

The University Library



*The Library of the
Department of Biochemistry
and Molecular Biology
ASTBURY BUILDING*

STRUCTURAL STUDIES OF
THE ARGININE REPRESSOR/ACTIVATOR
FROM *Bacillus subtilis*.

NICKOLAOS MENELAOU GLYKOS

Submitted in accordance with the requirements
for the degree of Doctor of Philosophy

The University of Leeds
The Department of Biochemistry and Molecular Biology

February 1995

The candidate confirms that the work submitted is his own
and that appropriate credit has been given where reference
has been made to the work of others.

Abstract

In the presence of L-Arginine, AhrC —the Arginine-dependent Repressor/Activator from *Bacillus subtilis*— represses the transcription of the genes encoding the anabolic and activates those encoding the catabolic enzymes of arginine metabolism. AhrC is a homohexamer of total molecular mass 105 kDa. It shows no homology to any of the characterised DNA-binding motifs or DNA-binding proteins with the exception of ArgR, the Arginine Repressor from *Escherichia coli*. ArgR does not act as a transcription activator but it has been shown to be a necessary accessory protein for the resolution —through site-specific recombination— of multimers of the ColE1 plasmid. Although the two proteins share only 29% identity and are from such taxonomically distinct prokaryotes, AhrC can complement *E. coli* ArgR⁻ strains both in the regulation of Arginine metabolism and the resolution of the ColE1 plasmid.

This thesis describes our attempts to determine the crystal structure of AhrC. Three different crystal forms have been produced and characterised. Useful derivatives have been prepared for two of these forms but the determination of their heavy atom structures proved impossible. An attempt to determine the low resolution structure of AhrC using Electron Microscopy has been unsuccessful. Molecular Replacement using as a search model the crystal structure of the hexameric core fragment of ArgR also failed to give a convincing solution.

Acknowledgements

I should like to thank Professor Simon Phillips for his advice, encouragement and support throughout the course of my studies.

Dr Amalia Aggeli for useful discussions and for recording and analysing a FTIR spectrum of AhrC.

Don Akrigg for his help with computational aspects of this work.

Les Child for the photographic work.

Dr Máire Convery for useful discussions and for making valuable comments on the first draft of this thesis.

Dr Greg van Duyne and Professor Paul Sigler for making the atomic coordinates of their ArgR model available to me before publication.

Professor Stavros Hamodrakas for suggesting Leeds and for his continued interest in my studies.

Carol Holtham, Coleen Miller, Isobel Parsons and Andrew Walsh for their help with the purification of AhrC and of the AhrC-ArgR chimaera.

Dr Andreas Holzenburg for useful discussions, especially about the Electron Microscopical studies of AhrC, and for reading critically Chapter 6.

Paul McPhie for his help with the Electron Microscopes.

Dr Mark Parsons for useful discussions and for making valuable comments on the manuscript.

Dr Jonathan Porter for useful discussions.

My family for their continued encouragement and support.

Table of Contents

| | |
|--|-------------|
| Abstract | i |
| Acknowledgements | ii |
| Table of Contents | iv |
| List of Figures | viii |
| List of Tables | xix |
| List of Acronyms | xx |
| List of Symbols | xxi |
| | |
| 1 Introduction | 1 |
| 1.1 Biochemical and Genetical Data | 7 |
| 1.2 Biophysical Data | 13 |
| | |
| 2 Protein Preparation, Crystallisation, and Preliminary Characterisation of AhrC Crystals | 17 |
| 2.1 Protein Preparation | 17 |
| 2.1.1 Protein Purification | 17 |
| 2.1.2 Stability of AhrC During Storage | 20 |
| 2.2 Crystallisation, and Preliminary Characterisation of AhrC Crystals | 21 |
| 2.2.1 Orthorhombic Form | 24 |

| | | |
|----------|---|-----------|
| 2.2.2 | Monoclinic Form | 28 |
| 2.2.3 | Trigonal Form | 31 |
| 3 | Preparation and | |
| | Preliminary Analysis of Heavy Atom Derivatives | 36 |
| 3.1 | Data Collection Strategy | 36 |
| 3.2 | Preparation of Heavy Atom Derivatives | 39 |
| 3.3 | Preliminary Analysis | 52 |
| 3.3.1 | The Niobium Cluster, Nb ₆ Cl ₁₄ | 53 |
| 3.3.2 | The H ₂ IrCl ₆ Derivative | 61 |
| 3.3.3 | Other Derivatives | 63 |
| 4 | Self Rotation Function Studies | |
| | and a Model of the Crystal Packing | 66 |
| 4.1 | Introduction | 66 |
| 4.2 | Self Rotation Function Studies | 70 |
| 4.2.1 | Native Crystals | 71 |
| 4.2.2 | Heavy Atom Structures | 74 |
| 4.3 | A Model of the Crystal Packing | 77 |
| 4.3.1 | Pseudo-origin Peaks | 77 |
| 4.3.2 | Very Low Resolution Permutation Syntheses | 81 |
| 4.3.3 | A Packing Function | 83 |
| 5 | Attempts to Determine the | |
| | Heavy Atom Structures of AhrC Derivatives | 86 |
| 5.1 | Introduction | 86 |
| 5.2 | Direct Methods | 87 |
| 5.2.1 | SHELXS-86 | 90 |

| | | |
|----------|--|------------|
| 5.2.2 | MULTAN77 | 92 |
| 5.2.3 | The Cochran & Douglas Method | 93 |
| 5.3 | Patterson Methods | 95 |
| 5.3.1 | HASSP | 102 |
| 5.3.2 | GROPAT | 103 |
| 5.3.3 | A Method by Nixon, P.E., 1978 | 105 |
| 5.3.4 | Image-seeking Methods | 107 |
| 5.3.5 | A Patterson Interpretation Programme : MSS | 114 |
| 5.3.6 | The Argos & Rossmann Method | 120 |
| 5.4 | Other Methods | 129 |
| 5.4.1 | Permutation Syntheses Method | 129 |
| 5.4.2 | Molecular Replacement | 133 |
| 5.4.3 | Direct Maximisation of an Electron Density Function | 134 |
| 5.4.4 | Monte Carlo Methods | 138 |
| 5.4.5 | An Attempt to Obtain Phase Information <i>ab initio</i> | 147 |
| 6 | Electron Microscopy of AhrC Crystals | 152 |
| 6.1 | Introduction | 152 |
| 6.2 | Image Reconstruction | 157 |
| 6.3 | Specimen Preparation, Data Collection and Image Processing | 162 |
| 6.4 | Characterised Projections | 165 |
| 6.4.1 | The [001] Projection | 165 |
| 6.4.2 | The [010] Projection | 170 |
| 6.4.3 | The [101] Projection | 174 |
| 6.4.4 | The [130] Projection | 176 |
| 6.5 | Other Projections | 176 |
| 6.6 | The Calculation of a 30Å Electron Density Map | 179 |

| | |
|--|------------|
| 7 Molecular Replacement | 186 |
| 7.1 The Model | 186 |
| 7.2 Orthorhombic Form | 189 |
| 7.2.1 The Programmes ALMN-TFFC and AMoRe | 189 |
| 7.2.2 X-PLOR | 190 |
| 7.3 Monoclinic Form | 195 |
| 7.3.1 A Model of the Crystal Packing | 195 |
| 7.3.2 Molecular Replacement | 200 |
| | |
| Discussion | 203 |
| Appendix A, Trigonal Form : A Redetermination | 207 |
| References | 209 |

List of Figures

| | | |
|-------------------|---|----|
| Figure 1.1 | Schematic diagrams of the crystal structures of some transcription factor-DNA complexes. | 3 |
| Figure 1.2 | Schematic diagrams of the crystal structures of some transcription factor-DNA complexes. | 4 |
| Figure 1.3 | Pathways of Arginine metabolism in <i>Bacillus</i> spp. | 8 |
| Figure 1.4 | Sequence alignment of AhrC and ArgR. | 10 |
| Figure 1.5 | Sequence alignment of the <i>E. coli</i> ARG box consensus sequence and three putative AhrC operator sequences located at the 5' region of the <i>argCJBD-cpa-argF</i> cluster. | 11 |
| Figure 1.6 | Model of how AhrC may induce the formation of a repression loop. | 12 |
| Figure 1.7 | Secondary structure prediction for AhrC. | 13 |
| Figure 1.8 | FTIR absorbance spectrum after solvent subtraction. | 14 |
| Figure 1.9 | Band-fitted Amide I region. | 16 |
| Figure 2.1 | Elution profile of the S-Sepharose column. | 19 |

| | | |
|--------------------|--|----|
| Figure 2.2 | High resolution SDS-PAGE of AhrC. | 19 |
| Figure 2.3 | SDS-PAGE of AhrC three months after purification. | 22 |
| Figure 2.4 | SDS-PAGE of AhrC crystals and precipitated AhrC four months after purification. | 22 |
| Figure 2.5 | A typical orthorhombic AhrC crystal. | 26 |
| Figure 2.6 | Orthorhombic form, 8° precession photographs of (A) the $hk0$ and (B) $h0l$ zones. | 26 |
| Figure 2.7 | A monoclinic AhrC crystal. | 30 |
| Figure 2.8 | Monoclinic form. 7° precession photograph of the $hk0$ zone. | 30 |
| Figure 2.9 | Typical trigonal AhrC crystals | 32 |
| Figure 2.10 | (A) “Still” X-ray photograph and (B) 6.5° precession photograph of a trigonal AhrC crystal ($hk0$ zone) | 33 |
| Figure 2.11 | Trigonal form, $hk1$ level. | 33 |
| Figure 2.12 | Trigonal form, $66\text{-}8\text{\AA}$ native Patterson projection along $[001]$. Four unit cells are shown. Contours every 5% of the origin peak; negative contours broken. | 35 |
| Figure 3.1 | 8° precession photographs of the $hk0$ zone from (A) Native orthorhombic AhrC crystals and (B) A crystal soaked for 20 hrs in 2 mM $\text{Nb}_6\text{Cl}_{14}$. | 56 |
| Figure 3.2 | (A) Plot of the mean fractional isomorphous difference versus resolution for an orthorhombic AhrC crystal soaked for 19 hrs in 0.7 mM $\text{Nb}_6\text{Cl}_{14}$. (B) Plot of the K_{emp} versus resolution for the same crystal. | 56 |

- Figure 3.3** 13–6Å $[E_{\Delta F_{isom}} \Delta F_{isom}]^2$ synthesis for a type II crystal 57
soaked in 0.9 mM Nb₆Cl₁₄ for 24 hrs. Contours every
2% of the origin peak.
- Figure 3.4** $v=0.0$ sections from (A) a ΔF_{isom}^2 and (B) a 59
 $[E_{\Delta F_{isom}} \Delta F_{isom}]^2$ synthesis for a type I crystal soaked
in 1.5 mM Nb₆Cl₁₄ for 21 hrs. Contours every 2% of
the origin peak.
- Figure 3.5** 8° precession photographs of the $hk0$ zone from (A) Na- 62
tive orthorhombic AhrC crystals and (B) Crystal
soaked for 12 hrs in 0.75 mM H₂IrCl₆.
- Figure 3.6** (A) Plot of the mean fractional isomorphous differ- 62
ence versus resolution for an orthorhombic AhrC crys-
tal soaked for 18 hrs in 0.8 mM H₂IrCl₆. (B) Plot of
the K_{emp} versus resolution for the same crystal.
- Figure 3.7** 13–7Å $[E_{\Delta F_{isom}} \Delta F_{isom}]^2$ Patterson synthesis for a type 64
II crystal soaked in 1.2 mM H₂IrCl₆ for 3 hrs. Contours
every 2% of the origin peak.
- Figure 3.8** Mean fractional isomorphous difference versus resolu- 65
tion for type II crystals soaked in (A) 5% sat. PICMB
for 25 hrs, (B) 0.1 mM Uranyl acetate for 1.5 hrs and
(C) 4mM KAu(CN)₄ for 16 hrs.
- Figure 4.1** Self rotation functions for native AhrC crystals. 72
- Figure 4.2** Self rotation functions for Nb₆Cl₁₄. 76
- Figure 4.3** Self rotation function for H₂IrCl₆. 77

- Figure 4.4** A : Symmetry elements in the [010] projection of orthorhombic AhrC crystals, B : Closest non-overlapping packing of 8 spherical molecules in the [010] projection. 79
- Figure 4.5** A : A hypothetical structure consisting of four atoms related by a non-crystallographic 2-fold axis parallel to a crystallographic 2-fold, B : Schematic diagram of the Patterson function of this structure. 79
- Figure 4.6** Permutation syntheses for the [010] projection using the four strongest, low resolution, $h0l$ reflections. 82
- Figure 4.7** Three sections from a packing function. 84
- Figure F.1** Additional packing arrangements. The lines enclose one unit cell. 84
- Figure 4.8** The crystal packing of the orthorhombic form. Views of the packing down the [100], [010] and [001] axes are shown. The radius of the spheres (each representing an AhrC hexamer) is 30Å. In all views 3×3 unit cells are shown. 85
- Figure 5.1** Harker sections $u=0.0$, $v=0.0$ and $w=0.5$ from a 13–7Å $[E_{\Delta F_{isom}} \Delta F_{isom}]^2$ synthesis for the Nb₆Cl₁₄ derivative. Contours every 2% of the origin peak. 96
- Figure 5.2** Implication diagram of the Nb₆Cl₁₄ derivative. Contours every 4% of the origin peak of the Patterson function. 97
- Figure 5.3** (A) [010] Patterson projection for Nb₆Cl₁₄ : 17-5Å ΔF^2 synthesis, (B) Sharpened synthesis, (C) Implication diagram using the sharpened function. 101

- Figure 5.4** Functions $y(x)$ (top curve) and $c \tanh(y(x)/c)$ (bottom curve). $y(x)$ consists of three Gaussians centered at $x=0.0$, 13.0 and -13.0 with $\sigma=1.5$, 3.0 and 5.0 correspondingly. In this example $c = 0.2$. 106
- Figure 5.5** (A) A hypothetical four-atom structure, (B) Its Patterson function, (C) The Patterson function seen as the superposition of four images of the structure. 108
- Figure 5.6** Vector superposition : The centrosymmetric case. 109
- Figure 5.7** Vector superposition : The non-centrosymmetric case. 111
- Figure 5.8** Atomic superposition : (A) A hypothetical two-atom structure in $p2$, (B) Its Patterson function, (C) Superposition of two copies of the Patterson function at the two equivalent positions of one atom (black circles). Double circles denote additional coincidences. 112
- Figure 5.9** $15\text{-}6\text{\AA}$ [001] Patterson projection for the Niobium derivative. 115
- Figure 5.10** $hk0$ weighted reciprocal lattice for $\text{Nb}_6\text{Cl}_{14}$. 115
- Figure 5.11** [001] Vector superposition functions for the Niobium derivative. 116
- Figure 5.12** [010] Atomic superposition function for the Niobium derivative. 116
- Figure 5.13** MSS : pseudocode. 118
- Figure 5.14** Argos & Rossmann, Type I search for $\text{Nb}_6\text{Cl}_{14}$, Molecular centre unknown. Contours every 1.5σ with first contour at 1.5σ above the mean. 123

- Figure 5.15** Argos & Rossmann, Type I search for Nb₆Cl₁₄. Molecular centre at $x = 0.13$, $y = 0.09$, $z = 0.225$. 124
- Figure 5.16** Argos & Rossmann, Type II search for Nb₆Cl₁₄. Three sections through the highest peak are shown. 125
- Figure 5.17** Observed and calculated Patterson functions for the Nb₆Cl₁₄ derivative. 127
- Figure 5.18** 10Å [001] projection of native AhrC crystals. Phases calculated from the Niobium derivative modelled as an octahedral arrangement of six sites. 128
- Figure 5.19** (A) A hypothetical 36 atom structure, (B) Its derivative, (C) The difference Patterson function, (D) A Fourier synthesis based on modelling the derivative as a single site at the molecular centre. 130
- Figure 5.20** [010] & [001] projections of native AhrC crystals based on modelling the Nb₆Cl₁₄ derivative as a single site at the position of the molecular centre. 131
- Figure 5.21** [001] permutation syntheses for the Niobium derivative. 133
- Figure 5.22** $u = 0.0$ Harker section from the sharpened isomorphous difference Patterson function for the Niobium derivative. Contours every 2% of the origin peak. 135
- Figure 5.23** Three sections from the translation function using BRUTE. 135
- Figure 5.24** Comparison of two sections from the observed (top row) and calculated Patterson function for a four-atom structure. 136

- Figure 5.25** Monte Carlo : Results from model calculations with a hypothetical 12-atom structure. (A) R-factor, and, (B) Free R value during minimisation, (C) The Patterson function, (D) The $F_o \exp(i\phi_c)$ synthesis. 145
- Figure 5.26** Monte Carlo : R-factor and R^{free} during two minimisations using 15-7Å $h0l$ terms for the Niobium derivative, (A) : $T = 0.007$, 12 atoms, (B) : $T = 0.005$, 16 atoms. 146
- Figure 5.27** 115–15Å $F_o \exp(i\phi_{\text{spheres}})$ synthesis. Phases from a sphere of constant density at the assumed position of the molecular centre. 148
- Figure 6.1** Linear aggregates of AhrC. 156
- Figure 6.2** Variation of the transfer function versus resolution (in nm^{-1}) for different values of Δf and $\Phi(\alpha)$: (A) 0 nm, 0 rad (B) 90,0 (C) 500,0 and (D) Superposition of 90,0 and 90,0.2 159
- Figure 6.3** Focus series of a thin, negatively stained AhrC crystal down the [001] axis. Approximate defocus values are (A) in focus, (B) 640 nm, (C) 1200 nm, (D) 1900 nm, (E) 3250 nm, (F) > 3500 nm. 161
- Figure 6.4** Electron micrograph of crushed, negatively stained AhrC crystals. 163
- Figure 6.5** Modulus of the Fourier transform of a well preserved image of the [001] projection. The positions of two weak but observed reflections corresponding to spacings of 21.5 (530) and 18.3Å (040) are indicated. 166

- Figure 6.6** CCD image of the [001] projection recorded at $\Delta f \approx 640$ nm. 166
- Figure 6.7** (A) Modulus of the Fourier transform of an image of the [001] projection, and, (B) The **phases** (in degrees) and *amplitudes* (arbitrary units) of all observed reflections. 167
- Figure 6.8** Results from the origin search using the 11 strongest *hk0* reflections. Contours every 0.5σ with first contour at 0.5σ above the mean (45°). 168
- Figure 6.9** (A) and (B) : The [001] projection at 25\AA resolution using X-ray amplitudes and EM phases, (C) : Magnified area from the filtered image. 169
- Figure 6.10** Electron micrograph of an image of the [010] projection. Note the presence of one- and two-cell thick areas near the bottom right-hand corner of the crystal. 171
- Figure 6.11** (A) Modulus of the Fourier transform of the image shown in Figure 6.10, and, (B) The phases and amplitudes of all observed reflections. 172
- Figure 6.12** (A) and (B) : The [010] projection at 25\AA resolution using X-ray amplitudes and EM phases, (C) : Magnified area from the filtered image. 173
- Figure 6.13** Electron micrograph of the [101] projection. 174
- Figure 6.14** [101] projection : (A) Modulus of the Fourier transform, (B) Filtered image, (C) View of the model of the crystal packing, (D) 4° precession photograph of the *hkh* plane, and, (E) The intensity-weighted reciprocal lattice. 175

- Figure 6.15** [130] projection : (A) Modulus of the Fourier transform, (B) Filtered image, (C) View of the model of the crystal packing, (D) 50Å synthesis using X-ray amplitudes and EM phases. 177
- Figure 6.16** Moduli of the Fourier transforms (first column) and filtered images (second column) of two projections which could not be characterised. 178
- Figure 6.17** Electron micrograph showing different periodicities within the same area. 179
- Figure 6.18** Sections through (A) : the density of a hypothetical 32 hexamer at 30Å resolution, (B) : a sphere of constant density, (C) : the map calculated with the mixed phase set, (D) : the difference map (C)-(A). 182
- Figure 6.19** Sections through (A) : the density of a hypothetical molecule at 30Å resolution, (B) : a sphere of constant density, (C) : the map calculated with the mixed phase set, (D) : the difference map (C)-(A). 183
- Figure 6.20** (A) : Surface representation of the 30Å electron density map, (B) : a stack of sections from the same map, and, (C) the corresponding area from the [010] projection. 184
- Figure 7.1** Schematic diagrams of the structure of the hexameric core of ArgR. 187
- Figure 7.2** Pairwise alignment of the C-terminal region of ArgR and AhrC. 188
- Figure 7.3** The value of the PC metric, (A) before, and, (B) after PC refinement of the orientation of the hexamer for the best 298 solutions from a rotation function calculated with low resolution data. 192

- Figure 7.4** The value of the PC metric, (A) before, and, (B) after PC refinement of the orientation of the ArgR trimer for the best 198 orientations from the rotation function. 193
- Figure 7.5** The value of the PC metric after PC refinement of the orientation of the ArgR hexamer for the best 216 orientations from the “Direct Rotation Function”. 194
- Figure 7.6** Results from a brute force search using X-PLOR. (A) 9-7Å, (B) 7-5Å. 195
- Figure 7.7** Monoclinic form : (A) 202-12, and, (B) 202-15Å [001] native Patterson projection. Contours every 3% of the origin peak. 196
- Figure 7.8** Monoclinic form : [001] Permutation syntheses. 198
- Figure 7.9** Results from a systematic search along x using discs of constant density, (A) Correlation coefficient, (B) R -factor. 199
- Figure 7.10** [010] Projection : A 60Å permutation synthesis (6 reflections : $\bar{1}01$, 100, 101, 200, 201 and 300). 200
- Figure 7.11** Values of the PC metric before (A and C) and after (B and D) PC-refinement of the best solutions from two rotation functions calculated using 12-7 and 8-5Å data. 201
- Figure 7.12** Monoclinic form : Results from two translation function calculations in X-PLOR (A) 12-7, (B) 8-5Å. Contours every 0.5σ above mean. 202

- Figure D.1** Lane 1 : Intact AhrC, Lanes 2, 3, 4, ..., samples taken after 2, 4, 8, ..., minutes of incubation in the presence of 2% (w/w) V8 at 38°C. 204
- Figure A.1** 8° precession photograph of a trigonal AhrC crystal in a direction perpendicular to the crystallographic 3-fold axis. 207
- Figure A.2** Magnified area of the precession photograph shown in Figure A.1. 208

List of Tables

| | | |
|------------------|---|----|
| Table 1.1 | Enzymes involved in Arginine metabolism. | 9 |
| Table 1.2 | Summary of the FTIR assignments. | 16 |
| Table 2.1 | Crystallisation conditions for the orthorhombic form. | 25 |
| Table 2.2 | Crystallisation conditions for the monoclinic form. | 29 |
| Table 2.3 | Crystallisation conditions for the trigonal form. | 32 |
| Table 3.1 | Native data sets. | 38 |
| Table 3.2 | Heavy atom soaking experiments : Orthorhombic form. | 41 |
| Table 3.3 | Heavy atom soaking experiments : Monoclinic form. | 48 |
| Table 3.4 | Data collections : Orthorhombic form. | 49 |
| Table 3.5 | Data collections : Monoclinic form. | 52 |

List of Acronyms

| | |
|----------|---|
| DNA | Deoxyribonucleic acid. |
| IPTG | Isopropyl-1-thio- β -D-galactopyranoside. |
| MIR | Multiple isomorphous replacement. |
| MIRAS | Multiple isomorphous replacement with anomalous scattering. |
| MPD | 2-methyl-2,4-pentanediol. |
| MR | Molecular replacement. |
| NMR | Nuclear magnetic resonance. |
| PEG | Polyethylene glycol. |
| RNA | Ribonucleic acid. |
| SDS-PAGE | Sodium dodecyl sulphate polyacrylamide gel electrophoresis. |
| SIRAS | Single isomorphous replacement with anomalous scattering. |

List of Symbols

| | |
|---|--|
| \mathbf{x} | Vector in real space. |
| \mathbf{u} | Vector in Patterson space. |
| \mathbf{h} | Reciprocal lattice vector corresponding to a reciprocal lattice point with coordinates hkl . |
| $\mathbf{F}_{\mathbf{h}}$ | Structure factor corresponding to the Bragg reflection \mathbf{h} . |
| $F_{\mathbf{h}}$ or $ \mathbf{F}_{\mathbf{h}} $ | Amplitude of the structure factor corresponding to the Bragg reflection \mathbf{h} . |
| $\Delta F_{\mathbf{h}}$ | Isomorphous difference corresponding to the Bragg reflection \mathbf{h} . |
| \mathbf{F}_H | Structure factor of the heavy atom structure of a derivative. |
| F_H or $ \mathbf{F}_H $ | Amplitude of a structure factor of the heavy atom structure of a derivative. |
| \mathbf{F}_{PH} | Structure factor of a derivative. |
| F_{PH} or $ \mathbf{F}_{PH} $ | Amplitude of a structure factor of a derivative. |

I should not like to leave an impression that all crystal structures are easy to solve.

I seem to have spent much more of my life not solving structures than solving them.

Dorothy Crowfoot Hodgkin
NOBEL LECTURE, 1964.

Chapter 1

Introduction

Life is evolution and evolution is heritable changes of information. There are two sources of information in biological systems : information encoded in their genetic material and information contained in pre-existing structures. It is the study of this former source that has led to the explosive growth of Biology in recent years. Nucleic acids —at least as we know them today— are chemically rather inert molecules (catalytic RNA excluded). Each and every step in their life cycle involves interactions with other macromolecules : replication, recombination, transcription, translation, regulation, packaging, repair, all require recognition by and interaction with proteins or other macromolecular assemblies (such as ribosomes or *snRNPs*). Clearly, understanding protein-nucleic acid interactions is understanding some of the most important events in the life of a cell.

Regulation of gene expression is of prime importance not only for its role in determining the pattern of cellular processes, but also, for its role in cell growth and differentiation. Gene activity in both prokaryotes and eukaryotes is regulated primarily at the level of transcription. The most common mechanism for this is through binding of proteins (transcription factors in eukaryotes, repressors and activators in prokaryotes) to specific DNA sequences. These proteins exert their regulatory effect either by causing (or stabilising) local changes in the struc-

ture of DNA, or by interacting with proteins involved in transcription, or by a combination of both.

The last five years have witnessed a significant increase in the amount of structural detail available for these systems : well over 30 crystal or NMR structures are currently available for transcription factors and their complexes. The majority of these proteins can be classified in six relatively well defined families [Helix-Turn-Helix, Homeodomain, Zinc-binding domains (3 classes), Basic Region Leucine Zipper, Basic Region / Helix-Loop-Helix / Leucine Zipper and the β -Ribbons-Helix-Helix family]. A number of transcription factors (such as the TATA-box binding protein or the papillomavirus E2 protein) show no similarity to any of these families. Figures 1.1 and 1.2 show schematic diagrams of the crystal structures of some complexes for which coordinates were available from the Protein Data Bank at the time of writing. Their beauty and diversity (two unifying themes in protein-DNA complexes) are immediately obvious.

Although a detailed description of these DNA-binding motifs will not be given here (excellent reviews can be found in Steitz, T.A., 1990, Harrison, S.C. & Aggarwal, A.K., 1990, Harrison, S.C., 1991, Freemont, P.S., Lane, A.N. & Sanderson, M.R., 1991, Pabo, C.O. & Sauer, R.T., 1992, Berg, J.M., 1993, Burley, S.K., 1994, Ellenberg, T., 1994, Wright, P.E., 1994, Phillips, S.E.V., 1994), some of the most interesting results to emerge from these studies will be discussed in some detail.

The most frequently observed mode of interaction between DNA and transcription factors involves insertion of a secondary structure element (such as an α -helix or a pair of β -strands) into the major groove of DNA and formation of (i) DNA-sequence-specific hydrogen bonds between protein side or main chain atoms and the exposed edges of the DNA base pairs, and, (ii) DNA-structure-dependent hydrogen bonds or salt bridges between side or main chain atoms (not necessarily from the recognition element) and the non-esterified phosphodiester

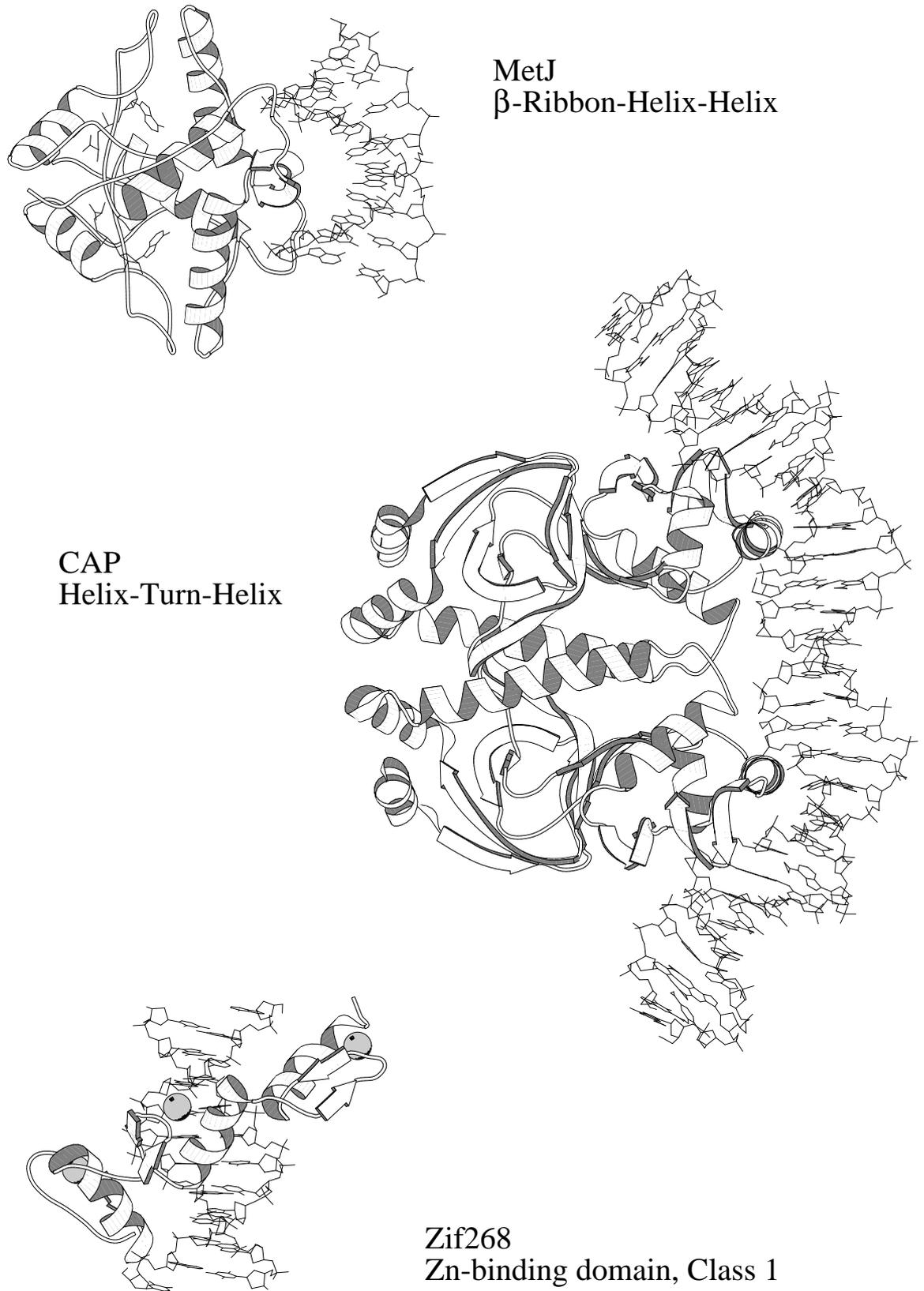


Figure 1.1: Schematic diagrams of the crystal structures of some transcription factor-DNA complexes.

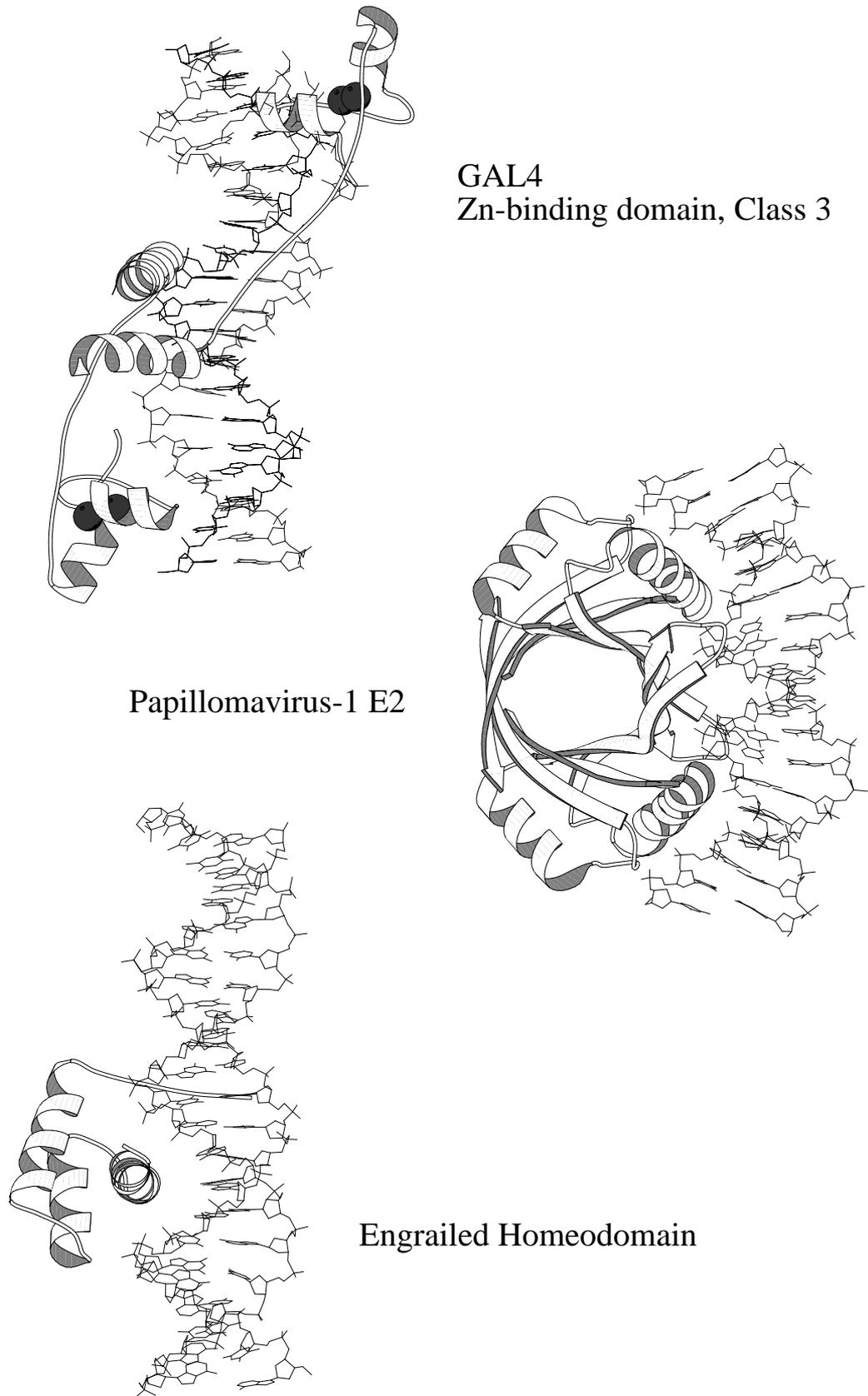


Figure 1.2: Schematic diagrams of the crystal structures of some transcription factor-DNA complexes.

oxygens of DNA. This second set of contacts can help ‘align’ the protein on the DNA and may also account for the ability of transcription factors to find their targets by a one-dimensional random walk on the DNA. In addition, when the structure of a given target sequence deviates from that of the canonical B-form (by being bent or kinked or inherently flexible), the sequence-specificity of a protein that can ‘sense’ these deviations will be enhanced. Other types of interactions (such as hydrophobic interactions with the $-\text{CH}_3$ group of thymine or the deoxyribose rings in the DNA backbone, hydrogen bonds mediated by water molecules, contacts in the minor groove, etc.) have also been observed, but are less frequent.

The formation of hydrogen bonds between protein side chains and the edges of the DNA base pairs is probably the most important source of sequence specificity. One interesting result that emerged from the structural studies of transcription factor-DNA complexes is the absence of a “recognition code” : the same side chain can form hydrogen bonds with different DNA bases, and the same base can be recognised by different side chains. This is not to say that all protein side chains or DNA bases are used with similar frequencies. The side chains of arginine, asparagine, glutamine and lysine account for the majority of the observed contacts. Similarly, most of the hydrogen bonds are directed towards purines and especially guanine. It is worth noting that the most frequently observed contacts (arginine-guanine, asparagine-adenine and glutamine-adenine) all involve a pair of hydrogen bonds between the side-chain and the base and their existence had been predicted theoretically (Seeman, N.C., Rosenberg, J.M. & Rich, A., 1976). In retrospect, the absence of a “universal recognition code” is not surprising : the evolutionary pressure on these relatively small regulatory circuits is not as high as in the case of, say, the translational apparatus or other systems with a more general and immediate effect on the cell.

One last point concerns the conservation of symmetry in the known transcrip-

tion factor-DNA complexes : if a regulatory protein has intramolecular symmetry, then, in general, its DNA target will also have 2-fold or pseudo 2-fold symmetry (depending on whether the DNA sequence is an exact palindrome or not), and in their complex the intramolecular symmetry axes will coincide. Given the number of transcription factors that are dimers or tetramers, symmetry conservation appears to be an evolutionarily successful mechanism for increasing the thermodynamic stability (and hence, specificity) of protein-DNA complexes.

This thesis describes our attempts to determine the crystal structure of AhrC, the arginine repressor/activator from *Bacillus subtilis*. There are several reasons which make AhrC a very interesting target for a structure determination. The first, and probably the most important, is its functional multiplicity : in the presence of L-Arginine, AhrC represses the genes encoding for the biosynthetic and activates those encoding for the catabolic enzymes of arginine metabolism. Furthermore, when AhrC is expressed in *Escherichia coli* cells it acts not only as a regulator of the arginine metabolism but also as a necessary accessory protein for the resolution (through site-specific recombination) of multimers of the ColE1 plasmid. Understanding the structural basis of the observed functional multiplicity is clearly an exciting prospect.

Other unique features of AhrC include (i) its hexameric organisation [AhrC and its *E. coli* homologue (ArgR), are at the time of writing the only known examples of hexameric regulatory proteins], (ii) its unusual (inherently bent) operator sequence, and, (iii) the absence of significant homology to any of the characterised DNA-binding motifs or DNA-binding proteins.

What follows is a more detailed discussion of the biochemical, genetical and biophysical data available for AhrC and closely related proteins.

1.1 Biochemical and Genetical Data.

The pathways of arginine metabolism in *Bacillus* spp. and the enzymes involved are shown in Figure 1.3 and Table 1.1 respectively. Starting from glutamate (which is synthesised from NH_4^+ and α -ketoglutarate, a citric acid cycle intermediate), arginine can be obtained in either seven (ArgBCDJFGH) or eight (ArgABCDEFGH) steps depending on whether the product of the *argJ* gene [which simultaneously removes the acetyl group from N-acetylornithine (to give ornithine) and transfers it to glutamate (to form N-acetylglutamate)] is active. It is not clear which of these two anabolic pathways is active in *Bacillus subtilis*, although recent evidence suggests that it is the ArgJ pathway which is used for ornithine synthesis (Baumberg, S. & Klingel, U., 1993 and references therein).

There are two major catabolic pathways for arginine : the first goes through Arginase and Ornithine aminotransferase to glutamate γ -semialdehyde and glutamate (**1, 2, 3** in Figure 1.3). The second pathway is the reverse of the biosynthetic reactions and goes via citrulline (through the action of Arginine deiminase) to ornithine and carbamoyl phosphate [catalyzed by Ornithine carbamoyltransferase, (**4, 5, 6** in Figure 1.3)].

The *B. subtilis* genes encoding for the enzymes involved in the anabolic pathway of arginine metabolism have been mapped and cloned (Harwood, C.R. & Baumberg, S., 1977, Baumberg, S. & Harwood, C.R., 1979, Mountain, A. & Baumberg, S., 1980, Baumberg, S. & Mountain, A., 1984, Mountain, A., Mann, N.H., Munton, R.N. & Baumberg, S., 1984, Mountain, A., McChesney, J., Smith, M.C.M. & Baumberg, S., 1986, Smith, M.C.M., Mountain, A. & Baumberg, S., 1986a). They are organised in two clusters : the first (*argCJBD-cpa-argF*) comprises the genes involved in the synthesis of up to and including citrulline, whereas the second (*argGH*) encodes for the enzymes that catalyse the last two steps in the pathway. Analysis of arginine hydroxamate-resistant (Ahr) mutants allowed the

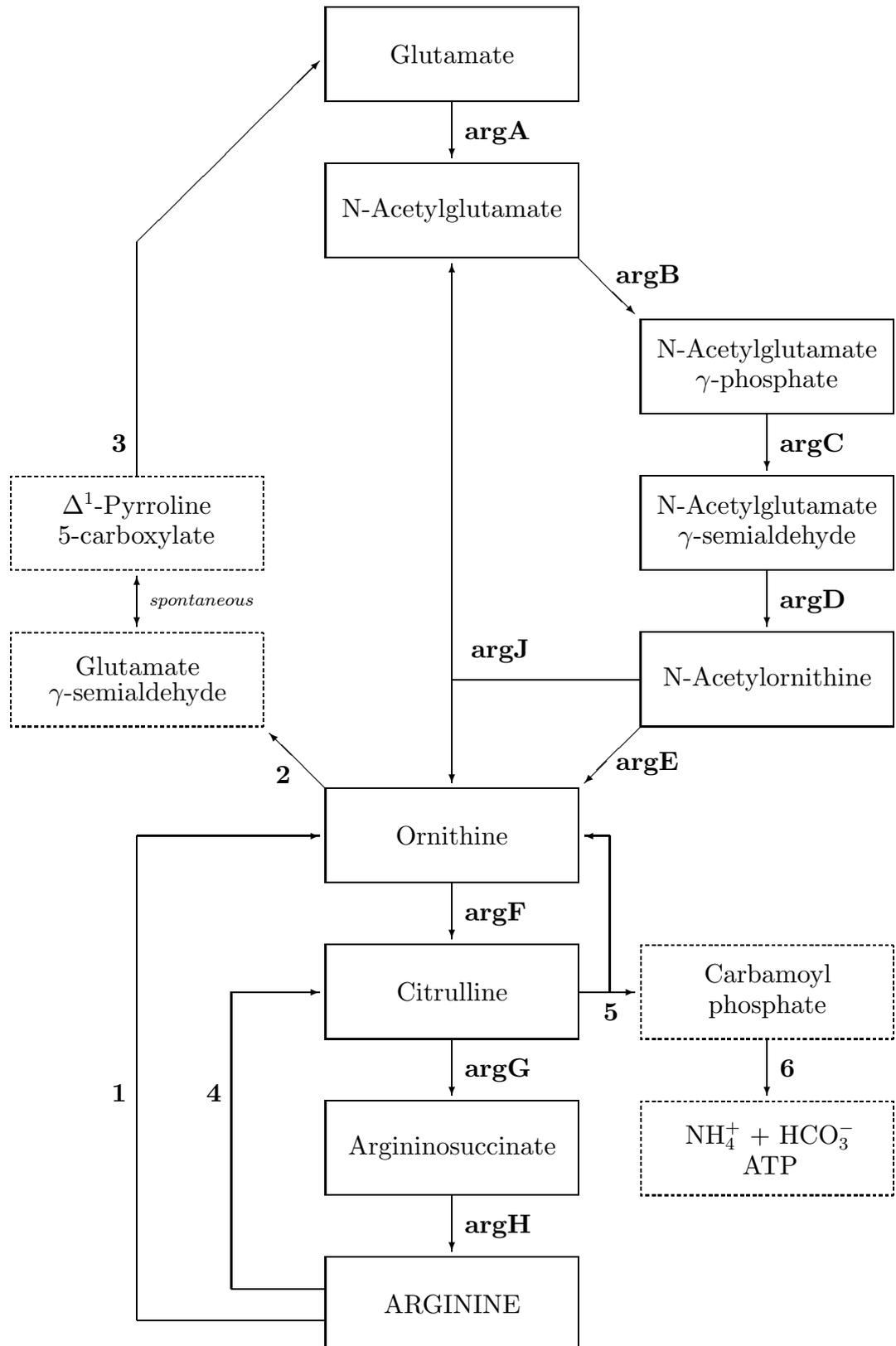


Figure 1.3: Pathways of arginine metabolism in *Bacillus* spp. Dashed boxes enclose intermediates of the catabolic pathways. Names or symbols for the genes encoding the corresponding enzymes are shown in bold type.

| Gene | Enzyme |
|-------------|---|
| argA | N-Acetylglutamate synthetase |
| argB | N-Acetylglutamokinase |
| argC | N-Acetylglutamylphosphate reductase |
| argD | N-Acetylornithine δ -transaminase |
| argE | N-Acetylornithinase |
| argF | Ornithine carbamoyltransferase (anabolic) |
| argG | Argininosuccinate synthetase |
| argH | Argininosuccinase |
| argJ | Ornithine acetyltransferase |
| 1 | Arginase |
| 2 | Ornithine aminotransferase |
| 3 | Δ^1 -Pyrroline 5-carboxylate dehydrogenase |
| 4 | Arginine deiminase |
| 5 | Ornithine carbamoyltransferase (catabolic) |
| 6 | Carbamate kinase |

Table 1.1: Enzymes involved in arginine metabolism

identification and (serendipitous) cloning of the *ahrC* gene whose product (AhrC) has since been shown to be the arginine repressor/activator (Smith, M.C.M., Mountain, A. & Baumberg, S., 1986a, North, A.K., Smith, M.C.M. & Baumberg, S., 1989, Smith, M.C.M., Czaplewski, L., North, A.K., Baumberg, S. & Stockley, P.G., 1989, Czaplewski, L.G., North, A.K., Smith, M.C.M., Baumberg, S. & Stockley, P.G., 1992).

AhrC, as shown by analytical ultracentrifugation (Czaplewski, L.G., North, A.K., Smith, M.C.M., Baumberg, S. & Stockley, P.G., 1992, hereafter referred to as CNSBS), is a stable homohexamer of total molecular mass 105kDa. Each

subunit consists of 149 amino acids whose sequence shows significant homology only to ArgR, the arginine repressor from *E. coli* (Figure 1.4). ArgR does not act as a transcription activator but it has been shown to be a necessary accessory protein for the resolution —through site-specific recombination— of multimers of the ColE1 plasmid (Lim, D., Oppenheim, J.D., Eckhardt, T. & Maas, W.K., 1987, Stirling, C.J., Szatmari, G., Stewart, G., Smith, M.C.M. & Sherratt, D.J., 1988). ArgR is not a resolvase, but may be implicated in synapse formation. Although the two proteins share only 29% identity and are from such taxonomically distinct prokaryotes, AhrC can complement *E. coli* ArgR⁻ strains both in the regulation of arginine metabolism and the resolution of the ColE1 plasmid (it is worth noting that the reverse is not true : ArgR can not complement *B. subtilis* AhrC⁻ strains in the regulation of arginine metabolism).

```

ArgR  MRSSAKQEELVKAFKALLKEEFSSQGEIVAALQEQGFDNINQSKVSRMLTKFGAVRTRN
AhrC  MNKGQRHIKI----REIITSNEIETQDELVDMLKQDGY-KVTQATVSRDIKELHLVKVPT
      * . . . . . . . . . . * . * . * . * . . . * . * . * . * .
      . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

ArgR  AKMEMVYCLPAEL---GVPTTSSPLKNLVLDIDYNDVAVVIHTSPGAAQLIARLLDSLKG
AhrC  NNGSYKYSLPADQRFNPLSKLKRALMDAFVKIDSASHMIVLKTMPGNAQAIGALMDNLDW
      . . . . . * . * . * . . . . . * . . . . * . * . * . * . * . * .

ArgR  AEGILGTIAGDDTIFTTPANGFTVKDLYEAILELFDQEL -156
AhrC  DE-MMGTICGDDTILII CRTPEDTEGVKNRLELL -149
      * . . * . * . * . * . * . * . * . * . * . * . * . * . * .

