Structural Studies of Nitrogen Regulation
in Escherichia coli

A thesis submitted for the degree of Master of Science
of
The Australian National University

Peter Mark Suffolk
December 1996
I would like to acknowledge Peter Mark Suffolk for his supervision. I would also like to thank the members of the Department of Civil Engineering who have helped me throughout the research process. Special thanks to Paul Carr and Danny Dearden. The support and encouragement of my family has helped me to complete this thesis. Finally and most importantly, I would like to thank my partner. Your unwavering support and encouragement have been invaluable. Your love and belief in me have enabled me to achieve the most fulfilling academic experience and for that I will forever be grateful.

Declaration

Except as otherwise noted, the work presented in this thesis is the work of the author. This work has not been presented, in whole or in part, for any other degree.

Peter Mark Suffolk
Acknowledgments

I would like to acknowledge and thank David Ollis for his supervision. Much thanks and appreciation must also go to Nick Dixon for his helpful and light-hearted advice on approaching the “grey-areas” of science and life. His words have kept me focused on the prize!

Thanks must also go to John Barton, Paul Carr and Jenny Daniher. Their individual help with so many things both inside and outside the lab made graduate school an enjoyable experience.

My gym partner, Richard Baldwin also deserves much thanks. Our friendship and matching desire for the ultimate work-out kept us sane as we both dealt with the frustration’s embodied by science.

I would especially like to thank the members of the Dixon/Ollis/Torda research groups. Each in their own way helped me to face the new challenges with renewed vigour and enthusiasm. The mountain of support made mole-hills of my ‘Everests’!

I would also like to thank my family. Their encouragement and support throughout my education has helped me at many needy times.

Finally and most importantly I would like to thank Sara. Her unmatched love, enthusiasm and zeal for life in its fullest form has inspired and supported me in the most trialing times. Sara, her love and her belief in me have projected me towards the most fulfilling world imaginable and for that I will always be thankful.
Dedication

To Sara, my best friend.

“No bird soars too high, if he soars with his own wings.”

~William Blake
Abstract

This thesis describes the determination of the high resolution X-ray crystal structure of the P\textsubscript{II} protein from \textit{Escherichia coli}. This structure is then compared to other proteins containing a double \( \beta\-\alpha\-\beta \) motif. These comparisons yield further insight into the nature of these motifs and their relevance in P\textsubscript{II}. Also reported are the attempts to co-crystallise P\textsubscript{II} with 2-ketoglutarate, glutamate and the nitrogen regulatory protein, NR\textsubscript{II}.

The P\textsubscript{II} protein is the product of the \textit{glnB} gene found at the 55 minute region within the \textit{E. coli} chromosome. At the beginning of this work the purification protocol for P\textsubscript{II} existed and initial crystallisation had been achieved. This study details the extension and further contributions to this investigation that resulted in the solution of the P\textsubscript{II} structure. In this investigation a better separation of P\textsubscript{II} from its uridylylated counterpart was achieved, resulting in better diffraction quality crystals. Sufficient native and heavy-atom derivative data were collected allowing solution of P\textsubscript{II} to 2.7 \( \AA \).

The central feature of the monomer is a double \( \beta\-\alpha\-\beta \) motif containing a four-stranded anti-parallel \( \beta \) sheet with two anti-parallel \( \alpha \) helices positioned behind the sheet. P\textsubscript{II} forms a trimer in a tricuspid arrangement that positions the helices outside of the molecule. Loop regions containing the functionally important residue tyrosine 51 (Tyr51) were found to be positioned away from the main body of the trimer.
The double β-α-β motif from P_I was compared to a number of different proteins that interact with phosphates or phosphate esters. Of these the S6 ribosomal protein was found to be the most similar to P_I. The result emerging from this comparison was that the double β-α-β motif does present a stable motif for binding phosphates, but in the case of P_I the nature of the trimer made such interactions unlikely. Instead a cleft formed between adjacent monomers in the P_I trimer seemed the most probable region for binding effector ligands such as 2-ketoglutarate. Co-crystallisation of P_I with 2-ketoglutarate, glutamate and NR_I were trialed in an attempt to further elucidate the nature of its interactions with other molecules that bind to the protein.
Publications


Contents

Declaration........................................................................... i

Acknowledgments.................................................................. ii

Dedication ........................................................................... iii

Abstract .............................................................................. iv

Publications......................................................................... vi

Contents............................................................................... vii

Chapter 1. Introduction....................................................... 1

1.1 Summary of work to be presented in this thesis.............. 1

1.2 Background..................................................................... 1

1.3 Aims .............................................................................. 10

Chapter 2. Purification and crystallisation of PII ............... 12

2.1 Introduction.................................................................... 12

2.2 Materials and Methods.................................................... 13

2.2.1 Fast Protein Liquid Chromatography....................... 13

2.2.2 Crystallisation apparatus......................................... 13

2.2.3 Mass Spectrometry................................................ 13

2.2.4 Overexpression of glnB gene product....................... 14

2.3 Purification....................................................................... 14

2.4 Molecular Weight Determination................................... 17

2.4.1 Mass Spectrometry................................................ 17

2.4.2 Sedimentation equilibrium.................................... 17

2.5 Crystallisation............................................................... 19

2.5.1 Crystal Morphology............................................... 19
Chapter 3. X-ray data collection .......................................................... 23
  3.1 Introduction ............................................................................ 23
  3.2 Materials and Methods .......................................................... 23
  3.3 Native data ........................................................................ 26
    3.3.1 Crystal form characterisation ....................................... 26
    3.3.2 Native data collection .................................................. 31
    3.3.3 Twinning by merohedry ............................................... 34
  3.4 Heavy-atom multiple isomorphous derivative data .................. 40
    3.4.1 Preparation of heavy-atom derivatives ......................... 40
    3.4.2 Derivative data collection and processing .................... 42
    3.4.3 Multiple isomorphous replacement ............................. 42

Chapter 4. Structure determination of P11 at 2.7 Å.............................. 53
  4.1 Introduction ........................................................................ 53
  4.2 Materials and Methods ........................................................ 53
  4.3 Oligomeric and low resolution structure .............................. 55
  4.4 Solvent flattening .............................................................. 55
  4.5 Cα chain trace and side-chain insertion ................................ 57
  4.6 P11 structure at 2.7 Å .......................................................... 64
    4.6.1 Monomeric structure ................................................ 64
    4.6.2 Trimeric structure .................................................... 68
    4.6.3 Crystal packing forces ............................................... 79

Chapter 5. Structural description of P11 ............................................. 80
  5.1 Introduction ........................................................................ 80
  5.2 Materials and methods ....................................................... 80
  5.3 Atomic refinement ............................................................ 81
  5.4 Features of P11 ................................................................ 83
    5.4.1 Monomeric structure ............................................... 83
    5.4.2 Trimeric structure .................................................... 83
5.4.3 Crystal packing forces ........................................ 87

5.5 Comparison of PII to other proteins ........................................ 90

5.5.1 Structural comparison of PII to other double
β-α-β motif proteins ............................................. 90

5.5.1.1 Introduction ........................................... 90

5.5.1.2 Comparison with S6 ................................ 91

5.5.1.3 Comparison with E2 ................................ 91

5.5.1.4 Comparison with NDPK .................................. 96

5.5.1.5 Comparison with ar-ATCase ....................... 100

5.5.1.6 Comparison with the activation
segment of PCPA ................................................. 100

5.5.1.7 Comparison of the U1 A RNA-binding
domain ............................................................... 101

5.5.1.8 Comparison with HPr .................................. 101

5.5.2 Relationship between double β-α-β motif and
protein function .................................................. 102

Chapter 6. Beyond the PII structure ............................................. 106

6.1 Introduction .................................................................. 106

6.2 Materials and methods ............................................. 107

6.3 PII with bound 2-ketoglutarate ................................. 108

6.3.1 Crystallisation ...................................................... 108

6.3.2 Data collection and refinement ............................. 108

6.4 PII with bound glutamate ........................................ 115

6.4.1 Crystallisation .............................................................. 115

6.4.2 Data collection ...................................................... 116

6.5 PII structure at pH 9.5 .................................................. 119

6.5.1 Crystal preparation and data collection ............ 119

6.5.2 Structure refinement ............................................. 120
Chapter 7. Overexpression and purification of \textit{glnL} and \textit{glnG} gene products .................................................................................................................. 127

7.1 Introduction ........................................................................................................... 127
7.2 Materials and Methods ....................................................................................... 128
    7.2.1 Bacterial Strains ............................................................................................... 128
    7.2.2 Growth Media .................................................................................................. 128
    7.2.3 Rapid Plasmid DNA Isolation .......................................................................... 128
    7.2.4 Restriction Endonuclease Digestion of DNA .............................................. 128
    7.2.5 Agarose Gel Electrophoresis .......................................................................... 130
    7.2.6 Ligation of DNA Fragments into Vectors .................................................... 130
    7.2.7 SDS Polyacrylamide Gel Electrophoresis ..................................................... 131
7.3 Insertion of \textit{Nd}e\text{I} restriction sites via oligonucleotide-directed mutagenesis ................................................. 132
7.4 Overexpression of NR\textsubscript{II}, gene product of \textit{glnL} .................................................. 134
    7.4.1 Expression from pND707 .............................................................................. 134
    7.4.2 Expression in pCL476 .................................................................................... 137
7.5 Purification of NR\textsubscript{II} ...................................................................................... 141
7.6 Attempted co-crystallisation of P\textsubscript{II} with NR\textsubscript{II} ................................................. 144
7.7 Overexpression of NR\textsubscript{I}, gene product of \textit{glnG} ................................................ 145
    7.7.1 Expression in pPL450 .................................................................................... 145
    7.7.2 Expression in pND707 ................................................................................... 147
    7.7.3 Expression in pETMCS-I .............................................................................. 151
    7.7.4 Expression in pCL476 .................................................................................... 153

Chapter 8. Conclusions and Discussions ........................................................................ 154

8.1 Role of P\textsubscript{II}-UMP in crystallisation of P\textsubscript{II} .............................................. 154
8.2 Structure of P\textsubscript{II} ......................................................................................... 155
8.3 Relevance of the double $\beta$–$\alpha$–$\beta$ motif in P\textsubscript{II} ........................................... 156
8.4 Overall conclusions ............................................................................................... 159
Chapter 1. Introduction

1.1 Summary of work to be presented in this thesis

The work described in this thesis is focused upon elucidating at an atomic level the way in which signal transduction is carried out by the P11 protein from Escherichia coli. Chapter 2 describes the improved purification of P11 and the crystallisations that resulted. The collection and treatment of native and derivative data forms the basis of Chapter 3, which is then extended by the structural solution of P11 at 2.7 Å in Chapter 4. The refined model of P11 and the comparison of this high resolution structure with other proteins containing double β-α-β motifs is the subject of Chapter 5. Chapters 6 and 7 report investigations regarding attempts to co-crystallise P11 with 2-ketoglutarate, glutamate and with NR11.

1.2 Background

The ability of enteric bacteria such as Escherichia coli to survive using ammonia as a sole source of nitrogen has been known for some time. The roles of various proteins and effectors in controlling appropriate levels of ammonia uptake have been investigated and reviewed (Magasanik, 1982; Stock et al., 1989 & Magasanik, 1993). Glutamine, glutamate and aspartate are the key intermediates for the biosynthesis of nitrogen compounds, with the last two being precursors for various amino acid biosynthesis pathways. These three compounds are formed by the following enzyme-catalysed reactions:

\[
\begin{align*}
\text{NH}_4^+ + 2\text{-ketoglutarate} + \text{NADPH} & \leftrightarrow \text{glutamate} + \text{NADP}^+ + \text{H}_2\text{O} \quad (1) \\
\text{glutamate} + \text{NH}_3 + \text{ATP} & \rightarrow \text{glutamine} + \text{ADP} + \text{P}_i \quad (2) \\
\text{glutamine} + 2\text{-ketoglutarate} + \text{NADPH} & \rightarrow 2\times \text{glutamate} + \text{NADP}^+ \quad (3) \\
\text{glutamine} + \text{oxaloacetate} & \leftrightarrow \text{aspartate} + 2\text{-ketoglutarate} \quad (4)
\end{align*}
\]
Under conditions of excess environmental NH$_3$ 85% of cellular nitrogen is derived from glutamate (Magasanik, 1993) via the glutamate dehydrogenase (GDH) catalysed reaction (1). When environmental ammonia levels are limiting, the reaction catalysed by GDH is thermodynamically unfavourable and an alternative NH$_3$-assimilation pathway is needed. In this situation incorporation of NH$_3$ as glutamine, in reaction (2) becomes the major mechanism of ammonia assimilation.

Reaction (2) is catalysed by the dodecameric enzyme glutamine synthetase (GS). The control of transcription and activation of GS is via the “two-component” regulatory system (Nixon et al., 1986; Ronson et al., 1987) illustrated in Figure 1.1. When the GS subunits are increasingly adenylylated by adenylyltransferase (ATase) GS increasingly loses activity with GS-(AMP)$_{12}$ being totally inactive (Rhee et al., 1985). In conditions of excess NH$_3$, GS is kept fully adenylylated as the GDH reaction supplies metabolic nitrogen to the cell in the form of glutamate. When levels of NH$_3$ drop, the GDH reaction is unfavourable and the mechanism shown in Figure 1.1 activates GS. It can be seen from this figure that the P$_{II}$ protein has a central role in controlling and directing these pathways.

P$_{II}$ is a 112 amino acid (a.a.) protein that is reversibly uridylylated at tyrosine 51 (Tyr51) (Son & Rhee, 1987) and was thought to exist as a tetramer (Brown et al., 1971). In response to concentrations of environmental ammonia (Kamberov et al., 1995), P$_{II}$ controls the levels of transcription of the Ntr regulon, which contains the glnALG operon. This operon contains the glnA (encoding GS), glnL (encoding
Figure 1.1 Schematic diagram representing the regulation of transcription and activation of glutamine synthetase (GS). Coloured in red are the sections of this pathway involved during activation of GS and transcription of \textit{glnA} from the strong promoter \textit{glnAp2}. Coloured in blue are those sections of the pathway involved in deactivation of GS.
Enhancement of GS NA II transcription at the strong promoter \( glnAp2 \)

\[ \text{NR}_1 \rightarrow \text{NR}_1 \text{-P} \]

\[ \text{NR}_1 + \text{P}_1 \rightarrow \text{NR}_2 \]

Repression of promoters \( glnAp1 \) and \( glnLp \)

\[ \text{UTase/UR} \]

\[ \text{ATase} \]

\[ \text{GS} \text{-(AMP)}_{12} \]

\[ \text{UTase/UR} \]

\[ \text{ATase} \]

\[ \text{P}_1 \text{-UMP} \]
the nitrogen regulatory protein II (NR\textsubscript{II}) and \textit{glnG} (encoding the nitrogen regulatory protein I (NR\textsubscript{I})) genes.

When levels of ammonia fall below 1 mM, the bifunctional uridylyltransferase/uridylyl-removing enzyme (UTase/UR) brings about uridylylation of P\textsubscript{II} at Tyr51 (Stadtman \textit{et al.}, 1980). P\textsubscript{II}-UMP acts in two separate ways to increase the level of activation and transcription of GS. The first way is via the binding of P\textsubscript{II}-UMP to ATase. The interaction of P\textsubscript{II}-UMP with ATase stimulates the de-adenylylation of GS, thus maintaining the activity of GS. The second pathway for activational and transcriptional control of GS involves NR\textsubscript{II}. In its uridylylated form P\textsubscript{II} has no ability to bind to NR\textsubscript{II}. In the absence of such interaction, NR\textsubscript{II} acts as a kinase phosphorylating NR\textsubscript{I} which then binds the strong promoter \textit{glnAp2} which is found upstream of \textit{glnA} (Magasanik, 1988). This promoter then stimulates transcription of the \textit{glnALG} operon, thereby increasing cellular levels of GS, NR\textsubscript{II} and NR\textsubscript{I}.

When the concentration of environmental ammonia increases above 1 mM the same high level activity of GS is not required because it becomes energy inefficient to maintain high levels of ATP hydrolysis required for GS activity. Subsequently P\textsubscript{II} acts to increasingly deactivate GS and reduce transcription of \textit{glnA}. Under these conditions de-uridylylation of P\textsubscript{II} is catalysed by UTase/UR. Unmodified P\textsubscript{II} is able to bind to NR\textsubscript{II} and promotes the kinase activity of NR\textsubscript{II} leading to dephosphorylation of NR\textsubscript{I} (Ninfa & Magasanik, 1986). NR\textsubscript{I} represses transcription of both \textit{glnA}—at a site near the \textit{glnAp2} promoter—and its own gene, \textit{glnG}, near the \textit{glnLp} promoter. In parallel to this P\textsubscript{II} still maintains interactions with ATase but in its
unmodified form stimulates the increasing adenylylation of the GS
dodecamer, thus inactivating GS.

The signal transduction role of the P_{II} protein in controlling the levels
of activation and transcription of GS is a crucial one in the
“two-component” regulatory system associated with nitrogen
assimilation. It is interesting however that P_{II}-like roles are not seen
in the “two-component” regulatory systems of chemotaxis, phosphate
regulation and osmoregulation (Appendix A). In phosphate regulation
PhoU stimulates the phosphatase activity of PhoR but has no influence
on the kinase activity of PhoR-P. Elucidation of the structure of P_{II}
might yield further information regarding the need of nitrogen
regulation for such a signal flagging protein, which appears
redundant in chemotaxis, phosphate regulation and osmoregulation.

A number of effectors are known to influence the function of P_{II},
depending on the status of cellular nitrogen (Engleman & Francis,
1978; Francis & Engleman, 1978). These ligands included
2-ketoglutarate and ATP. More recently it has been determined that
2-ketoglutarate in the presence of ATP is required for uridylylation of
P_{II} by UTase/UR and that glutamine is required by UTase/UR to inhibit
uridylylation of P_{II} and stimulate deuridylylation of P_{II}-UMP
(Kamberov et al., 1995).

It seems apparent from this brief outline above, that the signal
transduction function of P_{II} is the key flag for directing and controlling
the activity levels of GS in E. coli. As already mentioned the role of P_{II}
in regulation of nitrogen uptake in E. coli has been extensively studied.
The involvement of P_{II} in other biochemical pathways has also been
studied. P\textsubscript{II} and its homologues have been studied from the following organisms: \textit{Klebsiella pneumoniae} (Holtel & Merrick, 1988), \textit{Bradyrhizobium japonicum} (Martin \textit{et al.}, 1989), \textit{Rhizobium leguminosarum} (Colonna-Romano \textit{et al.}, 1987), \textit{Rhodobacter capsulatus} (Kranz \textit{et al.}, 1990), \textit{Rhodobacter sphaeroides} (Genbank accession number X71659), \textit{Azospirillum brasilense} (De Zamaroczy \textit{et al.}, 1990), \textit{Synechococcus} sp. (Tsinoramas \textit{et al.}, 1991) and \textit{Bacillus subtilis} (Wray \textit{et al.}, 1994). The a.a. sequences of these various P\textsubscript{II} proteins are aligned in Figure 1.2. This comparison shows the highly conserved nature of these sequences and suggests that in each case a similar mode of action is involved.

The P\textsubscript{II} proteins from \textit{K. pneumoniae}, \textit{B. japonicum}, \textit{R. leguminosarum}, \textit{R. capsulatus}, \textit{R. sphaeroides}, \textit{A. brasilense} and \textit{Synechococcus} sp. are all involved in regulation of cellular nitrogen assimilation. Most are thought to function in the same way as \textit{E. coli} P\textsubscript{II}. There are only a few exceptions among this group.

One such exception is found for \textit{B. subtilis}. The P\textsubscript{II} protein from this bacteria has the lowest identity to its \textit{E. coli} analog (41.1\%) and is the only member of this group which does not have a tyrosine at position 51. Instead of Tyr51, \textit{B. subtilis} has an isoleucine 51 which is at the start of a region of nine residues which lack any similarity or identity to the other P\textsubscript{II} analogues. Wray \textit{et al.} (1994) suggest that \textit{B. subtilis} P\textsubscript{II} participates in nitrate utilisation or adaptation to growth in this medium and the sequence differences reflect the differing function.
Figure 1.2  Alignment of PII sequences from nine species. Conserved residues across the nine species are highlighted in green. Bsu will generally be excluded from the comparison due to its low identity with Eco. Sequences were aligned using the BESTFIT program from the GCG package (Devereux et al., 1984).
<table>
<thead>
<tr>
<th></th>
<th>Identity (%)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eco</td>
<td>100.0</td>
<td>[A] Vasudevan et al., 1991</td>
</tr>
<tr>
<td>Bja</td>
<td>70.3</td>
<td>[C] Martin et al., 1989</td>
</tr>
<tr>
<td>Rle</td>
<td>65.2</td>
<td>[D] Colonna-Romano et al., 1987</td>
</tr>
<tr>
<td>Rca</td>
<td>67.8</td>
<td>[E] Kranz et al., 1990</td>
</tr>
<tr>
<td>Rsp</td>
<td>65.2</td>
<td>[F] GenBank (accession #X71659)</td>
</tr>
<tr>
<td>Abr</td>
<td>66.2</td>
<td>[G] De Zamaroczy et al., 1990</td>
</tr>
<tr>
<td>Ssp</td>
<td>70.5</td>
<td>[H] Tsinoremas et al., 1991</td>
</tr>
<tr>
<td>Bsu</td>
<td>41.1</td>
<td>[I] Wray et al., 1994</td>
</tr>
</tbody>
</table>

Abbreviations:  

References:  
Other differences include the fact that in *R. leguminosarum* expression of PII is itself nitrogen regulated (Colonna-Romano *et al.*, 1993) unlike the constitutive expression in *E. coli* (van Heeswijk *et al.*, 1993). PII from *Synechococcus* sp. differs from *E. coli* PII in that it signals fluctuations in nitrogen availability via phosphorylation of a serine residue (Forchhammer & de Marsac, 1994) despite having high sequence identity with PII from *E. coli* and containing an analogous Tyr51.

The information presented suggests a pathway for the regulation of nitrogen uptake in bacteria involving PII proteins that function in similar manners. At the start of and at the submission of this work there were no reported structural investigations regarding any of these PII homologues, but a preliminary crystallisation note regarding *E. coli* PII (Suh & Rhee, 1983) had been published. Further, despite the wealth of biochemical information regarding the systems of nitrogen regulation in *E. coli*, there was little known of the structural aspects of this bicyclic cascade mechanism.

1.3 Aims

At the beginning of the work presented in this thesis X-ray diffraction quality crystals of PII existed. The aims of this work were as follows. Firstly, to achieve greater separation of PII and PII-UMP during the preparation of the protein, in an attempt to achieve better quality crystals for X-ray studies. Secondly, to collect better native PII data and then gain sufficient phase information for the native structure factors, using the technique of multiple isomorphous replacement (MIR) and subsequent phase combination with models. Thirdly, to use this information to build a model of PII from the electron density maps generated from the X-ray data. Extensions of these investigations were
to study the structural changes to $P_{II}$ upon binding various biologically important ligands. Further to describe interactions between $P_{II}$ and $NR_{II}$ on both biochemical and atomic levels. It was hoped that the results of these investigations would further elucidate the way in which $P_{II}$ acts as a signal transducing protein and thus brings about regulation of GS.
Chapter 2. Purification and crystallisation of $P_{II}$

2.1 Introduction

The biochemical role of $P_{II}$ has been studied extensively (see Chapter 1 and Kamberov et al., 1994). The sedimentation equilibrium experiments of Adler, Purich and Stadtman (1975) indicated that $P_{II}$ was a tetrameric protein of molecular weight 44 kDa. A later study showed that $P_{II}$ was specifically uridylylated at tyrosine 51, not tyrosine 46, the only other tyrosine residue in the $P_{II}$ sequence (Son & Rhee, 1987).

