 2-5 Transfer of magnetisation during the HNCA and HNCOCA experiments.

CBCACONNH/HNCACB

CBCACONNH experiments correlate α and β ^{13}C chemical shifts with the amide ^{15}N and ^1H resonances of the succeeding residue i, whilst the HNCACB experiment correlates α and β ^{13}C chemical shifts for residue i and i-1 with the ^{15}N and ^1H resonances from residue i (Figure 2-6). In the ^{13}C dimension the α and β resonances are anti-phase making the assignment process easier.



CBCACONNH	HNCACB

Figure 2-6 Transfer of magnetisation during the CBCACONNH and HNCACB experiments.

2.6.3 HA chemical shift determination using HNHA experiments

To complete the assignment of the protein backbone the Hα proton chemical shift must be identified for each amino acid. Once assigned this nucleus provides vital links to the side-chain resonances in TOCSY type spectra. The HNHA is a three dimensional, heteronuclear experiment which requires the protein sample to be ^{15}N labelled. It is used to measure $^{3}J_{HN\text{-}H\alpha}$ coupling constant from the ratio of HN-HN cross-peak to HN-HA cross-peak intensities. The intensity of the cross-peak is related to the strength of the J coupling and this coupling is dependent on the main chain torsion angle Φ. The HNHA experiment is usually more sensitive than a TOCSY spectrum because the magnetisation from the amide proton is only passed as far as the Hα proton. In contrast, in the TOCSY experiment, once the magnetisation is transferred from the HN to the Hα proton it is rapidly transferred on to the remaining protons in the side-chain causing a reduction in signal intensity.

2.6.4 Side-chain chemical shift assignment

^{15}N TOCSY-HSQC and HCCH TOCSY spectra of *apo* FAS ACP were recorded to assign all side-chains of *apo* FAS ACP with the exception of aromatic side-chains. Total Correlation SpectroscopY (TOCSY) shows correlations between protons within the same spin system (within the same residue). In TOCSY the mixing between spins does not occur instantaneously, and the extent of mixing can be adjusted by modifying the isotropic mixing period of the experiment. It is therefore possible to control the

number of correlated spins within a spin system. A short mixing time TOCSY may only record cross peaks arising from adjacent protons and not from more distant protons. The incorporation of magnetic nuclei (^{15}N and ^{13}C) allow for the TOCSY to be recorded in three dimensions, with the ^{15}N HSQC TOCSY being used in conjunction with the HCCH TOCSY to identify the nitrogen, carbon and proton chemical shifts for each residue. The ^{15}N TOCSY allows for the assignment of the side chain proton chemical shifts, using the back-bone amide chemical shift from the ^{15}N HSQC as a reference. The HCCH TOCSY allows for both side-chain proton and carbon chemical shift assignment. In the ^{15}N TOCSY the whole spin system can be seen on one 'slice' of the spectrum corresponding to the amide nitrogen chemical shift whereas in the HCCH TOCSY the protons resonances for each spin system will appear upon slices corresponding to each carbon resonance found within the side-chain, and as such will appear more than once (except for glycines).

2.6.5 Phi and Psi angle determination

Another key parameter used for structure generation is dihedral angles. A representation of a dipeptide, showing the locations of the phi (φ) and psi (ψ) angles can be seen in Figure 2-7. The peptide bond is planar and as such has standard values independent of the secondary structural element in which it is found. α helical regions of a protein have phi angles in the region of -58° and psi angles of approximately -47°. By calculating the phi and psi angles for a protein the secondary structure may be predicted.

The software program TALOS (*27*) was used to predict back-bone phi and psi angles for *apo* FAS ACP using the chemical shift assignments for Hα, Cα, Cβ, CO and N and the FAS ACP protein sequence. It uses the principle that chemical shifts are highly correlated to their secondary structure. TALOS uses the chemical shift information for triplets of residues and compares all 15 chemical shifts found in this triplet with those found within a database, in order to make predictions for the phi and psi angles for the central residue of the triplet. The similarity is measured using a score based on the weighted sum of the square differences between shifts in the target protein and those in the data. The score also includes a small qualitative term, which is biased towards matching similar sequences.

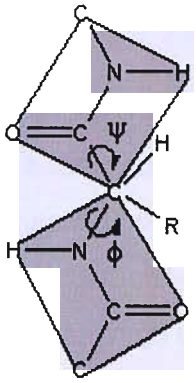


Figure 2-7 Phi (ϕ) and Psi (ψ) angles found within proteins. The peptide bond, omega Ω, is planar (represented by the grey boxes) and as such has standard values independent of the secondary structural element in which it is found. α helical regions of a protein have phi angles in the region of –58° and psi angles of approximately –47°. By calculating the phi and psi angles for a protein the secondary structure may be predicted.

2.6.6 NOE data

Unlike the other NMR experiments used during this work, Nuclear Overhauser Enhancement SpectroscopY (NOESY) takes advantage of through space dipolar couplings. These experiments measure the distance-dependent nuclear Overhauser effect (NOE) which exists between protons separated by less than 5 Å, and is vital for protein structure determination. NOEs are typically seen from any protons within ~5 Å of each other. As such, in ^{15}N HSQC NOESY, NOEs are most commonly observed between the backbone amide proton of residue i and protons (a) within its own spin system and (b) within the preceding residue, i-1. In α helices NOEs are usually observed between residue i, and the third residue away, i+3, (Hα-NH i,i+3)and (Hα-βH i,i+3) due to the hydrogen bonding seen in α helices. NOEs are often seen between Hα-NH i,i+4 although these tend to weaker than the i,i+3 NOEs and may be used to distinguish between α helices and 3_{10} helices which tend to be richer in i,i+2 contacts. Long range NOE's are seen between residues which, although far apart in the amino acid sequence, are located close in the 3D structure of the protein. The intensity of the NOE decreases as a function of $1/r^6$ where r is the interatomic distance. NOEs therefore provide a rough indication of how close in 3D space are the two protons giving rise to that NOE are, with more intense peaks relating to closer proximities.

2.6.7 Structure Calculations

The compiled NOE table was fed into the program XPLOR (*28*), which attempts to fold an extended strand *apo* FAS ACP into a three dimensional form which satisfies the NOE data. XPLOR uses a simulated annealing protocol in an attempt to form the lowest energy structure possible. Subsequent calculations utilised the lowest energy model from previous rounds of calculations. 20,000 high and low temperature iterations were applied in calculations using the extended strand protein conformation. 8000 iterations were used when the starting model was a folded structure produced from a previous run. Each class of NOE had centred averaging applied to it during structure calculations. Multiple structures were generated for each run, allowing for subsequent analysis of NOE violations. NOEs with violations over 1.0 Å were removed from subsequent calculations. As the model improved this NOE violation threshold was reduced to 0.1 Å. The excluded NOEs were rechecked in the relevant spectrum to ensure the correct assignment had been made and the correct errors applied. Families of structures were analysed using the program MOLMOL (*29*), where they were either overlaid throughout the whole molecule or over their secondary structure elements only and the RMSD values examined.

^{15}N TOCSY-HSQC and HCCH-TOCSY spectra were used for side chain resonance assignment and H-H NOESY, ^{15}N-HSQC-NOESY and CN-NOESY-HSQC spectra were used to determine NOE restraints. Phi and Psi restraints were calculated from Hα, Cα, Cβ, CO and N chemical shifts using TALOS (*30*).

2.7 References

1. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A laboratory Manual.*, Cold Spring Harbour Laboratory Press., New York.
2. Hopwood, D. A., Bibb, M.J., Chater, K.F., Kieser, T., Bruton, C.J., Kieser, H.M., Lydiate, D.J., Smith, C.P., Ward, J.M. and Schrempf, H. (1985) *The Genetic Manipulation of Streptomyces: A Laboratory Manual.*, The John Innes Foundation, Norwich.
3. Studier, F. W., Roseenberg, A.H., Dunn, J.J. and Dubendorff, J.W. (1990) *Methods Enzymol. 185*, 60-89.
4. Laemelli, U. K. (1970) *Nature 227*, 248-254.
5. Kissinger, C. R., Gehlhaar, D.K. and Fogel, D.B. (1999) *Acta. Cryst. D55*, 484-491.
6. Steller, I. (1997) *Journal of Applied Crystallography 30*, 1036-1040.
7. Leslie, A. G. W., MRC laboratory of molecular biology, Cambridge.
8. Fox, G. C. (1966) *Acta. Cryst. 20*, 886-891.
9. French, S., and Wilson, K. (1978) *Acta. Cryst. A34*, 517-525.
10. Vagin, A. (1997) *J. Appl. Cryst. 30*, 1022-1025.
11. Read, R. J. (1999) *Acta Cryst. D55*, 1759-1764.
12. Read, R. J. (2001) *Acta. Cryst. D57*, 1373-82.
13. Chothia, C., and Lesk, A. M. (1986) *EMBO J. 5*, 823-826.
14. Navaza, J. (1987) *Acta. Cryst. A43*, 645-653.
15. Navaza, J. (1990) *Acta. Cryst. A46*, 619-620.
16. Navaza, J. (1994) *Acta. Cryst. A50*, 157-163.
17. Navaza, J. (1993) *Acta. Cryst. D49*, 588-591.
18. Hirshfeld, F. L. (1968) *Acta. Cryst. A24*, 301-311.
19. Collaborative Computational Project, N. (1994) *Acta. Cryst. D50*, 760-763.
20. Navaza, J. (1994) *Acta. Cryst. A50*, 157-163.
21. Crowther, R. A. (1972) *The Molecular Replacement Method*, Gordon and Breach, New York.
22. Vagin, A. (1989) *Protein Crystallography 29*, 117-121.
23. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) *J. Biomol. NMR 6*, 277-293.
24. Garrett, D. S., Powers, R., Gronenborn, A. M., and Clore, G. M. (1991) *J. Magn. Reson. 95*, 214-220.
25. Zimmerman, D., Kulikowski, C., Wang, L., and Lyons, B. (1994) *J. Biomol. NMR 4*, 241.
26. Zimmerman, D. E., Kulikowski, C.A., Huang, Y., Feng, W., Tashiro, M., Shimotakahara, S., Chien, C., Powers, R., Montelione, G.T.,. (1997) *J. Mol. Biol. 269*, 592-610.
27. Cordier, F., Dingley, A. J., and Grzesiek, S. (1999) *J. Biomol. NMR 13*, 175-80.
28. Brunger, A. T. (1995), Newhaven, CT.
29. Koradi, R., Billeter, M., and Wüthrich, K. (1996) *J. Mol. Graphics. 14*, 51-55.
30. Cornilescu, G., Delaglio, F., and Box. A. (1999) *J. Biomol. NMR 13*, 289-302.

Chapter 3

Solution Structure of Fatty Acid Synthase *apo* ACP from *Streptomyces coelicolor*

3. Solution structure of Fatty Acid Synthase *apo* ACP from *Streptomyces coelicolor.*

3.1 Introduction

The solution structure of the polyketide synthase ACP from the actinorhodin system (*act*) in *Streptomyces coelicolor* was solved by Crump *et al.* (*1*) in 1996. As described in Chapter 1, there are many similarities between the PKS and FAS systems. However there are also differences which prevent 'cross-talk' between elements of the PKS and FAS systems. A structural comparison of two components, one from a FAS and one from a PKS within the same organism has, as yet, not been published. During the course of this chapter, the solution structure of *apo* FAS ACP from *Streptomyces coelicolor* is presented and a comparison between this and the *act* PKS ACP structure from the same species is discussed.

3.2 Practical Methodology and Results

3.2.1 *apo* FAS ACP expression and purification

apo FAS ACP was initially expressed using 10 % glycerol stock cells provided by Bristol University for both the production of unlabelled and ^{15}N labelled protein. *E. coli* cultures, with an A_{600} of 0.8, were induced for 3 hours in LB media for unlabelled samples and for 18 hours in M9 media for ^{15}N labelled samples. ACP samples were purified using Q sepharose high performance anionic exchange technology, with *apo* FAS ACP eluting in ~520 mM NaCl and *holo* FAS ACP eluting in 560 mM NaCl. Protein yields for the unlabelled ACP were ~20 mg/L and for the ^{15}N labelled protein ~10 mg/L.

However the original glycerol stock of *E.* coli BL21 DE3 expressing *apo* FAS ACP deteriorated and expression was lost. Molecular biology techniques were used to regain a functioning expression system in-house. Initially the glycerol stock was plated onto agar plates containing ampicillin and colonies selected for and cultured in LB media. After centrifugation, cell pellets underwent plasmid DNA extraction using Wizard®Plus SV Minipreps DNA purification system. 2 % agarose gel electrophoresis indicated the presence of the plasmid DNA as shown in Figure 3-1. Another sample of the same

extract was digested using BamHI and NedII (corresponding to the restriction sites surrounding the insert encoding the *apo* FAS ACP) before being run on an agarose gel to check for the presence of the DNA insert as seen in Figure 3-1. Insert DNA was excised from the agarose gel and cleaned, using Bio 101 Gene Clean Kits, before being ligated back into a fresh a pET11c vector. This new plasmid was then re-transformed into BL21 DE3 cells and induction checked in a test growth in LB media. The plasmids which yielded the highest induction were then carried through into larger scale growths. The new plasmids were shown to express in 400 mL LB, ^{15}N M9 and $^{15}N/^{13}C$ M9 cultures after a 3 hour induction, compared to an 18 hour induction in 10 mL test cultures. A time course expression experiment, using 400 mL unlabelled M9 and ^{15}N M9 cultures, was carried out, which revealed that maximal protein expression occurred much more quickly, possibly due to improved aeration from shaking in baffled flasks. It was observed that in 400 mL cultures induction took place between 1-2 hours with complete loss of protein after approximately 5 hours. Doubly labelled expression cultures were therefore induced for 2 hours prior to harvesting and purified as for standard ACP samples, with a yield of ~3 mg/L being obtained.

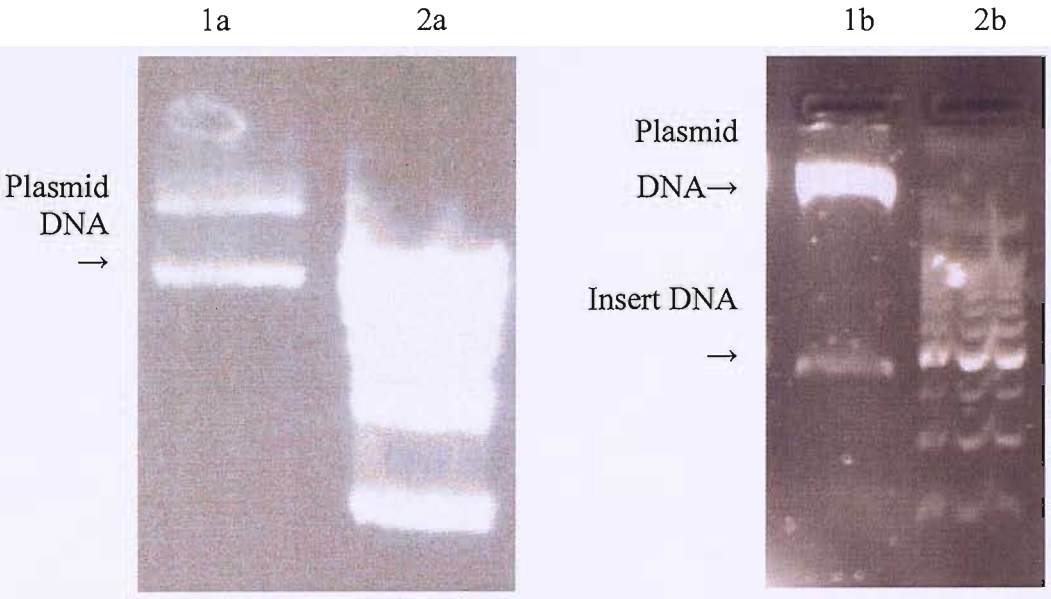
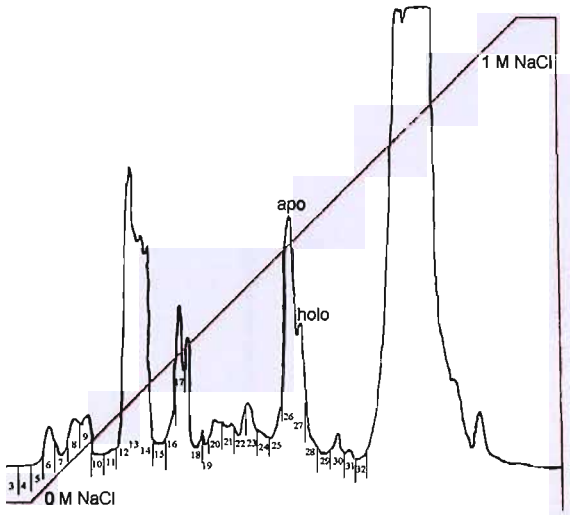


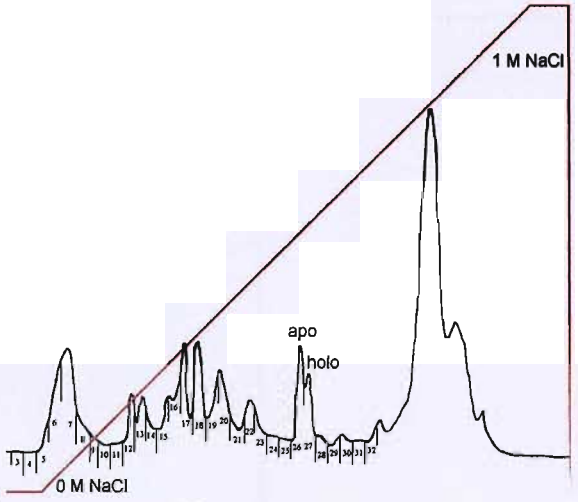
Figure 3-1 a) 2 % Agarose gel showing the presence of plasmid DNA (lane 1a) and 1 kb DNA markers (lane 2a), b) 2 % Agarose gel showing the presence of insert encoding for FAS ACP and plasmid DNA (lane 1b) and 100 bp DNA ladder (lane 2b).

Figure 3-2 shows the FPLC traces from a Q Sepharose high performance anionic exchange column produced from running unlabelled protein (60 % - 90 % ammonium sulphate cut), ^{15}N protein and $^{15}N/^{13}C$ protein respectively, eluted using a 0 – 1 M

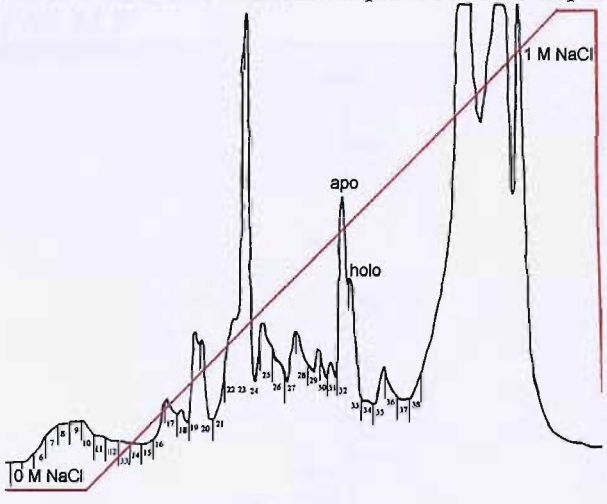
sodium chloride gradient, showing relative quantities of protein produced. Total mass of protein decreases with increased degree of labelling. The ratio of *apo* to *holo* ACP can be gauged from these traces since the *holo* ACP elutes immediately after the *apo*. The increase in the relative amount of *holo* ACP during the labelled expressions may be due to the lower overall amount of ACP available as substrate for the native acpS, which converts *apo* ACP to *holo*. Electrospray ionisation- mass-spectroscopy (ESI-MS) was used to confirm that the correct species had been isolated and that incorporation of the isotopes had been successful. The recorded mass-spectra for the three FAS ACP species are shown in Figure 3-3. The observed mass of 8787 Da for unlabelled FAS ACP agrees well with the expected mass of 8783 Da. ^{15}N labelled *apo* FAS ACP (expected mass 8865 Da) and $^{15}N/^{13}C$ labelled *apo* FAS ACP (expected mass 9250 Da) yielded observed masses of 8869 Da and 9249 Da respectively. Incorporation levels for the ^{15}N label appear to exceed 100% although the increase of 4 Da over the expected mass is consistent with that observed for the unlabelled protein and likely to be proton adducts since samples are treated with formic acid prior to MS analysis. For the $^{15}N/^{13}C$ labelled ACP mass spectra indicate an incorporation level almost 100 %.



a) Anion exchange FPLC trace of unlabelled *apo* FAS ACP. *apo:holo* ratio ~ 4:1

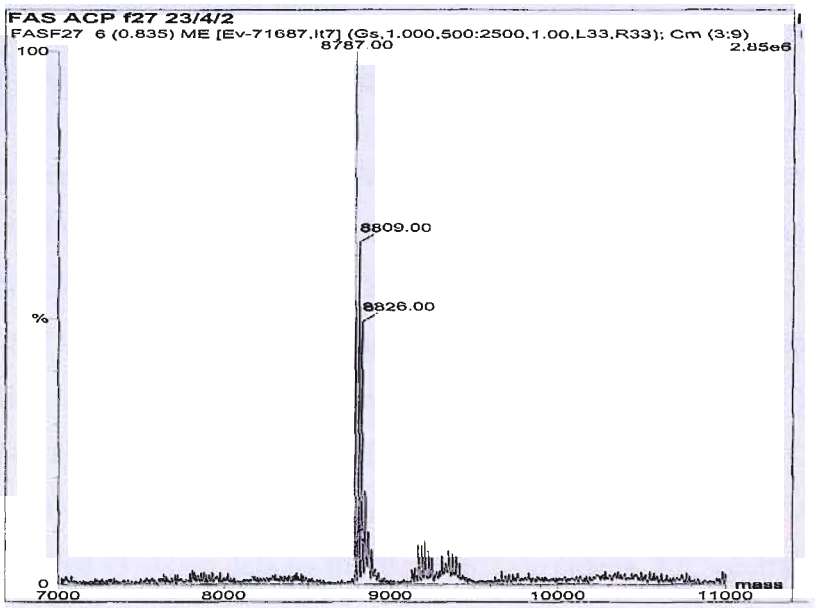


b) Anion exchange FPLC trace of ^{15}N labelled *apo* FAS ACP. *apo:holo* ratio ~ 3:1

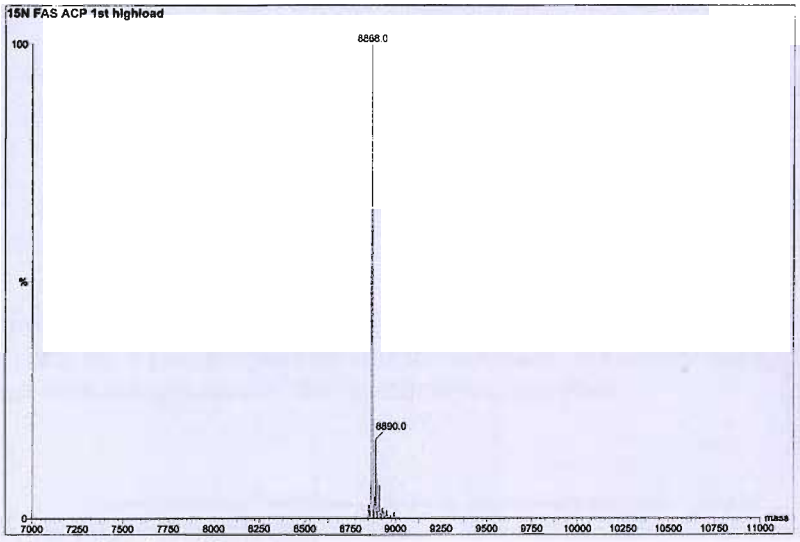


c) Anion exchange FPLC trace of $^{15}N/^{13}C$ labelled *apo* FAS ACP. *apo:holo* ratio ~2:1

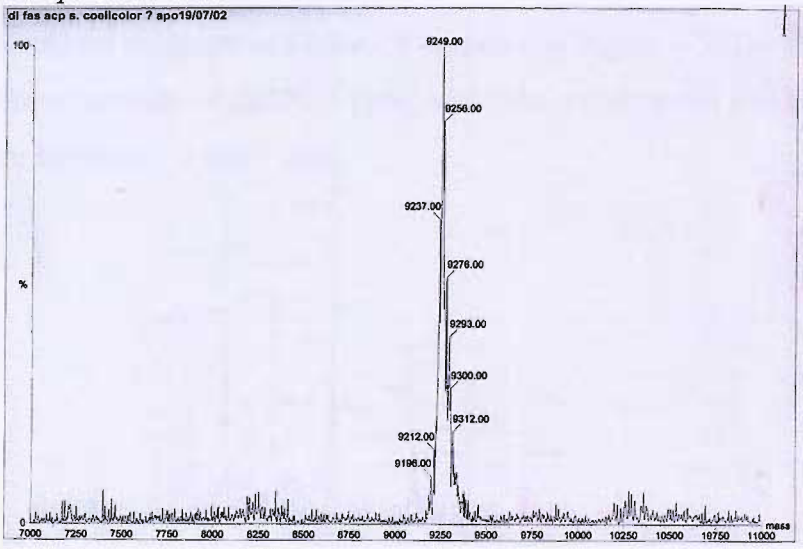
Figure 3-2 Q sepharose high performance anionic exchange column FPLC trace for a) unlabelled *apo* FAS ACP (fraction 26), b) ^{15}N labelled *apo* FAS ACP (fraction 32) and c) $^{15}N/^{13}C$ labelled *apo* FAS ACP (fraction 26), showing protein yield when eluted using a 0-1 M NaCl gradient (red line).



a) Unlabelled *apo* FAS ACP



b) ^{15}N labelled *apo* FAS ACP



c) $^{15}N/^{13}C$ labelled *apo* FAS ACP

Figure 3-3 Mass spectra of a) unlabelled *apo* FAS ACP [Expected mass: 8783 Da, observed mass: 8787 Da] b) ^{15}N labelled *apo* FAS ACP [Expected mass: 8865 Da, observed mass: 8869 Da] and c) $^{15}N/^{13}C$ labelled *apo* FAS ACP [Expected mass: 9250 Da, observed mass: 9249 Da].

3.3 NMR Studies on FAS ACP

ACP samples were prepared at a concentration of 6 mg/mL in 10 mM phosphate buffer, pH 8, with 5 % D_2O present in the 600 µL sample. ^{15}N labelled samples were used to record ^{1}H-^{15}N HSQC (*2*), ^{15}N TOCSY HSQC (*3-6*) (59 ms mixing time), ^{15}N NOESY HSQC (*5, 7, 8*) (150 ms and 40 ms mixing time), HNHA (*9*) and HNHB spectra. $^{15}N/^{13}C$ labelled samples were used to record HNCO (*10-12*), HNCA (*10, 11, 13*), HNCACB (*14*), HNCOCA (*11, 15*), HNCOCB, CBCACONNH (*16*), HCCH TOCSY (*17*) and CN NOESY HSQC (*18*) spectra. Similar samples of unlabelled ACP, in 100 % D_2O, were prepared to obtain data on the aromatic side chains. All spectra were processed using NMRpipe (*19*). The amino acid sequence for FAS ACP can be seen in Figure 3-4.

```
 AATQEEIVA GLAEIVNEIA GIPVEDVKLD KSFTDDLDVD SLSMVEVVVA AEERFDVKIP
         10          20          30          40          50          60

DDDVKNLKTV GDATKYILDH QA
          70         80
```

Figure 3-4 Primary amino acid sequence of FAS ACP from *Streptomyces coelicolor*. Serine 41 (highlighted in red) has a phosphopantetheine moiety added covalently attached by ACPS when converted from the *apo* form of the protein to the *holo* form.