```

Figure 1.4: Sequence alignment of AhrC and ArgR (CLUSTAL-V, Higgins, D.G., Bleasby, A.J. & Fuchs, R., 1992). Identities (*) and similarities (.) are indicated.

Mutational analysis of ArgR (Tian, G. & Maas, W.K., 1994, Burke, M., Merican, A.F. & Sherratt, D.J., 1994) suggested that the C-terminal region is implicated in arginine binding and oligomerisation whereas the N-terminal region is involved in DNA recognition and binding. The implied organisation of AhrC and ArgR in two functionally and/or structurally distinct domains is consistent with the the non-uniform distribution of homology between them (19% identity for residues 1-80, 34% identity for residues 81-152) and the sensitivity of AhrC to proteolytic cleavage.

Sequence analysis of the region 5' to the *argCJBD-cpa-argF* cluster revealed the presence of three putative operator sites similar to the *E. coli* arginine operator sequences (Figure 1.5, Smith, M.C.M., Mountain, A. & Baumberg, S., 1986b).

| | | | |
|------------------------------|-----------|-----------|----------------|
| ARG box consensus sequence : | AATGAATAA | | TNATNCANT |
| AhrC operator sequence R1 : | AATGTTAAA | T | AATTTTACACA |
| | R2 : | ATTGAATTA | ATTT TTATTCATG |
| | R3 : | AATGAATAA | AA ATATTAAAT |

Figure 1.5: Sequence alignment of the *E. coli* ARG box consensus sequence and three putative AhrC operator sequences located at the 5' region of the *argCJBD-cpa-argF* cluster.

DNase I and hydroxyl radical footprinting experiments (CNSBS, 1992) showed that AhrC protects (in an arginine-dependent manner) two of these regions (R1 and R2 in Figure 1.5) and a further site that lies within the coding sequence of the *argC* gene (hereafter referred to as *argC_{O2}*). The same experiments suggested that (i) AhrC interacts with one face of the DNA over a length approximately equal to five helical turns, and, (ii) three groups of nucleotides that are hyper-

sensitive to hydroxyl radical cleavage in the presence of AhrC, remain so even in its absence. The assumed DNA bending or kinking at these sites is consistent with their sequence characteristics, such as the presence of T-A steps embedded in A-tracts (Goodsell, D.S., Kaczor-Grzeskowiak, M. & Dickerson, R.E., 1994).

Based on the results from these experiments, a model of repressor binding has been put forward (CNSBS, 1992) in which AhrC binds with four of its subunits to the R1 and R2 operator sequences and with the remaining two to the $argC_{O2}$ sequence, thus forming a repression loop (Figure 1.6).



Figure 1.6: Model of how AhrC may induce the formation of a repression loop.

The biological significance of the $argC_{O2}$ site has been demonstrated by measuring ornithine carbamoyltransferase (ArgF) activity in a *B. subtilis* strain containing a mutant $argC_{O2d}$ operator sequence which was obtained by deleting four base pairs from the centre of the wild type operator (CNSBS, 1992). This strain showed a reproducible fourfold decrease in the repression ratio for ArgF, suggesting that $argC_{O2}$ is indeed involved in the regulation of the $argC_{JBD}$ - cpa - $argF$ cluster *in vivo*. Interestingly, not only could the mutant $argC_{O2d}$ operator still be protected by AhrC, but also, the footprint boundaries were identical with that of the wild type operator (the protected region had shrunk by 4 base pairs). This implies that AhrC is sufficiently flexible to be able to recognise its target DNA sequences even when their relative positions on the DNA have changed by as much as a $\approx 140^\circ$ rotation about the helix axis.

1.2 Biophysical Data.

Figure 1.7 is a graphical representation of the secondary structure prediction for AhrC (based on eight different methods, programme PREDICT, Eliopoulos, E., Geddes, A.J., Brett, M., Pappin, D.J.C. & Findlay, J.B.C., 1982, and references therein).

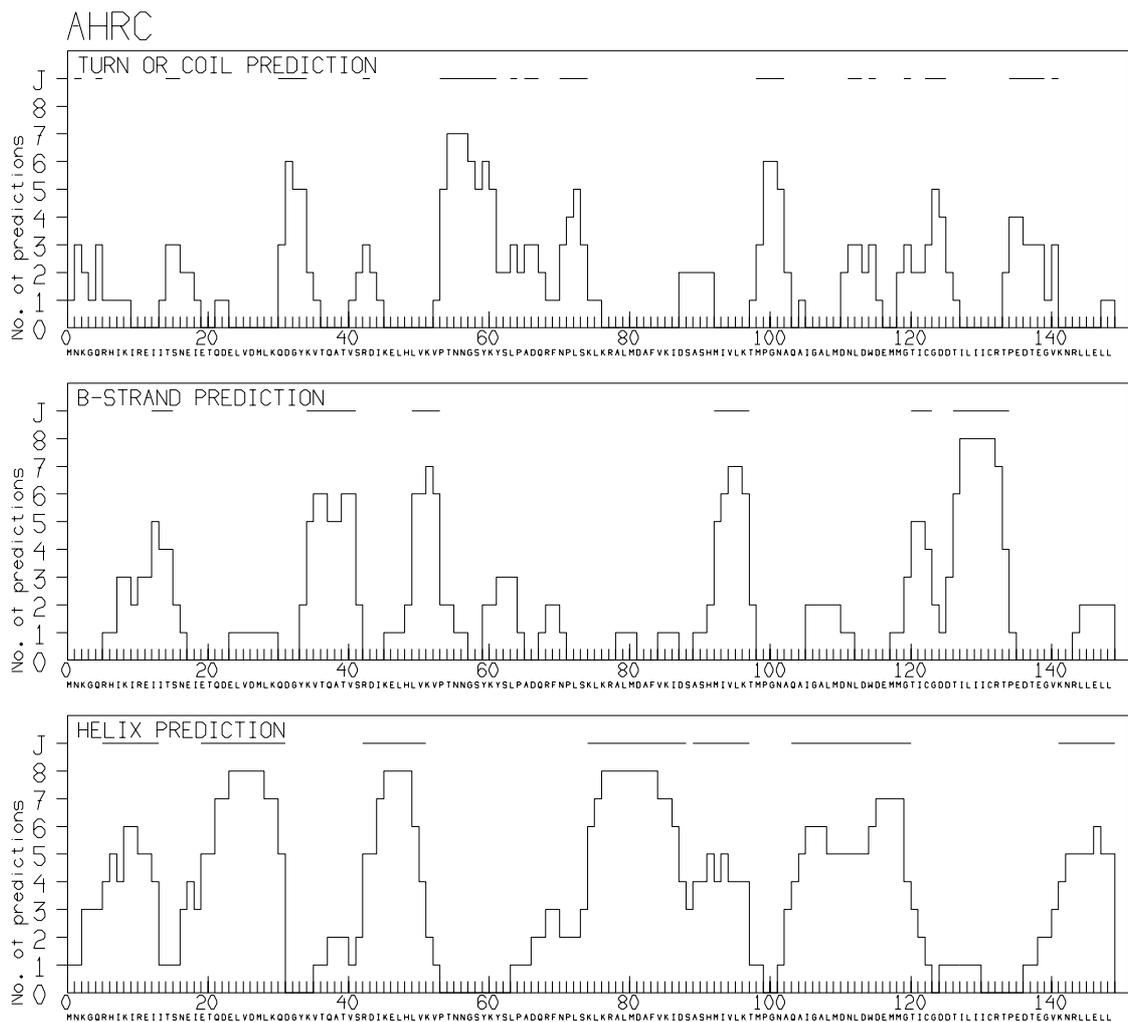


Figure 1.7: Secondary structure prediction for AhrC.

This prediction suggests that AhrC is an α/β or an antiparallel β structure with a high α -helical content. Although some very common secondary structure motifs are immediately obvious (such as the β -strand prediction centered at residue 38 which is followed by a turn or coil prediction, an α -helix and another

β -strand), the relatively low accuracy of these methods precludes a confident assignment of the structural class of AhrC.

During the course of this investigation, Dr Amalia G. Aggeli (Department of Chemistry, University of Leeds) kindly agreed to record and analyse a Fourier Transformed Infra-Red (FTIR) spectrum of a D₂O solution of AhrC in 200 mM (NH₄)₂SO₄. Some of her results are reproduced below.

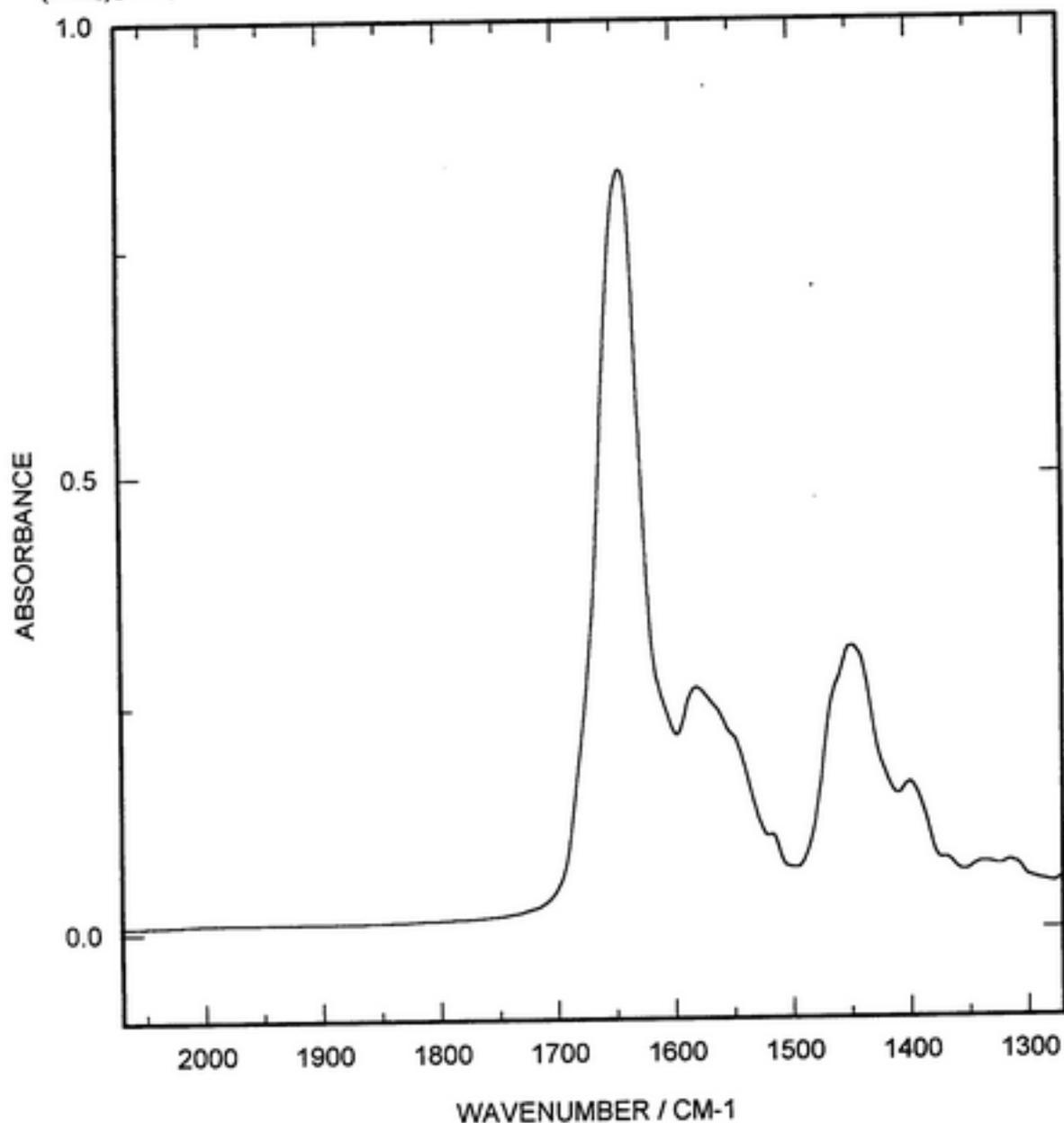


Figure 1.8: FTIR absorbance spectrum after solvent subtraction.

Figure 1.8 shows the absorbance spectrum after solvent subtraction and Figure 1.9 is a comparison between the observed (black line I) and band-fitted

(blue line) Amide I region. The wavenumbers corresponding to the peaks of the five bands are 1607 cm^{-1} (black line II), 1637 cm^{-1} (green), 1650 cm^{-1} (cyan), 1662.4 cm^{-1} (magenta) and 1679 cm^{-1} (red) (also shown in Table 1.2). The bands at 1637 and 1679 cm^{-1} are indicative of the presence of antiparallel β -sheet structure, whereas the 1650 cm^{-1} band is characteristic of α -helical structure. The 1662.4 cm^{-1} band is characteristic of turns and the 1607 cm^{-1} band probably arises from side-chain vibrations. A summary of these assignments is given in Table 1.2.

These results, together with the indications from the secondary structure prediction algorithms, suggest that AhrC has an antiparallel β structure. It is worth noting that the intensities of the 1637 – 1679 and 1650 cm^{-1} bands suggest that the dominant secondary structure elements are β -strands organised in antiparallel β -sheets and not α -helices (as indicated by the secondary structure prediction methods).

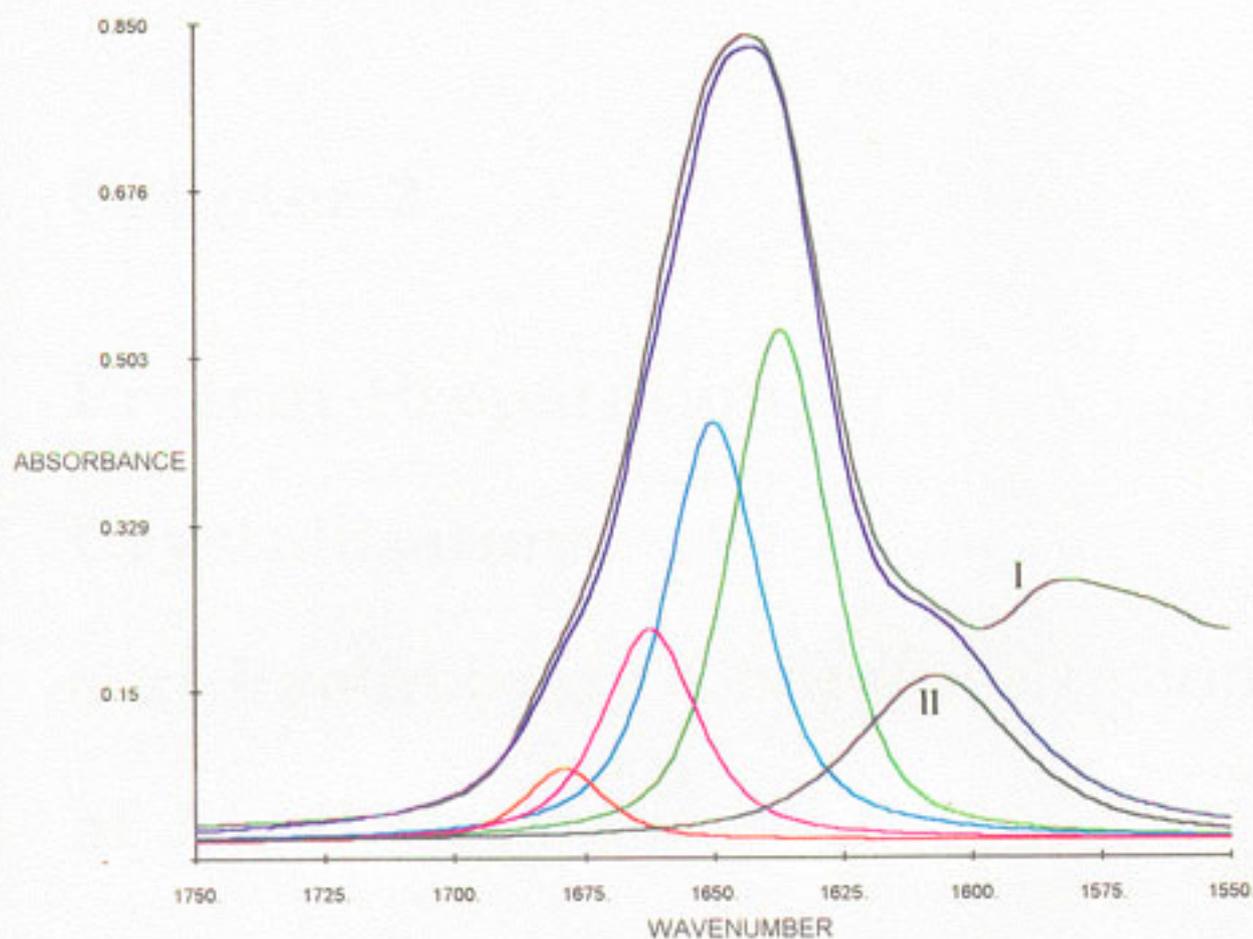


Figure 1.9: Band-fitted Amide I region.

| Wavenumber (cm^{-1}) | Colour code | Assignment |
|---------------------------------|-------------|-----------------------------|
| 1637 | Red | Antiparallel β -sheet |
| 1679 | Green | |
| 1650 | Cyan | α -helix |
| 1662 | Magenta | Turns |
| 1607 | Black | Side chains (?) |

Table 1.2: Summary of the assignments.

Chapter 2

Protein Preparation, Crystallisation, and Preliminary Characterisation of AhrC Crystals

2.1 Protein Preparation.

2.1.1 Protein Purification.

AhrC was purified as described by Czaplowski, L.G., *et al*, & Stockley, P.G., 1992. Differential precipitation from low ionic strength solutions is used twice during the purification, allowing an efficient initial separation before the final chromatographic step.

Escherichia coli strain DS903(pUL2202) was grown at 37°C in rich medium. At an optical density at 600 nm of 1.5 the cells were induced by the addition of IPTG and the incubation continued for 3 hrs. Cells were harvested by centrifugation and were thawed and resuspended in Arg buffer (20 mM Tris-HCl, 10 mM MgCl₂, 10 mM 2-mercaptoethanol, 1 mM PMSF, 25 μM TPCK, pH 7.5).

The cell suspension was sonicated on ice to prepare a cell extract and the insoluble material (which included AhrC) was harvested by centrifugation. The pellet was resuspended in Arg buffer and treated with DNase I. AhrC was solubilised by the addition of solid NaCl to a final concentration of 0.5 M, and the suspension incubated at 37°C for 15 min.

The supernatant after DNase I and NaCl treatments was obtained by centrifugation and extensively dialysed against Arg buffer containing 75 mM NaCl at 4°C. The precipitate which formed consisted mainly of AhrC. It was harvested by centrifugation and resuspended in 250 mM NaCl Arg buffer. The protein was loaded onto a S-Sepharose cation-exchange column with 250 mM NaCl Arg buffer, and the column developed using a 250-600 mM NaCl linear gradient in Arg buffer.

A typical elution profile of the S-Sepharose column is shown in Figure 2.1. Peak I is the flow through of proteins that do not bind to the column and peak II corresponds to proteolytically cleaved AhrC (Figure 2.2). The main peak (III), is—as shown by high resolution SDS-PAGE (Schägger, H. & Jagow, G., 1987), Figure 2.2—a mixture of intact AhrC and a faster migrating band which was thought to represent AhrC with one or two amino acids missing (Stockley, P.G., personal communication). Attempts to separate those two bands using preparative isoelectric focusing have been unsuccessful (Walsh, A.P., personal communication). It is worth noting, that a SDS-PAGE of washed and subsequently dissolved AhrC crystals (orthorhombic form, Section 2.2) showed that these crystals consist of only one protein species.

Later in this study, the plasmid carrying the *ahrC* gene was reconstructed. This resulted in the apparent loss of the faster migrating band. Crystallisation trials with this protein preparation showed that the conditions for optimum crystal growth of the orthorhombic form had changed slightly. Furthermore, the crystals produced from the new protein preparation were non-isomorphous with

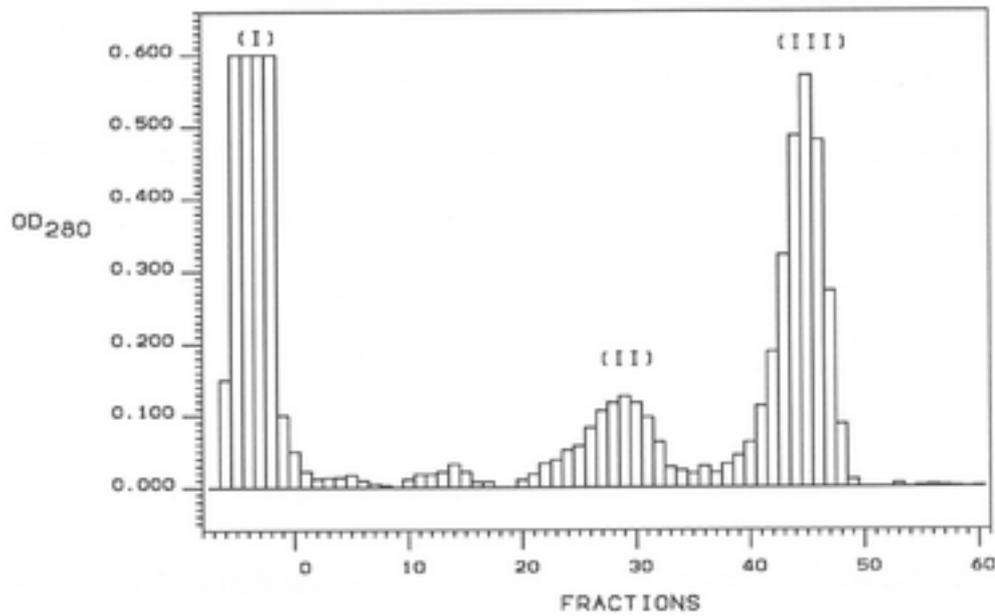


Figure 2.1: Elution profile of the S-Sepharose column.

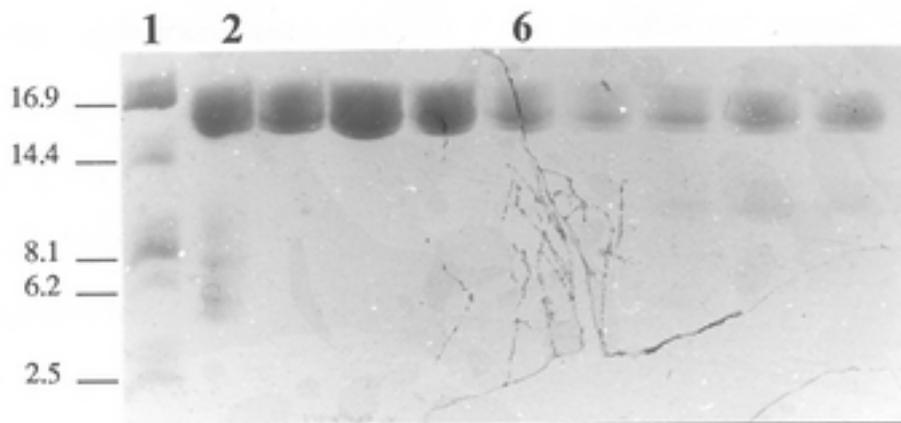


Figure 2.2: High resolution SDS-PAGE of fractions collected from peaks (II) & (III). Lane 1 : Molecular weight markers (16.9, 14.4, 8.1, 6.2 and 2.5 kDa). 2 : Standard AhrC solution. 3 to 5 : Fractions collected from peak (III). 6 to 10 : Fractions collected from peak (II).

those grown in the past (the mean fractional isomorphous difference for all data to 4Å was 20%). It is worth noting that the unit cell dimensions of these two types of the orthorhombic form are virtually identical.

The protein concentration was determined by measuring the absorbance at 280 nm, assuming that an OD₂₈₀ of 1.0 corresponds to a protein concentration of 1 mg ml⁻¹ (Czaplewski, L.G., *et al.*, & Stockley, P.G., 1992). Typically, 7 lt of culture would give approximately 20 mg of protein. AhrC was precipitated with 23% (w/v) PEG 6000 and stored at 4°C for further use.

2.1.2 Stability of AhrC During Storage.

Due to problems encountered with the reproducibility of the crystallisation experiments (Section 2.2), it was decided to monitor the the behaviour of the purified protein over the course of few months using SDS-PAGE. The main conclusions from these experiments are presented below.

AhrC is very sensitive to proteolytic cleavage. Although protease inhibitors such as PMSF or TPCK are present at all stages of protein purification and also during storage, approximately 5 months after the protein preparation, low molecular weight bands appear on SDS-PAGE (Figure 2.2, Lane 2). Futhermore, after 2 months of storage higher molecular weight species start appearing on SDS-PAGE of the purified protein (Figure 2.3). The molecular weights of these species (as judged from their electrophoretic mobility, Figure 2.3) suggest that they may correspond to covalently linked dimers and trimers of AhrC. The dominance of the 32 kDa band is consistent with chemical cross-linking experiments with bismido esters (Czaplewski, L.G., *et al.*, & Stockley, P.G., 1992). Although it is not clear what the nature of the bonds that stabilise these multimers is, the formation of disulphide bridges can probably be excluded since the protein is heated to 100°C in the presence of 2% (v/v) 2-mercaptoethanol during sample preparation.

It was not surprising to find that crystallised AhrC is more stable than the

precipitated protein. Figure 2.4 shows that AhrC crystals which have been stored for 4 months show little material outside the expected 16 kDa band, whereas a sample from the same protein preparation which was stored as a precipitate contains a significant amount of the covalently linked dimer.

An attempt to re-purify —using cation-exchange chromatography— the native AhrC hexamer from a protein preparation that had been stored for 4 months, showed that all species present in the sample eluted as a single peak (data not shown). This suggests that neither the molecular weight nor the charge of the hexameric molecule changes during storage.

2.2 Crystallisation, and Preliminary Characterisation of AhrC Crystals.

The crystallisation of AhrC in a form suitable for a complete three-dimensional X-ray structure determination has been reported (Boys, C.W.G., *et al*, & Stockley, P.G., 1990). Unfortunately, crystallisation under the conditions described therein was not reproducible (Boys, C.W.G., personal communication).

When this project started, it was decided that an attempt to crystallise AhrC based on its very low solubility in low ionic strength solutions was a worthwhile exercise. This approach proved very successful : all three crystal forms of AhrC described in this thesis are grown from low ionic strength solutions.

Although conditions that produced crystalline material were found soon after this project started, the production of crystals of a quality suitable for crystallographic studies was rather more difficult. In retrospect, the major problem was the instability of AhrC upon storage (Section 2.1.2) : the quality of the crystals and the reproducibility of the crystallisation experiments was inversely related to

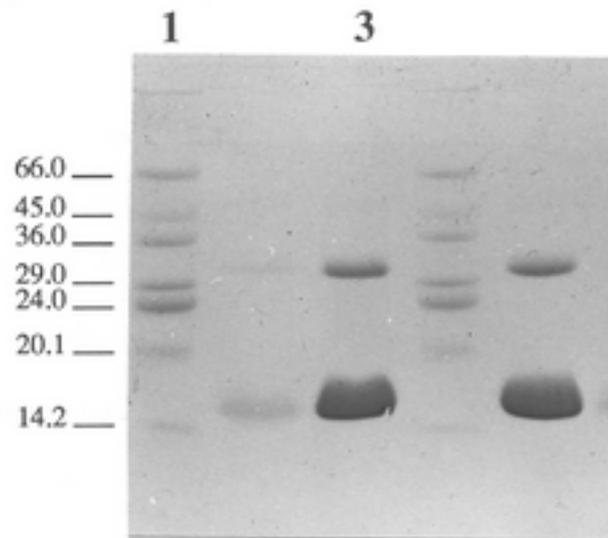


Figure 2.3: SDS-PAGE of AhrC three months after purification. Lanes 3 and 5 : AhrC, 1 and 4 : molecular weight markers(66, 45, 36, 29, 24, 20.1 and 14.2 kDa).

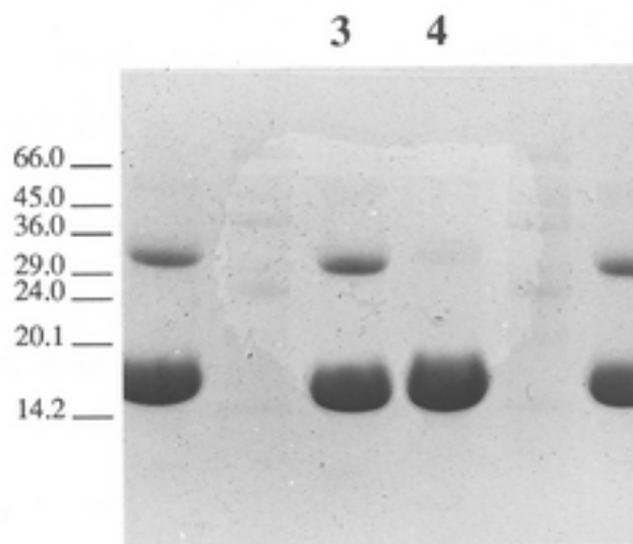


Figure 2.4: SDS-PAGE of AhrC crystals and precipitated AhrC four months after purification. Lanes 1,3,6 : AhrC which was stored as a precipitate, 4 : AhrC crystals (orthorhombic form) 2 and 5 : molecular weight markers(66, 45, 36, 29, 24, 20.1 and 14.2 kDa).

the length of time that the protein had been stored. Most of these problems disappeared when it was realised that the protein should be used as soon as possible, and in no case later than three weeks after its preparation.

All the crystallisation experiments were performed by hanging drop vapour diffusion (McPherson, A., 1982, Ducruix, A. & Giegé, R., 1992, Blundell, T.L. & Johnson, L.N., 1976). The effects of some crystallisation parameters which were found to be important for the three crystal forms grown from low ionic strength are discussed below.

Temperature Crystallisation trials have been set up at 4°C, 20°C and 28°C. The solubility of AhrC increases with decreasing temperature but crystals of similar quality can be grown at all these temperatures under slightly different conditions. For technical reasons, most of the crystallisation experiments were performed at 20°C.

Ionic strength The ionic strength of the well solution was adjusted with ammonium sulphate. Concentrations in the range 0 to 150 mM have given useful results, with different crystal forms growing at different concentrations.

Buffer and pH range The pH of the protein solution was adjusted to 7.5 using 30 mM phosphate buffer. The pH of the well solution was varied in the range 4.5–8.5 in steps of 0.2 units using phosphate, citrate or cacodylate buffer. Best results have been obtained from phosphate buffer at pH 4.9 (which is very close to the isoelectric point of AhrC).

PEG concentration Although PEG is not required for crystallisation, it was found that the addition of a small amount of a medium molecular weight PEG improved the morphology and size of the crystals. Crystallisation trials have been set up using 0.4, 1, 2, 4, 6, and 8 kDa PEG at concentrations varying from 0 to 15% (w/v) for both the well and protein solution. Unfortunately, the variation in the composition of the commercially available

PEGs led to a proportional variation of the optimum conditions for crystal growth. Best results for the orthorhombic form have been obtained from PEG 4000 at concentrations close to 5% (w/v).

Concentration of isopropanol Due to the sensitivity of AhrC to proteolytic cleavage, small amounts of the protease inhibitors PMSF and TPCK were included in all crystallisation trials. Because of their instability in aqueous solutions, both inhibitors were prepared in isopropanol. It was later found that isopropanol was affecting the crystal growth rate. Best results have been obtained from 1% (v/v) isopropanol.

2.2.1 Orthorhombic Form.

Details of the crystallisation conditions are given in Table 2.1. It should be noted that the conditions for optimal crystal growth may vary for different protein batches and will almost certainly be different for different brands of PEG. It was found necessary to optimise the concentration of ammonium sulphate and PEG in the well solution for each protein batch individually.

These crystals grow as rectangular blocks elongated along [010] and bounded on the (001) and (100) faces (Figure 2.5). Their typical size is $0.50 \times 0.15 \times 0.20$ mm³ but crystals with dimensions up to $1.50 \times 0.30 \times 0.35$ mm³ have been obtained. The ratio of the dimensions of the crystals is inversely proportional to the ratio of the unit cell dimensions with the longest crystal axis being parallel to the shortest unit cell translation.

The space group was determined from precession photographs and found to be C222₁ with $a=231.3\text{\AA}$, $b=74.4\text{\AA}$ and $c=138.0\text{\AA}$ (Figure 2.6¹).

This crystal form is identical to the one reported by Boys, C.W.G., *et al*, & Stockley, P.G., 1990. Assuming that the equivalent of one hexamer is present in

¹All zero level precession photographs were recorded with a crystal to film distance of 100 mm and are reproduced in this thesis on a scale of 1:1.

| | |
|------------------|---|
| Protein solution | 50 μM TPCK 1.25 mM DTT 1.2 mM PMSF 30 mM Phosphate buffer, pH 7.5 150 mM Ammonium sulphate 1 % (v/v) isopropanol 10 mg ml^{-1} AhrC |
| Well solution | 50 μM TPCK 1.25 mM DTT 1.2 mM PMSF 100 mM Phosphate buffer, pH 4.9 4% PEG 4000 60 mM Ammonium sulphate 1 % (v/v) isopropanol |
| Drops | 4 μl protein plus 4 μl well solution |

Table 2.1: Crystallisation conditions for the orthorhombic form.

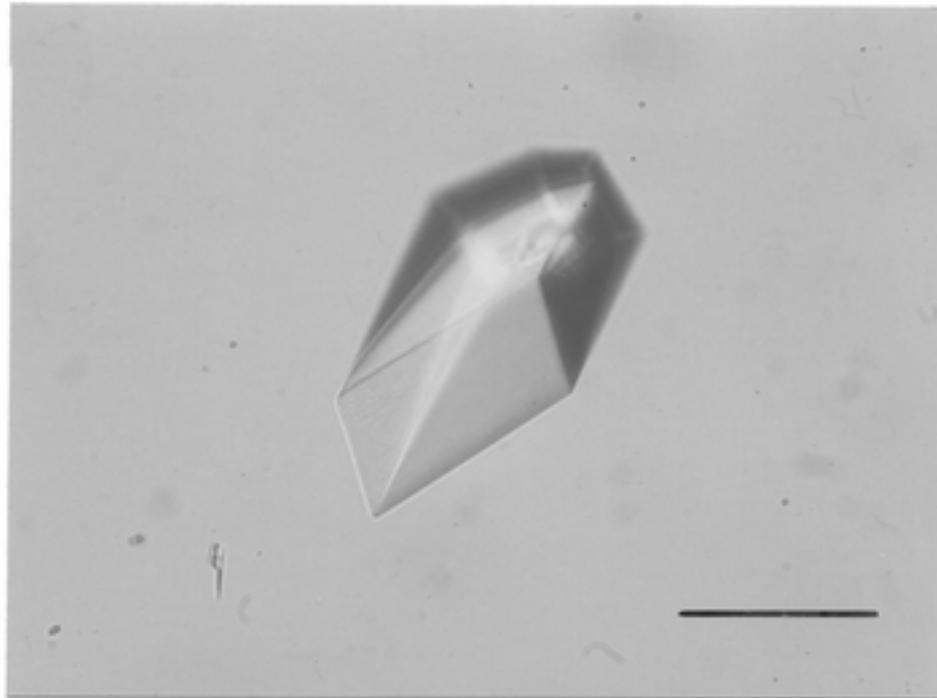


Figure 2.5: A typical orthorhombic AhrC crystal. Scale bar 0.25 mm.

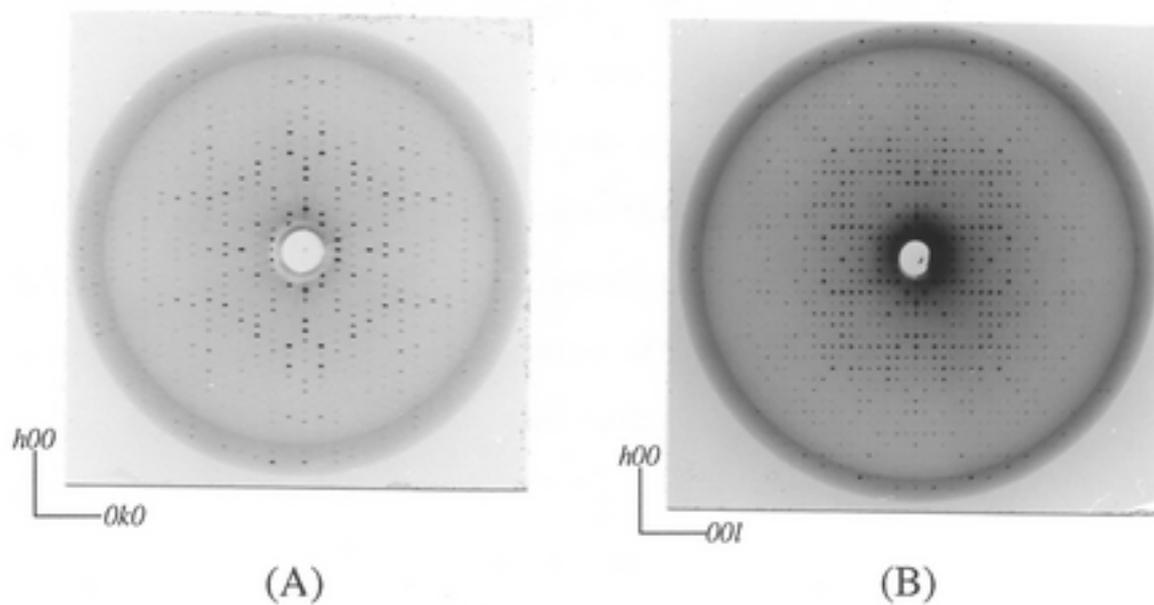


Figure 2.6: Orthorhombic form, 8° precession photographs of (A) the $hk0$ and (B) $h0l$ zones.

the asymmetric unit, the estimated solvent content is 55% (Matthews, 1977).

These crystals diffract X-rays to 3Å resolution using laboratory sources and to 2.7Å using synchrotron radiation (Chapter 3). They are sensitive to radiation damage : after 12 hours of exposure to 2.7 kW, monochromatised (CuK α) X-rays at room temperature, only low ($d_{\min} \geq 4.5\text{\AA}$) resolution reflections could be observed. An attempt was made to increase their useful life-time by soaking them in solutions containing free-radical scavengers, such as styrene or methyl methacrylate (Zaloga, G., & Sarma, R., 1974). Unfortunately, no improvement was found. This was not the case when the crystals were cooled to 4°C using a device originally described by Marsh, D.J. & Petsko, G.A., 1973. The useful crystal life-time was almost doubled. Most of the data sets described in this thesis were collected at this temperature. Alignment of the cold stream of air with the axis of the X-ray capillary proved to be more difficult than expected resulting in several crystals being lost due to water condensation. Most of these problems disappeared when it was decided to mount the crystals in a solution containing low gelling temperature agarose which set on cooling to maintain constant hydration during data collection (Richmond, T.J., *et al*, & Klug, A., 1984).

An artificial mother liquor consisting of 10% MPD and 100 mM acetate buffer at pH 4.9 has been developed. This was necessary for several reasons : Firstly, under the crystallisation conditions described above the crystals are not stable. Approximately 6 weeks after their appearance, they start dissolving, possibly due to the presence of a higher concentration of PEG in the well solution which leads to a gradual increase of the ionic strength in the hanging drops. Secondly, the presence of ammonium sulphate and phosphate ions can cause serious problems in heavy atom screening experiments (Blundell, T.L. & Johnson, L.N., 1976, McPherson, A., 1982). A stabilising solution consisting of MPD and acetate ions presents fewer problems. Finally, MPD at high concentrations can act as a cryo-protective mother liquor (Petsko, G.A., 1975). It proved possible to transfer

AhrC crystals to solutions containing up to 45% (v/v) MPD without any obvious problems. Such high concentrations of MPD should allow cooling of these crystals to at least -40°C which can help to reduce their sensitivity to radiation damage (Hope, H., 1988, Singh, T.P., *et al*, & Huber, R., 1980, Young, A.C.M. & Dewan, J.C., 1990). All orthorhombic crystals used in this study were transferred to the stabilising solution at least three days before experimentation.

2.2.2 Monoclinic Form.

The crystallisation conditions for this form are given in Table 2.2. These crystals could be grown reproducibly from only one protein batch. It is not clear why this is so, but, a possible explanation is that both bands seen on a SDS-PAGE of protein purified at the beginning of this project are needed for crystallisation (Section 2.1.1, Figure 2.2). The crystals grow as rhombic prisms elongated along $[010]$ with well developed (100) and (011) faces (Figure 2.7).

The space group, as determined from precession photographs, is $P2_1$ with unit cell dimensions $a=202.7\text{\AA}$, $b=72.6\text{\AA}$, $c=73.0\text{\AA}$ and $\beta=97.8^{\circ}$. They diffract X-rays to better than 4\AA using laboratory sources and to 3.2\AA using synchrotron radiation. Assuming that the crystallographic asymmetric unit contains two hexamers, the estimated solvent content is 45%.

The pattern of strong, low resolution reflections in the $hk0$ zone (Figure 2.8) is worth noting : strong reflections are present if $2h + k = 4n$, which suggests the presence (in the $[001]$ projection) of a centered superlattice with $a=202.7\text{\AA}$ and $b'=36.3\text{\AA}(=b/2)$.

Although the crystallisation conditions are very similar to those used for the orthorhombic form, the monoclinic crystals are not stable in the artificial mother liquor described in the previous section and most of these crystals are also unstable in their well solution. These problems made the heavy atom screening experiments for this form rather adventurous (Chapter 3).

| | |
|------------------|---|
| Protein solution | 50 μM TPCK 1.25 mM DTT 0.7 mM PMSF 30 mM Phosphate buffer, pH 7.5 150 mM Ammonium sulphate 1 % (v/v) isopropanol 10 mg ml^{-1} AhrC |
| Mixing solution | 50 μM TPCK 1.25 mM DTT 0.7 mM PMSF 30 mM Phosphate buffer, pH 7.5 150 mM Ammonium sulphate 1 % (v/v) isopropanol 8 % PEG 6000 |
| Well solution | 50 μM TPCK 1.25 mM DTT 0.7 mM PMSF 100 mM Phosphate buffer, pH 5.2 4.4% PEG 6000 16 mM Ammonium sulphate 1 % (v/v) isopropanol |
| Drops | 4 μl protein plus 4 μl mixing solution |

Table 2.2: Crystallisation conditions for the monoclinic form.

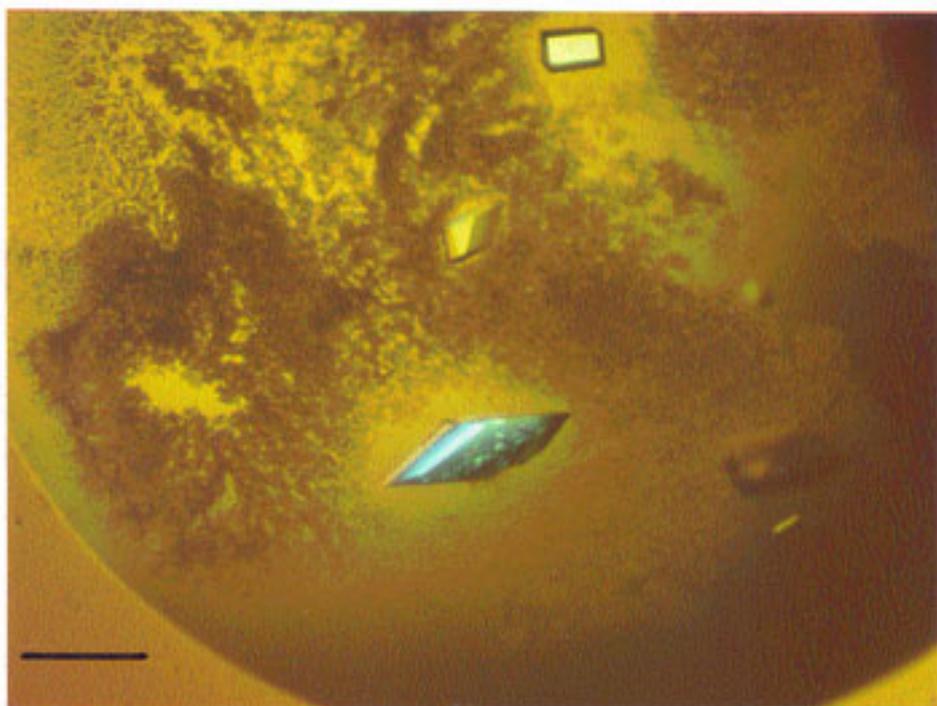


Figure 2.7: The diamond shaped crystal at the centre of the picture is a monoclinic AhrC crystal. All the other crystals are orthorhombic. Scale bar 0.25 mm.

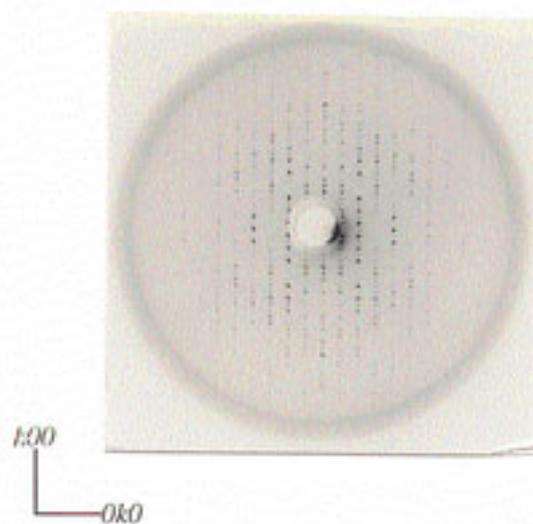


Figure 2.8: Monoclinic form. 7° precession photograph of the $hk0$ zone.

2.2.3 Trigonal Form.

Boys, C.W.G., *et al*, & Stockley, P.G., 1990 reported the growth of AhrC crystals with a habit of “small, triangular rods” from 15% MPD, 100 mM phosphate buffer, pH 7.5 and an 80-fold molar excess of L-arginine hydrochloride. These crystals were too small for a space group determination (Boys, C.W.G., personal communication).

In this study, crystals with a similar morphology have been obtained from low ionic strength solutions (Figure 2.9). Refinement of the crystallisation conditions (Table 2.3), allowed us to grow crystals of a size suitable for a preliminary characterisation, but, due to the inherent disorder of this form, we have been unable to collect a complete three-dimensional data set.

Figure 2.10(A) shows a “still” photograph taken with the long axis of the crystal 8° away from the direct beam. The absence of well defined reflections and the presence of almost continuous intensity in the various levels (most clearly seen in the -1 level), suggests that these crystals are disordered. It was a surprise to find that a zero level precession photograph from the same crystal (Figure 2.10(B)) showed well defined reflections out to 8\AA . The broadening of the higher resolution reflections seen in this photograph indicates the presence of rotational disorder about the morphological 3-fold.

An upper level ($hk1$) precession photograph (Figure 2.11) established that the crystal system is trigonal. The order is preserved to a much lower resolution with individual reflections merging to form arcs.