The picture that we had at the beginning of this study was as follows. Depending on nitrogen availability, $P_{II}$ would exist either in its unmodified or its uridylylated form ($P_{II}$-UMP). In the presence of excess nitrogen $P_{II}$ remains unmodified and NR$_1$/NR$_{II}$ represses transcription of the $glnA$ gene, which encodes glutamine synthetase (GS). In conditions of nitrogen starvation the reverse of this cycle brings about increased transcription of $glnA$. $P_{II}$ thus acts as a signal transduction protein in a bicyclic cascade mechanism for the regulation of GS activity (see Chapter 1 and Atkinson et al., 1994). Despite the wealth of information regarding the biochemical aspects of $P_{II}$, there had been little in the literature regarding structural investigations of this protein.

At the start of this investigation there existed one preliminary X-ray study of crystals of $E. coli$ $P_{II}$ (Suh & Rhee, 1983). In this study $P_{II}$ crystallised in the I23 space group. These crystals diffracted poorly with only a few medium resolution spots recorded. No improvements on these crystallisation conditions or subsequent data collected has
been reported. A later report published during our studies (De Mel et al., 1994), detailed successful crystallisation of PII in both the P21 and P212121 space groups. Both of these crystal forms are reported to diffract beyond 3.0 Å, but no further structural investigations have been reported.

At the start of this investigation there were crystals of PII available that were known to be twinned and to be contaminated by the uridylylated form of this protein. As twinning was hampering our attempts at structure elucidation, this study set out to improve the purity of the PII protein and also to improve the quality of subsequent crystals.

2.2 Materials and Methods

2.2.1 Fast Protein Liquid Chromatography

The Fast protein liquid chromatography (FPLC) was carried out using a Pharmacia GP-250 duel pump FPLC with an UV-MII spectrophotometer and a paper chart recorder.

2.2.2 Crystallisation apparatus

The crystals used in this investigation were grown by vapour diffusion by either the hanging drop technique (McPhearson, 1990) or from sitting drops using the HR Micro-Bridge (HR3-312).

2.2.3 Mass Spectrometry

The molecular mass of PII was determined by matrix assisted laser desorption mass spectrometry using a VG TofSpec mass analyser, with 2-(4-Hydroxyphenylazo)benzoic acid as the matrix (Juhasz et al., 1993).
2.2.4 Overexpression of glnB gene product

The overproducing strain of the PII protein was provided by Dr S.G. Vasudevan (Vasudevan et al., 1991). This strain was constructed by transformation of AN1459 with pCG646. pCG646 was constructed by subcloning the 1338 bp EcoRI-SmaI fragment from pPL246 into the similar sites in the expression vector pPL450 (Lilley et al., 1993). The insert contained the entire coding sequence for PII, with the glnB gene under the transcriptional control of the tandem heat inducible λ promoters PR and PL provided by pPL450 (Lilley et al., 1993). pPL450 also carries the heat inducible λcl857 repressor gene.

2.3 Purification

The protein used in this work was purified according to the protocol previously described (Vasudevan et al., 1994). The identity of the PII protein was confirmed by sequencing of the N-terminal ten residues. Sequencing was performed by the Biomolecular Resource Facility (BRF) at the ANU. The N-terminal sequence confirmed that this was PII and indicated that the N-terminal methionine was present. Incomplete separation of PII and PII-UMP on Fractogel TSK DEAE-650 (Merck) led to contamination of unmodified PII with PII-UMP. The two forms of PII were incompletely separated, emerging from the Fractogel column as two overlapping peaks (Figure 2.1).

These two peaks eluted from the DEAE column between 0.17 M and 0.19 M NaCl, the first major peak being later determined as unmodified PII and the second, minor peak being PII-UMP. Although the same protocol was used in this study, a slight under-loading of the DEAE column resulted in a better separation (Figure 2.1). As PII and PII-UMP could not be distinguished visibly via sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), the improved
Figure 2.1 FPLC traces for the separation of P_{II} and P_{II}-UMP on a Fractogel TSK DEAE-650 resin as eluted by a NaCl gradient. The traces show the absorbance of protein at 280 nm across a NaCl gradient shown in percentage. In both separations unmodified P_{II} is eluted first, shouldered closely by P_{II}-UMP. Separation A was shown to have a 20.2 % level of P_{II}-UMP, while separation B had a 7.1 % degree of uridylylation.
separation was qualified via inspection of the FPLC trace of A_{280} and quantified by UV absorption spectroscopy. The trace of A_{280} for these two separations shows that the P_{II}:P_{II-UMP} ratio for separation A approaches 2:1 while the same ratio for separation B is close to 3:1 (Figure 2.1).

The level of P_{II-UMP} was determined according to a variant of the formula derived by Son and Rhee (1987) and which relates A_{340}, A_{280} and A_{260}. Assuming that P_{II} exists as a trimer (see Chapter 4) and re-solving their simultaneous equations, a new pair of formulae for the concentrations of P_{II-UMP} (Equation 1) and of unmodified P_{II} (Equation 2) were derived.

\begin{align*}
\text{Equation 1} & \quad \text{P}_{\text{II-UMP}} = \frac{A_{260} - 0.692A_{280} - 1.41836A_{340}}{7.2496} \\
\text{Equation 2} & \quad \text{P}_{\text{II}} = \frac{A_{260} + A_{280} - 17.74[\text{P}_{\text{II-UMP}}] - 5.09A_{340}}{2.89}
\end{align*}

Using these equations the major peak of Prep A was 20.2 % P_{II-UMP}, while the major peak of Prep B was 7.1 % P_{II-UMP}.

The protein purified from both separations was used in separate crystallisation experiments. This was done as the level of uridylylation was not at that time known and SDS-PAGE indicated that these preparations were essentially free from contamination. It was also
thought that the major component of the mixture, unmodified P$_{II}$, would crystallise in preference to the minor component, P$_{II}$-UMP.

2.4 Molecular Weight Determination

2.4.1 Mass Spectrometry
The monomeric molecular mass of the P$_{II}$ protein was determined using matrix assisted laser desorption mass spectrometry. Samples of P$_{II}$ were prepared and given to Dr Nick Dixon for analysis. The internal standards used were bovine insulin (M$_r$ 5 733.8) and chicken egg white lysozyme (M$_r$ 14 311.8), which yielded the subunit mass of unuridylylated P$_{II}$ of 12 435±7 Da (Figure 2.2). This value is the mean of the three independent measurements and all three measurements are within 2σ of this value. The predicted subunit molecular weight of unuridylylated P$_{II}$ based on its nucleotide sequence is 12 427 Da (Vasudevan et al., 1991) which is within our experimental error.

2.4.2 Sedimentation equilibrium
As described in Chapter 4, a low resolution map generated for P$_{II}$ indicated that this protein had formed a trimer in the crystalline state. It was at this point that sedimentation equilibrium experiments of the meniscus depletion type (Yphantis, 1964) were carried out by Dr Peter Jeffrey, in an attempt to determine the oligomeric state of P$_{II}$ in solution. These experiments indicated that the molecular weight of native P$_{II}$ was 35 000±1 000 Da. This suggested that in solution P$_{II}$ is a trimeric molecule (Jeffrey, 1991) and not a tetramer as previously suggested by Adler, Purich and Stadtman (1975). These experiments confirmed our earlier determination of the monomeric molecular weight by mass spectrometry.
Figure 2.2 Three independent mass spectra of the P11 protein. These spectra show the percent abundance of the respective ions detected at particular mass-to-charge (M/Z) values. Each of the above spectra were measured using protein from the purification that had a calculated 7.1% level of uridylylation. The concentration of the sample protein was 1.8 mg/ml.
2.5 Crystallisation

2.5.1 Crystal Morphology

Unmodified PII was crystallised by the hanging drop vapour diffusion technique (McPhearson, 1990) and has been described by Vasudevan et al. (1994). Crystals were grown from 10 µl drops (22 mg/ml initial concentration of protein) containing 1 M phosphate buffer at pH 7.0 and 1 M NaCl. These drops were sealed over 1 ml reservoirs of 2 M phosphate, pH 7.0. Crystals formed after 2 weeks at 4°C. These crystals were too small to be used for X-ray diffraction and were left to determine whether further crystal growth would occur. After an additional 2 weeks these crystals reached their maximum size.

These hexagonal crystals ranged in size from 0.1 mm x 0.1 mm x 0.1 mm up to 0.3 mm x 0.3 mm x 0.4 mm (Figure 2.3A). These crystals were suitable for X-ray data collection. Crystals were either mounted directly, or harvested from the drop and stored in one of two stabilisation buffers. The two stabilisation buffers were either 0.53 M NaCl and 1.6 M phosphate buffer, or 0.53 M NH₄Cl and 1.6 M phosphate buffer. Both stabilisation buffers were at pH 7.0. The NaCl and NH₄Cl salts were added to the stabilisation buffers as they appeared to both help crystal growth and minimise deterioration of the crystals with time (Dr Eong Cheah, pers. comm.). Due to the long term effect of the NH₄Cl buffer on crystals (see later) and also the need to standardise crystal stabilisation conditions, only the buffer containing NaCl was used. The NaCl buffer was chosen in preference to the NH₄Cl buffer as the sodium salt was less likely to interact with PII.
Figure 2.3 Photographs showing the various crystal morphologies found for P$_{II}$. (A) A double ended P$_{II}$ crystal grown using the protein from the FPLC separation shown in Figure 2.1A, (B) The diamond shaped morphology found using the protein from the FPLC separation shown in Figure 2.1B, (C) A double ended crystal, similar to (A), while being degraded by the NH$_4$Cl buffer, (D) Degradation of double ended P$_{II}$ crystals, by the NH$_4$Cl containing buffer. Shown are the progressive stages of crystal degradation leading ultimately to single diamond shaped crystals.
When samples of P11 with lower levels of P11-UMP contamination were used for crystallisation (refer to Figure 2.1), crystal morphology varied slightly from the initial hexagonal crystals (Figure 2.3B). The initial crystals were hexagonally shaped and had an apex at either end (Figure 2.3A). This new crystal morphology resembled the end sections seen in the previous crystals. It seemed apparent that the level of P11-UMP in the crystallisation experiments had affected the crystal morphology. With increased contamination, the midsection of the crystal extends to form a more elongated shape.

An interesting effect was found when bi-apex crystals were placed in the NH4Cl stabilisation buffer. When left in this buffer for a period of approximately two weeks, the midsection of these crystals degraded (Figure 2.3C), eventually leaving two small diamond shaped crystals (Figure 2.3D). The morphology of these resultant crystals was identical to that shown in Figure 2.3B. Furthermore, these new crystal “fragments” were as robust as their parent crystals had been prior to treatment with NH4Cl. It was also observed that the crystals with the morphology shown in Figure 2.3B exhibited a lower degree of twinning by merohedry (see section 3.3.3).

It seemed apparent that, in the absence of contaminating P11-UMP, unmodified P11 crystallised with diamond-shaped morphology while in the presence of P11-UMP the morphology was extended to give an elongated bi-apex hexagonal morphology. The P11-UMP “generated” midsection of these elongated crystals was physically much weaker than the P11 proper and treatment with the NH4Cl buffer mentioned previously, disrupted and removed this section.
The work described in this chapter yielded $\text{P}_{\text{II}}$ protein that had greatly reduced levels of contamination from $\text{P}_{\text{II}}$-UMP. The 13% reduction in $\text{P}_{\text{II}}$-UMP contamination ultimately resulted in crystals that had lower twin fractions. The implications of these results and the advantages for structure elucidation of native $\text{P}_{\text{II}}$ form the basis of Chapter 3.
Chapter 3. X-ray data collection

3.1 Introduction
With crystals large enough for X-ray data collection, it was feasible to determine the structure of P_{11} using the technique of multiple isomorphous replacement (MIR). For this technique, both native and heavy-atom derivative data needed to be collected. This chapter describes the collection and processing of both native and derivative datasets, and assesses the quality of this data. The phasing power of derivatives used to determine the structure of P_{11} at 2.7 Å is also described.

3.2 Materials and Methods
Space group and cell dimensions were initially determined by precession photography. The precession camera used was mounted on a Rigaku RU200 rotating anode X-ray generator operating at 50 kV and 180 mA with a Cu anode. X-ray diffraction data were collected with a Rigaku R-AXIS-IIC image plate detector mounted on a Rigaku RU200 rotating anode X-ray generator with a Cu anode and operating at 50 kV and 100 mA. Data were reduced with the software provided by the distributor, Molecular Structure Corporation (MSC). The MSC programs corrected the data for Lorentz-polarisation and the process of scaling compensated for decay and absorption. Derivative data were scaled to native data with the program PROTEIN (Steigemann, 1992). PROTEIN was also used for calculation of difference Patterson functions using the method of Rossmann (1960), refinement of heavy atom positions (Dickerson et al., 1961) and the calculation of Fourier syntheses as per Stryer et al. (1964).
During the course of data collection $P_{11}$ crystals were found to slip. Crystals that had morphologies similar to those shown in Figure 2.3A had minimal slippage while crystal types represented in Figure 2.3B or Figure 2.3D slipped considerably. This effect caused major problems with data collection and so an alternate method of crystal mounting was employed. Thus crystals used for data collection were mounted in one of two ways. The first involved mounting crystals in 0.8 mm or 1.0 mm diameter quartz capillaries by drawing the crystal into the capillary with mother liquor via a 0.5 ml insulin syringe (Becton-Dickinson). The crystals were positioned with fine glass fibres, the mother liquor drawn off with the syringe and then dried using fine strips of filter paper. Where necessary the orientation of the crystals in the capillary was optimised using glass fibres. A small amount of liquor was left some distance from the crystals to maintain a moist environment for the crystal, and then both ends of the capillary were sealed with wax. This sealed capillary was then mounted on a goniometer head and data collected. For crystals like that shown in Figure 2.3A the above method was appropriate.

The second method of crystal mounting used an apparatus similar to that employed in Laue stop-flow diffraction experiments (Carr et al., 1992). This apparatus was used to stop the crystal slippage which occurred with some crystal morphologies (see section 3.3.3 and above). Figure 3.1 shows how syringes were attached to capillary tubes packed with glass wool (Figure 3.1A). Glass wool was drawn out into long strands and packed into the quartz capillaries such that the wool fibres were packed parallel to the tube (Figure 3.1B). Crystals and liquid were drawn into the capillary until the crystal was immobilised against the
Figure 3.1 Apparatus similar to that used in Laue stop-flow experiments. Diagram A shows how an insulin syringe is attached to polythene tubing and then to a quartz capillary which is packed with glass fibres. Diagram B shows an expanded view of the boxed region and illustrates how the crystals were finally positioned for diffraction.
glass wool. With the glass wool packed appropriately (Figure 3.1B) there was negligible extraneous X-ray absorption.

3.3 Native data

3.3.1 Crystal form characterisation
A hexagonal “bi-apex” (Figure 2.3A) 0.1 mm x 0.1 mm x 0.3 mm P11 crystal was mounted on the R-AXIS and data were collected at 4 °C to determine both cell dimensions and space group. The crystal was aligned with its 6-fold axis along the φ-axis of the oscillation camera; approximately perpendicular to the X-ray beam. Four thirty minute still images were collected. From these it was determined that the crystals were in the P6 space group—or related space groups P61, P62, P63, P64, P65—and had cell dimensions of a = 61.67 Å and c = 56.30 Å, as determined by the MSC software (Higashi, 1990).

A similar crystal was mounted on the precession camera by Dr Paul Carr to confirm the initial space group determination. The precession photograph of the hko layer (Figure 3.2) confirmed the correct a and b cell dimensions. A precession photograph of the hol layer (Figure 3.3) confirmed c and showed axial extinctions that indicated the reflection condition of \( l = 2n+1 \) for \( (h,k = 0,0) \). This suggested that the space group was in fact P63.

When a third crystal was used for a second hko precession photograph (Figure 3.4) it seemed that an approximate 2-fold axis existed perpendicular to the 6-fold axis; this additional symmetry would mean that these crystals are in the P6322 space group. It was thought that the true space group of these crystals was P63 and that the additional pseudo-symmetry was most likely due to twinning by merohedry. The
Figure 3.2 X-ray precession photograph ($\mu = 15^\circ$) of a native P$_{II}$ crystal ($hkO$ zone). The morphology of this crystal was similar to that shown in Figure 2.3A.
Figure 3.3  X-ray precession photograph ($\mu = 15^\circ$) of a native P$_{II}$ crystal ($h0l$ zone). The morphology of this crystal was similar to that shown in Figure 2.3A. The extinctions along $00l$ indicated that the space group was in fact P6$_3$. 

Figure 3.4  X-ray precession photograph ($\bar{\mu} = 15^\circ$) of a native PII crystal (hk0 zone). The morphology of this crystal was similar to that shown in Figure 2.3A. An additional local two-fold axis is indicated by this precession photograph.
merohedric twinning relates a particular reflection \( hkl \) to a pseudo-symmetry related position \( k,h,l \), via the pseudo 2-fold axis (International Tables for X-ray crystallography, Vol. C, 1992). A space group specific program was written to routinely detwin data from the R-AXIS software using the method of Fisher and Sweet (1980). These programs are discussed more fully in section 3.3.3.

With the unit cell dimensions measured, the unit cell volume could be determined via the following;

\[
V = (a)^2(c)\sin(120°)
\]

where \( a \) & \( c \) are the unit cell dimensions (see above)

Using these cell dimensions the unit cell volume was determined to be \( 1.851 \times 10^5 \text{ Å}^3 \) and the volume of the asymmetric-unit (a.s.u.) \( 0.308 \times 10^5 \text{ Å}^3 \). Using this value for the volume of the a.s.u. and the monomeric weight of \( \text{P}_{11} \) (12 435 Da) \( V_m \) was calculated to be \( 2.5 \text{ Å}^3/\text{Da} \), which is close to the average \( V_m \) of the values calculated for many different crystalline proteins in a previous study (Matthews, 1968).

With only one molecule in the a.s.u. it seemed likely that the additional pseudo 2-fold was arising from symmetry within the molecule and not a part of the \( \text{P}_{63} \) space group. The above calculations were not consistent with \( \text{P}_{11} \) being a tetramer.

Using an analysis of the solvent content of 116 different crystal forms of globular proteins, Matthews (1968) derived formulae for calculating approximate fractions for both the protein and solvent content of a protein crystal, given its \( V_m \). Applying these equations to the \( V_m \) value for \( \text{P}_{11} \) indicated that these crystals are comprised of 49% (v/v) protein
and 51 % (v/v) solvent. The value for solvent content, and thus protein content, lies within the range determined by Matthews (27 % (v/v) to 65 % (v/v) solvent) and close to the most frequently determined solvent value found in that study, 43 % (v/v).

3.3.2 Native data collection

Two native datasets were collected for the determination of the PII structure. The first native dataset was collected on a crystal stabilised in the NaCl buffer described in section 2.5.1. This crystal was hexagonally shaped with one blunt end and an apex along its 6-fold axis. At room temperature this crystal diffracted to 1.8 Å (Figure 3.5), but only the data to 2.2 Å was sufficiently complete for structure determination and preliminary refinement. This dataset (NAT1) contained 6 774 unique reflections and had an $R_{\text{merge}}$† of 2.5 %. The 5 733 reflections with $I > 5\sigma_I$ were used for structure solution and refinement. This dataset contained 93.3 % of the unique data available between 10.0 Å and 2.2 Å (Table 3.1).

This dataset—along with derivative datasets—was used to solve the initial structure of PII at 2.7 Å, but a higher resolution dataset was required to extend the refinement to 1.9 Å (see Chapter 5). The second native dataset was collected from a NaCl buffer stabilised crystal, but in this case the data was collected at 4 °C, the temperature at which the

$\dagger\quad R_{\text{merge}} = \frac{\sum |I(i) - <I(h)>|}{\sum I(i)}$  I(i) is the $i^{th}$ integrated intensity of a reflection with indices $h = hkl$.  $<I(h)>$ is the averaged intensity of a reflection with indices $h = hkl$. 


Figure 3.5 An oscillation image from the room temperature $P_{II}$ native data set, NAT1. Data was collected with 45 minute exposures with $\Delta \phi = 3.5^\circ$. The crystal to film distance was 80.0 mm which set the diffraction limits to a resolution of 1.8 Å at the centre of the edge of the imaging plates. It can be seen that this crystal diffracted past 1.8 Å.
Table 3.1 Completeness of the 2.2 Å native dataset (NAT1). Completeness of data (%) is shown by shells between 10.0 Å and 2.2 Å, along with the number of reflections in each shell.

<table>
<thead>
<tr>
<th>(d_{\text{min}}) (Å)</th>
<th>(d_{\text{max}}) (Å)</th>
<th>No. of reflections</th>
<th>% complete</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.62</td>
<td>10.00</td>
<td>317</td>
<td>100.0</td>
</tr>
<tr>
<td>4.60</td>
<td>3.62</td>
<td>317</td>
<td>99.5</td>
</tr>
<tr>
<td>4.06</td>
<td>4.60</td>
<td>313</td>
<td>99.5</td>
</tr>
<tr>
<td>3.71</td>
<td>4.06</td>
<td>305</td>
<td>99.3</td>
</tr>
<tr>
<td>3.46</td>
<td>3.71</td>
<td>301</td>
<td>99.3</td>
</tr>
<tr>
<td>3.26</td>
<td>3.46</td>
<td>308</td>
<td>99.2</td>
</tr>
<tr>
<td>3.10</td>
<td>3.26</td>
<td>304</td>
<td>98.9</td>
</tr>
<tr>
<td>2.97</td>
<td>3.10</td>
<td>287</td>
<td>98.7</td>
</tr>
<tr>
<td>2.86</td>
<td>2.97</td>
<td>312</td>
<td>98.5</td>
</tr>
<tr>
<td>2.76</td>
<td>2.86</td>
<td>289</td>
<td>98.1</td>
</tr>
<tr>
<td>2.68</td>
<td>2.76</td>
<td>279</td>
<td>97.8</td>
</tr>
<tr>
<td>2.60</td>
<td>2.68</td>
<td>294</td>
<td>97.4</td>
</tr>
<tr>
<td>2.53</td>
<td>2.60</td>
<td>274</td>
<td>96.7</td>
</tr>
<tr>
<td>2.47</td>
<td>2.53</td>
<td>280</td>
<td>96.4</td>
</tr>
<tr>
<td>2.42</td>
<td>2.47</td>
<td>272</td>
<td>93.9</td>
</tr>
<tr>
<td>2.37</td>
<td>2.42</td>
<td>276</td>
<td>93.5</td>
</tr>
<tr>
<td>2.32</td>
<td>2.37</td>
<td>261</td>
<td>94.8</td>
</tr>
<tr>
<td>2.28</td>
<td>2.32</td>
<td>262</td>
<td>94.4</td>
</tr>
<tr>
<td>2.24</td>
<td>2.28</td>
<td>268</td>
<td>93.8</td>
</tr>
<tr>
<td>2.20</td>
<td>2.24</td>
<td>234</td>
<td>93.3</td>
</tr>
</tbody>
</table>
crystals were grown and harvested. This dataset (NAT2) contained 11 079 unique reflections that had an $R_{\text{merge}}$ of 7.1 % and of these 7 976 reflections had $I > 3\sigma_I$. This data comprised 83.1 % of the unique data between 10.0 Å and 1.9 Å (Table 3.2) and was used for the refinement of the $P_{\text{II}}$ structure at 1.9 Å.

Both the NAT1 and NAT2 datasets had some level of merohedric twinning. The NAT1 dataset collected at room temperature had a twin fraction of 0.06. The second dataset, NAT2, had a twin fraction of 0.17. The method for detwinning datasets is described below.