3.3.1 1D FAS ACP Spectrum

The one dimensional spectrum of FAS ACP is shown in Figure 3-5. The amide proton resonances occur between ~6 and 10.5 ppm, whilst the α proton and side chain proton signals appear between ~ 0 and 5 ppm.

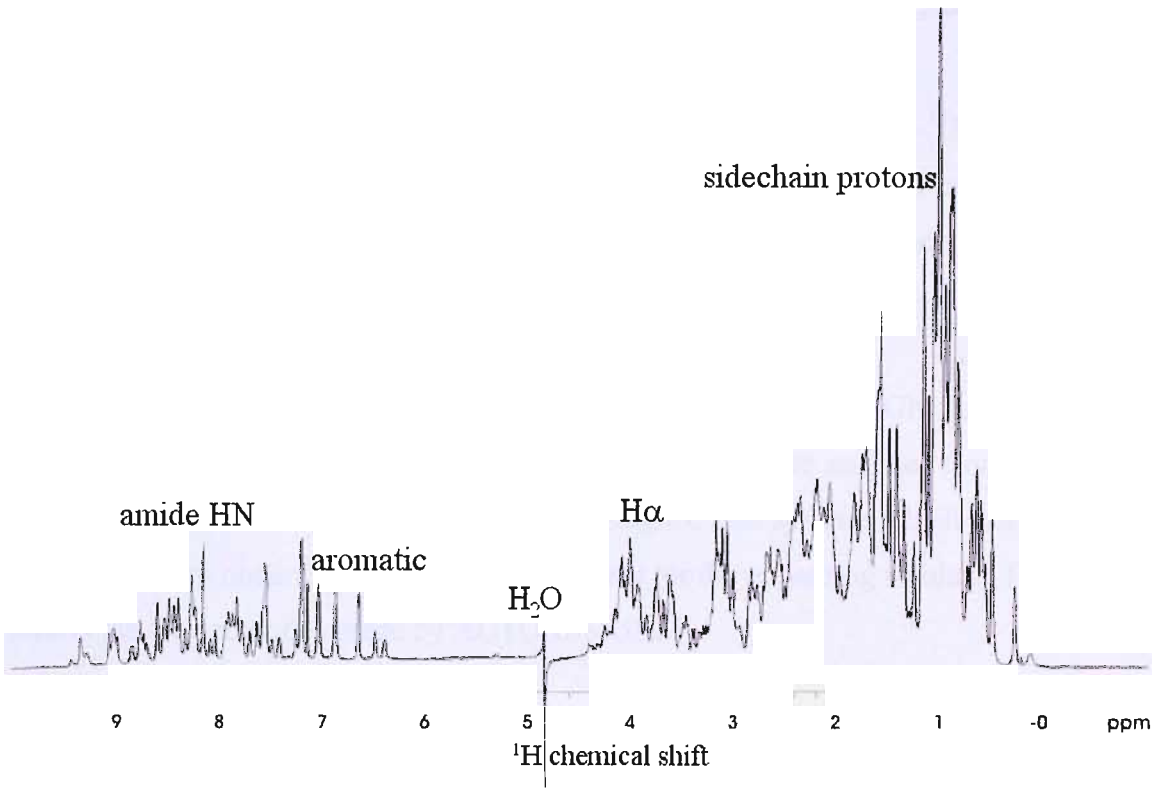


Figure 3-5 One dimensional proton spectrum of *apo* FAS ACP. The amide protons appear between ~6 and 9.5 ppm, whilst the α proton and side chain proton chemical shifts occur between ~ 0 and 5 ppm.

3.3.2 Chemical Shift Assignment

Critical to any analysis by NMR is the assignment of resonances. This is especially true in protein structure. Initially a ^{1}H-^{15}N HSQC spectrum was recorded on *apo* FAS ACP and revealed the protein to be folded, with a broad distribution of signals in both dimensions of the ^{1}H-^{15}N HSQC spectrum. The majority of resonances lie between 6 and 9.3 ppm in the ^{1}H dimension suggesting that the protein is likely to be predominantly α helical in structure, which was expected, since previously solved ACP structures are α helical bundles. Structures with a β sheet component display characteristic proton chemical shifts up to ~11 ppm.

3.3.3 Automated assignment of backbone atom (HN, N, CO, CA and CB) chemical shifts using AUTOASSIGN

Unassigned HN, N, CO, CA and CB chemical shifts, selected from ^{1}H-^{15}N HSQC, HNCA, HNCACB, HNCO, HNCOCA and CBCACONNH spectra, for *apo* FAS ACP were fed into AUTOASSIGN (*20*), which sequentially assigned the back bone of the protein.

The data in Table 3-1 shows the assignment success of AUTOASSIGN for *apo* FAS ACP. As can be seen, ~95 % of the backbone resonances were assigned from the data set. AUTOASSIGN also provides a list of unassigned spin systems which can be used subsequently to obtain the backbone assignments for the remaining residues. Figure 3-6 shows the connectivities made by AUTOASSIGN.

Resonance	Number Assigned	Total Assignments Possible	% Assigned
HN	74	78	95
N	74	78	95
CO (i-1)	73	78	94
CA	74	78	95
CB	69	75	92

Table 3-1 Assignment statistics for FAS ACP data analysed using AUTOASSIGN(*21*).

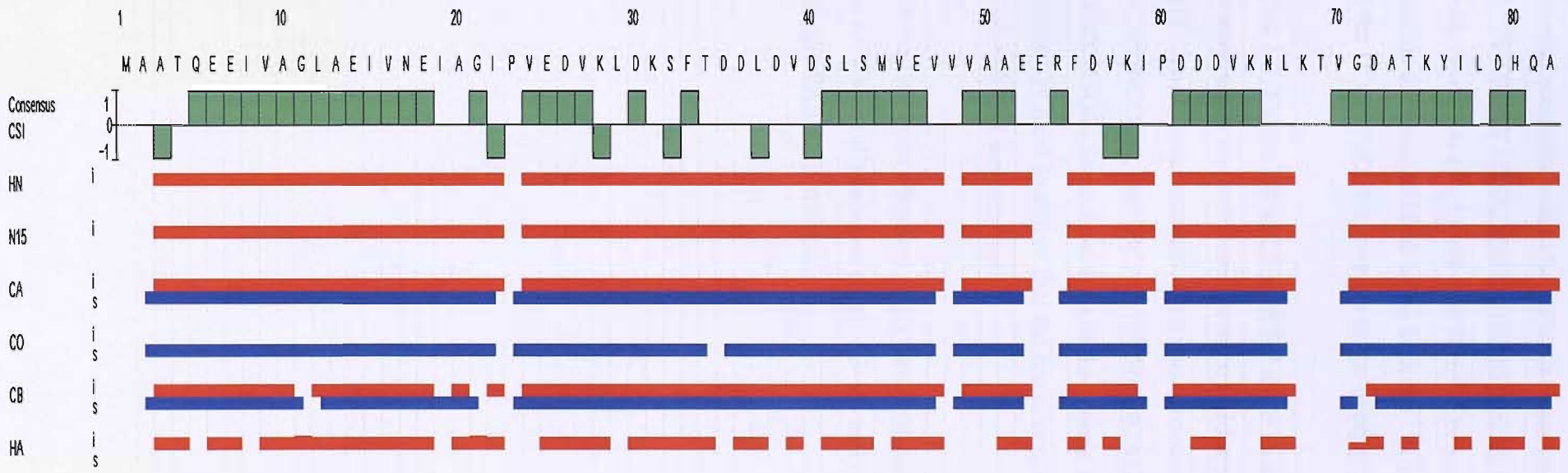


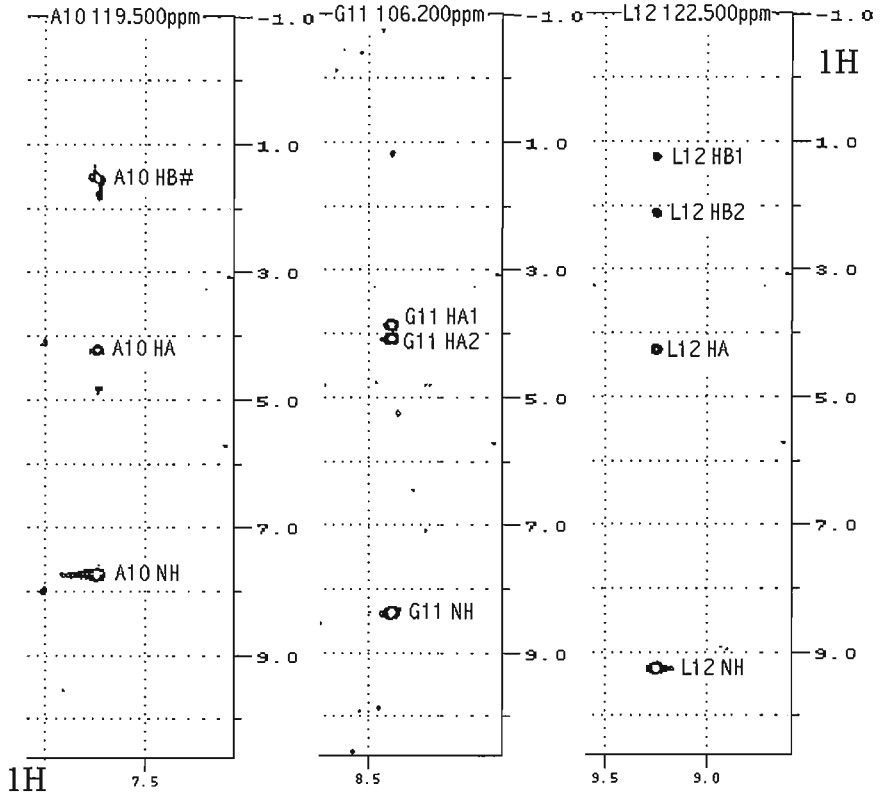
Figure 3-6 Graphical output of the sequential assignments and backbone assignment of FAS ACP output from AUTOASSIGN. Chemical shifts shown in red were assigned from intra-residue data; shifts shown in blue were sequentially assigned. The consensus chemical shift index shown in green predicts the secondary structure of the protein from the recorded chemical shifts, with 1 indicating α helical structures, 0 indicating no defined structure and –1 indicating a β sheet structure.

The areas within the amino acid sequence which caused problems for AUTOASSIGN were also seen observed to be difficult to assign manually using the ^{15}N NOESY HSQC and ^{15}N TOCSY HSQC spectra, due to overlap of similar spin systems. These included Val48, Glu53, Arg54, Lys68 and Thr69. Chemical shift assignments were achieved manually for these residues using both the ^{15}N and $^{15}N/^{13}C$ spectra previously described and the HCCH TOCSY to overcome the overlap problem.

Figure 3-7, Figure 3-8 and Figure 3-9 show an example of how sequential assignment was undertaken manually using a three residue sequence within *apo* FAS ACP, Ala10-Gly11-Leu12, as an example. This was carried out using pairs of spectra, the HNCA/HNCOCA, the CBCACONNH/HNCACB and the ^{15}N TOCSY HSQC and ^{15}N NOESY HSQC. Several sections of the protein sequence were selected and the spectra examined by hand to ensure that the resonances and connectivities assigned by AUTOASSIGN were correct.

Once the backbone chemical shifts were assigned the ^{1}H-^{15}N HSQC spectrum of *apo* FAS ACP was fully labelled (Figure 3-10) and used as a reference spectrum for the other *apo* FAS ACP spectra.

^{15}N TOCSY HSQC



^{15}N NOESY HSQC

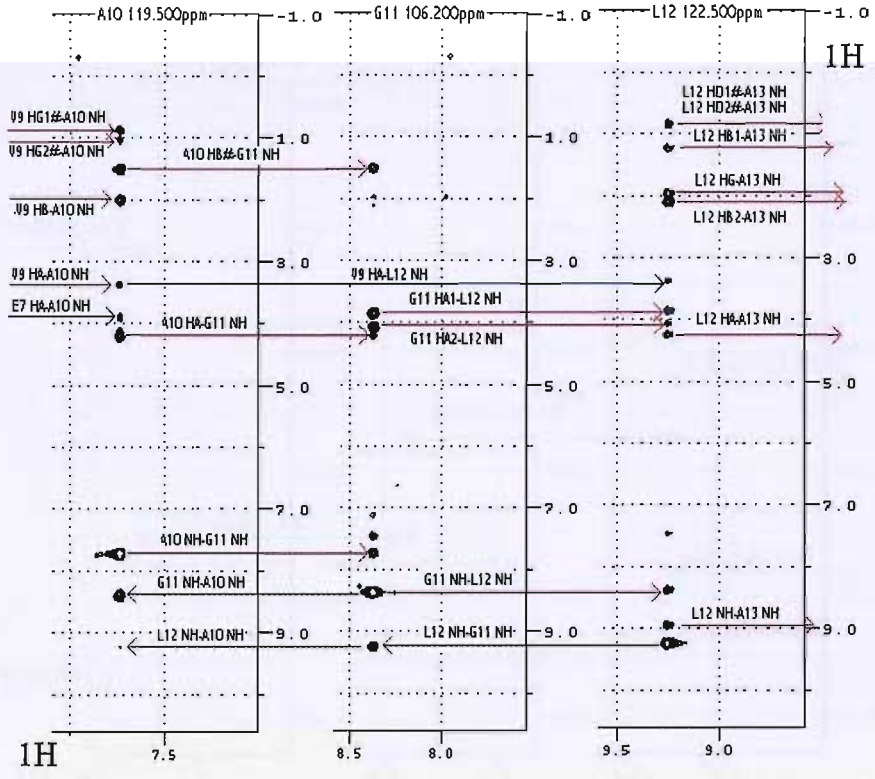
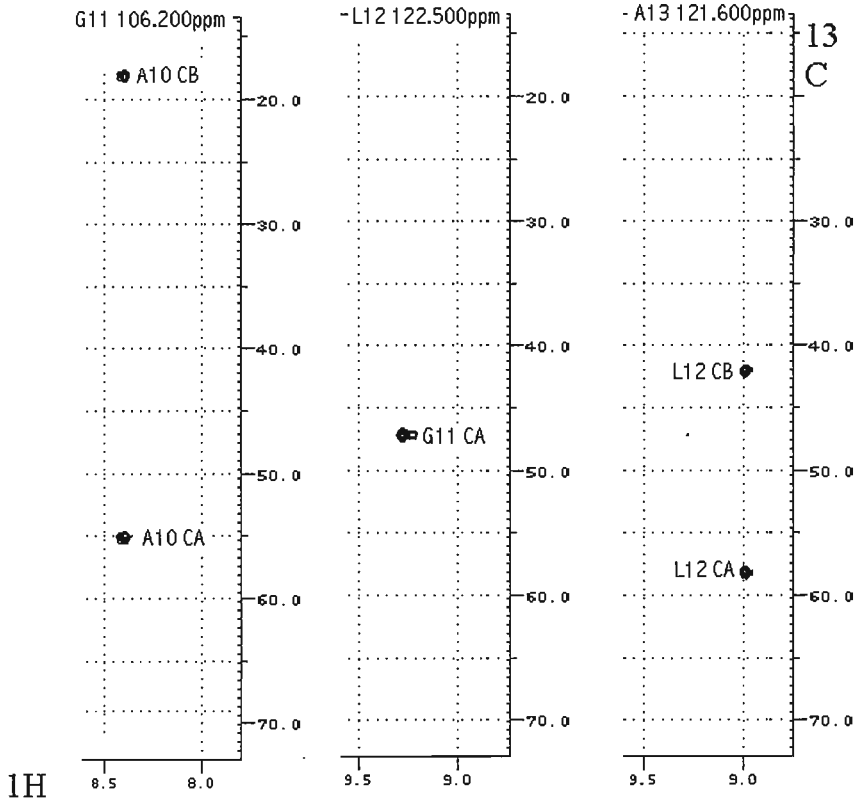


Figure 3-7 Sequential assignment of Ala10-Gly11-Leu12 in *apo* FAS ACP using ^{15}N NOESY HSQC and ^{15}N TOCSY HSQC spectra.

CBCACONNH



HNCACB

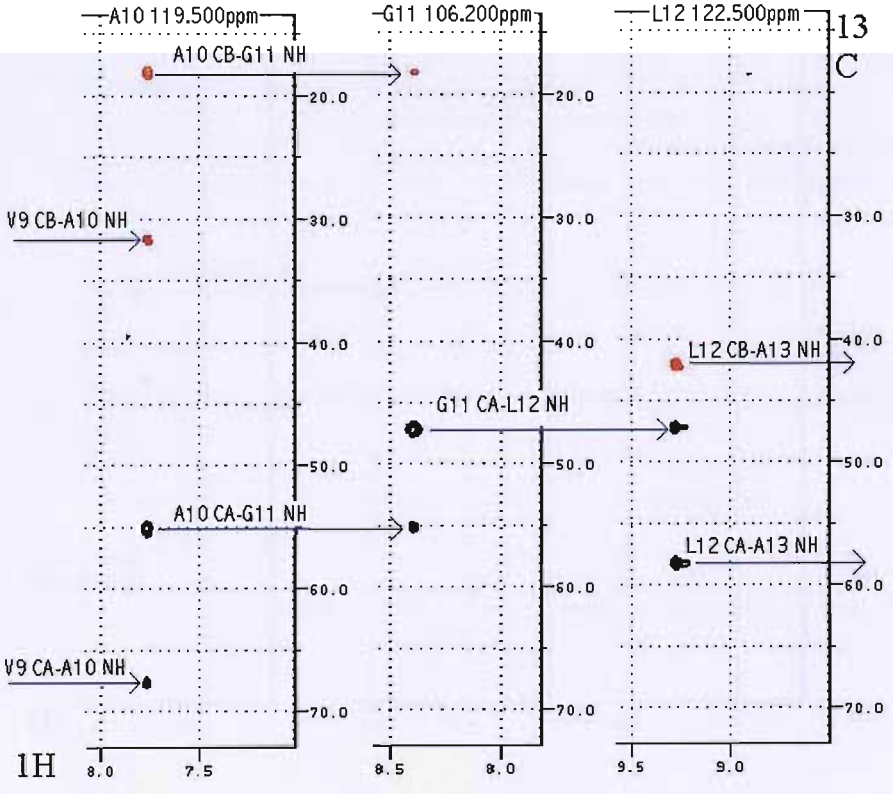
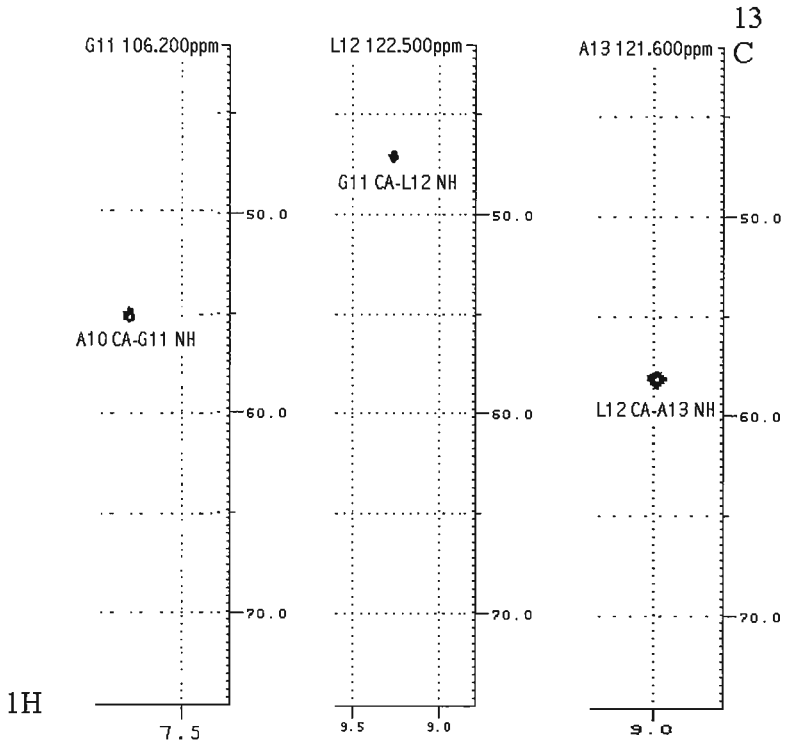


Figure 3-8 Sequential assignment of Ala10-Gly11-Leu12 in *apo* FAS ACP using CBCACONNH and HNCACB spectra. CB shifts are shown in red whilst CA shifts are shown in black.

HNCOCA



HNCA

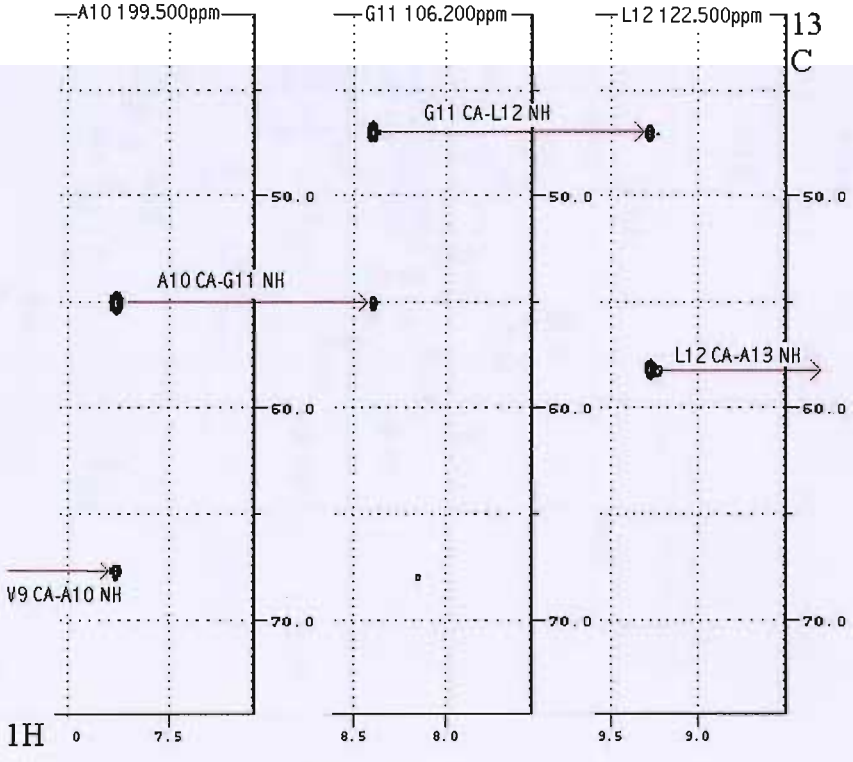


Figure 3-9 Sequential assignment of Ala10-Gly11-Leu12 in *apo* FAS ACP using HNCA and HNCOCA spectra.

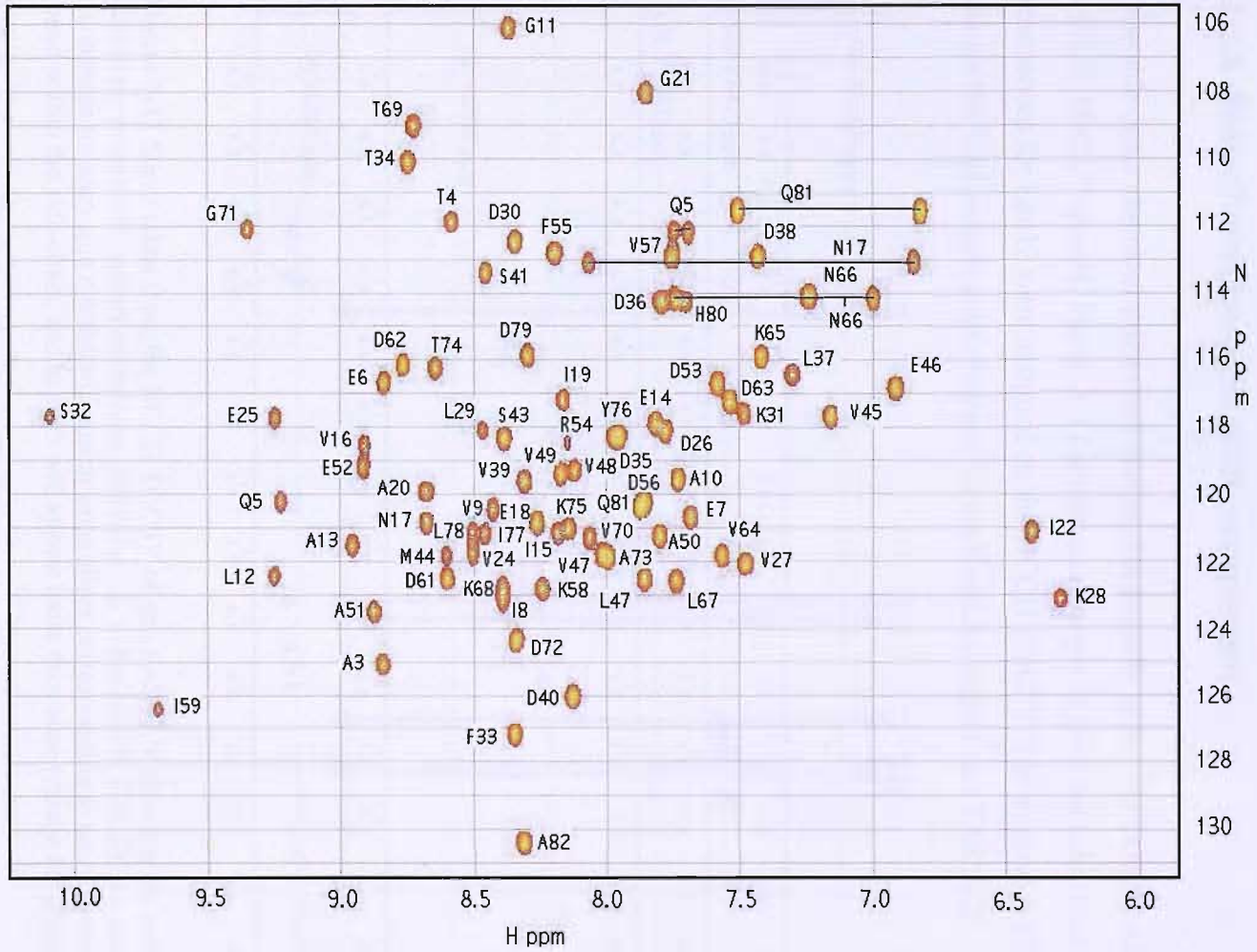


Figure 3-10 ^{1}H-^{15}N HSQC spectrum of *apo* FAS ACP from *Streptomyces coelicolor*. The residues giving rise to the amide proton chemical shift are labelled on the spectrum. The side chain amide proton peaks for residues N17, N66 and Q81 are also labelled.

3.3.4 HA chemical shift determination using HNHA experiments

A HNHA spectrum was used in conjunction with the ^{15}N HSQC and the assigned HN resonances to manually determine the Hα chemical shift assignments for all residues except Ala2, Pro23 and Pro60.