The space group determination was complicated by the crystal disorder and will be discussed in detail. The $hk0$ zero level photograph has symmetry $p6mm$. This means that the plane group of the projection of the electron density along the $[001]$ direction is either $p3m1$ or $p31m$. The only enantiomorphic, trigonal, non-rhombohedral space groups consistent with either of these two plane groups (for the projection along the 3-fold) are $P312$, $P3_112$, $P3_212$, $P321$, $P3_121$ and

| | |
|------------------|---|
| Protein solution | 50 μ M TPCK 1.25 mM DTT 1.2 mM PMSF 30 mM Phosphate buffer, pH 7.5 150 mM Ammonium sulphate 1 % (v/v) isopropanol |
| Well solution | 50 μ M TPCK 1.25 mM DTT 1.2 mM PMSF 100 mM Phosphate buffer, pH 4.9 6% PEG 6000 30 mM Ammonium sulphate 1 % (v/v) isopropanol |
| Drops | 4 μ l protein plus 4 μ l well solution |

Table 2.3: Crystallisation conditions for the trigonal form.

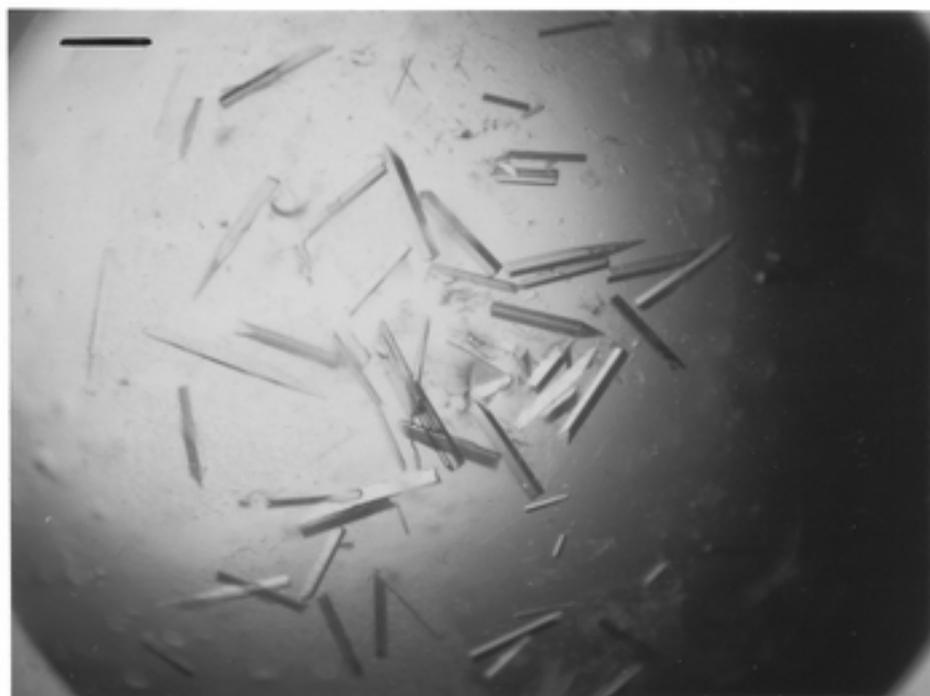
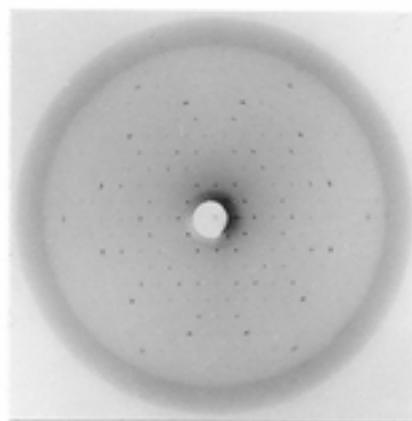


Figure 2.9: Typical trigonal AhrC crystals. Scale bar 0.25 mm.



(A)



(B)

Figure 2.10: (A) "Still" X-ray photograph and (B) 6.5° precession photograph of a trigonal AhrC crystal ($hk0$ zone).

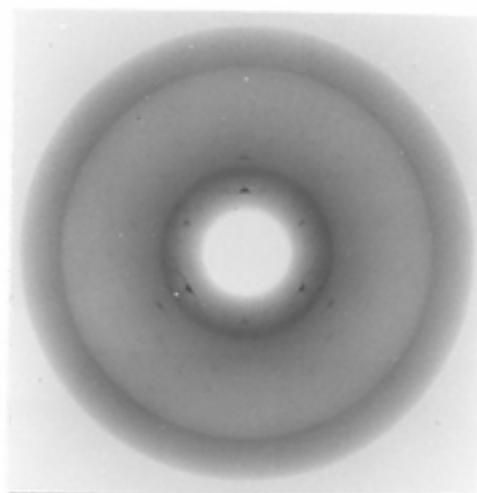


Figure 2.11: Trigonal form, $hk1$ zone.

P3₂21 (International Tables for X-ray Crystallography, Vol.I, 1952). The unit cell dimensions as determined from the X-ray photographs are $a=b=66.6\text{\AA}$ and $c=160\text{\AA}$. This small unit cell suggests the presence of only two hexamers per unit cell (with an estimated solvent content of 50%). The requirement for a space group with a set of two equivalent positions further reduces the possible choices to only two space groups : P312 (Wyckoff notation of possible sets : g, h, i) or P321 (c, d). In the absence of any information in directions perpendicular to the 3-fold, these two space groups can not be differentiated. In both space groups the point symmetry of the positions of these sets is 3 and the asymmetric unit contains two protomers. The crystal packing is fixed by the space group symmetry : the equivalent positions are $0, 0, z$ and $0, 0, \bar{z}$ with the crystallographic and molecular 3-fold coinciding and the molecules forming columns parallel to the [001] direction.

A low resolution native Patterson projection along the [001] direction is shown in Figure 2.12. The calculation of the Patterson projection involves no assumptions about the space group of the crystals and its consistency with the packing arrangement derived from symmetry considerations is an independent confirmation of the space group assignment².

The crystal packing as described above suggests a model for the observed disorder phenomena : the crystals suffer from translational disorder parallel to the 3-fold and to a lesser extent from rotational disorder about it. The translational component arises from the different relative positions of the protein columns in a direction parallel to the column axis (which coincides with the crystallographic 3-fold axis).

An important question is whether —based on the analysis above— the possibility of the protein having a molecular 6-fold axis of symmetry can be excluded. We believe that it can : due to the “special” position of the molecules, the order of

²The amplitudes of the $hk0$ reflections were estimated visually from the precession photograph shown in Figure 2.10(B). The 18 observed reflections were classified as “very strong”, “strong” and “weak”, and amplitudes were assigned to these as follows : “very strong” $\rightarrow F=6$ (1 reflection), “strong” $\rightarrow F=2$ (7 reflections) and “weak” $\rightarrow F=0.5$ (10 reflections).

the crystallographic $[001]$ axis depends (for the given crystal packing) on the order of the molecular axis. If the molecular axis was a 6-fold then the space group would be $P622$, but again, with two molecules per unit cell at $0, 0, z$ and $0, 0, \bar{z}$. We can reach the same conclusion based on a purely geometrical argument : if the order of the molecular axis was 6, then the individual protomers would have to be $\approx 80\text{\AA}$ long and with a diameter less than $\approx 20\text{\AA}$. Such dimensions are possible but highly unlikely.

It is unfortunate that the crystal form most suitable for a crystal structure determination is the most problematic of those described. Nevertheless, two important conclusions can be drawn from this preliminary analysis : (i) AhrC has maximal dimensions 66\AA by 66\AA by 80\AA and (ii) It possess an intramolecular axis of symmetry of order 3.

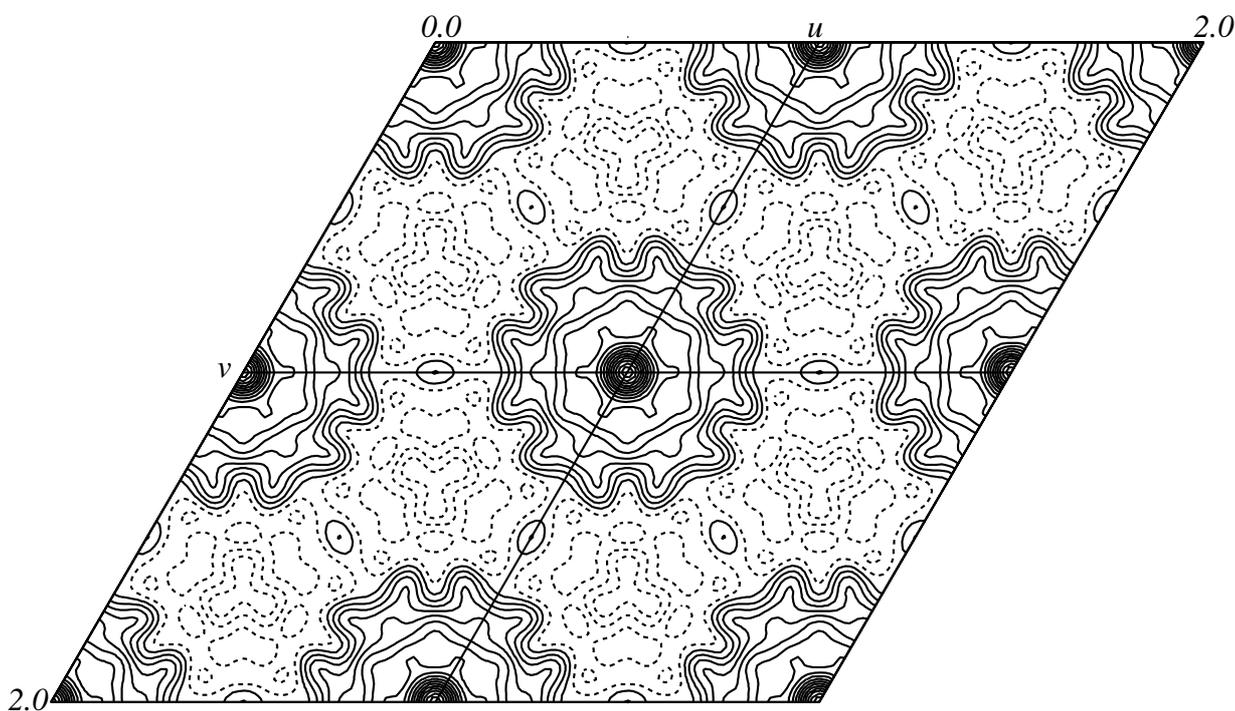


Figure 2.12: Trigonal form, $66\text{-}80\text{\AA}$ native Patterson projection along $[001]$. Four unit cells are shown. Contours every 5% of the origin peak; negative contours broken.

Chapter 3

Preparation and Preliminary Analysis of Heavy Atom Derivatives

3.1 Data Collection Strategy.

All data sets described in this thesis (with the exception of a medium (2.9Å) resolution data set collected using synchrotron radiation) have been collected using a X-100A Xentronics/Siemens multiwire, position sensitive, two-dimensional area detector. The X-ray source was graphite-monochromatised CuK_α radiation from a Rigaku RU200 rotating anode operating at 2.7 kW with a 200 μm focus. Each data set was collected from one crystal. The rotation method (Arndt, W. & Wonacott, A.J., 1977) was used for all data collections with an oscillation angle typically in the range 0.2 to 0.3°.

Care was taken to keep the geometry of the data collections as much as possible the same for the native and derivative data sets : most data sets from orthorhombic crystals have been collected with a crystal to detector distance of 17.5 cm and the detector set at an angle $2\theta=6.5^\circ$. With this setting all data

between 138 and 5.0Å (with some reflections to 3.9Å) can be recorded. The crystals were aligned (at $\omega=0^\circ$) with their [001] direction parallel to the X-ray beam and the [010] direction approximately 15° away from the rotation axis. The 15° tilt of the [010] axis has two consequences : (i) the Bijvoet pairs can not be measured simultaneously and (ii) a $\approx 95\%$ complete data set can be collected through a single 90° rotation without the need for an additional data collection with a different crystal orientation. The choice not to measure the hkl and $h\bar{k}l$ terms under as similar conditions as possible is justified on the grounds that the anomalous differences could not be measured accurately anyway : the crystals are sensitive to radiation damage and the data sets had to be collected as fast as possible, resulting in anomalous differences well below the noise level.

In the case of the monoclinic form, the crystal to detector distance was 25.5 cm and the detector swing angle was 10° . Due to the morphology of crystals, the [010] axis was parallel to the rotation axis necessitating the collection of 180° of data.

The data frames were processed using the programme XDS (Kabsch, W., 1988, 1993) and the intensities (corrected for Lorentz and polarisation factors) were converted to a .LCF (and later to a .MTZ) file for further processing using the CCP4 suite of programmes (Collaborative Computational Project, Number 4, 1994) : The programmes ROTAVATA and AGROVATA applied scales and isotropic temperature factors to continuous batches of data each corresponding to a 5° rotation. This step should compensate for (a) differences in the illuminated crystal volume, (b) radiation damage and (c) absorption, although differences in absorption are not expected to be significant since the crystals were embedded in an agarose gel with a linear absorption coefficient very similar to that of the crystals. Observations which differed by more than 3σ from the mean were rejected (usually less than 0.06% of the total number of measurements for data sets with an average multiplicity of 2). Data sets collected from crystals soaked in heavy atom solutions were brought to the same relative scale as the native data through

the application of an overall scale and temperature factor (programme ANSC).

Several data sets from native orthorhombic AhrC crystals have been collected (Table 3.1). At least one native data set was collected after each protein preparation and was compared with those from previous preparations. Data sets collected from different native AhrC crystals were not merged.

Two non-isomorphous types of the orthorhombic form have been used for data collections (Section 2.1.1). The Type I crystals had been obtained from protein preparations which showed two bands on a SDS-PAGE of the purified protein. Type II crystals have been obtained from more recent protein preparations. Data sets collected from native crystals of the same type are isomorphous (the average mean fractional isomorphous difference between data sets collected from native type II crystals is $\approx 5.0\%$ for all reflections between 30 and 4.0\AA).

| Form and Type | Resolution | R_{symm} | Completeness | Multiplicity |
|------------------|------------|-------------------|--------------|--------------|
| Orthorhombic, I | 4.9 | 5.6 | 73 | 2.1 |
| Orthorhombic, I | 4.9 | 4.9 | 85 | 2.2 |
| Orthorhombic, I | 4.0 | 7.0 | 75 | 1.8 |
| Orthorhombic, II | 6.0 | 4.6 | 92 | 2.3 |
| Orthorhombic, II | 3.6 | 10.6 | 62 | 2.8 |
| Orthorhombic, II | 4.6 | 4.6 | 92 | 1.6 |
| Orthorhombic, II | 3.4 | 5.0 | 100 | 1.9 ‡ |
| Orthorhombic, II | 3.7 | 9.6 | 100 | 1.8 ‡ |
| Orthorhombic, II | 3.5 | 7.3 | 80 | 1.2 |
| Orthorhombic, II | 2.9 | 6.6 | 88 | 1.6 † |
| Monoclinic | 4.5 | 4.9 | 78 | 1.7 |

Table 3.1: Native data sets.

‡ : Data collection of $hk0$ or $h0l$ terms only.

† : Data set collected using synchrotron radiation.

3.2 Preparation of Heavy Atom Derivatives.

The preparation of useful heavy atom derivatives of AhrC crystals proved to be both time consuming and frustrating. Most of the compounds tried damaged the crystals even at very low concentrations (Tables 3.2 and 3.3). This is especially true for compounds containing heavy metals such as mercury or platinum which can form covalent bonds with polarisable protein groups such as those found in the side-chains of methionine, cysteine or histidine. This is rather unfortunate, since the ability of those heavy atoms to form covalent complexes makes them highly specific and, thus, more useful for the preparation of well substituted and isomorphous heavy atom derivatives, with (hopefully) a small number of substitution sites.

The great majority of the heavy atom soaking experiments have been performed using orthorhombic AhrC crystals. This is due to (i) the presence of only one hexamer in their asymmetric unit and (ii) the availability of only a limited number of monoclinic AhrC crystals (Section 2.2.2).

The procedure followed for the heavy atom derivative search in the case of the orthorhombic AhrC crystals is outlined below.

A concentrated heavy atom solution was prepared in 10% MPD, 100 mM Acetate buffer, pH 4.9 (Section 2.2.1). A portion of this concentrated solution was diluted with a volume of the same artificial mother liquor to give 500 μ l of a heavy atom solution at the required concentration. A small crystal was soaked in this solution and its well being (or otherwise) was monitored every 2 to 4 hours. If the crystal survived the initial treatment without any obvious problems (appearance of cracks, loss of birefringence, etc), an attempt was made to take a precession photograph of the $hk0$ zone¹. Depending on the presence or otherwise

¹The $h0l$ zone would have been a better choice, but, due to the morphology of crystals (Section 2.2.1) mounting with the $[010]$ direction parallel to the X-ray beam requires manual re-orientation of the crystal and was therefore avoided.

of significant differences between this precession photograph and the native $hk0$ pattern, either a crystal was soaked using the same or higher concentration of the heavy atom solution and a data set was collected from it, or the procedure was repeated, but this time with a lower concentration of the heavy atom containing compound.

In some experiments, concentrations as low as $0.2 \mu\text{M}$ have been used. In others (marked as “m/p” in Table 3.2) stoichiometric amounts of the heavy atom containing compounds were used to give a specific number of heavy atoms per protein protomer². At such low concentrations, the signal from specific heavy atom binding can be very weak and might not be detected through comparison of precession photographs. In these cases a data set was collected immediately.

The instability of the monoclinic crystals in both the previously described artificial mother liquor and in their well solution, made the heavy atom soaking experiments with this crystal form rather inaccurate : the coverslips with their hanging drops were removed from the crystallisation plate, solid grains of the heavy atom containing compounds were added directly to the drops and the coverslips were replaced back on the crystallisation plate.

Attempts to crystallise AhrC in the presence of KHgI_4 , Ethylmercury phosphate or p-chloromercuribenzenesulphonic acid have also been unsuccessful. The protein precipitated even in the presence of very low concentrations of these compounds.

Tables 3.4 and 3.5 give details of the data sets collected from AhrC crystals soaked in solutions containing heavy atoms. The methods used and the collection strategy adopted have already been described. Data sets marked with (†) in Table 3.4 have been collected from Type I crystals.

²Due to the relatively simple morphology of the orthorhombic crystals, fairly accurate measurements of their volumes are possible, allowing an equally good estimate of the number of molecules per crystal to be obtained.

| Compound | Conditions | | Result |
|---------------------------------------|--------------|---------|----------------|
| Baker's dimercurial | 1 mM | 12 hrs | Cracked |
| | 0.1 mM | 5.5 hrs | Cracked |
| | 30 μ M | 20 hrs | Cracked |
| | 4 μ M | 20 hrs | Cracked |
| | 2 μ M | 20 hrs | Cracked |
| HgCl ₂ | 0.3 mM | 5.5 hrs | Cracked |
| | 60 μ M | 20 hrs | Cracked |
| | 4 μ M | 21 hrs | Disordered |
| | 2 μ M | 24 hrs | Disordered |
| HgI ₂ | 0.4 % sat. | 15 hrs | Disordered |
| | 0.1 % sat. | 23 hrs | Data Collected |
| pCMB | 1 % sat. | 20 hrs | Disordered |
| | 0.2 % sat. | 69 hrs | Disordered |
| Mersalyl acid | 0.4 % sat. | 15 hrs | Cracked |
| | 0.1 % sat. | 20 hrs | Cracked |
| KAuCl ₄ | 1 mM | 3 hrs | Disordered |
| | 0.6 mM | 1.5 hrs | Disordered |
| | 0.5 mM | 16 hrs | Cracked |
| | 0.1 mM | 18 hrs | Disordered |
| | 2 μ M | 46 hrs | Disordered |
| H ₂ IrCl ₆ | 0.03 μ M | 78 hrs | No Differences |
| | 0.75 mM | 12 hrs | Differences |
| | 0.6 mM | 4 hrs | Data Collected |
| | 1.2 mM | 3 hrs | Data Collected |
| (CH ₃ COO) ₂ Hg | 5 mM | 16 hrs | Cracked |

Table 3.2: Heavy atom soaking experiments : Orthorhombic form

| | | | |
|---|------------|---------|----------------|
| Mercurochrome | 0.2 % sat. | 20 hrs | No Differences |
| | 5 % sat. | 3 hrs | No Differences |
| | 2.5 % sat. | 16 hrs | Disordered |
| | 3 % sat. | 12 hrs | Disordered |
| | 100 % sat. | 3 hrs | Data Collected |
| | 100 % sat. | 5 hrs | Data Collected |
| Sm(NO ₃) ₃ | 1 mM | 160 hrs | Differences |
| | 2 mM | 18 hrs | Disordered |
| | 0.4 mM | 21 hrs | Data Collected |
| K ₂ PtCl ₆ | 2 mM | 2 hrs | Disordered |
| | 1 mM | 2 hrs | Disordered |
| | 0.2 mM | 17 hrs | Disordered |
| Pb(NO ₃) ₂ | 0.3 mM | 19 hrs | No Differences |
| | 3 mM | 15 hrs | Disordered |
| Pr(NO ₃) ₃ | 0.6 mM | 17 hrs | Differences |
| | 0.6 mM | 17 hrs | Data Collected |
| NdCl ₃ | 2 mM | 16 hrs | Differences |
| | 2 mM | 18 hrs | Data Collected |
| (NH ₄) ₂ OsCl ₆ | 0.2 mM | 6 hrs | Cracked |
| | 0.04 mM | 17 hrs | Cracked |
| Ta ₂ O ₅ | 10 % sat. | 16 hrs | No Differences |
| | 100 % sat. | 16 hrs | Data Collected |
| DCMNP | 4 % sat. | 15 hrs | Disordered |
| | 4 % sat. | 2 hrs | Disordered |
| K ₃ UO ₂ F ₅ | 0.2 mM | 25 hrs | Differences |
| | 0.2 mM | 20 hrs | Data Collected |

Table 3.2: Heavy atom soaking experiments : Orthorhombic form

| | | | |
|---|-----------------|--------|----------------|
| VOSO_4 | 2 mM | 16 hrs | Cracked |
| CH_3COOTl | 0.8 mM | 25 hrs | No Differences |
| | 0.8 mM | 17 hrs | Data Collected |
| $\text{Ce}(\text{NO}_3)_3$ | 1 mM | 16 hrs | Cracked |
| | 1 mM | 2 hrs | Cracked |
| | 1 mM | 60 min | Disordered |
| TAMM | 0.05 mM | 15 hrs | Disordered |
| | 0.01 mM | 15 hrs | Disordered |
| | 0.05 mM | 30 min | Data Collected |
| | 0.1 mM | 25 min | Data Collected |
| Thimerosal | 0.2 mM | 25 hrs | Disordered |
| $\text{H}_3\text{PO}_4 \cdot 12\text{WO}_3$ | 0.2 mM | 25 hrs | Disordered |
| | 0.2 mM | 6 hrs | Data Collected |
| | 1 m/p | 20 hrs | Data Collected |
| $\text{Co}[\text{Hg}(\text{SCN})_4]$ | 100 % sat. | 20 hrs | Cracked |
| Erythrosin B | 0.1 mM | 40 hrs | Disordered |
| | 0.05 mM | 24 hrs | Disordered |
| KReO_4 | 0.4 mM | 20 hrs | No Differences |
| | 0.8 mM | 40 hrs | No Differences |
| | 0.8 mM | 40 hrs | Data Collected |
| AgNO_3 | 2 μM | 18 hrs | Cracked |
| Gd_2O_3 | 0.5 mM | 20 hrs | Disordered |
| $(\text{CH}_3\text{COO})_2\text{Pb}$ | 0.2 mM | 6 hrs | No Differences |
| | 0.2 mM | 29 hrs | Disordered |
| | 0.2 mM | 18 hrs | Data Collected |
| $\text{Hg}(\text{NO}_3)_2$ | 2 % sat. | 6 hrs | Cracked |

Table 3.2: Heavy atom soaking experiments : Orthorhombic form

| | | | |
|------------------------------------|------------|---------|----------------|
| K ₂ Au(CN) ₄ | 0.02 mM | 20 hrs | Data Collected |
| | 0.05 mM | 14 hrs | Data Collected |
| | 0.1 mM | 16 hrs | Data Collected |
| | 0.2 mM | 14 hrs | Data Collected |
| | 0.4 mM | 16 hrs | Data Collected |
| | 1 mM | 16 hrs | Data Collected |
| | 4 mM | 2 hrs | Data Collected |
| | 4 mM | 16 hrs | Data Collected |
| | 10 mM | 1.5 hrs | Data Collected |
| Ag ₂ SO ₄ | 2 % sat. | 6.5 hrs | Cracked |
| K ₂ HgI ₄ | 0.4 mM | 6.5 hrs | Cracked |
| pCMBS | 0.2 mM | 20 hrs | Cracked |
| | 0.2 mM | 6.5 hrs | Cracked |
| | 0.2 mM | 2 hrs | Disordered |
| | 1 m/p | 18 hrs | Data Collected |
| | 3 m/p | 18 hrs | Data Collected |
| | 6 m/p | 18 hrs | Data Collected |
| | 12 m/p | 3 hrs | Data Collected |
| | 16 m/p | 3 hrs | Cracked |
| | 9 m/p | 16 hrs | Data Collected |
| EMP | 0.1 mM | 6.5 hrs | Disordered |
| CMPN | 100 % sat. | 6.5 hrs | No Differences |
| | 100 % sat. | 29 hrs | No Differences |
| | 100 % sat. | 48 hrs | No Differences |
| AgI | 100 % sat. | 48 hrs | Disordered |
| | 2 % sat. | 24 hrs | Disordered |

Table 3.2: Heavy atom soaking experiments : Orthorhombic form

| | | | |
|---|------------|---------|----------------|
| CH_3COOAg | 0.4 mM | 6.5 hrs | Disordered |
| $\text{Nb}_6\text{Cl}_{14}$ | 0.03 mM | 24 hrs | Differences |
| | 1.5 mM | 21 hrs | Data Collected |
| | 0.7 mM | 48 hrs | Data Collected |
| | 0.7 mM | 20 min | Data Collected |
| $\text{UO}_2(\text{NO}_3)_2$ | 0.4 mM | 24 hrs | Differences |
| | 0.8 mM | 11 hrs | Data Collected |
| | 0.4 mM | 3 hrs | Data Collected |
| | 0.04 mM | 16 hrs | Data Collected |
| Eu_2O_3 | 0.6 mM | 3 hrs | Disordered |
| | 0.2 mM | 20 hrs | Disordered |
| $(\text{CH}_3)_3\text{PbCH}_2\text{COOH}$ | 5 mM | 6 hrs | Differences |
| | 3 mM | 2 hrs | Data Collected |
| | 3 mM | 16 hrs | Data Collected |
| | 5 mM | 16 hrs | Data Collected |
| | 10 mM | 2 hrs | Data Collected |
| | 14 mM | 16 hrs | Data Collected |
| | 20 mM | 18 hrs | Data Collected |
| TaCl_5 | 100 % sat. | 6 hrs | Cracked |
| pHMBA | 100 % sat. | 48 hrs | Differences |
| | 100 % sat. | 20 hrs | Data Collected |
| PdCl_2 | 100 % sat. | 2 hrs | Cracked |
| Tl_2CO_3 | 1 mM | 20 hrs | No Differences |
| AgCN | 100 % sat. | 2 hrs | No Differences |
| | 100 % sat. | 20 hrs | Cracked |
| $\text{Hg}(\text{NO}_3)_2$ | 100 % sat. | 2 hrs | Cracked |

Table 3.2: Heavy atom soaking experiments : Orthorhombic form

| | | | |
|---|------------|--------|----------------|
| HgBr ₂ | 100 % sat. | 2 hrs | Cracked |
| HgO | 100 % sat. | 2 hrs | Cracked |
| Ce ₂ (SO ₄) ₃ | 100 % sat. | 20 hrs | Cracked |
| Cl-Hg-(C ₆ H ₄)-I | 2 % sat. | 25 hrs | Data Collected |
| TiNO ₃ | 1 mM | 20 hrs | No Differences |
| | 2 mM | 18 hrs | No Differences |
| | 3.8 mM | 20 hrs | No Differences |
| | 2 mM | 3 hrs | Data Collected |
| Tl ₂ SO ₄ | 1 mM | 44 hrs | No Differences |
| TlCl | 100 % sat. | 72 hrs | No Differences |
| K ₂ WO ₄ | 1 mM | 20 hrs | Cracked |
| K ₂ PtBr ₆ | 0.2 mM | 20 hrs | Disordered |
| | 2 m/p | 20 hrs | Data Collected |
| | 3 m/p | 20 hrs | Data Collected |
| | 50 μM | 35 min | Disordered |
| Y(NO ₃) ₃ | 2 mM | 20 hrs | Cracked |
| La(NO ₃) ₃ | 1 mM | 2 hrs | Cracked |
| K ₂ PdCl ₄ | 0.2 mM | 2 hrs | Cracked |
| Nd ₂ (SO ₄) ₃ | 0.4 mM | 20 hrs | No differences |
| | 0.8 mM | 20 hrs | Disordered |
| [(CH ₃ Hg) ₃ O]OH | 1 mM | 2 hrs | Disordered |
| CdSO ₄ | 0.4 mM | 2 hrs | No Differences |
| | 0.6 mM | 48 hrs | No Differences |
| SiO ₂ ·12WO ₃ | 1 mM | 6 hrs | Cracked |
| | 0.1 mM | 6 hrs | Cracked |
| | 0.5 m/p | 20 hrs | Data Collected |

Table 3.2: Heavy atom soaking experiments : Orthorhombic form

| | | | |
|--|------------|---------|----------------|
| Er_2O_3 | 100 % sat. | 42 hrs | No Differences |
| Dy_2O_3 | 100 % sat. | 70 hrs | No Differences |
| $\text{UO}_2(\text{CH}_3\text{COO})_2$ | 0.2 mM | 16 hrs | Differences |
| | 0.1 mM | 1.5 hrs | Data Collected |
| | 0.2 mM | 16 hrs | Data Collected |
| Na_2WO_4 | 1 mM | 4 hrs | Cracked |
| $(\text{C}_6\text{H}_5)\text{Hg-Ac}$ | 100 % sat. | 4 hrs | Cracked |
| $\text{K}_2\text{Pt}(\text{NO}_3)_4$ | 0.4 mM | 4 hrs | Disordered |
| $\text{K}_2\text{Pt}(\text{CN})_4$ | 0.4 mM | 15 hrs | Differences |
| | 0.4 mM | 20 hrs | Data Collected |
| $\text{Pt}(\text{H}_2\text{NCH}_2\text{CH}_2\text{NH}_2)\text{Cl}_2$ | 100 % sat. | 2 hrs | Cracked |
| $\text{Ag}_3\text{C}_6\text{H}_5\text{O}_7$ | 0.8 mM | 2 hrs | Cracked |
| PbCO_3 | 100 % sat. | 20 hrs | No Differences |
| | 100 % sat. | 72 hrs | No Differences |
| $\text{W}_{12}\text{H}_{48}\text{N}_{12}\text{O}_{41}$ | 100 % sat. | 20 hrs | Disordered |
| Hg-dUTP | 0.2 mM | 14 hrs | Cracked |
| | 0.2 mM | 60 min | No Differences |
| $(\text{NH}_4)_6[\text{Mo}_7\text{O}_{24}]$ | 0.6 mM | 20 hrs | Cracked |
| $\text{Mo}_{12}\text{H}_3\text{O}_{40}\text{P}$ | 1 mM | 2 hrs | Cracked |
| Arg-Cys-Hg | 4 mM | 20 hrs | No Differences |
| | 4 mM | 76 hrs | No Differences |
| Acetamide & $(\text{CH}_3\text{COO})_2\text{Hg}$ | 2 mM | 2 hrs | |
| | 0.4 mM | 2 hrs | Cracked |
| Acetamide & pCMBS | 4 mM | 2 hrs | |
| | 0.4 mM | 20 hrs | Disordered |

Table 3.2: Heavy atom soaking experiments : Orthorhombic form

| | | | |
|---|----------------|------------------|----------------|
| HgSO ₄ | 100 % sat. | 2 hrs | Cracked |
| Nb ₆ Cl ₁₄ & UO ₂ (NO ₃) ₂ | 2 mM 0.4 mM | 20 hrs 4 hrs | Data Collected |
| Nb ₆ Cl ₁₄ & H ₂ IrCl ₆ | 2 mM 0.5 mM | 20 hrs 20 hrs | Data Collected |

Table 3.2: Heavy atom soaking experiments : Orthorhombic form

| Compound | Conditions | | Result |
|------------------------------------|------------|--------|----------------|
| Nb ₆ Cl ₁₄ | ≈1 mM | 3 hrs | Differences |
| | ≈1 mM | 6 hrs | Data Collected |
| TAMM | 100 % sat. | 3 hrs | Cracked |
| H ₂ IrCl ₆ | ≈0.5 mM | 5 hrs | Cracked |
| pCMBS | ≈0.5 mM | 20 hrs | Disordered |
| K ₂ Pt(CN) ₄ | ≈1 mM | 20 hrs | Disordered |

Table 3.3: Heavy atom soaking experiments : Monoclinic form

- m/p** : Molecules of the heavy atom
containing compound per protein protomer.
- Baker's** : 1,4-diacetoxymercuri-2,3-dimethoxybutane.
- pCMB** : p-chloromercuribenzoic acid.
- pCMBS** : p-chloromercuribenzenesulphonic acid.
- pHMBA** : p-hydroxymmercuribenzoic acid.
- EMP** : Ethylmercury phosphate.
- CMPN** : 1-(4-Chloromercuriphenylato)-2-naphthol.
- TAMM** : Tetrakis-(acetoxymmercuri)methane.

| Compound | Conditions | Res (Å) | R _{symm} (%) | Comp (%) | Diff (%) | K _{emp} | |
|---|--------------------|------------|--------------------------|-------------|-------------|------------------|---|
| HgI ₂ | 0.1 % sat., 23 hrs | 4.9 | 6.4 | 82 | 13.0 | 2.2 | † |
| H ₂ IrCl ₆ | 0.7 mM, 10 hrs | 4.9 | 5.7 | 69 | 23.0 | 2.5 | † |
| | 0.3 mM, 4 hrs | 5.0 | 3.3 | 95 | 5.6 | 2.0 | |
| | 0.6 mM, 4 hrs | 5.0 | 3.5 | 99 | 8.0 | 2.6 | |
| | 1.2 mM, 3 hrs | 5.0 | 3.5 | 77 | 10.0 | 1.8 | |
| | 0.8 mM, 18 hrs | 3.6 | 3.9 | 100 | 18.0 | 2.1 | ‡ |
| Mercurochrome | 100 % sat., 3 hrs | 5.0 | 4.2 | 55 | 7.0 | 1.7 | |
| | 100 % sat., 20 hrs | 4.9 | 5.5 | 73 | 9.0 | 1.9 | † |
| | 100 % sat., 5 hrs | 5.0 | 4.4 | 94 | 12.0 | 1.8 | |
| Sm(NO ₃) ₃ | 0.4 mM, 21 hrs | 6.0 | 4.0 | 99 | 6.0 | 1.2 | † |
| Pr(NO ₃) ₃ | 0.6 mM, 17 hrs | 4.9 | 3.5 | 89 | 12.0 | 1.7 | † |
| NdCl ₃ | 2 mM, 18 hrs | 4.9 | 5.6 | 37 | 15.0 | 2.0 | † |
| Ta ₂ O ₅ | 100 % sat., 16 hrs | 4.9 | 6.4 | 80 | 16.0 | 1.4 | † |
| K ₃ UO ₂ F ₅ | 0.2 mM, 20 hrs | 6.0 | 5.5 | 90 | 10.0 | 2.4 | † |
| | 0.4 mM, 20 hrs | 6.5 | 5.5 | 76 | 19.0 | 3.7 | † |
| CH ₃ COOTl | 0.8 mM, 17 hrs | 4.9 | 5.5 | 87 | 15 | 1.4 | † |
| TAMM | 0.05 mM, 30 min | 5.0 | 4.2 | 99 | 8.0 | 2.0 | |
| | 0.1 mM, 25 min | 5.0 | 2.9 | 98 | 4.5 | 1.5 | |
| H ₃ PO ₄ ·12WO ₃ | 0.2 mM, 6 hrs | 7.0 | 8.0 | 60 | 14.0 | 2.0 | † |
| | 1 m/p, 20 hrs | 4.4 | 3.7 | 97 | 4.5 | 1.2 | |
| KReO ₄ | 0.8 mM, 40 hrs | 4.9 | 5.4 | 82 | 10.0 | 1.7 | † |
| (CH ₃ COO) ₂ Pb | 0.2 mM, 18 hrs | 4.9 | 5.6 | 60 | 9.0 | 1.4 | † |
| Cl-Hg-(C ₆ H ₄)-I | 5 % sat., 25 hrs | 4.9 | 4.3 | 70 | 17.0 | 1.6 | † |
| | 2 % sat., 18 hrs | 4.5 | 4.0 | 91 | 14.0 | 3.6 | |

Table 3.4: Data collections : Orthorhombic form.

| Compound | Conditions | Res (Å) | R _{symm} (%) | Comp (%) | Diff (%) | K _{emp} | |
|----------------------------------|----------------------|------------|--------------------------|-------------|-------------|------------------|---|
| pCMBS | 1 m/p, 18 hrs | 4.5 | 3.8 | 98 | 4.0 | 1.5 | |
| | 3 m/p, 18 hrs | 5.0 | 3.4 | 94 | 5.4 | 1.6 | |
| | 6 m/p, 18 hrs | 5.0 | 3.9 | 80 | 7.3 | 1.9 | |
| | 12 m/p, 3 hrs | 4.8 | 3.5 | 97 | 7.5 | 1.4 | |
| | 9 m/p, 30 μM, 16 hrs | 5.0 | 4.0 | 90 | 7.9 | 1.7 | |
| | 9 m/p, 50 μM, 16 hrs | 4.9 | 3.6 | 94 | 8.0 | 1.6 | |
| Nb ₆ Cl ₁₄ | 1.5 mM, 21 hrs | 4.2 | 5.4 | 70 | 16.0 | 3.2 | † |
| | 1.0 mM, 48 hrs | 6.0 | 4.0 | 96 | 24.9 | 7.0 | |
| | 0.7 mM, 20 min | 6.5 | 4.4 | 90 | 8.9 | 2.9 | † |
| | 0.6 mM, 24 hrs | 5.0 | 3.9 | 5.4 | 15.0 | 2.3 | † |
| | 0.9 mM, 24 hrs | 5.0 | 3.6 | 96 | 20.0 | 6.0 | |
| | 1.2 mM, 20 hrs | 4.0 | 6.4 | 100 | 23.0 | 3.3 | ‡ |
| | 0.8 mM, 18 hrs | 6.0 | 6.5 | 100 | 13.0 | 2.1 | ‡ |
| | 1.2 mM, 26 hrs | 4.5 | 7.6 | 100 | 20.0 | 3.3 | ‡ |
| | 0.7 mM, 19 hrs | 3.5 | 4.0 | 100 | 18.0 | 2.7 | ‡ |
| | 0.7 mM, 23 hrs | 5.0 | 4.1 | 95 | 21.0 | 4.5 | |
| TiNO ₃ | 1 mM, 18 hrs | 5.5 | 4.9 | 87 | 14.8 | 1.2 | † |
| | 2 mM, 3 hrs | 5.0 | 3.0 | 46 | 4.5 | 1.5 | |
| TMLAc | 3 mM, 2 hrs | 5.0 | 3.8 | 99 | 11.0 | 1.2 | |
| | 3 mM, 16 hrs | 5.0 | 4.2 | 44 | 12.0 | 1.5 | |
| | 5 mM, 16 hrs | 5.5 | 5.0 | 87 | 15.1 | 1.4 | † |
| | 10 mM, 2 hrs | 5.0 | 4.1 | 38 | 9.0 | 1.6 | |
| | 14 mM, 16 hrs | 5.0 | 4.4 | 98 | 10.2 | 1.6 | |
| | 20 mM, 18 hrs | 5.0 | 3.5 | 97 | 8.8 | 1.5 | |

Table 3.4: Data collections : Orthorhombic form.

| Compound | Conditions | Res (Å) | R _{symm} (%) | Comp (%) | Diff (%) | K _{emp} | |
|---|--------------------------------|------------|--------------------------|-------------|-------------|------------------|---|
| KAu(CN) ₄ | 0.02 mM, 20 hrs | 4.5 | 3.9 | 84 | 6.4 | 1.6 | |
| | 0.05 mM, 14 hrs | 4.5 | 4.4 | 97 | 7.5 | 1.4 | |
| | 0.1 mM, 16 hrs | 5.0 | 3.9 | 77 | 6.5 | 1.9 | |
| | 0.2 mM, 14 hrs | 4.5 | 3.6 | 95 | 8.3 | 1.6 | |
| | 0.4 mM, 16 hrs | 4.5 | 4.8 | 95 | 8.5 | 2.0 | |
| | 1 mM, 16 hrs | 5.0 | 5.2 | 99 | 11.0 | 1.5 | |
| | 4 mM, 2 hrs | 5.0 | 4.8 | 93 | 15.0 | 1.7 | |
| | 4 mM, 16 hrs | 5.0 | 4.5 | 90 | 19.0 | 2.8 | |
| | 10 mM, 1.5 hrs | 5.0 | 5.0 | 96 | 12.3 | 1.8 | |
| | 4 mM, 18 hrs | 7.0 | 8.6 | 70 | 15.0 | 1.4 | † |
| UO ₂ (NO ₃) ₂ | 0.8 mM, 11 hrs | 6.0 | 5.4 | 85 | 32.0 | 6.2 | † |
| | 0.4 mM, 3 hrs | 5.0 | 3.5 | 90 | 15.0 | 1.4 | † |
| | 0.04 mM, 16 hrs | 5.0 | 4.0 | 95 | 6.0 | 1.9 | |
| K ₂ PtBr ₆ | 1 m/p, 20 hrs | 5.0 | 3.9 | 97 | 7.0 | 2.2 | |
| | 2 m/p, 20 hrs | 5.0 | 3.7 | 90 | 9.5 | 2.7 | |
| | 3 m/p, 20 hrs | 5.0 | 2.9 | 95 | 10.3 | 2.6 | |
| SiO ₂ ·12WO ₃ | 0.5 m/p, 20 hrs | 5.0 | 3.5 | 45 | 7.0 | 1.5 | |
| UO ₂ (CH ₃ COO) ₂ | 0.1 mM, 1.5 hrs | 5.0 | 5.0 | 81 | 15.0 | 2.0 | |
| K ₂ Pt(CN) ₄ | 0.4 mM, 20 hrs | 4.8 | 5.1 | 82 | 10.0 | 1.8 | † |
| Nb ₆ Cl ₁₄ & UO ₂ (NO ₃) ₂ | 2 mM, 20 hrs 0.4 mM, 4 hrs | 5.0 | 3.7 | 92 | 26.5 | 5.0 | |
| Nb ₆ Cl ₁₄ & H ₂ IrCl ₆ | 2 mM, 20 hrs 0.5 mM, 20 hrs | 5.0 | 3.7 | 96 | 33.0 | 4.2 | |
| pHMBA | 100 % sat., 20 hrs | 5.0 | 3.7 | 85 | 13.0 | 1.4 | † |

Table 3.4: Data collections : Orthorhombic form.

| Compound | Conditions | Res (Å) | R _{symm} (%) | Comp (%) | Diff (%) | K _{emp} |
|----------------------------------|-------------|------------|--------------------------|-------------|-------------|------------------|
| Nb ₆ Cl ₁₄ | 1 mM, 6 hrs | 7.0 | 5.0 | 99 | 23.0 | 10.0 |

Table 3.5: Data collection : Monoclinic form.

- † : Type I orthorhombic crystals.
- ‡ : Data collection of $hk0$ or $h0l$ terms only.
- Res** : Resolution cutoff for data set (Å).
- R_{symm}** : R-factor between symmetry related reflections (%).
- Comp** : Completeness of data set (%).
- Diff** : Mean fractional isomorphous difference (%).
- K_{emp}** : Overall $K_{emp} (= 2 \sum \Delta F_{iso} / \sum \Delta F_{ano})$.
- TMLAc** : $(CH_3)_3PbCH_2COOH$.
- pCMBS** : p-chloromercuribenzenesulphonic acid.
- pHMBA** : p-hydroxymercuribenzoic acid.
- TAMM** : Tetrakis-(acetoxymmercuri)methane.

3.3 Preliminary Analysis.

Good derivatives are those which give interpretable protein electron density maps, but, when interpretation of the observed isomorphous differences in terms of the correct heavy atom structure is complicated in itself, other criteria are needed if time and effort is to be saved.

Several such criteria have been proposed (see for example Leslie, A.G.W., 1991, Jones, Y. & Stuart, D., 1991). Most of these try to address the two most important questions about the quality of a derivative : isomorphism and degree

of substitution. For a good derivative the unit cell dimensions should remain unchanged, the mean fractional isomorphous difference —assuming a random distribution of the heavy atom sites— should decrease monotonically with resolution and in a manner very similar to the form factor of the replacement groups, the isomorphous difference Patterson maps calculated using data from different resolution shells should be very similar, the isomorphous and anomalous difference Patterson map should be closely related, and finally, both the isomorphous and the anomalous signal should be significantly above the noise level. When the heavy atom structure is known other statistics such as the phasing power or the figure of merit can give more reliable estimates of the “goodness” of a derivative (Ito, N., 1991).

The list of possibly useful derivatives of AhrC crystals is disappointingly short. Only five compounds gave any signs of specific binding and from those only one (the Niobium cluster, $\text{Nb}_6\text{Cl}_{14}$) has the characteristics of a well substituted and isomorphous (to at least 6\AA) derivative. A short analysis of the possibly useful data sets collected from orthorhombic AhrC crystals soaked in solutions containing heavy atoms is given below.

3.3.1 The Niobium Cluster, $\text{Nb}_6\text{Cl}_{14}$.

The lower halide complexes of Niobium and Tantalum (Nb_6X_{14} and Ta_6X_{14} , X : Cl or Br) have a long history. They were first synthesised at 1907 and their exact chemical formulas caused considerable discussion in the early literature (especially about the oxidation number of the metals). Vaughan, P.A, Sturdivant, J.H. and Pauling, L., 1950, determined the structures of the $\text{Nb}_6\text{Cl}_{12}^{++}$, $\text{Ta}_6\text{Cl}_{12}^{++}$ and $\text{Ta}_6\text{Br}_{12}^{++}$ groups through X-ray diffraction studies of their concentrated solutions in ethanol (two of the fourteen chlorine atoms are easily removed as was shown by Harned, H.S, 1913). The six Niobium atoms are at the corners of a regular octahedron whose edges are 2.98\AA long. The twelve chlorine atoms

are located on the radial perpendicular bisectors of the octahedron edges with a shortest Nb-Cl distance of approximately 2.36Å. The dimensions of the tantalum complex are very similar. It was later proposed that the similarity both in charge and dimensions of these two complexes could be used to form a pair of precisely isomorphous protein crystals, the only difference being the presence of one or the other of these clusters. Preliminary experiments with lysozyme crystals showed the feasibility of the method, but, due to the non-isomorphism between the modified and the native crystals, no further progress was made (Corey, R.B., *et al*, & Kay, L.M., 1962, Stanford, R.H., *et al*, & Corey, R.B., 1962). More recently, the Ta₆Br₁₄ cluster was used to prepare useful derivatives of two large proteins, transketolase and ribulose-1,5-bisphosphate carboxylase/oxygenase (Schneider, G. & Lindqvist, Y., 1994). The Nb₆Cl₁₄ cluster has been used in the structure determination of glutamine synthetase (Almasy, R.J., *et al*, & Eisenberg, D., 1986).