3.3.3 Twinning by merohedry

As indicated by both precession photography and native data collection, the majority of $P_{\text{II}}$ crystals had some degree of twinning by merohedry (Buerger, 1960). This twinning had to be computationally corrected in order to determine the structure of $P_{\text{II}}$. To do this, local space group specific programs incorporating the methods of Fisher and Sweet (1980) were written by Dr Paul Carr and Dr Eong Cheah to detwin the data output from the R-AXIS software. Datasets were routinely detwinned prior to being scaled with other datasets via the program PROTEIN or before being used in X-PLOR for crystallographic refinement.

A twin fraction (t.f.) quantitates the contribution of a twin to the reflections measured for a particular crystal. These fractions range from 0.0 to 0.5, where the maximum contribution from a twin can be half of the intensities. Detwinning involves determining these fractions for each dataset.
Table 3.2 Completeness of the 1.9 Å native dataset (NAT2). Completeness of data (%) is shown in twenty bins shells between 10.0 Å and 1.9 Å, along with the number of reflections in each shell.

<table>
<thead>
<tr>
<th>$d_{\text{min}}$ (Å)</th>
<th>$d_{\text{max}}$ (Å)</th>
<th>No. of reflections</th>
<th>% complete</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.95</td>
<td>10.00</td>
<td>405</td>
<td>80.9</td>
</tr>
<tr>
<td>4.01</td>
<td>4.95</td>
<td>397</td>
<td>82.4</td>
</tr>
<tr>
<td>3.53</td>
<td>4.01</td>
<td>409</td>
<td>84.3</td>
</tr>
<tr>
<td>3.22</td>
<td>3.53</td>
<td>410</td>
<td>84.4</td>
</tr>
<tr>
<td>3.00</td>
<td>3.22</td>
<td>415</td>
<td>84.9</td>
</tr>
<tr>
<td>2.82</td>
<td>3.00</td>
<td>397</td>
<td>83.7</td>
</tr>
<tr>
<td>2.68</td>
<td>2.82</td>
<td>400</td>
<td>83.5</td>
</tr>
<tr>
<td>2.57</td>
<td>2.68</td>
<td>410</td>
<td>83.1</td>
</tr>
<tr>
<td>2.47</td>
<td>2.57</td>
<td>405</td>
<td>84.7</td>
</tr>
<tr>
<td>2.39</td>
<td>2.47</td>
<td>418</td>
<td>86.7</td>
</tr>
<tr>
<td>2.31</td>
<td>2.39</td>
<td>419</td>
<td>86.4</td>
</tr>
<tr>
<td>2.25</td>
<td>2.31</td>
<td>404</td>
<td>87.1</td>
</tr>
<tr>
<td>2.19</td>
<td>2.25</td>
<td>425</td>
<td>86.4</td>
</tr>
<tr>
<td>2.14</td>
<td>2.19</td>
<td>388</td>
<td>82.9</td>
</tr>
<tr>
<td>2.09</td>
<td>2.14</td>
<td>398</td>
<td>81.1</td>
</tr>
<tr>
<td>2.05</td>
<td>2.09</td>
<td>381</td>
<td>82.6</td>
</tr>
<tr>
<td>2.00</td>
<td>2.05</td>
<td>370</td>
<td>77.6</td>
</tr>
<tr>
<td>1.97</td>
<td>2.00</td>
<td>398</td>
<td>80.6</td>
</tr>
<tr>
<td>1.93</td>
<td>1.97</td>
<td>360</td>
<td>78.1</td>
</tr>
<tr>
<td>1.90</td>
<td>1.93</td>
<td>367</td>
<td>77.3</td>
</tr>
</tbody>
</table>
Through the process of collecting heavy-atom derivative data it became apparent that there was a relationship between crystal morphology and the exhibited twin fraction of a crystal. When data were collected on the “bi-apex” crystals of the type shown in Figure 2.3A, a high twin fraction was usually observed. These twin fractions routinely ranged between 0.12 and 0.25. By comparison, when data were collected on diamond shaped crystals or the crystals degraded with NH$_4$Cl, shown in Figures 2.3B and 2.3D respectively, twin fractions between 0.00 and 0.10 were regularly found.

Correctly determining the t.f. for a dataset was important as an incorrect value would lead to the unnecessary exclusion of data. Part of the detwinning program used yielded a plot of the number of negative intensities calculated for values of t.f. between 0.0 and 0.45. Data with values of t.f. greater than 0.45 would be useless for our purposes. Ideally this plot should be linear with the x-intercept indicating the t.f. of the data. In practice a line of best-fit would have to be assigned. An investigation into one dataset—PII11 (Table 3.3)—was used to check that our assignment of the best fit yielded the optimal t.f.

The PII11 data were collected from a native P$_{II}$ bi-apex crystal soaked in saturated p-chloromercuriphenylsulfonic acid (PCMBS) for 7 hours at 4 °C. Prior to detwinning this 2.8 Å dataset had 21.0 % differences from the native. When this data was detwinned three possible lines could be ascribed to the resulting t.f. plot (Figure 3.6), which indicated values of 0.12, 0.09 and 0.07 for the t.f. Each of these values were used to detwin PII11 and the three resulting datasets loaded into PROTEIN. Each dataset was found to have similar differences from the native—approximately 21.0 %—but had less reflections as the t.f. was
Figure 3.6  Plot showing the number of intensities calculated to be negative as a function of twin fraction for the P1111 dataset. Shown are the possible lines of best-fit resulting in t.f. values of 0.07 (blue), 0.09 (black) and 0.12 (red).
overestimated. Difference Patterson maps showed marginally less detail in both \( z = 0 \) and \( z = 0.5 \) Harker sections (Figure 3.7)(see section 3.4.3) as t.f. increased.

These results seem to indicate that underestimation of t.f. is the most appropriate treatment and that our earlier line of best-fit applications had been optimal. For the PII11 dataset the t.f. value used to process the data was 0.09, given the improved signal-to-noise ratio indicated by Figure 3.7. A recent study within this lab has been carried out using the refined 1.9 Å P\(_{11}\) structure to test the validity of our detwinning methods and has confirmed the previous results.

The merohedric twinning in P\(_{11}\) is a macroscopic effect that seems to arise from the midsection of the elongated hexagonal crystals. The diamond shaped ends of the “bi-apex” crystals described and illustrated in Chapter 2 (Figure 2.3A) seemed to contain areas of low twin fraction in a crystal. In many instances the crystals were large enough to allow crystal alignment and subsequent data collection at one end of a crystal without interference from the high twin fraction midsection. In other instances it was possible to physically separate the two ends of an elongated crystal and to mount only the individual diamonds.

It also seemed that there was a relationship between the crystal morphology and the likelihood of that crystal slipping during data collection. If a “bi-apex” P\(_{11}\) crystal (Figure 2.3A) was mounted in a capillary without support and with its six-fold axis parallel to the axis of the capillary, then no slippage was detected. Diamond shaped or NH\(_4\)Cl degraded crystals (Figure 2.3B and Figure 2.3D) mounted without support, were more likely to slip. It was because of this
Figure 3.7 Difference Patterson maps generated for the P111 dataset. Shown are the $z = 0$ and the $z = 0.5$ Harker sections calculated using values of twin fraction of A) 0.07, B) 0.09 and C) 0.12. Arrows indicate the heavy-atom positions (see section 3.4.3). All contours in the maps shown are at $1 \sigma$ intervals.
physical instability that apparatus derived from the Laue stop-flow cells (Figure 3.1) was employed.

As a result of these observations it was now possible to easily and reliably choose low-twin crystals of P11 before mounting. If crystals were not large enough to be physically separated or to allow the X-ray beam through only the low-twinned section of a “bi-apex” crystal, then the crystal could be chemically treated with the NH₄Cl buffer to remove the highly twinned midsection. When the smaller, low twinned crystals were used for data collection the support of the glass wool-containing capillaries minimised crystal slippage.

3.4 Heavy-atom multiple isomorphous derivative data

3.4.1 Preparation of heavy-atom derivatives

Heavy-atom derivatives were prepared by soaking native crystals of P11 in various solutions of heavy-atom compounds for varying lengths of time. The conditions for the derivatives were iteratively refined. Successful derivatisation was first indicated by small superficial cracking within the weaker midsection of the crystal and then confirmed by X-ray diffraction. Of the crystals that survived soaking, 97 possible derivatives had data collected from them. The large number of datasets collected reflects the iterative nature of heavy-atom derivatisations. In the course of trying to gain sufficiently good phases for P11, twenty three different heavy-atom compounds were trialed including a number of double-atom soaks. Each of these different compounds were soaked until a suitable dataset was collected. Six of these datasets were used for MIR phasing of the native data. The conditions for the preparation of these six heavy-atom derivatives are summarised in Table 3.3.
Table 3.3 Heavy-atom derivatives used in isomorphous replacement.

<table>
<thead>
<tr>
<th>Derivative</th>
<th>Heavy-atom compound</th>
<th>Concentration</th>
<th>Soaking time</th>
</tr>
</thead>
<tbody>
<tr>
<td>PII11</td>
<td>PCMBS(^a)</td>
<td>saturated</td>
<td>7</td>
</tr>
<tr>
<td>PII23</td>
<td>Na(_2)[rCl(_6)]</td>
<td>20 mM</td>
<td>3.5</td>
</tr>
<tr>
<td>PII25</td>
<td>PCMBS + Na(_2)[rCl(_6)] + 10 mM</td>
<td>50% sat.</td>
<td>2</td>
</tr>
<tr>
<td>PII79</td>
<td>PCMBS + K(_2)[Pt(CN)(_4)] + 15 mM</td>
<td>50% sat.</td>
<td>1.25 (^b)</td>
</tr>
<tr>
<td>PII89</td>
<td>K(_2)[Pt(CN)(_6)]</td>
<td>20 mM</td>
<td>19.5 (^b)</td>
</tr>
<tr>
<td>PII90</td>
<td>K(_2)[Pt(CN)(_6)]</td>
<td>20 mM</td>
<td>20</td>
</tr>
</tbody>
</table>

\(^a\) PCMBS = p-chloromercuriphenylsulfonic acid.

\(^b\) Soaks were carried out sequentially in the order indicated.
3.4.2 Derivative data collection and processing

Derivative data were collected at different temperatures. PII11 was collected at room temperature while both PII23 and PII25 were collected at 7 °C. The final three derivatives used—PII79, PII89 and PII90—were all collected at 4 °C. Data were collected at 4 °C to minimise the temperature change the crystal experienced. As described in Chapter 2, crystallisation was at 4 °C and the change in temperature at the point of data collection seemed to both increase the mosaic spread of crystals and decrease their life in the X-ray beam. To curb this, crystals were mounted at 4 °C and placed in a thermally isolated container for transfer to the cold stream at the R-AXIS goniostat.

A summary of the intensity statistics for the final six derivatives used to solve the phase problem of the native dataset NAT1 are shown in Table 3.4. Rmerge values calculated using the total data for each data set were all less than 11 %. Native and derivative datasets were scaled together using the method of Rae (Rae, 1965; Rae & Blake, 1966) as incorporated in the program PROTEIN (Steigemann, 1992).

3.4.3 Multiple isomorphous replacement

The six heavy-atom derivatives used for MIR phase determination had large intensity differences from the NAT1 dataset (Table 3.4) with an average difference value of approximately 18 %. Heavy-atom sites were located by calculation of difference Patterson maps and Fourier syntheses (Rossmann, 1960; Stryer et al., 1964) as per methods outlined in Blundell and Johnson (1976). The heavy-atom sites found were refined by least-squares methods (Dickerson et al., 1961) as
Table 3.4  Statistics for NAT1 and the heavy-atom derivatives used to solve the PII structure at 2.7 Å. For each data set is listed the number of unique reflections collected and their multiplicity, along with the resolution to which the data was complete. Also given are the $R_{\text{merge}}$ values and the difference of each derivative to NAT1.

<table>
<thead>
<tr>
<th>Data set</th>
<th>Number of observations</th>
<th>Number of unique reflections</th>
<th>Resolution (Å)</th>
<th>Twin fraction</th>
<th>$R_{\text{merge}}^a$ (%)</th>
<th>Difference $b$ (%) from NAT1</th>
</tr>
</thead>
<tbody>
<tr>
<td>NAT1</td>
<td>39 582</td>
<td>6 578</td>
<td>2.2</td>
<td>0.06</td>
<td>2.50</td>
<td>-</td>
</tr>
<tr>
<td>PII11</td>
<td>51 518</td>
<td>5 563</td>
<td>2.8</td>
<td>0.09</td>
<td>10.31</td>
<td>21.0</td>
</tr>
<tr>
<td>PII23</td>
<td>68 057</td>
<td>6 774</td>
<td>3.0</td>
<td>0.04</td>
<td>7.68</td>
<td>15.3</td>
</tr>
<tr>
<td>PII25</td>
<td>52 772</td>
<td>5 572</td>
<td>4.5</td>
<td>0.15</td>
<td>7.21</td>
<td>19.8</td>
</tr>
<tr>
<td>PII79</td>
<td>43 331</td>
<td>5 260</td>
<td>2.8</td>
<td>0.08</td>
<td>8.04</td>
<td>19.7</td>
</tr>
<tr>
<td>PII89</td>
<td>25 493</td>
<td>3 639</td>
<td>2.8</td>
<td>0.05</td>
<td>8.11</td>
<td>17.4</td>
</tr>
<tr>
<td>PII90</td>
<td>44 011</td>
<td>4 233</td>
<td>2.3</td>
<td>0.05</td>
<td>5.50</td>
<td>14.5</td>
</tr>
</tbody>
</table>

$a \quad R_{\text{merge}} = \sum \frac{\Sigma |I(h,i)| - <I(h)>|}{\Sigma I(i)} \quad I(h,i) = \text{observed intensity of the reflection } h. I(i) = i^{th} \text{ integrated intensity of a reflection.}$

$b \quad \text{Difference} = \frac{\Sigma |I_d - I_n|}{\Sigma (I_d + I_n)/2}, \text{ where } I_d \text{ is the observed intensity from the derivative and } I_n \text{ the observed intensity from the native summed over all common reflections.}$
implemented by the phase refinement subroutine in the program PROTEIN.

Native P11 crystallises in the P63 space group. This space group is characterised by the following six symmetry operators:

\[
\begin{align*}
\mathbf{a} & : (x, y, z) \\
\mathbf{b} & : (y, x-y, z) \\
\mathbf{c} & : (x+y, x, z) \\
\mathbf{d} & : (x, y, z+rac{1}{2}) \\
\mathbf{e} & : (y, x+y, z+rac{1}{2}) \\
\mathbf{f} & : (x-y, x, z+rac{1}{2})
\end{align*}
\]

There are therefore ten inter-site vectors relating a particular site \((x, y, z)\) to its symmetry equivalents. These ten vectors are the five listed below, and their inverses.

\[
\begin{align*}
\mathbf{g} & : (-y-x, x-2y, 0) \\
\mathbf{h} & : (-2x+y, -x-y, 0) \\
\mathbf{i} & : (-2x, -2y, \frac{1}{2}) \\
\mathbf{j} & : (y-x, -x, \frac{1}{2}) \\
\mathbf{k} & : (-y, x-y, \frac{1}{2})
\end{align*}
\]

It can be seen that these inter-site vectors result in peaks on the \(w = 0\) and \(w = 0.5\) planes in Patterson space and thus define the two Harker planes for P63 (Harker, 1936). Typical Harker sections from the difference Patterson map are shown in Figure 3.8.

It can be seen from Figure 3.8 that the a.s.u. contains one unique peak of relative magnitude 1 on the \(w = 0\) section (Figure 3.8A), while in the \(w = 0.5\) section there are two (Figure 3.8B), one of relative magnitude 1 and the other of relative magnitude 0.5. The reason that one heavy-atom site gives rise to this pattern of peaks is as follows: In P63 Patterson space the interatomic vectors \(\mathbf{g}\) and \(\mathbf{h}\) are equivalent as they are related by operator \(\mathbf{b}\) plus inversion. Likewise, vectors \(\mathbf{i}\) and \(\mathbf{j}\) are
Figure 3.8  Harker sections for PI111. (A) shows the $w = 0$ Harker section with the arrow indicating a $7\sigma$ peak from the Hg-Hg interatomic vector. (B) shows the $w = 0.5$ Harker section which contains two Hg-Hg vector peaks; a $7\sigma$ (1) and a $3\sigma$ peak (2). Both A and B have the boundaries of the asymmetric subunit indicated by heavy lines.
related by symmetry operator $c$ and inversion. The result of this is that the pair of equivalent vectors $g$ and $h$ contribute to one peak, as does the pair of vectors $i$ and $j$. Vector $k$ has no such equivalent and therefore gives rise to a peak half the magnitude of the other two peaks.

A summary of the refined parameters for the six heavy-atom derivatives used for phasing the NAT1 data is shown in Table 3.5. There were four individual sites that exhibited metal binding within the $P_{11}$ crystal. The first was a PCMB site which had high occupancy and relatively low B factor. The second site was an Ir site which was close to the third Pt site. The fourth site was also a Pt site but arose from $K_2[Pt(CN)_4]$.

The first successful heavy-atom derivative of $P_{11}$ was PII11 (PCMB). Native $P_{11}$ crystallised in the $P6_3$ space group which has no symmetry elements perpendicular to $z$. When the binding site of the Hg atom was located via difference Pattersons, as the first derivative its $z$-coordinate was set to 0.0. All subsequent derivatives were given values for $z$ relative to the Hg site. The $x/y$-coordinates of a second heavy-atom–Ir in PII23—were determined using difference Pattersons but determining the location of the $z$-coordinate relative to Hg, was difficult. Initially difference Fouriers with phases calculated from a single derivative did not give clear, unambiguous values for $\Delta z$ and subsequently double soak derivatives—such as PII25 and PII79 (Table 3.5)—were used successfully. The correct enantiomorph was assigned by calculation of the figure of merit (f.o.m.) values for both possibilities. The phases resulting in the best f.o.m. was assigned the correct enantiomorph. Periodically, throughout the course of additional phase collection and refinement, these calculations were
Table 3.5 Final heavy-atom derivative parameters. Listed are the number of sites identified in each derivative, their positions in fractional coordinates and the corresponding relative occupancy of each site as calculated in PROTEIN. Also listed are the isotropic temperature factors (\(B\)) for each site.

<table>
<thead>
<tr>
<th>Derivative</th>
<th>Heavy-atom compound</th>
<th>Number of sites</th>
<th>Occupancy</th>
<th>(B) ((\text{Å}^2))</th>
<th>(x)</th>
<th>(y)</th>
<th>(z)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PII11</td>
<td>PCMBS</td>
<td>1</td>
<td>29.1523</td>
<td>17.869</td>
<td>0.185</td>
<td>0.303</td>
<td>0.000</td>
</tr>
<tr>
<td>PII23</td>
<td>Na(_2)rCl(_6)</td>
<td>1</td>
<td>26.3087</td>
<td>33.464</td>
<td>0.501</td>
<td>0.464</td>
<td>0.122</td>
</tr>
<tr>
<td>PII25</td>
<td>PCMBS</td>
<td>2</td>
<td>22.0555</td>
<td>15.449</td>
<td>0.186</td>
<td>0.305</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>+ Na(_2)rCl(_6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PII79</td>
<td>PCMBS</td>
<td>2</td>
<td>28.1440</td>
<td>13.331</td>
<td>0.187</td>
<td>0.305</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>+ K(_2)[Pt(CN)(_4)]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PII89</td>
<td>K(_2)[Pt(CN)(_6)]</td>
<td>2</td>
<td>19.9377</td>
<td>28.856</td>
<td>0.224</td>
<td>0.395</td>
<td>0.358</td>
</tr>
<tr>
<td>PII90</td>
<td>K(_2)[Pt(CN)(_6)]</td>
<td>2</td>
<td>15.4071</td>
<td>28.700</td>
<td>0.225</td>
<td>0.396</td>
<td>0.357</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>11.2717</td>
<td>44.057</td>
<td>0.426</td>
<td>0.497</td>
<td>0.122</td>
</tr>
</tbody>
</table>
re-evaluated to confirm our original results. Eventually these phases resulted in electron density maps that were sufficiently detailed to find right-handed helices and thus confirm our enantiomorphic assignment.

When K$_2$[Pt(CN)$_4$] was used as a derivative two Pt sites were found. The first site was novel, while the second was the same site found for PII79. The new site had greater occupancy and lower B values than did the previously identified site. These facts indicate that Pt had a preference to bind to the new site and the existence of Pt in another site in double soaks represented the competitive nature of the Hg binding.

Heavy-atom refinement was initially carried out independently for each derivative. After refinement to convergence alone, refinement was repeated with all six derivatives included. This two stage approach to refinement limited the likelihood of an incorrect initial refinement corrupting the phase and position information for other derivatives. A resolution cut off of 2.7 Å was applied to this data due to the drop in $F_{h/E}$ values for data beyond 2.7 Å. Figure 3.9 shows a plot of $F_{h/E}$ with respect to resolution for the six derivatives used to phase PII (Table 3.3). It can be seen that each of the six derivatives contributed to phasing by yielding information. For example PII25 has good low resolution $F_{h/E}$ values which diminish rapidly toward high resolution, while PII89 yields good phase information at both low and high resolution.

Sufficient phase information for the majority of these derivatives was to 2.5 Å, but in two cases—PII23 and PII25—was limited to much lower resolution. To conserve the usefulness of the phase information over
Figure 3.9 Plot of $F_h/E$ with respect to resolution for the six heavy-atom derivatives used provide phasing for P11. Resolution is plotted as its inverse ($\AA^{-1}$). $F_h$ is the root mean square (rms) heavy-atom structure factor amplitude and $E$ the rms lack of closure error. The 2.7 Å cut off is indicated with a vertical back line.
the six derivatives, the resolution limit to 2.7 Å was applied. Including data between 10.0 Å and 2.2 Å the overall figure of merit was 0.69, while limiting both native and derivative data to 2.7 Å yielded an overall figure of merit of 0.79. The overall heavy-atom derivative statistics are shown in Table 3.6. While the information from Table 3.3 and Figure 3.9 give insight into the values shown in Table 3.6, there are some further comments that should be made.

The low value of $Fh/E$ for PII25 data at 2.7 Å arises from the fact that the crystal only diffracted out to 4.5 Å. As mentioned previously, this derivative did contribute phasing power to the low resolution terms and as such was a necessary inclusion. This data also had a 0.15 twin fraction which also resulted in loss of data. For derivatives PII79 and PII89 lower degrees of oscillation data were collected. Both derivatives had 80° of data collected before loss of useful diffraction, while the other four derivatives had at least 120° of data collected. With data limited to between 10.0 Å and 2.7 Å the NAT1 dataset was left 97.7 % complete.

There were a number of problems that hampered attempts to determine phase information for the native PII data. The first and most difficult was the twinning of the crystals. The presence of this effect resulted in the loss of data through the detwinning process since only reflections for which twin mates were observed could be used in subsequent calculations. The twin correction probably caused an increase in the error of the anomalous differences. We could have applied local scaling to reduce this error, but there was insufficient time to write the appropriate software.
Table 3.6 Heavy-atom derivative refinement statistics. Values for $F_h/E$ were determined via phase combination using reflections between 20.0 Å and 2.7 Å, where $F_h$ is the root mean square (rms) heavy-atom structure factor amplitude and $E$ the rms lack of closure error. The Cullis R-factors (Cullis et al., 1961), $R_c$, have also been calculated using data between 20.0 Å and 2.7 Å.