3.3.5 Side-chain chemical shift assignment

The side chain chemical shifts were assigned mannually using ^{15}N TOCSY-HSQC and HCCH TOCSY spectra. Figure 3-11 shows an example of how the side chain resonances for Leu12 were assigned on the HCCH TOCSY. The appendix (Chapter 7) shows the final assignments obtained for each residue within *apo* FAS ACP.

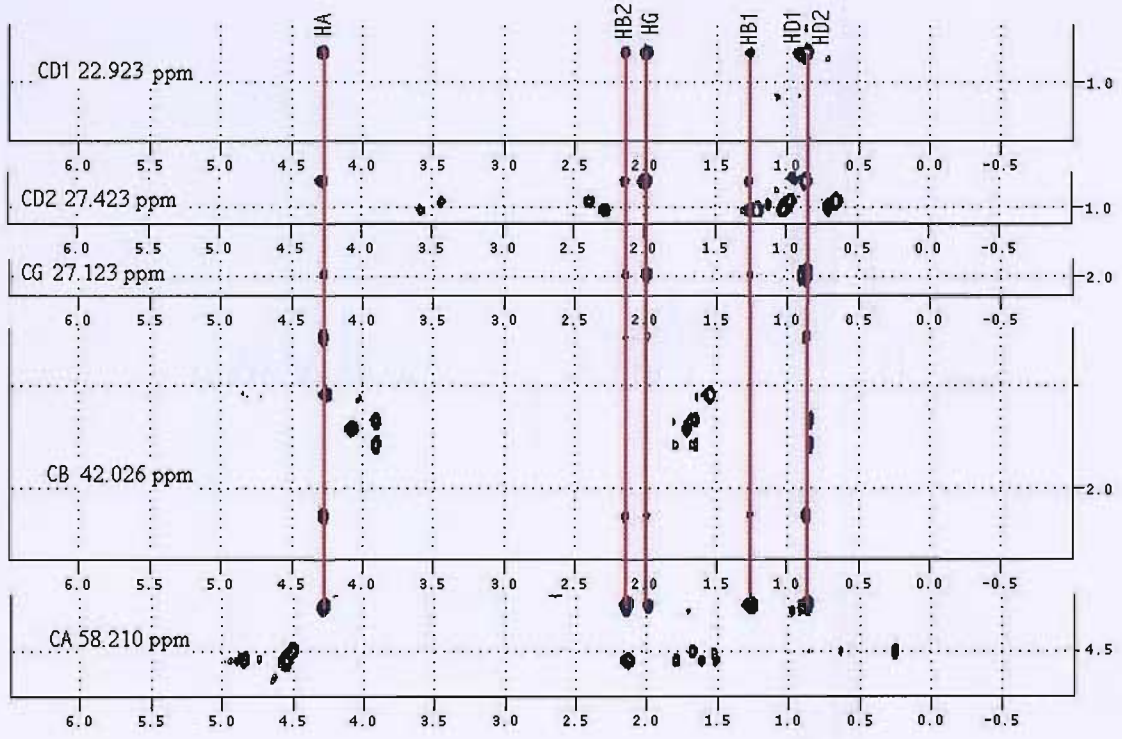


Figure 3-11 Slices taken from the HCCH TOCSY of *apo* FAS ACP showing the assignment the side-chain proton and carbon resonances of leucine 12. In the HCCH TOCSY the protons resonances for each spin system will appear upon slices corresponding to each carbon resonance found within the side-chain, and as such will appear more than once (except for glycines).

3.3.6 Aromatic chemical shift determination

To complete proton assignment for *apo* FAS ACP it was necessary to assign the chemical shifts for the aromatic side chains. These are usually buried within the structure and give rise to many NOEs which are vital to the final structure determination. A sample of unlabelled *apo* FAS ACP was dissolved in ~ 100 % D_2O at

a concentration of 6 mg/ml. A 1D spectrum was immediately recorded and the sample was left overnight at 25 °C so that the backbone amide protons exchanged with ^{2}H. This left just the aromatic protons and thus removed the usual overlap problem which prevents them being seen normally. A further 1D spectrum was then recorded to ensure that all the amide protons had exchanged. 2D TOCSY and NOESY spectra were then recorded so that the aromatic resonances could be assigned. An overlay of the 1D spectra recorded after 0 hours and 18 hours in ~ 100 % D_2O shows the amide proton exchange (Figure 3-12), whilst the assignment of the 2D TOCSY showing the aromatic region of the spectrum is illustrated in Figure 3-13.

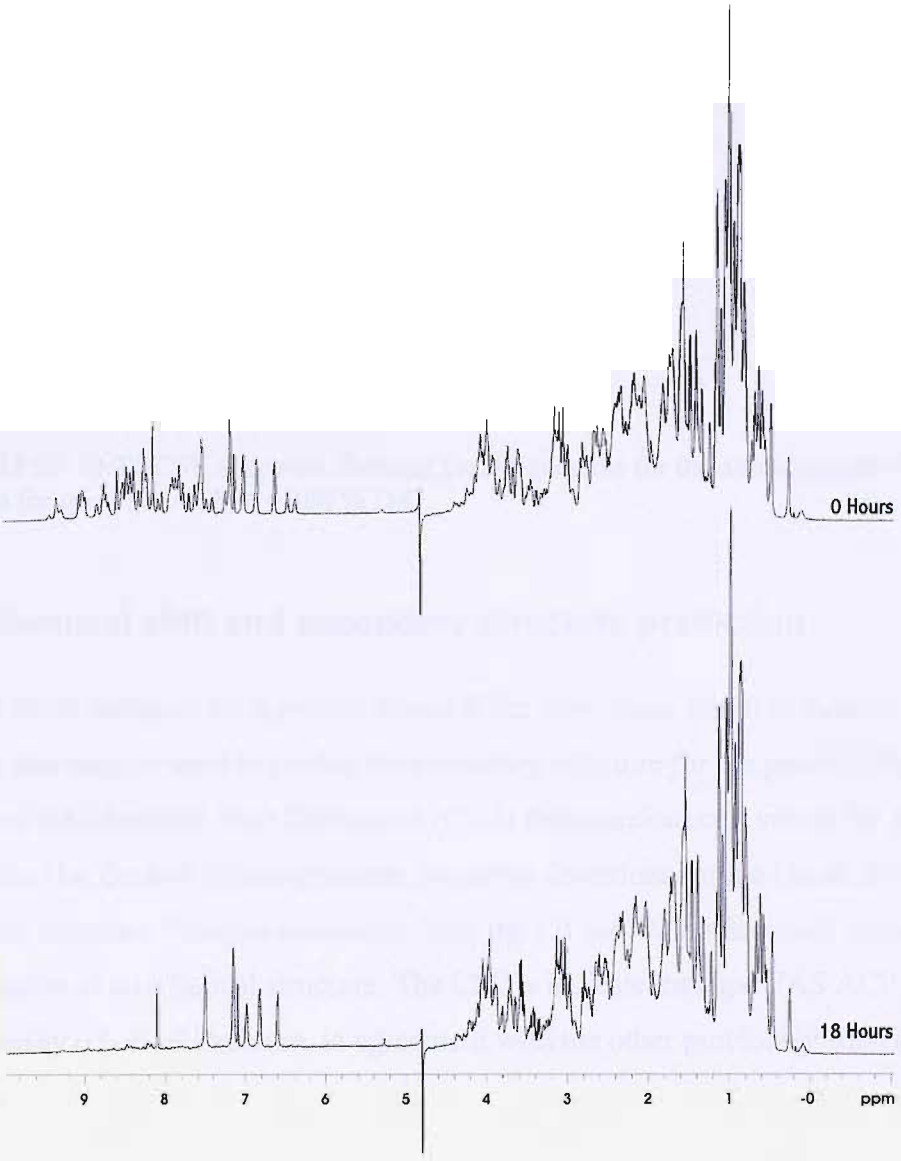


Figure 3-12 1D spectra of *apo* FAS ACP recorded after 0 hours and 18 hours in ~ 100 % D_2O. The majority of resonances seen in the 6-10 ppm region disappear after 18 hours due to proton-deuterium exchange. The remaining resonances seen in this range correspond to aromatic protons resonances which remain as the protons in the aromatic rings are unable to exchange with deuterium.

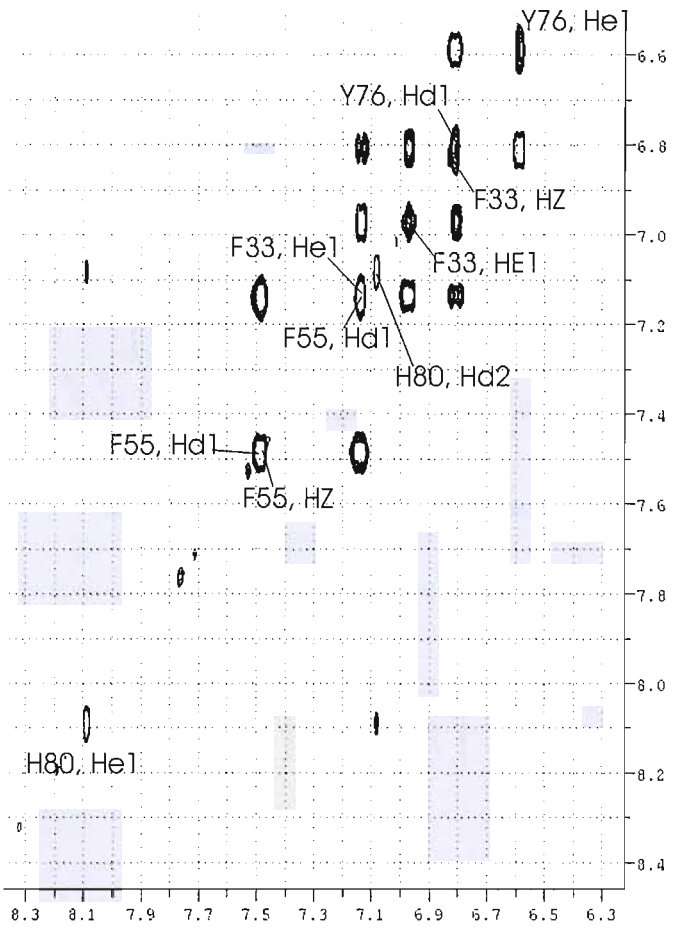
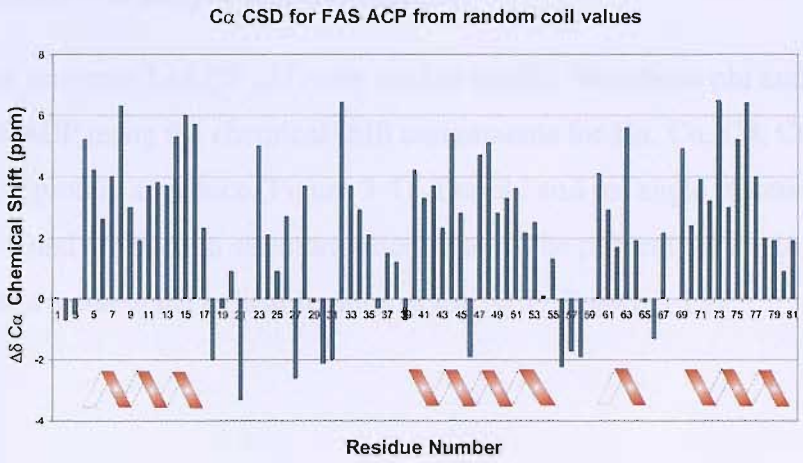
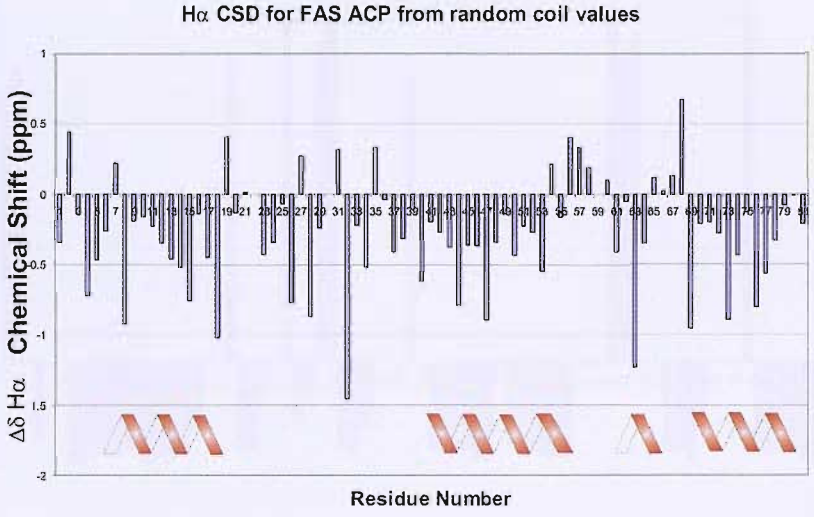


Figure 3-13 2D ^{1}H TOCSY spectrum showing the assignments for the aromatic proton resonances for *apo* FAS ACP in ~ 100 % D_2O.

3.3.7 Chemical shift and secondary structure prediction

Chemical shifts assigned for a protein which differ from those found in random coil structures and may be used to predict the secondary structure for the protein. Figure 3-14 shows the Chemical Shift Deviations (CSD) from random coil values for *apo* FAS ACP for the Hα, Cα and Co assignments. Negative deviations for the Hα shifts indicate an α helical structure. Positive deviations from the Cα and Co random coil values are also indicative of an α helical structure. The CSD's indicate that *apo* FAS ACP has a predominantly α helical structure, in agreement with the other previously solved ACP structures.





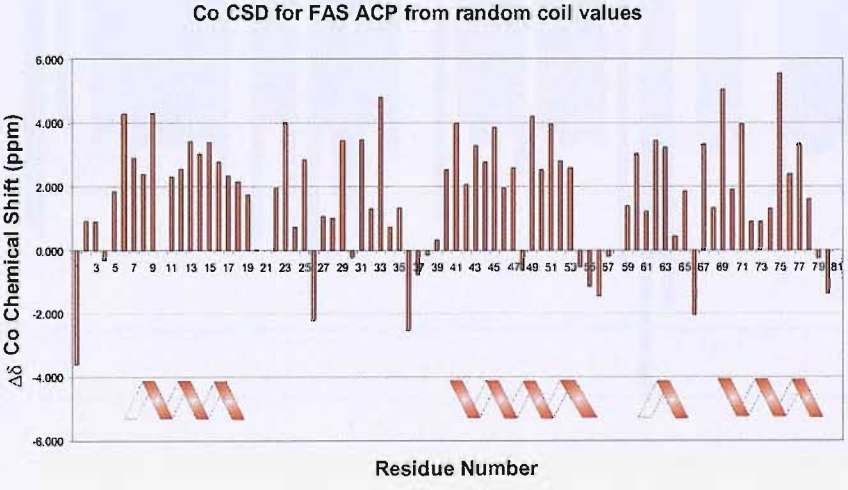


Figure 3-14 Chemical shift deviations for Cα (green), Hα (blue) and CO (red) shifts for *apo* FAS ACP indicating a predominantly α helical structure. Positive chemical shift deviations for Cα and CO and negative chemical shift deviations for Hα indicate an α helical structure.

3.3.8 Phi and Psi angle determination

The software program TALOS (*22*) was used to predict back-bone phi and psi angles for *apo* FAS ACP using the chemical shift assignments for Hα, Cα, Cβ, CO and N and the FAS ACP protein sequence (Figure 3-4). The phi and psi angle restraints are output from TALOS and are used in structure calculations. The phi and psi angle restraints may be plotted to show the helical regions of the ACP (Figure 3-15).

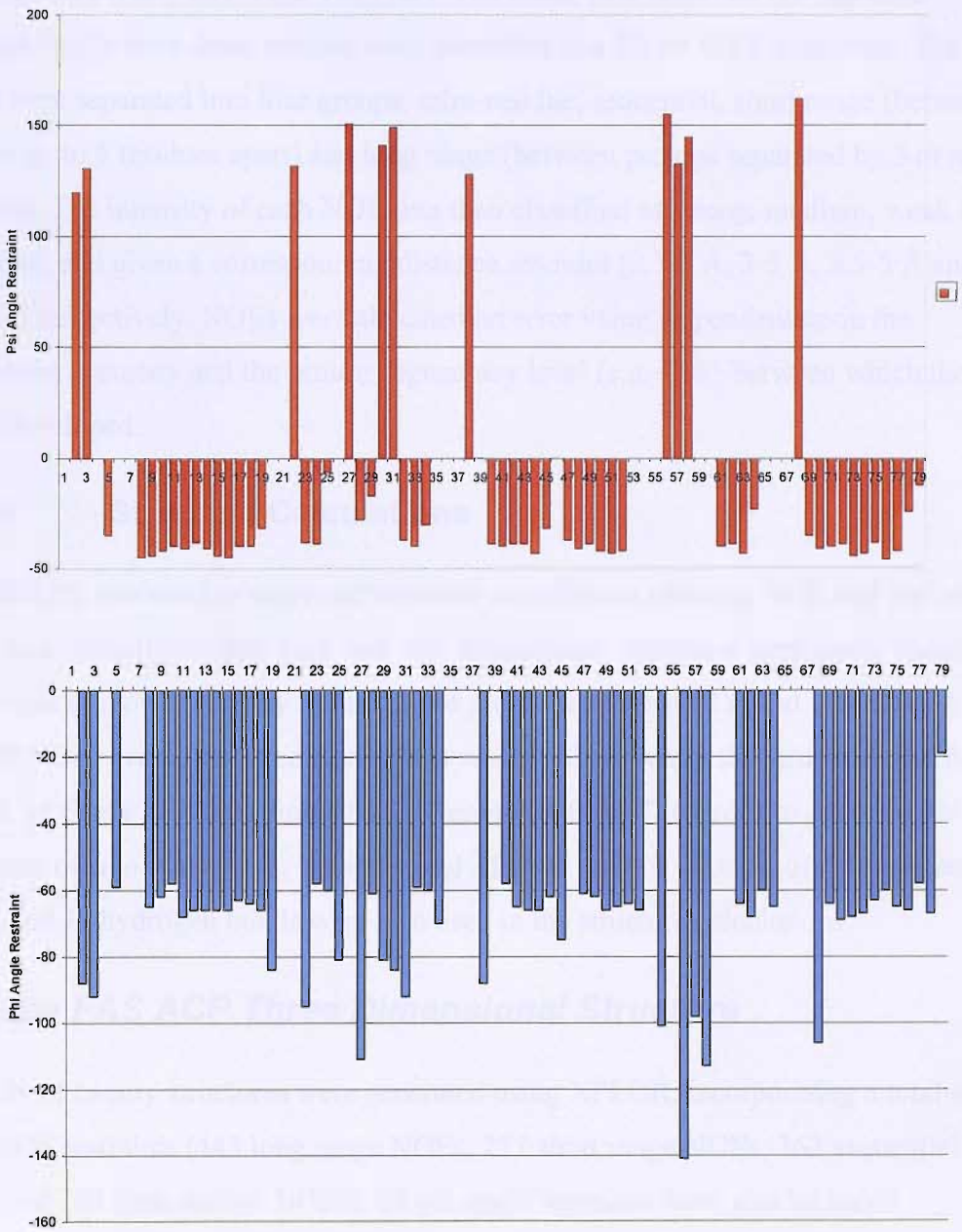


Figure 3-15 Graphical representation of the phi and psi angle restraints applied during structure calculations, with the alpha helical regions indicated. α helical regions of a protein have phi angles in the region of –58° and psi angles of approximately –47°.

3.3.9 NOE data

Backbone and side-chain NOE's

A three dimensional ^{15}N NOESY HSQC was acquired with 150 ms mixing time and used to identify NOEs between the backone amide protons and side-chain protons. A CN NOESY HSQC was also recorded to provide proton-proton NOEs edited by the attached ^{13}C chemical shift. Once the resonances for the aromatic residues (F33 (Hδ1, Hε1, Hζ), F55 (Hδ1, Hε1, Hζ), Y76 (Hδ1, Hε1), and H80 (Hδ1, Hε1)), had been assigned NOEs from these protons were identified in a 2D NOESY spectrum. The NOEs were separated into four groups, intra-residue, sequential, short range (between protons up to 5 residues apart) and long range (between protons separated by 5 or more residues). The intensity of each NOE was then classified as strong, medium, weak or very weak and given a corresponding distance restraint (2.3-5 Å, 3-5 Å, 3.5-5 Å and 4.5-5 Å) respectively. NOEs were allocated an error value, dependent upon the assignment accuracy and the proton degeneracy level (e.g. CH_3) between which the NOEs developed.

3.3.10 Structure Calculations

XPLOR (*23*) was used to carry out structure calculations utilising NOE and phi and psi angle data. Initially 20,000 high and low temperature iterations were used, reduced to 8,000 once the lowest energy model of the previous refinement round was used. A total of 1309 NOEs were identified during this work. Subsequently the structure was further refined by Chris Williams from the University of Bristol in order to produce the final structures of *apo* FAS ACP, using a total of 1811 NOEs. A total of 120 phi and psi angles and 39 hydrogen bonds were also used in the structure calculations.

3.4 apo FAS ACP Three Dimensional Structure

A family of twenty structures were generated using XPLOR, incorporating a total of 1811 NOE restraints (443 long range NOEs, 237 short range NOEs, 362 sequential NOEs and 769 intra residue NOEs). 64 phi angle restraints were also included. Statistical data on the *apo* FAS ACP structure is shown in Table 3-2. A Ramachandran plot of the ensemble of 20 structures (Figure 3-16) indicates that the majority of residues lie in the most favoured and additionally allowed regions (excluding glycine residues). The distribution of phi and psi angles from the final set of 20 *apo* FAS ACP structures are shown in Figure 3-17, with each dot representing the psi and phi angle of

an amino acid residue in the structures. Each set of angles for a particular amino acid is enclosed by a vertical rectangle, the size of which reflects the spread of angles present in the 20 structures. The smaller the rectangle the better defined the angle. The dihedral angles are similar to those seen in Figure 3-15. The alpha helical regions of the protein have smaller variance in the angles when compared to the loop regions, as would be expected, due to the more regimented secondary structure preventing flexibility in the helical regions.

apo FAS ACP was shown to possess a predominantly α helical structure (Figure 3-18). The α helical bundle consists of helix I (residues 5-19), helix II (residues 41-55), helix III (residues 61-64) and helix IV (residues 70-78). The helices are arranged so that helix I runs top to bottom of the structure from the front elevation, helix II runs bottom to top, helix III runs top to bottom and helix IV runs bottom to top. The 20 structures were overlaid over the whole structure using the program MOLMOL (*24*) (Figure 3-19). The family of structures superimposes with an RMSD value of 0.426 Å for backbone atoms and 0.651 Å for all heavy atoms. When the structures were overlaid over just the main helices (I, II, IV) these values fell to 0.265 Å and 0.633 Å. Helix I overlays with RMSD values of 0.114 Å and 0.633 Å, helix II with 0.170 Å and 0.642 Å and helix IV with 0.098 Å and 0.414 Å respectively. The lower RMSD value over just the secondary structural elements is to be expected as the loop regions have more flexibility and thus higher RMSD values. The mean structure generated during this work was superimposed with the final mean model refined by C. Williams (Figure 3-20) and gave a RMSD value of 1.506 Å over the whole structure and a , RMSD value of 0.633 Å over the backbone atoms for residues 5-19, 41-55 and 72-80. This difference may be explained by the extra 502 NOE restraints used during structure calculations, which would help provide a better overall structure.

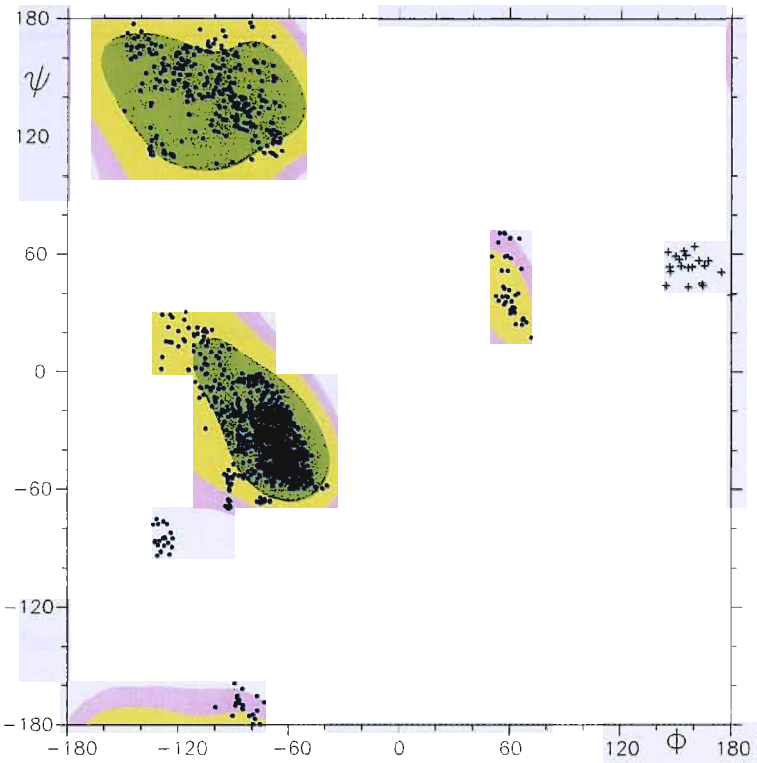


Figure 3-16 Ramachandran plot of the ensemble of 20 structures. The majority of residues lie in the most favoured and additionally allowed regions (excluding glycine residues).