Comparison of a 8° precession photograph of the *hk0* zone from an orthorhombic crystal soaked for 20 hours in 2 mM Nb₆Cl₁₄ with the native pattern (Figure 3.1) shows significant differences, especially at low angles (as would be expected from a cluster). The unit cell dimensions are similar to those of the native crystals with an average fractional change of less than 0.3%. Although changes of this magnitude make the derivative not very useful for spacings less than 4Å (Crick, F.H.C. & Magdoff, B.S., 1956), the difference for the resolution range used in subsequent studies should be very small.

Several data sets have been collected from orthorhombic AhrC crystals soaked in solutions containing Nb₆Cl₁₄ with concentrations ranging from 0.5 to 2 mM (Table 3.4 and 3.5). In all cases the mean fractional isomorphous difference was significantly above the noise level (5%, equal to the mean fractional isomorphous difference between native crystals) and the same is true for the overall K_{emp}, although for most of the data sets no attempt was made to measure the anomalous

differences accurately (Section 3.1). A plot of the mean fractional isomorphous difference versus resolution and of the K_{emp} versus resolution is shown in Figure 3.2 for an orthorhombic crystal soaked for 19 hrs in 0.7 mM $\text{Nb}_6\text{Cl}_{14}$.

The isomorphous difference Patterson maps are very similar for different data collections. The linear (Pearson) correlation coefficients between isomorphous difference Patterson functions calculated using different data sets or different resolution ranges from the same data set (usually 13-9Å and 9-6Å) were in the range 0.60 to 0.77. Also, when a data set collected from a crystal soaked for 24 hrs in 0.6 mM $\text{Nb}_6\text{Cl}_{14}$ was treated as native and was compared with a data set collected from a crystal soaked for 21 hrs in 1.5 mM $\text{Nb}_6\text{Cl}_{14}$, the sets were found to be significantly different and the isomorphous difference Patterson functions were very similar to those obtained when the true native data set was used (suggesting again that the observed differences arise not from some sort of non-isomorphism, for example a rotation of the molecules, but from specific heavy atom binding). The agreement between the isomorphous and anomalous difference Patterson maps was poor for the majority of the data sets. This was expected since (i) the anomalous signal was well below the noise level and (ii) no attempt was made to measure the hkl and $h\bar{k}l$ terms under the same conditions (Section 3.1).

Figure 3.3 shows a sharpened difference Patterson function for $\text{Nb}_6\text{Cl}_{14}$ calculated using coefficients $[E_{\Delta F_{\text{isom}}} \Delta F_{\text{isom}}]^2$ where $E_{\Delta F_{\text{isom}}}$ are the normalised structure factors corresponding to the observed isomorphous differences $\Delta F_{\text{isom}} = || \mathbf{F}_{PH} | - | \mathbf{F}_P ||$. The normalised structure factors have not been calculated in the usual way, $E_{\Delta F_{\mathbf{h}}} = \Delta F_{\mathbf{h}} / (\varepsilon \sum_{j=1}^N f_j^2)^{1/2}$, because the number N of heavy atoms with an atomic scattering factor f is not known in advance. Instead, their calculation (programme ECALC) is based on their second most important property, namely that their distribution is independent of the complexity of the structure with $\langle E_{\mathbf{h}}^2 \rangle = 1$ (Woolfson, M.M., 1961). The use of sharpened Patterson functions was first suggested and their advantages demonstrated (using the [001]

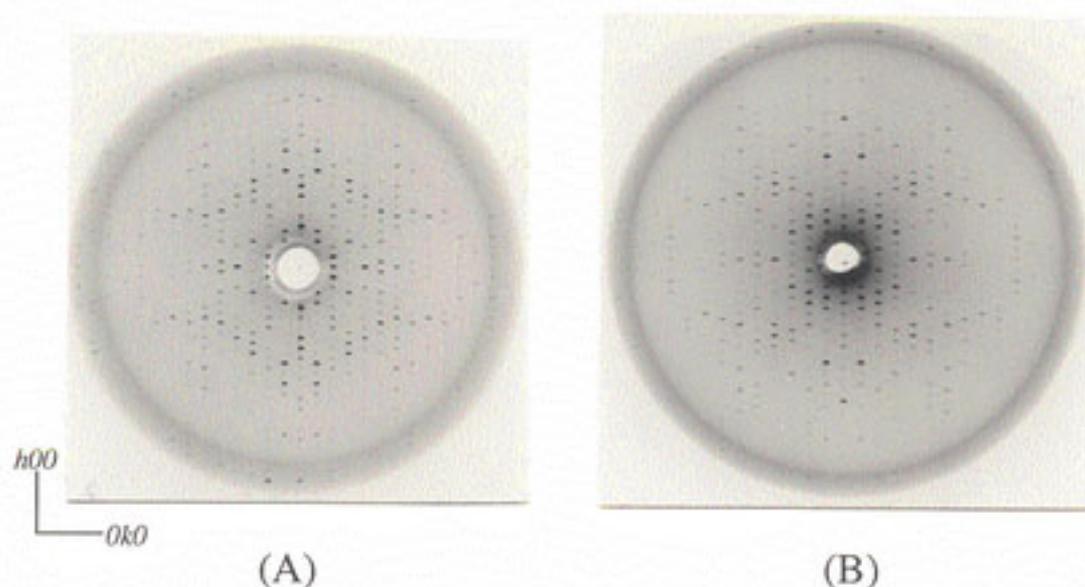


Figure 3.1: 8° precession photographs of the $hk0$ zone from (A) Native orthorhombic AhrC crystals and (B) Crystal soaked for 20 hrs in 2 mM $\text{Nb}_6\text{Cl}_{14}$.

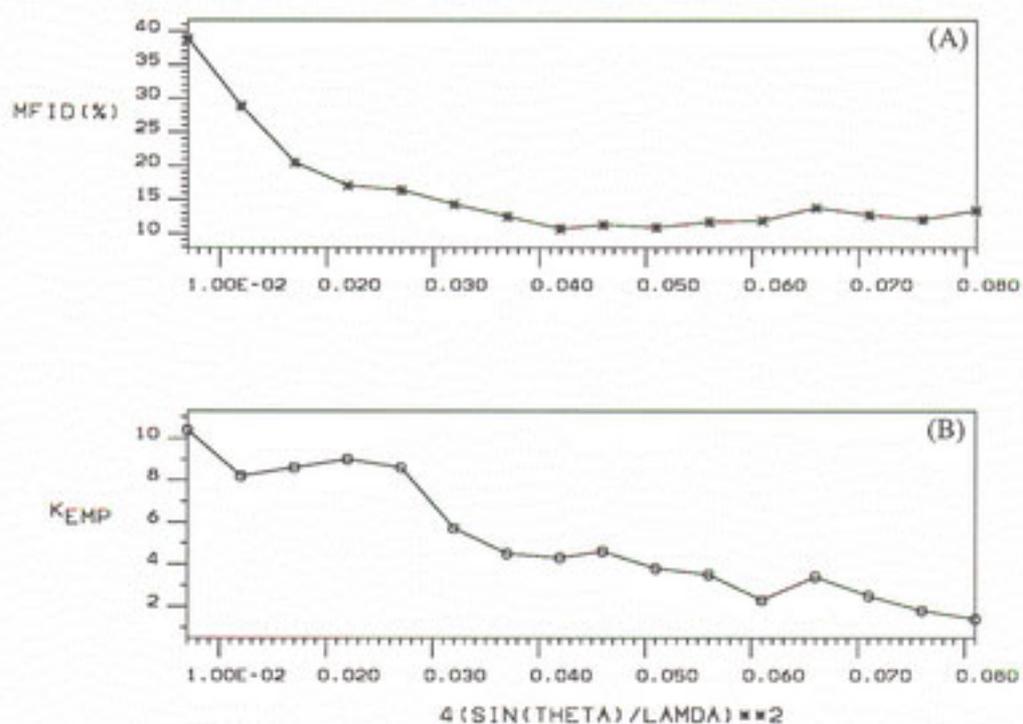


Figure 3.2: (A) Plot of the mean fractional isomorphous difference versus resolution for an orthorhombic AhrC crystal soaked for 19 hrs in 0.7 mM $\text{Nb}_6\text{Cl}_{14}$. (B) Plot of the K_{emp} versus resolution for the same crystal.

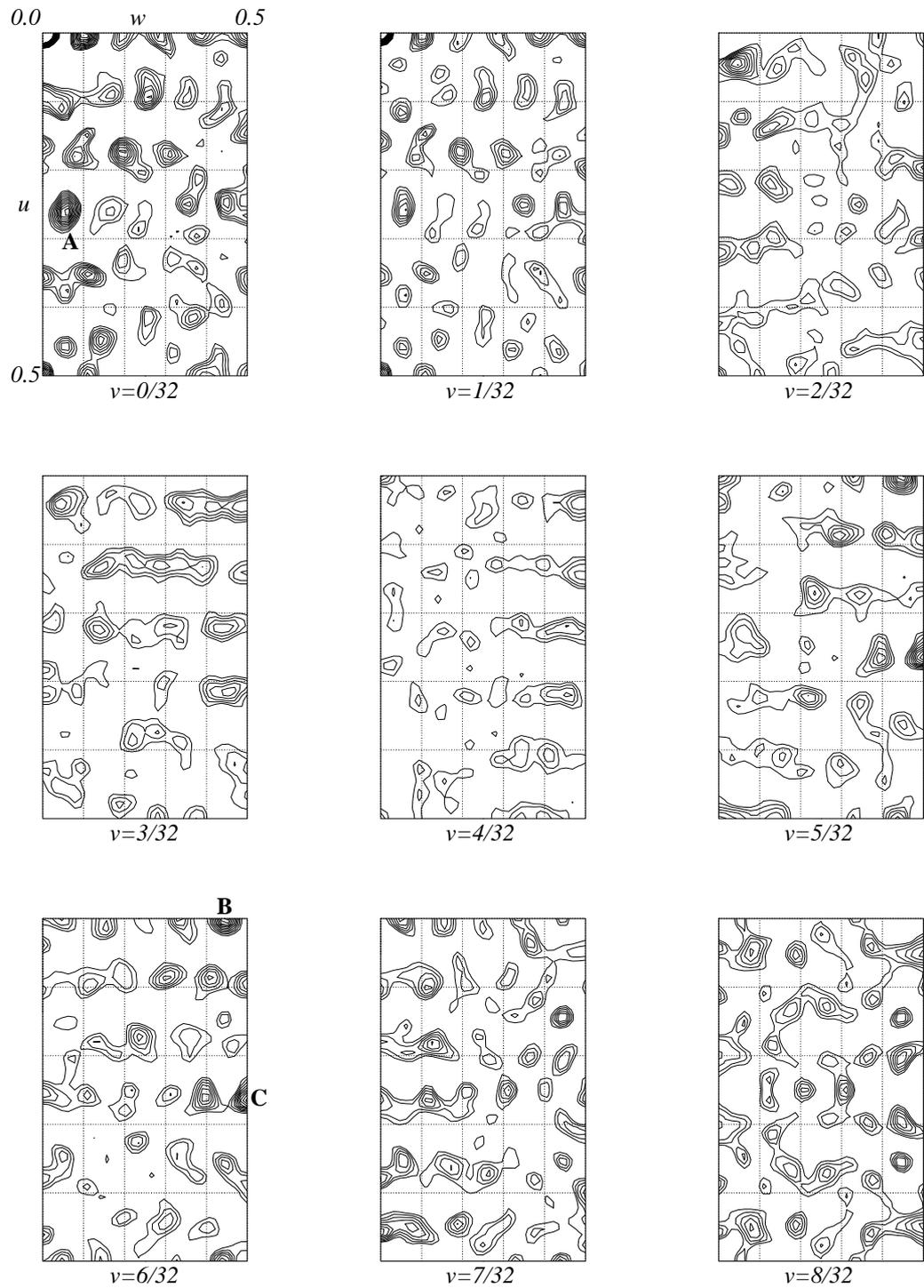


Figure 3.3: $13\text{-}\hat{\text{A}} [E_{\Delta F_{isom}} \Delta F_{isom}]^2$ synthesis for a type II crystal soaked in 0.9 mM $\text{Nb}_6\text{Cl}_{14}$ for 24 hrs. Contours every 2% of the origin peak.

Patterson projection of copper sulphate) by Patterson himself (Patterson, A.L., 1935). Although the calculation of sharpened Patterson functions is a standard procedure in small molecule crystallography (Hodgkin, D.C., 1987, Robinson, W. & Sheldrick, G.M., 1988, Sheldrick, G., 1986), their use in macromolecular crystallography is rather limited. With one or two hexamers per asymmetric unit in the case of AhrC, the interpretation of the isomorphous difference Patterson functions was expected to be difficult and the use of sharpened syntheses was thought to offer considerable advantages mainly due to their effectively higher resolution. This is more so for the $\text{Nb}_6\text{Cl}_{14}$ derivative since the scattering power of the Niobium cluster falls rapidly with resolution and a ΔF_{isom}^2 synthesis would be dominated by the large, low resolution terms, making the interpretation more difficult and less accurate. Figure 3.4 shows the $v=0.0$ section from (A) a ΔF_{isom}^2 and (B) a $[E_{\Delta F_{isom}} \Delta F_{isom}]^2$ synthesis for a type I orthorhombic crystal soaked in 1.5 mM $\text{Nb}_6\text{Cl}_{14}$ for 21 hrs. The approximate mirror plane at $w=0.25$ (arising from the presence of non-crystallographic symmetry) is obvious in the sharpened map but not in the ΔF_{isom}^2 synthesis. Other sharpened Patterson functions such as $[(E_{\Delta F_{isom}}^3 \Delta F_{isom}^2)^{1/2}]^2$ have also been examined but it was found that in some cases too much weight was given at relatively high resolution reflections with rather inaccurate measurements.

The space group of the Patterson function is Cmmm with 16 equivalent positions and the conventional asymmetric unit is $u : 0.0$ to 0.50 , $v : 0.0$ to 0.25 and $w : 0.0$ to 0.50 . The peaks marked as A, B, and C in Figure 3.3 are located on the Harker sections $u=0.0$, $v=0.0$ and $w=0.5$ and they form a self consistent set of Harker vectors corresponding to a site with coordinates $x=0.13$, $y=0.09$ and $z=0.225$ or equivalent by Patterson symmetry. Their heights relative to the origin peak are 27.5% 26% and 20% correspondingly. It was initially thought —mainly due to the appearance of the unsharpened Patterson function, Figure 3.4— that a derivative with one major substitution site had been obtained. Least squares

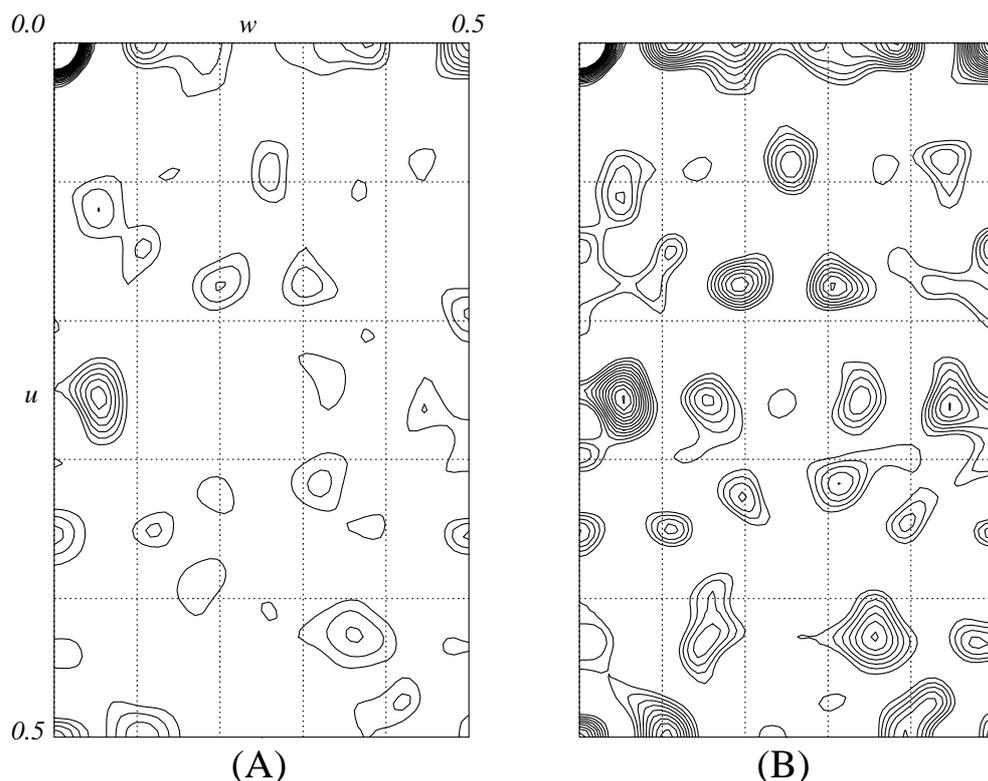


Figure 3.4: $v=0.0$ sections from (A) a ΔF_{isom}^2 and (B) a $[E_{\Delta F_{isom}} \Delta F_{isom}]^2$ synthesis for a type I crystal soaked in 1.5 mM Nb_6Cl_{14} for 21 hrs. Contours every 2% of the origin peak.

refinement of this site against the observed isomorphous differences for all centric reflections between 35 and 5Å resolution gave statistics of acceptable quality (programme REFIN from the CCP4 suite) :

The Cullis R-factor (Cullis, A.F., *et al*, & North, A.C.T., 1961), defined as

$$R_c = \frac{\sum |F_{H(obs)} - F_{H(calc)}|}{\sum F_{H(obs)}}$$

where $F_{H(obs)}$ and $F_{H(calc)}$ are the observed and calculated amplitudes of the heavy atom structure factors and the sums are taken over centric terms only , was 0.59. The weighted Cullis R-factor (wR_c) was 0.57 (the weights used were $1/\sigma^2$, where σ is the standard deviation of the observation based on counting statistics). The linear (Pearson) correlation coefficient between the observed and calculated

amplitudes of the heavy atom structure factors, defined as

$$C = \frac{\sum(F_{H(obs)} - \langle F_{H(obs)} \rangle)(F_{H(calc)} - \langle F_{H(calc)} \rangle)}{\sqrt{\sum(F_{H(obs)} - \langle F_{H(obs)} \rangle)^2 \sum(F_{H(calc)} - \langle F_{H(calc)} \rangle)^2}}$$

was 0.48. Finally, the gradient (slope) of the least squares line for a $F_{H(obs)}$ versus $F_{H(calc)}$ scatter plot was 0.43 (ideally it should be 1.0 corresponding to angle of 45°).

It should be noted that the heavy atom site was modelled using a single Nb^{3+} ion, with a form factor as given in International Tables for X-ray Crystallography, Vol. IV, 1974. Due to the high correlation between occupancy and temperature factor, they were refined independently in alternating cycles of least squares refinement. The temperature factor converged to a value $B=157\text{\AA}^2$ for a site with occupancy 14.5 (arbitrary units). This large temperature factor accounts for the increase of the scattered intensity at low angles due to the “in phase” scattering from the individual atoms of the cluster (the value $B=157\text{\AA}^2$ is in very good agreement with the temperature factor of 150\AA^2 for a $\text{Nb}_6\text{Cl}_{14}$ derivative used in the structure determination of Glutamine Synthetase (Almasy, R.J., *et al*, & Eisenberg, D., 1986)).

Single isomorphous replacement with anomalous scattering (SIRAS) phases were calculated from the refined atomic parameters using the programme PHASE from the CCP4 suite of programmes. The overall phasing power $\sqrt{\langle (F_{H(obs)}/\epsilon)^2 \rangle}$, where ϵ is the residual lack of closure, $\epsilon = F_{PH(obs)} - F_{PH(calc)}$, was 2.1 for all data between 35 and 6\AA . The mean figure of merit was $\langle m \rangle = 0.42$ with

$$m = \frac{\int_{\alpha=0}^{2\pi} P(\alpha) \exp(i\alpha) d\alpha}{\int_{\alpha=0}^{2\pi} P(\alpha) d\alpha} \quad \text{and} \quad P(\alpha) = \exp(-\sum \epsilon(\alpha)^2 / 2E^2)$$

where $P(\alpha)$ is the normalised probability of the phase α being correct and E is the total error which was estimated from centrosymmetric projections : $\langle E^2 \rangle = \langle (|F_{PH} - F_P| - F_{H(calc)})^2 \rangle$ (Blow, D.M. & Crick, F.H.C, 1959, Ramachandran, G.N. & Srinivasan, R., 1970, Blundell, T.L. & Johnson, L.N., 1976).

Several attempts have been made to determine the heavy atom structure of other possibly useful derivatives using cross difference Fourier syntheses, $m(F_{PH} - F_P)e^{i\alpha P}$, followed by comparison of the observed isomorphous difference Patterson functions with those calculated from the heavy atom structures suggested by the difference Fourier syntheses. In all cases the agreement between the observed and the calculated Patterson functions was poor. Double difference Fourier syntheses, $m(|F_{PH(obs)} - F_{PH(calc)}|)e^{i\alpha P}$, allowed the identification of five possible minor sites for the $\text{Nb}_6\text{Cl}_{14}$ derivative. Inclusion of these sites improved the quality of the statistics both in least squares refinement and phase calculation but not the quality of the phases themselves.

These problems suggested that the heavy atom structure of the $\text{Nb}_6\text{Cl}_{14}$ derivative was more complex than initially thought. Further attempts to determine the heavy atom structure of this derivative are described in Chapter 5.

3.3.2 The H_2IrCl_6 Derivative.

Orthorhombic AhrC crystals soaked in solutions containing H_2IrCl_6 with concentrations ranging from 0.6 to 1.2 mM were found to be significantly different from the native crystals. A 8° precession photograph of the $hk0$ zone from a crystal soaked for 12 hours in 0.75 mM H_2IrCl_6 is shown in Figure 3.5. A plot of the mean fractional isomorphous difference versus resolution and of the K_{emp} versus resolution for an orthorhombic crystal soaked for 18 hrs in 0.8 mM H_2IrCl_6 is shown in Figure 3.6. It is obvious that Iridium is also behaving like a cluster with very large isomorphous differences at low angles and a rapid fall off with increasing resolution. Although H_2IrCl_6 is not expected to be a cluster, Ito, N., 1991, has also found that this compound binds as a cluster with three Iridium atoms in the major substitution site of Galactose Oxidase. It is worth noting that H_2IrCl_6 has a rather complicated photochemistry and that no precautions had been taken to keep the compound in the dark.

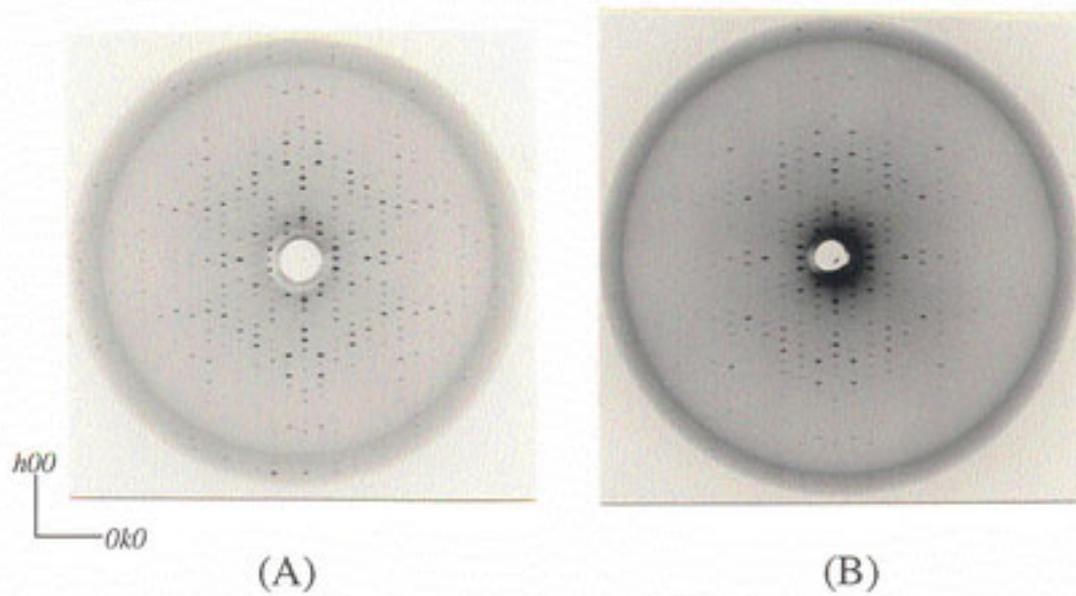


Figure 3.5: 8° precession photographs of the $hk0$ zone from (A) Native orthorhombic AhrC crystals and (B) Crystal soaked for 12 hrs in $0.75 \text{ mM H}_2\text{IrCl}_6$.

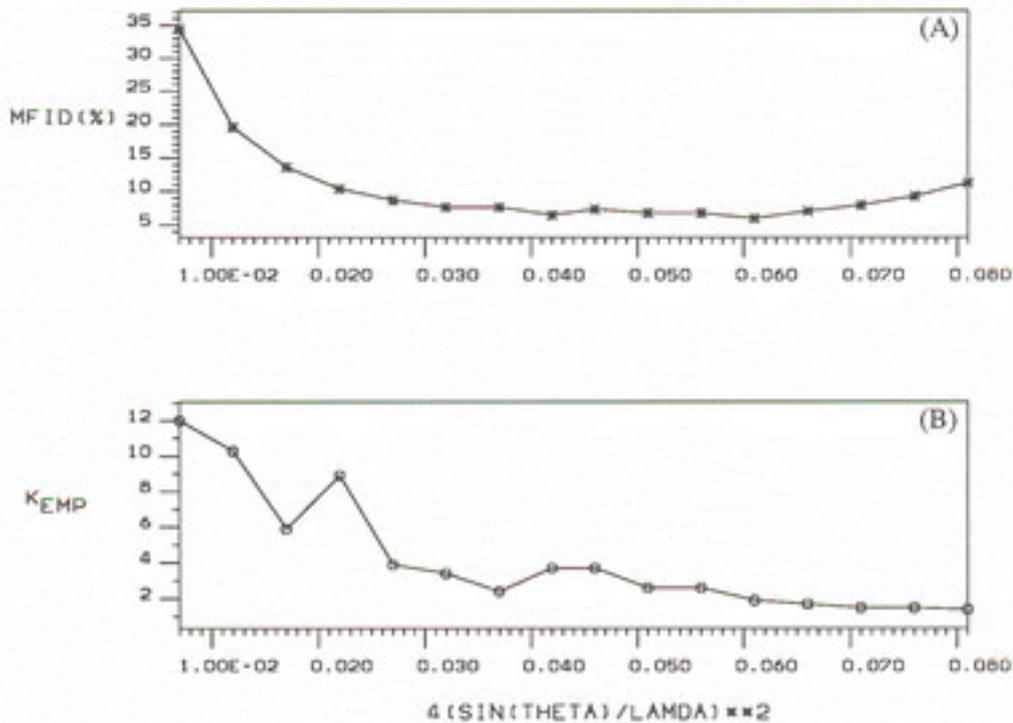


Figure 3.6: (A) Plot of the mean fractional isomorphous difference versus resolution for an orthorhombic AhrC crystal soaked for 18 hrs in $0.8 \text{ mM H}_2\text{IrCl}_6$. (B) Plot of the K_{EMP} versus resolution for the same crystal.

The $[E_{\Delta F_{isom}} \Delta F_{isom}]^2$ Patterson map (Section 3.3.1) for H_2IrCl_6 is shown in Figure 3.7. It is obvious that this Patterson function can not be immediately interpreted in terms of the heavy atom structure. Attempts to determine the heavy atom structure of this derivative are described in Chapter 5.

3.3.3 Other Derivatives.

As well as the Niobium and Iridium clusters, Uranyl nitrate, Uranyl acetate, $\text{KAu}(\text{CN})_4$ and p-iodochloromercuribenzoic acid (pICMB) showed signs of specific heavy atom binding. The isomorphism for all these derivatives is not well preserved (Figure 3.8). The isomorphous difference Patterson maps for most of these data sets and especially for the uranyl derivatives were featureless. This was not surprising as the low specificity of Uranyl is well documented (Blundell, T.L. & Johnson, L.N., 1976). The $\text{KAu}(\text{CN})_4$ and pICMB derivatives were expected to be more useful but, again, the isomorphous difference Patterson functions consisted of a more or less uniform distribution of peaks. Some of the methods described in Chapter 5 have also been applied to these two derivatives.

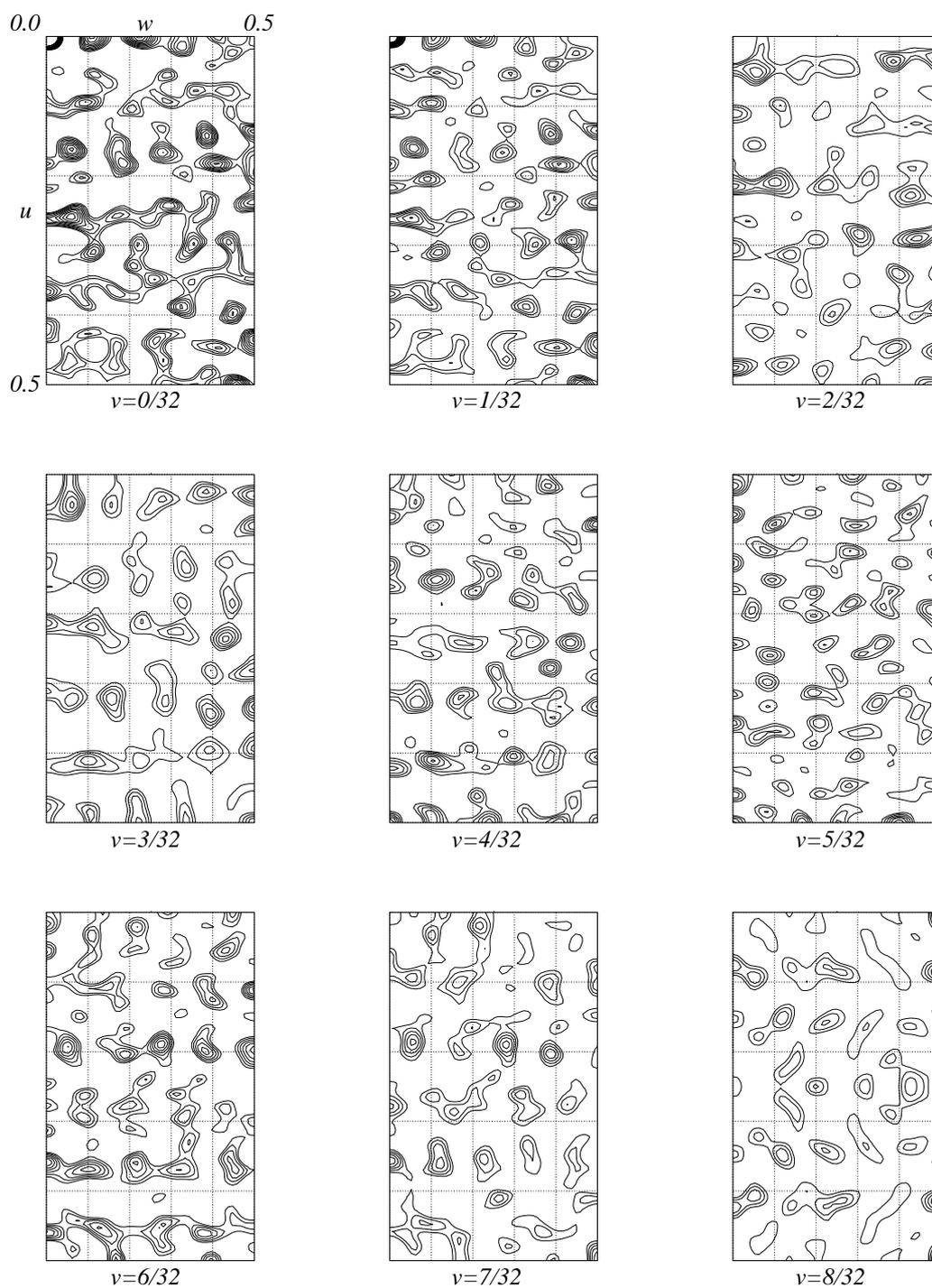


Figure 3.7: $13\text{-}\text{\AA}$ $[E_{\Delta F_{isom}} \Delta F_{isom}]^2$ Patterson synthesis for a type II crystal soaked in 1.2 mM H_2IrCl_6 for 3 hrs. Contours every 2% of the origin peak.

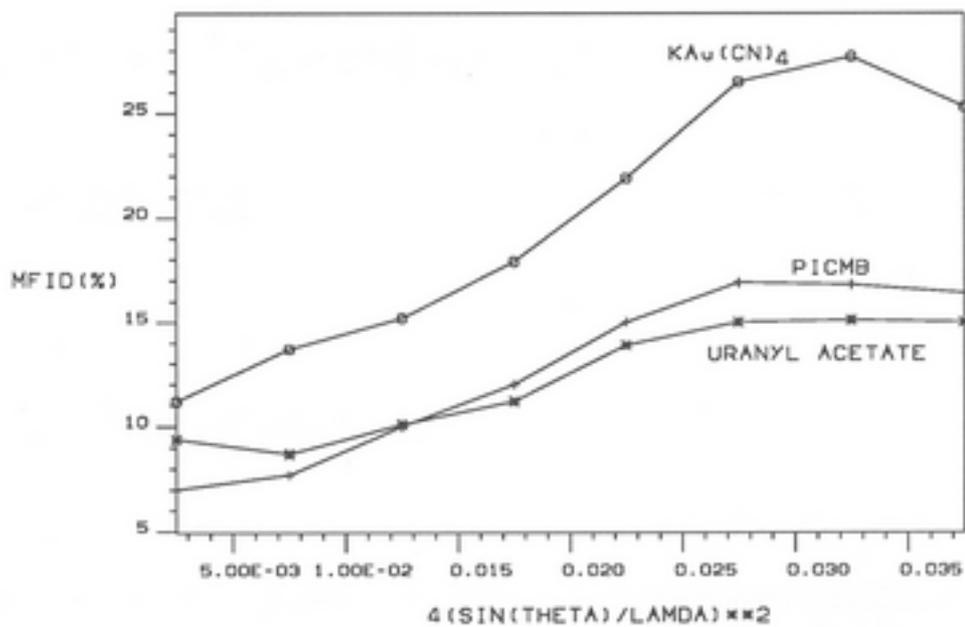


Figure 3.8: Mean fractional isomorphous difference versus resolution for type II crystals soaked in (A) 5% sat. PICMB for 25 hrs, (B) 0.1 mM Uranyl acetate for 1.5 hrs and (C) 4mM $\text{KAu}(\text{CN})_4$ for 16 hrs.

Chapter 4

Self Rotation Function Studies and a Model of the Crystal Packing

4.1 Introduction.

Our attempts to prepare derivatives of the orthorhombic AhrC crystals useful for a medium to high resolution three-dimensional structure determination using multiple isomorphous replacement have been rather unsuccessful (Chapter 3). The most useful derivative is the Niobium cluster $\text{Nb}_6\text{Cl}_{14}$ whose phasing power at resolution higher than about 5\AA is very low. If high non-crystallographic symmetry were not present, a low resolution study of AhrC would probably be the only option available. But the presence of a six-fold redundancy in the observed data together with the relatively high solvent content of the orthorhombic form, makes possible the refinement and subsequent extension of a low resolution phase set to the limit of the observed diffraction of the native crystals (2.9\AA). The theory and practice of phase refinement and extension using real space averaging and solvent flattening (or its reciprocal space equivalent in the form of the Molecular

Replacement equations) is very well documented and will not be discussed in detail (Theory and Reviews : Rossmann, M.G. & Blow, D.M., 1963, Rossmann, M.G. & Blow, D.M., 1964, Main, P. & Rossmann, G.M., 1966, Crowther, R.A., 1967, Main, P., 1967, Crowther, R.A., 1969, Bricogne, G., 1974, Bricogne, G., 1976, Rossmann, M.G., 1990, Lawrence, M.C., 1991, Rossmann, M.G. *et al*, & Choi, H.-K. & Lynch, R.E., 1992, Brünger, A.T. & Nilges, M., 1993, Applications : Jack, A., 1973, Argos, P., Ford, G.C. & Rossmann, M.G., 1975, Champness, J.N., Bloomer, A.C., Bricogne, G., Butler, P.J.G. & Klug, A., 1976, Winkler, F.K., Schutt, C.E. & Harrison, S.C., 1977, Bloomer, A.C., Champness, J.N., Bricogne, G., Staden, R. & Klug, A., 1978, Harrison, S.C., Olson, A.J., Schutt, C.E. & Winkler, F.K., 1978, Rayment, I., *et al*, & Johnson, J.E., 1978, Nordman, C.E., 1980, Rayment, I., Baker, T.S., Caspar, D.L.D. & Murakami, W.T., 1982, Rayment, I., 1983, Rayment, I., Baker, T.S. & Caspar, D.L.D., 1983, Gaykema, W.P.J. *et al*, & Beintema, J.J., 1984, Rossmann, M.G., *et al*, & Vriend, G., 1985, Almassy, R.J., *et al*, & Eisenberg, D., 1986, Arnold, E., *at al*, & Rossmann, M.G., 1987, Luo, M., *et al*, & Palmenberg, A.C., 1987, Luo, M., Vriend, G., Kamer, G. & Rossmann, M.G., 1989, Acharya, R., *et al*, & Brown, F., 1989, Jones, E.Y., Walker, N.P.C. & Stuart, D.I., 1991, Wu, H., Keller, W., Rossmann, M.G., 1993, Fry, E., Acharya, R. & Stuart, D.I., 1993, Tête-Favier, F., Rondeau, J.-M., Podjarny, A. & Moras, D., 1993).

For the method to be applicable it is required that, (i) both the position and the orientation of the non-crystallographic symmetry axes (with respect to the crystallographic frame) are known, (ii) a sufficiently detailed envelope of the repeating unit (or the whole assembly if the local axes form a point group) is available and (iii) a starting phase set has been determined by other methods.

The orientation of the local symmetry axes is usually determined through a study of the self rotation function of the native crystals or from the heavy atom positions of a derivative (if these can be determined independently). The posi-

tions of the local axes in the crystallographic frame can be determined from the heavy atom positions or ‘packing considerations’ or from the special positions of the molecules. An initial envelope is obtained from the starting phase set or from a low resolution model determined for example from three-dimensional reconstruction using electron microscopy. The starting phase set is usually determined from multiple isomorphous replacement or from a model, although in cases of very high non-crystallographic symmetry convergence can be achieved even from random initial phase sets (Jack, A., 1973).

The strategy for the structure determination of AhrC can now be formulated as follows : (A) Identification of the non-crystallographic symmetry axes and determination of their orientation, (B) Determination of the heavy atom positions either independently or based on the knowledge of the orientation of the local symmetry axes (and assuming that the heavy atom structure exhibits the point group symmetry of the macromolecular assembly), (C) Determination of the position of the molecular centre from the heavy atom positions (assuming that the local symmetry axes form a point group), (D) Calculation of the initial isomorphous replacement protein map to as high resolution as possible (depending on the quality of the phases available), (E) Determination of an approximate envelope from the averaged isomorphous replacement map, (F) Refinement of phases at constant resolution through real space averaging and solvent flattening with optional redetermination of the envelope and refinement of the position of the molecular centre and of the orientation of the local axes, (G) Gradual phase extension and refinement to as high resolution as possible.

The power of the above procedure can be further increased, if needed, through simultaneous averaging using both the orthorhombic and monoclinic forms giving a total of an eighteen-fold redundancy in the observed data¹.

¹A major limitation here is the availability of only low, 4Å resolution data for the monoclinic form. This is due to the non-reproducibility of the crystallisation trials which resulted in only a small number of crystals being available for further analysis, Section 2.2.2.

In the presence of high non-crystallographic symmetry, the interpretation of isomorphous difference Patterson functions in the traditional way (identification of possible heavy atom sites from the Harker sections, elimination of spurious sites and reference of the correct sites to a common origin through examination of cross vectors) can become very complicated or impossible. In the case of the orthorhombic form of AhrC, a derivative with two sites per protomer would give 9120 non-origin vectors. Accidental (or systematic) overlap of those vectors can only make things worse. The standard method for the interpretation of the Patterson function in such cases has been described by Argos & Rossmann in two papers (Argos, P. & Rossmann, M.G., 1974, Argos, P. & Rossmann, M.G., 1976, but see also Tong, L. & Rossmann, M.G., 1993). A full description of the method is given in Section 5.3.6. In short, the procedure is based on using the known orientation of the local axes (possibly from a study of the self rotation function), to reduce the size of the problem to that of determining the position of the heavy atom sites in the *non-crystallographic* asymmetric unit. In the example given above with two sites per protomer of AhrC, the problem would be reduced to that of determining the positions of only two instead of twelve sites.

The preceding discussion makes clear that the determination of the non-crystallographic symmetry is of utmost importance not only for the process of phase refinement and phase extension but also for obtaining a phase set in the first place. Unfortunately, the identification and determination of the orientation of the local axes in the case of the orthorhombic form of AhrC proved difficult. For most of this study, and mainly due to the problems encountered with the interpretation of the self rotation functions, the non-crystallographic symmetry present in these crystals was considered to be unknown. Section 4.2 describes our attempts to determine the orientation of the local symmetry axes. Section 4.3 describes a model of the crystal packing of the orthorhombic form. Chapter 5 is an account of our attempts to determine the heavy atom structures of some of

the most promising heavy atom derivatives of orthorhombic AhrC crystals.

4.2 Self Rotation Function Studies.

A homo-hexamer like AhrC is expected to self-assemble in a regular (symmetrical) manner. The most common point group for homo-hexamers is 32 although the possibility of a point group 6 can not be excluded². The analysis of the trigonal form (Section 2.2.3) suggested that an intramolecular 3-fold is present, but these crystals are disordered and it is not improbable that the molecular symmetry and the disorder phenomena have a cause-effect relationship, in which case, no conclusions can (safely) be drawn about the symmetry of the molecule.

The self-rotation function (Rossmann, M.G. & Blow, D.M., 1962, Rossmann, M.G., 1972),

$$R(\mathbf{C}) = \int_{-\infty}^{+\infty} P(\mathbf{u})P(\mathbf{C}\mathbf{u})U(\mathbf{u})dV$$

where $P(\mathbf{u})$ is the value of the Patterson function at the end of the vector \mathbf{u} , \mathbf{C} is a 3×3 rotation matrix and $U(\mathbf{u})$ is a ‘shape’ function defining the volume of the integration around the origin of the Patterson function, will have a large value when \mathbf{C} corresponds to a symmetry operator (crystallographic or non-crystallographic) of the Patterson function. A systematic search of the asymmetric unit of the rotation function (that is, the calculation of $R(\mathbf{C})$ for all unique (for a given Laue group) rotation matrices \mathbf{C}) should give a map of the orientation of all symmetry axes present in the Patterson function (within the chosen integration volume).

In practice, interpretation of the self rotation function can be anything from straightforward, to very difficult (eg. Jones, E.Y., Walker, N.P.C & Stuart, D.I., 1991) or even misleadingly straightforward (eg. Åkervall, K., *et al*, & Moring,

²Strictly speaking, it is incorrect to use the crystallographic point group symbols to refer to molecular point group symmetry. On the understanding that the meaning of the symbol “32” is more obvious —both for the writer and the prospective readers— than that of the corresponding Schoenflies symbol, “ D_3 ”, we will continue using this incorrect symbolism throughout this thesis. The Schoenflies symbols corresponding to the 32 crystallographic point groups can be found in Table 3.9.1 (page 44) in the International Tables for X-ray Crystallography, Vol. I, 1952.

1971 but see also Klug, A., 1971 in the same volume). “Special” orientations of the non-crystallographic axes with respect to the crystallographic, “special” packing of the molecules in the cell, etc., can result in an uninterpretable, or (worse), a misleading self-rotation function. It is worth noting in this respect, that a recent survey of 129 protein crystal structures with more than one molecule per asymmetric unit, showed that in 83% of the structures examined, the orientation of the non-crystallographic symmetry axes was “special”, that is, parallel or orthogonal to face diagonals, body diagonals or unit cell edges (Wang, X. & Janin, J., 1993).

All self-rotation functions have been calculated using the ‘fast’ rotation function (Crowther, R.A., 1972) as implemented in the programme POLARRFN (written by Kabsch, W., and distributed with the CCP4 suite). The results will be presented in the form of a polar stereographic projection with ω defining the latitude and ϕ the longitude of a rotation axis of order $(360^\circ/\kappa)$ where κ is the angle of rotation (in degrees) about this axis. The orientation of the crystallographic frame is such that the x axis is horizontal ($\omega=90^\circ$, $\phi=0^\circ$), the y axis is vertical ($\omega=90^\circ$, $\phi=90^\circ$) and the z axis is normal to the plane of the paper ($\omega=0^\circ$, $\phi=0^\circ$). In all diagrams, contours are plotted every 5% of the origin peak with first contour at 5% (the origin peak corresponds to no rotation or a rotation which is equivalent to a crystallographic symmetry operator). The self rotation function has symmetry mmm, but four asymmetric units will always be shown.

4.2.1 Native Crystals.

Figure 4.1 is a collection of sections $\kappa=120^\circ$ (search for 3-folds) and $\kappa=180^\circ$ (search for 2-folds) from self rotation functions calculated using different resolution ranges or integration radii. All functions shown have been calculated using a native data set collected from one orthorhombic, type II, AhrC crystal. Very similar results have been obtained from all native data sets examined.