<table>
<thead>
<tr>
<th>Derivative</th>
<th>Resolution (Å)</th>
<th>mean $F_h/E$</th>
<th>$R_c^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>PII11</td>
<td>2.7</td>
<td>1.97</td>
<td>0.65</td>
</tr>
<tr>
<td>PII23</td>
<td>2.7</td>
<td>1.86</td>
<td>0.66</td>
</tr>
<tr>
<td>PII25</td>
<td>2.7</td>
<td>1.93</td>
<td>0.45</td>
</tr>
<tr>
<td>PII79</td>
<td>2.7</td>
<td>1.98</td>
<td>0.57</td>
</tr>
<tr>
<td>PII89</td>
<td>2.7</td>
<td>1.84</td>
<td>0.51</td>
</tr>
<tr>
<td>PII90</td>
<td>2.7</td>
<td>1.93</td>
<td>0.49</td>
</tr>
</tbody>
</table>

$R_c = \frac{\Sigma |F_{PHC} - F_{PH}|}{\Sigma |F_{PH} - F_P|}$, which is the summation over all centric reflections. $F_{PHC}$ and $F_{PH}$ are the calculated and observed derivative structure factors, while $F_P$ is the native structure factor amplitude.
The heavy-atom derivatives described above yielded sufficient phasing power for the calculation of electron density maps from the NAT1 dataset, which were of sufficient quality to allow model building. The MIR phases from these derivatives were used until the model phases were of greater quality. This process is the subject of Chapter 4.
Chapter 4. Structure determination of P\textsubscript{II} at 2.7 Å

4.1 Introduction
With the phases determined for the P\textsubscript{II} native structure factors it was now possible to calculate electron density maps. In the early stages of construction of the P\textsubscript{II} model only four heavy-atom derivatives were used: PII11, PII23, PII25 and PII79. In the course of model building the final two derivatives, PII89 and PII90, were added. The initial maps discussed below were calculated using the first four derivatives while the last two derivatives were used in conjunction with phase combination from a preliminary model. The construction of an initial model into these maps and the subsequent model refinement is the subject of this chapter. The P\textsubscript{II} structure was determined to a crystallographic R-factor of 24.5 % using data between 10.0 Å and 2.7 Å (Cheah et al., 1994). This model would later form the basis for the further refinement of P\textsubscript{II} at 1.9 Å (see Chapter 5).

4.2 Materials and Methods
The program PROTEIN (Steigemann, 1992) was used for phase combination (Remington et al., 1982) of MIR phases (Blow & Crick, 1959) with model phases. These combined phases were then used to calculate electron density maps which allowed the model to be further improved.

Initial MIR maps from PROTEIN were calculated for a 50 Å × 50 Å × 50 Å volume. These maps were then sectioned in Z and printed. The printed maps were placed on 5 mm thick Perspex\textsuperscript{TM} sheets, where each layer represented a 2 Å section of the map in Z. The individual layers were then stacked to form a three dimensional representation of the electron density of the crystallised protein.
Except for the initial "stack-maps", phase combined maps were displayed using the Alberta/Caltech version of the TOM/FRODO program (Jones, 1985) running on either a Silicon Graphics IRIS 4D/310VGX or Silicon Graphics Indy workstation. This program was also used to build models to the displayed maps.

The solvent flattening (Wang, 1985) performed involved determining a molecular envelope based on the solvent content of the crystal and the intensity averaged over a 8 Å sphere on each point of a grid constructed within the electron density maps. All points outside the molecular boundary (solvent region) are reset to a constant low density. Any points within the molecular envelope which exhibit electron density lower than that of the solvent was reset above this level. Iteration of this procedure resulted in improved maps.

Crystallographic refinement of the model coordinates were carried out using version 3.1 of the program X-PLOR (Brünger, 1992) as implemented on a Fujitsu VP2200 supercomputer. Access to these facilities were made possible by a grant of time from the ANU Supercomputing Facility.

Models were checked for deviations from known protein geometry using the program PROCHECK (Laskowski et al., 1993). This program was also used to produce Ramachandran plots (Ramachandran & Sasisekharan, 1968).

Secondary structure was assigned using a modified form of the program DSSP (Kabsch & Sander, 1983).
Schematic diagrams of the PII structure were generated by the programs MOLSCRIPT (Kraulis, 1991) and MIDAS (Ferrin et al., 1988).

4.3 Oligomeric and low resolution structure

Using MIR phases from the heavy-atom derivatives PII11, PII23, PII25 and PII79 (Table 3.3) initial electron density maps were calculated. These maps were calculated at 2.7 Å resolution. “Stack-maps” from this data were not of sufficient quality to allow confident identification of a continuous main chain trace and thus no initial model could be determined.

An MIR map was calculated using 5 Å data and was plotted as a stack-map. Figure 4.1 shows this map with the P63 space group operators labelled. It is easy to see that the PII protein is forming around a crystallographic three fold axis to form a homotrimer. This result was in conflict with our earlier light scattering experiments which suggested PII was a tetramer in solution (Dr Eong Cheah, unpublished). It was at this point sedimentation equilibrium experiments were carried out by Dr Peter Jeffrey, as described in Chapter 2. As already mentioned these results indicated that PII did in fact exist as a trimer in solution. In addition to showing the trimeric nature of crystalline PII, the map calculated using 5 Å data showed some additional broad structural detail, but was still not of sufficient quality to trace backbone atoms continuously.

4.4 Solvent flattening

In an attempt to increase the definition of the MIR maps the process of solvent flattening (Wang, 1985) was employed by Dr Paul Carr. After
Figure 4.1 An electron density map calculated using 5 Å data from the P11 native data set NAT1. Labelled are the crystallographic symmetry operators for the P63 space group: 6-fold screw axis, 3-fold axis, & 2-fold screw axis.
solvent flattening, the maps had a much lower level of noise and showed continuous backbone density (Figure 4.2). These maps also showed some side-chain density for residues in the central regions of the protein. An unfortunate side effect of solvent flattening is that the side-chain density on the outer surface of the protein may be poorly defined. A number of rounds of solvent flattening were performed and it was the final electron density map resulting from this treatment that were used to trace the backbone atoms of this protein.

4.5 Ca chain trace and side-chain insertion

As mentioned above, the solvent flattened maps allowed the backbone atoms to be fitted into density to obtain an initial trace of the main chain. At this stage maps were calculated and plotted out as stack-maps. Fitting of Ca atoms to these maps comprised the initial stages of model building. Figure 4.3 shows sections of the MIR map used to build an initial model for P11. As can be seen (Figure 4.3) the MIR maps were very noisy at this stage and the outer regions of the molecule were poorly defined, thus limiting the number of Ca atoms that could be fitted to the maps. The Ca model coordinates and solvent flattened MIR maps were displayed with TOM/FRODO and a polyalanine model was built to fit the electron density. This model contained 75% of the main chain atoms with no side chain atoms. The areas of the MIR map that were not sufficiently continuous to fit main chain atoms included the later defined flexible 3-4 loop and the N- and C-termini.

Coordinates from this preliminary model were used for combination of model and MIR phases (Remington et al., 1982) within PROTEIN. The resulting phase combined map was sufficiently improved such that 95% of the main chain atoms could be fitted. Model phases were again
Figure 4.2 An MIR stack map after one round of solvent flattening. This map is viewed down the Z axis, which is parallel to the crystallographic three-fold axis of PII. One whole PII trimer can be seen within the boundaries of the map.
Figure 4.3 An MIR stack map used for the initial model chain tracing for the $P_{II}$ molecule. This map is viewed down the three-fold axis (parallel to Z) of the molecule and illustrates one trimer.
combined with MIR phases but the maps remained poorly defined for the exterior sections of the mapped molecule. It was at this stage that the continuing search for derivatives with better phasing power yielded another two derivatives. The phases from the new PII89 and PII90 derivatives (Table 3.3) were combined with the previous MIR phases, to yield an overall figure of merit of 0.79.

The combined MIR and model phases were improved to the point where some side-chain atoms could be seen in maps. Despite many poorly defined regions in the map (including the later defined 3-4 loop) part of the PII sequence was fitted to the density. Side-chains were then built into our model where the density was sufficient and these new models used in further phase combination. This process was continued until all of the side-chain atoms had been placed. At this point the difficulty in model building became apparent. The quality of the map was limited by a number of features of this molecule. These limiting features included the flexibility of the 3-4 loops, the nature of the β-sheet interaction between monomers (see later in section 4.6.2) and the close proximity of the N- and C-termini. These led to confusion in defining the start and the end of this molecule. The final model was consistent with the predicted amino acid sequence (Vasudevan et al., 1991). Figure 4.4 shows a representative region of the 2F₀-Fₐ density and the final model.

The model built into the 2F₀-Fₛ map density mentioned above was then refined using X-PLOR (Brünger, 1992). This refinement comprised of positional refinement followed by simulated annealing. Simulated annealing was from an initial T of 4000 K to a final T of 300 K in 25 K steps at 0.5 fs/step. The resultant model was then treated with
Figure 4.4 2F_o-F_c maps calculated using data between 10.0 Å and 2.7 Å and contoured at 1.0 σ. This figure illustrates the definition of this map for both main chain and side-chain atoms. This figure was generated from the TOM/FRODO program.
positional refinement followed by unrestrained B-factor refinement. The final P\textsubscript{II} model at 2.7 Å had an R-factor of 24.5 % for data between 10.0 Å and 2.7 Å. On bond lengths this model had an rms deviation from ideality of 0.01 Å and on bond angles an rms deviation of 1.492°. A Luzzati plot (Luzzati, 1952) suggested that this model had a mean coordinate error of 0.35 Å. As an additional measure of reliability a Ramachandran plot (Ramachandran & Sasisekharan, 1968), was calculated for the 2.7 Å refined P\textsubscript{II} structure (Figure 4.5). By comparison of φ and ψ angles from this model with an analysis based on 118 structures (Laskowski \textit{et al.}, 1993) it was found that 90 of the 112 residues resided in the most favoured regions, 3 in additionally allowed and 1 residue in generously allowed region. There were three residues that were in disallowed regions. These residues were Tyr46, Ala49 and Glu50. All proline and glycine residues were in their allowed regions.
Figure 4.5 A Ramachandran plot for the P_{11} structure at 2.7 Å. Labelled are non-terminal, non-glycine residues that are outside the most favoured regions (A, B, & L). Also labelled are the additionally allowed regions (a, b, l, & p) and the generously allowed regions (~a, ~b, ~l, & ~p).
4.6 P_{II} structure at 2.7 Å

4.6.1 Monomeric structure

A ribbon diagram of the P_{II} monomer is shown in Figure 4.6. The monomer is comprised of eight strands and two helices, which combine to form a structure that is approximately 36 Å × 30 Å × 60 Å. The assignment of secondary structural elements was made by the DSSP program (Kabsch & Sander, 1983)(Table 4.1). A stereo view of a Cα trace for the P_{II} monomer is shown in Figure 4.7.

In referring to the P_{II} molecule the notation β for strand and α for helix will be used. There are four main features in the monomeric structure that are especially noteworthy. The first is the central region which has an anti-parallel β-sheet and two anti-parallel α-helices. The central β-sheet is composed of four anti-parallel strands, namely strands: β2, β5, β1 and β6 (Figure 4.6). This sheet is twisted with a twist of 44° between strand β2 and strand β6, determined using the method of Ploegman et al. (1978). Behind this sheet are located the two helices which form the double β–α–β motif (Pastore et al., 1992) in P_{II}. The connectivity of this fold in P_{II} is β1–αA–β2, β5–αB–β6. The double β–α–β motif is a stable fold seen in a number of proteins with diverse functions (see Chapter 5).

An important finding with this monomeric structure was the existence of a pseudo two-fold axis perpendicular to the β-sheet, between β5 and β1 (Figure 4.6). When this two-fold axis is applied 78.6% of Cα atoms are aligned with other original Cα atomic positions with a rms deviation of 2.13 Å. Strands β1 and β6 overlay with good agreement onto strands β2 and β5, respectively as does helix A onto B. This local
Figure 4.6 A MOLSCRIPT (Kraulis, 1991) ribbon diagram of the PII monomer. β-sheets are labelled with numbers, and the helices are labelled with capital letters. Tyr51, the site of uridylylation is shown at the top of the figure. The names of the three “random coil” regions are also shown.
Table 4.1  Secondary structural elements for the P$_{II}$ structure at 2.7 Å as determined by the program DSSP.

<table>
<thead>
<tr>
<th>Secondary structural element</th>
<th>Residues range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strand 1</td>
<td>4 - 9</td>
</tr>
<tr>
<td>Helix A</td>
<td>13 - 23</td>
</tr>
<tr>
<td>Strand 2</td>
<td>29 - 33</td>
</tr>
<tr>
<td>Strand 3</td>
<td>35 - 40</td>
</tr>
<tr>
<td>Strand 4</td>
<td>55 - 56</td>
</tr>
<tr>
<td>Strand 5</td>
<td>59 - 65</td>
</tr>
<tr>
<td>Helix B</td>
<td>71 - 80</td>
</tr>
<tr>
<td>Strand 6</td>
<td>90 - 95</td>
</tr>
<tr>
<td>Strand 7</td>
<td>99 - 102</td>
</tr>
<tr>
<td>Strand 8</td>
<td>105 - 109</td>
</tr>
</tbody>
</table>
Figure 4.7 A stereo MOLSCRIPT (Kraulis, 1991) diagram of the Cα trace of the PⅡ monomer calculated at 2.7 Å. Labelled are the N- and C- termini, Tyr51 and every fifth residue.
pseudo-symmetry encompasses $\beta 1, \beta 2, \beta 5, \beta 6, \alpha A,$ and $\alpha B$ but does not extend to the other regions within $P_{II}$. Twinning most commonly arises when two molecules are in the a.s.u. and are related by a non-crystallographic symmetry axis. In the case of $P_{II}$, however, the existence of an approximate two-fold axis within the monomer seems to be the cause of the observed merohedric twinning.

The second feature is the 3-4 loop which connects $\beta 3$ to $\beta 4$ (Figure 4.6). This loop contains Tyr51, the residue that is modified as a part of signal transduction within the cell. This 3-4 loop is positioned away from the remainder of the protein which appears spatially conserved. The ramifications of this positioning will be discussed later in light of the trimeric structure. With the exception of strands $\beta 3$ and $\beta 4$ there is no stabilisation between monomers of a particular trimer arising from the 3-4 loop. It is not surprising therefore that the highest B-factors are found within the 3-4 loop. Although not well determined at this stage, the B-factors for the monomeric structure of $P_{II}$ determined at 2.7 Å indicated that the 3-4 loop was the most mobile section of the structure.

The third and fourth features of the $P_{II}$ monomer are the B-6 and 7-8 loops respectively (Figure 4.6). These two loops have an important role to play within the trimer in that they help to form elliptical clefts that are repeated around the three-fold axis. The roles that these loops play in cleft formation is discussed below.

4.6.2 Trimeric structure

$P_{II}$ is a trimer both in solution and in the crystalline state. The $P_{II}$ trimer is found centred about the crystallographic three-fold axis as
can be seen in Figure 4.8. The P\textsubscript{II} trimer is arranged such that the central $\beta$-sheet from each monomer packs around the three-fold axis leaving the $\alpha$-helices exposed on the outer surfaces (Figure 4.8). The arrangement of the strands from each monomer brings about sheet extension around the triad (Figure 4.9). The central sheet from any monomer is extended by strands $\beta7$ and $\beta8$ from a second monomer and by strands $\beta3$ and $\beta4$ from the third. The end result is that within the trimer twenty four strands interact to form three $\beta$-sheets each containing eight strands (Figure 4.9). The contribution of strands from each monomer to each of the $\beta$-sheets is an uncommon feature within trimers, as is the tricuspid arrangement of the central cavity (Figure 4.8).

There are a couple of points to note about this trimeric structure. The first is the nature of the central sheets from each monomer. As already mentioned each sheet has an internal twist angle of 44°, but in the trimer the extended eight-stranded sheets have a twist angle of 150°. It is the ability of the central sheets to twist that allows P\textsubscript{II} to form a trimer in this tricuspid arrangement. The extension of these sheets probably also confers the stability of this structure.

The second point to mention would be the existence of elliptical clefts between the 3-4 and B-6 loops from one monomer and the 7-8 loop from another monomer (Figures 4.8 and 4.10). These clefts have one side composed of uncharged residues while the other side is lined with basic residues. These basic residues include Arg38, Lys40 (from the 3-4 loop), Lys85, Lys90 (from the B-6 loop) from one monomer and Arg101 and Arg103 (from the 7-8 loop) from a second monomer. Electrostatic calculations were carried out on the P\textsubscript{II} trimer using the Biosym
Figure 4.8  A MOLSCRIPT schematic diagram of the P_{II} trimer calculated at 2.7 Å. The trimer is viewed down its crystallographic three-fold axis. Each monomer is coloured differently.
Figure 4.9  A schematic connectivity diagram for the P₁ trimer with the secondary structural units defined as per Figure 4.6. This diagram illustrates the formation of three eight-stranded β-sheets about the three-fold axis, with the α-helices residing around the outside of the trimer.
Figure 4.10 A) A MOLSCRIPT schematic diagram of the PII trimer calculated at 2.7 Å. The trimer is viewed perpendicular to its crystallographic three-fold axis. Each monomer is coloured differently. This diagram shows clearly the positioning of Tyr51 residues from each monomer well away from the main body of the trimer. B) A MIDAS schematic Cα diagram of the PII trimer calculated at 2.7 Å and in the same orientation as A). The Van der Waals surfaces of the residues surrounding the cleft mentioned in text are shown.
DELPHI\textsuperscript{a} package of programs. An electrostatic potential was generated in a grid of 1 Å intervals, using a dielectric of 2.0 for protein and 80.0 for solvent. The value of solvent ionic strength used was 0.145 M. Full charges were assigned to the amino acids lysine, arginine, aspartic acid and glutamic acid, while a half charge was assigned to histidine residues. The results of these calculations indicates that the cleft regions described above are surrounded by large regions of positive charge (Figure 4.11). These charged cleft regions of P\textsubscript{II} seem likely to participate in interactions with the various effectors known to be involved in the function of this protein. These results will be discussed in more detail later (see Chapter 8).

A view perpendicular to the three-fold of the P\textsubscript{II} trimer (Figure 4.10) shows a very obvious and potentially very important structural feature; the positioning of the Tyr51 residue by the 3-4 loop. The majority of the P\textsubscript{II} trimer exists in a compact structure that is approximately 50 Å in diameter and 30 Å high. In contrast to this the Tyr51 residue is positioned 13 Å away from the main body of the trimer on a mobile loop. This loop is stabilised to some extent by crystal contacts with other trimers as described below.

When the secondary structural elements are juxtaposed with the sequence alignments from Chapter 1, some trends can be seen (Figure 4.12). It seems apparent that strands are more highly conserved across the nine species compared than are the helices. The most highly conserved region is the 3-4 loop which contains Tyr51.

Figure 4.11 The electrostatic potential map of the PII trimer at 2.7 Å. This map has been contoured at ± 2 kT e⁻¹; positive regions have been shown in red and negative regions in blue. All clefts in the trimer are situated beneath the large positive fields.
Figure 4.12 Alignment of P11 sequences from nine species. Secondary structural elements are shown in red above the appropriate sequences. Strands are shown as arrows and helices as rectangles. Conserved residues are highlighted in green. Sequences were aligned using the BESTFIT program from the GCG package (Devereux et al., 1984).


Excluding P\textsubscript{II} from Bacillus subtilis, of the fourteen residues that comprise this loop eight are fully conserved and Tyr51 is conserved throughout. The reasons for excluding P\textsubscript{II} from B. subtilis will be discussed in Chapter 8, as will further discussion of the patterns of sequence conservation in P\textsubscript{II}.

A MIDAS (Ferrin et al., 1988) wire-frame diagram of the P\textsubscript{II} structure determined at 2.7 Å with the residues conserved across eight species is shown in Figure 4.13. There are a number of biologically relevant features and possible conclusions to be drawn from Figure 4.12 and Figure 4.13.

Firstly it can be seen that the P\textsubscript{II} sequences are well conserved, with an overall identity of 51%. This rises to 57% for the 3-4 loop. Inspection of the trimer reveals a number of structural features which seem likely to have biological importance. The first is the high degree of sequence conservation in the flat upper surface of the trimer (Figure 4.13). A second area of possible importance is the cleft regions (Figure 4.13 and Figure 4.10), which are also highly conserved. Another feature to be noted is the 3-4 loop regions that contain the Tyr51 residues. Despite the highly conserved nature of these loops, the section of the loop which extends away from the main body of the trimer does not maintain the high level of identity. Each of these regions seem likely to be involved in interactions with other proteins or ligands to bring about the signal transducing function of this protein. The possible consequences of these structural features will be discussed in greater detail in Chapter 8.
Figure 4.13  MIDAS (Ferrin et al., 1988) stick diagrams of the \( \text{P}_{11} \) monomer and orthogonal views of the trimer determined at 2.7 Å. Coloured in blue are the residues conserved across the eight known bacterial \( \text{P}_{11} \)-like proteins. Coloured in black are the Tyr51 residues for each subunit. These species are described in Figure 4.12 and the sub-selection of the eight species used described in text.
4.6.3 Crystal packing forces

Crystallisation contacts have an important role in stabilising the 3-4 loop that contains Tyr51. It is the symmetry mate of one P_{II} trimer that interacts with the 3-4 loop and helps to stabilise it via lattice contacts. These crystal contacts support the 3-4 loop at residues Arg47, Tyr51 and Val53.

The idea that this 3-4 loop was very mobile in the crystalline state raised questions about how this loop would exist in solution. Flexible loops, such as the 3-4 loop, most often exist close to the main body of the protein in order to evade proteolytic attack and gain stability via hydrogen bonding. When needed for function, these loops are often positioned by the binding of effector ligands. In an effort to gain further insight into how this loop existed in solution an NMR investigation was undertaken by Dr Max Keniry at the Research School of Chemistry, ANU. This study could not find any evidence for conformational mobility of the 3-4 loop of P_{II} (Dr Max Keniry, unpublished). This suggests that in solution the 3-4 loop may adopt a conformation such that the loop is stabilised by the rest of the trimer. How this conformational change—as indicated by crystal packing—possibly relates to the function of the P_{II} protein will be discussed further in Chapter 8.
Chapter 5. Structural description of PII

5.1 Introduction
With the structure of E. coli PII solved to 2.7 Å there remained the need for a more detailed description of this protein which would involve solution of this structure at higher resolution. To achieve this a higher resolution dataset was needed. The 2.7 Å structure could then be used as the initial model for refinement. This chapter briefly describes how the PII structure at 2.7 Å was refined to 1.9 Å and then using this refined model goes on to make structural comparisons with other proteins that contain an interlocking double β-α-β motif.

5.2 Materials and methods
Refinement of the final PII coordinates were carried out using the X-PLOR program version 3.1 (Brünger, 1992) as implemented on a Fujitsu VP2200 supercomputer. Access to these facilities were made possible by a grant of time from the ANU Supercomputing Facility.

Secondary structure was assigned using a modified form of the program DSSP (Kabsch & Sander, 1983).

The diagrams in this chapter representing protein structures were generated by the program MOLSCRIPT (Kraulis, 1991).

The PII coordinates used for comparison with other proteins are those that were determined for 1.9 Å data. These coordinates were as submitted to the Brookhaven Protein Data Bank (PDB) minus water molecules. The comparison of the PII structure with the other proteins containing a double β-α-β motif was carried out on a topological level.
from information provided in the appropriate references. All coordinates used for comparisons were retrieved from the PDB.

DNA sequences for the PII-like proteins listed were retrieved from the GenBank/EMBL database using the GCG programs (Devereux, Haeberli & Smithies, 1984). The sequences used were predicted from nucleotide sequences using the GCG package. Calculations of percent identity and similarity were also performed using the GCG package of programs.

5.3 Atomic refinement

Dr Paul Carr collected a new PII native dataset to 1.9 Å and then used the structure of PII at 2.7 Å as the initial starting model for refinement. Rounds of positional refinement using the Powell conjugate gradient method (Powell, 1977) and then cycles of individual restrained isotropic B-factor refinement were carried out. As the quality of the maps improved waters were added to the model. The final number of bound waters fitted to the 2F₀-Fc and F₀-Fc maps generated with this model was 312. The method used is described in detail in Carr et al. (1996), but is outlined briefly here.