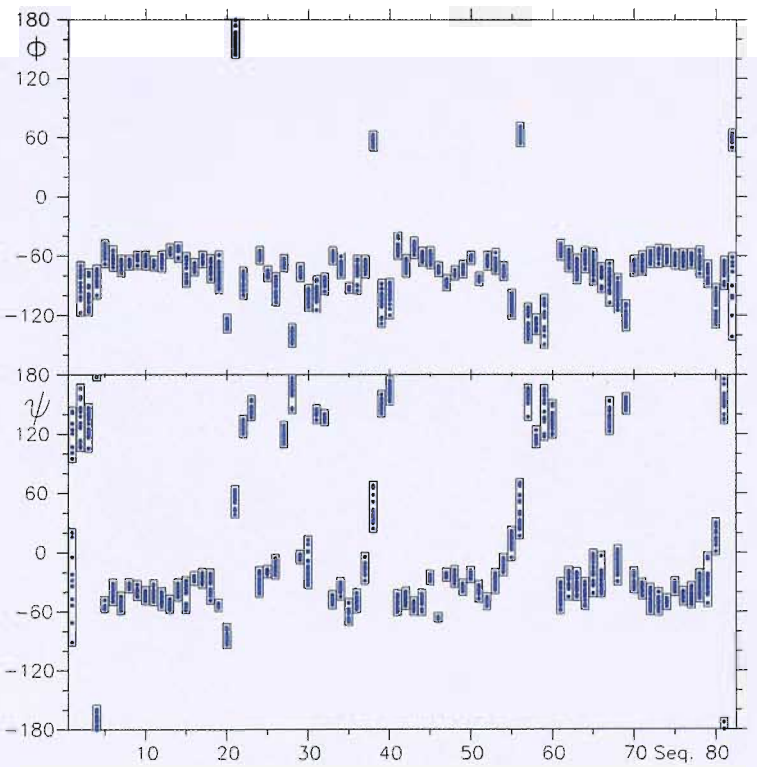
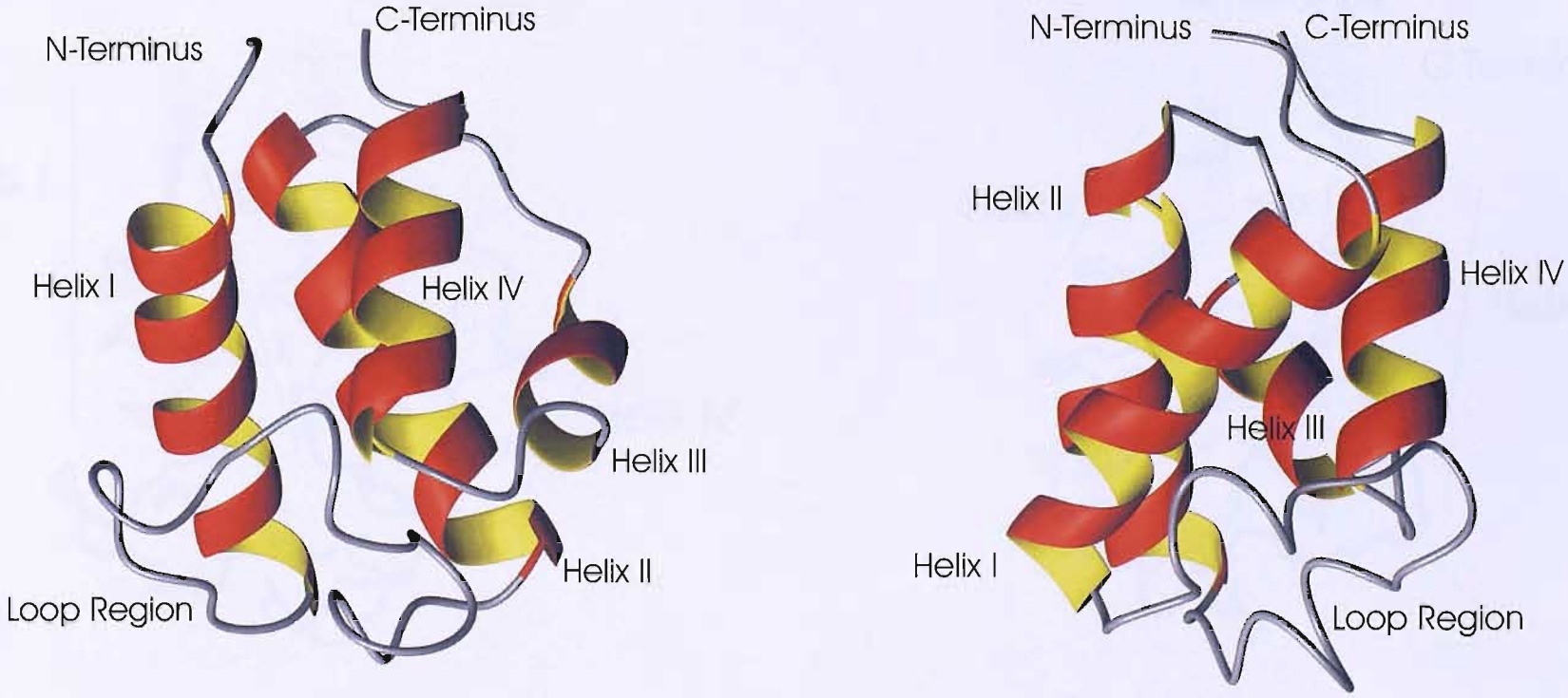


Figure 3-17 The distribution of phi and psi angles throughout the set of 20 structures. Each dot represents the psi and phi angle of an amino acid residue in the structures. Each set of angles for a particular amino acid is enclosed by a vertical rectangle, the size of which reflects the spread of angles present in the 20 structures. The smaller the rectangle the better defined the angle.

		Unambiguous	Ambiguous
Number of NOE distance restraints	intra-residue	769	38
	sequential ($\|i-j\|=1$)	362	53
	short ($1<\|i-j\|<5$)	237	54
	long ($\|i-j\|\geq 1$)	443	213
	H-bonds	39	
Number of angle restraints	dihedral angles	120	
RMSD (Å)	back-bone atoms (2-82)	0.34 +/- 0.06	
	heavy atom (2-82)	1.29 +/- 0.18	
RMSD from ideal geometry			
Angles (°)		0.5389 +/- 0.0132	
Improper (°)		0.3187 +/- 0.0185	
(phi/psi) in Ramachandran plot (%)		Representative Structure	Ensemble
residues (2-82)	core region	93.3	91.8
	additionally allowed regions	6.7	7.9
	generously allowed regions	0.0	0.3
	forbidden regions	0.0	0.0

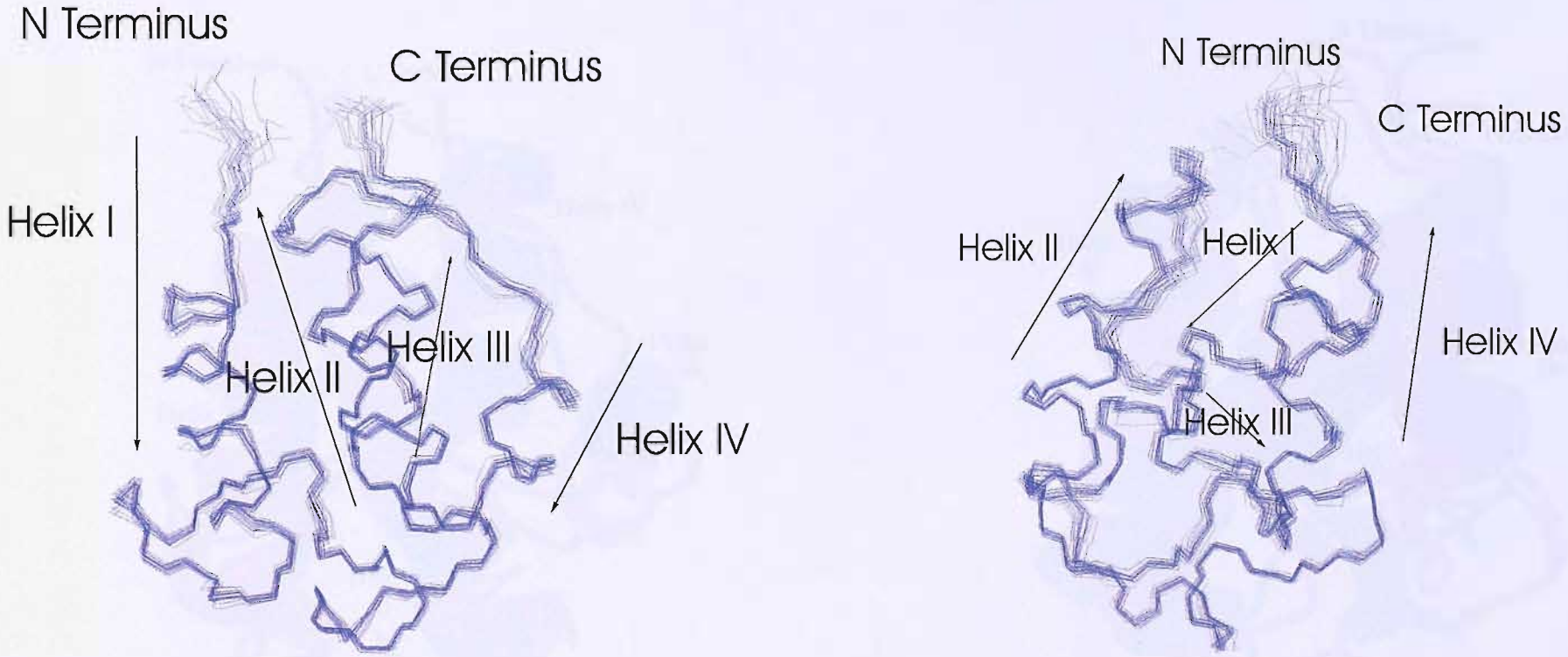
Table 3-2-Structural statistics and atomic RMS Differences for 20 Calculated *apo* FAS ACP Structures.



a) Front elevation

b) Side elevation

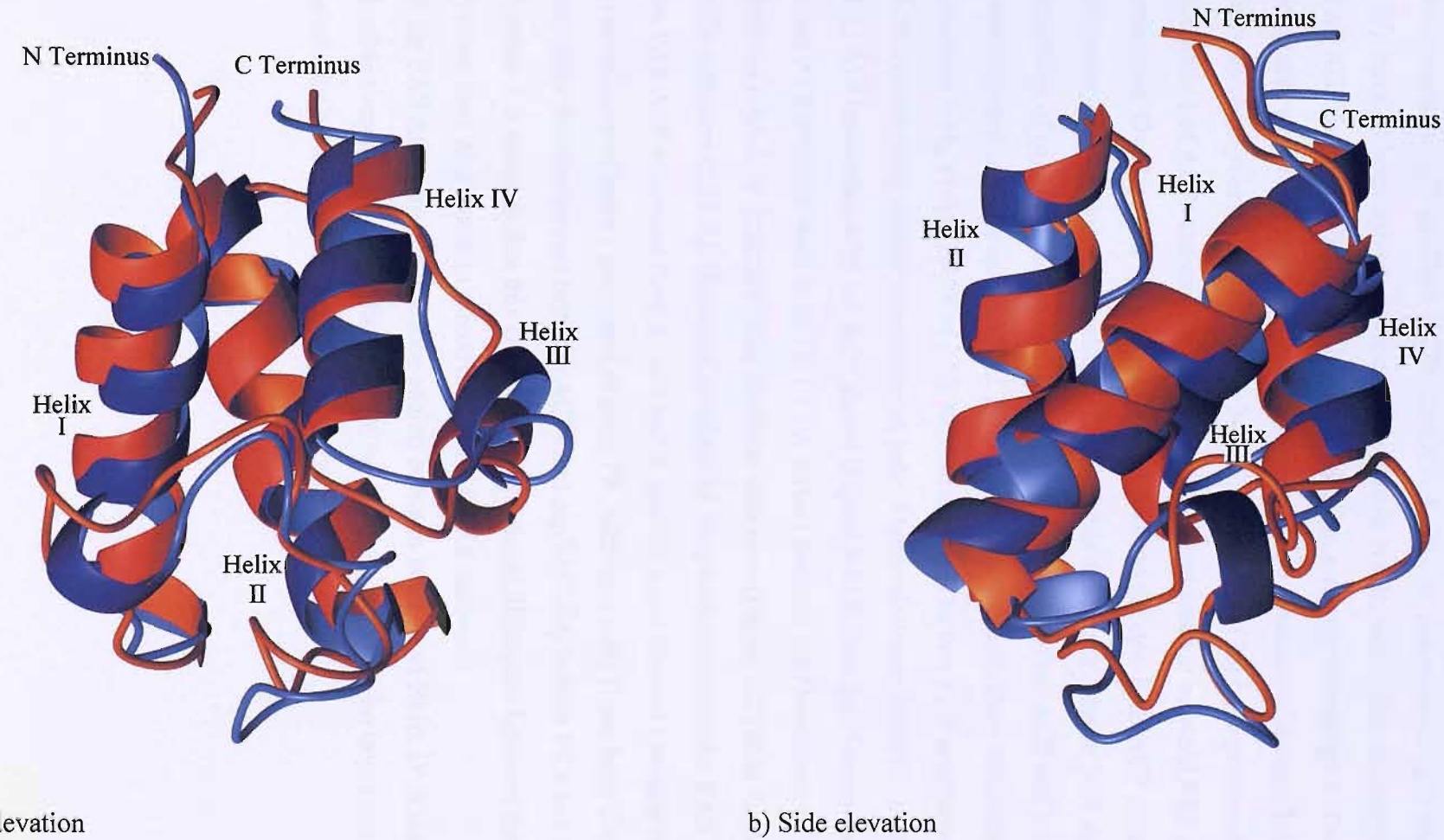
Figure 3-18 Front elevation and side elevation of the lowest energy structure of *apo* FAS ACP from *Streptomyces coelicolor*. Helix labels are shown and are used as standard throughout. The α helical bundle consists of helix I (residues 5-19), helix II (residues 41-55), helix III (residues 61-64) and helix IV (residues 70-78). The helices are arranged so that helix I runs top to bottom of the structure from the front elevation, helix II runs bottom to top, helix III runs top to bottom and helix IV runs bottom to top.



a) Front elevation

b) Side elevation

Figure 3-19 Overlays of 20 lowest energy *apo* FAS ACP structures from *Streptomyces coelicolor*. Structures were superimposed over backbone atoms for residues 5-19, 41-55 and 72-80 giving a RMSD value of 0.244 Å. The lower RMSD value over just the secondary structural elements is to be expected as the loop regions have more flexibility and thus higher RMSD values.



a) Front elevation

b) Side elevation

Figure 3-20 Overlays of the mean *apo* FAS ACP structure from *Streptomyces coelicolor* produced during this work (blue) and that of C. Williams (red). The structures were superimposed over all residues giving an RMSD 1.506 Å, and over backbone atoms for residues 5-19, 41-55 and 72-80 giving a RMSD value of 0.633 Å. This difference may be explained by the extra 502 NOE restraints used during structure calculations, which would help provide a better overall structure.

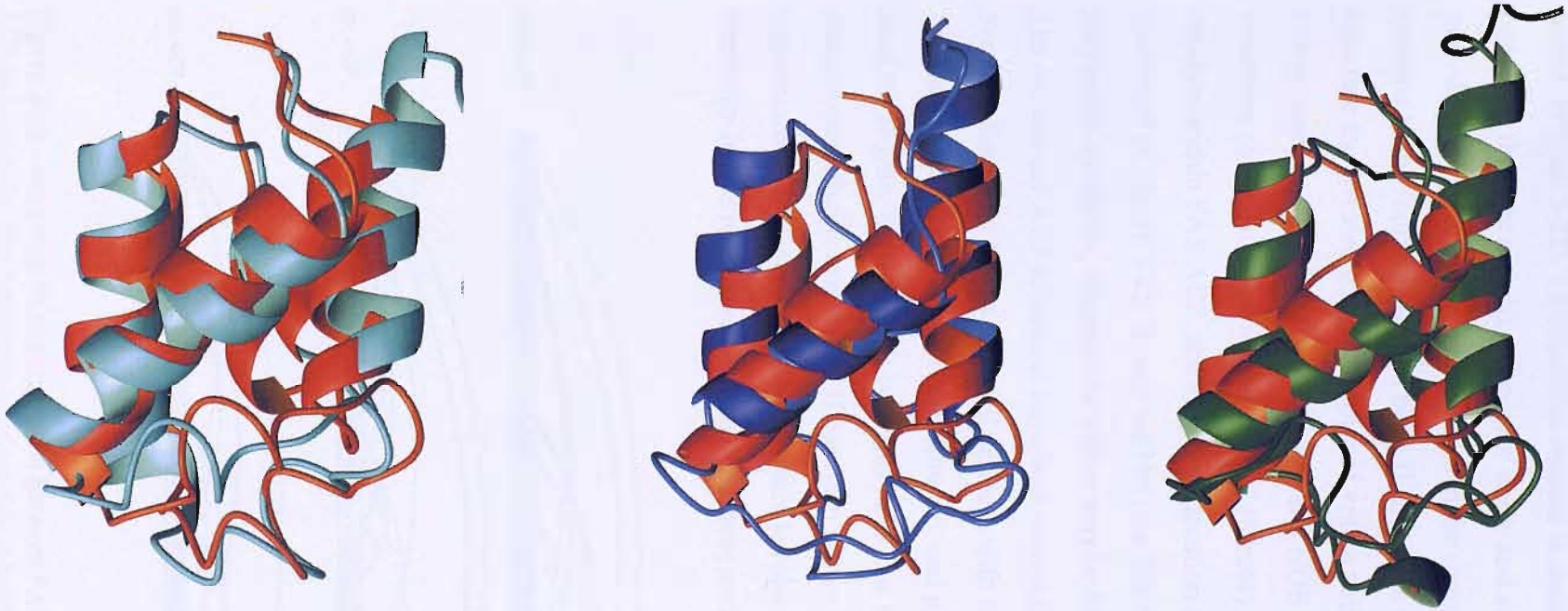
3.5 Comparison of FAS ACP Structure with other ACPs

The structures of the type II PKS ACPs from actinorhodin (*25*), oxytetracycline (*26*) and frenolicin (*27*) and FAS ACPs from *E.coli* (*28*), *M. tuberculosis* (*29*) and *B. subtilis* (*30*) have all been solved previously by solution NMR, whilst the structure of butyryl FAS ACP from *E. coli* (*31*) has been solved using x-ray crystallography. These structures were superimposed onto the lowest energy structure of the *apo* FAS ACP from *Streptomyces coelicolor* (Table 3-3). Findlow *et al.* (*26*) have previously noted that helix I of ACPs shows a distinct variability in orientation between PKS and FAS structures. Overlays of PKS and FAS structures onto the *apo*-FAS ACP structures from *Streptomyces coelicolor* have also highlighted this difference. Figure 3-21 shows the structures of (a) *Mycobacterium tuberculosis* FAS ACP, (b) *act* ACP and (c) *otc* ACP superimposed onto the *apo*-FAS ACP structure, over the secondary structural elements (residues 5-19, 41-55, 70-78 of FAS ACP structure). The two FAS structures (Figure 3-21 a) show very similar orientation of helix I with relation to helix IV (a difference of 0.63 Å).The overlay with *act* ACP shows (Figure 3-21 b) that the N-terminus of helix I in act ACP interacts with helix IV 1.72 Å further towards the N-terminus of that helix than the FAS ACP structure. This shift was also seen (Figure 3-21 c) in the *otc* PKS ACP structure (2.77 Å). However, overlays of *Streptomyces coelicolor* FAS ACP onto the FAS ACP structures from *E. coli* and *B. subtilis* again showed virtually no variation in orientation of helix I compared to helix IV. Although helix II has been shown to be important for interactions between ACPs and acpS (*32-34*) in both PKS and FAS systems it is possible that this conserved orientational difference between the two systems may play a role in selectivity for FAS/PKS enzymes.

In the FAS ACP structure the loop region between helix I and Helix IV is relatively in-flexible compared to that of the PKS ACPs. This may also play an important part in selectivity between the components of the two systems.

RMSDs relative to mean/Å	Full length	Res 5-78	Res 5-19, 41-55, 70-78
act ACP *S. coelicolor*	4.725	4.546	3.562
otc ACP *S. rimosus*	3.517	3.561	2.930
fren ACP *S. roseofulvus*	5.624	5.589	5.859
FAS ACP *M. tuberculosis*	2.077	1.943	1.581
FAS ACP *B. subtilis*	8.938	9.050	7.661
FAS ACP *E. coli*	8.868	8.898	7.539
Butyryl FAS ACP *E. coli*	9.004	9.041	7.617
Type I FAS ACP from Rat	8.121	7.970	7.617

Table 3-3 Summary of RMDS values derived from superimposing structurally characterised type II fatty acid and polyketide ACPs over various regions of the FAS ACP from *Streptomyces coelicolor*. (All values were calculated using the "fit" routine in MolMol (*24*) over heavy atoms).



a) FAS ACP and FAS ACP *M. tuberculosis* b) FAS ACP and *act* ACP c) FAS ACP and *otc* ACP

Figure 3-21 *apo* FAS ACP from *Streptomyces coelicolor* (red) overlaid with a) FAS ACP from *Mycobacterium tuberculosis* (with the c-terminal tail removed)-pale blue, b) *act* ACP from *Streptomyces coelicolor*-dark blue and c) *otc* ACP from *Streptomyces rimeous* (with the c-terminal tail removed) - green. (Structures overlayed over residues 5-19, 41-55, and 70-78 of FAS ACP). The loop region between helix I and helix II is better conserved between the two FAS structures than between the FAS ACP and the PKS ACP. The two FAS structures (a) show very similar orientation of helix I with relation to helix IV (a difference of 0.63 Å).The overlay with *act* ACP shows (b) that the N-terminus of helix I in *act* ACP interacts with helix IV 1.72 Å further towards the N-terminus of that helix than the FAS ACP structure. This shift was also seen (c) in the *otc* PKS ACP structure (2.77 Å).

In both the *act* and *otc* structures, a tyrosine residue is found two residues prior to the start of helix II (two residues before the serine to which the phosphopantetheine is added) (Figure 3-22). This tyrosine residue is seen to have an NOE interaction to a residue at the base of helix I (Leucine in *otc* and valine in *act*), and to a phenylalanine preceding the tyrosine by five residues in the sequence There is also an NOE from the tyrosine to the fourth residue in helix II, a leucine in both cases. In the *apo*-FAS ACP structure the tyrosine has been replaced with a valine, which has no NOE to end of helix I or to phenylalanine 33. There is however a NOE matching those seen within the other structures to the fourth residue in helix II (Met44). A graphical representation of these residues within FAS ACP, showing their location and orientation within the structure, is illustrated in Figure 3-23. It may well be that this tyrosine plays an important role in polyketide synthesis, whereas the valine may be important in fatty acid biosynthesis. The *otc* and *act* ACP structures have both shown that, when families of structures were overlaid, that helix II was less well defined, with higher RMSD values. The *apo*-FAS ACP structure also shows this phenomenon, and it is possible that this is a result of the need for slightly more flexibility in this helix due to the addition of the phosphopantetheine arm. Helix II has been shown to be the location at which interactions take place with other enzymes involved in PKS/FAS synthesis, and the flexibility in this region would help with these interactions.

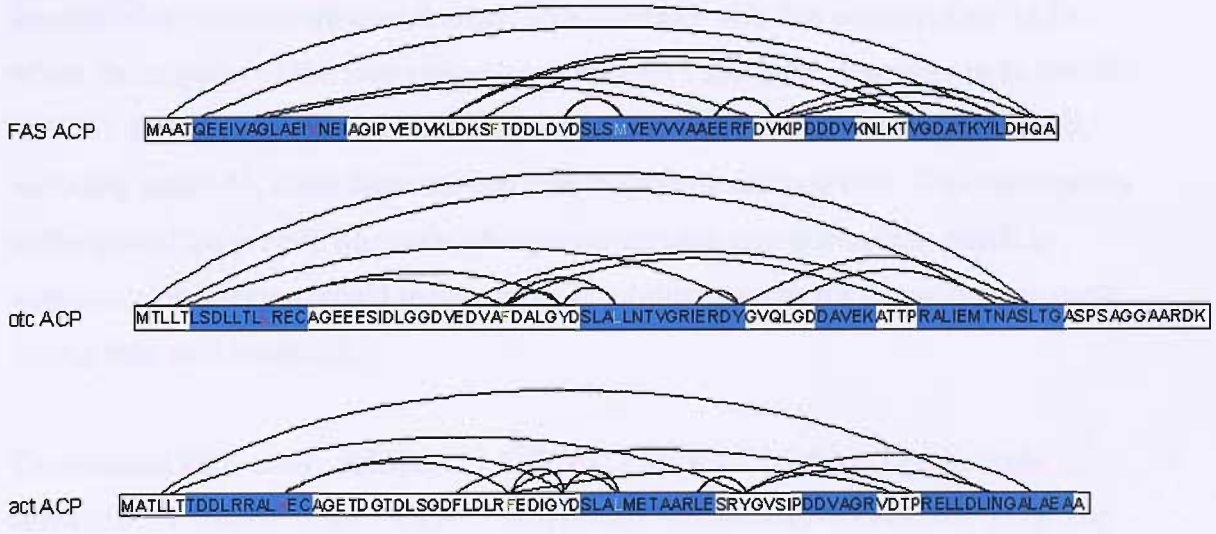


Figure 3-22 Long range NOE comparison between FAS ACP from *Streptomyces coelicolor, otc* ACP and *act* ACP. The tyrosine residue preceding helix II by 2 residues, present in *otc* and *act* ACPs is replaced by a valine residue in FAS ACP (Blue). The NOEs seen from these tyrosine residues in *act* and *otc* are not present in the FAS data. However the observed NOE from the tyrosine to the fourth residue of helix II is also seen in the FAS structure from valine 39 to methionine 44.

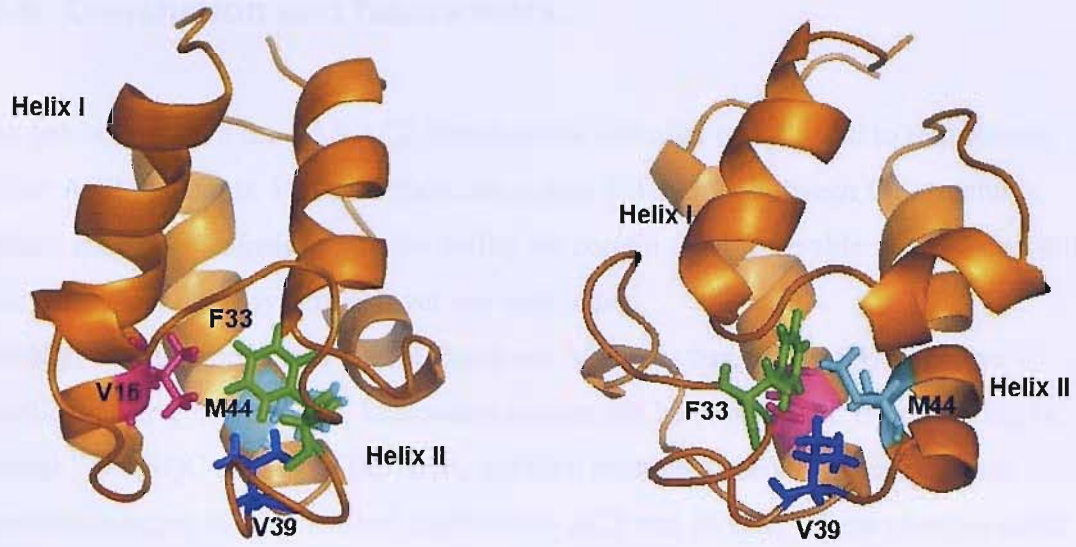


Figure 3-23 Illustration of val39 (blue) in FAS ACP from *Streptomyces coelicolor*, and met44 to which a long range NOE is found. This NOE is consistent with those seen in *act* and *otc* between the equivalent tyrosine and leucine residues. In the *act* and *otc* structures there exists NOEs from the tyrosine residue to a phenylalanine residue (phe33 in FAS ACP) (green) and to a residue, equivalent to val16 in FAS ACP (pink). These two NOEs are not present in the FAS ACP structure.(*35*)

Dynamic studies have been undertaken on a ^{15}N *apo*-FAS ACP sample, by Peter Dunford, the University of Southampton. These studies have shown that FAS ACP is a fairly rigid molecule, with few residues exhibiting fast internal motion or a susceptibility to exchange contribution. This contrasts with that seen in other ACPs, where the majority of the loop region between helix I and helix II appears to be flexible (*27, 36*). The c-terminal end of this loop region and the first few residues of helix II, including serine 41, show faster motion than the rest on the molecule. This corresponds to the area of the protein where the phosphopantetheine arm is attached, which is expected as this region would require more flexibility in order for interactions to occur during fatty acid synthesis.