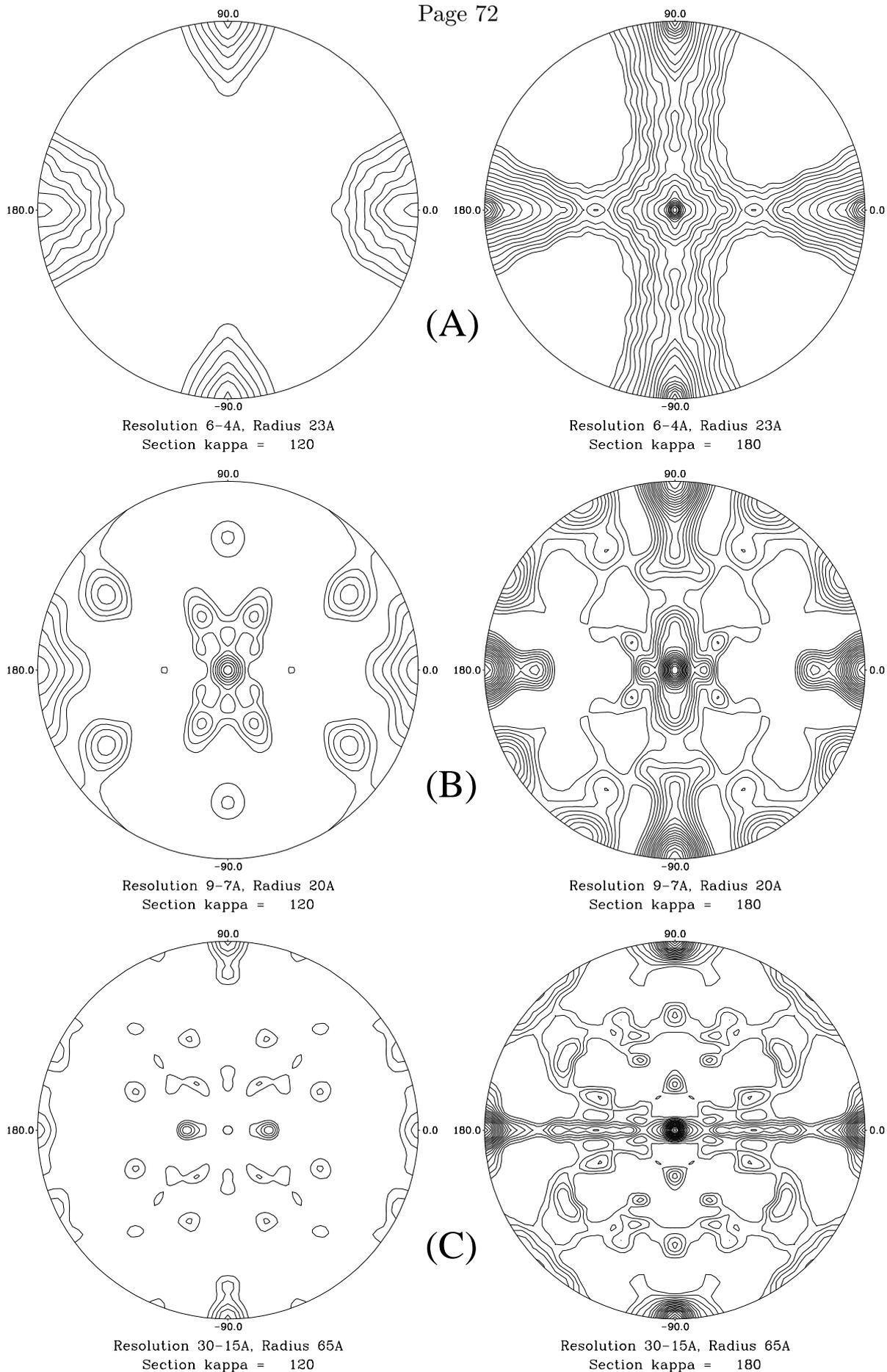
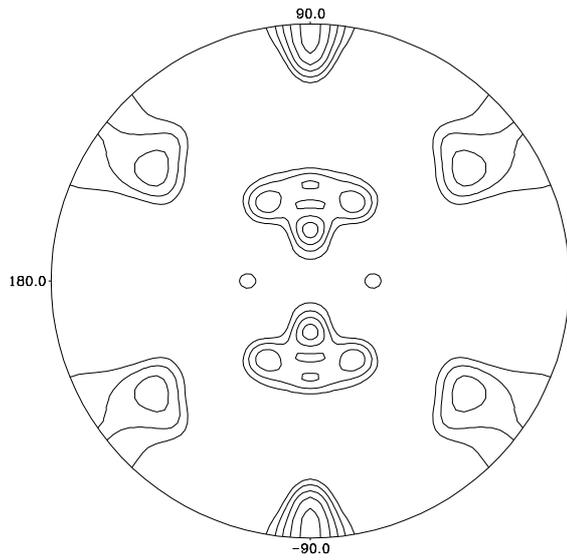
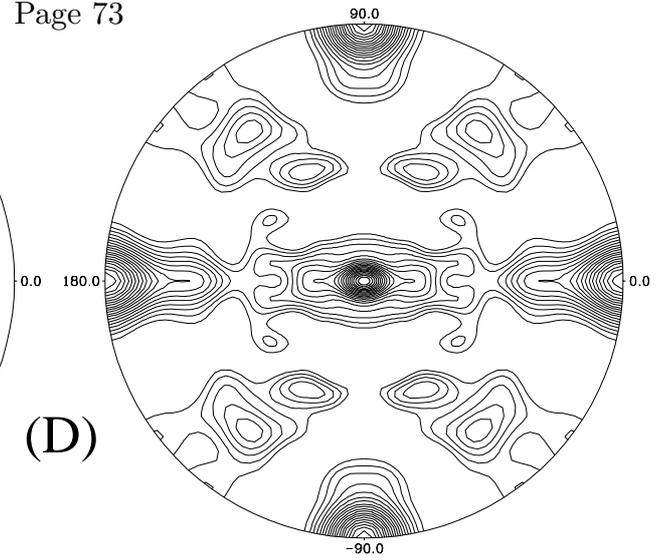


Figure 4.1: Self Rotation Functions for native AhrC crystals.

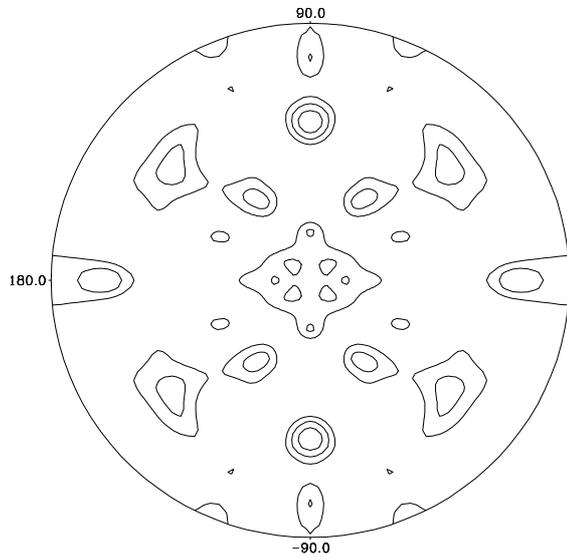


Resolution 17-8A, Radius 25A
Section kappa = 120

(D)

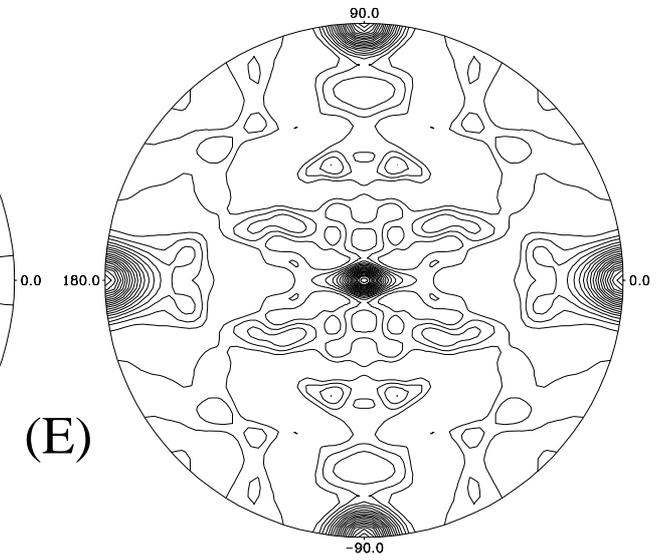


Resolution 17-8A, Radius 25A
Section kappa = 180

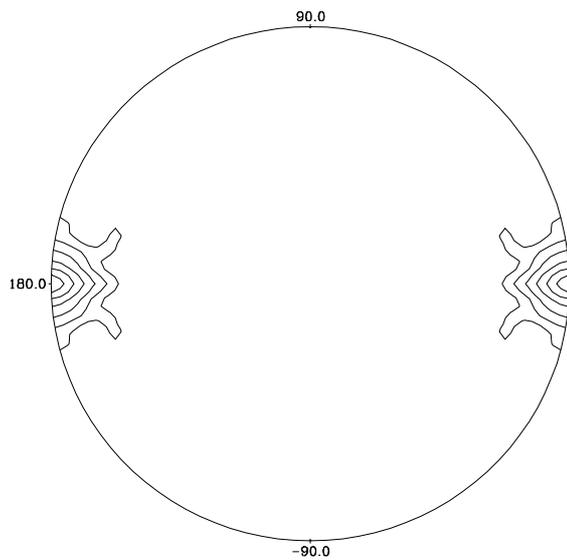


Resolution 13-7A, Radius 30A
Section kappa = 120

(E)

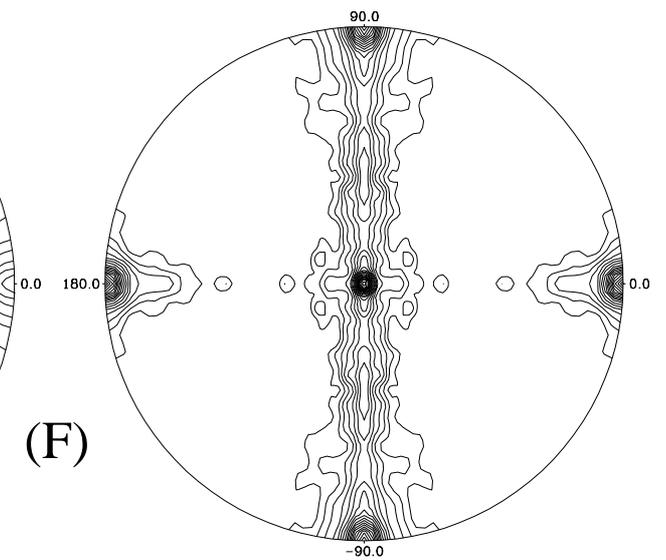


Resolution 13-7A, Radius 30A
Section kappa = 180



Resolution 8-5A, Radius 29A
Section kappa = 120

(F)



Resolution 8-5A, Radius 29A
Section kappa = 180

Figure 4.1: Self Rotation Functions for native AhrC crystals.

It is obvious that the appearance of the self-rotation function changes considerably when different resolution ranges or integration radii are used : in the various $\kappa=120^\circ$ sections shown there are indications of a non-crystallographic 3-fold along x or y or z or at $\omega=60^\circ$, $\phi=90^\circ$ or at $\omega=15^\circ$, $\phi=0^\circ$ etc. Although such a variability precludes a confident assignment of the orientation of the non-crystallographic symmetry axes, the consistency of the function when only medium to high resolution data are used is worth discussing :

All self rotation functions calculated using medium to high resolution data (Figure 4.1, A & F) indicate the presence of a non-crystallographic 3-fold axis parallel to x . The $\kappa=180^\circ$ sections from the same functions show strong peaks on the yz plane as would be expected from a 32 hexamer. The peak at $\omega=30^\circ$, $\phi=90^\circ$, $\kappa=180^\circ$ suggests that the orientation of the non-crystallographic 2-folds is such that one of them is parallel to y . This would require the presence of a strong (pseudo-origin) peak on the $v=0$ Harker section of the Patterson function for the native crystals (see Section 4.3 and Figure 4.5). A native Patterson function which was calculated using all data between 30 and 12Å showed no major features (the highest peak was 4% of the origin peak and it was located on the $w=0.5$ Harker section). The absence of strong features in the 12Å native Patterson function suggests that (i) there are no even-fold non-crystallographic symmetry axes parallel to the crystallographic, and (ii) there are no simple translations relating parts of the structure. It should be noted, however, that the above considerations can not exclude the possibility that the non-crystallographic 3-fold is parallel to x with one of the non-crystallographic 2-folds nearly, but not exactly, parallel to y .

4.2.2 Heavy Atom Structures.

An analysis of the non-crystallographic symmetry present in the heavy atom structure of a derivative is based on the same concepts as for the native crystals,

the only difference being the use of the isomorphous difference Patterson function instead of the native Patterson. For reasons that have already been discussed (Section 3.1.1), most of the calculations have been performed using a sharpened Patterson function.

Figure 4.2 shows the sections $\kappa=120^\circ$ and $\kappa=180^\circ$ from self rotation functions for the Niobium derivative which have been calculated using different resolution ranges. The strongest peaks on the $\kappa=120^\circ$ section are at $\omega=90^\circ$, $\phi=60^\circ$ and $\omega=60^\circ$, $\phi=90^\circ$. Although the exact positions and relative strengths of the various peaks are different when different resolution ranges are used for the calculation, the overall appearance of the functions is preserved. The same sections from the self rotation function for the Iridium derivative are shown in Figure 4.3. The major peak on the $\kappa=120^\circ$ section is again at $\omega=90^\circ$, $\phi=60^\circ$. Taken together, the self rotation functions for the Niobium and Iridium derivatives suggest the presence of a non-crystallographic 3-fold at $\omega=90^\circ$, $\phi=60^\circ$. Although this result is not consistent with any of the self rotation functions for the native crystals (Figure 1.4), several attempts were made to determine the heavy atom structure using the Argos & Rossmann method and assuming that the indications from the self rotation functions for the heavy atom structures are correct. These are discussed in Section 5.3.6.

In summary, a confident determination of the non-crystallographic symmetry of the orthorhombic AhrC crystals is not possible. The strongest indication from the native self rotation functions (a non-crystallographic 3-fold parallel to x) is not consistent with the self rotation functions calculated using the isomorphous difference Patterson function for two of the most promising derivatives.

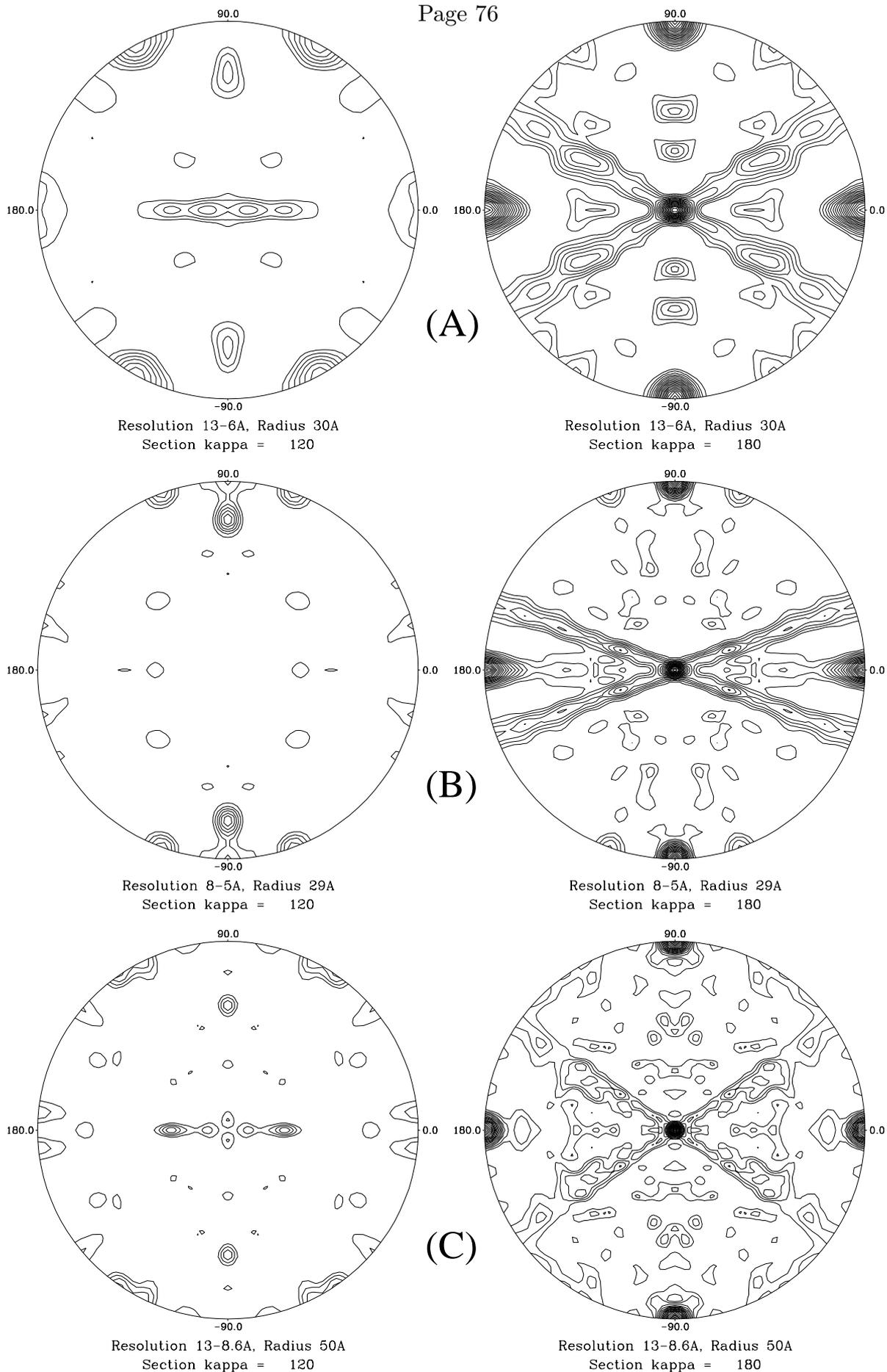


Figure 4.2: Self Rotation Functions for $\text{Nb}_6\text{Cl}_{14}$.

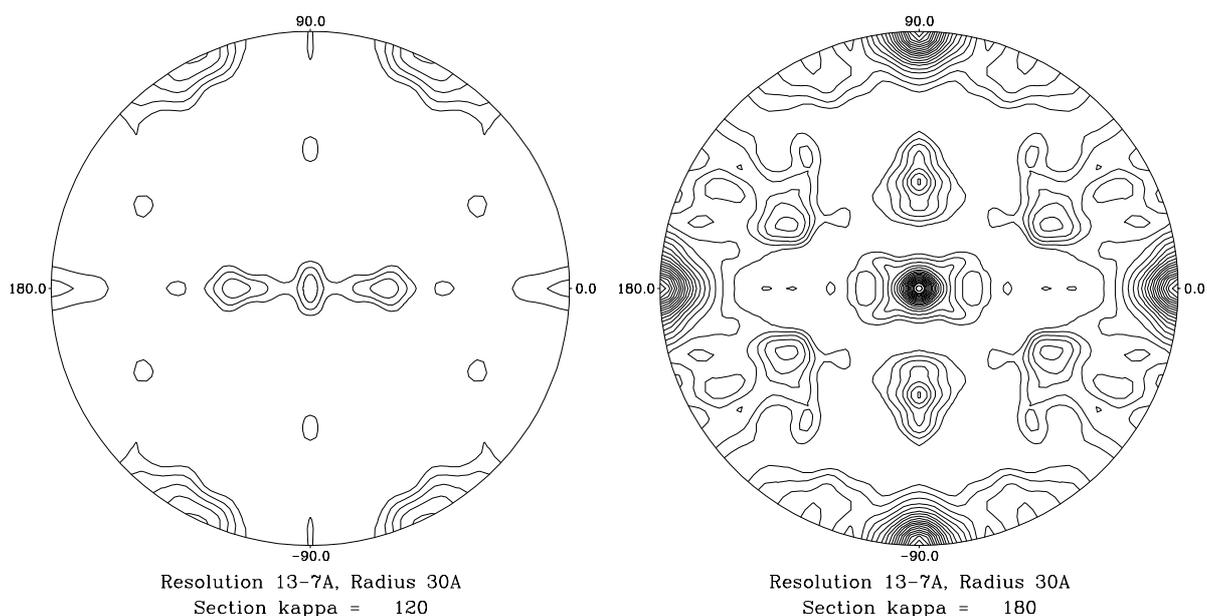


Figure 4.3: Self Rotation Function for H_2IrCl_6 .

4.3 A Model of the Crystal Packing.

4.3.1 Pseudo-origin Peaks.

The unit cell dimensions of the orthorhombic form AhrC crystals are $a=231.3\text{\AA}$, $b=74.4\text{\AA}$ and $c=138.0\text{\AA}$. If it is assumed that the AhrC hexamer (at very low resolution) is approximately spherical, then its estimated radius is $\approx 32\text{\AA}$. The very short b translation would suggest that in this direction the molecules do not overlap. An attempt was made to determine the approximate location of the molecules in this projection. The plane group of the projection of the electron density along $[010]$ is pmg . Although the origin for this plane group is conventionally taken at the 2-fold axis, we preferred to keep the origin at the same position as for the space group, that is, at the intersection of a mirror and glide line as shown in Figure 4.4A. Figure 4.4B shows the packing arrangement that minimises the overlap between crystallographically related molecules (each circle in this diagram represents the projection of an AhrC hexamer). The radius of

the circles is 29.2\AA and the fractional coordinates of the centre of the circle which is closest to the origin are $x=0.126$ and $z=0.212$. We saw in Chapter 3 that the fractional coordinates of an atom at the position indicated by the very strong peaks in the isomorphous difference Patterson function for the Niobium derivative are $x=0.13$, $y=0.09$ and $z=0.225$. The difference in the x and z coordinates is only 0.8 and 1.8\AA respectively.

The agreement between the position of a “heavy atom” as determined from the isomorphous difference Patterson for the Niobium cluster, and the position of the molecular centre deduced from packing considerations, suggests that the large peaks in the difference Patterson function are not due to the presence of a single major site, but arise from the “special” orientation of the non-crystallographic symmetry axes of the heavy atom structure : when even-fold non-crystallographic symmetry axes are parallel to crystallographic even-fold axes, large pseudo-origin peaks will be observed on the corresponding Harker sections (if the crystallographic axis is a 3-fold, then the heavy atom structure must have a non-crystallographic 3- or 6-fold parallel to the crystallographic for pseudo-origin peaks to be present). Figure 4.5A is a two-dimensional example showing a four-atom heavy atom structure with a non-crystallographic 2-fold axis parallel to a crystallographic one. Figure 4.5B is a schematic diagram of the Patterson function of this structure. The large non-origin Patterson peaks arise from the presence of a simple translation relating the two copies of the heavy atom structure. The position of these peaks depends only on the position of the non-crystallographic axes in the crystallographic frame and not on the heavy atom structure itself. The title ‘pseudo-origin’ peaks can now be justified : the pattern of vectors around the true origin peak and the large non-origin peaks is identical. To put this in a different way, the observed Patterson function is the convolution of the Patterson function of the isolated heavy atom structure with a periodic function consisting of only the origin and pseudo-origin peaks (which

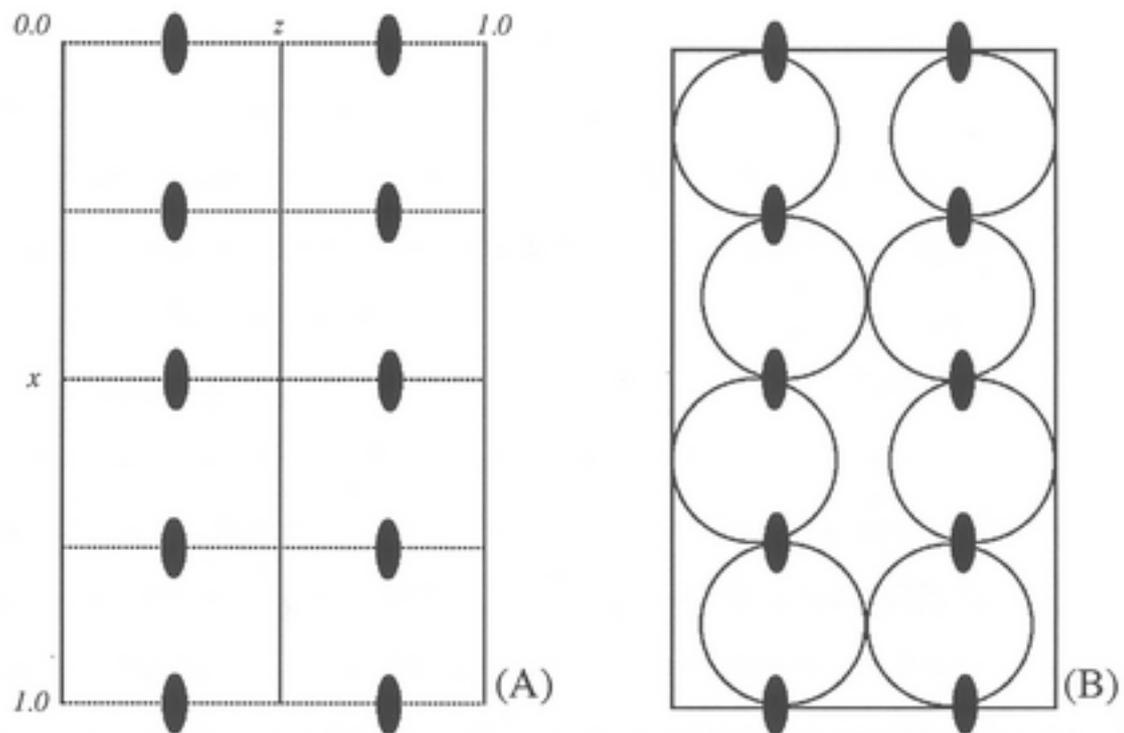


Figure 4.4: A : Symmetry elements in the $[010]$ projection of orthorhombic AhrC crystals, B : Closest non-overlapping packing of 8 spherical molecules in the $[010]$ projection.

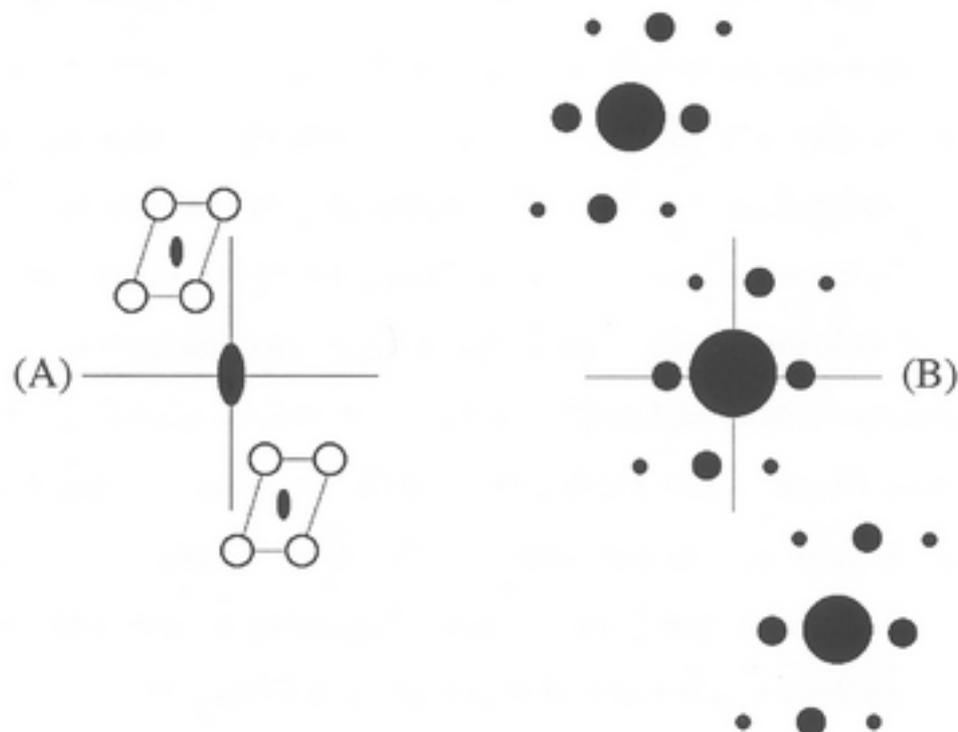


Figure 4.5: A : A hypothetical structure consisting of four atoms related by a non-crystallographic 2-fold axis parallel to a crystallographic 2-fold, B : Schematic diagram of the Patterson function of this structure.

are 'structure-independent'). This provides a means for obtaining the Patterson function of the heavy atom structure alone (as if there was no crystallographic symmetry present). A more complete discussion of the peculiarities arising from the presence of non-crystallographic symmetry axes which are parallel to crystallographic is given in Chapter 5.

This interpretation of the isomorphous difference Patterson function for the Niobium derivative suggests that the heavy atom structure has three intersecting and mutually perpendicular even-fold non-crystallographic axes. There are several point groups that satisfy this requirement (222, mmm, 422, $\bar{4}2m$, 4/mmm, 622, 6/mmm, 432, m3m) but none of them belongs to the trigonal system. If it is assumed that the heavy atom structure has the 32 point group as a subgroup, then the unavoidable conclusion is that the point group of the heavy atom structure must be 622, 6/mmm, 432 or m3m. The first two (hexagonal) point groups would require the presence of a non-crystallographic 6-fold parallel to one of the major axes, but this conclusion is not supported by the self rotation function of the heavy atom structure (Figure 4.2). The two cubic point groups are not entirely inconsistent with the observed self rotation function. Attempts to determine the heavy atom structure using the Argos & Rossmann method and assuming a cubic point group symmetry for the heavy atom structure, are described in Chapter 5.

The preceding analysis suggests that the molecular centre of an AhrC hexamer in the orthorhombic form is at $x=0.13$, $y=0.09$ and $z=0.225$ or equivalent by Patterson symmetry (the choice of enantiomorph is arbitrary at this stage). Further evidence supporting this model of the crystal packing came from (i) Examination of very low resolution permutation maps in the [010] projection (Section 4.3.2), (ii) A systematic search of the asymmetric unit using a packing function as a criterion (Section 4.3.3), and (iii) Electron microscopical studies of AhrC crystals (Chapter 6).

4.3.2 Very Low Resolution Permutation Syntheses.

The basis of the permutation syntheses method is very simple : given a set of observed amplitudes for a number of structure factors, Fourier syntheses are calculated for all unique phase combinations, and these are then examined for the presence of correct —or, as is usually the case, for the absence of unreasonable— features. The method is practical when (i) only a small number of centrosymmetric terms is to be examined, and (ii) a suitable selection criterion is available.

Woolfson, M.M., 1954, showed that examination of only 16 selected sign combinations of 7 terms will ensure that at least one of the syntheses will be calculated with 6 signs out of 7 correct (the total number of sign combinations for 7 terms is 128). Woolfson’s method has been of some use in small molecule crystallography (see for example Wright, W.B., 1958a and 1958b).

In protein crystallography, Boyes-Watson, J. and Perutz, M.F., 1943, used the permutation syntheses method to obtain “the first direct picture of a small protein molecule” by determining the signs of the four observed $h0l$ reflections from air-dried haemoglobin crystals.

We applied the method to the $[010]$ projection of the orthorhombic form. The total number of unique phase combinations for the four strongest low resolution reflections (201, 202, 002 and 402) is only 4 : the 002 and 402 terms are structure semi-invariants (that is, their signs are independent of the choice of origin) and all their sign combinations must be examined, but the 201 and 202 reflections belong to the two parity groups needed to fix the origin and signs can thus be allotted to them at will³. The four unique permutation maps are shown in Figure 4.6. The syntheses B, C and D are unreasonable since, (i) there are peaks with very high protein density either on the 2-folds or the mirror lines (B, C and D), (ii) the solvent areas are unreasonably large (B and D), and (iii) the density is

³It should be noted that because the origin is not at a 2-fold, not all reflections have phases 0 or π : the origin has been shifted by 0.25 along z , and so all reflections with $l = 2n + 1$ have phases $\pm\pi/2$

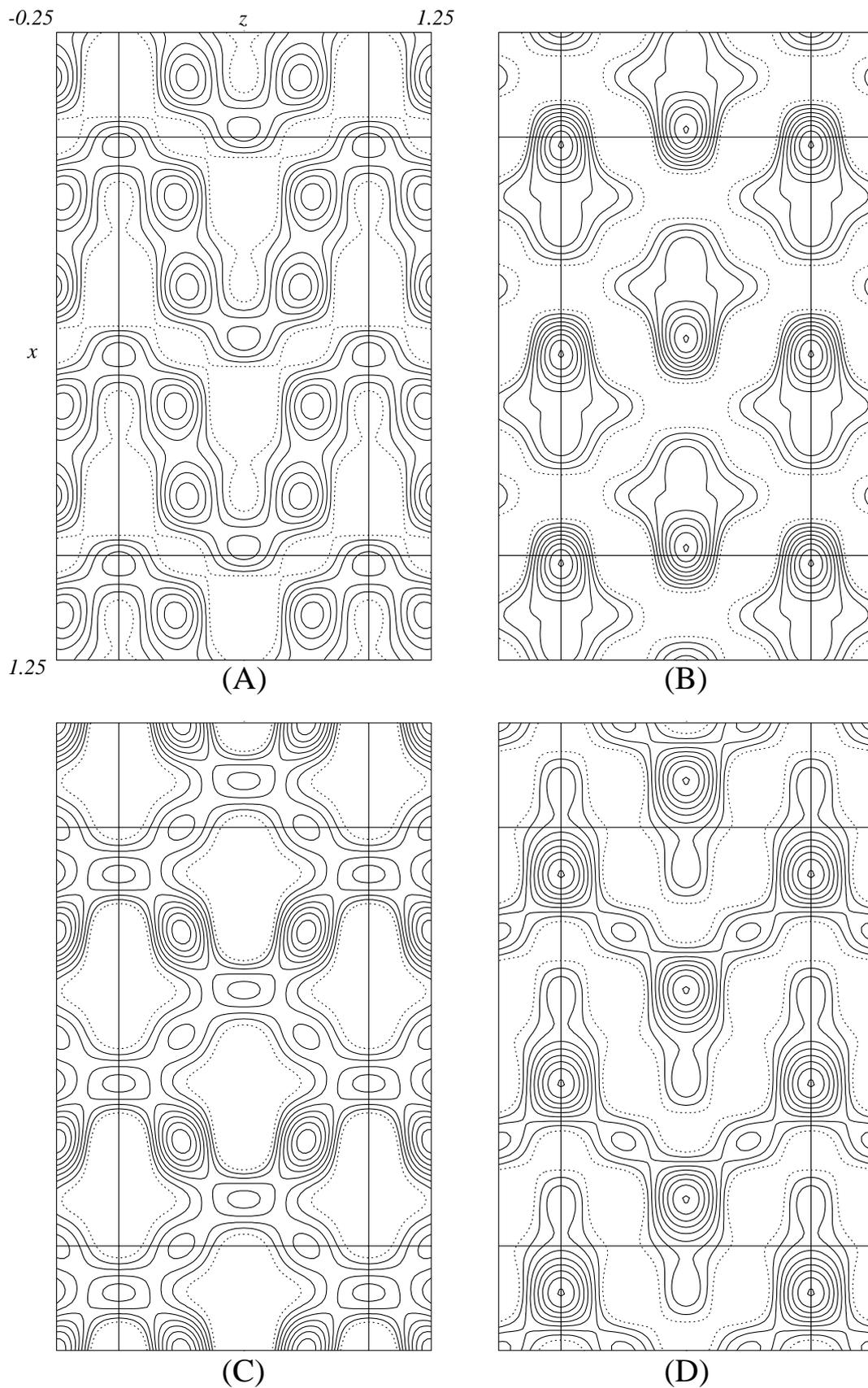


Figure 4.6: Permutation syntheses for the $[010]$ projection using the four strongest, low resolution, $h0l$ reflections. The lines enclose one unit cell.

not connected (B and D). Synthesis A shows a more or less uniform distribution of connected density with a reasonably small solvent area.

The position of the highest peak in this map is consistent with the packing model of the orthorhombic AhrC form and, as was found later in the course of the investigation, map A is as a whole in very good agreement with low resolution images of the [010] projection as seen in electron micrographs of negatively stained, crushed AhrC crystals (Chapter 6).

4.3.3 A Packing Function.

A programme was written which, for every unique packing arrangement of spheres of a given radius in the known orthorhombic unit cell, calculated the amount of overlap (expressed as fraction of the volume of the asymmetric unit) between crystallographically related molecules.

If the radius of the spheres is sufficiently large, the result from this search will be the closest, non-overlapping packing of spheres in the given cell. The search was conducted on a grid covering the asymmetric unit of the search function, which, since the origin and the enantiomorph are not fixed, is only 1/8th of the asymmetric unit of the space group. Three sections from the search with a sphere of radius 32\AA are shown in Figure 4.7. The peak seen on these sections correspond to an overlap between symmetry related spheres of less than 0.001% of the volume of the asymmetric unit. The position of the peak on the $y=7/36$ section is at $x=0.133$, $z=0.220$, in very good agreement with the packing model. The position along y is not well defined : all sections from $y=0.0$ to $y=0.25$ show the same peak with more or less the same amount of overlap between crystallographically related spheres. This is not surprising, since the unit cell along the [010] direction is only one molecule thick. The conclusion, therefore, is that although the x and z coordinates of the molecular centre can be determined from packing considerations alone, the y coordinate can only be determined accurately from the

pseudo-origin peaks on the $u=0.0$ and $w=0.5$ Harker sections of the isomorphous difference Patterson function for the Niobium derivative⁴.

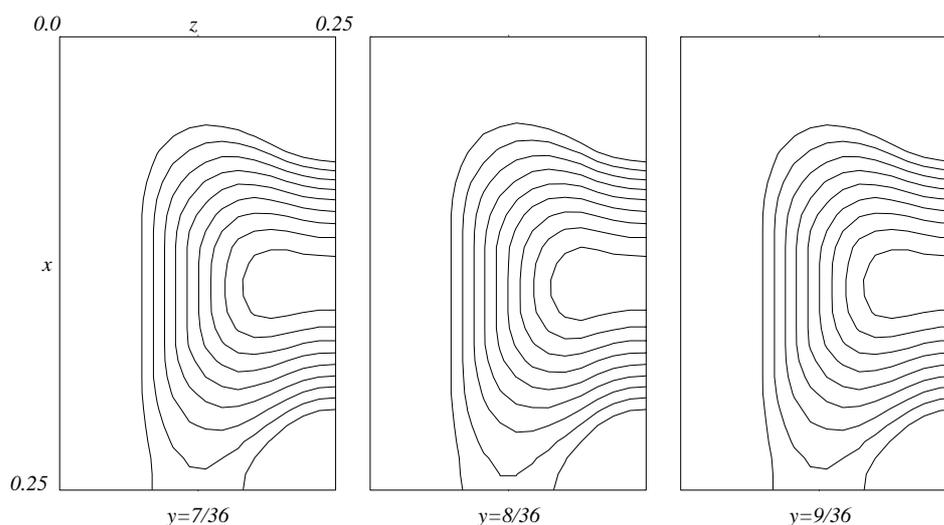


Figure 4.7: Three sections from the packing function. The radius of the spheres used for this search was 32\AA .

⁴ The validity of the results obtained from both the packing function (Section 4.3.3) and the packing considerations (discussed in Section 4.3.1), depends on whether the implicit assumption that the crystallographic asymmetric unit contains one hexamer (and not two independent trimers) is correct. There is no *a priori* reason why this should be so, and indeed, it is possible to arrive at sensible packing arrangements by assuming that crystallographic 2-fold axes coincide with non-crystallographic 2-fold axes of a 32 hexamer. The space group $C222_1$ has two sets of equivalent positions with point symmetry 2 : one set is at $x,0,0$ (and positions equivalent by crystallographic symmetry) and the second is at $0,y,0.25$ (and equivalent). If it is assumed that the molecules are approximately spherical and that they do not overlap in the $[010]$ projection, then there are only two additional packing arrangements that must be examined.

In the first (marked as A in Figure F.1) the two hexamers are at $x_1, 0, 0$ and $x_2, 0, 0$, and the molecules form columns parallel to x . In the second (F.1(B)) the two hexamers are at $x_1, 0, 0$ and $0, y_2, 0.25$, and the projection down the $[010]$ axis has an approximately hexagonal appearance. The calculation of the permutation syntheses shown in Figure 4.6 involves no assumptions about the packing arrangement and their inconsistency with both F.1(A) and F.1(B) suggests that the crystallographic asymmetric unit contains a single hexamer and not two independent trimers. This conclusion is consistent with the presence (and position) of the pseudo-origin peaks in the isomorphous difference Patterson function for the $\text{Nb}_6\text{Cl}_{14}$ derivative and the Electron Microscopical studies of AhrC crystals.

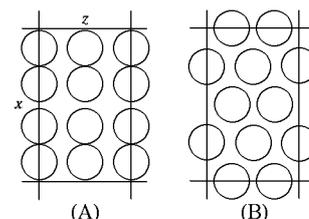


Figure F.1: Additional packing arrangements. The lines enclose one unit cell.

Figure 4.8 shows views of the crystal packing down the $[100]$, $[010]$ and $[001]$ axes. The molecules form layers parallel to the xy planes, with alternating layers being displaced by $\approx 13\text{\AA}$ in a direction parallel to y (most clearly seen by the zig-zag arrangement of molecules in the $[100]$ view). Large solvent channels ($\approx 40\text{\AA}$ in diameter) run parallel to z .

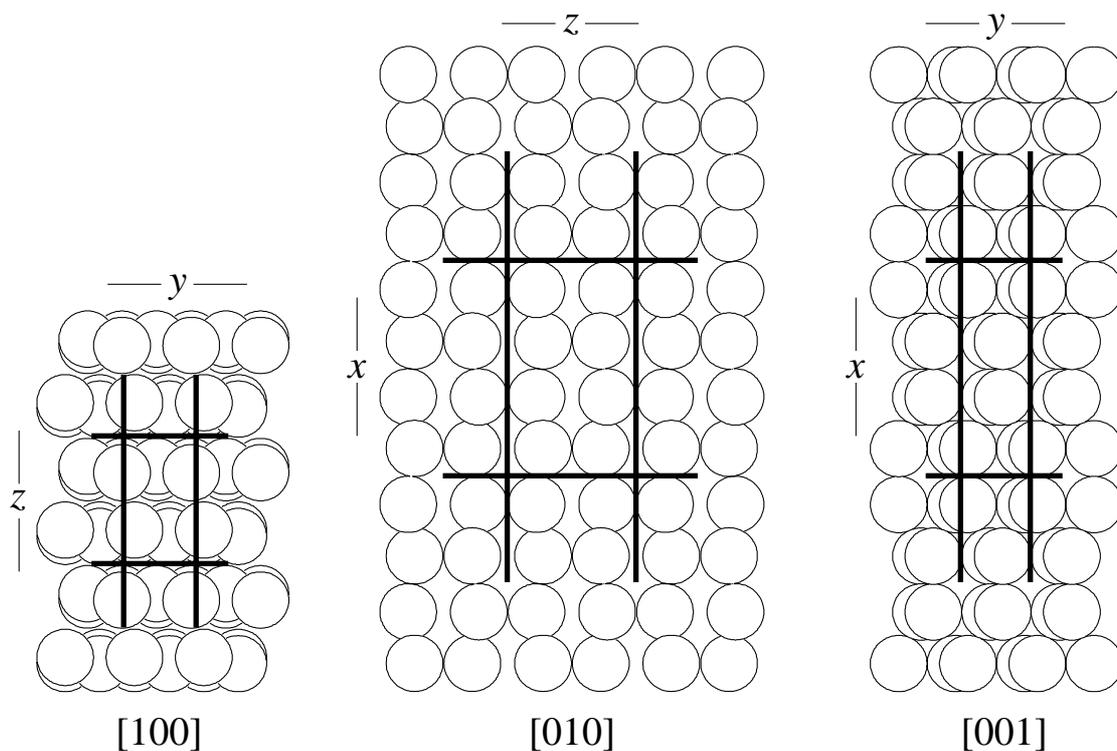


Figure 4.8: The crystal packing of the orthorhombic form. Views of the packing down the $[100]$, $[010]$ and $[001]$ axes are shown. The radius of the spheres (each representing an AhrC hexamer) is 30\AA . In all views 3×3 unit cells are shown.

Chapter 5

Attempts to Determine the Heavy Atom Structures of AhrC Derivatives

5.1 Introduction.

Chapter 4 made clear that the determination of the crystal structure of AhrC had to continue in the absence of a firm knowledge of the orientation of the non-crystallographic symmetry axes. This meant that the difference Patterson maps had to be interpreted in all their complexity. This Chapter describes the methods used in the attempt to determine the heavy atom structures of AhrC derivatives.

For clarity, we divided the methods we used into three major categories. In the direct methods category are included all methods based on an algebraic formalism describing probabilistic phase relationships between structure factors of known amplitudes. The second category includes those methods in which a direct interpretation of the Patterson function is being attempted. A third category includes methods that are not directly related to either direct or Patterson based methods. It should be noted that although direct and Patterson methods have

been kept as separate categories, these methods are related and in some cases they are formally equivalent (for example the Σ_1 formula of Hauptman & Karle, 1953 is directly related to the presence of Harker peaks in the Patterson function (Cochran, W. & Woolfson, M.M., 1954, Vand, V. & Pepinsky, R., 1954, Cochran, W., 1954)).

Several of the methods described in this Chapter attempt to determine the heavy atom structure from the centrosymmetric projections only. Although overlap is a serious problem when working with projections, the fact that the phases—for the conventional choice of origin—are either 0 or π , makes all analyses far more accurate¹. The amplitudes of the heavy atom structure (assuming that care has been taken to exclude the terms for which a “crossover”, $F_H = F_{PH} + F_P$, is possible) can be determined with an accuracy that is limited only by the standard deviations of the individual measurements. This makes possible (i) the calculation of reliable Patterson projection maps and (ii) the refinement of trial structures without the problems associated with estimating F_H for non-centrosymmetric terms. Finally, phases can be calculated for the centrosymmetric terms of the native crystals from only one derivative. Although a projection of the native crystals is not very useful in the context of a three-dimensional structure determination, in our case, where the low resolution projection maps along [010] and [001] were known in advance (Chapters 4 and 6), such a calculation can be seen as an independent test of the correctness of a proposed heavy atom structure.

5.2 Direct Methods.

The use of direct methods for the determination of the heavy atom positions in derivatives of protein crystals is well documented (Steitz, T.A., 1968, Neidle,

¹In the presence of anomalous scattering, the phase of \mathbf{F}_{PH} will, in general, be different from 0 or π for the conventional choice of origin. For the rest of this thesis, the assumption will be made that the imaginary dispersion correction term ($\Delta f''$) is so small that the errors introduced by treating \mathbf{F}_P and \mathbf{F}_{PH} as colinear are negligible compared with the errors of measurement.

S., 1973, Navia, M.A. & Sigler, P.B., 1974, Wilson, K.S., 1978, Sheldrick, G.M., 1991). These studies showed that (i) it is feasible to use direct methods to determine the heavy atom structure of derivatives from isomorphous difference data, even at resolutions as low as 8\AA , (ii) whereas some relationships appear to work well with ΔF data (such as the triple-product sign relationship), others (eg. negative quartets) are less effective. It is clear that relationships in which active use is made of very weak terms can not work properly with ΔF data, since a small ΔF can still correspond to a large F_H if $\mathbf{F}_H \perp \mathbf{F}_P$. This means that most of the recent developments in the field of direct methods (see Woolfson, M.M., 1987 and references therein) are not expected to be very useful for the interpretation of isomorphous difference data. What follows is a short description of only those formulae that are needed for a discussion of our attempts to determine the heavy atom structures.