The final PII model had a crystallographic R-factor of 13.2 % for 3σ data between 10.0 Å and 1.9 Å. This model had rms deviations from ideal stereochemistry for bond lengths and bond angles of 0.017 Å and 1.98 Å, respectively. The Luzzati plot (Luzzati, 1952) for this structure indicated a mean coordinate error of 0.15 Å (Figure 5.1). This model was checked for stereochemical correctness using the PROCHECK program (Laskowski et al., 1993) and was found to be within acceptable
Figure 5.1 A Luzzati plot for the \( \text{P}_\text{II} \) model calculated at 1.9 Å resolution. Shown are the theoretical distributions (---) for mean coordinate errors of 0.1 Å, 0.15 Å and 0.2 Å as described by Luzzati (Luzzati, 1952).
limits of the values determined for 118 protein structures solved to 2.0 Å or better. The values for acceptability were according to Morris et al. (1992).

5.4 Features of \( \text{P}_{\text{II}} \)

5.4.1 Monomeric structure

The structure of \( \text{P}_{\text{II}} \) refined to 1.9 Å resolution does not vary considerably from the structure determined at 2.7 Å. The primary difference between these structures lies in the definition of the secondary structure. Figure 5.2 shows a MOLSCRIPT (Kraulis, 1991) diagram comparing both the 2.7 Å and the 1.9 Å structures for \( \text{P}_{\text{II}} \). The redefinition of secondary structure indicates the 2-3 strands now forming strand 2 while the 4-5 strands become strand 4. The sequence composition of the new secondary structural assignments for \( \text{P}_{\text{II}} \) are listed in Table 5.1.

Strand 2 has an interesting characteristic. There is a noticeable bend in this strand at Glu32 (Figure 5.3). The angle subtended between residues 27-32 and 32-36 was calculated to be 58.2° using the method of Ploegman et al. (1978). While the phi and psi angles for this residue (-91°,127° respect.) are still in the most favoured region of the Ramachandran plot, these values are the lowest within strand 2. A study by Daffner et al. (1994) indicated that 85% of the 247 non-homologous structures studied from the Protein Database at Brookhaven (Bernstein et al., 1977) exhibited a bend similar to the one found in \( \text{P}_{\text{II}} \).

5.4.2 Trimeric structure

The bend in strand 2 has an important role in regards to formation of the trimer. The gradual curve of strand 2 allows the extension of the
Figure 5.2 A MOLSCRIPT ribbon diagram showing a comparison between the $P_{II}$ structures determined at A) 1.9 Å and B) 2.7 Å resolution. Indicated are the redefinition's of secondary structural elements with the increased resolution.
Table 5.1  Composition of the secondary structural units of P$_{11}$ determined at 1.9 Å.

<table>
<thead>
<tr>
<th>Secondary structural element</th>
<th>Residues involved</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strand 1</td>
<td>2 - 8</td>
</tr>
<tr>
<td>Helix A</td>
<td>10 - 23</td>
</tr>
<tr>
<td>Strand 2</td>
<td>27 -36</td>
</tr>
<tr>
<td>Strand 3</td>
<td>56 - 65</td>
</tr>
<tr>
<td>Helix B</td>
<td>67 - 81</td>
</tr>
<tr>
<td>Strand 4</td>
<td>90 - 95</td>
</tr>
<tr>
<td>Strand 5</td>
<td>98 - 101</td>
</tr>
<tr>
<td>Strand 6</td>
<td>106 - 109</td>
</tr>
</tbody>
</table>
Figure 5.3 A stereo view of the Cα trace of the P\textsubscript{II} monomer calculated at 1.9 Å. Labelled are every fifth residue and Tyr51. A noticeable bend in strand 2 can be seen at Glu32.
central four-strand $\beta$-sheet around the three-fold axis (Figure 5.4). Figure 5.4 shows schematically how the formation of the three eight-stranded $\beta$-sheets around the crystallographic triad are now intertwined via the curved strand 2 from each monomer. These interactions further extend the stabilisation of the trimer previously discussed in Chapter 4.

5.4.3 Crystal packing forces

The crystal packing forces that were mentioned for the P$_{II}$ structure determined at 2.7 Å remain the same for the higher resolution structure. As mentioned in Chapter 4, it is the symmetry mate of one P$_{II}$ trimer that interacts with the T-loop and helps to stabilise it. Figure 5.5 shows the interactions that residues within the T-loop have with symmetry related residues. Designation of the individual subunits of the two P$_{II}$ trimers involved will be as follows: A, B & C refer to the subunits from one trimer, while D, E & F refer to the subunits of the P$_{63}$ symmetry related trimer. The interactions discussed involve subunits A, D and F. As shown in Figure 5.5 the interactions that stabilise the 3-4 loop are hydrogen bonds between HE A47 & O51 D71, H A51 & O51 D75 / O52 D75, and H A53 & O F105. These crystal contacts support the T-loop at residues A47, A51 and A53.

Concentration of high temperature factors is within the T-loop (formerly the 3-4 loop) of the P$_{II}$ molecule. The mean B-factor values over the whole molecule for main-chain and side-chain atoms were 17.03 Å$^2$ and 20.23 Å$^2$ respectively, while for the T-loop the corresponding values were 31.26 Å$^2$ and 34.46 Å$^2$ (respectively). Glu44 exhibited the highest main-chain and side-chain B-factor values (43.23 Å$^2$ and 49.75 Å$^2$ respectively).
Figure 5.4  A schematic connectivity diagram for the P11 trimer calculated at 1.9 Å, with the secondary structural units defined as per Table 5.1. This diagram indicates the formation of three eight-stranded β-sheets about the three-fold axis as was seen in the 2.7 Å structure (Figure 4.9) with the added feature of the extended strands 2 and 3 adding greater stability to the trimer.
Figure 5.5 A schematic diagram showing the crystal contacts between monomers that maintain the T-loop in the P63 crystal form. The loop illustrated is from a monomer designated A. The other subunits (not shown) are denoted by B and C. The monomers from the symmetry related trimer are labelled D, E and F.
5.5 Comparison of PII to other proteins

5.5.1 Structural comparison of PII to other double β-α-β motif proteins

5.5.1.1 Introduction
As described earlier in this chapter and also in Chapter 4, the PII protein has an interesting folding pattern that brings about formation and stabilisation of the trimer. The folding of peptide to form the monomer, particularly the central double β-α-β motif, has been seen previously. Swindells et al. (1993) communicated a comparison of folding for the histidine-containing phosphocarrier protein (HPr) with acylphosphatase (APt). The results of this investigation suggested that both of these proteins had adopted a double β-α-β skeleton in order to present a stable phosphate-binding motif. Given the observed structural similarities between the PII and HPr monomers, and also the fact that PII is known to bind ATP (Kamberov et al., 1995), an investigation was undertaken to further detail the relationship between a protein containing a double β-α-β motif and the function that fold facilitates. While there are a number of proteins that contain this structural motif (Orengo et al., 1993) a subset has been selected for the purposes of this discussion.

The seven proteins discussed in this chapter are the ribosomal protein S6 from Thermus thermophilus, the DNA-binding domain of the transcriptional regulator E2 of the bovine papillomavirus-1 (BPV-1), Dictyostelium discoideum nucleoside diphosphate kinase (NDPK), the allosteric effector domain of the regulatory subunit of aspartate transcarbamylase (ar-ATCase) from E. coli, the activation segment of porcine pancreatic procarboxypeptidase A (PCPA), the U1 A domain of the ribonulceoprotein U1 snRNP from E. coli and the
histidine-containing phosphocarrier protein (HPr) from *Bacillus subtilis*. The topology's of the proteins to be compared with PII are shown in Figure 5.6.

5.5.1.2 Comparison with S6

S6 is one of more than fifty proteins that combine with three molecules of RNA to help form procaryotic ribosomes (Wittman, 1982). The X-ray structure of this 101 amino acid (a.a.) protein was determined at 2.0 Å by Lindahl *et al.* (1994). S6 consists of a four-stranded anti-parallel β-sheet with two anti-parallel α-helices on one side of the sheet. In all subsequent comparisons the Ca atoms used are within the six secondary elements of the double β-α-β motif: namely β1, α1, β2, β3, α2 and β4. The elements compared are listed and defined for each protein in Table 5.2. Forty four PII Ca atoms were compared to equivalent S6 atoms and this yielded a root-mean square (rms) distance of 1.85 Å. Figure 5.7 shows the similarity of the PII and S6 structures. It can also be seen from this figure that, as in PII, there exists additional two-fold symmetry perpendicular to the plane of the β-sheet in the S6 protein. Despite these structural similarities there is no significant sequence identity between these proteins.

5.5.1.3 Comparison with E2

The E2 protein is a dominant transcriptional regulator of BPV-1 (Ustav & Stenlund, 1991; Yang *et al.*, 1991). The crystal structure of the 85 a.a. DNA-binding domain of this protein bound to a 16 base-pair DNA sequence was solved to 1.7 Å (Hedge *et al.*, 1992). It was found that the DNA-binding domain of E2 was comprised of five strands and two α-helices. These elements were organised as a 4-stranded anti-parallel β-sheet, two anti-parallel α-helices, the second helix was followed by the short fifth strand. These domains formed a barrel shaped
Figure 5.6 Topology diagrams of proteins containing a double β-α-β motif. Shown are schematic diagrams for the appropriate sections of the following proteins: *E. coli* P11, *T. thermophilus* S6, BPV-1 E2 DNA-binding domain, *D. discoideum* NDPK, *E. coli* ATCase, porcine pancreatic PCPA activation segment, *E. coli* U1 snRNP A RNA-binding domain and *B. subtilis* HPr.
P_{II}

S6 ribosomal protein

E2 DNA-binding domain

NDPK

ar-ATCase

PCPA

U1 A domain of U1 snRNP

HPr
Table 5.2  Residues compared between P\textsubscript{II} and the proteins listed. The labels $\beta_1, \alpha_1, \beta_2, \beta_3, \alpha_2$ and $\beta_4$ refer to the order of secondary structural units as they appear in the double $\beta-\alpha-\beta$ motif.

<table>
<thead>
<tr>
<th>Protein (ptn)</th>
<th>RMS dist (Å)</th>
<th>$\beta_1$ ptn</th>
<th>P\textsubscript{II}</th>
<th>$\alpha_1$ ptn</th>
<th>P\textsubscript{II}</th>
<th>$\beta_2$ ptn</th>
<th>P\textsubscript{II}</th>
<th>$\beta_3$ ptn</th>
<th>P\textsubscript{II}</th>
<th>$\alpha_2$ ptn</th>
<th>P\textsubscript{II}</th>
<th>$\beta_4$ ptn</th>
<th>P\textsubscript{II}</th>
</tr>
</thead>
<tbody>
<tr>
<td>S6</td>
<td>1.85</td>
<td>4-9</td>
<td>3-8</td>
<td>22-31</td>
<td>12-21</td>
<td>39-43</td>
<td>29-33</td>
<td>60-65</td>
<td>58-63</td>
<td>70-81</td>
<td>68-79</td>
<td>87-91</td>
<td>90-94</td>
</tr>
<tr>
<td>E2</td>
<td>2.05</td>
<td>327-333</td>
<td>3-9</td>
<td>341-350</td>
<td>12-21</td>
<td>355-358</td>
<td>28-31</td>
<td>374-380</td>
<td>58-64</td>
<td>386-391</td>
<td>70-75</td>
<td>399-405</td>
<td>88-94</td>
</tr>
<tr>
<td>NDPK</td>
<td>2.39</td>
<td>7-12</td>
<td>3-8</td>
<td>21-30</td>
<td>12-21</td>
<td>38-42</td>
<td>29-33</td>
<td>73-79</td>
<td>57-63</td>
<td>83-93</td>
<td>69-79</td>
<td>117-121</td>
<td>88-92</td>
</tr>
<tr>
<td>ar-ATCase</td>
<td>2.46</td>
<td>15-19</td>
<td>3-7</td>
<td>24-34</td>
<td>11-21</td>
<td>41-46</td>
<td>28-33</td>
<td>57-60</td>
<td>59-62</td>
<td>67-75</td>
<td>70-78</td>
<td>81-86</td>
<td>90-95</td>
</tr>
<tr>
<td>hnRNP A</td>
<td>2.93</td>
<td>11-15</td>
<td>3-7</td>
<td>26-36</td>
<td>12-22</td>
<td>41-46</td>
<td>27-32</td>
<td>54-58</td>
<td>59-63</td>
<td>67-73</td>
<td>70-76</td>
<td>83-86</td>
<td>90-93</td>
</tr>
<tr>
<td>HPr*</td>
<td>5.55</td>
<td>65-61</td>
<td>4-8</td>
<td>N/A</td>
<td>N/A</td>
<td>39-43</td>
<td>29-33</td>
<td>33-36</td>
<td>60-63</td>
<td>N/A</td>
<td>N/A</td>
<td>2-6</td>
<td>88-92</td>
</tr>
</tbody>
</table>

* Only the strands of the double $\beta-\alpha-\beta$ motif in HPr were used in comparisons (see text).
Figure 5.7 Stereo plot of Ca-trace comparison for the *E. coli* P$_{11}$ and S6 proteins. Labelled for P$_{11}$ (black) are residues 1 and 112, as well as Tyr51. Labelled for S6 (red) are its N- and C-termini, residues A1 and A97, respectively.
dimer that involved the two 4-stranded sheets. This dimer bound the DNA in the central core of the barrel. When 41 Ca atoms were compared as listed in Table 5.2 a rms distance of 2.05 Å was calculated. Stereo diagrams showing the similarity of P11 to E2, NDPK, ar-ATCase, PCPA and U1 snRNP A are shown in Figure 5.8. No sequence identity was found between P11 and any of the proteins listed in Table 5.2.

5.5.1.4 Comparison with NDPK

NDPK functions to exchange a γ-phosphate between nucleoside tri- and diphosphates in what is termed a “ping-pong” mechanism (Parks & Agarwal, 1973). This kinase is the major source of GTP and with the exception of ATP derivatives, is also the source of most RNA and DNA precursors. The structure of NDPK from *Dictyostelium discoideum* was solved to 2.2 Å (Dumas *et al.*, 1992). The NDPK monomer was found to contain 4 anti-parallel strands and 4 α-helices. From these the double β-α-β motif was formed by β1-α1-β2 and β3-α3-β4. This 100 kDa protein exists as a symmetrical hexamer, composed of staggered trimers which maintain dimer contacts similar to those seen in ATCase (Stevens *et al.*, 1990). It was shown that the fourth strand histidine (His122) of NDPK was phosphorylated by nucleoside triphosphates (Wallet, 1992) and that this caused the nucleoside triphosphates to bind to the side rather than the face of the β-sheet. A comparison of 44 equivalent Ca atoms from P11 and NDPK yielded a rms distance of 2.39 Å (Figure 5.8).
Figure 5.8  MOLSCRIPT stereo Cα plots comparing $P_{II}$ (drawn in black) to the following proteins (drawn in red): E2 (A), NDPK (B), ar-ATCase (C), PCPA (D) and U1 A (E). Labelled on the $P_{II}$ Cα trace are residues 1, 51 and 112.
A) E2

B) NDPK

C) ar-ATCase
D) PCPA

E) U1A
5.5.1.5 Comparison with ar-ATCase

ATCase catalyses the reaction between carbamoyl phosphate and L-aspartate to give phosphate and N-carbamoyl-L-aspartate (Jones et al., 1955), the latter of which is used in the synthesis of pyrimidines (Reichard & Hanshoff, 1956). Stevens et al. (1990) have solved the structure of unligated, ATP- and CTP-complexed ATCase to 2.6 Å and found that in each case this protein existed as a dodecamer which contained the six catalytic and six regulatory chains (Weber, 1968; Wiley & Lipscomb, 1968). It is the allosteric domains (residues 14-97) of the regulatory subunits of ATCase that binds both CTP and ATP in their PII-like fold. Forty one PII Cα atoms align with ar-ATCase Cα atoms from the double β-α-β fold (Figure 5.8) with a rms distance of 2.46 Å.

5.5.1.6 Comparison with the activation segment of PCPA

PCPA is the inactive Zn-containing carboxypeptidase that is activated by trypsin. Digestion with trypsin catalyses the loss of a 95 a.a. activation segment (residues 4-99) in a two-step cascade mechanism (Rinderknecht, 1986). The remaining 307 a.a. form the active exopeptidase carboxypeptidase A (CPA), whose activity is regulated by the previously cleaved activation segment. The X-ray structure of porcine pancreatic PCPA was solved to 2.0 Å (Guasch et al., 1992) and was found to be similar to the structure of PCPB (Coll et al., 1991). The activation segment of PCPA consists of three α-helices and a four-stranded anti-parallel β-sheet. When the six components of the double β-α-β motif were compared with those found in PII, 36 Cα atoms overlayed (Figure 5.8) with a rms distance of 2.53 Å. Of these only 2 PCPA residues conserved identity with PII.
5.5.1.7 Comparison of the U1 A RNA-binding domain

There are many small nuclear ribonucleoproteins (snRNP) that function to bind RNA in order to facilitate splicing, storage and transport of RNA; U1 snRNP is one such snRNP (Maniatis and Reed, 1987). U1 snRNP encorporates the U1 A protein that contains the RNP specific RNA-binding motif (Lutz-Freyermuth et al., 1990). The RNA-binding domain of U1 snRNP A from \textit{E. coli} has been solved to 2.8 Å (Nagai et al., 1990), which has allowed speculation as to the nature of this proteins interaction with RNA. As with all these proteins the double \(\beta-\alpha-\beta\) motif was compared to the equivalent elements in P\(_{II}\) (Figure 5.8). This yielded a rms distance of 2.93 Å for 38 Ca atoms.

5.5.1.8 Comparison with HPr

HPr is a 9 kDa protein that is involved in phosphorylation of carbohydrates to initiate their metabolism (reviewed in Meadow et al., 1990). The HPr protein from \textit{Bacillus subtilis} is known to be phosphorylated at the N\(\delta\) atom of His15 and at Ser46 by an ATP-dependant protein kinase (Kalbitzer et al., 1982). Herzberg et al. (1992) solved the structure of \textit{B. subtilis} HPr at 2.0 Å and it was seen that this structure contained a double \(\beta-\alpha-\beta\) motif. Inspection of the topology of HPr (Figure 5.6) indicates that to align the strands of the \(\beta\)-sheet with those of P\(_{II}\), the helices of HPr must be placed on the opposing side to those from P\(_{II}\). Because of this only strands were used in comparative alignments. This comparison indicated that 19 Ca atoms overlayed with a rms distance of 5.55 Å. The poor alignment arises from the opposite directions of the \(\beta\)-sheet twists in P\(_{II}\) and HPr.
5.5.2 Relationship between double $\beta$-$\alpha$-$\beta$ motif and protein function

Prior to this work it had been suggested that there was a relationship between the double $\beta$-$\alpha$-$\beta$ motif and phosphate binding (Swindells et al., 1993). At the time it was suggested that the mode of phosphate binding was via the presence of positively charged residues at the amino terminus of the first helix in the double $\beta$-$\alpha$-$\beta$ fold. It now seems apparent that the characteristics of phosphate binding by this motif are broader.

The *E. coli* P$_{II}$ protein is known to have two different interactions with phosphate-containing ligands. The first is via covalent uridylylation of Tyr51 by UTase/UR (Son & Rhee, 1987). The formation of this phospho-tyrosyl bond is a stable modification that is only removed enzymatically by the uridylyl-removing activity of UTase/UR. The second interaction is the binding of ATP to P$_{II}$ which facilitates the binding of 2-ketoglutarate (2-KG), which in turn allows uridylylation of P$_{II}$ (Kamberov et al., 1995). Current investigations within this group are determining the positions of both ATP and 2-KG after introduction into native P$_{II}$ crystals (Dr E. Cheah, in preparation). The results of these investigations will add further detail to the underlying theme of phosphate binding in proteins containing a double $\beta$-$\alpha$-$\beta$ motif.

S6 from *T. thermophilus* is like P$_{II}$ in that it contains both a double $\beta$-$\alpha$-$\beta$ motif and the additional symmetry generated by the four strands. S6 is thought to bind phosphates in the form of ribosomal RNA (rRNA). S6, however, does not form a multimer but is thought to form a complex with S18—another ribosomal protein—as is the case in *E. coli* (Held et al., 1974; Prakash & Aune, 1978) and, as such, bind rRNA across the face of the extended $\beta$-sheet (Lindahl et al., 1994).
Overall *E. coli* ATCase is a dodecamer but the regulatory subunits form dimers (Stevens *et al.*, 1990). In forming a dimer the central β-sheet, which includes the double β-α-β motif, is extended to form a 10-stranded β-sheet. It is at the amino terminus of the original 4-stranded β-sheet that either the ATP or CTP binds. This mode of phosphate binding is similar to that in the S6/S18 complex. Although both ar-ATCase and P_{II} bind tri-phosphates, P_{II} differs in its mode of interaction with those phosphates. In P_{II} the extended β-sheets face inwards and are surrounded by the remainder of the protein. This structure makes binding of ATP in a fashion similar to either the S6/S18 complex or the ar-ATCase dimer unfavourable. Given this scenario it seems more likely that ATP and/or 2-KG would bind in the exposed cleft.

The activation segment of PCPA is the only protein in this study that does not have a functional role with either phosphates or phosphate esters. Porcine pancreatic PCPA loses a 95 a.a. polypeptide as a result of tryptic proteolysis (Guasch *et al.*, 1992). This segment then inhibits the remaining protein by binding to CPA through the face of the β-sheet contained in the double β-α-β motif. Despite the absence of phosphate involvement, it seems that the folding motif is still advantageous with regard to protein interactions. The activation segment of PCPA is the only member of this double β-α-β motif comparison group that is involved in protein-protein interactions. PCPA has a much higher level of hydrophobic residues than does any other of the comparison proteins, with the exception of P_{II}. This fact, in part, contributes to its ability to interact with other proteins.
The NDPK protein from *D. discoideum* is a hexamer that is assembled as a dimer of trimers (Dumas *et al.*, 1992). Unlike P11, the NDPK trimer is formed by contacts between loop sections of coil from each monomer, as is the case between the 3 regulatory dimers in ATCase (Stevens *et al.*, 1990). In this trimer helices are both inside and outside of the β-sheets. The α-helices involved in the double β-α-β motif are positioned on the inside of the sheet. Histidine122—the site of phosphorylation in NDPK—is located at the end of strand 4.

In the case of the E2 DNA-binding domain of BPV-1, the protein forms a dimeric barrel that is composed of 8 strands (Hedge *et al.*, 1992). The formation of this dimer positions the 2 α-helices on the outside of the barrel and it is the amino-terminal α-helix that interacts with the DNA—via the major groove—to wrap it around the barrel.

U1 snRNP A forms a dimer by packing the double β-α-β motif around a crystallographic dyad with the helices packed inside of the β-sheets (Nagai *et al.*, 1990). Unlike many of the other proteins described here there is no extension of the monomeric β-sheet in the dimer. Rather, dimer contacts are made through the helices. The U1 snRNP A dimer binds RNA between the edge of the β3 strands and the side of the α1 helices. This binding mode seems to embrace the notion that both the β-sheet and the positively charged amino-terminal α-helix are pivotal in phosphate interactions.

The picture to emerge from this comparison is that the double β-α-β motif does seem to have an important role in interactions with phosphates or phosphate esters. What seems apparent is that it is the structural motif and not sequence conservation that underlies the
ability of these proteins to interact with phosphates. This interaction is likely to occur in one of three regions within this motif. The first is interaction with the face of the β-sheet, as is the case in S6, NDPK and ar-ATCase. While not involving phosphates or nucleotides, the interaction of the activation segment of PCPA with CPA still exhibits interaction of a similar nature to that shown in S6, NDPK and ar-ATCase. The second would be interaction with the amino-terminal α-helix, that would contain large numbers of positively charged residues, as seen in the E2 protein. The third region for possible phosphate interaction with a protein containing a double β-α-β motif would appear to be between the edge of a β-sheet and the amino-terminal α-helix, as is the case in the U1 snRNP A protein.