The minimal PKS assay, utilising *act* ACP and PKS KS/CLF from *Streptomyces coelicolor*, is inactive if the PKS ACP is replaced with the native FAS ACP (*37*). The *act* ACP has been shown to be able to self-malonate (*38, 39*) *in vitro*, where the FAS ACP does not. The actinorhodin gene cluster in *Streptomyces coelicolor* is lacking a gene encoding for MCAT which carries out this transfer, although *act* ACP has been shown to undergo malonation by the FAS MCAT *in vitro* at a rate higher than that observed in self-malonation.

3.6 Conclusion and future work.

As has been shown the FAS ACP structure has a similar overall fold to that seen in other ACP structures. However there are subtle differences between the structures which may give an insight into the ability for certain ACP to be able to cross-talk with enzymes from other system and yet not with others.

Ideally, now that both the FAS ACP and *act* ACP structures from *Streptomyces coelicolor* have been solved, interaction studies can be undertaken. By recording an initial ^{15}N HSQC spectra of the ACP, and then titrating a second protein into the sample, changes in the chemical shifts of the ACP can be seen. These changes occur when the environment surrounding the proton changes, as occurs when binding happens. In this way the region of the ACP that is involved in the binding to the second molecule can be identified. By carrying out these experiments on enzymatic components from both FASs and PKSs e.g. Acps, MCAT, KS/CLF, KR, it is hoped that the differences, and similarities, between the two systems may be elucidated.

3.7 References

1. Crump, M. P., Crosby, J., Dempsey, C. E., Murray, M., Hopwood, D. A., and Simpson, T. J. (1996) *FEBS Lett. 391*, 302-6.
2. Bodenhausen, G., and Ruben, D. J. (1980) *Chem. Phys. Letts. 69*, 185.
3. Fesik, S. W., and Zuiderweg, E. P. R. (1990) *Quart. Rev. Biophys. 23*, 97-131.
4. Bax, A., and Grzesiek, S. (1993) *Accounts of Chemical Research 26*, 131.
5. Marion, D. D., P.C., Kay, L.E., Wingfield, P.T., Bax, A., Gronenborn, A.M. and Clore, G.M. (1989) *Biochemistry 28*, 6150-6156.
6. Jahnke, W., Baur, M., Gemmecker, G., and Kessler, H. (1995) *J. Mag. Reson. 106*, 86.
7. Marion, D., Kay, L.E., Sparks, S.W., Torchia, D.A. and Bax, A. (1989) *J. Am. Chem. Soc. 111*, 1515-1517.
8. Zuiderweg, E. R., and Fesik, S. W. (1989) *Biochemistry 28*, 2387-2391.
9. Vuister, G. W. a. B., A. (1993) *J. Am. Chem. Soc. 115*, 7772-7777.
10. Kay, L. E., Ikura, M., Tschudin, R. and Bax, A. (1990) *J. Magn. Reson. 89*, 496-514.
11. Grzesiek, S., and Bax, A. (1992) *J. Magn. Reson. 96*, 432-440.
12. Muhandiram, D. R., and Kay, L. E. (1994) *J. Mag. Reson. 103*, 203.
13. Farmer, B. T., 2nd, Venters, R. A., Spicer, L. D., Wittekind, M. G., and Muller, L. (1992) *J. Biomol. NMR 2*, 195-202.
14. Wittekind, M., and Mueller, L. (1993) *J. Magn. Reson. 101*, 201.
15. Bax, A. a. I., M. (1991) *J. Biomol. NMR 1*, 99-104.
16. Grzesiek, S. a. B., A. (1992) *J. Am. Chem. Soc. 144*, 6291-6293.
17. Bax, A., Clore, G.M. and Gronenborn, A.M. (1991) *J. Magn. Reson. 88*, 425.
18. Sattler, M., Maurer, M., Schleurer, J. and Griesinger, C. (1995) *J. Biomol. NMR 5*, 97.
19. F. Delaglio, S. G., G. W. Vuister, G. Zhu, J. Pfeifer and A. Bax:. (1995) *J. Biomol. NMR 6*, 277-293.
20. Zimmerman, D., Kulikowski, C., Wang, L., and Lyons, B. (1994) *J. Biomol. NMR 4*, 241.
21. Zimmerman, D. E., Kulikowski, C.A., Huang, Y., Feng, W., Tashiro, M., Shimotakahara, S., Chien, C., Powers, R., Montelione, G.T.,. (1997) *J. Mol. Biol. 269*, 592-610.
22. Cordier, F., Dingley, A. J., and Grzesiek, S. (1999) *J. Biomol. NMR 13*, 175-80.
23. Brunger, A. T. (1995), Newhaven, CT.
24. Koradi, R., Billeter, M., and Wüthrich, K. (1996) *J. Mol. Graphics. 14*, 51-55.
25. Crump, M. P., Crosby, J., Dempsey, C. E., Parkinson, J. A., Murray, M., Hopwood, D. A., and Simpson, T. J. (1997) *Biochemistry 36*, 6000-8.
26. Findlow SC, W. C., Simpson TJ, Crosby J, Crump MP. (2003) *Biochemistry 42*, 8423-33.
27. Li, Q., Khosla, C., Puglisi, J.D. and Liu, C.W. (2003) *Biochemistry 42*, 4648-4657.
28. Holak, T. A., Nilges, M., Prestegard, J. H., Gronenborn, A. M., and Clore, G. M. (1988) *Eur. J. Biochem. 175*, 9-15.
29. Wong, H. C., Liu, G., Zhang, Y. M., Rock, C. O., and Zheng, J. (2002) *J. Biol. Chem. 277*, 15874-80.
30. Xu, G. Y., Tam, A., Lin, L., Hixon, J., Fritz, C. C., and Powers, R. (2001) *Structure (Camb) 9*, 277-87.

31. Roujeinikova, A., Baldock, C., Simon, W. J., Gilroy, J., Baker, P. J., Stuitje, A. R., Rice, D. W., Slabas, A. R., and Rafferty, J. B. (2002) *Structure (Camb) 10*, 825-35.
32. Worsham, L. M., Earls, L., Jolly, C., Langston, K. G., Trent, M. S., and Ernst-Fonberg, M. L. (2003) *Biochemistry 42*, 167-76.
33. Flaman, A. S., Chen, J. M., Van Iderstine, S. C., and Byers, D. M. (2001) *J. Biol. Chem. 276*, 35934-9.
34. Parris, K. D., Lin, L., Tam, A., Mathew, R., Hixon, J., Stahl, M., Fritz, C. C., Seehra, J., and Somers, W. S. (2000) *Structure Fold. Des. 8*, 883-95.
35. DeLano, W. N. (2002), DeLano Scientific, San Carlos, CA, USA.
36. Findlow, I. S. F., Winsor, C., Simpson, T. J., Crosby, J., and Crump, M. P. (2003) *Biochemistry 42*, 8423-33.
37. Revill, W. P., Bibb, M. J., and Hopwood, D. A. (1996) *J. Bacteriol. 178*, 5660-7.
38. Matharu, A. L., Cox, R. J., Crosby, J., Byrom, K. J., and Simpson, T. J. (1998) *Chem. Biol. 5*, 699-711.
39. Hitchman, T. S., Crosby, J., Byrom, K. J., Cox, R. J., and Simpson, T. J. (1998) *Chem. Biol. 5*, 35-47.

Chapter 4

Crystallisation and Structural Studies of *gris* ACP from the Polyketide Synthase of *Streptomyces griseus*

4. Crystallisation and Structural Studies of gris ACP from the Polyketide Synthase of *Streptomyces griseus.*

Current structural work on PKS and FAS ACPs has, with the exception of the X-ray structure of butyryl ACP from *E. coli* (*1*), been solved using NMR. However, NMR structures of ACPs have only shown data consistent with monomeric ACP forms, while experimental data derived by alternative means indicate that ACPs associate to form dimeric structures *in vitro*(*2-4*).

Work by Zhou *et al.*(*2*) has shown that the ACPs involved in the biosynthesis of the polyketides tetracenomycin and frenolicin appear in a multimeric form (possibly dimeric) after gel filtration experiments in the presence of DTT. Cysteine free PKS ACPs have also been analysed using electrospray-mass spectrometry (*3, 4*) and native polyacrylamide gel electrophoresis, again indicating that they are able to form dimeric structures.

It is hoped that by using X-ray crystallography to solve a PKS ACP structure that essential interactions between monomers may be identified.

4.1 Practical methodology and results

4.1.1 Crystallisation trials

apo gris ACP was expressed in a modified pET7 vector, supplied by The University of Bristol, by heat induction of a 1.5 A_{600} culture of *E. coli* at 42 °C for thirty minutes followed by growth at 30 °C for three hours. ACP was then purified by anionic exchange chromatography and the purity evaluated using 12% SDS PAGE analysis (Figure 4-1) prior to crystallisation. A faint contaminating band is visible at ~7 kDa, which is possibly a degradation product.

Molecular Dimension screens 1 and 2 (MD1 and MD2) were used for initial crystallisation condition screening. Screens were set up using the hanging drop method, with drops initially comprising 2 µL protein solution (25 mg/mL *gris* ACP in 25 mM

Tris, pH 8) and 2 μL well solution. Initial hits resulting in crystal formation are shown in Table 4-1.

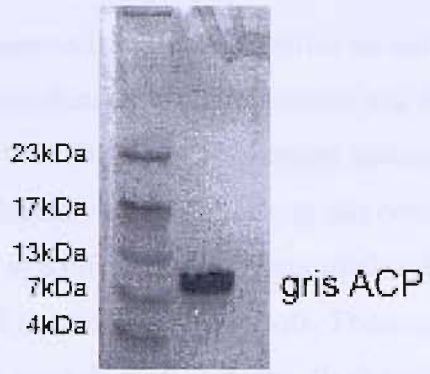


Figure 4-1SDS PAGE analysis (12 %) of *apo gris* ACP, showing a faint contaminating band at ~7 kDa.

Hit Number	Time for crystal growth	Precipitant	Salt	Buffer	pH	Diffraction Results
Hit 1	1 day	30 % PEG 4000	0.2 M $MgCl_2.6H_2O$	0.1 M Tris.HCl	8.5	Salt
Hit 2	2 days	50 % v/v MPD	0.2 M $NH_4H_2PO_4$	0.1 M Tris.HCl	8.5	Salt
Hit 3	2 days	1.6 M $(NH_4)_2SO_4$	0.1 M NaCl	0.1 M HEPES	7.5	Salt
Hit 4	1 month	0.1 M NaH_2PO_4, 0.1 M KH_2PO_4	2 M NaCl	0.1 M MES	6.5	Salt
Hit 5	2 months	30 % v/v MPD	0.2 M $(CH_3COO)_2Mg.4H_2O$	0.1 M sodium cacodylate	6.5	Poor Diffraction
Hit 6	2 weeks – 6 months	$(NH_4)_2SO_4$	NaCl	0.1 M Tris		2.5 Å Data Set
Hit 7	~ 1 month	$(NH_4)_2SO_4$	NaCl	0.1 M Tris		1.8 Å Data Set

Table 4-1 Initial MD1 and MD2 conditions producing crystals for the apo *gris* ACP screens. The crystals from hit 1 and hit 2 were sent to the ESRF synchrotron in Grenoble and the crystals from the other conditions were processed in-house using the in house rotating anode generator/ MAR detector system. Crystals in hit conditions 6 and 7 formed over a variety of pH values.

Crystals produced in hits 1 and 2 were taken to the ESRF Synchrotron source at Grenoble, France where low temperature diffraction data indicated that they were composed of salt. Diffraction data on crystals produced in hits 3 and 4 were wet mounted and recorded in-house and was also identified as salt. Refinement of the conditions for hit 5, using a combination of grid screens and full factorial screens produced crystals in 30 % v/v MPD, 0.2 M magnesium acetate tetrahydrate, 0.1 M sodium cacodylate, pH 6.0. Initially one crystal from this condition was flash cooled using a cryostream to 100 K and diffraction data recorded in-house. Diffraction data (not shown) took the form of rings of smudged spots. These spots were not consistent with those resulting from ice crystals formed during flash cooling in a cryostream, or with those expected from salt crystals. Salt crystals diffract giving disperse diffraction spots originating from tiny unit cells, giving rise to large spot separations in reciprocal space. Diffraction data were also obtained from just the loop used to support the crystal in case there was contamination, however no diffraction was observed. Closer examination of the crystals and screens suggested that the crystals were indeed protein in nature due to the diffraction type, but were very badly ordered and with a small unit cell.

In an attempt to produce higher quality crystals for use in diffraction studies a 30 mg/mL sample of *gris* ACP protein was prepared in 20 mM Tris, 100 mM NaCl at pH 8 and applied to a calibrated Superose 12 gel filtration column equilibrated in the same buffer. A protein peak (Figure 4-2) was seen to elute with a molecular weight of ~20,000 Da indicating that the *gris* ACP exists in the form of a dimer under these conditions (there are no cysteine residues in *gris* ACP therefore this dimer is not due to the formation of disulphide bonds). A smaller contaminating peak was seen to elute at a much lower molecular weight than the main peak. Bradford assays showed that only the first peak contained protein. The first peak was concentrated to ~25 mg/mL using a Vivaspin with molecular weight cut off 5,000 Da at 3,000 g, and new grid and full factorial screens were set-up. However, no crystals were seen to form with the new protein sample.

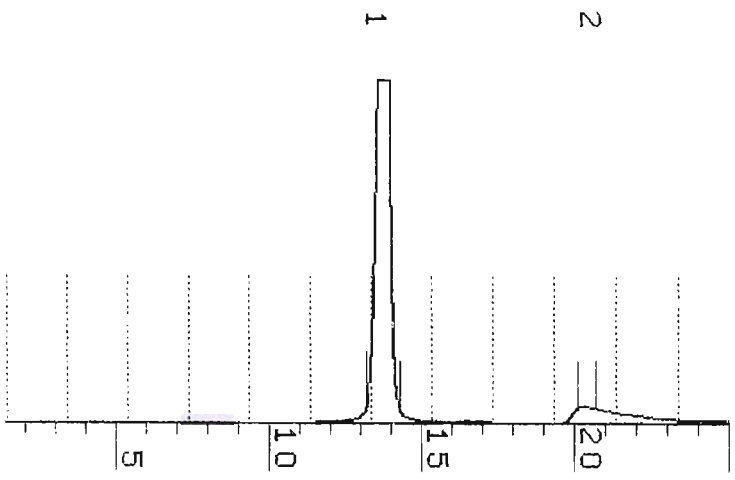


Figure 4-2 Integrated FPLC trace obtained from a Superose 12 gel filtration column of *apo gris* ACP at 30 mg/mL in 20 mM Tris, 100 mM NaCl at pH 8, showing possible dimeric form eluting at ~13.5 mL (~20 kDa) and a second lower molecular weight contaminant eluting just after 20 mL.

Small, rod shaped crystals were observed from hit 6, in conditions containing ammonium sulphate, Tris HCl and NaCl. Grid screens and full factorial screens were set-up around these conditions in an attempt to produce better quality crystals, using protein purified using a Superose-12 gel filtration column. Crystals were formed under a variety of different conditions, in a seemingly random fashion, independent of precipitant concentration and pH. It was noted that in the conditions in which crystals did form, the wells were not fully sealed by the vacuum grease, allowing greater vapour diffusion to occur than would normally be found with a hanging drop condition. Attempts to narrow down the crystallisation conditions were unsuccessful as crystallisation did not occur in fully sealed wells. Crystallisation was attempted using a modified hanging drop method, where a greater distance between the mother liquor and the drop was used. This method also proved unsuccessful. Following reports (*5-8*) that opening wells for a short time may induce nucleation, some of the sealed conditions not containing crystals were opened to the air for up to 30 seconds before being resealed. Crystals were then seen to form in some of these conditions (examples of which are shown in Figure 4-3) which were taken forward for data collection.

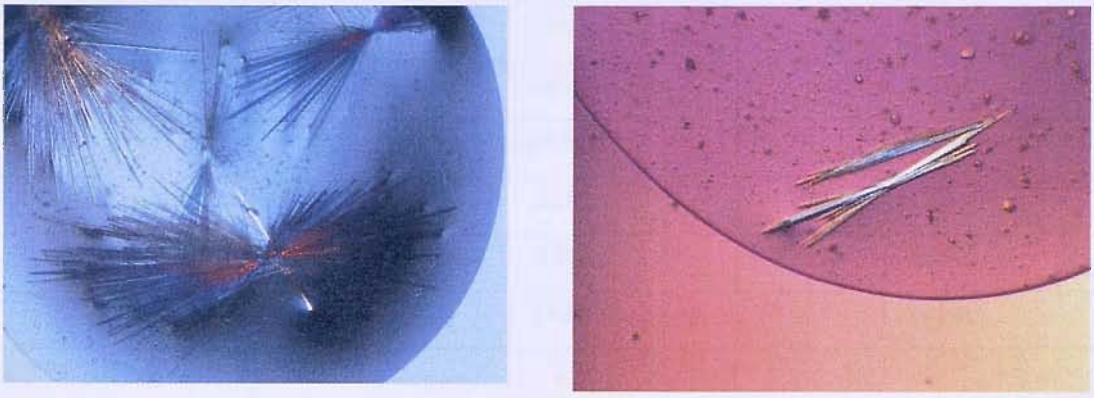


Figure 4-3 Examples of typical *apo gris* ACP crystals grown in conditions containing ammonium sulphate, tris HCl and NaCl used to collect diffraction data in house.

4.1.2 Data Collection and Processing

Initially a crystal from the Molecular Dimensions screen (hit 6) was wet mounted in a capillary tube and data was collected in-house using an Enarf-Nonius CuK_{α} rotating anode generator/ MAR detector system. The crystal was shown to diffract to 2.5 Å and a data set of 180 images was recorded. A crystal taken from one grid screen condition (exact condition unknown due to drying of well condition) was subsequently flash-cooled in a mohair loop, using a cryostream, in mother liquor containing: 1.6 M ammonium sulphate, 0.1 M Tris pH 7.5, 0.1 M NaCl and 20 % glycerol, the latter being applied to the mother liquor as a cryo-protectant. This crystal was shown to diffract to ~1.8 Å and a data set was collected in-house. Both data sets were then processed using the methods described in Chapter 5.

Data collection parameters for the two data sets can be seen below in Table 4-2, with the beam centre being determined using IPDISP.

Maximum Resolution	2.5Å	1.8Å
Xc	150.714	150.44
Yc	149.846	149.95
Crystal to Detector distance	230	120
Number of images	180	208
λ	1.54180	1.54180
Oscillation angle (ΔΦ)	2	1

Table 4-2 Data collection parameters for the two *gris* ACP data sets, processed using IPDISP and MOSFLM.

The resolution of the data was initially estimated by adjusting the background of one diffraction image and noting the highest resolution of those spots which could easily be observed. The optimal resolution was subsequently determined from the I/sigI tables output by the data processing package MOSFLM.

Both data sets were processed using MOSFLM. The distortion index tables determined using the REFIX algorithm are listed in Table 4-3 and Table 4-4. Usually the correct spacegroup, determined from the distortion index table, is the one with the lowest penalty and highest symmetry, i.e. there will usually be a significant jump in penalty from the correct solution to the next one. This jump can be seen in both Table 4-3 and, Table 4-4 however there are several solutions which are very close in penalty.

No	PENALTY	LATT	a	b	c	alpha	beta	gamma	Possible spacegroups
15	154	hP	44.18	44.72	84.64	90.3	90.2	90.9	P3,P31,P32,P312,P321,P3112,P3121,P3212,P3221P6,P61,P65,P62,P64,P63,P622,P6122,P6522,P6222, P6422,P6322
14	153	mC	174.81	44.18	44.72	90.9	90.5	75.6	C2
13	153	mC	44.18	99.15	84.64	89.7	89.8	64.4	C2
12	150	oC	44.18	99.15	84.64	89.7	90.2	115.6	C222,C2221
11	149	mC	99.15	44.18	84.64	90.2	90.3	64.4	C2
10	8	tP	44.18	44.72	84.64	90.3	90.2	90.9	P4,P41,P42,P43,P422,P4212,P4122,P41212,P4222,P42212,P4322,P43212
9	6	oC	62.38	63.33	84.64	90.0	90.3	89.3	C222,C2221
8	6	mC	62.38	63.33	84.64	90.0	90.3	90.7	C2
7	5	oP	44.18	44.72	84.64	90.3	90.2	90.9	P222,P2221,P21212,P212121
6	4	mC	62.38	63.33	84.64	90.0	90.3	89.3	C2
5	4	mP	44.18	44.72	84.64	90.3	90.2	90.9	P2,P21
4	3	mP	44.72	44.18	84.64	90.2	90.3	90.9	P2,P21
3	2	mP	44.18	84.64	44.72	90.3	90.9	90.2	P2,P21
2	1	aP	44.18	44.72	84.64	89.7	90.2	89.1	P1
1	0	aP	44.18	44.72	84.64	90.3	90.2	90.9	P1

Table 4-3 Distortion index table for native *gris* ACP crystal data to a resolution of 2.5 Å. The solution with the lowest penalty and highest symmetry is shown in red.

No	PENALTY	LATT	a	b	c	alpha	beta	gamma	Possible spacegroups
15	151	mC	43.64	99.14	83.96	90.2	90.4	64.3	C2
14	151	mC	99.14	43.64	83.96	90.4	90.2	64.3	C2
13	149	oC	43.64	173.23	44.66	89.9	90.4	104.2	C222,C2221
12	149	mC	173.23	43.64	44.66	90.4	90.1	75.8	C2
11	149	mC	43.64	173.23	44.66	90.1	90.4	75.8	C2
10	10	tP	43.64	44.66	83.96	90.0	90.4	90.4	P4,P41,P42,P43,P422,P4212,P4122,P41212,P4222,P42212,P4322,P43212
9	9	oC	62.23	62.66	83.96	89.8	90.3	88.7	C222,C2221
8	9	mC	62.23	62.66	83.96	89.8	90.3	88.7	C2
7	9	mC	62.66	62.23	83.96	90.3	90.2	91.3	C2
6	3	mP	44.66	43.64	83.96	90.4	90.0	90.4	P2,P21
5	3	oP	43.64	44.66	83.96	90.0	90.4	90.4	P222,P2221,P21212,P212121
4	2	mP	43.64	83.96	44.66	90.0	90.4	90.4	P2,P21
3	1	mP	43.64	44.66	83.96	90.0	90.4	90.4	P2,P21
2	0	aP	43.64	44.66	83.96	90.0	89.6	89.6	P1
1	0	aP	43.64	44.66	83.96	90.0	90.4	90.4	P1

Table 4-4 Distortion index table for native *gris* ACP crystal data to a resolution of 1.8 Å. The solution with the lowest penalty and highest symmetry is shown in red.

Both data sets were therefore processed in the P1 spacegroup and the reflection files opened using HKLVIEW and pseudo-precession pictures were examined to look for mirror planes, systematic absences and symmetry related intensities. The pseudo-precession picture with superimposed mirror planes for the 2.5 Å data set is shown in Figure 4-4.

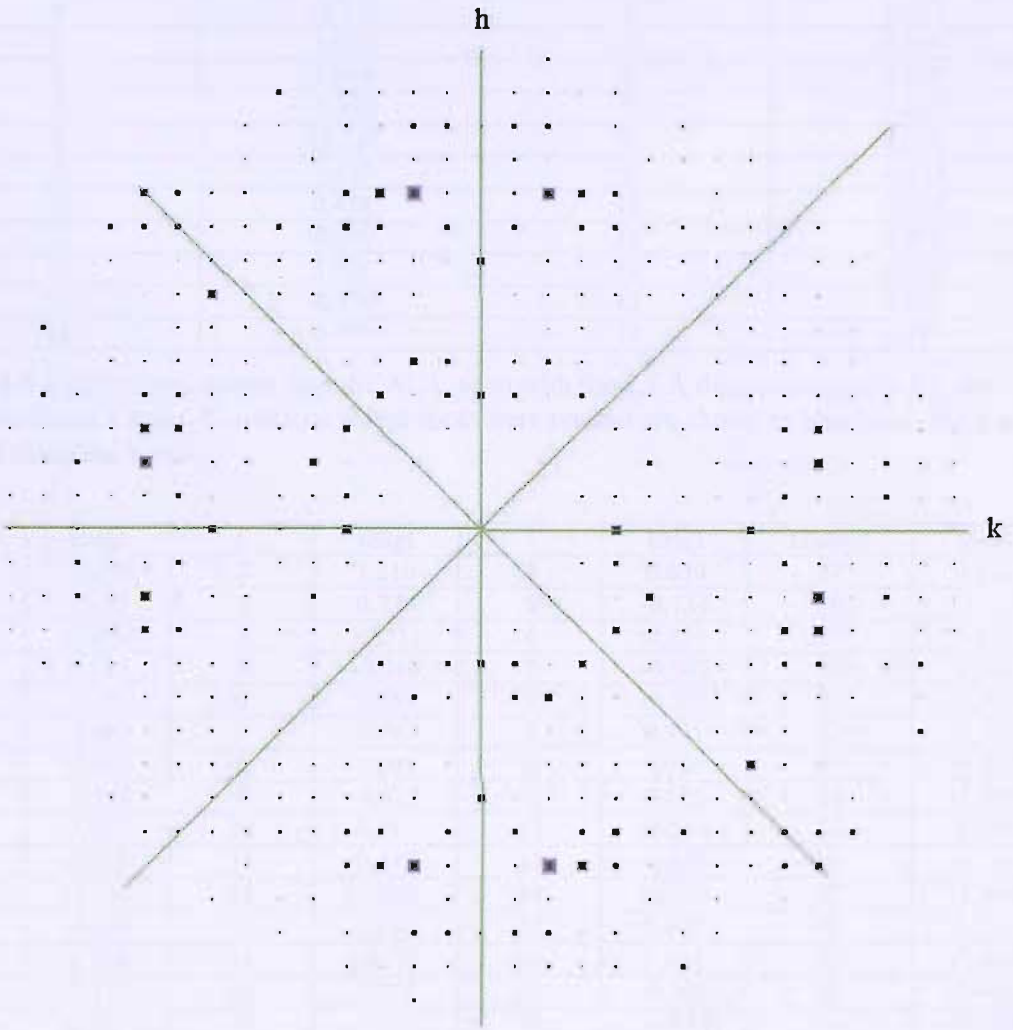


Figure 4-4 Pseudo-precession picture of 2.5 Å data set, processed in space group P1, viewed along the l=0 axis. Three mirror planes can be seen represented by the green lines.