Three papers, published in the same issue of *Acta Crystallographica*, showed how the constraints of both positivity (which alone leads to the determinantal inequalities of Karle & Hauptman, 1950) and atomicity of the electron density give rise to relationships between structure factors : Sayre, D., 1952 used the convolution theorem to show that for a structure consisting of equal resolved atoms :

$$\mathbf{F}_h = \frac{1}{VS_h} \sum_{\mathbf{k}} \mathbf{F}_k \mathbf{F}_{h-k} \quad (5.1)$$

where S_h is a function which accounts for the change of the atomic scattering factor of the atoms in the “squared” (ρ^2) structure and the summation is taken over all terms to infinity. (It should be noted that Sayre’s equation is an exact equality relationship valid for both centrosymmetric and non-centrosymmetric structures). Cochran, W., 1952 suggested that for such a structure, a Fourier synthesis calculated using only a subset of structure factors will be characterised by a large value of the integral $\int_V \rho^3 dV$. He then showed that the requirement

' $\int_V \rho^3 dV$ is maximum positive' is equivalent to

$$Q = \sum_{\mathbf{h}} \sum_{\mathbf{k}} \mathbf{F}_{\mathbf{h}} \mathbf{F}_{\mathbf{k}} \mathbf{F}_{\mathbf{h}+\mathbf{k}}, \quad \text{is maximum positive} \quad (5.2)$$

For a centrosymmetric structure, the value of Q will be maximum positive when most of the products $\mathbf{F}_{\mathbf{h}} \mathbf{F}_{\mathbf{k}} \mathbf{F}_{\mathbf{h}+\mathbf{k}}$ are positive. This led him to the triple-product sign relationship

$$s(\mathbf{h})s(\mathbf{k})s(\mathbf{h} + \mathbf{k}) \approx +1 \quad (5.3)$$

where $s(\mathbf{h})$ means 'sign of $\mathbf{F}_{\mathbf{h}}$ ' and \approx means 'probably equals'. Zachariasen, W.H., 1952 extended equation 5.3 to the general case when several pairs of \mathbf{k} and $\mathbf{h} + \mathbf{k}$ terms give sign indications for the \mathbf{h} term :

$$s(\mathbf{h}) \approx s \left[\sum_{\mathbf{k}} s(\mathbf{k})s(\mathbf{h} + \mathbf{k}) \right] \quad (5.4)$$

and used this equation to solve the structure of metaboric acid (this was the first application of what later became known as "The Symbolic Addition Procedure" (Karle, I.L. & Karle, J., 1963, Karle, I.L. & Karle, J., 1964, Karle, J. & Karle, I.L., 1966)). Equation 5.4 is equivalent to the Σ_2 formula of Hauptman & Karle, 1953. The derivation of the probability formulas for relationships 5.3 and 5.4 was not straightforward (see Klug, A., 1958 and references therein). Cochran & Woolfson, 1955 gave the first practically correct derivation, although it was Klug, A., 1958 who gave the complete (and complex) treatment. Cochran & Woolfson's probability formula for the most useful equation 5.4 is :

$$P_+(\mathbf{h}) \approx \frac{1}{2} + \frac{1}{2} \tanh \left[(\sigma_3 \sigma_2^{-3/2}) | E_{\mathbf{h}} | \sum_{\mathbf{k}} E_{\mathbf{k}} E_{\mathbf{h}+\mathbf{k}} \right] \quad (5.5)$$

where

$$\sigma_n = \sum_{j=1}^N Z_j^n$$

In the above equations Z_j is the atomic number of the j th atom, and $E_{\mathbf{h}}$ is the normalised structure factor.

For the case of a non-centrosymmetric crystal, Cochran, 1955 showed that

$$\phi(\mathbf{h}) \approx \phi(\mathbf{k}) + \phi(\mathbf{h} - \mathbf{k}) \quad (5.6)$$

where $\phi(\mathbf{h})$ is the phase angle of $\mathbf{F}_\mathbf{h}$. This is the triple-phase relationship. Although the exact formula for the probability distribution of the above equation will not be given here, it can be said that its general form is very similar to a Gaussian centered on the expected value of $\phi(\mathbf{h})$ (Cochran, 1955). Finally, Karle & Hauptman, 1956 produced the so called ‘tangent formula’ which can be used to estimate $\phi(\mathbf{h})$ when several pairs of known phases are available :

$$\tan(\phi(\mathbf{h})) \approx \frac{\sum_{\mathbf{k}} |E_{\mathbf{k}} E_{\mathbf{h}-\mathbf{k}}| \sin[\phi(\mathbf{k}) + \phi(\mathbf{h}-\mathbf{k})]}{\sum_{\mathbf{k}} |E_{\mathbf{k}} E_{\mathbf{h}-\mathbf{k}}| \cos[\phi(\mathbf{k}) + \phi(\mathbf{h}-\mathbf{k})]} \quad (5.7)$$

The last formula needed for the discussion of the programmes and methods used in the study of AhrC, comes from the theory of ‘nested neighborhoods’. Hauptman, H., 1975 showed that the expected value of the quartet :

$$\Phi = \phi(\mathbf{h}) + \phi(\mathbf{k}) + \phi(\mathbf{l}) + \phi(\mathbf{m}) \quad (5.8)$$

where

$$\mathbf{h} + \mathbf{k} + \mathbf{l} + \mathbf{m} = \mathbf{0}$$

is $\Phi = 0$ if all $|E_{\mathbf{h}}|, |E_{\mathbf{k}}|, |E_{\mathbf{l}}|, |E_{\mathbf{m}}|, |E_{\mathbf{h}+\mathbf{k}}|, |E_{\mathbf{k}+\mathbf{l}}|, |E_{\mathbf{l}+\mathbf{h}}|$ are large, but it is $\Phi = \pi$ if the ‘cross’ terms $|E_{\mathbf{h}+\mathbf{k}}|, |E_{\mathbf{k}+\mathbf{l}}|, |E_{\mathbf{l}+\mathbf{h}}|$ are all small. When some of the cross terms are small and other are large the probability distribution is bimodal. The indications $\Phi = \pi$ (known as negative quartets) are very useful for plane or space groups that lack translational symmetry (also known as symmorphic groups), since they drive the development of phases away from the trivial solution $\phi(\mathbf{h}) = 0$ for all \mathbf{h} .

5.2.1 SHELXS-86.

An analysis of all data sets collected from possibly useful heavy atom derivatives of AhrC crystals was attempted using the programme SHELXS-86 (Sheldrick, G.M., 1985, Sheldrick, G.M., 1986, Robinson, W. & Sheldrick, G.M., 1988). The programme is based on three relationships : the triple-phase relationship (which

reduces to the triple-product sign relationship for centrosymmetric terms), the negative quartets and a weighted tangent formula. SHELXS starts by forming phase permutations of a subset of reflections chosen by the user and then refines each permutation using the tangent formula. The best 10% of those permutations (as judged by two figures of merit, see below) is then refined with all data above a user defined threshold $|E_{min}|$. The two figures of merit used in SHELXS are related to (i) how well the refined phases agree with their estimated values from the known probability distributions, and (ii) how well the phase sets agree with the indications from the negative quartets. The phase set with the best combined figure of merit is then improved by a partial structure expansion procedure.

During our attempts to obtain a solution (mainly for the Niobium and Iridium derivatives), the following parameters have been varied: the $|E_{min}|$ threshold, the expected number of atoms (a large number reduces the probabilities of the phase relationships, Equation 5.5), the atom types, the inclusion or not of negative quartets, the number of phase permutations to be examined, the composition of the subset, the low and high resolution cutoff, the omission of the largest E -values, and, finally, the use of F_{HLE} instead of ΔF data.

The results from SHELXS-86 were evaluated as follows. If the best solution was a ‘uranium atom solution’ (that is, the best E -map contained only one or two very large peaks, usually at special positions or at the assumed molecular centre, see Chapter 4), no further consideration was given to it. Most of the solutions examined belonged to this category. If the best E -map contained several approximately equal peaks, the Patterson function corresponding to these sites was calculated and compared with the observed Patterson. In the few cases where the agreement was acceptable, the trial heavy atom structure was refined and SIRAS phases were calculated for the native crystals as discussed in Section 3.3.1. These phases were used to calculate a cross difference Fourier, $m(\Delta F_{lr})e^{i\phi_{SIRAS}}$, for a different derivative. The atomic sites corresponding to the top six peaks in this

map were then used to calculate a Patterson function and the agreement between the observed and calculated Patterson functions was assessed. None of the solutions examined required any further calculations to confirm their incorrectness (or otherwise).

As already discussed, the centrosymmetric ΔF terms are probably the most accurate isomorphous difference data available when the heavy atom structure is unknown. We have collected the $hk0$ and $h0l$ terms both from native crystals and crystals soaked in $\text{Nb}_6\text{Cl}_{14}$ and H_2IrCl_6 to as high resolution as possible and as accurately as practically feasible (Tables 3.1 & 3.4). Attempts to phase these reflections using SHELXS have also been unsuccessful.

5.2.2 MULTAN77.

The approach used in MULTAN77 (Main, P., 1985, Woolfson, M.M., 1987) is outlined below : A list of all triple-phase relationships involving large E -values is prepared. The CONVERGENCE procedure (Germain, G. & Woolfson, M.M., 1968, Germain, G., Main, P. & Woolfson, M.M., 1970) is then used to exclude reflections that are not well linked (through phase relationships) to many other reflections. The result from this procedure is a small number of strong reflections which, when assigned phases, will reliably propagate the phase information to a large number of other terms. A starting phase set is obtained by defining the origin and the enantiomorph and by permuting the phase values for a small number of reflections chosen from the CONVERGENCE procedure. The refinement and extension of the starting phase set is based on a weighted tangent formula.

The results from MULTAN were evaluated as described in the previous section. Again, the great majority of the E -maps examined contained essentially one peak either at the molecular centre or at a special position. It must be noted, however, that the various parameters have not been tested as extensively as for SHELXS-86.

5.2.3 The Cochran & Douglas Method.

The Cochran & Douglas method is one of the oldest computer-based direct method algorithms (Cochran & Douglas, 1953, Cochran & Douglas, 1954). It is essentially a direct maximisation of Q in Equation 5.2 : if there are n centrosymmetric structure factors related by m triple-product sign relationships (Equation 5.3), then for each unique combination of signs (2^n in total) the value of

$$\chi = \sum_{\mathbf{h}} \sum_{\mathbf{k}} U_{\mathbf{h}} U_{\mathbf{k}} U_{\mathbf{h}+\mathbf{k}} \quad (5.9)$$

where U is the unitary structure factor, is evaluated and its largest values stored (the summations are over all m triple-product sign relationships). The sign combination with the largest χ -value will be the one for which most of the strongest phase indications are satisfied. If the assumptions behind the probability distribution formulae are valid for the structure under examination, this will be the most probable phase set.

Given that programmes as powerful as SHELXS or MULTAN have not found a convincing solution for either of the two most promising derivatives, it would appear that this algorithm is bound to fail. Nevertheless, Wright, W.B., 1958a and 1958b, showed that the Cochran & Douglas method succeeded where several other methods (including symbolic addition) failed (it is fair to note, however, that the compound she examined (Glutathione) contained a ‘heavy’ atom, the presence of which increased substantially the probability that the phase relationships hold).

A programme was written which for a given number (≤ 32) of centrosymmetric terms (i) prepared a list of all unique triple-product sign relationships, and (ii) for every unique sign combination the sum in Equation 5.9 was evaluated (using normalised instead of unitary structure factors) and if its value was greater than a preset limit, the combination was stored. At the end of the procedure, all stored sign combinations were listed in order of decreasing χ -values.

The programme was used with various sets of centrosymmetric $hk0$ or $h0l$ ΔF terms for the Niobium derivative. The number of reflections in these sets was in the range 24 to 32 and the number of phase relationships was between 40 and 72. For the [001] projection, 30 terms are about 20% of all reflections to 7\AA resolution, and model calculations showed that this number of reflections is quite adequate for solving an artificial structure consisting of 6 atoms.

The results were evaluated as follows : Fourier syntheses were calculated for the top 100 phase sets and these were examined for the presence of a reasonable number (> 2) of approximately equal peaks. For the most promising of these syntheses, a trial heavy atom structure was constructed with atomic positions corresponding to the largest peaks in the map, and these were refined as described in Section 3.3.1. If the refinement statistics were of acceptable quality, phases were calculated for the native crystals and these were used to obtain a projection of the native crystals along [001] (or [010]). This projection was then examined for the presence of solvent channels at the positions expected from the model of the crystal packing (Chapters 4,6) and taking into account a possible origin shift. None of the maps examined showed any sign of solvent channels at the expected positions.

In conclusion, all our attempts to determine the heavy atom structure of AhrC derivatives using direct methods have been unsuccessful. From a theoretical standpoint this is exactly what we should expect : the ΔF data will be a rather bad estimate of F_H for half of the reflections, only low resolution data are used, the possible non-isomorphism is not taken into account, etc. However, these problems were present even in those cases where the procedure was found to work. In the case of AhrC there is an additional problem. Due to the presence of high non-crystallographic symmetry, the heavy atoms are not 'randomly and uniformly' distributed in the unit cell. Since the probabilistic approach is based on the

assumption that the atomic positions are random variables, this ‘non-randomness’ makes the validity of the probability distribution formulae questionable. Schenk, H., 1973, Krieger, W. & Schenk, H., 1973 showed —using model calculations— how systematic overlap in the Patterson function leads to considerable deviations from the theoretical models. Systematic overlap in the Patterson function is exactly what we observe in the case of our best derivative (Section 4.3.1).

5.3 Patterson Methods.

Several Patterson function based methods have been used in the attempt to determine the heavy atom structure of AhrC derivatives. Before discussing the individual methods, it is useful to illustrate the size and peculiarities of the problem that these methods had to tackle.

Figure 5.1 shows the three Harker sections from the sharpened isomorphous difference Patterson function for the Niobium derivative. The meaning of the three large peaks (marked A,B & C) has already been discussed (Section 3.3.1 and Section 4.3.1). It is obvious that the Harker sections contain several significant peaks apart from the pseudo-origin peaks. It was thought that it would be useful to know how many possible heavy atom sites are consistent with those Harker sections and what their distribution would be. A programme was written which for every position in the crystallographic asymmetric unit, calculated the Harker vectors that would be observed if an atom was present at that position. If the value of the Patterson function was greater than zero for all three Harker vectors, then the sum of those values was stored at the corresponding position of a map which was a copy of the asymmetric unit. Figure 5.2 is the result of such a search using the sharpened Patterson function for the Niobium derivative (Section $y=2/32$ contains no significant peaks and has been omitted). This map is very similar to what Buerger, M.J., 1959 called an ‘implication diagram’. This

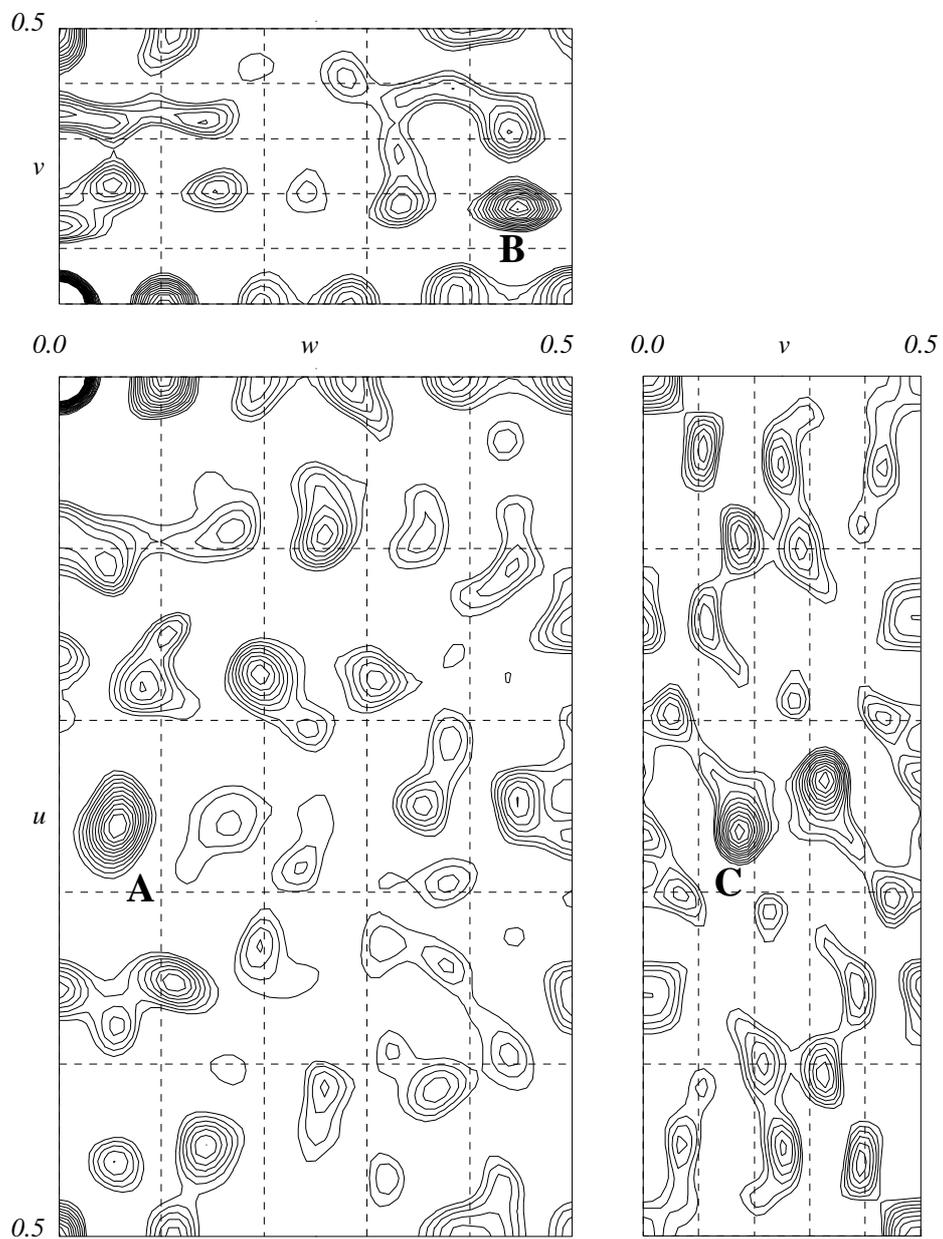


Figure 5.1: Harker sections $u=0.0$, $v=0.0$ and $w=0.5$ from a $13\text{-}7\text{\AA}$ $[E_{\Delta F_{isom}} \Delta F_{isom}]^2$ synthesis for the $\text{Nb}_6\text{Cl}_{14}$ derivative. Contours every 2% of the origin peak.

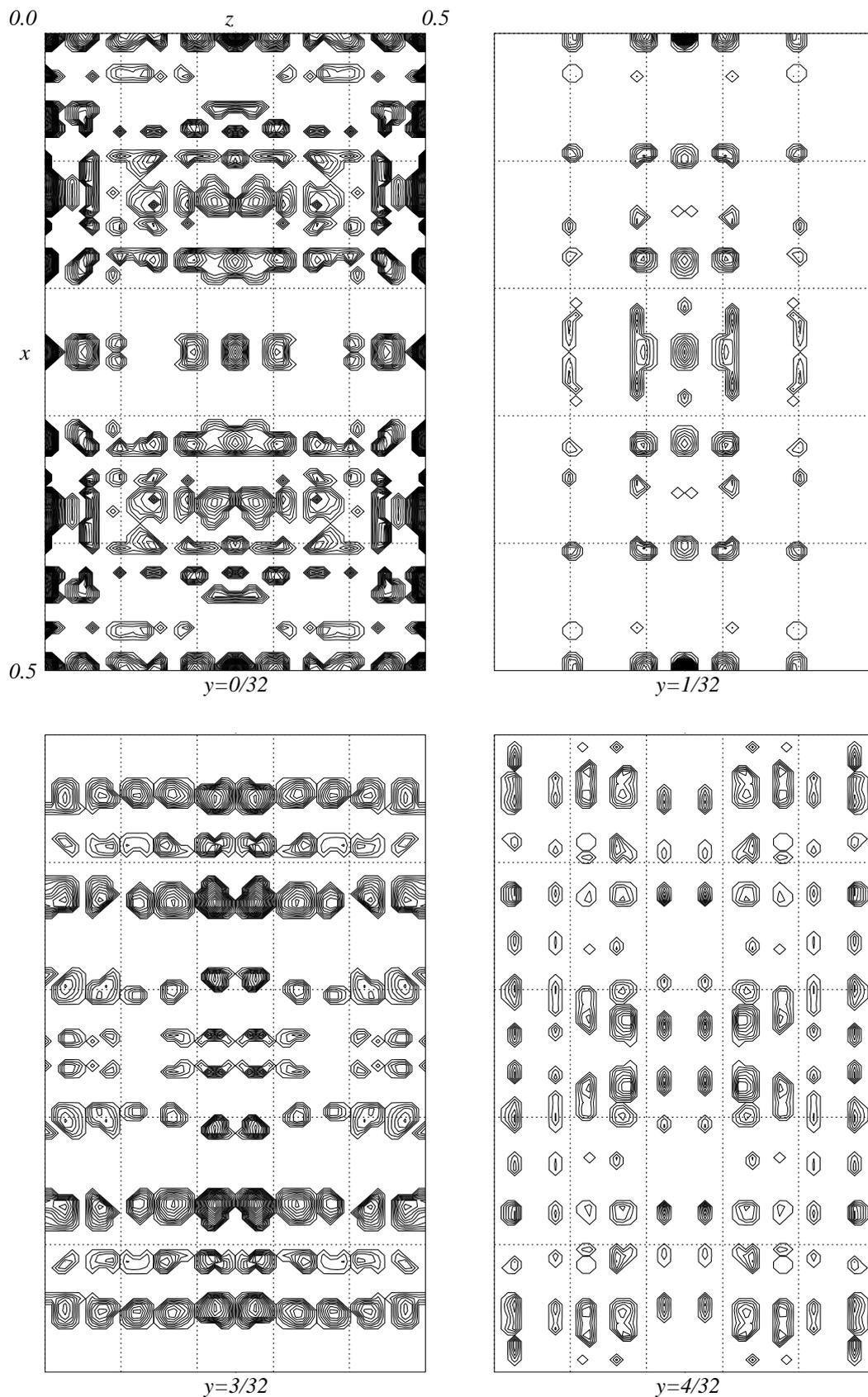


Figure 5.2: Implication diagram of the $\text{Nb}_6\text{Cl}_{14}$ derivative. Contours every 4% of the origin peak of the Patterson function.

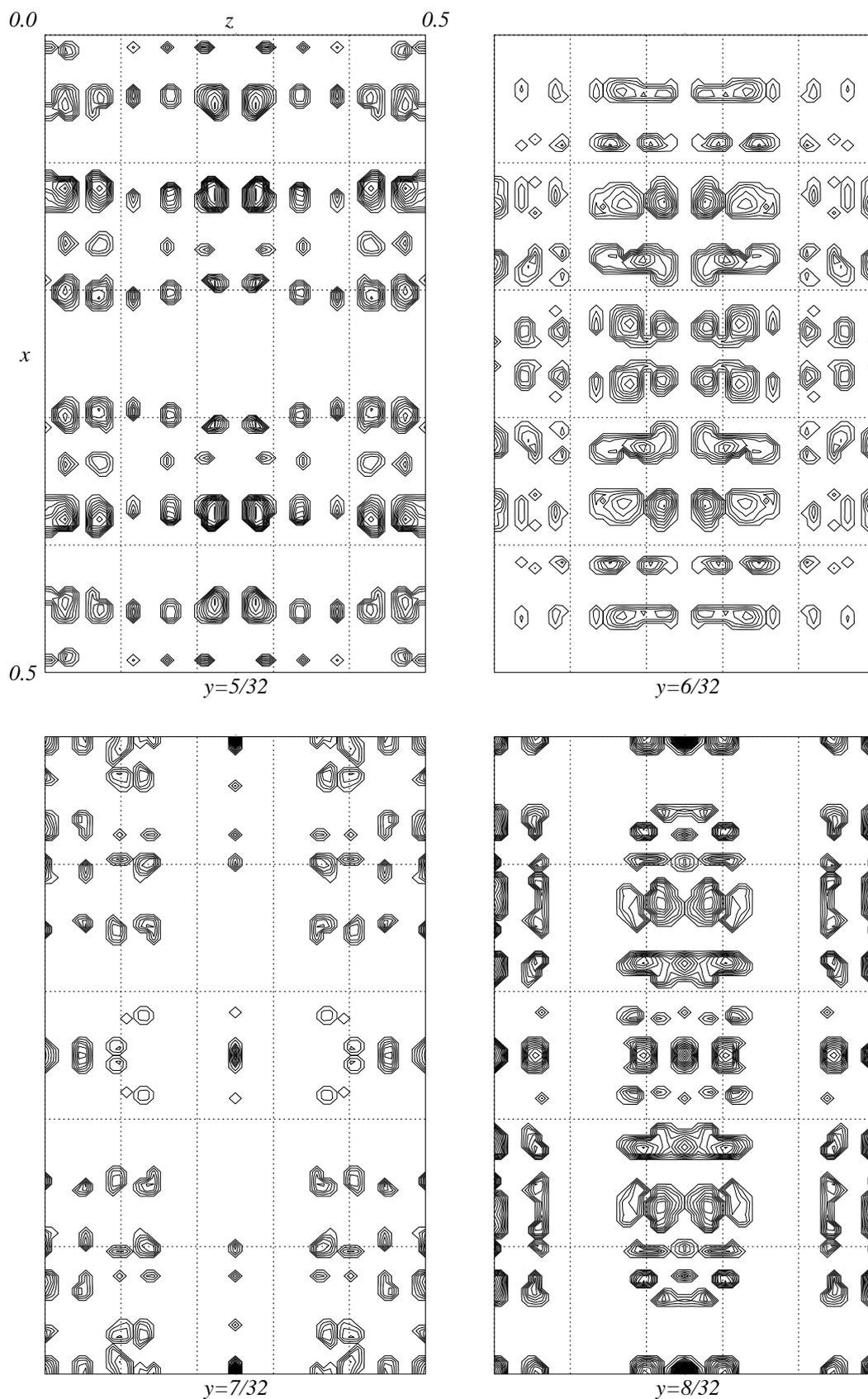


Figure 5.2: Implication diagram of the $\text{Nb}_6\text{Cl}_{14}$ derivative. Contours every 4% of the origin peak of the Patterson function.

search is also similar to the 'single atom search' as used in several Patterson solution programmes (such as HASSP or GROPAT), although in this case it is not required that significantly large Harker peaks are present on all three Harker sections.

Several features of this map are worth discussing. The first is that the implication diagram has extra symmetry : its asymmetric unit is 1/8th of the crystallographic asymmetric unit. This is because neither the enantiomorph nor the origin are fixed in this type of search. The second point is the large number of possible heavy atom positions : well over 80 peaks can be found in the asymmetric unit of the diagram. Most of these peaks arise from the presence of cross vectors on the Harker sections of the Patterson function. In the case of the Niobium derivative, the presence of cross vectors on the Harker sections is not accidental. Because non-crystallographic symmetry axes are parallel to all three crystallographic 2-folds, the non-crystallographic Harker sections coincide with the crystallographic. In other words, the Harker sections contain vectors between atoms related by both crystallographic and non-crystallographic symmetry. A third point is the presence of very large peaks on section $y=0.0$. These arise because no correction has been made for the different multiplicities of the various Patterson peaks. The last point is the very regular distribution of peaks in the implication diagram : the majority of significant peaks are arranged on lines parallel to z . These lines are again arranged in a regular manner : they form layers perpendicular to the y axis. It is well known since the early days of protein crystallography (see for example Dickerson, R.E, *et al*, & Weinzierl, J.E., 1967) that a small number of bad measurements of relatively low resolution reflections can lead to isomorphous difference Patterson functions with a very regular arrangement of peaks which are nothing more than ripples arising from those bad measurements. There is evidence which suggests that the regular arrangement of peaks in the difference Patterson map for Niobium arise from the heavy atom

structure and not from errors of measurement : (i) the Patterson function remains virtually unchanged when different native or derivative data sets are used for the calculation, (ii) the presence of some of the largest ΔF terms has been confirmed from comparison of precession photographs, (iii) the correlation coefficient between Patterson functions calculated using different resolution ranges is acceptably high (Section 3.3.1), and finally, (iv) the implication diagram corresponding to the [010] Patterson projection (Figure 5.3) shows the same pattern of peaks although only the $h0l$ terms have been used for this calculation and the presence of the largest ΔF terms has been confirmed from examination of precession photographs of that zone.

It is interesting to note that several authors have suggested that all ΔF terms greater than four (or even three) times the mean value of the observed isomorphous differences should be excluded from all subsequent calculations. Their argument is of a statistical nature and is based on calculating the probability of $\Delta F > n\overline{\Delta F}$. In the absence of any information about the heavy atom structure, the best that can be done is to assume that the heavy atoms are distributed randomly and uniformly in the unit cell. In the presence of non-crystallographic symmetry, the distribution of heavy atoms will be non-random and since most of these reagents are ionic (and, thus, bind on the surface of the macromolecule), their distribution will also be non-uniform. This is not to imply that bad measurements which result in very large differences should not be monitored and excluded. The solution adopted in the case of AhrC, was to collect and compare at least two data sets from each possibly useful derivative.

One final point that has to be made, concerns the evaluation of the results. All methods described in this section are bound to produce trial heavy atom structures whose Patterson function will be related to the observed Patterson. This is more so for the Niobium derivative because the presence of regularities in the heavy atom structure makes it very easy to obtain solutions which, although

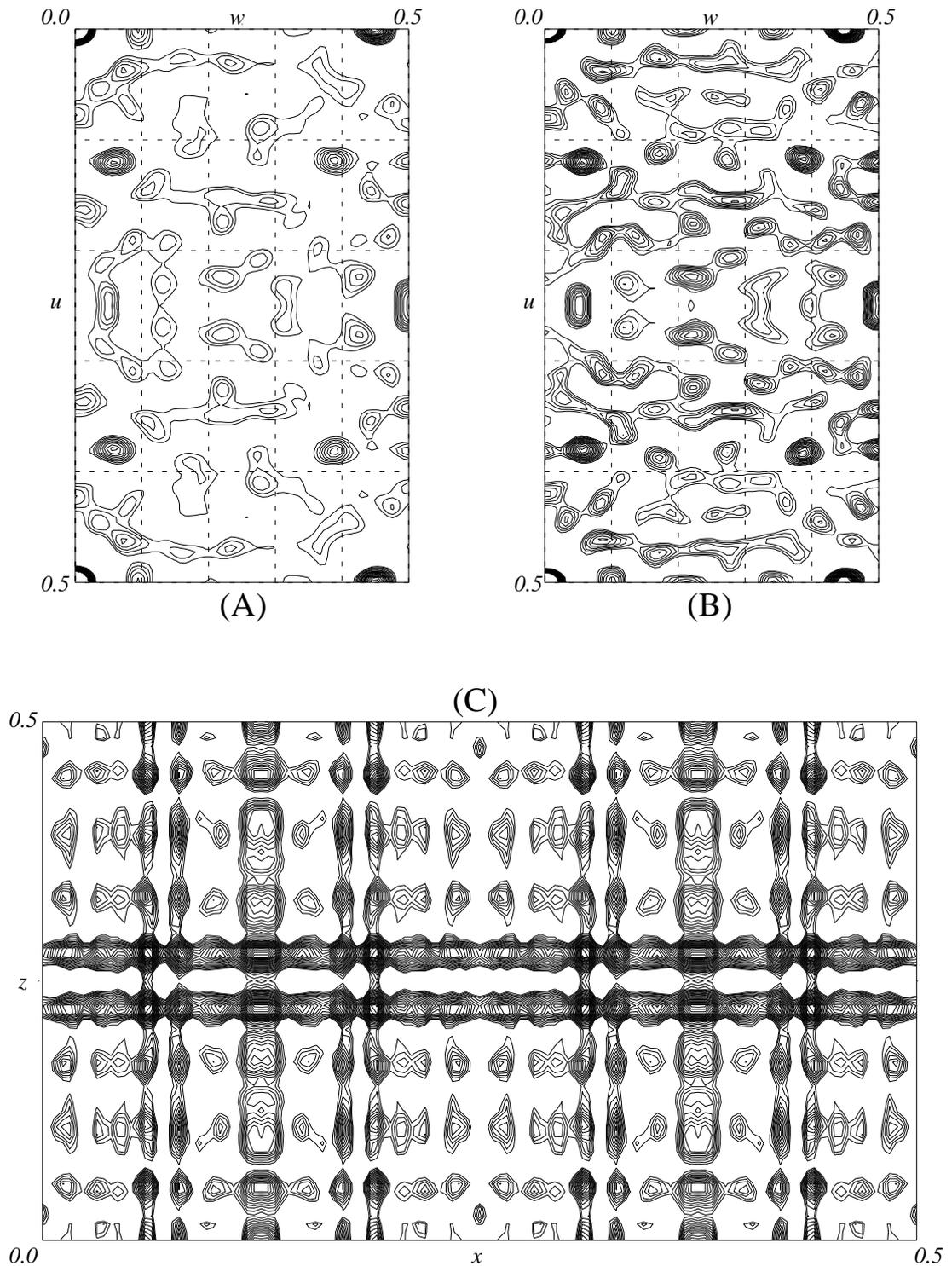


Figure 5.3: (A) [010] Patterson projection for $\text{Nb}_6\text{Cl}_{14}$: 17.5\AA ΔF^2 , (B) Sharpened synthesis, (C) Implication diagram using the sharpened function.

wrong, can reproduce several of the strong peaks in the observed difference Patterson. In these cases where the agreement with the observed Patterson function was not sufficiently bad for the solution to be discarded immediately, refinement of the trial structure followed by calculation of cross difference Fourier syntheses for other derivatives proved to be a very powerful test of the correctness of the proposed heavy atom structure (the paper from Dickerson, R.E, *et al*, & Weinzierl, J.E., 1967 is again illuminating).

5.3.1 HASSP.

Terwilliger, T.C., Kim, S.-H. & Eisenberg, D., 1987 proposed a method of determining heavy atom positions which is based on using isolated Patterson peaks as an aid to identify possible two-site solutions which are then extended through a systematic search of the asymmetric unit for additional sites consistent with the starting pair. They have implemented their ideas in the programme HASSP, a short description of which is given below.

The programme starts with a systematic search for single-site solutions using only the Harker sections. Probability measures are applied in order to identify potentially significant solutions, but these possible single-site solutions are not used in the subsequent calculations. The next step is to identify isolated peaks in the Patterson function and to use each one of these in turn as a possible ‘single-weight’ cross vector in a systematic search of the crystallographic asymmetric unit : if \mathbf{u} is the vector from the origin of the Patterson function to the isolated cross peak, then for every crystallographically unique atomic position \mathbf{x} , the Harker vectors corresponding to a pair of atoms at \mathbf{x} and $\mathbf{x} + \mathbf{u}$ are calculated. If the density of the Patterson function at these Harker vectors satisfies certain probability criteria, then these atoms form a possible two-site solution. From all possible two-site solutions for any given cross peak, the one which gives the most significant Harker vectors is extended by searching the asymmetric unit for an

additional atom whose Harker vectors and cross vectors with the starting pair are significantly above the noise level. The procedure is repeated until no significant additional sites can be found, or the limit of 6 possible atomic sites per solution is reached.

Most of the data sets which showed signs of specific heavy atom binding have been examined using HASSP. Attempts were made with both unsharpened ΔF^2 and sharpened $[E_{\Delta F_{isom}} \Delta F_{isom}]^2$ syntheses using various resolution ranges, typically 13-6Å and 10-5Å. The single-site search was not very useful : In the case of the Niobium derivative, only four significant single-site solutions could be identified, two of which were on special positions while a third was at the molecular centre. The reason for the very small number of possible single-atom solutions is the requirement that all Harker peaks should be significantly above the noise level. In the two-site search, all isolated peaks that the programme could identify (usually about 40 peaks) were used. The solutions produced from this search, were examined as described in the previous Section. Although several of the most promising solutions for the Niobium derivative gave good statistics during centric refinement (typical values were $wR_c = 0.55$, $C = 0.48$, Gradient 0.45, see Section 3.3.1 for definitions), no further progress could be made with any of them.

A possible explanation for these problems can be traced to the selection of the “best” pair of atoms for any given cross vector. As the number of sites increases, the probability that an incorrect solution can give a better fit to the Harker sections increases. Furthermore, the Harker sections, especially for Niobium, are dominated by cross vectors and can not be trusted in the early stages of the development of a solution. A more cautious approach would be to compile a list of all pairs which give positive Harker vectors and then to expand each one of these in turn. The expansion step would also have to keep track of more than just the “best” solution. It is obvious that this is a hierarchical tree structure

in which the number of possibilities grows too fast for the calculation to be practical (If at every step the best n solutions are kept and the decision as to which combination is best is postponed until each combination consists of m atoms, the total number of combinations for only one cross vector is n^m . For 40 cross vectors and $n = m = 8$, it would be necessary to search the crystallographic asymmetric unit 83,886,080 times and at the end of the procedure 671,088,640 combinations of 8 sites would have to be compared).

5.3.2 GROPAT.

Jones, Y. & Stuart, D., 1991 proposed a ‘brute force’ approach to the solution of complex difference Patterson maps. Their algorithm is summarised below.

For every position in the crystallographic asymmetric unit, the values of the Patterson function at the end of the predicted Harker vectors are used to calculate an overall ‘probability’ that the position under examination corresponds to a heavy atom site. Six criteria are used in calculating this probability : the sum, product and minimum of the Patterson densities, their root mean square, the mean probability for observing those values and, finally, the product of these probabilities. The probabilities are calculated from the expression $\exp[-(P_{ideal} - P_{obs})^2/\sigma^2]$, where, P_{obs} is the observed Patterson density at a given grid point, P_{ideal} is a user input parameter which should be equal to the Patterson density expected from a single Harker vector, and σ is the expected error level in the Patterson map (this is again a user-defined parameter but the authors suggest that the root mean square deviation of the Patterson map can be used as an estimate of the expected error). All pairwise combinations of the best 200 positions from this search are then examined and the best 200 pairs (as judged from the criteria outlined above but now considering cross vectors) are written out. This list of best pairwise combinations forms the basis of the subsequent analysis : if all heavy atom sites are present in the initial list of putative single-site solutions,

then, the list of pairwise combinations should contain all possible pairs of those sites. The problem is reduced to that of finding all possible sets of atomic sites whose pairwise combinations are present in the list. These are possible heavy atom structures.

Searching through a list of 200 pairs of atomic codes for the presence of all pairwise combinations of an initially unknown number of sites invites mistakes. A programme was written therefore, which searched this list for all possible 3-, 4-, 5- and 6-atom solutions. Even if some pairs of, say, a 6-atom solution are missing from the initial list, the solution will not be missed since most of the 5-atom combinations will be present.

GROPAT has been tested with both ΔF^2 and $[E_{\Delta F}\Delta F]^2$ syntheses for the Niobium and Iridium derivatives. Again, as with HASSP, no convincing solution could be identified. Considerations similar to those described in the previous section apply here also : As the complexity of the problem increases, the probability that incorrect pairs of atoms (from the single-site search) will give cross vectors on regions of the Patterson function with high density, is also increasing. Furthermore, the presence of cross vectors on the Harker sections for Niobium, makes the probability-based single-site search less reliable.

5.3.3 A Method by Nixon, P.E., 1978

The effect of non-randomness of the atomic positions on the probability distribution of the basic phase-determining formulae has already been discussed (Section 5.2.3). We have also seen that this non-randomness is best expressed in terms of overlap of Patterson peaks. Nixon, P.E., 1978 suggested a method for lessening the problems associated with overlapping Patterson peaks. Although his method is based on using direct methods to solve the phase problem, it will be discussed in this section since it involves a modification of the Patterson function. In this approach, the observed Patterson function $P(\mathbf{u})$ is modified to give

a function $P'(\mathbf{u})$ as follows :

$$\text{if } P(\mathbf{u}) \leq 0 \text{ then } P'(\mathbf{u}) = P(\mathbf{u}) \text{ otherwise } P'(\mathbf{u}) = c \tanh[P(\mathbf{u})/c] \quad (5.10)$$

where the value of c is adjusted so that only the strongest peaks are affected. The effect of this modification, as shown in Figure 5.4, is to reduce the density of the most prominent peaks of the Patterson function. The modified Patterson function is then back-transformed to yield new estimates for the intensities which can then be used as observed data in a direct methods programme.

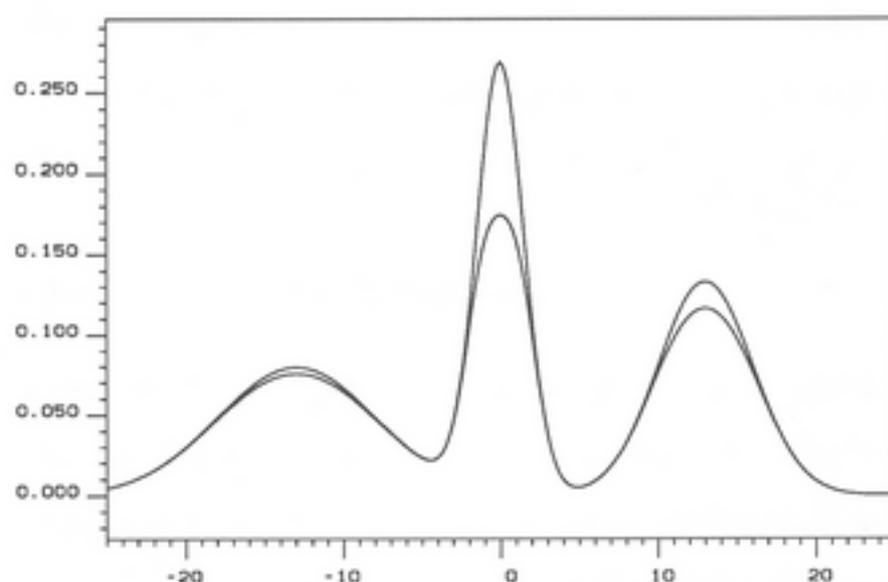


Figure 5.4: Functions $y(x)$ (top curve) and $c \tanh(y(x)/c)$ (bottom curve). $y(x)$ consists of three Gaussians centered at $x=0.0$, 13.0 and -13.0 with $\sigma=1.5$, 3.0 and 5.0 correspondingly. In this example $c = 0.2$.

The basis of the method can not be easily justified when thinking in terms of Patterson vectors : the large peaks arise from the presence of regularities in the structure, and the correct atomic arrangement will reproduce these peaks at their observed height. The method is best understood in terms of the distribution of the largest E -values : when the atomic positions are not independent random variables, the large E -values will not be distributed uniformly over all classes

of reflections. In some cases, the classes of reflections that are systematically weak or strong can be identified and the E -values (within each class) can be re-normalised (although more complex algorithms do exist (Hai-Fu, F., Jia-Xing, Y., Main, P. & Woolfson, M.M., 1983)). Nixon's method is an indirect way of re-normalising the E -values for those classes of reflections that are systematically strong without having to identify them explicitly.

A programme was written to modify a given Patterson function according to Equation 5.10. The modified function was back-transformed using the programme SFC from the CCP4 suite. The resulting intensities were input to the direct methods programme SHELXS86 (Section 5.2.1). Most of the solutions examined belonged to the 'uranium atom' category. No convincing solutions could be identified.

5.3.4 Image-seeking Methods.

The Patterson function of a structure consisting of N atoms can be viewed as the superposition of N images of the structure, where the i th image ($i = 1, \dots, N$), is obtained by placing the i th atom at the origin of the function, as shown in Figure 5.5. Image-seeking methods attempt to retrieve any one of these images. Several different names (Atomic superposition, Vector superposition, Image-seeking) have been proposed for methods that are in principle equivalent. Although these procedures represent the most logical way to proceed in the analysis of a Patterson function, the power and automation of direct methods, made them largely redundant. The theory and practice of image-seeking methods is very well documented and will not be discussed in great detail (Buerger, M.J., 1951, Buerger, M.J., 1959, Buerger, M.J., 1964, Ramachandran, G.N. & Srinivasan, 1970, Fridrichsons, J. & Mathieson, A.McL., 1962, Shoemaker, D.P., Barieau, R.E., Donohue, J. & Lu, C.-S., 1953, Donohue, J. & Bryden, J.H., 1955, Simpson, P.G., Dobrott, R.D. & Lipscomb, W.N., 1965, Germain, G. & Woolfson, M.M., 1966, and more

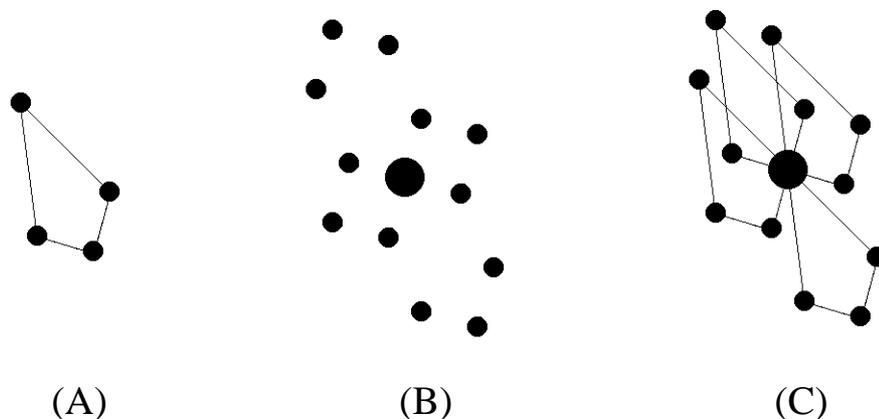


Figure 5.5: (A) A hypothetical four-atom structure, (B) Its Patterson function, (C) The Patterson function seen as the superposition of four images of the structure (Note the presence of four enantiomorphic images).

recently, Richardson, J.W. & Jacobson, R.A., 1987, Sheldrick, G.M., 1991).

The principle of the vector-superposition method is illustrated in Figures 5.6 and 5.7 for the centrosymmetric and non-centrosymmetric cases respectively. Figure 5.6(A) shows a hypothetical two-dimensional centrosymmetric structure consisting of 5 atoms, and Figure 5.6(B) is a schematic diagram of its Patterson function. In Figure 5.6(C) two copies of the Patterson function (one represented with open circles, the second with filled) have been superimposed so that the origin of the second function coincides with the peak marked “I” in (B). It can be seen that there are only five places where peaks from the two functions coincide, and these are shown in Figure 5.6(E). This is the original structure. Peak “I” is a single-weight Patterson peak which corresponds to the vector between the two atoms that are furthest apart in the structure. If the multiple peak marked as “II” in (B) is used for the superposition (shown in Figure 5.6(D)) then Patterson peaks coincide at 8 places. Diagram (F) shows that these can be interpreted as two images of the structure related by a simple translation. It is an inter-

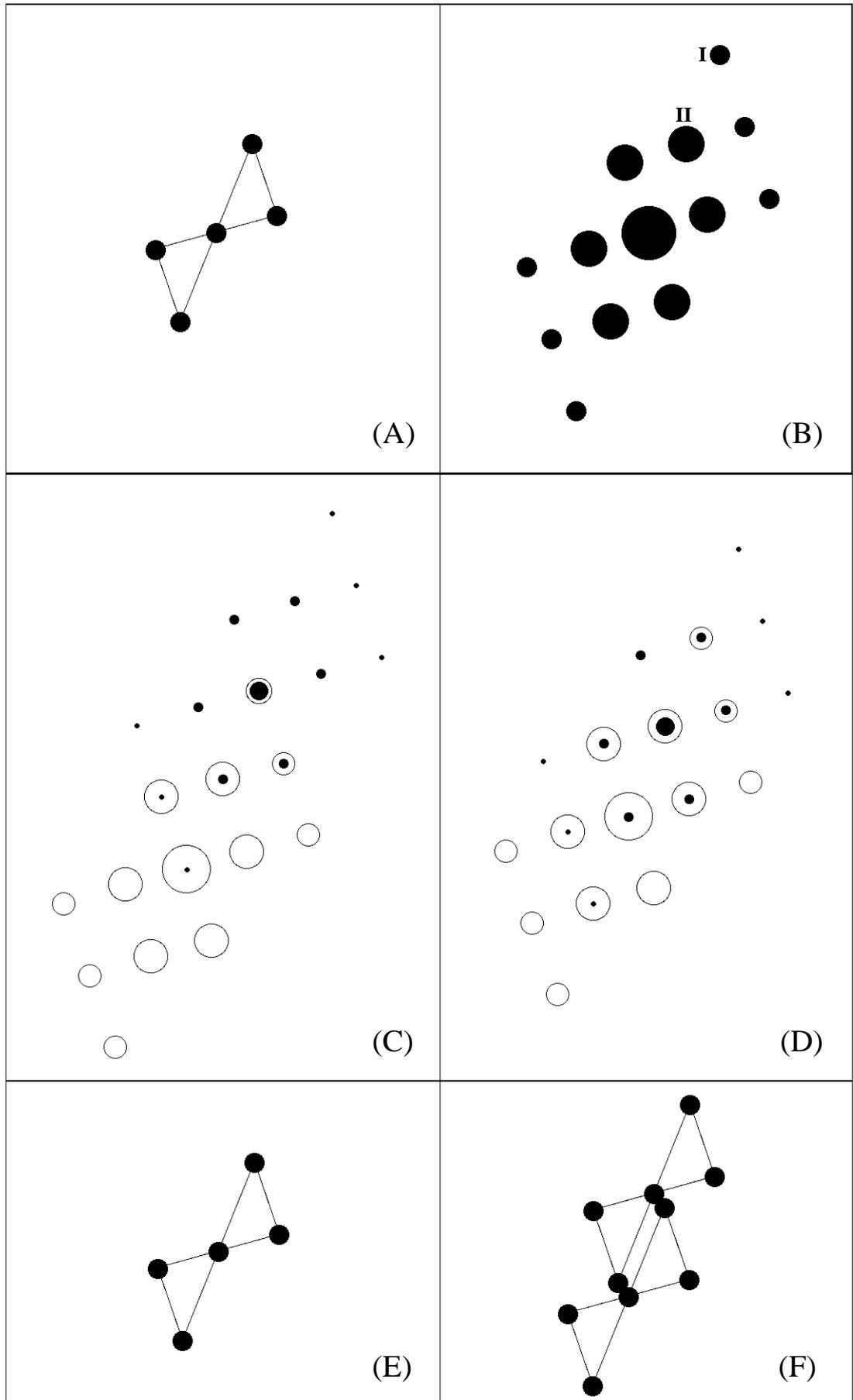


Figure 5.6: Vector superposition : The centrosymmetric case.

esting result that in the case of a centrosymmetric structure (or projection) a superposition on a single-weight Patterson peak can reveal the structure.