These trends suggest strongly that the double β-α-β motif is inherently involved in interactions with phosphates or nucleotides. In the case of P11 however, formation of the trimer makes those regions most likely to bind ATP inaccessible. In compensation the trimer yields three cleft regions that might carry out this function. As we await the complete refinement results of the P11/ATP/2-KG structure, it seems likely that the ATP moiety and possibly the 2-KG will be found in the cleft regions mentioned above and thus add further support to the hypotheses suggested.
Chapter 6. Beyond the P\textsubscript{II} structure

6.1 Introduction

The structure of P\textsubscript{II} suggests a number of ways in which this protein might carry out its signal transducing function (a complete treatment can be found in Chapter 8). The features of P\textsubscript{II} most likely to be involved in signal transduction or protein-protein interaction are the T-loop, the conserved upper-surface of the trimer and the elliptical cleft that is formed by contributions from any two adjacent monomers.

At the beginning of this study there were a number of effector ligands that were thought to influence the function of P\textsubscript{II}, depending on the status of cellular nitrogen (Engleman & Francis, 1978; Francis & Engleman, 1978). These ligands included 2-ketoglutarate, ATP and also glutamine. It had been determined that 2-ketoglutarate in the presence of ATP was required for uridylylation of P\textsubscript{II} by UTase/UR. Glutamine was required by UTase/UR to inhibit uridylylation of P\textsubscript{II} and to stimulate deuridylylation of P\textsubscript{II}-UMP. With the structure of P\textsubscript{II} determined the question remained: How does the interaction of these ligands with P\textsubscript{II} affect its structure so that the transduced signal is reversed? Attempts to describe further this situation formed the basis of the work reported in this chapter.

This chapter describes the structural investigation of P\textsubscript{II} co-crystallised with 2-ketoglutarate. More recent biochemical information indicated glutamate also interacted with P\textsubscript{II} (Kamberov et al., 1994) and subsequent investigation into co-crystallisation of P\textsubscript{II} and glutamate was undertaken. Due to the conditions at which these co-crystals were
achieved, it was found necessary to re-solve the native P_{II} structure at pH 9.5, by step-wise transfer of native crystals to new buffers.

6.2 Materials and methods

The chemicals used in this study were purchased from either Sigma, Aldrich, Fluka Biochemika, ICN or the British Drug Houses Ltd. Glutamate was bought from the British Drug Houses Ltd and 2-ketoglutarate from Aldrich.

Refinement of the model coordinates were carried out using the X-PLOR program version 3.1 (Brünger, 1992) as implemented on a Fujitsu VP2200 supercomputer. Access to these facilities were made possible by a grant of time from the ANU Supercomputing Unit.

Our models were checked for deviations from ideality using the program PROCHECK (Laskowski et al., 1993). This program was also used to produce Ramachandran plots (Ramachandran & Sasisekharan, 1968).

The secondary structural assignments for the P_{II} structure in this chapter were made using the DSSP package of programs (Kabsch & Sander, 1983).

The diagrams in this chapter representing the P_{II} structure were generated by the programs MOLSCRIPT (Kraulis, 1991) and MIDAS (Ferrin et al., 1988) using PDB formatted coordinates.
6.3 P11 with bound 2-ketoglutarate

6.3.1 Crystallisation
Co-crystals of P11 with 2-ketoglutarate (2-KG) were achieved via the Magic 94 crystallisation screens kindly provided by Dr Jim Remington. The first crystal was grown using the Magic 94 screen number 56, which contained 1.2 M A.S., 0.2 M NaCl and 0.1 M 2-(cyclohexylamino)ethanesulfonic acid (CHES) at pH 9.5. A P11/2-KG mixture was made containing 99 µl P11 (initial concentration 18 mg/ml) and 1 µl 2-KG (initial concentration 20 mg/ml). Per drop 4 µl of this protein/ligand mixture was added to 4 µl of the Magic 94 screen number 56. Crystallisation was achieved at 4 °C by the Micro-Bridge Sitting Drop method from Hampton Research (HR3-312). This crystal was 0.1 mm x 0.1 mm x 0.3 mm and had a hexagonal morphology. A dataset was collected on this crystal at 4 °C (see below).

A second co-crystal of P11 with 2-ketoglutarate was grown using the same conditions described above. This crystal also had a hexagonal morphology but was larger than the previous crystal (Figure 6.1A), with dimensions of 0.3 mm x 0.2 mm x 0.3 mm. A 4 °C dataset was collected on this crystal in the same manner as the first (see below).

6.3.2 Data collection and refinement
The first P11 / 2-ketoglutarate co-crystal was grown from the Magic 94 screen number 56 by Dr Karen Edwards. This crystal grew in a hexagonal morphology, similar to that shown in Figure 2.3B (Chapter 2). This morphology had the same apex and overall shape as in Figure 2.3B, but with one end slightly elongated. Such a morphology suggested that data collected from this crystal would exhibit a low–less than 0.1–twin fraction. This crystal diffracted to 1.9 Å and was
Figure 6.1 Crystal morphology of P11 co-crystallised with; A) 2-ketoglutarate and B) glutamate.
indexed in the P6\textsubscript{3} space group with an acceptance of 88.33 \% and had cell dimensions of $a = 60.9 \text{ Å}$ and $c = 58.9 \text{ Å}$. Sixteen 4° oscillations were collected for 45 minutes each which netted a dataset that was 82.6 \% complete for data between 10.0 Å and 2.5 Å (PIIAKG1)(Table 6.1).

The PIIAKG1 dataset contained 27 556 observations of 7 392 unique reflections which had an overall $R_{\text{merge}}$ of 4.9 \% and of these 3 544 reflections had $I > 3\sigma_I$. When treated with the detwinning process described in Chapter 3 this data was found to have no significant twin fraction. This level of twinning was expected given the morphology of the crystal.

The second P\textsubscript{11}/2-ketoglutarate co-crystal was grown from repeat conditions of the first crystal. This hexagonal crystal was larger than the first—0.3 mm × 0.3 mm × 0.3 mm—and can be seen in Figure 6.1A. This crystal diffracted to 2.0 Å (Figure 6.2) and was indexed in the P6\textsubscript{3} space group with an acceptance of 87.21 \%. Cell dimensions of $a = 61.2 \text{ Å}$ and $c = 58.8 \text{ Å}$ were determined. Twenty oscillations of 4° were collected for 40 minutes each, which netted a dataset that was 85.1 \% complete for data between 10.0 Å and 2.5 Å (PIIAKG2)(Table 6.2).

The PIIAKG2 dataset contained 41 550 observations of 8 349 independent reflections, which had an $R_{\text{merge}}$ of 6.5 \% and of these 3 689 reflections had $I > 3\sigma_I$. When this data was detwinned it was found to have a 0.11 twin fraction. This value was indicative of the crystal morphology observed (Figure 6.1A).
Table 6.1 Completeness of the $\text{P}_{11}/2$-ketoglutarate co-crystal dataset (PIIAKG1). Completeness of data (%) is shown collected in twenty bins equally distributed between 10.0 Å and 2.5 Å, along with the number of reflections residing within each bin.

<table>
<thead>
<tr>
<th>$d_{\text{min}}$ (Å)</th>
<th>$d_{\text{max}}$ (Å)</th>
<th>No. of reflections</th>
<th>% complete</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.22</td>
<td>10.00</td>
<td>179</td>
<td>79.8</td>
</tr>
<tr>
<td>5.15</td>
<td>6.22</td>
<td>185</td>
<td>80.8</td>
</tr>
<tr>
<td>4.57</td>
<td>5.15</td>
<td>179</td>
<td>81.8</td>
</tr>
<tr>
<td>4.19</td>
<td>4.57</td>
<td>179</td>
<td>82.0</td>
</tr>
<tr>
<td>3.91</td>
<td>4.19</td>
<td>176</td>
<td>82.8</td>
</tr>
<tr>
<td>3.69</td>
<td>3.91</td>
<td>186</td>
<td>82.9</td>
</tr>
<tr>
<td>3.51</td>
<td>3.69</td>
<td>187</td>
<td>83.4</td>
</tr>
<tr>
<td>3.37</td>
<td>3.51</td>
<td>184</td>
<td>83.9</td>
</tr>
<tr>
<td>3.24</td>
<td>3.37</td>
<td>180</td>
<td>83.9</td>
</tr>
<tr>
<td>3.13</td>
<td>3.24</td>
<td>182</td>
<td>84.0</td>
</tr>
<tr>
<td>3.04</td>
<td>3.13</td>
<td>177</td>
<td>84.2</td>
</tr>
<tr>
<td>2.95</td>
<td>3.04</td>
<td>175</td>
<td>83.7</td>
</tr>
<tr>
<td>2.88</td>
<td>2.95</td>
<td>180</td>
<td>83.9</td>
</tr>
<tr>
<td>2.81</td>
<td>2.88</td>
<td>156</td>
<td>83.4</td>
</tr>
<tr>
<td>2.75</td>
<td>2.81</td>
<td>191</td>
<td>83.5</td>
</tr>
<tr>
<td>2.69</td>
<td>2.75</td>
<td>173</td>
<td>83.4</td>
</tr>
<tr>
<td>2.64</td>
<td>2.69</td>
<td>166</td>
<td>83.4</td>
</tr>
<tr>
<td>2.59</td>
<td>2.64</td>
<td>178</td>
<td>83.1</td>
</tr>
<tr>
<td>2.54</td>
<td>2.59</td>
<td>169</td>
<td>82.9</td>
</tr>
<tr>
<td>2.50</td>
<td>2.54</td>
<td>162</td>
<td>82.5</td>
</tr>
</tbody>
</table>
Figure 6.2 An oscillation from the P$_{11}$/2-ketoglutarate dataset, PIIAKG2. Data were collected at room temperature as 40 minute exposures of 4° oscillations. The crystal to film distance was 90.0 mm which set the diffraction limits to a resolution of 2.0 Å at the edge of the imaging plates. This figure indicates that this crystal diffracted out to these limits.
Table 6.2 Completeness of the PII/2-ketoglutarate co-crystal dataset (PIIAKG2). Completeness of data (%) is shown collected in twenty bins equally distributed between 10.0 Å and 2.5 Å, along with the number of reflections residing within each bin.

<table>
<thead>
<tr>
<th>$d_{min}$ (Å)</th>
<th>$d_{max}$ (Å)</th>
<th>No. of reflections</th>
<th>% complete</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.22</td>
<td>10.00</td>
<td>175</td>
<td>78.1</td>
</tr>
<tr>
<td>5.15</td>
<td>6.22</td>
<td>181</td>
<td>79.5</td>
</tr>
<tr>
<td>4.57</td>
<td>5.15</td>
<td>190</td>
<td>81.3</td>
</tr>
<tr>
<td>4.19</td>
<td>4.57</td>
<td>188</td>
<td>82.9</td>
</tr>
<tr>
<td>3.91</td>
<td>4.19</td>
<td>177</td>
<td>82.9</td>
</tr>
<tr>
<td>3.69</td>
<td>3.91</td>
<td>185</td>
<td>83.5</td>
</tr>
<tr>
<td>3.51</td>
<td>3.69</td>
<td>183</td>
<td>83.7</td>
</tr>
<tr>
<td>3.37</td>
<td>3.51</td>
<td>194</td>
<td>84.1</td>
</tr>
<tr>
<td>3.24</td>
<td>3.37</td>
<td>190</td>
<td>84.6</td>
</tr>
<tr>
<td>3.13</td>
<td>3.24</td>
<td>196</td>
<td>85.1</td>
</tr>
<tr>
<td>3.04</td>
<td>3.13</td>
<td>179</td>
<td>85.1</td>
</tr>
<tr>
<td>2.95</td>
<td>3.04</td>
<td>196</td>
<td>85.5</td>
</tr>
<tr>
<td>2.88</td>
<td>2.95</td>
<td>187</td>
<td>85.5</td>
</tr>
<tr>
<td>2.81</td>
<td>2.88</td>
<td>187</td>
<td>85.5</td>
</tr>
<tr>
<td>2.75</td>
<td>2.81</td>
<td>178</td>
<td>85.5</td>
</tr>
<tr>
<td>2.69</td>
<td>2.75</td>
<td>187</td>
<td>85.6</td>
</tr>
<tr>
<td>2.64</td>
<td>2.69</td>
<td>191</td>
<td>85.6</td>
</tr>
<tr>
<td>2.59</td>
<td>2.64</td>
<td>180</td>
<td>85.5</td>
</tr>
<tr>
<td>2.54</td>
<td>2.59</td>
<td>177</td>
<td>85.4</td>
</tr>
<tr>
<td>2.50</td>
<td>2.54</td>
<td>168</td>
<td>85.1</td>
</tr>
</tbody>
</table>
Attempts to solve the PII/2-KG co-crystallised structure initially involved the PIIAKG1 and then finally the PIIAKG2 datasets. Both approaches followed parallel routes and as such the method will be described only once. The initial model used for this refinement was the refined structure of PII at 1.9 Å, with all water removed. Using 3σ data between 8.0 Å and 2.0 Å the model underwent cycles of positional refinement till convergence. This step reduced the R-factor from 49.2 % to 40.5 %. Following this B-factors were refined to convergence which reduced the R-factor to 37.3 %.

At this point the current model was checked against 2F_o-F_c and F_o-F_c maps. The trace of the model seemed correct but indicated some areas of change. These regions were the T- and B-loops. Both areas were rebuilt according to these maps and 73 waters were added to spherical 2F_o-F_c density greater than 2σ. After these modifications the model coordinates were again refined with a round of positional and then B-factor refinement to convergence, which resulted in an R-factor of 31.2 %. All further refinement consisted of positional, then B-factor refinement, inspection of the model with 2F_o-F_c and F_o-F_c maps, and then rebuilding the model, which included adding waters.

This further refinement did not reduce the R-factor to below 30 %. It seemed possible that the T- and the B-loops, which were the most altered in the PII structure, might be the cause. These two regions were subsequently removed from refinement in order to gain a better overall trace for the molecule, at the expense of some side-chain detail. The quality of these further maps improved to the point where the T- and the B-loops were able to be traced but refinement of these models maintained R-factors above 30 %. It seemed the 2-ketoglutarate had
bound non-specifically to \( \text{P}_{\text{II}} \) causing disorder within the T- and B-loops. The failure to successfully trace these regions may have reflected the multiplicity of the structures within the crystal packing.

It was at this time that Kamberov et al. (1995) published results of an investigation into the ability of \( \text{P}_{\text{II}} \) to bind 2-KG and glutamate in the presence and absence of ATP. Their findings suggested that while the presence of ATP enhances the ability of \( \text{P}_{\text{II}} \) to bind both 2-KG and glutamate, in its absence both ligands bind non-specifically to the \( \text{P}_{\text{II}} \) trimer. These results seem to explain in part the difficulties that were encountered with the refinement of the two \( \text{P}_{\text{II}}/2\text{-KG} \) datasets. Due to time constraints further investigation of this problem was not possible.

### 6.4 \( \text{P}_{\text{II}} \) with bound glutamate

#### 6.4.1 Crystallisation

Co-crystals of \( \text{P}_{\text{II}} \) with glutamate were achieved from a number of conditions that had been screened around the conditions for the Magic 94 screen number 56 or the fast screen number 39 from Hampton Research. Fast screen number 39 contains 2 M A.S., 2 % (v/v) PEG 400 and 0.1 M HEPES at pH 7.5. Each crystallisation drop contained 4 µl of the well solution and 4 µl of a protein/ligand mix. Each 4 µl of the protein mix contained 2.8 µl of \( \text{P}_{\text{II}} \) at 9.2 mg/ml and 1.2 µl of 100 mM glutamate.

The first set of \( \text{P}_{\text{II}}/\text{glutamate} \) co-crystals were grown over a well that contained 2.1 M A.S., 3 % (v/v) PEG 400 and 0.1 M HEPES at pH 7.5. These crystals were either very small crystallites or were visibly mosaic. In either case these crystals were not of sufficient quality for
X-ray analysis. Conditions were varied in an attempt to improve the crystals, but this was not achieved satisfactorily.

The second set of crystals for P_{II}/glutamate co-crystals were grown from 1.4 M A.S., 0.3 M NaCl and 0.1 M CHES at pH 9.5. These crystals were small—less than 0.1 mm × 0.1 mm × 0.1 mm—and intergrown. The crystals did have visible physical separations that allowed singular crystals to be separated and mounted for X-ray data collection. As already mentioned these crystals were too small and did not diffract. Optimisation of these conditions did yield crystals of a sufficient size and quality for X-ray analysis.

Improved P_{II}/glutamate co-crystals were grown from 1.8 M A.S., 0.3 M NaCl and 0.1 M CHES at pH 9.5 using the same protein/ligand mix described previously. These crystals were 0.2 mm × 0.2 mm × 0.4 mm (Figure 6.1B) and had a “bi-apex” hexagonal morphology similar to those found for native P_{II} (Figure 2.3A). An interesting feature of these crystals was a straight indentation found in the centre of these crystals perpendicular to their 6-fold axis. By analogy with the physical separation found for the crystals described in Chapter 2 (Figure 2.3A) it was thought that this separation may define the high and low twinned areas of the crystal. Accordingly, prior to X-ray analysis these double-ended crystals were physically separated at their midsection and one end analysed at a time.

6.4.2 Data collection

Three twenty minute still photographs were taken of the P_{II}/glutamate co-crystal at 4 °C and with a crystal to film distance of 90.0 mm. Each still photograph showed reflections out to 2.0 Å. These stills were indexed in the P6_{3} space group, with an acceptance of 88.34 %.
determined cell dimensions for this co-crystal were $a = 61.6 \, \text{Å}$ and $c = 59.0 \, \text{Å}$. These figures indicated that as was the case with the $\text{P}_{11}/2$-ketoglutarate co-crystals, the length of the $c$ axis in the unit cell of the $\text{P}_{11}/\text{glutamate}$ co-crystal was extended relative to the native $\text{P}_{11}$ cell dimensions.

Oscillation data were collected as 60 minute exposures of $3^\circ$ changes in phi at $4 \, ^\circ \text{C}$. The first oscillation photograph of this crystal showed reflections to approximately $2.5 \, \text{Å}$ (Figure 6.3A), but by the fifth photograph the crystal had stopped diffracting (Figure 6.3B). Inspection of the crystal indicated that dehydration did not seem to be the cause of the short term diffraction. Further attempts of co-crystallisation and data collection have not yet yielded better X-ray quality crystals.
Figure 6.3 Oscillation photographs of the P11/glutamate co-crystal. Both oscillations were collected for 60 minutes at 4 °C and represent a Δφ of 3°. Crystal to film distance was 90.0 mm representing 1.9 Å at the edges of the plate. A) The 0–3° oscillation; B) The 9–12° oscillation.
6.5 P\text{II} structure at pH 9.5

6.5.1 Crystal preparation and data collection

In an attempt to determine changes the P\text{II} structure might undergo at higher pH, a native crystal similar to that shown in Figure 2.3B was grown at pH 7.0—as per Chapter 2—and sequentially soaked in mixture buffers until the final storage solution was the Magic 94 screen number 56 (M56) provided by Dr Jim Remington. The exchange of one buffer for the other was carried out according to the following procedure.

A native P\text{II} crystal was placed in a mixture containing 800 µl of the NaCl stabilisation buffer and 200 µl of M56 for 15 minutes. Next 500 µl of this mixture, including the crystal, was withdrawn and added to 500 µl of M56 for a further 15 minutes. The crystal was then removed from this solution in 200 µl of the mixture and added to a further 800 µl of M56 for an hour. At the end of this time the crystal was taken from this mixture and placed in M56 alone overnight (16 hours).

The crystal resulting from the treatment above was mounted in the stop-flow type apparatus mentioned (Figure 3.1) and data were collected. This crystal was found to diffract out to 2.2 Å and was indexed in the P6\text{3} space group, with cell dimensions of a = 61.8 Å and c = 56.2 Å. These dimensions vary from those found for NAT2 (a = 61.6 Å and c = 56.3 Å) by 0.3 % in a and 0.2 % in c. These figures indicated no major change to the crystal packing due to the change in both buffer and pH.
A dataset was collected for this crystal that comprised of thirty, 4° oscillations each of 45 minutes. This resulted in 50,923 observations of 8,723 unique reflections, which had an $R_{\text{merge}}$ of 8.1%. Of the unique reflections observed 3,018 had $I > 3\sigma_I$ between 10.0 Å and 2.5 Å, which were used in refinement. Table 6.3 shows the completeness of the high pH native data set, NAT95, and the number of reflections found in bins sorted according to resolution. A twin fraction of 0.09 was determined for this data. This value adhered to our hypothesis relating the morphology of a crystal to its associated t.f.

6.5.2 Structure refinement
The initial model for refinement of the P$_{II}$ structure at pH 9.5 was the final P$_{II}$ model refined at 1.9 Å with all waters removed. Using the measured structure factors and the model phases, initial positional and B-factor refinements were carried out. After 400 cycles of positional and 20 cycles of B refinement the R-factor was 20.3%. The geometry of this model was checked via the PROCHECK program (Laskowski et al., 1993) and problem residues flagged. Both $2F_o-F_c$ and $F_o-F_c$ electron density maps were calculated and side chains rebuilt according to the maps and the PROCHECK output. A further round of 400 cycles positional and 20 cycles of B-factor refinements were carried out resulting in a 19.9% R-factor for $I > 3\sigma_I$ data between 10.0 Å and 2.5 Å.

At this stage of the refinement, bound waters were fitted to uninhabited spherical density in $3\sigma$ $2F_o-F_c$ or $2\sigma F_o-F_c$ maps calculated using the current model. In this first round of water fitting 174 waters were placed. With these waters included into the model a further round of 400 cycles positional and 20 cycles of temperature factor/
Table 6.3 Completeness of the PII native dataset collected at pH 9.5 after changing the buffer from pH 7.0 (NAT95). Completeness of data (%) is shown collected in twenty bins equally distributed between 10.0 Å and 2.5 Å, along with the number of reflections residing within each bin.

<table>
<thead>
<tr>
<th>d_{min} (Å)</th>
<th>d_{max} (Å)</th>
<th>No. of reflections</th>
<th>% complete</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.22</td>
<td>10.00</td>
<td>154</td>
<td>71.0</td>
</tr>
<tr>
<td>5.15</td>
<td>6.22</td>
<td>157</td>
<td>71.5</td>
</tr>
<tr>
<td>4.57</td>
<td>5.15</td>
<td>161</td>
<td>72.4</td>
</tr>
<tr>
<td>4.19</td>
<td>4.57</td>
<td>149</td>
<td>72.1</td>
</tr>
<tr>
<td>3.91</td>
<td>4.19</td>
<td>139</td>
<td>70.7</td>
</tr>
<tr>
<td>3.69</td>
<td>3.91</td>
<td>147</td>
<td>70.6</td>
</tr>
<tr>
<td>3.51</td>
<td>3.69</td>
<td>136</td>
<td>69.4</td>
</tr>
<tr>
<td>3.37</td>
<td>3.51</td>
<td>133</td>
<td>68.8</td>
</tr>
<tr>
<td>3.24</td>
<td>3.37</td>
<td>127</td>
<td>68.1</td>
</tr>
<tr>
<td>3.13</td>
<td>3.24</td>
<td>129</td>
<td>67.4</td>
</tr>
<tr>
<td>3.04</td>
<td>3.13</td>
<td>134</td>
<td>66.8</td>
</tr>
<tr>
<td>2.95</td>
<td>3.04</td>
<td>112</td>
<td>65.8</td>
</tr>
<tr>
<td>2.88</td>
<td>2.95</td>
<td>125</td>
<td>65.4</td>
</tr>
<tr>
<td>2.81</td>
<td>2.88</td>
<td>88</td>
<td>63.9</td>
</tr>
<tr>
<td>2.75</td>
<td>2.81</td>
<td>99</td>
<td>62.5</td>
</tr>
<tr>
<td>2.69</td>
<td>2.75</td>
<td>79</td>
<td>61.2</td>
</tr>
<tr>
<td>2.64</td>
<td>2.69</td>
<td>76</td>
<td>59.5</td>
</tr>
<tr>
<td>2.59</td>
<td>2.64</td>
<td>62</td>
<td>57.9</td>
</tr>
<tr>
<td>2.54</td>
<td>2.59</td>
<td>79</td>
<td>56.9</td>
</tr>
<tr>
<td>2.50</td>
<td>2.54</td>
<td>80</td>
<td>55.9</td>
</tr>
</tbody>
</table>
water occupancy (BQ) refinements were carried out, resulting in an R-factor of 14.5% calculated using 3σ data between 10.0 Å and 2.5 Å.