The 2.5 Å data set shows three mirror planes, represented by green lines in the pseudo-precession picture indicating that the crystal was tetragonal. Systematic absences were seen in both data sets as shown in Table 4-5 and Table 4-6. These indicate that two screw axes were present in the 2.5 Å data set and a 3 screw axes was present in the 1.8 Å data set. Intensities of symmetry related coordinates were then compared to the space group criteria as found in the International Tables for X-ray crystallography, resulting in the conclusion that the 2.5 Å data set was in spacegroup $P4_12_12$ whilst the 1.8 Å data set

was in spacegroup $P2_12_12_1$. The data sets were then reprocessed using MOSFLM in their respective space groups.

h	I/sigI	k	I/sigI
1	2.530	1	1.732
2	1.023	2	1.429
3	-0.343	3	1.625
4	10.142	4	8.357
5	1.079	5	1.134
6	5.844	6	0.169
7	-0.180	7	0.015
8	10.318	8	8.376
9	1.417	9	-0.215
10	0.438	10	-2.008
11	-0.247	11	0.915
12	5.001	12	6.530
13	-0.174	13	0.901
14	0.797	14	0.471

Table 4-5 I/sigI values, output from SCALA, seen with the 2.5 Å data, presented in P1, set along the h and k axies. Conditions where spots were present are shown in blue along the h axis and red along the k axis.

h	I/sigI	k	I/sigI	l	I/sigI	l (cont)	I/sigI
2	-3.236	2	1.210	4	0.630	27	0.892
3	1.757	3	0.778	5	-0.134	28	12.706
4	12.252	4	12.311	6	4.221	29	1.971
5	2.415	5	-0.040	7	-0.542	30	0.308
6	11.972	6	2.351	8	16.348	31	-0.982
7	-0.960	7	-0.567	9	0.731	32	15.543
8	12.495	8	9.988	10	7.046	33	-0.802
9	-0.956	9	-0.522	11	0.625	34	-1.006
10	2.470	10	0.341	12	19.684	35	1.756
11	0.225	11	0.413	13	-0.935	36	11.386
12	14.182	12	12.442	14	13.524	37	-1.360
13	-1.643	13	1.112	15	0.645	38	5.363
14	6.756	14	10.179	16	2.054	39	-0.004
15	-2.113	15	1.137	17	-0.009	40	0.449
16	0.897	16	5.330	18	14.854	41	-0.758
17	0.751	17	-0.098	19	-1.807	42	-0.357
18	2.539	18	1.198	20	6.213	43	-0.257
19	1.037	19	0.732	21	0.839	44	2.847
20	3.323	20	-0.330	22	2.296	45	-1.206
21	-0.812	21	-0.295	23	-0.734	46	0.297
22	10.269	22	-0.063	24	3.098		
23	1.963	23	1.112	25	1.184		
24	5.903	24	6.575	26	3.298		

Table 4-6 I/sigI values, output from SCALA, seen with the 1.8 Å data set, processed in P1, along the h, k and l axes. Conditions where spots were present are shown in blue along the h axis, red along the k axis and green along the l axis.

Each reflection file output from MOSFLM was then sorted using SORTMTZ and then scaled using SCALA. The scaled data was then truncated using TRUNCATE to

generate structure factor amplitudes. Table 4-7 shows the statistics derived by SCALA and TRUNCATE.

Parameter	2.5 Å Data set	1.8 Å Data set
Spacegroup	$P4_12_12$	$P2_12_12_1$
I/sigI Total	3.4	3.9
I/sigI Max resolution	2.8	0.8
Multiplicity Total	13.5	7.9
Multiplicity Max resolution	14.6	7.8
Observed Reflections	57110	238577
Unique Reflections	619	15921
Completeness Total (%)	99.7	100
Completeness Max resolution (%)	99.7	100
Rmeas Total	0.183	0.183
Rmeas Max resolution	0.239	1.042
Rsym Total	0.177	0.171
Rsym Max resolution	0.231	0.973
Solvent content (%)	34	38
Molecules per Asymmetric unit	1	2
a	44.5981	44.0543
b	44.5981	44.0543
c	84.1606	84.1246
α	90	90
β	90	90
γ	90	90

Table 4-7 Scaling statistics output from SCALA and TRUNCATE for both the 2.5Å and the 1.8 Å *apo gris* ACP data sets.

The output from TRUNCATE gives the solvent content of the unit cell and as such can be used to determine the number of protein molecules in the asymmetric unit. In the 2.5 Å data set there was a solvent content of 34 % indicating one *gris* ACP molecule per asymmetric unit. However in the 1.8 Å data set, if the calculation was carried out assuming one molecule in the asymmetric unit, the solvent content was 66 %, which is generally thought of as too high for a protein crystal, and in this case the crystals were very rigid and resistant to damage, in contrast to higher water content crystals. If the calculation was carried out with 2 molecules in the asymmetric unit then the solvent content was 38 % so the 1.8Å data was truncated assuming two molecules in the asymmetric unit.

The programs MOLREP, BEAST, AmoRe and EPMR were all employed for molecular replacement of both the data sets. A variety of different ACP structures were used as models in an attempt to obtain a molecular replacement solution. These can be seen in Table 4-8.

ACP Name	**FAS/PKS**	**Species**	**PDB code**	**Swiss-Prot code**	**Solution Method**	**Sequence homology**
otc ACP	PKS	*Streptomyces rimosus*	1NQ4	ACPX STRRM	NMR	53%
act ACP	PKS	*Streptomyces coelicolor*	1AF8	ACPX STRCO	NMR	51%
fren ACP	PKS	*Streptomyces roseofulvus*	1ORS	068916	NMR	~25%
fas ACP	FAS	*E. coli*	1ACP	ACP ECOLI1	X-RAY	~25%
fas ACP	FAS	*Mycobacterium tuberculosis*	1KLP	P71603	NMR	~25%
fas ACP	FAS	*Bacillus subtilis*	1HY8	ACP BACSU	NMR	~25%
fas ACP	FAS	*Streptomyces coelicolor*		P72393	NMR	~25%

Table 4-8 Previously solved structures of ACP's with greater than 25 % sequence homology to the *gris* ACP sequence, their PDB code and their Swiss-Prot codes, which were used during molecular replacement.

The NMR structural models (random selection taken from the ensemble for each ACP entry in the PDB data base) were tested as individual structures as well as an overall averaged structure. Each model was tested using its full coordinate set and its secondary structure, excluding flexible regions. Backbone mimicking poly-ala and poly-ser models were also tested. However none of the above structures resulted in a correct molecular replacement solution. It was therefore decided to carry out heavy atom soaks in order to use derivatives to obtain experimental phases.

4.1.3 NMR studies of divalent cation binding to *gris* ACP

Analysis of the *gris* ACP sequence (Figure 4-5) shows that none of the standard residues utilised in heavy metal derivative formation was present.

SKQEFTLED	LKRILLEGAG	ADEGVDLDGD	ILDTDFEELG	YESLALLETG
10	20	30	40	50
GRIEREYGIT	LDDDVLADSR	TPRSLIAAIN	AAFQGLVEA	
60	70	80		

Figure 4-5 *gris* ACP amino acid sequence from *Streptomyces griseus*.

Previous studies by Prestegard *et al.* (*9*) in the early 1990's indicated that ACPs bound divalent cations (Ca^{2+}, Mg^{2+}, Mn^{2+}) which could prove useful for isomorphous phasing, a crystallographic technique used to solve the phase problem. A ^{15}N labelled *apo gris* ACP sample was used to carry out NMR binding studies with the aim to show that *gris* ACP also displayed this ability. A ^{1}H-^{15}N HSQC spectrum was recorded of 0.6 mM *gris* ACP in 10 mM sodium phosphate buffer, pH 7.5, 25 $^{\circ}C$. Initially $NiCl_2$ was titrated into the sample, and after each addition another ^{1}H-^{15}N HSQC spectrum was recorded. If these spectra are overlaid changes in resonance chemical shifts indicate interactions between the nickel ions and the ACP. During the nickel titration the protein was soluble with increasing ion concentrations until the addition of 1mM $NiCl_2$, at which point further titration was abandoned due to protein precipitation. The ^{1}H-^{15}N HSQC spectrum recorded at 0mM (black) and 500 μM $NiCl_2$ (red) (Figure 4-6), shows movement/loss of chemical shifts of residues 61-63 and hence implying interactions between the Ni^{2+} ions and the ACP in the region of helix III (when compared to other ACP structures) at this concentration.

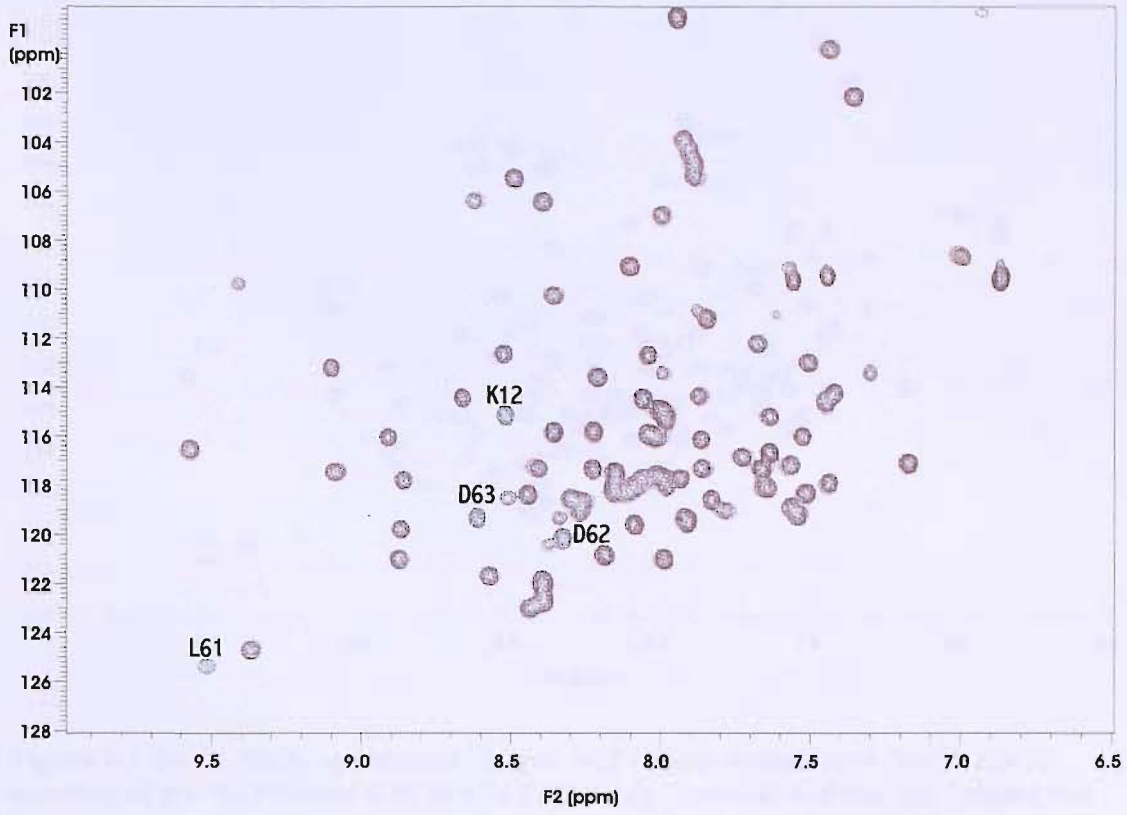


Figure 4-6 ^{1}H-^{15}N HSQC spectrum of ^{15}N *gris* ACP (black) overlaid with ^{1}H-^{15}N HSQC spectrum of *gris* ACP titrated with 500 μM $NiCl_2$ (red). The chemical shifts for residues K12, L61, D62 and D63 all disappear during the titration, indicating that these residues are interacting with the $NiCl_2$.

A fresh ACP sample was prepared under the same conditions and the titration was repeated using $CaCl_2$. No protein precipitation was seen with increasing calcium concentration. The overlay between 0 mM (black) and 80 mM $CaCl_2$ (red) in Figure 4-7 shows movements in the backbone amide chemical shifts. When compared to sequence alignments with other known ACP structures these chemical shift changes indicate that interactions/movements were occurring in the region of the c-terminus of helix I, loop 1 and the region between the C-terminal end of helix 2 and the N-terminal end of helix 4 (Figure 4-8).

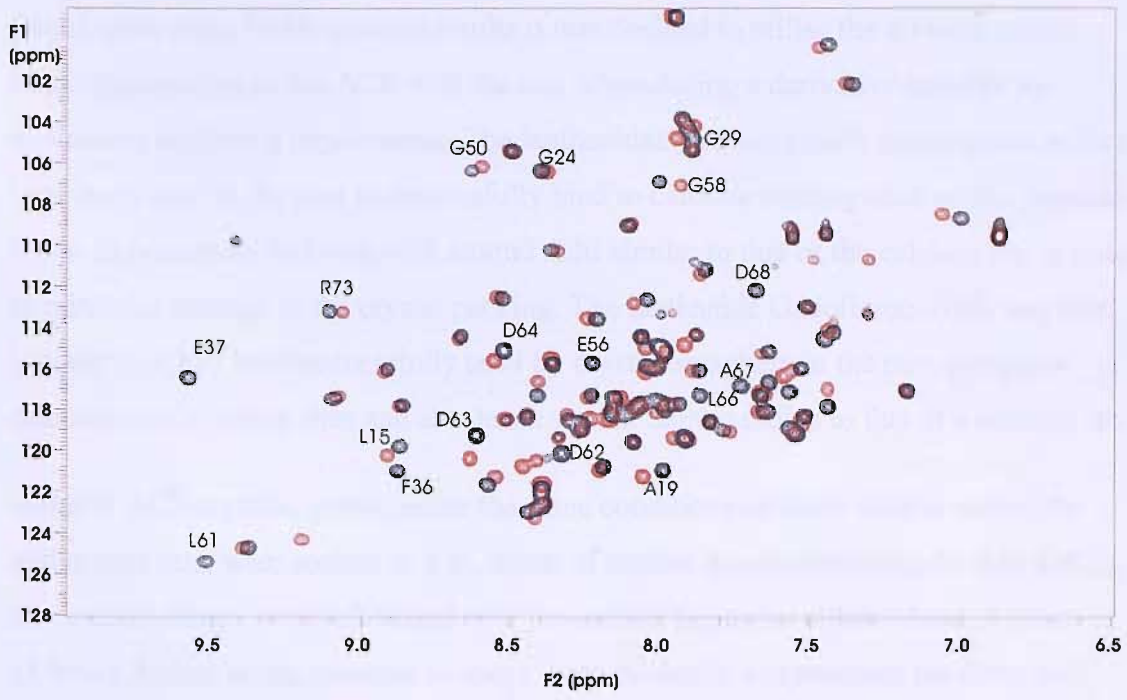


Figure 4-7 ^{1}H-^{15}N HSQC spectrum of ^{15}N *gris* ACP (black) overlaid with ^{1}H-^{15}N HSQC spectrum of *gris* ACP titrated with 80 mM $CaCl_2$ (red). Chemical shift changes indicate that interactions/movements were occurring in the region of loop 1 and the region between the C-terminal end of helix 2 and the N-terminal end of helix 4.

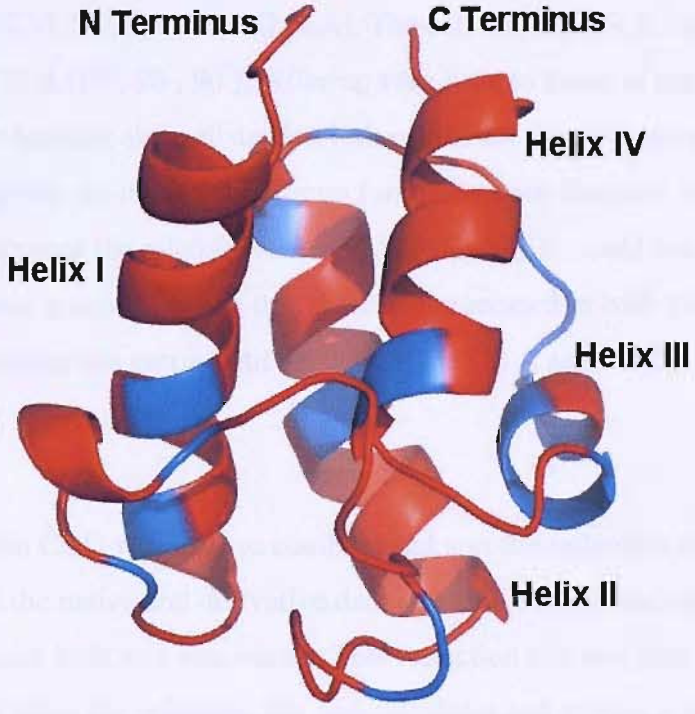


Figure 4-8 Illustration of the residues whose chemical shifts changed when *gris* ACP was tritrated with $CaCl_2$. These residues are shown relative to the FAS ACP from *Streptomyces coelicolor* as the *gris* ACP structure has yet to be solved.

Based upon these NMR titration results it was decided to utilise the divalent cation binding properties of the ACP with the aim of producing a derivative suitable for anomalous scattering experiments. The lanthanides were originally investigated as they have been used in the past to successfully bind to calcium binding sites within proteins. It was important to find ions with atomic radii similar to that of the calcium ion in order to minimise damage to the crystal packing. The lanthanide Gadolinium (Gd) was used initially as it had been successfully used by crystallographers in the past to replace calcium ions in active sites and also has a similar atomic radius to that of a calcium ion.

apo gris ACP crystals, grown under the same conditions as those used to record the native data sets, were soaked in 5 µL drops of mother liquor containing 10 mM $GdCl_2$. The crystals/drops were left sealed over the mother liquor for either 1 hour, 3 hours or 18 hours, before being mounted in loops, flash cooled in a cryostream (as described above) and exposed to the X-ray beam in house. Diffraction data was only obtained from the crystal soaked for 18 hours, with diffraction seen to 3.4 Å, and a data set comprising 150 images with a 1° oscillation was collected.

This data set, as with the native data set, was processed using the CCP4 suite of programs (MOSFLM, SORTMTZ, SCALA). The cell dimensions for the data set were 44.84, 44.84, 82.70 Å (90°, 90°, 90°), differing very little to those of the two native data sets. Again, when looking at the distortion index table for the derivative data it was difficult to distinguish the exact space group for the data set. Because of the possibility of anomalous scattering the relative intensities in HKLVIEW could not be used to determine the space group. The data was therefore processed in both $P4_12_12$ and $P2_12_12_1$ and further processing was carried out using both the 1.8 Å and the 2.5 Å native data sets.

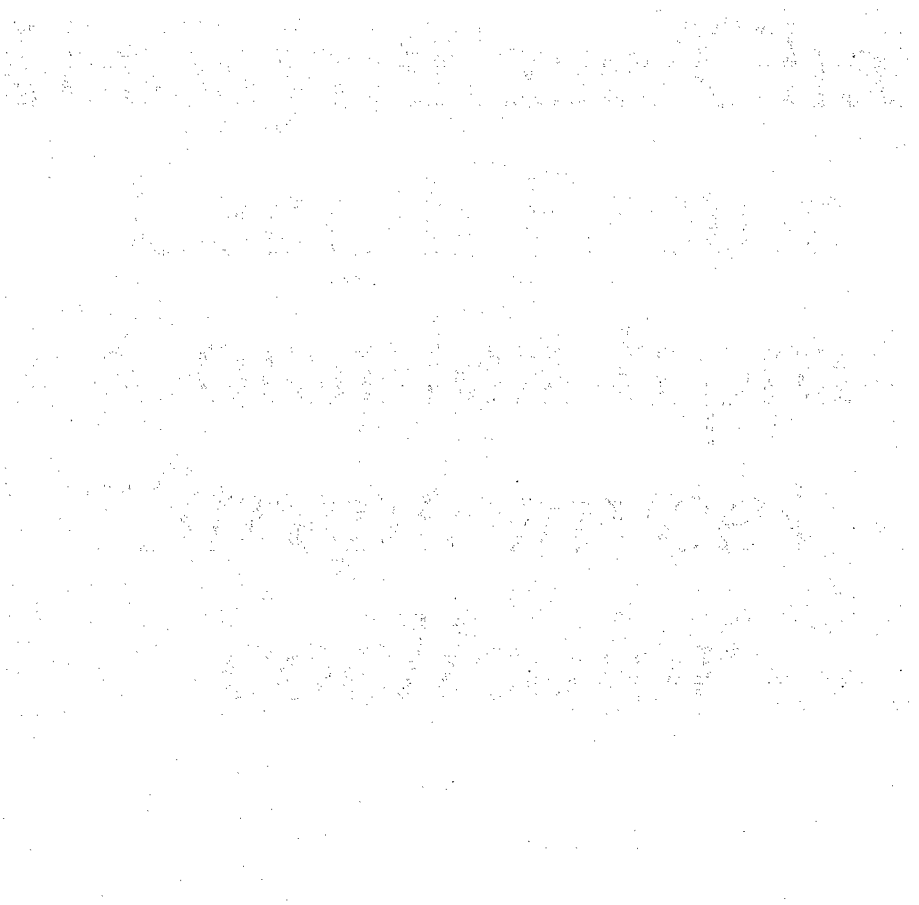
The CCP4 program CAD was used to combine and sort the reflection data output from SCALA, for both the native and derivative data sets, and a combined reflection file containing data from both sets was output. This reflection file was then input into SCALEIT, which takes the reflection file and calculates and applies a native scaling function to ensure that the two data sets are comparable. The scaled reflection file is then ready to attempt a search for the heavy metal binding site. To do this FFT is used, which utilises a fast Fourier transform to calculate difference Pattersons between the two data sets.

The CCP4 program NPO is then used to visualise the output difference density maps. In order to determine whether a heavy atom is present it is necessary to look at specific positions within the Patterson maps for difference density. To do this it is necessary to look at the Harker sections for the individual space groups. If a derivative is present then there should be difference density observable at these specific positions. However for the Gd derivative data set no evidence of heavy metal binding was observed.

No solution to this problem has currently been found. Future plans include trying several different approaches to attempt to obtain a phasing solution. Firstly crystal screens will be set-up under the same conditions as the native crystals but with the heavy metal being added to the drop, in the hope that the heavy metal will bind to the protein prior to crystal formation. Secondly soaks containing alternative heavy metals will be attempted. Soaks using samarium (Sm) at various concentrations have been tried (20 mM – 1 M) without success, as the protein crystals appear to melt immediately upon addition of the heavy metal to the drop. However, as this shows that the Sm is having an effect on the crystal, and hence the protein, it may be possible to bind the heavy metal successfully prior to crystal growth, although the crystals may not be isomorphous with those of the underivatised protein. The possibility of using Ca^{2+} ions bound to the ACP is another option. Intrinsic calcium ions have been used in SAD experiments (single wavelength anomalous diffraction) and it may be possible to utilise the divalent cation binding properties of the *gris* ACP in this way, in order to solve the phase problem.

4.2 References

1. Roujeinikova, A., Baldock, C., Simon, W. J., Gilroy, J., Baker, P. J., Stuitje, A. R., Rice, D. W., Slabas, A. R., and Rafferty, J. B. (2002) *Structure (Camb) 10*, 825-35.
2. Zhou, P., Florova, G., and Reynolds, K. A. (1999) *Chem. Biol. 6*, 577-84.
3. Matharu, A. L., Cox, R. J., Crosby, J., Byrom, K. J., and Simpson, T. J. (1998) *Chem. Biol. 5*, 699-711.
4. Crosby, J., Byrom, K. J., Hitchman, T. S., Cox, R. J., Crump, M. P., Findlow, I. S., Bibb, M. J., and Simpson, T. J. (1998) *FEBS Lett. 433*, 132-8.
5. Giacovazzo, C. (1992) *Fundamentals of Crystallography*, Oxford University Press, NEW YORK.
6. Chayen, N. E. (1997) *J. Appl. Cryst. 30*, 198-202.
7. Mikol, V., Rodeau, J-L. and Giege, R. (1990) *Analytical Biochemistry 186*, 332-339.
8. Jovine, L. (2000) *J. Appl. Cryst. 33*, 988-989.
9. Frederick, A. F., Kay, L. E., and Prestegard, J. H. (1988) *FEBS Lett. 238*, 43-8.



Chapter 5

Ketosynthase/Chain Length Factor Complex from *Streptomyces coelicolor*

5. The Polyketide Synthase Ketosynthase/Chain Length Factor complex from *Streptomyces coelicolor*

Although there have been ketosynthase (KS) crystal structures published from FAS systems (*1-8*) there had not been any polyketide KS structures published when this work was commenced. In 2004 Keating-Clay *et al.* (*9*) published the crystallographic structure of the *actinorhodin* KS/CLF complex from *Streptomyces coelicolor*. The work described below was undertaken prior to the publication of that paper.

The KS/CLF complex from the actinorhodin (act) polyketide system in *Streptomyces coelicolor* has been looked at during this work, with the aim of solving the 3D structure using x-ray crystallography as the protein complex is too large to use current NMR techniques. Minimal PKS assays *in vitro* indicate that the FAS ACP from *Streptomyces coelicolor* does not function as part of the assay, where as the act ACP does. Comparisons of the two ACPs have shown a similar overall fold, as discussed in Chapter 3, and it is hoped that by solving the KS/CLF complex structure, both with and/or without act ACP bound, will enable determination of the ACP binding site on the KS. Knowledge of a polyketide KS will also allow for comparison with those from FAS systems, hopefully revealing differences between the two which may be exploited in production of new polyketide products, or as selective targets of antimicrobial therapies.

5.1 Practical Methodology and Results

5 1.1 KS/CLF Crystallisation and Characterisation

KS/CLF from *Streptomyces coelicolor* was provided by Dr T. Nicholson from The University of Bristol. Upon elution from a nickel affinity column the protein, although pure, was shown to precipitate. It was discovered that, in the presence of free nickel ions eluted from the affinity column, protein aggregation occurred through co-ordination of nickel to the protein. This phenomenon may be overcome by the addition of a 1 mM solution of the chelating agent EDTA to the collection tubes to prevent the nickel from binding to the HIS tags and thus increasing protein recovery. Attempts to resolubilise the aggregated KS/CLF via denaturation with 6 M guanidine hydrochloride were unsuccessful as aggregation was seen once the guanidine was dialysed out. Denaturation in 6 M urea followed by a step dialysis into 4 M, 2 M, 1 M and 0 M urea showed no precipitate formation. However on subsequent running of an anionic

exchange column and SDS-PAGE no protein was recovered, resulting in the conclusion that the protein had been lost at some point during dialysis. Finally dialysis of KS/CLF in the presence of urea but in the absence of EDTA produced in a brown precipitate due to interactions between the urea and nickel ions. A second batch of KS/CLF supplied by The University of Bristol, in 50 mM sodium phosphate, 2 mM DTT, 1 mM EDTA and 10 % glycerol pH 7.3 at a concentration of ~4-6 mg / mL. This purity was checked by SDS-PAGE as can be seen in Figure 5-1. This protein solution was dialysed, using a slide-a-lyser, into the same buffer but containing Tris instead of phosphate to avoid false positives within subsequent crystal screens.

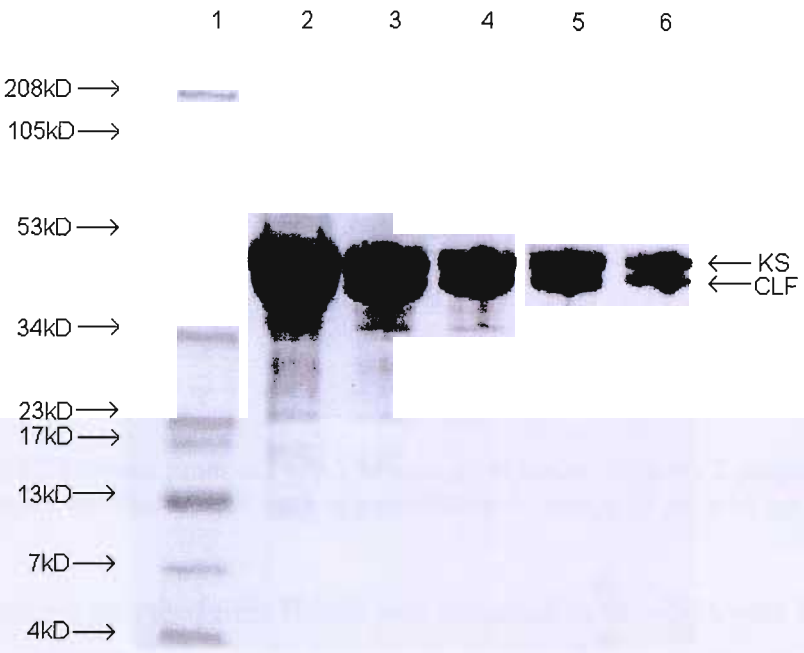


Figure 5-1 SDS-PAGE gel showing purity of KS/CLF. Lane 1: molecular weight markers, lane 2: 25 µl sample, lane 3: 20 µl sample, lane 4: 10 µl sample, lane 5: 5 µl sample, lane 6: 2 µl sample. (Sample: 50 µl KS/CLF supplied by Bristol University + 50 µl 1xgel loading buffer, boiled for 5 minutes).