Figure 5.7(A) shows a hypothetical non-centrosymmetric 4-atom structure and 5.7(B) is its Patterson function. The superposition on a multiple peak (marked as “II” in (B)) is shown in Figure 5.7(C). In (E) these coincidences are interpreted as arising from two images of the structure (related by a simple translation) plus their enantiomorphs. In (D), a single-weight vector was used for the superposition. The result (shown in (F)) is the superposition of an image of the structure and its enantiomorph. The ambiguity arising from the presence of both enantiomorphs can be resolved either by a second superposition on a different peak or by “hand exorcism” of one of the enantiomorphic images.

It is worth noting that, in general, the space (plane) group of the superposition function is P1 (p1). Identification of a permissible origin (for the given space or plane group) can be achieved either by visual examination of the superposition map or by a systematic search for the position (in the superposition map) which if chosen as a crystallographic origin will maximise the agreement between the observed peaks.

If one or more atomic positions are known (from examination of Harker sections or lines, for example), then these can be used in a variation of the superposition method known as ‘Atomic Superposition’. If for a structure that belongs to a plane or space group with m equivalent positions, the coordinates (with respect to a permissible origin) of n atoms are known, then mn copies of the Patterson function are superimposed with their origin peaks at all equivalent positions of the n atoms, and the resulting superposition function is examined for the presence of additional (and initially unknown) atomic sites (Figure 5.8). Since the space or plane group symmetry is used explicitly, the atomic superposition function has the space (plane) group of the structure.

Our first attempts with the vector superposition method were based on su-

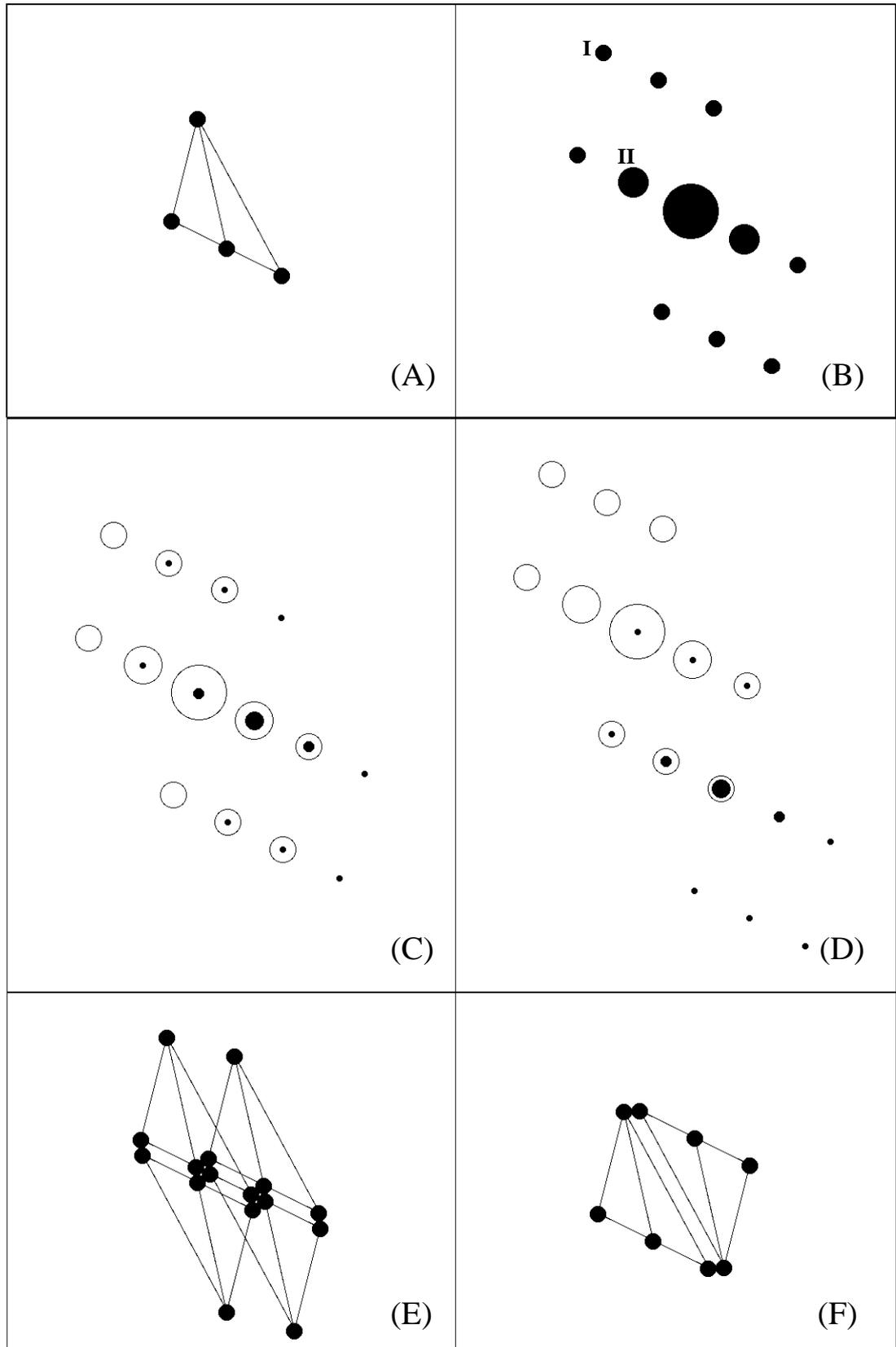


Figure 5.7: Vector superposition : The non-centrosymmetric case.

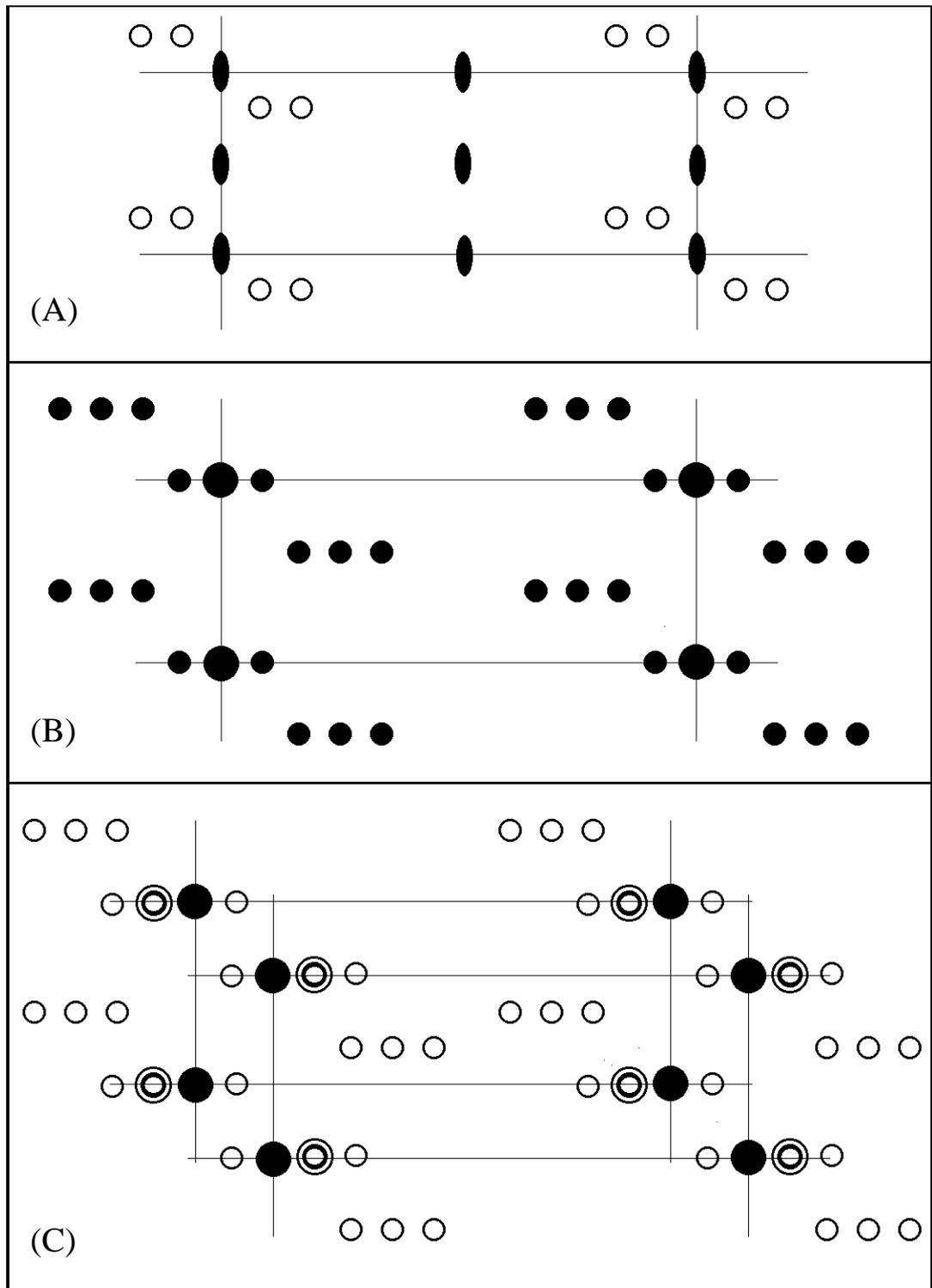


Figure 5.8: Atomic superposition : (A) A hypothetical two-atom structure in p2, (B) Its Patterson function, (C) Superposition of two copies of the Patterson function at the two equivalent positions of one atom (black circles). Double circles denote additional coincidences.

perimposing manually contoured Patterson projection functions. It was soon realised that an automatic procedure was needed if the superpositions were to be carried out accurately a large number of times : Two programmes have been developed. The first deals with the two-dimensional (centrosymmetric) case. Given a Patterson projection \mathcal{P} and a number n of grid points, the programme will superimpose (using the sum, product or minimum function) n copies of \mathcal{P} with their origin peaks at the specified grid points. The superposition function is written as a CCP4 map file which can be plotted using the programme PLUTO (from the CCP4 suite of programmes). This superposition programme can be used for both vector and atomic superposition.

The second programme deals with the general case of a three-dimensional Patterson function and implements some of the ideas presented by Richardson, J.W. & Jacobson, R.A., 1987 and Sheldrick, G.M., 1991. Two copies of the Patterson function are superimposed on a given Patterson peak using the minimum function. The position (in the superposition function) which if chosen as an origin of the known space group will maximise the agreement between grid points related by crystallographic symmetry is then identified through a systematic search : the criterion used is the sum of the minimum density of all crystallographically related grid points, for every given choice of origin. Once the “best” origin has been identified, the superposition function is averaged using the crystallographic symmetry and the final (averaged) map is written out in the CCP4 map format.

Most of our attempts with the superposition method concentrated on the [001] and [010] Patterson projections of the Niobium and Iridium derivatives. Although the [001] projection is down a 138Å long axis, it was hoped that useful results could still be obtained : the implication diagram (Figure 5.2) suggests that the heavy atom sites of the Niobium derivative are arranged on lines parallel to z and would, thus, overlap in the [001] projection. The [001] Patterson projection (Figure 5.9) clearly indicates the presence of a regular arrangement of heavy

atom sites. The ΔF -weighted reciprocal lattice representation of the $hk0$ level (Figure 5.10) shows a hexagonal arrangement of clusters of large ΔF terms with two outstandingly large differences for the reflections 11,5,0 and 12,6,0. After removal of these two terms, the projection remained virtually unchanged, suggesting that the large differences are due to the underlying heavy atom structure and not due to errors of measurement. The presence of regularities in the heavy atom structure made the identification of single-weight Patterson peaks very difficult : Superpositions have been carried out on all relatively weak Patterson peaks with density greater than zero. Some of the most promising superposition functions showed an arrangement of peaks which was in very good agreement with the model of the crystal packing (Figure 5.11). Unfortunately, none of these solutions led to a projection of the native crystals with the expected arrangement of solvent channels.

Several attempts using the [010] projection have also been unsuccessful. One of the most promising atomic superposition functions is shown in Figure 5.12. The presence of a non-crystallographic 2-fold axis at the position of the assumed molecular centre is indicated. No progress could be made with this or any other solutions examined.

Finally, a number of attempts using the second of the programmes described above with the three-dimensional isomorphous difference Patterson function for the Niobium derivative have been unsuccessful. This was mainly due to the difficulty in identifying single-weight Patterson peaks and the presence of both enantiomorphs in the superposition function.

5.3.5 A Patterson Interpretation Programme : MSS.

The discussion in sections 5.3.1 and 5.3.2 suggested that accepting (or rejecting) partial two-site solutions very early during the development of a possible heavy atom structure is not necessarily the best way forward. The most systematic

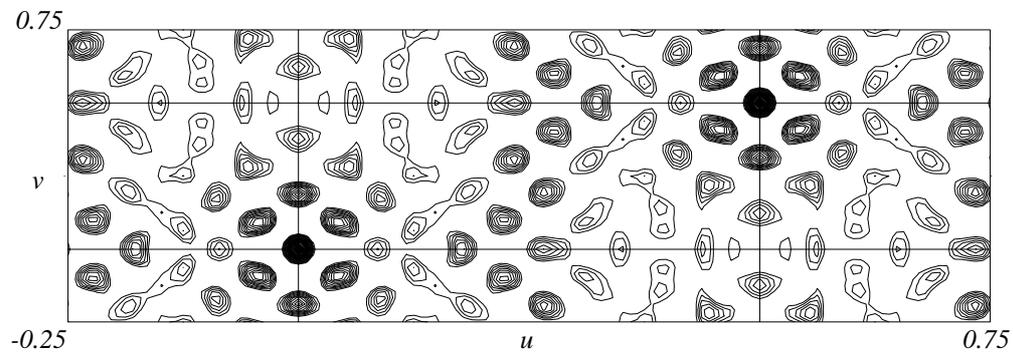


Figure 5.9: 15-6Å [001] Patterson projection for the Niobium derivative.

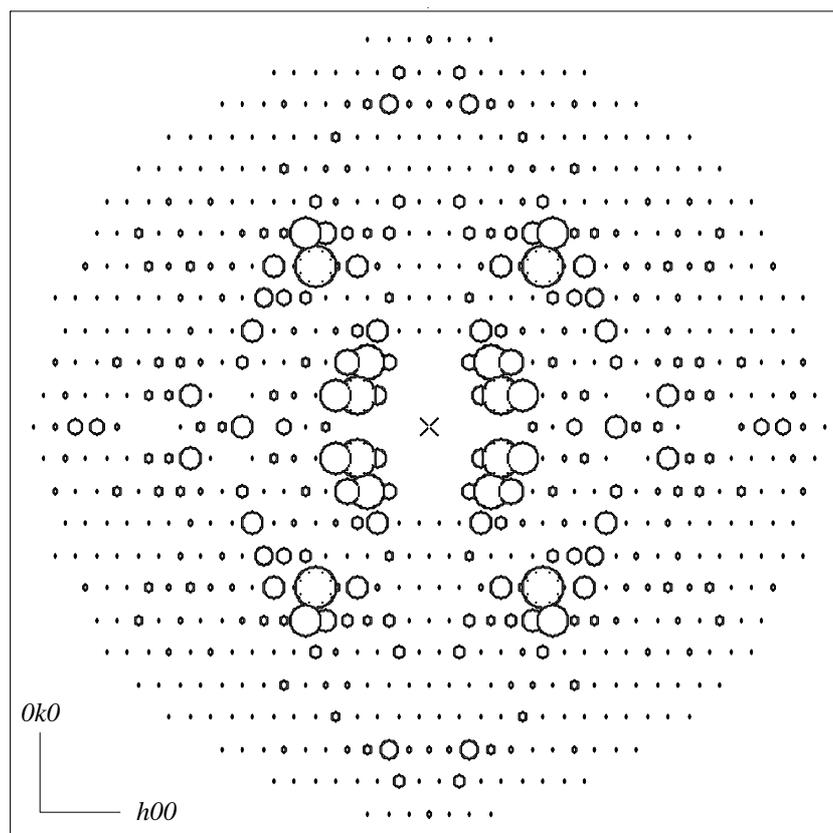


Figure 5.10: $hk0$ weighted reciprocal lattice for $\text{Nb}_6\text{Cl}_{14}$.

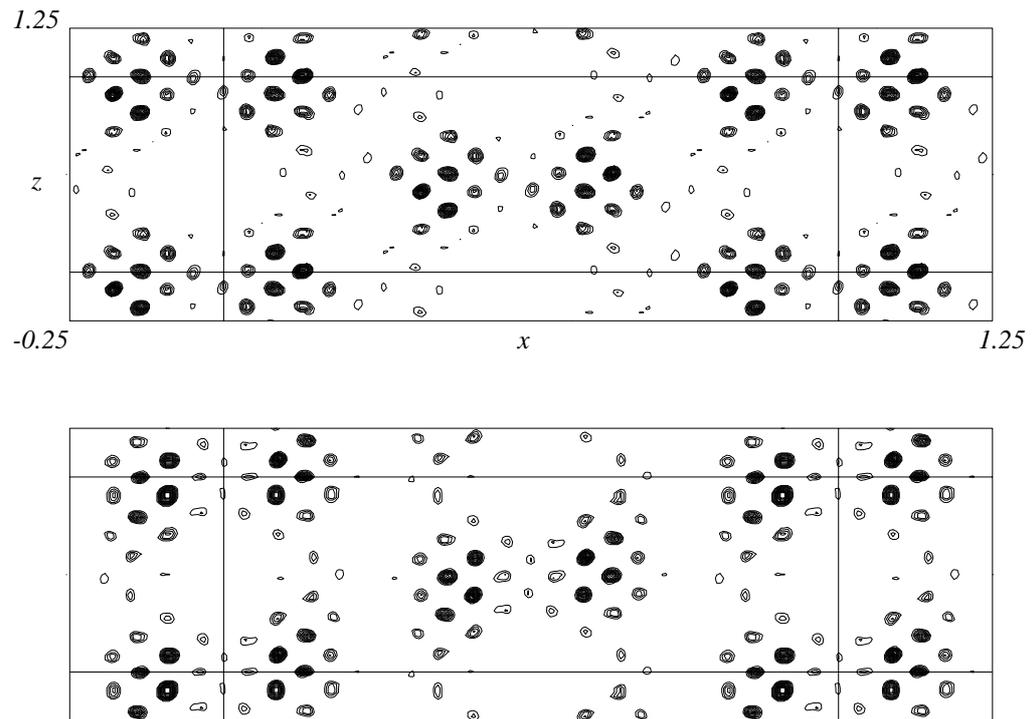


Figure 5.11: [001] Vector superposition functions for the Niobium derivative.

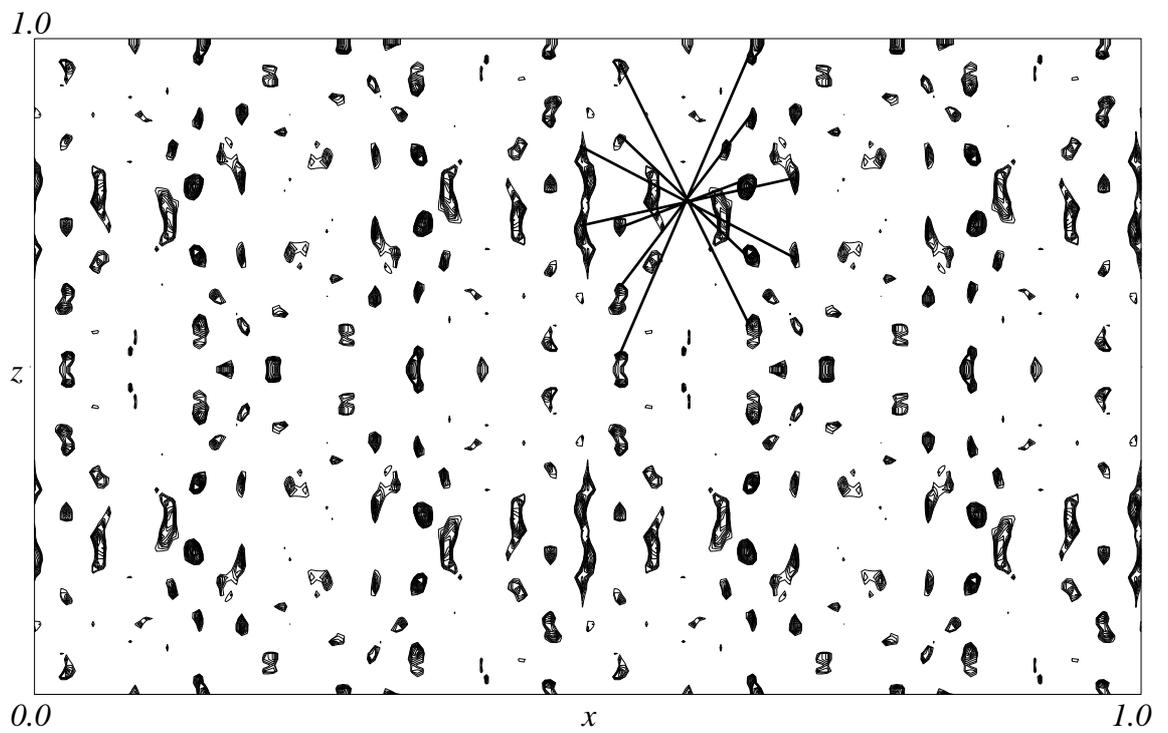


Figure 5.12: [010] Atomic superposition function for the Niobium derivative.

(but least elegant) approach for the interpretation of a Patterson function would be to evaluate the agreement between the observed Patterson function and the Patterson function calculated from every crystallographically unique arrangement of atoms. Such a procedure would be equivalent to a direct R-factor search, that is, calculation of the R-factor for all unique atomic arrangements. This approach is practical only for relatively small unit cells : in the case of the orthorhombic form of AhrC, a systematic search of the “best” six-atom structure (as judged by the R-factor or the correlation coefficient between the observed and calculated structure factor amplitudes), would require the calculation of structure factors $3.4 \cdot 10^{24}$ times (assuming that a 2\AA grid is used).

A Patterson function interpretation programme has been developed in which the number of possible heavy atom structures that have to be examined is reduced to a manageable size through the application of the following *a priori* assumptions about the Patterson function :

1. For every heavy atom position \mathbf{x} , the value of the implication diagram $\mathcal{I}(\mathbf{x})$ is greater than zero, that is, the Patterson function has positive values for all Harker vectors corresponding to \mathbf{x} .
2. If \mathbf{x} is a heavy atom position, then a pair of numbers n, m exists such that
 - (i) n out of the m atomic positions which give the largest sum of both Harker and cross vectors with \mathbf{x} , are also heavy atom positions, and, (ii) the sum of the values of the Patterson function for all vectors between the n heavy atom sites is the largest of all other combinations of n out of these m sites.

The pseudocode for this programme is shown in Figure 5.13.

Although the requirement of the positivity of the Patterson function for all Harker vectors of all heavy atom sites reduces significantly the volume of the crystallographic asymmetric unit that has to be searched for possible heavy atom

```

1. Calculation of the implication diagram  $\mathcal{I}$  :
  for (every grid point  $\mathbf{x}$  of  $\mathcal{I}$ )
    if (the density of the Patterson function ( $\mathcal{P}$ ) for all Harker vectors
        corresponding to a heavy atom site at  $\mathbf{x}$  is greater than zero)
      then  $\mathcal{I}(\mathbf{x}) = \sum \mathcal{P}(\text{Harker}(\mathbf{x}))$ 
      else  $\mathcal{I}(\mathbf{x}) = 0$ 

2. Development of  $n$ -site partial heavy atom structures :
  for (every  $\mathbf{x}$  in  $\mathcal{I}$ )
    if ( $\mathcal{I}(\mathbf{x}) > 0$ )
      then (find  $m$  other grid points which give the largest sum of both Harker
          and cross vectors with  $\mathbf{x}$ . Find the combination of  $n$  out of these  $m$ 
          sites which give the largest sum of all predicted vectors and save
          this  $n$ -site partial structure)
      else (next grid point)

3. Extension of all  $n$ -site partial structures :
  for (every  $n$ -site partial solution)
    (Find additional sites for which  $\mathcal{I}(\mathbf{x}) > 0$  and all predicted vectors with
    the already known sites lie in positive areas of the Patterson function.)

4. Refinement and calculation of useful metrics :
  for (every heavy atom structure)
    if (before refinement :  $\text{Corr}(E_o, E_c) > \text{limit}_1$ )
      then (Refine the positional parameters using Hart's algorithm. Target
          function is the correlation coefficient between  $E_o$  and  $E_c$  for
          centric terms)
        if (after refinement :  $\text{Corr}(E_o, E_c) > \text{limit}_2$ )
          then (calculate lengths of vectors between heavy atom sites
              and between heavy atoms and the molecular centre).

```

Figure 5.13: MSS : pseudocode.

sites ($\approx 10\%$ for the Niobium derivative), the algorithm is still practical only for relatively small values of n and m (typically $n \leq 6, m \leq 12$).

As an aid to the identification of possible correct solutions, the programme refines the positional parameters of the heavy atoms using Hart's algorithm (Hart, R.G., 1961) and calculates the lengths of the vectors between heavy atoms and between the molecular centre and the heavy atom sites (due to the presence of non-crystallographic symmetry, the correct solution was expected to show clusters of vectors of similar lengths).

The programme has been tested with hypothetical heavy atom structures consisting of up to 20 atoms in the crystallographic asymmetric unit (space group $C222_1$), some of which had non-crystallographic symmetry axes parallel to crystallographic. The correct solution could be identified for all structures consisting of less than approximately 14 atoms. For the more complex structures, only partially correct solutions were usually identified. The main reason for these problems was a violation of the assumption of the positivity of all Harker vectors : since the F_{000} term is not included in the calculation of the Patterson function, it was not unexpected to find that for some vectors the value of the Patterson function was negative. An attempt was made to lower the threshold for the acceptance of a possible vector; the result was a very large number of possible heavy atom structures which made the identification of the correct solution rather difficult.

Several attempts with both conventional and sharpened isomorphous difference Patterson functions for the most promising derivatives have been unsuccessful. As before, some solutions gave very good statistics during least squares refinement, but no further progress could be made with any of them.

5.3.6 The Argos & Rossmann Method.

Argos, P. & Rossmann, M.G., 1974,1976, were the first to realise that in the presence of known non-crystallographic symmetry, the problem of determining the positions of the heavy atoms in the crystallographic asymmetric unit of a derivative can be reduced to that of determining their positions in the non-crystallographic asymmetric unit (assuming that the heavy atom structure obeys the non-crystallographic symmetry of the macromolecular assembly).

When the position of the molecular centre in the crystallographic frame is not known, their method is divided in two steps.

In the first step (Type I search), an attempt is made to determine the positions of the heavy atoms with respect to the molecular centre : for every position \mathbf{x} in the non-crystallographic asymmetric unit, the positions of all non-crystallographically equivalent positions of \mathbf{x} are generated and the vectors between those sites are calculated. The sum of the density of the isomorphous difference Patterson function at the end of the predicted vectors is a measure of how well a heavy atom site at \mathbf{x} agrees with the observed Patterson function. It is obvious that at this stage, vectors between heavy atoms that are bound to crystallographically related molecules, as well as vectors between multiple heavy atom sites on the same protein subunit, are not been taken into account.

Once the positions of the substitution sites with respect to the molecular centre are known, the position of the molecular centre with respect to a permissible —for the given space group— origin, can be determined (Type II search) : For every position of the molecular centre in the unit cell of the corresponding Cheshire group (Hirshfeld, F.L., 1968), the sum of the values of the isomorphous difference Patterson function at the end of all predicted vectors (now taking into account the crystallographic symmetry) is evaluated. Clearly, this second step is a translation function very similar to those discussed in Sections 5.4.2 and 7.2.1.

If the position of the molecular centre in the crystallographic frame is known,

then it is possible to include this information in the Type I search : for every equivalent (by non-crystallographic symmetry) position of \mathbf{x} , the vector corresponding to one of the equivalent positions of the molecular centre is added and all crystallographically related sites are generated. The summation now is over all unique vectors between heavy atoms related by both non-crystallographic and crystallographic symmetry. The inclusion of vectors between heavy atoms that are bound to crystallographically related molecules offers a significant improvement of the signal to noise ratio.

If there is more than one heavy atom site per non-crystallographic asymmetric unit and the positions (with respect to the molecular centre) of one or more of these atoms have been determined, then, it is possible to repeat the Type I search with these heavy atoms 'fixed' at their known positions. This search can be performed with or without knowledge of the position of the molecular centre.

A set of programmes has been developed to carry out the calculations required for these two types of searches. Since in our case the amount of calculation involved is not as large as in the case of icosahedral viruses (discussed in detail by Argos, P. & Rossmann, M.G., 1976), no attempt was made to reduce the number of vectors that are used in the various steps of the procedure. The programmes allow the Type I search to be conducted with or without knowledge of the position of the molecular centre or of any 'fixed' heavy atom sites. In the translation function step (Type II search), all known heavy atom sites in the non-crystallographic asymmetric unit can be used.

Although the Type I search need only be done within the non-crystallographic asymmetric unit (that is, in the molecular as opposed to the crystallographic frame), in the case of AhrC where the non-crystallographic symmetry was not known with certainty, a more general scheme was used : the search was carried out in the (orthogonal) crystallographic frame and a volume equal to a hemisphere around the molecular centre was examined (the radius of the sphere is a

user-defined parameter). This means that in those of our attempts where 32 symmetry was assumed, the amount of calculation involved was as much as 3 times that needed. The advantages of this approach are (i) any point group symmetry can be used without any need for modifying the programmes, and, (ii) the analysis of the results is simplified since no conversion to the crystallographic frame is necessary. The programmes have been tested with hypothetical heavy atom structures obeying various non-crystallographic point group symmetries (32, 622, 432, etc.). These calculations showed that when the number of heavy atoms in the non-crystallographic asymmetric unit was greater than about three, spurious peaks appeared in the Type I search conducted without knowledge of the molecular centre. When the position of the molecular centre was treated as known, the heavy atom positions could be easily identified in all cases examined.

All indications from the self rotation functions for both the native crystals and the heavy atom structures (Chapter 4), have been examined using the Argos & Rossmann method with the isomorphous difference Patterson functions for the most promising AhrC derivatives.

From these attempts, we will discuss in detail the results from the calculations for the Niobium derivative which were based on the assumption that the point group symmetry of the heavy atom structure is 32 with the 3-fold at $\omega = 90^\circ$, $\phi = 60^\circ$ and one of the 2-folds parallel to z (discussed in Section 4.2.2). Figure 5.14 shows the results from the Type I search conducted without knowledge of the position of the molecular centre. Figure 5.15 is the same search but this time assuming that the molecular centre is at $x = 0.13$, $y = 0.09$ and $z = 0.225$ (Because the programme calculates a volume equal to a hemisphere around the molecular centre, some of the peaks in these maps are related by non-crystallographic symmetry. Such peaks are marked by the same numeral).

The similarity of these two functions suggests that the assumed non-crystallographic symmetry is consistent with the previously determined position of the

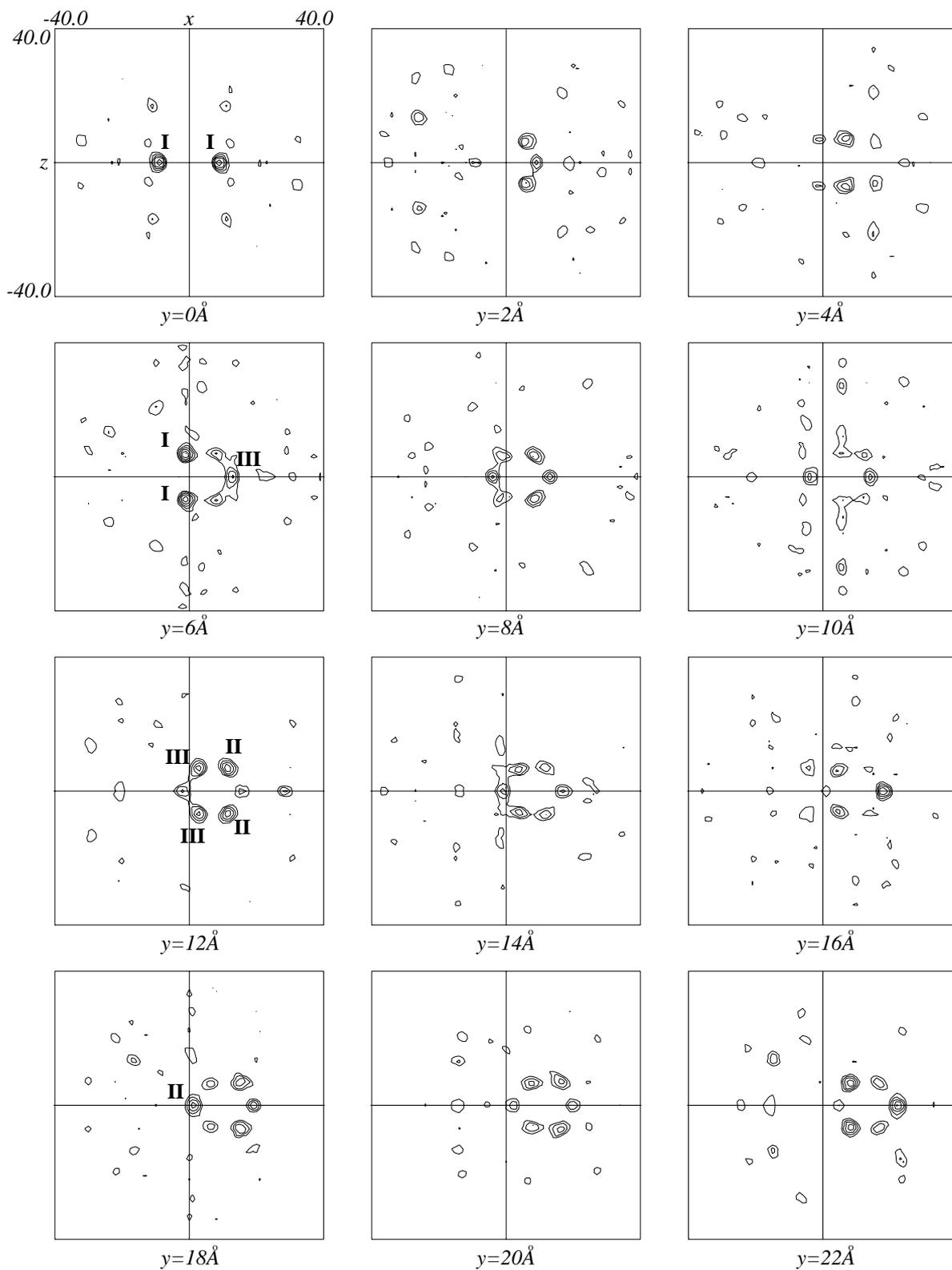


Figure 5.14: Argos & Rossmann, Type I search for $\text{Nb}_6\text{Cl}_{14}$, Molecular centre unknown. Contours every 1.5σ with first contour at 1.5σ above the mean.

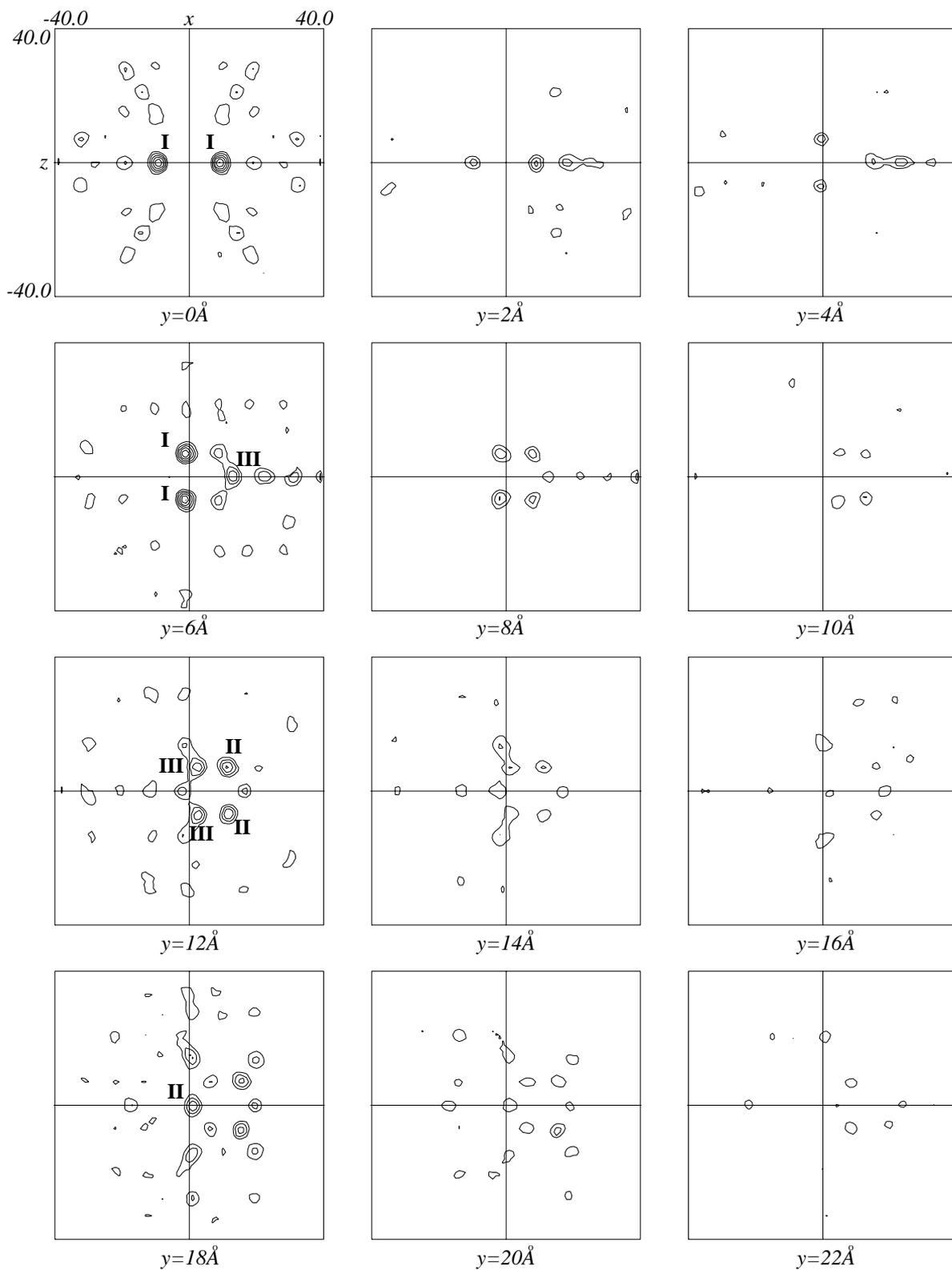


Figure 5.15: Argos & Rossmann, Type I search for $\text{Nb}_6\text{Cl}_{14}$. Molecular centre at $x = 0.13$, $y = 0.09$, $z = 0.225$.

molecular centre. Figure 5.16 shows the results from the Type II search using only the six atoms corresponding to the peak marked as (I) in Figures 5.14 and 5.15. The highest translation function peak is at $x = 0.13$, $y = 0.09$ and $z = 0.22$ as expected. Peak (I) in Figures 5.14 and 5.15 is clearly the most significant feature of these maps. The six heavy atom sites corresponding to it (through application of the non-crystallographic symmetry operators) are at the corners of an approximately regular octahedron whose edges are $\approx 14\text{\AA}$ long. The orientation of the octahedron is such that three orthogonal, intersecting, non-crystallographic 2-folds are parallel to the three crystallographic 2-fold axes, thus generating large pseudo-origin peaks on all three Harker sections of the ΔF_{calc}^2 Patterson function².

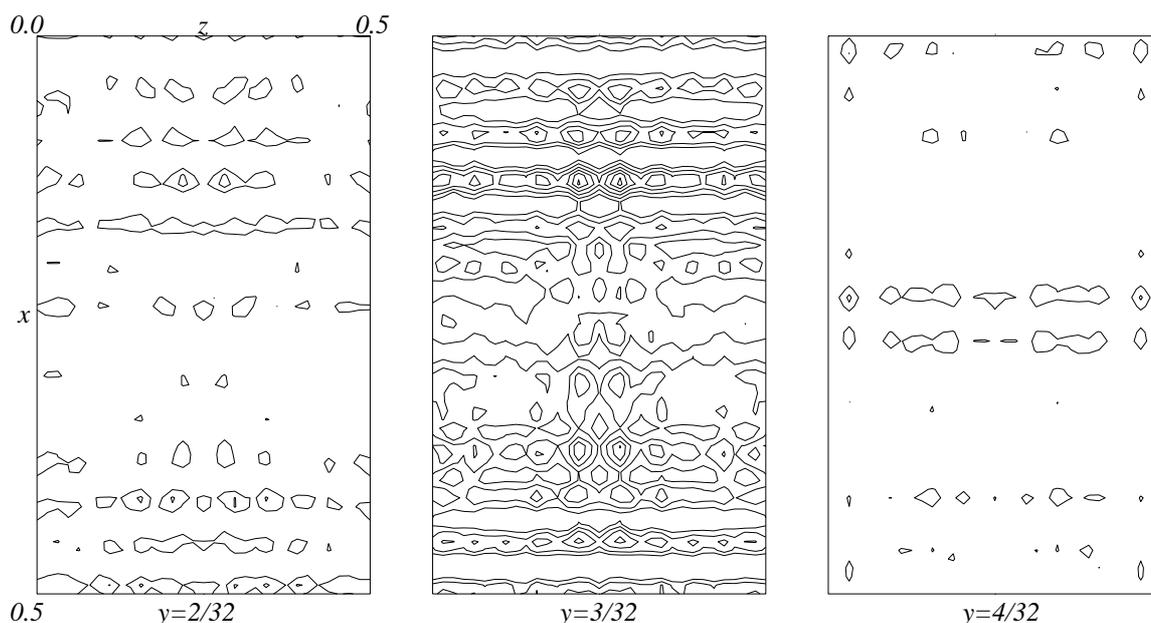


Figure 5.16: Argos & Rossmann, Type II search for $\text{Nb}_6\text{Cl}_{14}$. Three sections through the highest peak are shown.

A comparison of the observed $13\text{-}7\text{\AA}$ sharpened isomorphous difference Patterson function for the $\text{Nb}_6\text{Cl}_{14}$ derivative with the Patterson function calculated

²The three orthogonal non-crystallographic 2-folds are generated by the extra symmetry that an octahedral arrangement of heavy atoms possesses: the point group symmetry of a regular octahedron is $m\bar{3}m$, which has the 32 point group as a subgroup.

from these six heavy atom sites alone, is shown in Figure 5.17 (for clarity, the contouring level for the observed Patterson has been increased to 3% of the origin peak). Although several of the most prominent peaks can be accounted for, some of the predicted peaks are not present or are misplaced by 0.025 along w . Least squares refinement of these six sites against all centric terms between 15 and 6Å resolution converged to rather high R-factors and a very low correlation coefficient between the observed and calculated amplitudes for the heavy atom structure.

An attempt to determine the heavy atom structure of the Iridium derivative using SIRAS phases calculated from the unrefined six atom structure for the Niobium derivative was unsuccessful. These problems suggested that the indications from the self rotation function for the Niobium derivative could not lead to a successful structure determination.

It was a surprise to find that a Fourier synthesis for the native crystals which was calculated using SIRAS phases from the unrefined six atom structure for the $\text{Nb}_6\text{Cl}_{14}$ derivative (Figure 5.18) showed well defined solvent channels with a distribution of density which was in very good agreement with the model of the crystal packing and the electron microscopical studies of AhrC crystals. The calculation of reasonable low resolution protein density maps from incorrect heavy atom structures was unexpected. Further investigation of this problem led to the conclusion that when the non-crystallographic point group symmetry of the heavy atom structure is the same as (or is a supergroup of) the crystallographic point group symmetry and the non-crystallographic and crystallographic symmetry axes are parallel, then, protein phases calculated from any heavy atom structure that obeys the non-crystallographic symmetry are not random, but will lead to a protein density map consisting of images of the origin-removed Patterson function of an isolated protein molecule. To put this in a more practical way, protein phases calculated from any heavy atom structure that can generate the

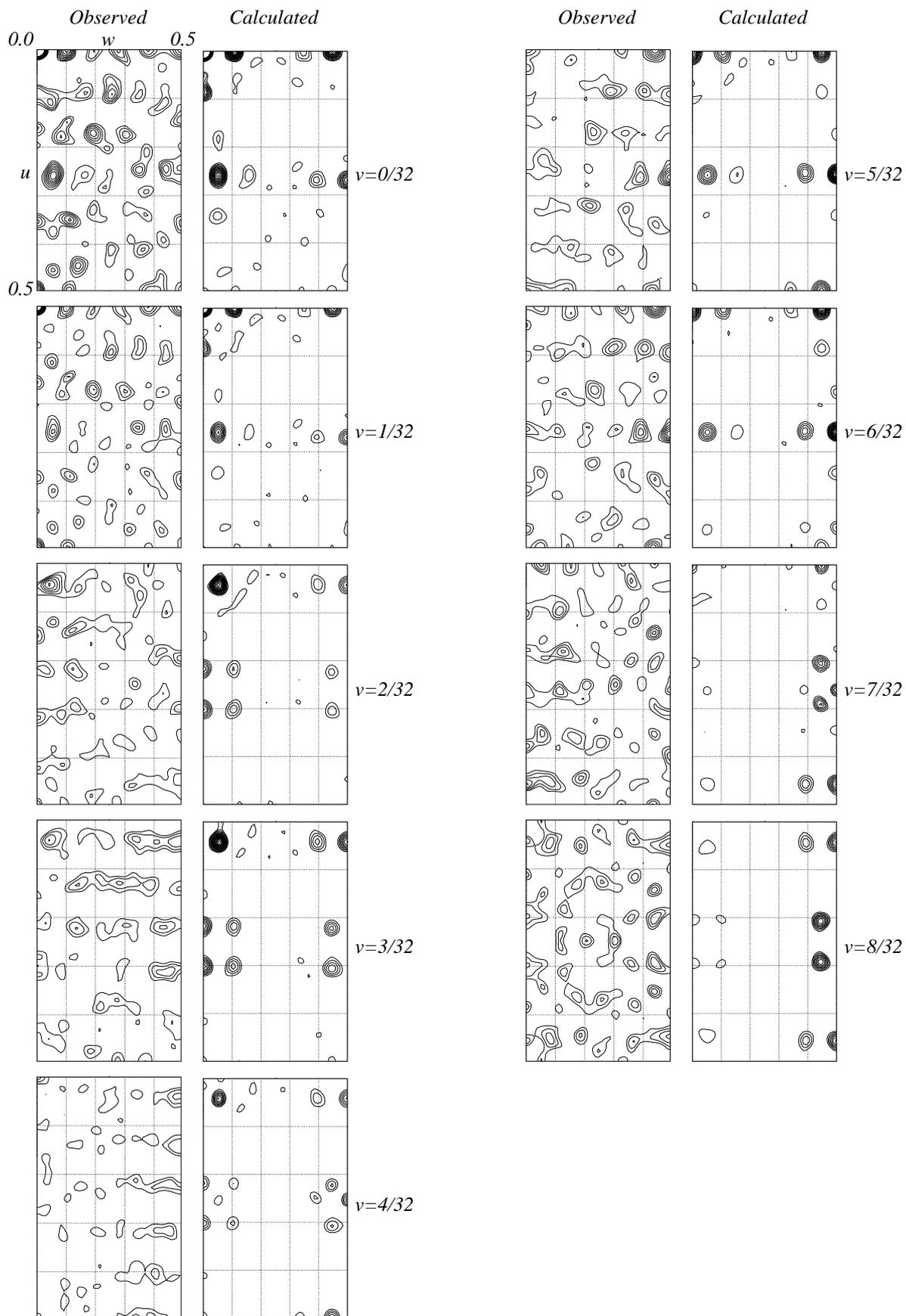


Figure 5.17: Observed and calculated Patterson functions for the $\text{Nb}_6\text{Cl}_{14}$ derivative.