Improved 2F₀-Fc and F₀-Fc electron density maps were calculated using the new model phases and the structure factors from the NAT95 dataset. These maps allowed residues flagged by PROCHECK to be re-fitted to electron density. Pre-existing waters were checked against these new maps, which resulted in 14 being removed with 15 new waters being positioned within new density.

This model then underwent one round of 400 cycles positional refinement, followed by 20 cycles each of BQB refinement and then cycles of positional refinement to convergence. The resultant model had an R-factor of 12.8% for 3σ data between 10.0 Å and 2.5 Å. Electron density maps generated from these model phases indicated 9 poorly fitted waters which were removed, leaving a total of 166 bound waters in the structure. A final round of 400 cycles positional refinement followed by 20 cycles each of BQB refinement were carried out which resulted in the final model. The final model of P11 resoaked at pH 9.5 had an R-factor of 10.6% for 3σ data between 10.0 Å and 2.5 Å and an R-factor of 15.8% for 3σ data between ∞ and 2.5 Å. Figure 6.4 shows representative electron density calculated for this final model.

Upon comparing this structure with the P11 structure solved to 1.9 Å at pH 7.0, there were no appreciable differences in backbone positions. A comparison of Cα, C, N, O and Cβ atoms of the P11 structures determined at pH 7.0 and pH 9.5 was undertaken. This investigation indicated that for the 550 atoms used in the comparison there was a rms distance of 0.31 Å between the structures. Figure 6.5 shows a Cα
Figure 6.4 A stereo view of the final $2F_o-F_c$ map for the $P_{11}$ model at pH 9.5. This model was calculated using $3\sigma$ data between 10.0 Å and 2.5 Å resolution. The map shown is contoured at 0.8$\sigma$. 
Figure 6.5  MOLSCRIPT stereo Cα trace comparison of the P11 structures determined at pH 7.0 (shown in black) and at pH 9.5 (shown in red).
trace comparison of the P_{11} structure solved at pH 9.5 and to 2.5 Å resolution and the structure solved to 1.9 Å at pH 7.0. It was found that both structures had mean coordinate errors of 0.15 Å, as indicated by their respective Luzzati plots (Luzzati, 1952) (Figures 5.1 and 6.6).

The structure of P_{11} at pH 9.5 suggests that this protein does not have a structure substantially different from that of P_{11} at pH 7.0. This result is important as it suggests any changes in the structure of P_{11} co-crystallised with ligands at higher pH is due to the presence of the ligands, not the change in pH. As refinement of P_{11}/ligand structures are completed (Edwards et al., 1996), we anticipate describing ligand alterations to the P_{11} protein that have biological significance and will help further describe the role of this protein in signal transduction.
Figure 6.6 A Luzzati plot for the P11 structure at pH 9.5, calculated at 2.5 Å resolution. Plotted are the R-factors for respective resolutions which are shown as their reciprocal. Also shown are the theoretical distributions (---) for mean coordinate errors of 0.1 Å, 0.15 Å and 0.2 Å as described by Luzzati (Luzzati, 1952).
Chapter 7. Overexpression and purification of \( \text{glnL} \) and \( \text{glnG} \) gene products

7.1 Introduction

The \( P_{II} \) protein interacts with a number of different proteins to bring about the regulation of GS activity and transcription (see Chapter 1). Regulation of GS activity is via either the direct interaction of \( P_{II} \) or \( P_{II} \text{-UMP} \) with ATase. \( P_{II} \) interacts with \( NR_{II} \) to regulate transcription of GS. The \( P_{II} \text{-NR}_{II} \) complex phosphorylates the translational enhancer nitrogen regulator protein I (\( NR_{I} \)) which then results in the enhanced transcription of \( glnA \), the gene for GS.

Though the model for the structure of \( P_{II} \) offered clues about its mode of action, the structure of the \( P_{II} \text{-NR}_{II} \) complex would yield greater insight into the manner in which \( P_{II} \) carries out its role as a signal transduction protein. This would also offer insights into how the uridylylation state of \( P_{II} \) ultimately determines the regulation of transcription of \( glnA \).

The main objective of this investigation was to describe structural interactions between \( P_{II} \) and \( NR_{II} \), and the functional interactions between the \( P_{II} \text{-NR}_{II} \) complex and \( NR_{I} \). This chapter describes the over-production of the nitrogen regulator proteins I and II (the respective gene products of the \( glnG \) and \( glnL \) genes), the purification of \( NR_{II} \) and attempts at co-crystallisation of \( P_{II} \) with \( NR_{II} \). Results published during our work (Kamberov et al., 1994; Kamberov et al., 1995), answered the questions we hoped to study and consequently the biochemical investigation mentioned above was not undertaken. The
findings of Kamberov et al. (1994) did however yield information that assisted our initial attempts of crystallising the P11-NR11 complex.

7.2 Materials and Methods

7.2.1 Bacterial Strains

*Escherichia coli* strains used in this study are listed in Table 7.1. The most commonly used strain was AN1459, while AN2666 was used for ssDNA production and BL21::DE3/pLysS was used for the IPTG induced pETMCS-I expression systems.

7.2.2 Growth Media

*E. coli* strains were grown in LB medium (Luria and Burrous, 1957) containing 25 µg/ml thymine (LBT). When required, ampicillin (LBTA medium) and chloramphenicol (LBTAC medium) were used at 50 µg/ml.

7.2.3 Rapid Plasmid DNA Isolation

Plasmid DNA was isolated from *E. coli* by alkaline extraction according to Silhavy et al. (1984) and stored at -20 °C in a mixture (TE) containing 10 mM Tris(hydroxymethyl)aminomethane (Tris) and 1 mM ethylenediaminetetraacetate (EDTA), supplemented with RNaseA at 20 µg/ml.

7.2.4 Restriction Endonuclease Digestion of DNA

In general 2.5 units of restriction enzyme (Boehringer-Mannheim) were used per µg of DNA for 1 h at 37 °C. Fragments of DNA resulting from digestions were isolated as described below.
Table 7.1 The *Escherichia coli* strains used in this study and their partial genotypes.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Genotype</th>
<th>Source/Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>AN1459</td>
<td><em>supE44 thi-1 leuB6 thr-1 ilvC hsdR recA srLA::Tn10</em></td>
<td>Elvin <em>et al.</em> (1986)</td>
</tr>
<tr>
<td>AN2666 (JM101recA)</td>
<td><em>supE44 thi-1 Δ(lac-proAB) recA srLA::Tn10</em></td>
<td>L. Hatch</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Yanisch-Perron <em>et al.</em>, 1985</td>
</tr>
<tr>
<td>BL21::DE3/pLysS</td>
<td><em>hsdS gal (λcI857 indI Sam7 min5 lac UV5-T7 gene1)</em></td>
<td>Studier &amp; Moffatt (1986)</td>
</tr>
</tbody>
</table>
7.2.5 Agarose Gel Electrophoresis

Both preparative and analytical agarose gel electrophoresis were carried out in a Davis system horizontal submarine apparatus (147 mm × 136 mm × 10 mm; Maniatis et al., 1982) with 1.5 mm sample wells. Agarose concentrations ranging from 0.7 % to 1.0 % (w/v) were used. Gels were prepared in a buffer containing 89 mM Tris, 89 mM boric acid, 2 mM EDTA (TBE; Sambrook et al., 1989) which contained 0.5 µg/ml ethidium bromide. Samples were loaded into cast gels and subjected to a 60 V field until the DNA had entered the gel. Voltage of 100 V was then applied to the gel until the desired resolution of DNA fragments was achieved. DNA fragments were visualised using either a long-wave UV desk lamp or a short-wave transilluminator (UV transilluminator Model TS-15, Ultraviolet Products). Photographs were taken on the transilluminator using a Model MP-4 Land Camera (Polaroid) with Polaroid type 57 high speed film.

DNA fragments were isolated from agarose gels by a modification of the procedure of Dretzen et al. (1981). Fragments were electrophoresed onto NA-45 membrane (Schleicher & Schuell) and recovered by elution into 400 µl of a solution of 1 M NaCl containing 50 mM arginine at 70 °C. After 1 h the membrane was removed and the DNA precipitated with two volumes of ethanol at - 70 °C. Samples subsequently underwent centrifugation (Eppendorf Microfuge); the pellet was dried in vacuo and then resuspended in 20 µl of TE and stored at - 20 °C.

7.2.6 Ligation of DNA Fragments into Vectors

Ligation of DNA fragments were carried out in T4 DNA ligase, where the concentrations of insert and vector used were adjusted per recommendations of Legerski and Robberson (1985). Where both insert
and vector had cohesive termini the ligation buffer described by Sambrook et al. (1989) was used. These reactions were carried out for 16 h at 14 °C in 20 µl total volume. When the fragments of DNA to be ligated had blunt-ended termini a blunt-end ligation buffer (Sambrook et al., 1989) was used. These reactions were run for 2 h at 30 °C in a total volume of 20 µl.

7.2.7 SDS Polyacrylamide Gel Electrophoresis
The sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) performed in this investigation were under denaturing conditions as described by Laemmli (1970). Gels were cast between glass plates resulting in a gel 200 mm × 150 mm × 1.5 mm. Both resolving and stacking gels were made as described by Maniatis et al. (1982). Resolving gels contained 10 ml of 30 % (w/v) acrylamide/bisacrylamide (30:2.7), 7.5 ml of 1.5 M Tris.HCl pH 8.8, 300 µl of 10 % (w/v) SDS, 10 µl of N,N,N',N'-Tetramethylethylethylenediamine (TEMED), 100 µl of ammonium persulphate and 12.1 ml of water. The stacking gels consisted of 1.12 ml of 30 % (w/v) acrylamide/bisacrylamide (30:2.7), 1.87 ml of 0.5 M Tris.HCl pH 6.8, 75 µl of 10 % (w/v) SDS, 6 µl of TEMED, 60 µl of ammonium persulphate and 4.35 ml of water.

The samples loaded onto the cast gels contained 8 % sample and 2 % SDS-loading buffer. Per ml this cracking buffer contained 300 µl of 1 % (w/v) bromophenol blue in 50 % (v/v) glycerol, 150 µl of 2 M Tris base, 5 µl of 1 M DTT, 200 µl of 10 % SDS and 300 µl of water. Prior to loading onto gels, samples were heated at 100 °C for two minutes. Apparent molecular weight markers were used as standards and run along with samples. The standards used were contained in the Low Molecular Weight Electrophoresis Calibration Kit from Pharmacia.
Electrophoresis was performed at a constant voltage of 140 V for approximately 5 h. At the end of this time gels were removed from their mold and stained for an hour in a solution (500 ml) of Brilliant Blue R (0.4 % (w/v); Aldrich) in 50 % (v/v) methanol, 10 % (v/v) acetic acid in water. The gels were destained using 500 ml of 10 % (v/v) acetic acid and 10 % (v/v) isopropanol in water.

7.3 Insertion of NdeI restriction sites via oligonucleotide-directed mutagenesis

The DNA containing the glnLG genes was extracted from the Kohara library clone 8D12 of the λ bacteriophage EMBL4 (provided by Mr J. Crowther). Digestion of this clone with StuI yielded a 3038 bp fragment containing glnLG, which was isolated and ligated into the vector pPL450 (Love et al., 1996), which had earlier been digested with SmaI and 5'-dephosphorylated with calf intestinal phosphatase (CIP). The resulting new plasmid pPS742 (Figure 7.1), was then transformed into AN1459. Transformants were selected by ampicillin resistance at 30 °C. Plasmid DNA was isolated and run on an agarose gel to check that pPS742 was 7.2 kbp. Digestion of pPS742 with SalI indicated the regeneration and hence correctly orientated insertion of the fragment containing the glnLG genes.

Preparation of single-stranded DNA for site-directed mutagenesis involved firstly the insertion of the glnLG+ fragment from pPS742 into the phagemid vector pTZ19U (Mead et al., 1986). The vector pPS742 was
Figure 7.1 Construction of the expression plasmid pPS742. This plasmid was formed by insertion of the \textit{glnL} and \textit{glnG} genes into a SmaI restriction site found within the polylinker region of the pPL450 vector (Lilley \textit{et al.}, 1993).
digested with the restriction endonucleases XhoI and EcoRI, and the resulting 3013 bp fragment containing glnLG was isolated. This fragment was then ligated into pTZ19U which had been previously digested with SalI and EcoRI. The newly formed plasmid pPS745 was selected in AN2666 at 37 °C.

Creation of Ndel restriction sites (5'-CATATG) at the beginning of both the glnL and glnG genes would allow easy manipulation and insertion of these genes into a number of existing expression vectors. To achieve this, two different 27-mer oligonucleotides were synthesised by the Biomolecular Resource Facility at the ANU. The sequences of these oligonucleotides are given in Figure 7.2. Site-directed mutagenesis (Taylor et al., 1985; Nakamaye & Eckstein, 1986) using the Sculptor™ in vitro mutagenesis system provided by Amersham (RPN 1526) was carried out following the suppliers instructions.

The successful creation of the two NdeI endonuclease restriction sites into pPS745 was confirmed by NdeI/KpnI digestion of DNA extracted from AN1459 transformants of pPS745. A physical map of the subsequently formed vector pPS747 is shown in Figure 7.3.

7.4 Overexpression of NRII, gene product of glnL

7.4.1 Expression from pND707
The glnL gene was isolated from pPS747 following its digestion with NdeI. The 1061 bp glnL+ fragment was isolated and ligated with the heat-inducible expression vector pND707 (Love et al., 1996), that had been previously linearized with NdeI (for 1.5 h at 37 °C). Insertion of glnL into the NdeI site contained in the polylinker region of pND707, placed this gene under the control of the strong promoters P_R and P_L,
Figure 7.2 Oligonucleotides synthesised for use in oligonucleotide-directed mutagenesis. RSCPS1 was used to place an NdeI restriction site at the beginning of the \textit{glnL} gene. RSCPS2 was used to place an NdeI restriction site at the end of \textit{glnL}, which was also the beginning of the \textit{glnG} gene.
Figure 7.3 Map of the plasmid pPS747, generated by oligonucleotide-directed mutagenesis of pPS745. It differs by having \textit{Ndel} restriction sites at the beginning and end of \textit{glnL} and at the beginning of \textit{glnG}.

\textbf{pPS747}  
5843 bp
and a ribosome binding site. The resulting 5345 bp plasmid was then transformed into AN1459; ampicillin-resistant transformants were selected at 30 °C. Plasmid DNA isolation and subsequent visualisation on an agarose gel identified the transformants that contained plasmids with a single insert. *EcoRI/XbaI* digestion was then used to determine the orientation of the insertion. The construction of the new plasmid pPS748 is illustrated in Figure 7.4.

Overproduction of the desired protein was tested at 42 °C and a subsequent time-course overproduction experiment indicated an optimal time of 3 h expression of *glnL* for cell growths (Figure 7.5).

### 7.4.2 Expression in pCL476

Overproduction of a chimeric version of the *glnL* product with six additional histidine residues inserted directly after the N-terminal methionine was undertaken. The additional histidine residues would offer the possibility of easy purification and heavy-atom derivitisation. An overexpression vector that produced such a chimeric protein (pCL476) was kindly provided by Mr C. Love (Love *et al.*, 1996).

The strategy for construction of a vector to overproduce N-Met-(His)₆-NR₁₂ parallels that for overproduction of NR₁₂ in pND707. The isolated 1061 bp fragment from an *NdeI* digestion of pPS747 was ligated into pCL476 that had been previously digested with *NdeI*. This 5333 bp vector was then transformed into AN1459, transformants were selected as being resistant to ampicillin at 30 °C. Visualisation of an appropriate size band on an agarose gel indicated only one insert had been incorporated into the *NdeI*-digested pCL476. Correct orientation of the insert was confirmed by digestion with *EcoRI* and *XbaI*. This product plasmid was designated pPS749 (Figure 7.6) and
Figure 7.4 Construction of the expression vector pPS748. This 5345 bp heat-inducible plasmid was generated by insertion of the \textit{glnL} gene into \textit{NdeI} restriction site in the polylinker region of pND707.
Figure 7.5 A 10% SDS-PAGE gel showing the time-course of induction of expression of the \textit{glnL} gene, NR\textsubscript{II}, directed by the heat-inducible plasmid pPS748. Shown are the proteins expressed by this vector at 30 °C and at 42 °C over a five-hour period. Molecular weight markers (kDa) are also shown.
Figure 7.6  The vector pPS749 for the overexpression of the \textit{glnL} gene product, NR\textsubscript{H}, with an N-terminal Met-(His)\textsubscript{6} extension. The \textit{glnL} gene has been inserted within the \textit{NdeI} restriction site in the polylinker region of pCL476 (Love et al., 1996).
overproduction of N-Met-(His)$_6$-NR$\text{II}$ was tested at 42 °C. A monitored
time-course of induction indicated a 3 h induction period of expression
to be optimal (Figure 7.7).

7.5 Purification of NR$\text{II}$

Purification of unmodified NR$\text{II}$ was from the plasmid pPS748
described earlier. The previously described purification protocols
(Ninfa et al., 1986; Keener & Kustu, 1988) were found not to be
appropriate for this vector due to its increased levels of expression.
Adaptation of these two published purification procedures resulted in a
method for purification of NR$\text{II}$.

Three litres of cells were grown to OD$_{595} = 0.5$ in LBTA at 30 °C and
then heated for 3 h at 42 °C. Cells were collected by centrifugation for
15 min. at 4 °C and 7000 rpm (GS-3 rotor). Cell pellets were frozen and
stored overnight at -70 °C, resuspended to OD = 50 in Buffer A (20 mM
N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid (HEPES) [pH 7.0],
1 mM βME, 20 % v/v glycerol and 600 mM KCl) and then lysed using a
French Press at 12,000 p.s.i. Cell debris were removed by
centrifugation (75 min. at 19,000 rpm (SS-34 rotor) and 4 °C). The
resulting supernatant was treated with ammonium sulfate (AS) at
0.2 g/ml. Ground AS was added over a 30 min. period and allowed to
dissolve for one hour. The precipitated protein was then collected by
centrifugation (60 min. at 15,000 rpm (SS-34 rotor) at 4 °C), dissolved in
a minimum volume of Buffer A and then dialysed against 1 litre of
Buffer A (2 changes). Figure 7.8 shows a 10 % SDS-polyacrylamide gel
indicating the separation achieved at the various stages of purification
of NR$\text{II}$.
Figure 7.7 A 10 % SDS-PAGE gel showing the time course of induction of chimeric protein N-Met-(His)$_6$-NR$_{II}$ from the expression plasmid pPS750. The proteins expressed at 30 °C and at 42 °C over a period of five-hours are shown.
Figure 7.8 10% SDS-PAGE showing the initial steps of purification for the nitrogen regulatory protein NR_{II}. Shown is the expression of NR_{II} after a three hour induction at 42°C compared with uninduced cells (30°C). Also shown are the proteins collected in the supernatant following centrifugation of the French press lysate (F.P. sup.), and both the pellet and supernatant resulting from AS precipitation (A.S. pellet and sup. respectively). Molecular weight markers (kDa) are also shown.
The dialysed protein was then loaded onto a phenyl-agarose column (80 ml) and a salt gradient (600 mM - 0 mM KCl over 3 h at 120 ml/h) applied. A single peak eluted at 480 mM KCl which absorbed poorly at 280 nm. Fractions which contained NR\textsubscript{II}, as determined by SDS-PAGE, were pooled and dialysed against 1 litre of Buffer A (2 changes). The resulting dialysate was concentrated in an Amicon Centriplus\textsuperscript{TM} 30 kDa cutoff concentrator to a concentration of \(~10\ \text{mg/ml} (28.1\ \text{mg NR}_{II}/3\ \text{ml})\) as determined by the sequence-based extinction coefficient and A\textsubscript{280} measurements. Quantification of protein was carried out by measuring A\textsubscript{280} on a Varian Cary 1E UV-VIS Spectrophotometer using a 1 ml quartz cuvette (Starna) of 1 cm path length. The molar extinction coefficient for NR\textsubscript{II} at 280 nm was estimated from its amino acid composition as described by Gill and von Hippel (1989).

The presence of the \textit{glnL} product was inferred initially by visualisation of a 35 kDa protein on 10 \% SDS-PAGE gels, and then confirmed by N-terminal sequence analysis at the Biomolecular Resource Facility (BRF) at ANU. These data confirmed the sequence as that described for NR\textsubscript{II} by Miranda-Rios \textit{et al.} (1987). They also indicated that the starting methionine is post-translationally cleaved from the protein. The protein thus starts at alanine 2.

\textbf{7.6 Attempted co-crystallisation of P\textsubscript{II} with NR\textsubscript{II}}

With purified NR\textsubscript{II} we were now in a position to attempt co-crystallisation of this protein with P\textsubscript{II}. The solution of the P\textsubscript{II}/NR\textsubscript{II} structure would further describe the role of this signal transducing protein in regulating the uptake of nitrogen in \textit{E. coli}. 
The crystallisation experiments were performed using the methods described in Chapter 2. Findings by Kamberov et al. (1995) and our own crystallisation results (Edwards et al., 1996) suggested that in order to co-crystallise these proteins a P\textsubscript{II}/ATP/2-KG/NR\textsubscript{II} mixture in the ratio of 1:4:4:1 was required. Accordingly a mixture containing 93 µl of 18.4 mg/ml P\textsubscript{II} and 32 µl of P\textsubscript{II}-buffer A (Vasudevan et al., 1994) was treated with 18 µl of 200 mM ATP and then 16 µl of 200 mM 2-KG. This mixture was left on ice for 5 min. and then to it was added 33.15 µl of 10.71 mg/ml NR\textsubscript{II}. At no stage during the mixing of proteins and ligands was any precipitant observed.

For the crystallisation attempts described, 4 µl of the P\textsubscript{II}/ATP/2-KG/NR\textsubscript{II} mixture was combined with 4 µl of the appropriate well solution. This 8 µl drop was then either placed on a cover-slip or on a HR Micro-Bridge (HR3-312) as described previously (Chapter 2). Crystallisation using the Crystal Screen II™ (Hampton Research) were initially attempted. After 4 weeks at 4 °C light to heavy precipitation had formed in half of the screen conditions (#1-18) while for the remainder there was no visible signs of any solid material.

Further attempts to co-crystallise P\textsubscript{II} with NR\textsubscript{II} are continuing in an ongoing attempt to more fully understand and describe the role P\textsubscript{II} has with NR\textsubscript{II} that brings about part of its signal transducing function.