Hanging drop crystal screens of KS/CLF were setup using Molecular Dimension crystal screen MD1-01 and MD1-02 using 2 µL protein solution : 2 µL mother liquor drops. Upon addition of protein to the drops for the screens it was noted that a dense cloudy precipitate was forming in a greater proportion of conditions than would usually be expected. Crystals were seen after 2 days in 0.2 M calcium chloride, 20 % v/v 2-propanol, 0.1 M sodium acetate pH 4.6, (Figure 5-2). When trying to refine the conditions it was discovered that, following storage at –20 °C, the protein would not dialyse into tris buffer without precipitating. When looking at the crystal screens it was noticed that no precipitate was present in wells containing an alcohol, which is not

entirely unexpected as the addition of alcohols, especially glycerol, is known to stabilise proteins in solution, but can therefore inhibit crystal formation. The protein solution was also successfully dialysed into the same tris buffer as before but containing 20 % glycerol in an attempt to stabilise the protein further. Grid screens and full factorial screens were then set up using this protein. New MD1-01 and MD1-02 screens were also set up using the protein in a higher glycerol content, in an attempt to keep the protein stable for a longer time, thus allowing more chance for nucleation to occur before the protein precipitated. The original crystal produced from the molecular dimensions screen, was wet mounted and placed into an x-ray beam in-house, and diffracted to give a characteristic pattern consistent with that observed from salt crystals.

Figure 5-2 KS/CLF crystal produced in 0.2 M calcium chloride, 20 % v/v 2-propanol, 0.1 M sodium acetate pH 4.6. The crystals were approximately 0.1 mm x 50 μm x 50 μm in size.

The protein sample provided from Bristol was estimated to be >95 % pure by SDS-PAGE, (Figure 5-1). Despite the high purity, a sample of the KS/CLF complex was purified further using gel filtration chromatography to check for homogeneity. A 200 μL sample of KS/CLF at ~5 mg/ml was applied to a Superose 12 gel filtration column which had been equilibrated in 100 mM Tris buffer pH 7.4 containing 100 mM NaCl, 1 mM EDTA, 2 mM DTT and 10 % glycerol and which shows what appears to be a higher molecular weight contaminant, which is possibly a higher order aggregate of the KS/CLF complex, shown in Figure 5-3. The retention time of the column was calibrated using a series of standards and the proposed molecular weight of the protein eluting in the main peak, the KS/CLF complex, is ~ 176 kDa, compared to that of the expected dimer at ~88 kDa. From this gel filtration data it therefore appears that under these conditions a proportion of the KS/CLF in the sample may exists as a dimer of heterodimers. This purified protein was concentrated to ~2.5 mg/mL using a vivaspin

5000 and entered into new Molecular Dimension screens and the full range of screens from Jena Bioscience.

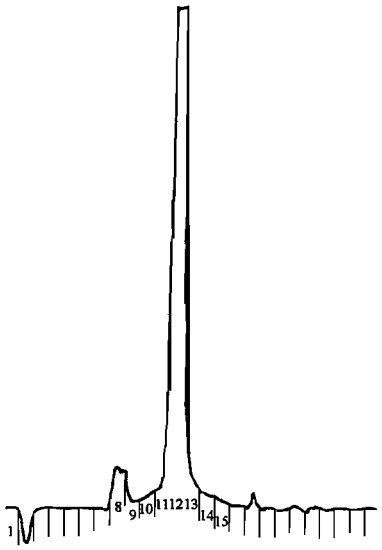
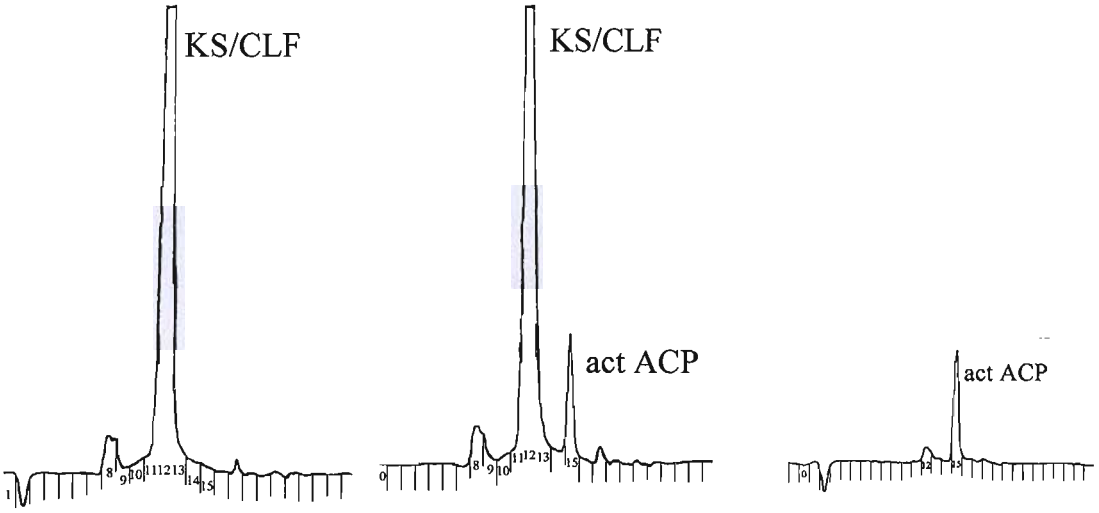


Figure 5-3 Gel filtration trace of KS/CLF showing a small peak representing an high order aggregate, possibly a dimer of KS/CLF, (eluting after 8 ml) and KS/CLF peak (eluting after 11 ml).

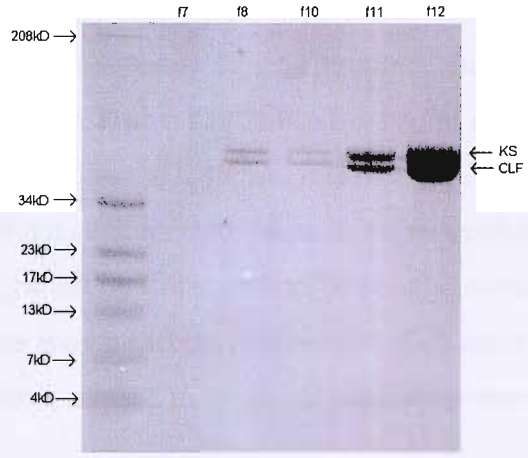
5 1.2 Gel filtration of KS/CLF with ACPs

In order to stabilise the KS/CLF complex it was decided to attempt to crystallise it with *apo act* ACP present as it is known that the three enzymes interact within the mPKS. 200 µL samples of ~4 mg/mL KS/CLF were prepared with and without *act* ACP (2 mg/ml), and in addition a 200 µL sample of 2 mg/mL *apo act* ACP was also prepared as a control. Samples were incubated overnight at +4 °C and then applied to a Superose 12 gel filtration column in 30 mM tris buffer, pH 7.3, 10 % glycerol, 1 mM DTT, 150 mM NaCl. Figure 5-4 shows the gel filtration traces from the three different samples, whilst Figure 5-5 shows the SDS-PAGE analysis of these results.

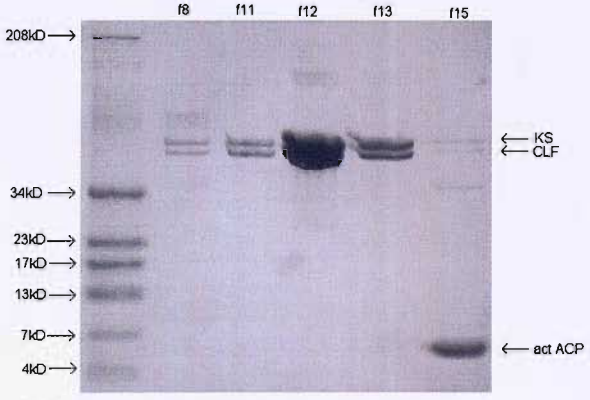


a) KS/CLF b) KS/CLF + *apo act* ACP c) *apo act* ACP

Figure 5-4 Superose 12 gel filtration column traces, run in 30 mM tris buffer, pH 7.3, 10 % glycerol, 1 mM DTT, 150 mM NaCl of a) KS/CLF, b) KS/CLF plus *apo act* ACP and c) *apo act* ACP. (Dimeric KS/CLF elutes after 8 minutes, monomeric KS/CLF elutes after 12 ml and *apo act* ACP elutes after 15 ml)



a) KS/CLF



b) KS/CLF + *apo act* ACP

Figure 5-5 SDS-PAGE gels showing fractions from gel filtration shown in Figure 5-4 of a) KS/CLF and b) KS/CLF plus *act* ACP. No ACP binding to the KC/CLF can be seen in the KS/CLF containing fractions (f8, f11, f12, f13) indicating that no binding between the two proteins occurs under the gel filtration conditions.

The elution profiles (Figure 5-4) of the KS/CLF remain consistent even with the addition of *act* ACP, with a peak retention volume of 11 mL for both KS/CLF and KS/CLF with *act* ACP present, indicating that there is no ACP-KS/CLF interaction under these conditions. The SDS-PAGE analysis shows no *act* ACP within the KS/CLF fraction (f11, f12, f13) and also indicating no binding.

^{1}H- ^{15}N HSQC titrations of KS/CLF into ^{15}N labeled *gris* ACP (Figure 5-6) showed slight movement in some of the amide peaks of the ACP, in the region of the c-terminus of helix I, the beginning of loop I and the beginning of helix II (compared to other solved ACP structures), consistent with binding/interactions between the proteins (Figure 5-7). This implies that this is the region of the ACP to which the KS/CLF binds and shows that cross-talk occurs between these two components of the two systems.

gris ACP was therefore added to KS/CLF under the same conditions as those used for the *act* ACP and applied to a Superose 12 gel filtration column. Figure 5-8 shows the elution profile for the *gris* ACP plus KS/CLF. SDS-PAGE analysis shown (Figure 5-9) clearly shows no ACP bound to the KS/CLF. It is therefore possible that the gel filtration conditions are not conducive to interactions between the KS/CLF complex and the ACP. Subsequently, work by Chris Arthur (personal communication) at The University of Bristol has shown that the minimal *act* PKS assay is not functional in the presence of 150 mM NaCl, as used in the gel filtration experiments.

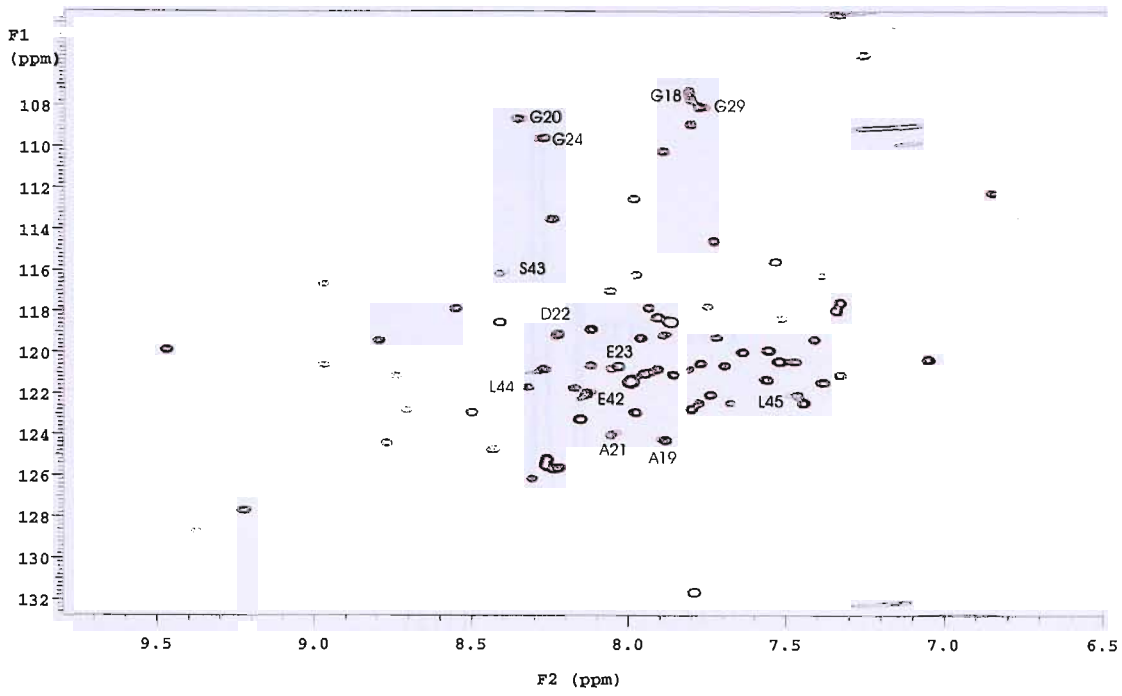


Figure 5-6 ^{1}H-^{15}N HSQC titration of KS/CLF into ^{15}N labelled *gris* ACP. The grey spectra shows the ^{1}H-^{15}N HSQC of the *gris* ACP on its own whilst the red spectra shows the ^{1}H-^{15}N HSQC of *gris* ACP titrated with KS/CLF.

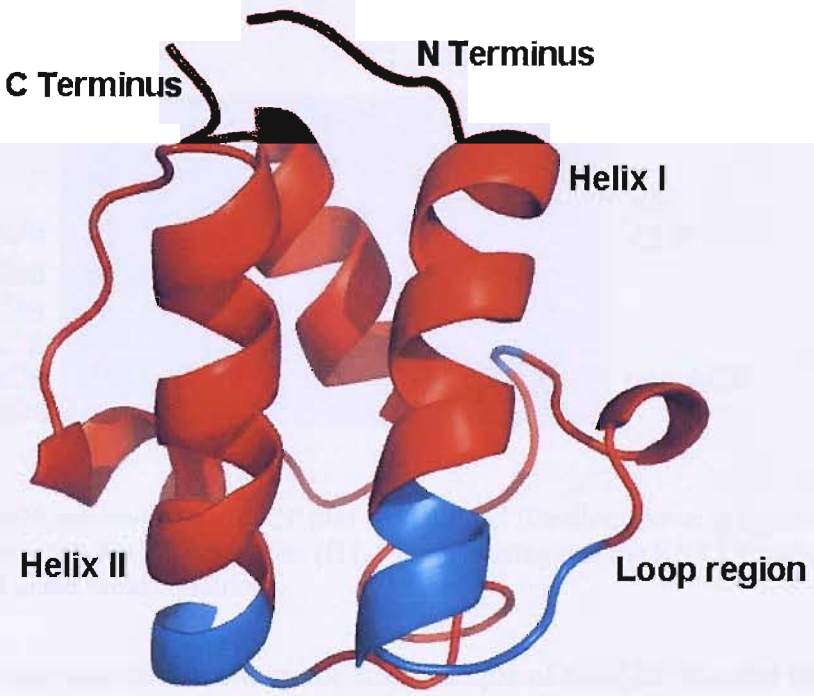


Figure 5-7 Illustration of the residues whose chemical shifts changed when KS/CLF was titrated into *gris* ACP. These residues are shown in blue and shown on the FAS ACP structure as the *gris* ACP structure has yet to be solved. The interactions occur at the c-terminal of helix I and the n-terminus of helix II, in the region of the serine to which the phosphopantetheine arm is attached.

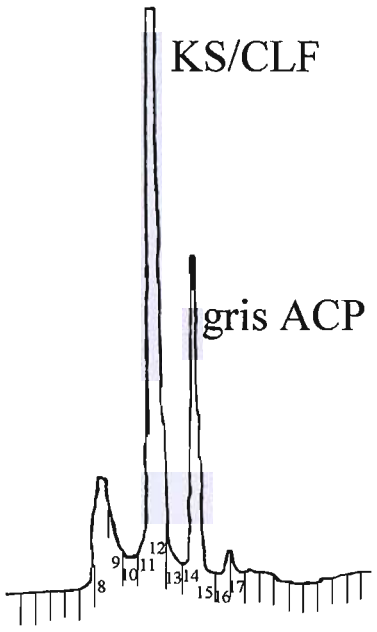


Figure 5-8 Elution profile from Superose 12 gel filtration column for *apo gris* ACP plus KS/CLF. The retention times for the KS/CLF components are the same (12 ml) as those seen for KS/CLF with no added ACP, indicating that no binding between the two proteins takes place.

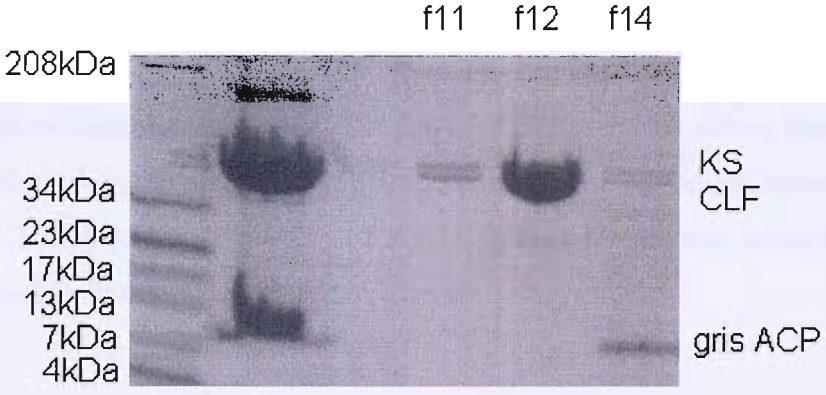


Figure 5-9 SDS-PAGE analysis of *gris* ACP plus KS/CLF gel filtration shown in Figure 5-8. No ACP can be seen in the KS/CLF fractions (f11, f12) indicating that the KS/CLF and *gris* ACP do not interact under these conditions.

A minimal PKS assay was carried out on the stock sample of KS/CLF in order to determine the activity of the sample. The extracted sample was analysed by HPLC using a C18 column as described in the methods section. The HPLC trace (Figure 5-10) shows no SEK4 production (~20-22 minutes) for two repeats the KS/CLF sample (pink and blue) compared to a standard sample showing SEK4 productions and hence activity (grey). This indicates that the KS/CLF sample used for gel filtration and crystallization studies was more than likely inactive which could explain the lack of binding on the gel filtration column and the problems with the crystallization.

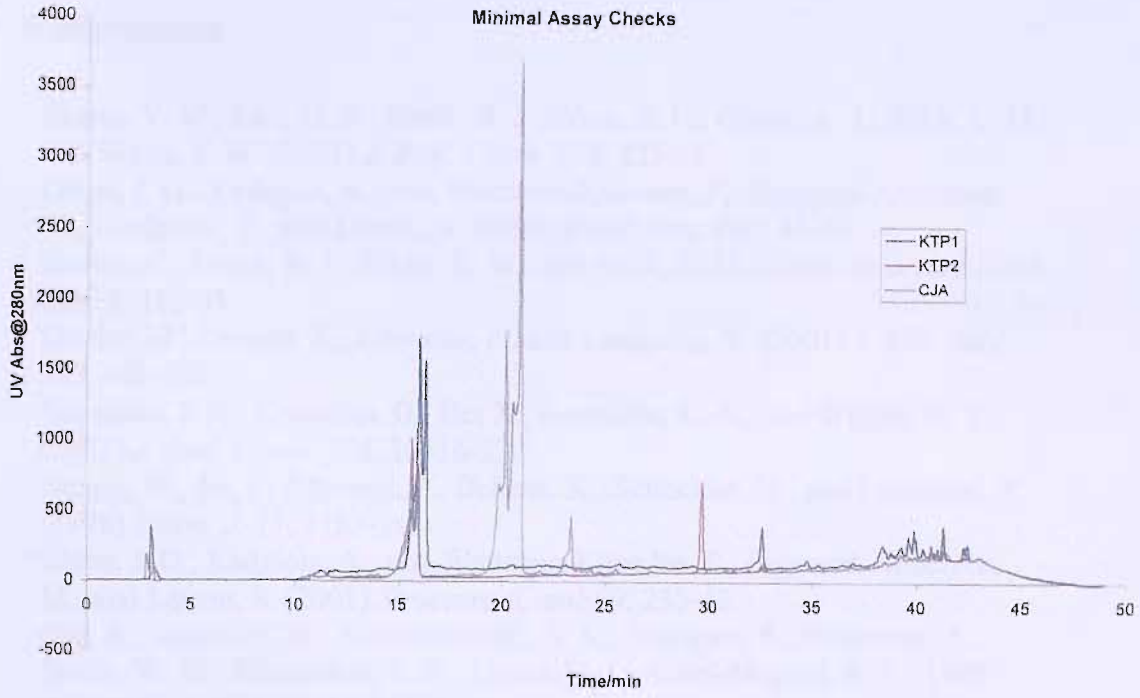


Figure 5-10 HPLC trace of minimal PKS assay showing no SEK4 production (blue and pink) for KS/CLF and SEK4 production (grey) for a standard sample at ~20 minutes.

The crystal structure of KS/CLF from *Streptomyces coelicolor* (*9*) was published before any further work could be carried out. The KS/CLF structure was solved from crystals grown over three days using the hanging drop method in 3.8 M sodium formate, at room temperature. The structure revealed that KS/CLF forms a homodimer under these crystallisation conditions, as discussed in Chapter 1.

5.2 References

1. Zhang, Y. M., Rao, M. S., Heath, R. J., Price, A. C., Olson, A. J., Rock, C. O., and White, S. W. (2001) *J. Biol. Chem. 276*, 8231-8.
2. Olsen, J. G., Kadziola, A., von Wettstein-Knowles, P., Siggaard-Andersen, M., Lindquist, Y., and Larsen, S. (1999) *FEBS Lett. 460*, 46-52.
3. Davies, C., Heath, R. J., White, S. W., and Rock, C. O. (2000) *Structure Fold. Des. 8*, 185-95.
4. Moche, M., Dehesh, K., Edwards, P., and Lindqvist, Y. (2001) *J. Mol. Biol. 305*, 491-503.
5. Scarsdale, J. N., Kazanina, G., He, X., Reynolds, K. A., and Wright, H. T. (2001) *J. Biol. Chem. 276*, 20516-22.
6. Huang, W., Jia, J., Edwards, P., Dehesh, K., Schneider, G., and Lindqvist, Y. (1998) *Embo. J. 17*, 1183-91.
7. Olsen, J. G., Kadziola, A., von Wettstein-Knowles, P., Siggaard-Andersen, M., and Larsen, S. (2001) *Structure (Camb) 9*, 233-43.
8. Qiu, X., Janson, C. A., Konstantinidis, A. K., Nwagwu, S., Silverman, C., Smith, W. W., Khandekar, S. K., Lonsdale, J., Abdel-Meguid, S. S. (1999) *J.Biol.Chem. 274*, 36465-36471.
9. Keatinge-Clay, A. T., Maltby, D. A., Medzihradsky, K. F., Khosla, C. and Stroud, R. M. (2004) *Nature Structural & Molecular Biology 11*, 888-893.

Chapter 6

Biological Implications of this Work

6. Biological Implications of this Work.

The solution structure of FAS ACP from *Streptomyces coelicolor* has, for the first time, allowed comparisons between fatty acid and polyketide ACPs from the same species as prior to this work all solved fatty acid and polyketides ACP structures originated from different species. This will allow for a range of further experiments to be undertaken to try to elucidate the crucial differences between the two ACP types and will hopefully reveal why the ACPs are not interchangeable between the fatty acid and polyketide systems, although they are often interchangeable within the same system from different species.

The determination of the FAS ACP structure from *Streptomyces coelicolor* has provided a further atomic resolution structure of an ACP for comparison and interaction studies. The FAS ACP structure had the same overall 4 α-helical structure as seen in other ACPs. Closer examinations of structure overlays of the FAS ACP from *Streptomyces* coelicolor with *act* ACP from *Streptomyces* coelicolor, FAS ACP from *Mycobacterium* tuberculosis and the PKS *otc* ACP from *Streptomyces rimeous* revealed subtle differences between individual structures and differences between the FAS and the PKS ACPs. These included differences in the orientation and length of the helices and the flexibility in the loop regions (especially loop 1).The differences observed in the orientation of Helix I between the FAS and PKS ACPs may be essential for the selectivity in recognition between the enzymes involved in fatty acid and polyketides synthesis and for the prevention of cross-talk between the two systems. There are also differences in the conservation of residues in the sequences between ACPs from the two systems. Of the residues highly conserves between the PKS ACPs (arg12, gly20, phe35, gly39, tyr40, thr70, pro71, arg72 and asn79, relative to the *act* ACP sequence) only phe35 (phe33 in the FAS ACP sequence) is conserved. This phenylalanine gives rise to a long range NOE in the FAS, *act* and *otc* structures and may be important in maintaining the three dimensional protein shape, whilts the non-conserved residues may be part of the reason for the difference in selectivity between the two systems.

Titration of FAS and *act* ACPs with expected interaction partners from *Streptomyces coelicolor* using NMR can now reveal the structural mechanisms of interaction and the location of essential residues. By using the proteins found early in either the FAS, the PKS or both, (e.g. Acps, MCAT, KS, KS/CLF, KR) it should be possible to discover why the two ACPs will both interact with the MCAT from the FAS whilst the ACPs are not interchangeable when interacting with other proteins (e.g. the minimal PKS).

The ability to identify the mechanisms of interaction between the enzymes, and the way in which the synthesis of either the fatty acid or the polyketide is carried out, may allow for the discovery of ways to alter the end product of the synthesis. In future this may give rise to the ability to modify the synthase to produce a range of different biologically active molecules. These may provide new anti-microbial agents to counteract the recent rise in antibiotic resistant bacteria, or other useful drugs to supplement those already synthesised by members of the *Streptomyces sp.*

The addition of the FAS ACP structure to those of the other ACPs already solved has also provided further scope for comparisons between different types of FAS and PKS. The ability to identify differences between the type I (mammalian) and type II (bacterial) fatty acid synthases would give an ideal opportunity to locate a target which could be used to selectively inhibit the type II FAS (essential for cell survival) without effecting a type I system. This would allow elimination of bacterial infections in mammals without affecting the host.

As with all new scientific results the long term wider effect can only be guessed at but it hoped that the ability to finally compare the FAS and PKS from the same species will help to elucidate the differences and similarities in the mechanism of fatty acid and polyketides biosynthesis.

Chapter 7

Appendix

7 Appendix

Group	Atom	Nuc	Shift
A2	CA	13C	52.026
A2	CB	13C	19.370
A2	CO	1H	173.219
A2	HA	1H	4.040
A2	HB#	1H	1.523
A2	HN	1H	7.716
A3	CA	13C	51.596
A3	CB	13C	20.661
A3	CO	1H	178.155
A3	HA	1H	4.844
A3	HB#	1H	1.569
A3	HN	1H	8.851
A3	N	15N	124.993
T4	CA	13C	60.368
T4	CB	13C	71.565
T4	CG2	13C	22.124
T4	CO	1H	175.383
T4	HA	1H	4.438
T4	HB	1H	4.744
T4	HG2#	1H	1.376
T4	HN	1H	8.586
T4	N	15N	111.927
Q5	CA	13C	60.820
Q5	CB	13C	28.786
Q5	CG	13C	34.253
Q5	CO	1H	177.053
Q5	HA	1H	3.703
Q5	HB1	1H	2.061
Q5	HB2	1H	2.208
Q5	HE21	1H	6.844
Q5	HE22	1H	7.519
Q5	HG1	1H	2.148
Q5	HG2	1H	2.244
Q5	HN	1H	9.232
Q5	N	15N	120.129
Q5	NE2	15N	111.596
E6	CA	13C	60.544
E6	CB	13C	29.037
E6	CG	13C	37.198
E6	CO	1H	179.730
E6	HA	1H	3.954
E6	HB1	1H	1.998
E6	HB2	1H	2.103
E6	HG1	1H	2.344
E6	HG2	1H	2.482
E6	HN	1H	8.847
E6	N	15N	116.616
E7	CA	13C	58.900
E7	CB	13C	30.443
E7	CG	13C	36.008
E7	CO	1H	180.178
E7	HA	1H	4.153
E7	HB1	1H	2.141
E7	HB2	1H	2.508
E7	HG1	1H	2.304
E7	HG2	1H	2.478
E7	HN	1H	7.695
E7	N	15N	120.597
I8	CA	13C	64.298
I8	CB	13C	37.601
I8	CD1	13C	19.199
I8	CG1	13C	29.020
I8	CG2	13C	14.346
I8	CO	1H	177.784
I8	HA	1H	3.871
I8	HB	1H	1.764
I8	HD1#	1H	0.589
I8	HG11	1H	0.088
I8	HG12	1H	1.520
I8	HG2#	1H	0.511
I8	HN	1H	8.402
I8	N	15N	123.063
V9	CA	13C	67.703
V9	CB	13C	31.557
V9	CG1	13C	21.085
V9	CG2	13C	24.213
V9	CO	1H	177.472
V9	HA	1H	3.381
V9	HB	1H	2.026
V9	HG1#	1H	0.906
V9	HG2#	1H	1.054
V9	HN	1H	8.442
V9	N	15N	120.429
A10	CA	13C	55.275
A10	CB	13C	18.206
A10	CO	1H	181.185
A10	HA	1H	4.235
A10	HB#	1H	1.532
A10	HN	1H	7.746
A10	N	15N	119.543
G11	CA	13C	47.192
G11	CO	1H	176.009
G11	HA1	1H	3.870
G11	HA2	1H	4.085
G11	HN	1H	8.387
G11	N	15N	106.040
L12	CA	13C	58.210
L12	CB	13C	42.026
L12	CD1	13C	22.923
L12	CD2	13C	27.423
L12	CG	13C	27.123
L12	CO	1H	178.694
L12	HA	1H	4.260
L12	HB1	1H	1.246
L12	HB2	1H	2.129
L12	HD1#	1H	0.879
L12	HD2#	1H	0.854
L12	HG	1H	1.985

L12	HN	1H	9.265
L12	N	15N	122.527
A13	CA	13C	55.991
A13	CB	13C	18.717
A13	CO	1H	179.344
A13	HA	1H	4.062
A13	HB#	1H	1.701
A13	HN	1H	8.977
A13	N	15N	121.410
E14	CA	13C	59.892
E14	CB	13C	29.577
E14	CG	13C	36.337
E14	CO	1H	179.311
E14	HA	1H	3.956
E14	HB1	1H	2.225
E14	HB2	1H	2.312
E14	HG1	1H	2.341
E14	HG2	1H	2.634
E14	HN	1H	7.830
E14	N	15N	117.865
I15	CA	13C	65.587
I15	CB	13C	38.941
I15	CD1	13C	13.230
I15	CG1	13C	29.298
I15	CG2	13C	19.234
I15	CO	1H	177.912
I15	HA	1H	3.802
I15	HB	1H	2.131
I15	HD1#	1H	0.767
I15	HG11	1H	1.062
I15	HG12	1H	1.984
I15	HG2#	1H	1.095
I15	HN	1H	8.194
I15	N	15N	121.180
V16	CA	13C	67.260
V16	CB	13C	31.529
V16	CG1	13C	23.297
V16	CG2	13C	23.655
V16	CO	1H	178.481
V16	HA	1H	3.567
V16	HB	1H	2.264
V16	HG1#	1H	0.996
V16	HG2#	1H	1.197
V16	HN	1H	8.924
V16	N	15N	118.426
N17	CA	13C	58.705
N17	CB	13C	41.281
N17	CO	1H	177.651
N17	HA	1H	4.636
N17	HB1	1H	2.774
N17	HB2	1H	3.110
N17	HD21	1H	6.870
N17	HD22	1H	8.082
N17	HN	1H	8.687
N17	N	15N	120.845
N17	ND2	15N	113.161
E18	CA	13C	59.558
E18	CB	13C	29.543
E18	CG	13C	35.832
E18	CO	1H	178.517
E18	HA	1H	3.960
E18	HB1	1H	2.171
E18	HB2	1H	2.274
E18	HG#	1H	2.508
E18	HN	1H	8.280
E18	N	15N	120.771
I19	CA	13C	63.432
I19	CB	13C	39.373
I19	CD1	13C	13.977
I19	CG1	13C	30.008
I19	CG2	13C	17.254
I19	CO	1H	177.024
I19	HA	1H	3.964
I19	HB	1H	1.656
I19	HD1#	1H	0.800
I19	HG1#	1H	1.357
I19	HG11	1H	1.347
I19	HG12	1H	1.640
I19	HG2#	1H	0.976
I19	HN	1H	8.159
I19	N	15N	118.994
A20	CA	13C	51.931
A20	CB	13C	21.813
A20	CO	1H	178.516
A20	HA	1H	4.820
A20	HB#	1H	1.547
A20	HN	1H	8.692
A20	N	15N	119.935
G21	CA	13C	46.318
G21	CO	1H	173.926
G21	HA1	1H	3.956
G21	HA2	1H	4.050
G21	HN	1H	7.863
G21	N	15N	108.061
I22	CA	13C	57.098
I22	CB	13C	39.058
I22	CD1	13C	11.423
I22	CG1	13C	27.710
I22	CG2	13C	17.143
I22	HA	1H	4.346
I22	HB	1H	1.660
I22	HD1#	1H	0.848
I22	HG11	1H	1.374
I22	HG12	1H	1.310
I22	HG2#	1H	1.013
I22	HN	1H	6.415
I22	N	15N	121.126
P23	CA	13C	63.551
P23	CB	13C	32.286
P23	CD	13C	51.213
P23	CG	13C	28.349
P23	CO	1H	178.044
P23	HA	1H	4.331
P23	HB1	1H	1.769
P23	HB2	1H	2.371
P23	HD1	1H	3.710

P23	HD2	1H	4.039
P23	HG1	1H	1.965
P23	HG2	1H	2.140
V24	CA	13C	66.485
V24	CB	13C	32.175
V24	CG1	13C	21.434
V24	CG2	13C	21.091
V24	CO	1H	179.131
V24	HA	1H	3.627
V24	HB	1H	2.057
V24	HG1#	1H	0.992
V24	HG2#	1H	1.023
V24	HN	1H	8.515
V24	N	15N	121.712
E25	CA	13C	58.435
E25	CB	13C	28.585
E25	CG	13C	36.029
E25	CO	1H	176.616
E25	HA	1H	4.065
E25	HB#	1H	2.013
E25	HG#	1H	2.304
E25	HN	1H	9.255
E25	N	15N	117.635
D26	CA	13C	54.709
D26	CB	13C	40.909
D26	CO	1H	176.401
D26	HA	1H	4.634
D26	HB1	1H	2.652
D26	HB2	1H	2.693
D26	HN	1H	7.790
D26	N	15N	117.977
V27	CA	13C	64.016
V27	CB	13C	29.615
V27	CG1	13C	22.014
V27	CG2	13C	21.099
V27	CO	1H	172.911
V27	HA	1H	3.555
V27	HB	1H	2.342
V27	HG1#	1H	0.748
V27	HG2#	1H	0.948
V27	HN	1H	7.489
V27	N	15N	122.031
K28	CA	13C	53.640
K28	CB	13C	36.491
K28	CD	13C	25.185
K28	CE	13C	42.595
K28	CG	13C	29.223
K28	CO	1H	177.055
K28	HA	1H	4.657
K28	HB1	1H	1.548
K28	HB2	1H	2.090
K28	HD1	1H	1.340
K28	HE#	1H	3.064
K28	HG1	1H	1.694
K28	HN	1H	6.309
K28	N	15N	123.007
L29	CA	13C	58.215
L29	CB	13C	42.934
L29	CD1	13C	24.195
L29	CD2	13C	25.572
L29	CG	13C	27.213
L29	CO	1H	177.391
L29	HA	1H	3.586
L29	HB1	1H	1.475
L29	HB2	1H	1.759
L29	HD1#	1H	0.822
L29	HD2#	1H	0.918
L29	HG	1H	1.778
L29	HN	1H	8.484
L29	N	15N	118.077
D30	CA	13C	53.751
D30	CB	13C	40.027
D30	CO	1H	176.219
D30	HA	1H	4.494
D30	HB1	1H	2.607
D30	HB2	1H	2.795
D30	HN	1H	8.351
D30	N	15N	112.387
K31	CA	13C	54.166
K31	CB	13C	32.431
K31	CD	13C	32.301
K31	CE	13C	42.395
K31	CG	13C	28.434
K31	CO	1H	175.758
K31	HA	1H	4.382
K31	HB1	1H	1.814
K31	HD1	1H	1.390
K31	HD2	1H	2.076
K31	HE1	1H	2.968
K31	HE2	1H	3.035
K31	HG1	1H	1.484
K31	HG2	1H	1.581
K31	HN	1H	7.496
K31	N	15N	117.482
S32	CA	13C	56.167
S32	CB	13C	63.370
S32	CO	1H	177.641
S32	HA	1H	4.877
S32	HB1	1H	3.858
S32	HB2	1H	4.108
S32	HN	1H	10.104
S32	N	15N	117.600
F33	CA	13C	63.667
F33	CB	13C	39.267
F33	CO	1H	176.281
F33	HA	1H	4.095
F33	HB1	1H	3.148
F33	HB2	1H	3.227
F33	HD#	1H	7.115
F33	HE#	1H	6.955
F33	HN	1H	8.362
F33	HZ	1H	6.795
F33	N	15N	127.158
T34	CA	13C	64.534
T34	CB	13C	69.365
T34	CG2	13C	22.118

T34	HA	1H	4.355
T34	HB	1H	4.109
T34	HG2#	1H	1.306
T34	HN	1H	8.761
T34	N	15N	109.999
D35	CA	13C	56.692
D35	CB	13C	41.217
D35	CO	1H	176.412
D35	HA	1H	4.628
D35	HB1	1H	2.456
D35	HB2	1H	2.784
D35	HN	1H	7.980
D35	N	15N	118.286
D36	CA	13C	56.015
D36	CB	13C	43.780
D36	CO	1H	177.019
D36	HA	1H	5.059
D36	HB1	1H	2.706
D36	HB2	1H	3.079
D36	HN	1H	7.808
D36	N	15N	114.219
L37	CA	13C	54.278
L37	CB	13C	41.466
L37	CD1	13C	26.458
L37	CD2	13C	22.973
L37	CG	13C	26.605
L37	CO	1H	173.874
L37	HA	1H	4.451
L37	HB1	1H	1.702
L37	HB2	1H	2.305
L37	HD1#	1H	0.593
L37	HD2#	1H	0.773
L37	HG	1H	1.470
L37	HN	1H	7.317
L37	N	15N	116.275
D38	CA	13C	55.351
D38	CB	13C	39.539
D38	CO	1H	174.926
D38	HA	1H	4.336
D38	HB1	1H	2.572
D38	HB2	1H	3.152
D38	HN	1H	7.431
D38	N	15N	112.779
V39	CA	13C	62.557
V39	CB	13C	31.843
V39	CG1	13C	22.462
V39	CG2	13C	22.091
V39	CO	1H	174.943
V39	HA	1H	3.999
V39	HB	1H	1.988
V39	HG1#	1H	0.838
V39	HG2#	1H	0.948
V39	HN	1H	8.328
V39	N	15N	119.578
D40	CA	13C	53.127
D40	CB	13C	41.608
D40	CO	1H	176.010
D40	HA	1H	4.627
D40	HB1	1H	2.797
D40	HB2	1H	3.151
D40	HN	1H	8.130
D40	N	15N	125.946
S41	CA	13C	62.438
S41	CB	13C	62.656
S41	CO	1H	176.720
S41	HA	1H	4.008
S41	HB#	1H	3.915
S41	HN	1H	8.478
S41	N	15N	113.387
L42	CA	13C	57.811
L42	CB	13C	41.330
L42	CD1	13C	23.666
L42	CD2	13C	24.848
L42	CG	13C	27.432
L42	CO	1H	180.361
L42	HA	1H	4.289
L42	HB1	1H	1.643
L42	HB2	1H	1.873
L42	HD1#	1H	0.904
L42	HD2#	1H	0.969
L42	HG	1H	1.697
L42	HN	1H	7.874
L42	N	15N	122.485
S43	CA	13C	61.733
S43	CB	13C	63.010
S43	CO	1H	176.245
S43	HA	1H	4.308
S43	HB1	1H	3.713
S43	HB2	1H	4.023
S43	HN	1H	8.391
S43	N	15N	118.275
M44	CA	13C	56.684
M44	CB	13C	30.870
M44	CG	13C	32.461
M44	CO	1H	177.872
M44	HA	1H	4.028
M44	HB1	1H	1.866
M44	HB2	1H	2.019
M44	HE#	1H	0.219
M44	HG1	1H	2.182
M44	HG2	1H	2.527
M44	HN	1H	8.616
M44	N	15N	121.704
V45	CA	13C	66.745
V45	CB	13C	31.859
V45	CG1	13C	21.168
V45	CG2	13C	22.295
V45	CO	1H	177.858
V45	HA	1H	3.526
V45	HB	1H	2.229
V45	HG1#	1H	0.952
V45	HG2#	1H	1.121
V45	HN	1H	7.174
V45	N	15N	117.525
E46	CA	13C	59.080
E46	CB	13C	29.360

E46	CG	13C	35.961
E46	CO	1H	178.228
E46	HA	1H	4.050
E46	HB1	1H	2.164
E46	HB2	1H	2.331
E46	HG1	1H	2.341
E46	HG2	1H	2.487
E46	HN	1H	6.938
E46	N	15N	116.654
V47	CA	13C	58.610
V47	CB	13C	39.508
V47	CG1	13C	19.717
V47	CG2	13C	22.965
V47	CO	1H	179.731
V47	HA	1H	3.953
V47	HB	1H	1.660
V47	HG1#	1H	0.651
V47	HG2#	1H	0.968
V47	HN	1H	8.178
V47	N	15N	117.067
V48	CA	13C	66.168
V48	CB	13C	31.204
V48	CG1	13C	19.754
V48	CG2	13C	22.996
V48	CO	1H	177.669
V48	HA	1H	3.416
V48	HB	1H	2.382
V48	HG1#	1H	0.641
V48	HG2#	1H	0.966
V48	HN	1H	8.043
V48	N	15N	121.661
V49	CA	13C	66.416
V49	CB	13C	31.995
V49	CG1	13C	21.340
V49	CG2	13C	22.902
V49	CO	1H	179.127
V49	HA	1H	3.986
V49	HB	1H	2.064
V49	HG1#	1H	1.000
V49	HG2#	1H	1.110
V49	HN	1H	8.199
V49	N	15N	119.298
A50	CA	13C	55.054
A50	CB	13C	19.013
A50	CO	1H	180.998
A50	HA	1H	4.258
A50	HB#	1H	1.553
A50	HN	1H	7.821
A50	N	15N	121.250
A51	CA	13C	55.341
A51	CB	13C	17.293
A51	CO	1H	179.294
A51	HA	1H	3.977
A51	HB#	1H	1.525
A51	HN	1H	8.894
A51	N	15N	123.402
E52	CA	13C	59.861
E52	CB	13C	30.284
E52	CG	13C	36.621
E52	CO	1H	179.851
E52	HA	1H	4.197
E52	HB1	1H	2.170
E52	HB2	1H	2.541
E52	HG1	1H	2.390
E52	HG2	1H	2.868
E52	HN	1H	8.928
E52	N	15N	119.116
E53	CA	13C	58.657
E53	CB	13C	29.862
E53	CG	13C	35.879
E53	CO	1H	178.676
E53	HA	1H	4.150
E53	HB1	1H	2.078
E53	HB2	1H	2.155
E53	HG1	1H	2.318
E53	HG2	1H	2.502
E53	HN	1H	7.606
E53	N	15N	116.615
R54	CA	13C	58.356
R54	CB	13C	30.732
R54	CD	13C	42.918
R54	CG	13C	27.032
R54	CO	1H	177.857
R54	HA	1H	3.945
R54	HB1	1H	1.092
R54	HB2	1H	1.575
R54	HD1	1H	2.731
R54	HD2	1H	2.890
R54	HG1	1H	0.702
R54	HG2	1H	0.872
R54	HN	1H	8.166
R54	N	15N	118.392
F55	CA	13C	57.307
F55	CB	13C	39.622
F55	CO	1H	174.459
F55	HA	1H	4.875
F55	HB1	1H	2.753
F55	HB2	1H	3.450
F55	HD#	1H	7.471
F55	HE#	1H	7.147
F55	HN	1H	8.207
F55	HZ	1H	7.465
F55	N	15N	112.795
D56	CA	13C	55.273
D56	CB	13C	39.055
D56	CO	1H	174.557
D56	HA	1H	4.531
D56	HB1	1H	2.599
D56	HB2	1H	3.261
D56	HN	1H	7.865
D56	N	15N	120.045
V57	CA	13C	59.291
V57	CB	13C	36.330
V57	CG1	13C	20.686
V57	CG2	13C	21.760
V57	CO	1H	173.658

V57	HA	1H	4.726
V57	HB	1H	2.010
V57	HG1#	1H	0.955
V57	HG2#	1H	1.004
V57	HN	1H	7.764
V57	N	15N	112.810
K58	CA	13C	54.564
K58	CB	13C	33.813
K58	CD	13C	24.807
K58	CG	13C	28.577
K58	CO	1H	175.800
K58	HA	1H	4.712
K58	HB#	1H	1.720
K58	HD#	1H	1.362
K58	HE#	1H	2.963
K58	HG#	1H	1.650
K58	HN	1H	8.274
K58	N	15N	122.905
I59	CA	13C	58.446
I59	CB	13C	37.794
I59	CD1	13C	16.772
I59	CG1	13C	26.694
I59	CG2	13C	12.121
I59	HA	1H	4.487
I59	HB	1H	1.672
I59	HD1#	1H	0.239
I59	HG11	1H	0.851
I59	HG12	1H	1.519
I59	HG2#	1H	0.623
I59	HN	1H	9.546
I59	N	15N	126.235
P60	CA	13C	62.688
P60	CB	13C	32.810
P60	CD	13C	51.286
P60	CG	13C	27.813
P60	CO	1H	177.478
P60	HA	1H	4.371
P60	HB1	1H	1.921
P60	HB2	1H	2.403
P60	HD#	1H	3.811
P60	HD1	1H	3.749
P60	HD2	1H	3.859
P60	HG#	1H	2.018
P60	HG1	1H	1.929
P60	HG2	1H	2.020
D61	CA	13C	58.054
D61	CB	13C	40.549
D61	CO	1H	178.710
D61	HA	1H	4.775
D61	HB#	1H	2.614
D61	HN	1H	8.613
D61	N	15N	122.523
D62	CA	13C	56.803
D62	CB	13C	40.393
D62	CO	1H	176.901
D62	HA	1H	4.328
D62	HB1	1H	2.510
D62	HB2	1H	2.649
D62	HN	1H	8.777
D62	N	15N	116.146
D63	CA	13C	55.456
D63	CB	13C	40.991
D63	CO	1H	178.523
D63	HA	1H	4.685
D63	HB#	1H	2.406
D63	HN	1H	7.552
D63	N	15N	117.068
V64	CA	13C	66.918
V64	CB	13C	31.719
V64	CG1	13C	21.568
V64	CG2	13C	24.466
V64	CO	1H	178.304
V64	HA	1H	3.085
V64	HB	1H	2.068
V64	HG1#	1H	0.776
V64	HG2#	1H	0.936
V64	HN	1H	7.566
V64	N	15N	121.705
K65	CA	13C	58.100
K65	CB	13C	31.942
K65	CD	13C	23.380
K65	CE	13C	42.290
K65	CG	13C	29.322
K65	CO	1H	176.424
K65	HA	1H	4.034
K65	HB#	1H	1.812
K65	HD1	1H	1.359
K65	HD2	1H	1.450
K65	HE#	1H	3.008
K65	HG#	1H	1.688
K65	HN	1H	7.449
K65	N	15N	115.770
N66	CA	13C	53.028
N66	CB	13C	39.066
N66	CO	1H	175.587
N66	HA	1H	4.878
N66	HB1	1H	2.776
N66	HB2	1H	3.079
N66	HD21	1H	7.020
N66	HD22	1H	7.746
N66	HN	1H	7.252
N66	N	15N	114.075
N66	ND2	15N	114.117
L67	CA	13C	53.311
L67	CB	13C	39.002
L67	CD1	13C	25.148
L67	CD2	13C	22.522
L67	CG	13C	25.148
L67	CO	1H	174.344
L67	HA	1H	4.511
L67	HB1	1H	0.927
L67	HB2	1H	2.142
L67	HD1#	1H	0.440
L67	HD2#	1H	0.541
L67	HG	1H	1.819
L67	HN	1H	7.753

L67	N	15N	122.490
K68	CA	13C	58.335
K68	CB	13C	33.537
K68	CD	13C	25.461
K68	CE	13C	42.384
K68	CG	13C	28.524
K68	CO	1H	179.317
K68	HA	1H	4.537
K68	HB#	1H	2.119
K68	HD1	1H	1.504
K68	HD2	1H	1.603
K68	HE#	1H	3.139
K68	HG#	1H	1.785
K68	HN	1H	8.391
K68	N	15N	122.587
T69	CA	13C	58.571
T69	CB	13C	73.707
T69	CG2	13C	21.447
T69	CO	1H	175.827
T69	HA	1H	5.252
T69	HB	1H	4.369
T69	HG1	1H	5.329
T69	HG2#	1H	1.063
T69	HN	1H	8.737
T69	N	15N	108.951
V70	CA	13C	66.259
V70	CB	13C	31.326
V70	CG1	13C	23.320
V70	CG2	13C	20.767
V70	CO	1H	180.125
V70	HA	1H	3.333
V70	HB	1H	2.270
V70	HG1#	1H	0.443
V70	HG2#	1H	0.642
V70	HN	1H	8.076
V70	N	15N	121.309
G71	CA	13C	47.731
G71	CO	1H	175.773
G71	HA1	1H	3.806
G71	HA2	1H	4.025
G71	HN	1H	9.365
G71	N	15N	111.986
D72	CA	13C	57.720
D72	CB	13C	41.864
D72	CO	1H	179.648
D72	HA	1H	4.527
D72	HB1	1H	2.964
D72	HB2	1H	3.188
D72	HN	1H	8.352
D72	N	15N	124.263
A73	CA	13C	55.347
A73	CB	13C	17.634
A73	CO	1H	177.671
A73	HA	1H	4.131
A73	HB#	1H	1.538
A73	HN	1H	8.009
A73	N	15N	121.717
T74	CA	13C	67.326
T74	CB	13C	68.405
T74	CG2	13C	21.018
T74	CO	1H	175.388
T74	HA	1H	3.682
T74	HB	1H	4.194
T74	HG2#	1H	1.099
T74	HN	1H	8.658
T74	N	15N	116.190
K75	CA	13C	59.222
K75	CB	13C	32.177
K75	CD	13C	25.016
K75	CE	13C	42.309
K75	CG	13C	28.896
K75	CO	1H	177.283
K75	HA	1H	3.945
K75	HB1	1H	1.939
K75	HB2	1H	1.972
K75	HD1	1H	1.463
K75	HD2	1H	1.510
K75	HE#	1H	3.054
K75	HG1	1H	1.673
K75	HG2	1H	1.785
K75	HN	1H	8.157
K75	N	15N	120.774
Y76	CA	13C	62.721
Y76	CB	13C	39.208
Y76	CO	1H	180.332
Y76	HA	1H	4.203
Y76	HB1	1H	3.047
Y76	HB2	1H	3.127
Y76	HD#	1H	6.820
Y76	HE#	1H	6.583
Y76	HN	1H	7.967
Y76	N	15N	118.279
I77	CA	13C	66.730
I77	CB	13C	37.832
I77	CD1	13C	13.846
I77	CG1	13C	30.213
I77	CG2	13C	17.201
I77	CO	1H	177.262
I77	HA	1H	3.537
I77	HB	1H	1.813
I77	HD1#	1H	0.682
I77	HG11	1H	2.031
I77	HG12	1H	0.953
I77	HG2#	1H	0.554
I77	HN	1H	8.472
I77	N	15N	121.194
L78	CA	13C	58.462
L78	CB	13C	42.502
L78	CD1	13C	25.505
L78	CD2	13C	25.432
L78	CG	13C	26.766
L78	CO	1H	179.751
L78	HA	1H	3.899
L78	HB1	1H	1.658
L78	HB2	1H	1.779
L78	HD1#	1H	0.830

L78	HD2#	1H	0.842	H80	N	15N	114.173
L78	HG	1H	1.637	Q81	CA	13C	56.399
L78	HN	1H	8.523	Q81	CB	13C	28.713
L78	N	15N	121.090	Q81	CG	13C	33.192
D79	CA	13C	55.679	Q81	CO	1H	174.340
D79	CB	13C	40.623	Q81	HA	1H	4.455
D79	CO	1H	177.273	Q81	HB1	1H	2.146
D79	HA	1H	4.410	Q81	HB2	1H	2.355
D79	HB1	1H	2.229	Q81	HE21	1H	7.762
D79	HB2	1H	2.530	Q81	HE22	1H	7.706
D79	HN	1H	8.304	Q81	HG1	1H	2.579
D79	N	15N	115.765	Q81	HG2	1H	2.717
H80	CA	13C	57.147	Q81	HN	1H	7.891
H80	CB	13C	28.986	Q81	N	15N	120.398
H80	CO	1H	174.141	Q81	NE2	15N	112.235
H80	HA	1H	4.683	A82	CA	13C	54.061
H80	HB1	1H	2.556	A82	CB	13C	20.193
H80	HB2	1H	3.638	A82	HA	1H	4.206
H80	HD2	1H	7.078	A82	HB#	1H	1.447
H80	HE1	1H	8.097	A82	HN	1H	8.317
H80	HN	1H	7.720	A82	N	15N	130.218

Table 7-1 Chemical shift assignments for *apo* FAS ACP.