7.7 Overexpression of NR\textsubscript{I}, gene product of \textit{glnG}

7.7.1 Expression in \textit{pPL450}

The plasmid pPS742 was tested for the expression of NR\textsubscript{I} in two hour trial inductions. Figure 7.9A shows the production of NR\textsubscript{I} from AN1459/pPS742 and also the minor expression of the gene product of
Figure 7.9 Overexpression of the *glnG* gene from three different plasmids. Molecular weight markers (kDa) are shown for each figure. (A) Expression of NR\textsubscript{I} from the expression plasmid pPS742 at 30 °C and at 42 °C. Minor expression of NR\textsubscript{II} was also detected from this plasmid at 42 °C. (B) Expression of NR\textsubscript{I} from the expression plasmid pPS743 at 30 °C and at 42 °C. (C) Expression of NR\textsubscript{I} from the expression plasmid pPS744 at 30 °C and at 42 °C.
In order to have expression of *glnG* only, *glnL* was removed from pPS742 by two similar procedures. In the first pPS742 was digested with *SalI*, followed by isolation of a 6386 bp band which contained both the important overexpression components from pPL450 and the *glnG* gene. This fragment was reclosed with T4 ligase and the resulting vector pPS743 transformed into AN1459 (Figure 7.10). Ampicillin-resistant transformants selected at 30 °C induced expression of *glnG* at 42 °C over two hours, to overproduce the protein NR₁ (Figure 7.9B).

The second approach to sole overexpression of NR₁ paralleled the first. pPS742 was digested with *SalI* and *NotI* (Figure 7.11). These digests resulted in a 6012 bp fragment which was subsequently end-filled and reclosed. The new vector pPS744 was transformed into AN1459 (Figure 7.11). Ampicillin-resistant transformants were selected at 30 °C and transcription of *glnG* was successfully induced at 42 °C over a two-hour period (Figure 7.9C).

AN1459/pPS744 was used by other members of the group for further investigations, including purification of the NR₁ protein.

### 7.7.2 Expression in pND707

The construction of a pND707-based vector for *glnG* expression followed analogous steps to those involved in the construction of the pND707-based vector for *glnL* expression. The *glnG* gene was isolated as a 1671 bp fragment from an *NdeI*/*KpnI* digestion of pPS747 (Figure 7.3). This fragment was then ligated with the vector fragment resulting from an *NdeI*/*KpnI* digestion of pND707 and the resultant plasmid pPS791 (Figure 7.12) was transformed into AN1459. Ampicillin resistant transformants were selected at 30 °C and plasmid
Figure 7.10 Construction of the vector pPS742 for the overexpression of the glnG gene product, NR1. The majority of the glnL gene was removed by digestion of the pPS742 vector with Sall.
1. SalI/NotI digestion
2. Isolation of linearised vector
3. Reclose vector

Figure 7.11 Construction of the vector pPS744 for the overexpression of the glnG gene product, NR1. The glnL gene was removed by digestion of the pPS742 vector with SalI and NotI. This double digestion of pPS742 removed more of the glnL gene than was removed by digestion with SalI alone (Figure 7.10).
Figure 7.12 Construction of the 5955 bp vector pPS791. This expression vector was generated by insertion of the 1671 bp \textit{glnG} gene between the \textit{NdeI} and \textit{KpnI} restriction sites in the polylinker region of the vector pND707.
DNA was extracted. Regeneration of both NdeI and KpnI restriction sites confirmed the correct orientation of glnG within the plasmid. Subsequent two-hour trial inductions indicated inducible production of the protein NR1. Time course inductions over five hours indicated that a three-hour induction would be appropriate for cell growth for preparation of protein NR1 (Figure 7.13).

7.7.3 Expression in pETMCS-I

Selective expression of the glnG gene was also achieved using the bacteriophage T7 RNA polymerase system designed by Studier and Moffatt (1986). In these systems, genes cloned into the polylinker region of the vector are under the control of T7 RNA polymerase, which is activated by IPTG. The pETMCS-I vector used was digested with NdeI and KpnI and ligated with the 1671 bp fragment isolated from the NdeI/KpnI digest of pPS747 (Figure 7.12), to form the new 6302 bp plasmid pPS793. Ampicillin-resistant AN1459 transformants were selected at 37 °C and plasmid DNA was obtained through alkaline extraction. Digestion of pPS793 with NdeI and KpnI confirmed the correct insertion of the glnG+ fragment into pETMCS-I. Plasmid DNA was again obtained by alkaline extraction and then retransfomed into BL21::DE3/pLysS. Transformants were screened for both ampicillin and chloramphenicol resistance at 37 °C. Two-hour trial inductions carried out using 20 µl of 500 mM IPTG per 10 ml LBTAC indicated successful overexpression of glnG. Three hours induction with IPTG yielded optimal overproduction of native NR1. The level of expression for this time period were similar to those shown in Figure 7.13.
Figure 7.13 A 10% SDS-PAGE gel showing the time course of induction for the NR1 protein from the expression plasmid pPS791. The proteins expressed at 30 °C and at 42 °C over a period of three-hours are shown.
7.7.4 Expression in pCL476

The expression of NR_I with a methionine and six histidine residues attached to its N-terminal methionine residue (N-Met-(His)_6-NR_I) was achieved using a procedure similar to that used for N-Met-(His)_6-NR_{II}. The difference between overexpression of N-Met-(His)_6-NR_I and its NR_{II} equivalent, was the initial ligation of the respective genes into pCL476. Digestion of pPS747 with NdeI and KpnI resulted in a 1671 bp fragment containing glnG. This fragment was then inserted into the pCL476 vector which had also been previously digested with NdeI and KpnI. The resultant plasmid was designated pPS792. After selecting successful transformants, a three-hour period of induction at 42 °C was seen to be optimal for overexpression of N-Met-(His)_6-NR_I (data not shown). The level of expression of this protein from pPS792 was similar to the levels of N-Met-(His)_6-NR_{II} expressed by the vector pPS750 (Figure 7.7).

Due to time constraints the AN1459/pPS791, AN1459/pPS793 and AN1459/pPS792 strains were given to other members of our lab to purify NR_I and N-Met-(His)_6-NR_I.
Chapter 8. Conclusions and Discussions

8.1 Role of P_{II}-UMP in crystallisation of P_{II}

At the beginning of this work the crystals of native P_{II} were contaminated by considerable levels of P_{II}-UMP. These crystals had a hexagonal morphology (Figure 2.3A). The X-ray data collected on these crystals revealed that they suffered from twinning by merohedry, with values for twin fraction (t.f.) between 0.1 and 0.3. Improved separation of P_{II} and P_{II}-UMP yielded more pure samples of native protein to be used in crystallisation. The resultant crystals had altered morphologies (Figure 2.3B) which gave rise to appreciable crystal slippage during data collection. This problem was overcome by the use of a special apparatus (Figure 3.1) which secured the crystal without interfering with X-ray diffraction.

Data collected from these crystals indicated t.f. values of less than 0.1. It seems apparent that the increasing levels of P_{II}-UMP in the crystallisation of P_{II} altered the crystal morphology and with that change came varying levels of twinning. While smaller diamond shaped crystals did suffer slippage, they exhibited lower values for t.f. than the elongated hexagonal crystals, which contained higher levels of P_{II}-UMP. Although not without exception, this empirical "rule-of-thumb" in general allowed the selection of crystals of lower t.f. Native and heavy-atom derivative data were collected which yielded sufficiently accurate phases to be determined for the structural solution of P_{II}.

Six heavy-atom derivative crystals were used to refine three heavy-atom positions and obtain MIR phases to 2.7 Å. This led to our initial model
for the structure of P11 (Cheah et al., 1994). This structure was further refined to 1.9 Å (Carr et al., 1996).

8.2 Structure of P11

The P11 monomer at 1.9 Å is comprised of 6 strands and 2 α-helices. Four of the strands are involved in an anti-parallel β-sheet which has 2 anti-parallel α-helices behind it (Figure 5.2). Tyr51, the residue post-translationally modified via uridylylation, is found on a stretch of random coil denoted the T-loop (residues 37-55). The central region of the monomer forms a double β-α-β motif which has been seen in a number of proteins that are known to bind phosphates or nucleotides (Swindells et al., 1993).

Between strands 1 and 3 of the monomer an approximate 2-fold axis can be seen. It is the existence of this additional symmetry that is the major contributor to the twinning observed for P11. Given that only 1 molecule per a.s.u. was determined for P11 the twinning had to arise from within the molecule and not from the P63 symmetry that had ordered its packing. In light of this structure the relationship between the level of P11-UMP contamination and the t.f. observed, it seems that the addition of the UMP moiety at Tyr51 alters the structure in such a way that the approximate 2-fold symmetry approaches a real dyad. Further work in this laboratory will determine if this is in fact the case.

Both in solution and in the crystalline state P11 exists as a homotrimer. P11 monomers interact around the three fold axis in a tricuspid arrangement with the helices facing outside of the central β-sheet (Figure 5.4). This arrangement allows interaction of the central β-sheet with 2 strands from each of the other two subunits. This interaction results in the formation of three 8-stranded β-sheets. The trimer—
the exception of the T-loops—is a compact barrel of about \(50 \, \text{Å} \times 30 \, \text{Å}\). These loops position Tyr51—the residue modified post-translation—approximately \(13 \, \text{Å}\) away from the main body of the molecule.

The positioning of Tyr51 on a flexible loop must be to facilitate a particular role within the signal transduction function of P\(_{II}\). This T-loop is most likely involved in protein-protein interactions, with either one or all of the NR\(_{II}\), UTase/UR or ATase enzymes. This loop region is also well conserved across the P\(_{II}\) proteins found in nine different bacteria. This suggests that this loop has an important role in the function of P\(_{II}\). The other two regions of this molecule that might possibly be involved with other proteins are the cleft—formed at the junction of two subunits (Figure 4.10)—and the flat, highly conserved upper surface of the trimer (Figure 4.13). Any or all of the P\(_{II}\) features mentioned above could be involved in protein-protein interactions. The highly conserved nature of the T-loop presents an appropriate ‘flag’ for signal transduction, while the flat conserved surface at the base of these loops suggests an area for interactions with other protein.

The nature of interactions of P\(_{II}\) with NR\(_{II}\) will continue to be investigated in this laboratory in an attempt to determine the regions of this molecule involved with signalling the changes of nitrogen availability. The high levels of conservation of the P\(_{II}\) proteins from various microorganisms (Figure 4.12) suggest that mechanistic insights found for \textit{E. coli} P\(_{II}\) might suggest similar paths for these homologues, especially for the trimeric P\(_{II}\) from \textit{Synechococcus} sp..

\textbf{8.3 Relevance of the double} \(\beta\)-\(\alpha\)-\(\beta\) \textbf{motif in P}\(_{II}\)

The interactions of P\(_{II}\) with a number of different effectors has been known for some time and more recently elucidated further by Kamberov
et al. (1995). Depending on the nitrogen status of the cell PII will interact with 2-KG, glutamate or with ATP and one of the two ligands mentioned. While it is known that the presence of ATP is required for high affinity binding of either 2-KG or glutamate to PII, there is no information that has indicated where these interactions might occur within the structure of this protein. Having solved the X-ray structure of *E. coli* PII and found that it contains a double β-α-β motif, there are a number of potential areas that seem likely for this protein to bind these ligands, particularly ATP and 2-KG, and also to interact with other proteins.

The first region of interaction of either ATP or 2-KG might possibly be across the face of the β-sheet of any particular monomer. Of the six proteins compared in Chapter 5 four interact in this fashion. The S6, NDPK, and ar-ATCase proteins bind or interact with phosphates across the face of their β-sheet, which are extended upon complex formation or oligomerisation. The structure of the PII trimer suggests that such interactions would be unlikely. The tricuspid arrangement of PII (Figure 4.8) points the faces of the β-sheets inwards making these sheets inaccessible to either proteins or ligands.

The second possible region of protein or ligand interaction might be with the amino-terminal end of the A helix. As with most of the proteins described here, upon oligomerisation the E2 DNA-binding domain of BPV-1 forms an extended β-sheet but it is the amino-terminal end of its ‘A helix’ that interacts with DNA. These interactions are through the dipole created within the helix and help this protein to bind to the phosphate backbone of the DNA. It seems unlikely that this type of interaction is taking place between PII and ATP or 2-KG. The ‘open’
and conserved nature of this area of the trimer may be involved in interactions with other proteins.

By analogy with the discussed examples of proteins containing a double β-α-β motif, a third possible area for ligand interaction might involve both the β-sheet and the amino-terminal end of the A-helix. Ribosomal RNA is bound in this fashion to U1 snRNP A. Dimer contacts are made through helices in U1 snRNP A and the rRNA binds between β3 strand and α1 helix (Table 6.3). In the trimeric structure of P11 however, this region is partly covered (Figure 4.10) and would be unlikely to facilitate interaction with either ATP or 2-KG.

The most likely regions for interaction of ATP and 2-KG with P11 would be the clefts created by formation of the trimer (section 4.6.2 and Figure 4.10). These regions are highly conserved across the P11 species and contains appropriate electrostatics to facilitate such interactions. Given that ATP and 2-KG are required to bind to P11 in order for Tyr51 to be uridylylated it seems likely that alterations in the cleft upon binding these two ligands positions the T-loop in such a way to promote and accommodate uridylylation. Further study in this laboratory will continue to investigate these interactions in an attempt to gain further insight conformational changes in P11 brought about by binding of ATP and 2-KG.

The overall picture that emerges is that the double β-α-β motif does present a stable framework for the interaction of various proteins with a number of different ligands containing phosphates or phosphate esters. Historically these interactions have been described as involving the β-sheet, the amino-terminal end of the first α-helix or a combination of
159

the two. The structure of the P11 trimer suggests that binding of ATP and 2-KG does not involve these types of interactions. More probable is the binding of these two ligands within the cleft regions of the P11 trimer, which consequently affects the position of Tyr51. Further work on co-crystals of P11/ATP/2-KG in this laboratory will yield greater insight into the nature of this motif and this protein in binding phosphate-containing ligands.

8.4 Overall conclusions

The work presented in this thesis has culminated in the structure determination of the P11 protein from *Escherichia coli*. This protein contains six strands and two helices, which are arranged as a four-stranded anti-parallel β-sheet with two anti-parallel α-helices positioned behind the sheet. This β-sheet is involved in the formation of a double β-α-β motif.

The primary goal of this investigation was to solve the structure of P11. This goal was achieved with solution of the structure to 2.7 Å resolution. This model then formed the basis for further refinement of P11 to 1.9 Å. The secondary goals of this study focused on trying to further describe the function of P11 by elucidating its interactions with 2-KG, glutamate and also with NR11. The initial results from these investigations have been promising, but require further work to describe, at an atomic level, the interactions P11 has with these ligands and the NR11 protein. One result that arose from these initial studies was that the structure of P11 at pH 9.5 was found not to differ appreciably from the structure at physiological pH. Given the successful co-crystallisation of P11 with a number of ligands at pH 9.5, this result indicates the changes upon ligand binding are a
consequence of their presence, and not the pH at which the binding took place.

With the solution of the P\textsubscript{II} structure comes the need for description of the structure with respect to its function. In relation to protein-protein interaction, P\textsubscript{II} has the T-loop, conserved upper surface and cleft regions of the trimer which offer potential sites. The double β-α-β motif contained in P\textsubscript{II} seems unlikely to be involved in the binding of ATP and 2-KG, with the cleft region within the trimer being a more probable candidate. As further co-crystallised structures are solved in this laboratory a more detailed picture will arise indicating how P\textsubscript{II} achieves its multipurpose function as a signal transduction protein and how its interactions with ATP relate to the present picture of phosphate binding in proteins containing double β-α-β motifs.
Appendix A. "Two-component" Regulatory Systems in *E. coli*

A.1 Introduction
As discussed in Chapter 1 PII functions as a signal transducing protein in a bicyclic cascade mechanism that regulates nitrogen uptake in the cell. While being individual in its primary function, there are a number of other bicyclic cascade mechanisms or "two-component" regulatory systems (Nixon *et al.*, 1986; Ronson *et al.*, 1987) which serve their purposes similarly. Each of these systems have a kinase/phosphatase and a response regulator (Ninfa, 1991), but the signalling for their activity varies widely.

In this chapter I will describe three other "two-component" regulatory systems with respect to regulation of the Ntr regulon in order to draw comparisons between their signalling processes and that of PII. The regulatory systems I will discuss are chemotaxis, phosphate regulation and osmoregulation. Each of these cases are discussed for *Escherichia coli*. The discussion presented here is a brief summary for the purposes of general comparison with the Ntr system. It is only the NR1 and NR11 proteins that are the direct interest of this study. As such, fuller explanations of symbols and detailed definitions are to be found in Stock *et al.* (1989).

A.2 Chemotaxis
Flagellated bacteria such as *E. coli* have the ability to move in a tumbling manner toward higher concentrations of nutrients (attractants) or away from concentrations of noxious chemicals (repellents). These bacteria do this through the process of chemotaxis. Chemotaxis has been well investigated and more recently reviewed in Ninfa (1991). Figure A.1 summarises the main points involved in
Figure A.1 Two-component pathway for the regulation of chemotaxis in *E. coli*. Tar, Tsr, Trg and Tap are membrane chemoreceptor-transducer proteins. FliM, FliN and FliG form a complex that acts as the flagellar switching complex. Adapted from Stock *et al.*, 1989.
chemotaxis. Upon stimulation by attractants or repellents the receptor-transducer proteins—Tar, Tsr, Trg and Tap—bind CheW and CheA. CheA autophosphorylates at His48 in its sequence (Hess et al., 1988) and when CheY binds to it the phosphoryl group is transferred from CheA to Asp57 of CheY (Sanders et al., 1989). P-CheY then interacts with the flagella switch complex—FliM/FliN/FliG—to bring about the appropriate mode of tumbling.

The NR_{II}- and NR_{I}-like components in the chemotaxis mechanism are easily identified. It seems apparent that CheA is acting as the NR_{II}-like kinase bringing about phosphorylation of the NR_{I}-like response-regulator CheY. From its phosphatase-like interaction with CheY-P it is apparent that CheZ is also acting as a NR_{II} analogue. CheZ brings about reduced interaction of CheY-P with the flagellar motor components to reduce tumbling, but does not have a role in promoting the interaction of CheY-P with the tumbling flagellar motor complex.

Despite this system bearing many similarities to the regulation of the Ntr regulon, in terms of function, there is no known P_{II} analogue. While NR_{II} and NR_{I} share similarities with CheA and CheY, in their respective functions, domains and sequences (Ninfa, 1991), the systems they help to regulate differ markedly. It appears that this system has lost the need for a signalling protein in controlling cell movement.

A.3 Phosphate regulation

There are several *E. coli* proteins that assist in or actively transport phosphate through the cell envelope and into the cytoplasm. High-affinity phosphate transport systems like the phosphate specific transport (PST) system (Nakata et al., 1987), actively carry P\textsubscript{i} across the
inner and outer cell membranes. A combination of anion-specific porins, such as PhoE, periplasmic alkaline phosphatase (PhoA) and phosphate binding proteins, such as PstS, provide an access route for environmental Pᵢ into the cytoplasm. The mechanism for Pᵢ transport involving PhoE and PhoA is reviewed in Stock et al. (1989) and is summarised in Figure A.2.

As is the case in nitrogen regulation and chemotaxis, the proteins involved in this "two-component" system are translated from the Pho regulon (Wanner, 1987). The kinase/phosphatase in this system is PhoR and the response regulator, PhoB. In conjunction with the PST system PhoU acts to reduce transcription of the Pho regulon. Although the role of PhoU is uncertain, given its observed interaction with PhoR/PhoB, it seems most probable that PhoU binds to PhoR and as such stimulates the phosphatase activity of PhoR. In this scenario PhoU would be functioning in an analogous sense to the interaction of P_I with NR_{II}, to bring about de-phosphorylation of the response regulator.

A major way in which the roles of P_{II} and PhoU differ is that P_{II} has an active role in regulation of GS activity under all environmental conditions, either at an activational or transcriptional level (Chapter 1). In contrast PhoU is not essential for the uptake of Pᵢ, only for phosphate repression via repressed transcription of the Pho regulon (Makino et al., 1985). It seems that the control of nitrogen assimilation has been more tightly controlled with regards to the signalling mechanism that regulates the final level of GS activity. This may be due in part to the existence of other pathways that can assist in
Transcriptional activation of the Pho regulon

Figure A.2  Schematic diagram for the mechanism used to regulate uptake of phosphate in *E. coli*. Adapted from Stock *et al*., 1989.
phosphate regulation, as opposed to the crucial role GS plays under conditions of limited environmental ammonia. One pathway that assists phosphate regulation is osmoregulation.

A.4 Osmoregulation

In bacteria, such as *E. coli*, the nature of the inner membrane does not support an osmotic pressure gradient and so the cell wall-outer membrane does this via osmoregulation. The process of osmoregulation was reviewed by Csonka (1989) and will be discussed to allow comparisons with the “two-component” nitrogen regulatory pathway in *E. coli*. Within this comparison the analogies with the Che and Pho regulatory mechanisms will be made. A schematic diagram outlining osmoregulation is shown in Figure A.3.

The inner membranes of *E. coli* have active transport systems and receptors to move osmotic molecules both in and out of the cytoplasm. In response to environmental signals (Cronan, Jr *et al.*, 1987), the membrane bound EnvZ protein, via OmpR-P, regulates transcription of the permeability of the cell to various solutes in the surrounding medium (reviewed in Csonka, 1989; Forst & Inouye, 1988). OmpF and OmpC are the porin homotrimers that form channels and thus regulate permeability properties of the cell (Nikaido & Vaara, 1985).

EnvZ is an inner membrane protein that functions as an osmosensor (Hall & Silhavy, 1981) through its autophosphorylation at His243 (Aiba *et al.*, 1989). EnvZ-P then phosphorylates OmpR. High levels of OmpR-P are thought to cause repression of *ompF* expression and increased expression of *ompC* by the binding of OmpR-P to low-affinity sites (Jo *et al.*, 1986). Expression from the *ompC* promoter stimulates transcription of *micF* RNA which acts to decrease the synthesis of
Figure A.3 Schematic diagram illustrating osmoregulation in *E. coli*. Regulation of *ompF* and *ompC* is via interaction of OmpR-P with high-affinity (open box) or low-affinity (filled-box) promoters situated upstream of these genes. This figure was adapted from Stock *et al.*, 1991.
OmpF (Matsuyama & Mizushima, 1985; Mizumo et al., 1984). When there is little Omp-P, activated transcription of ompF occurs via the ompF promoter (Figure A.3).

With this "two-component" mechanism the kinase is EnvZ and the response regulator is OmpR. The roles of these proteins closely parallel those of NRII and NR1, respectively, from the Ntr regulation pathway. Despite this there appears to be no phosphatase-facilitating function analogous to the role of PII in nitrogen regulation or PhoU in phosphoregulation. However, as is the case with NRII, the cytoplasmic domain of EnvZ can in the presence of ATP catalyse dephosphorylation of Omp-P (Aiba et al., 1989). The lack of PII-analogue in osmoregulation may be due to the fact that OmpF and OmpC tend to be regulated in a reciprocal fashion, which maintains a constant total level of porin protein (Lugtenberg et al., 1976). This control must be sufficient for osmoregulation whereas for nitrogen uptake, in the form of glutamine, a faster, more responsive regulation pathway is required. Similar reasoning might explain the role of PhoU in phosphoregulation.

It seems that regulation of nitrogen assimilation requires an additional level of control, which is maintained by the PII protein. Regulation of chemotaxis, phosphate uptake and osmotic balance appear not to need such a finely tuned metabolic switch. The supplementary nature of phosphate regulation and osmoregulation may be the reason why a PII-like protein is not required for pathway control. The tumbling towards or away from certain environmental conditions does not require quick reversal of the chemotaxis
mechanism and therefore would not need a metabolic switch similar to that required for regulation of levels of active GS.
References


metabolic transformation of the P$_{11}$-regulatory protein. 


Steigemann, W. (1992). PROTEIN v.3.1, Max-Planck Institut fuer Biochemie, Martinsried bei Muenchen, FRG.