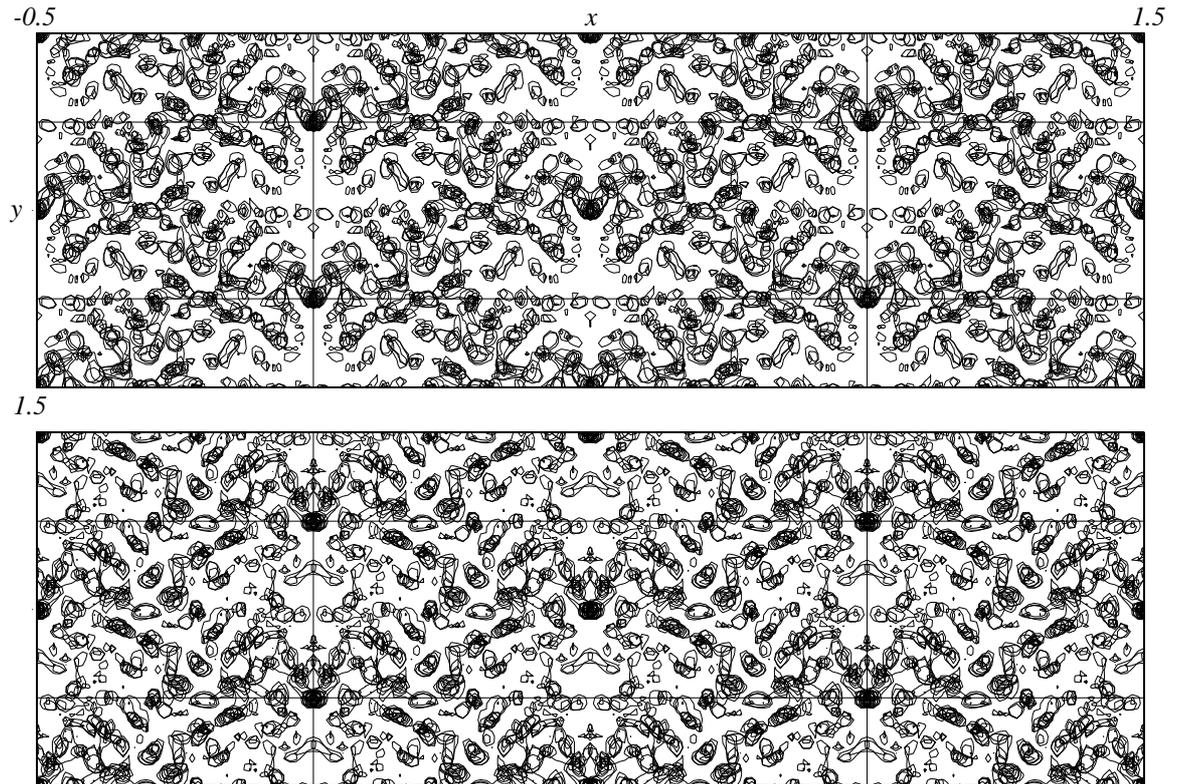


Figure 5.18: 10Å [001] projection of native AhrC crystals, Top : (+), Bottom : (−) hand. Phases calculated from the Niobium derivative modelled as an octahedral arrangement of six sites.

large pseudo-origin peaks seen in the ΔF^2 Patterson function for the Niobium derivative, are not random and are bound to generate reasonable, but incorrect low resolution protein density maps.

Figure 5.19 shows the results of model calculations which were undertaken in order to confirm the argument presented above. Figure 5.19(A) is the [010] projection of the electron density of a hypothetical “native” crystal structure consisting of 36 atoms obeying 32 symmetry with all non-crystallographic symmetry axes at general orientations. Figure 5.19(B) is the projection of a “derivative” of the structure shown in (A) : four “heavy” atoms have been added in such a way that the heavy atom structure has point group symmetry 222 with the non-crystallographic 2-fold axes parallel to all three crystallographic 2-folds (the space

group of these hypothetical structures is $C222_1$). Figure 5.19(C) shows the $v = 0$ section from the isomorphous difference Patterson function for this hypothetical derivative. The large pseudo-origin peak at $u = 0.25$, $w = 0.05$ arises from the parallelism between the non-crystallographic and crystallographic 2-folds. Least squares refinement of a “heavy atom” site at the position of the molecular centre against all centric terms between 13 and 6\AA resolution gave statistics of acceptable quality ($R_c = 0.55$, $C = 0.59$). SIRAS phases were calculated from this refined atomic position and a Fourier synthesis for the “native” crystals was calculated using coefficients $m(F_p)e^{i\alpha_{SIRAS}}$. This is shown in Figure 5.19(D). It is obvious that this map has high density at the correct position but the density itself is incorrect (it is an image of the origin removed Patterson function of an isolated hypothetical molecule).

It was not a surprise to find that a Fourier synthesis for the native AhrC crystals which was calculated using SIRAS phases from the Niobium derivative modelled with one site at the molecular centre (Section 3.1.1), showed the now familiar distribution of density down the $[010]$ and $[001]$ directions (Figure 5.20). The agreement between the permutation map (A) in Figure 4.6 with the $[010]$ projection in Figure 5.20 is worth noting.

In conclusion, the presence of non-crystallographic symmetry axes which have a “special” orientation with respect to the crystallographic, makes a reasonable low resolution protein map a necessary but not sufficient condition for the correctness of a proposed heavy atom structure.

5.4 Other Methods.

5.4.1 Permutation Syntheses Method.

The application of the permutation syntheses method to centrosymmetric, low resolution terms, has been described (Section 4.3.2). The extension of the method

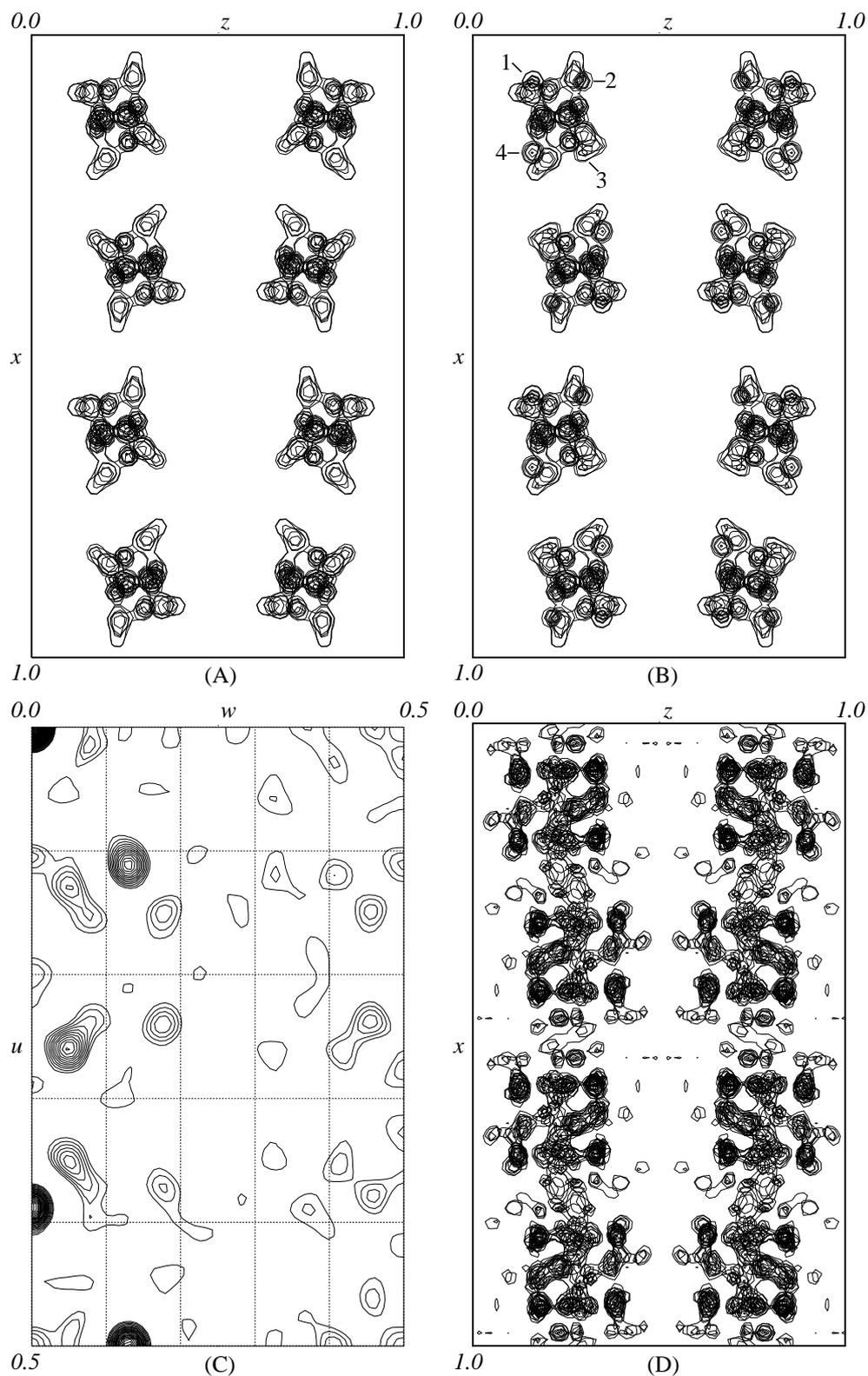


Figure 5.19: (A) A hypothetical 36 atom structure, (B) Its derivative, (C) The difference Patterson function, (D) A Fourier synthesis based on modelling the derivative as a single site at the molecular centre.

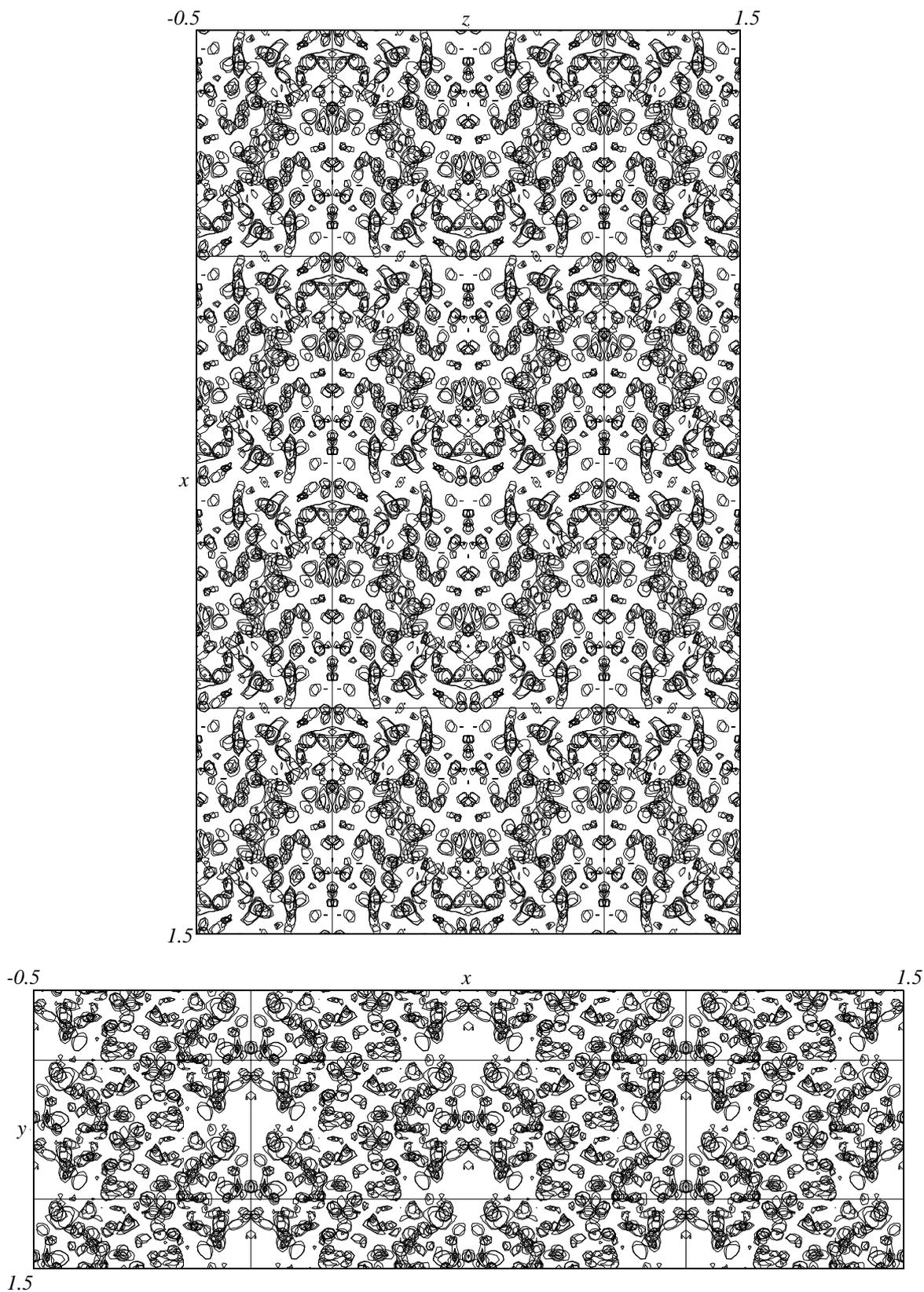


Figure 5.20: [010] & [001] projections of native AhrC crystals based on modelling the $\text{Nb}_6\text{Cl}_{14}$ derivative as a single site at the position of the molecular centre.

in the special case of ΔF data is straightforward : Fourier syntheses are calculated for all unique phase combinations of a small number of large ΔF terms, and the most promising of these are further analysed by other methods (least squares refinement, cross difference Fourier syntheses for other derivatives, etc.). The resolution range over which these terms are chosen is important. Low resolution ΔF data are greatly affected by differences in the electron density of the solvent of crystallisation and are not very useful for the determination of the heavy atom positions. High resolution ΔF terms are sensitive to lack of isomorphism and tend to give permutation maps with a grid pattern of fringes (this is due to the small number of terms included in the syntheses). For most of our attempts, data between 13 and 7Å resolution were used.

The application of the method to ΔF data is complicated by the fact that the only criteria for selecting those phase combinations which are likely to be correct are (i) the absence of heavy atom sites from what was known to be solvent areas, and, (ii) the concentration of electron density at few atomic sites. Since in the [010] projection the solvent channels are less well defined, our efforts focused on the [001] projection. Figure 5.21 shows two of the most promising permutation syntheses for the [001] projection of the Niobium derivative using the seven largest $hk0$ terms between 13 and 8Å resolution. The agreement with the model of the crystal packing and the concentration of the electron density at few atomic sites is obvious, but, as further examination of these solutions suggested, it can only be taken to mean that the data are not inconsistent with our presumptions.

Several sets of $hk0$ and $h0l$ ΔF terms for the Niobium and Iridium derivative have been examined. Although some of the heavy atom structures suggested by the permutation maps could be refined with very good statistics, none of these solutions led to protein projection maps with the expected distribution of density.

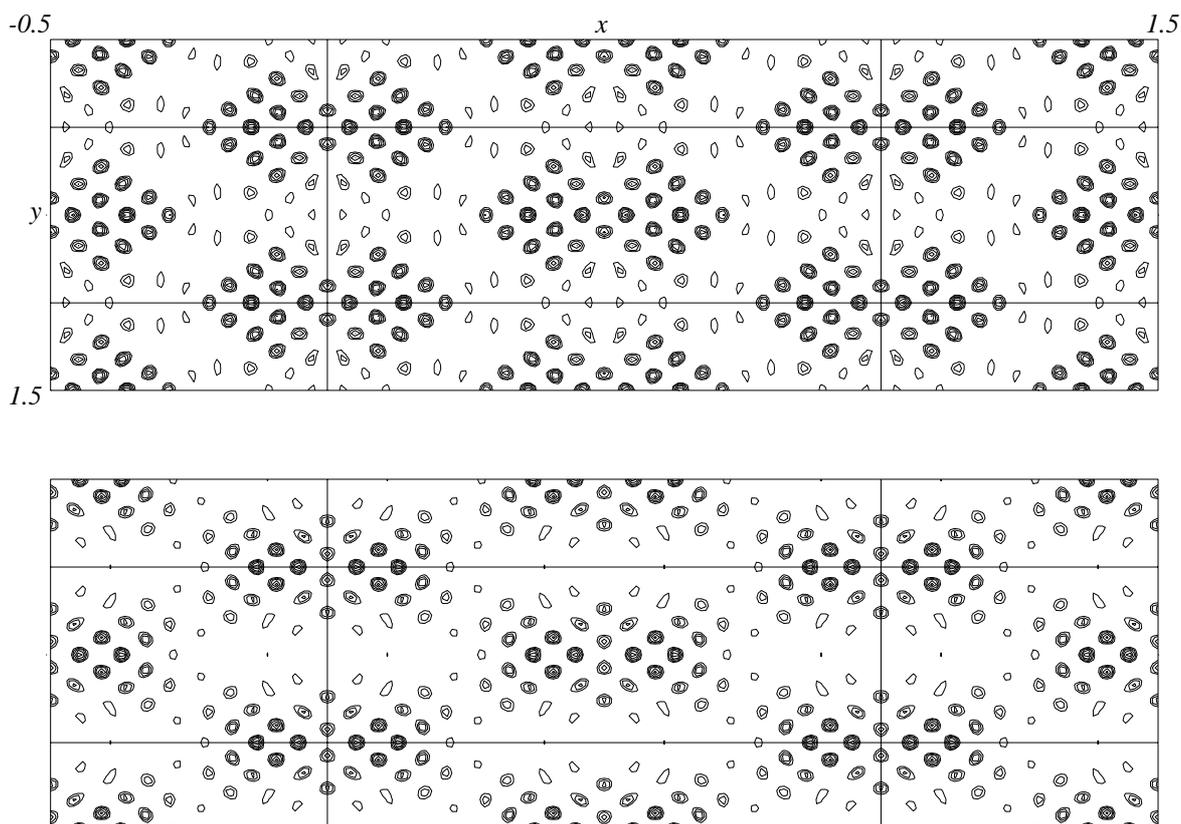


Figure 5.21: [001] permutation syntheses for the Niobium derivative.

5.4.2 Molecular Replacement.

Examination of the sharpened isomorphous difference Patterson for the Niobium and Iridium derivatives at the region around the origin of the function, led to the development of several possible (partial) heavy atom structures. For example, the $u = 0.0$ Harker section from the sharpened isomorphous difference Patterson function for the Niobium derivative (Figure 5.22) shows strong peaks at the corners of an approximately regular hexagon centered at the origin of the function. The average length of the vectors corresponding to those peaks is 14\AA . A partial heavy atom structure consisting of three sites at the corners of an equilateral triangle with the 3-fold parallel to x and one of the edges parallel to z can account for these features. Similarly, the $u = 0.0, v = 0.0, w$ line shows a regular arrangement of equidistant Patterson peaks which can be accounted for by four heavy atom sites arranged on a line parallel to z with interatomic distances equal

to 13.8, 41.1 and 13.8Å.

The determination of the position of such models in the crystallographic frame was attempted using the translation function programme BRUTE (Fujinaga, M. & Read, R.J., 1987). For every crystallographically unique position of the search model in the known cell, BRUTE calculates the correlation coefficient and the R-factor between the observed and calculated intensities of the structure factors³. Figure 5.23 shows three sections through the highest peak of the translation function for the second of the models described above. The major peak is at $x = 0.13$, $y = 0.09$, $z = 0.22$ with a corresponding value of the linear correlation coefficient of 0.42. The four heavy atom sites are symmetrically arranged about the assumed position of the molecular centre. A comparison of the observed Patterson function for the Niobium derivative with the Patterson function calculated from these four heavy atom sites is shown in Figure 5.24. Because all four sites are on a line parallel to z , the Patterson function calculated from this structure has peaks only on sections $v = 0.0$ and $v = 0.18$. Although the agreement with the observed Patterson function is acceptable (for a partial heavy atom structure), these four sites could not be refined. Furthermore, a difference Fourier synthesis for the Iridium derivative which was calculated using SIRAS phases from the unrefined heavy atom positions, showed a more or less uniform distribution of peaks that could not be interpreted in terms of a heavy atom structure. Very similar results have been obtained from all models examined.

5.4.3 Direct Maximisation of an Electron Density Function.

The [001] permutation syntheses in Figure 5.21 showed the level of clarity to be expected even from few correctly phased $hk0$ terms. Our attempts to determine

³The procedure is made practical through the use of partial structure factors for each crystallographically related copy of the search model. In this way, the computationally expensive Fourier transforms can be avoided (Nixon, P.E. & North, A.C.T., 1976)

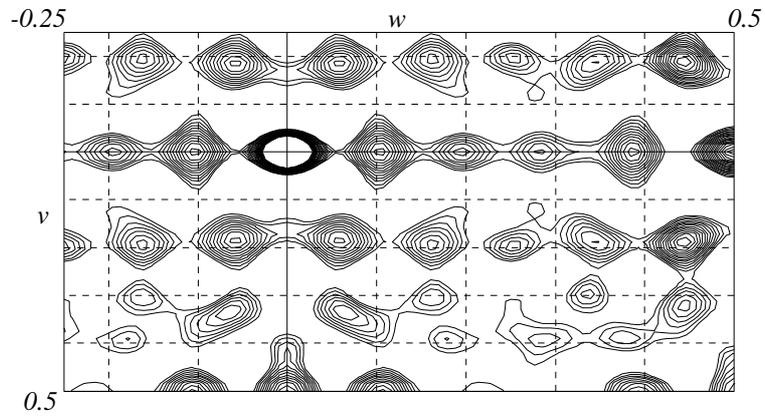


Figure 5.22: $u = 0.0$ Harker section from the sharpened isomorphous difference Patterson function for the Niobium derivative. Contours every 2% of the origin peak.

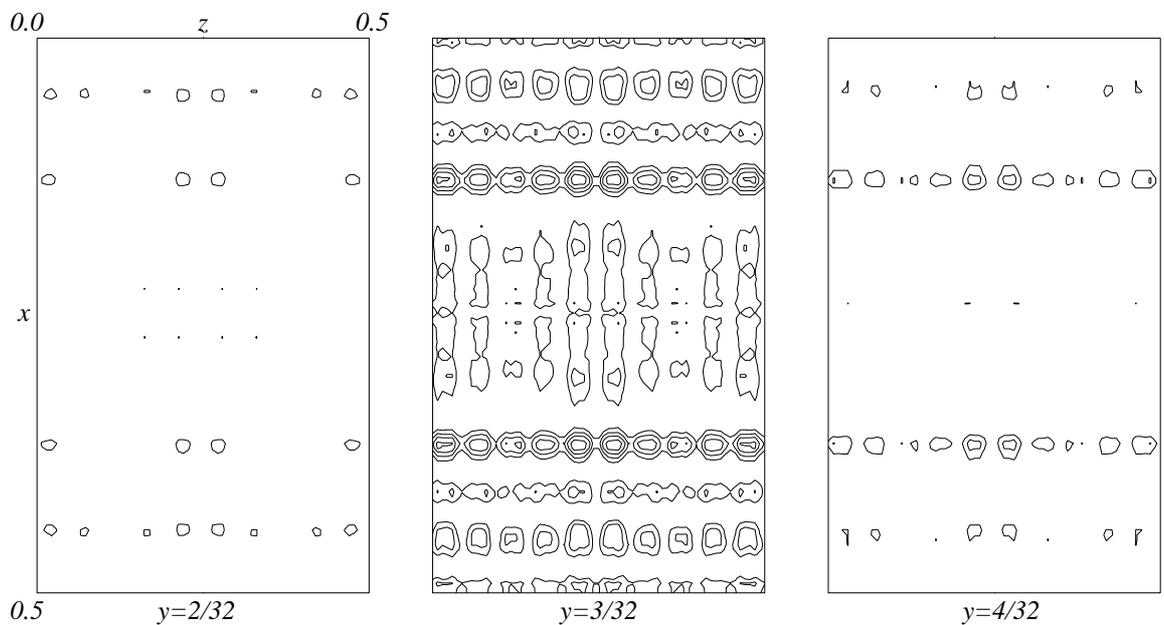


Figure 5.23: Three sections from the translation function using BRUTE.

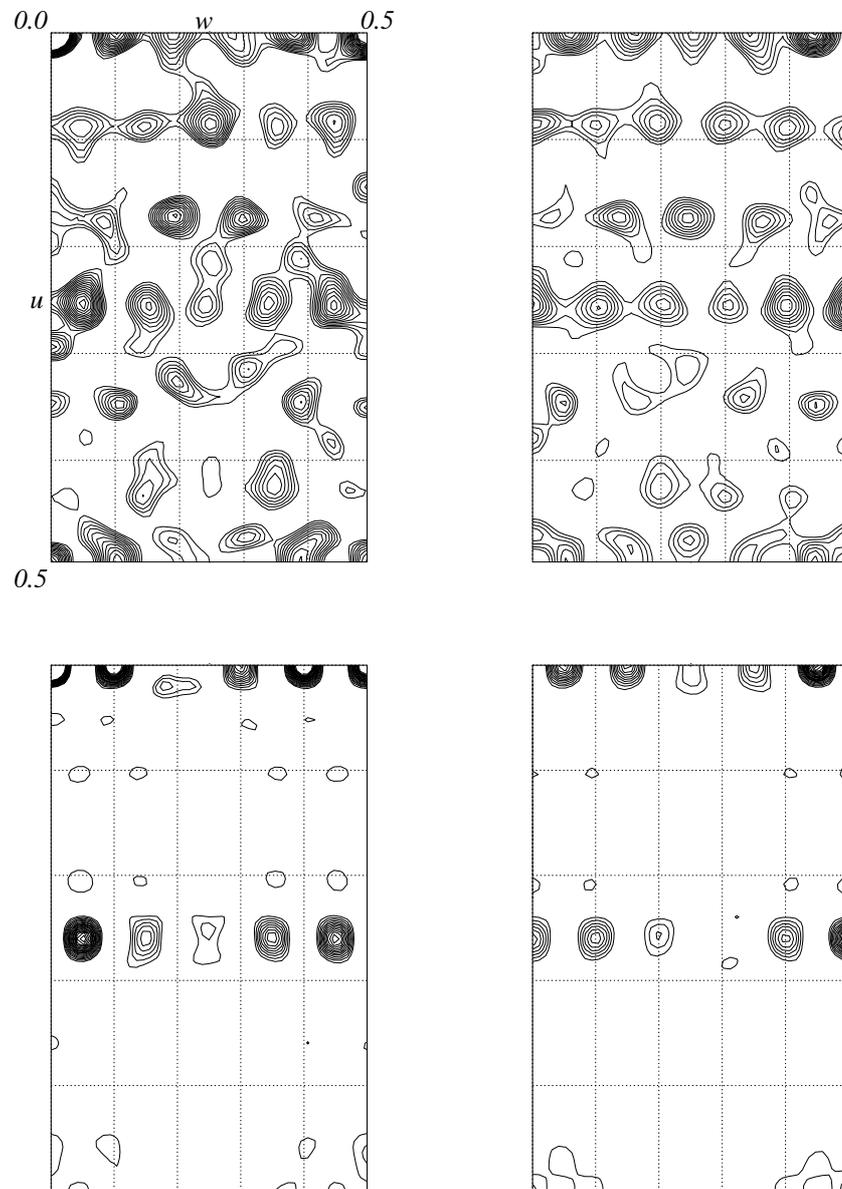


Figure 5.24: Comparison of two sections from the observed (top row) and calculated Patterson function for a four-atom structure.

the phases of the $hk0$ terms using either the direct methods programme SHELXS-86 (Section 5.2.1) or the Cochran & Douglas method (Section 5.2.3) have been unsuccessful. The main problem with the application of the Cochran & Douglas method to the $hk0$ terms, appears to be the absence of translational symmetry in this projection. The plane group of the $[001]$ projection is cmm for which $F(hk) = F(\bar{h}\bar{k}) = F(\bar{h}k) = F(h\bar{k})$, that is, all crystallographically related structure factors have the same phase and, thus, all triple sign products are satisfied with the trivial solution $s(F_{\mathbf{h}}) = +1$ for every \mathbf{h} (This is not the case, for example, with the $[010]$ projection (plane group pmg) for which $F(hk) = F(\bar{h}\bar{k}) = -F(\bar{h}k) = -F(h\bar{k})$ when $l = 2n + 1$). In other words, the “best” phase set from the Cochran & Douglas method will be the one with all signs positive, and the phase sets that follow will be linear combinations of the indices of the reflections that enter the triple sign products. The solution with all signs positive will give a map with a very large peak at the origin and will, thus, maximise the value of the integral $\int_V \rho^3 dV$ (Equation 5.2).

In the case of AhrC derivatives, where a large number of substitution sites are present, the function $\int_V \rho^3 dV$ —which reaches a maximum for the map with the smallest number of peaks consistent with the data—is not very useful. An attempt was made to identify an electron density function that would be more appropriate for the given problem. The most promising results have been obtained from the function

$$\int_V \rho^3 dV / \max(\rho) \text{ is maximum positive.} \quad (5.11)$$

Model calculations showed that this function will take its maximum value for the phase set which gives the largest number of peaks consistent with the data. The same calculations showed that when a large number of sites is present, Equation 5.11 can give better results than Equation 5.2, that is, the correct phase set can be closer to the top of the list of best solutions for Equation 5.11 than for Equation 5.2 (Stanley, E.. 1979, 1986 also discussed various electron den-

sity functions, but his conclusion was that the most useful integral is $\int_V \rho^3 dV$. Function 5.11 was not considered in his papers).

A programme was written, which for every unique phase combination of a given number of terms, calculated the value of the function 5.11. At the end of the procedure, the “best” phase sets were saved for further examination. Since there is no analytical approximation to the value of $\max(\rho)$, it is necessary to calculate a Fourier synthesis for each phase combination. This means that the number of terms that can be examined is not as large as is the case for the Cochran & Douglas method.

The programme has been tested with both $hk0$ and $h0l$ ΔF data (the 21 largest differences between 13 and 7Å resolution were used for both calculations). The top 100 phase sets from each run were evaluated as discussed in Section 5.2.3. No convincing solutions could be identified.

5.4.4 Monte Carlo Methods.

We have seen how all analytical approaches to the problem of determining the heavy atom structures of the most promising AhrC derivatives were unsuccessful. It was hoped that a method based on a modified direct minimisation of a suitable statistical metric (such as the R -factor or $(1-C)$, where C is the linear correlation coefficient), although less elegant, would avoid the need for any assumptions other than the basic crystallographic requirement that the global minimum of the chosen statistic corresponds to the true crystal structure or one of its homometric pairs⁴.

Our first attempt with such a method was based on refinement of randomly chosen atomic configurations : The random number generator was used to prepare a list of fractional coordinates for a given number of atoms in the crystallographic

⁴Strictly speaking, this requirement can only be true in the case of very accurate (small molecule) data. When the errors of measurement are significant, the correct treatment is given by the Maximum Entropy methods.

asymmetric unit. The positional parameters of those atoms were refined using an implementation of Hart's algorithm with target function the linear correlation coefficient between the observed and calculated structure factor amplitudes. If the value of the correlation coefficient after refinement was greater than a preset limit, the refined positional parameters were stored for further use. The procedure was repeated for a new random structure until a user-defined number of iterations had been completed. In the final step, all atomic sites saved during the main part of the the calculation were plotted on a map that was a copy of the crystallographic asymmetric unit. If the procedure had been successful, these should cluster at few, hopefully correct, atomic positions⁵.

Several tests with data calculated from hypothetical heavy atom structures showed that the method could converge to the correct structure only when the number of heavy atom sites was less than about six per asymmetric unit. This is due to (i) the relatively small number of random structures that can be examined (in the order of few thousands), and, (ii) the rather inefficient sampling of the parameter space as will be discussed below. Several attempts with ΔF data for the Niobium derivative showed no sign of convergence.

Khachaturyan, A., Semenovskaya, S. & Vainshtein, B., 1981 and Semenovskaya, S.V., Khachaturyan, K.A. & Khachaturyan, A.G., 1985 (hereafter referred to as SKK) proposed a much more elegant and thoughtful Monte Carlo approach to the determination of a crystal structure. They started by noting the analogy between the determination of the thermodynamic equilibrium in statistical mechanics and the optimisation problem for a function of many variables, and they argued that the problem of determining a crystal structure can be viewed as that of determining the low-temperature state of a model gas composed of a known number of atoms in a known unit cell, with the unit cell and the R-factor

⁵It is worth noting that following the development of this algorithm, an abstract of a paper by Vand, V., Niggli, A. & Pepinsky, 1960 was discovered in which a closely related method was described.

being regarded as a vessel and a potential energy term respectively. They suggested that the low-temperature state can be determined with the variation of the Monte Carlo method proposed by Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. & Teller, E., 1953 (but see also Rosenbluth, M.N. & Rosenbluth, A.W., 1954, Wood, W.W. & Parker, F.R., 1957, Alder, B.J. & Wainwright, T.E., 1959, Hammersley, J.M. & Handscomb, D.C., 1964) :

When a classical system is in thermodynamic equilibrium with its surroundings, the expectation \bar{F} of any state-function $F(n)$ is given by

$$\bar{F} = \frac{\int F(n) \exp(-E_n/kT) dn}{\int \exp(-E_n/kT) dn} \quad (5.12)$$

where E_n is the energy of the state n , k is Boltzmann's constant, T is the absolute temperature of the surroundings and dn is a volume element in the multidimensional phase space (the dimension of the integral in Equation 5.12 equals the number of degrees of freedom of the system. For a two-dimensional structure consisting of N particles, the integral is $2N$ -dimensional). The simplest Monte Carlo method for calculating \bar{F} would be to choose a configuration n randomly, calculate its energy E_n , give F_n a weight $\exp(-E_n/kT)$ and iterate. When the number of configurations tends to infinity, the weighted average of F_n s will tend to \bar{F} . This procedure is inefficient because with high probability we choose low-weight states (that is, configurations for which $\exp(-E_n/kT)$ is very small). The adaptation of this rather primitive Monte Carlo method to the crystallographic problem is obvious : an atomic configuration n is chosen randomly, the structure factors $\mathbf{F}_{\mathbf{h},n}$ are calculated, the agreement between the observed and calculated amplitudes—in the form of the R-factor—is evaluated, the $\mathbf{F}_{\mathbf{h},n}$ are given a weight $\exp(-R_n/kT)$ and the procedure is iterated. When the number of structures examined tends to infinity, the weighted average of the structure factors will tend to its thermodynamic average both in amplitude and phase. As the temperature of the system tends to zero, the thermodynamic average of the structure factors will tend to the true (crystal structure) value. Clearly, this approach is

very similar to our first attempt with a Monte Carlo method as described above.

The objective of the Monte Carlo method proposed by Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. & Teller, E., 1953 is to generate a series (formally : a Markov chain) of configurations in which each state n recurs with frequency $\exp(-E_n/kT)$. When such a chain of configurations has been attained, the expectation \bar{F} of any state-function $F(n)$ can be found by taking the average of F over an infinitely long segment of the chain with all configurations being given the same weight, or as the authors put it, “instead of choosing configurations randomly and weighting them with $\exp(-E/kT)$, we choose configurations with a probability $\exp(-E/kT)$ and weight them evenly”.

Their algorithm for choosing states with probability $\exp(-E/kT)$, as modified by SKK for the crystallographic problem, is as follows. Initial coordinates for all (N) atoms in the crystallographic asymmetric unit are chosen randomly and uniformly. The random number generator selects one of these N atoms, and one of the six nearest grid points to which that atom may migrate (the atoms are constrained to move on a fine rectangular grid inscribed in the asymmetric unit). Whether the atom will move to this new position or not is decided as follows : If the new position is occupied by a different atom no migration occurs and the same atomic configuration is used as a starting point for the next iteration. If the new position is vacant, the change in the R-factor (ΔR) resulting from the migration to the new site is calculated. If $\Delta R \leq 0$, that is, the R-factor for the new configuration is lower, the move is accepted. If on the other hand, $\Delta R > 0$ the move may be realised with probability $\exp(-\Delta R/T)$: the random number generator chooses a number ξ such that $0 < \xi < 1$. If $\xi < \exp(-\Delta R/T)$ the move is accepted, otherwise no migration will occur and the previous configuration is taken to be the new configuration. Finally, a new atom is chosen and the procedure is iterated. The proof that this algorithm chooses configurations with probability $\exp(-R/T)$ will not be given here (see Wood, W.W. & Parker, F.R., 1957 and

Hammersley, J.M. & Handscomb, D.C., 1964). It is worth noting that since the paper by SKK, this method of minimising the difference between the observed and calculated amplitudes of the structure factors has been termed “the reverse Monte Carlo method” and found uses in such diverse areas of science, as structural modelling of glasses (Keen, D.A. & McGreevy, 1990 and references therein) and *ab initio* phasing of low resolution data from protein crystals (Subbiah, S., 1991, 1993, David, P.R. & Subbiah, S., 1994).

What we are interested in, is the average value of the structure factors over a segment of this chain of configurations that is sufficiently remote from its starting point. If we start taking our averages at a ‘time’ t_0 from the starting point of the chain and the averages are calculated for the time interval $(t_0, t_0 + t)$, then at $t \rightarrow \infty$ the average value of any structure factor will tend to its thermodynamic average. If the temperature T is low, the system will spend most of its time close to the ground state (that is, close to the global minimum of the R-factor) and the average value of the structure factors (both amplitudes and phases) will tend to their true values. SKK noted that the value of

$$\eta(\mathbf{h}, t) = \frac{|\langle \mathbf{F}_{\mathbf{h},t} \rangle|}{|\mathbf{F}_{\mathbf{h}}|_{obs}} \quad (5.13)$$

where $|\langle \mathbf{F}_{\mathbf{h},t} \rangle|$ is the amplitude of the average value of the structure factor $\mathbf{F}_{\mathbf{h}}$ at a time t and $|\mathbf{F}_{\mathbf{h}}|_{obs}$ is its observed value, will tend to 1 when the system approaches the global minimum and to 0 when the system is in a disordered state (it should be noted that there is no physical ‘time’ or ‘temperature’ involved in these calculations. ‘Time’ is a variable counting successive atomic configurations, whereas ‘temperature’ is a control parameter defining the average mobility of the atoms). SKK used L-proline (which consists of 8 atoms excluding hydrogens) as a model structure for testing their algorithm. For their calculations they used a $15 \times 15 \times 15$ grid (the average distance between grid points was 0.3 \AA), the averaging procedure was started at $t = 2.24 \cdot 10^6$ and lasted $7.04 \cdot 10^5$ time units. The maximum deviation of the final phases from their true values was only 23°

and the phases of all reflections for which $\eta(\mathbf{h}, t) > 0.4$ were correct.

A programme was developed to implement the ideas presented above. Although the basic algorithmic steps are identical with those given in SKK, some minor alterations were found to be necessary. The first modification is the calculation of the free R value (R^{free}) for a randomly chosen subset of reflections (usually 10% of the total) that were not allowed to enter the expression for the R-factor during minimisation (Brünger, A.T., 1992a). Because R^{free} is an unbiased indicator of the information content of the atomic model, it can be used to differentiate between a minimum of the R-factor that corresponds to the correct solution and a local minimum with no structural significance. R^{free} can also be helpful in deciding the approximate number of heavy atom sites which are likely to be present in the derivative. The second modification concerns the computational efficiency of the algorithm : the limiting step of the whole procedure is the structure factor calculation. In their paper, SKK used a space-group specific geometrical structure factor calculation algorithm (International Tables for X-ray Crystallography, Vol. I, 1952). When the number of atoms and reflections is small, this approach can be faster than FFT. What escaped their notice, however, is that because at each step only one atom is being moved, the calculation can be made independent of the number of atoms : If at $t = 0$ the contribution of each atom to each reflection is calculated, then for the rest of the minimisation it is only necessary to calculate the contribution to the structure factors of only the atom that is being tested. The method is fairly efficient : for 170 reflections, a DEC 4000 can go through 2314 configurations in one second of CPU time independently of the number of atoms involved (the same algorithmic improvement is also applicable to the method described by Subbiah, S., 1991, 1993, David, P.R. & Subbiah, S., 1994).

Figure 5.25 shows the results from model calculations using a hypothetical 12-atom structure in plane group pmg : (A) and (B) show the variation of the

R-factor and the free R value during the first 4,200,000 time units of the minimisation (the programme calculates and saves the average value of these two metrics over successive segments of the chain. In this example, the averages were calculated every 20,000 time units). It can be seen, that at $t = 4,000,000$ the system reached a deep minimum with $R=18\%$ and $R^{\text{free}}=24\%$ (the R-factor is not 0 due to the finite size of the grid used. For this calculation we used 7\AA data with a grid spacing of about 1.6\AA). This minimum corresponds to the true structure, shown in Figure 5.25(D) (the noise in this map is due to the omission of 56 weak reflections from the minimisation). Figure 5.25(C) shows the Patterson function for this hypothetical 12-atom structure. It is interesting to note that several attempts to determine the same structure using the direct methods programme SHELXS-86 have been unsuccessful : the violation of the assumptions behind direct methods is so serious that even error-free data can not help.

Figure 5.26 shows typical examples of the results obtained from minimisations using ΔF data from the most promising AhrC derivatives : both R-factor and R^{free} show a more a less uniform distribution with relatively high average values. For most of the reflections, the $\eta(\mathbf{h}, t)$ values (Equation 5.13) at the end of the minimisation were very close to zero and the $\Delta F_o \exp(i\phi_{\text{averaged}})$ maps could not be interpreted in terms of a heavy atom structure. Numerous attempts with different temperatures, number of atoms (6 to 16), chain lengths (4,000,000 to 50,000,000), averaging intervals (2,000,000 to 5,000,000), number of reflections, cooling protocols, etc., have all given very similar (negative) results.

There are several different possible explanations for these problems : the number of heavy atom sites could be larger than 16, the chain length or the averaging interval may be too short for the given problem, the errors of measurement may be too large for the method to work, etc. Model calculations with hypothetical structures consisting of up to 18 atoms showed that as the number of degrees of freedom of the system increases, the distribution of the R-factor (for the same

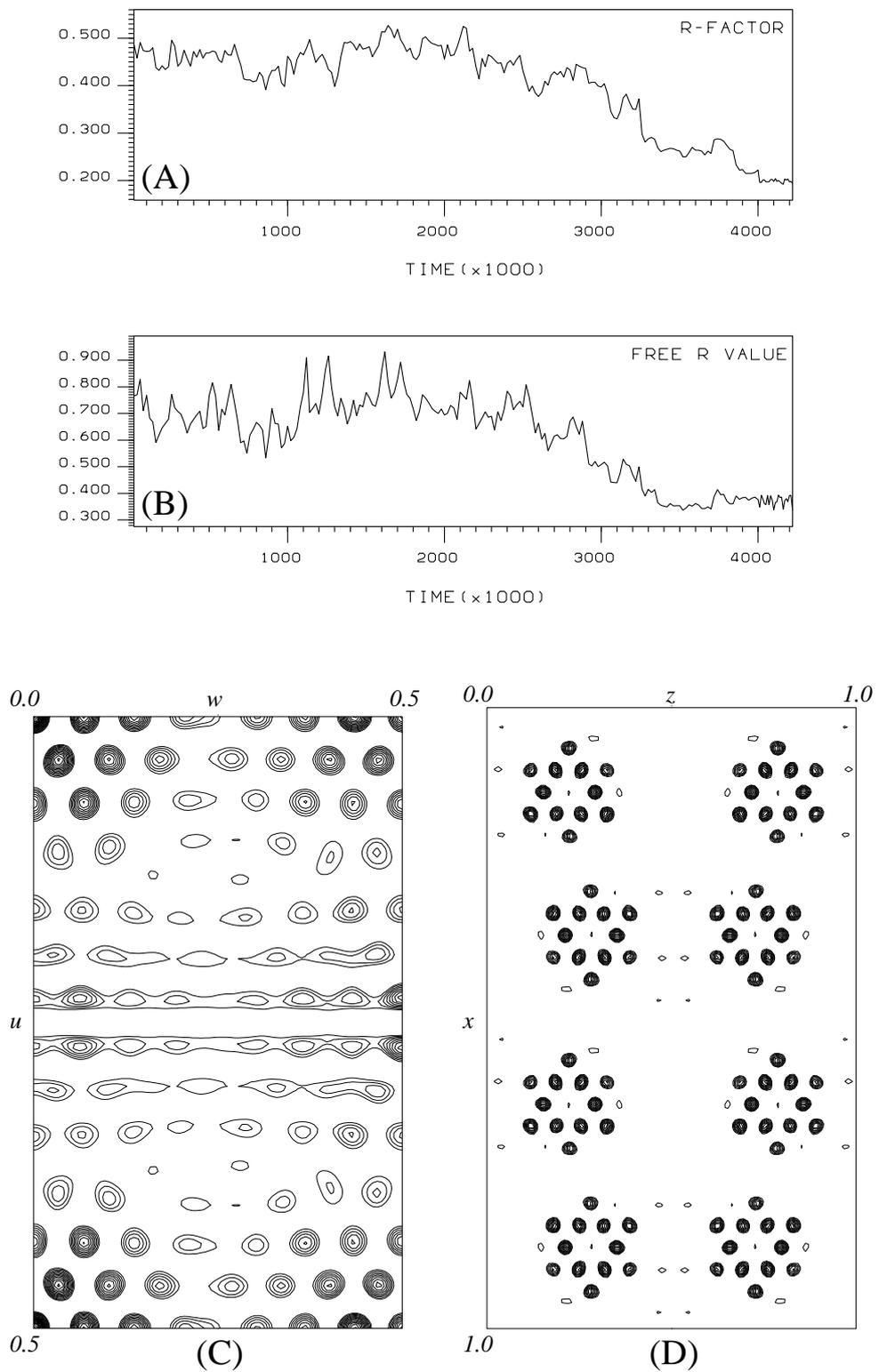


Figure 5.25: Monte Carlo : Results from model calculations with a hypothetical 12-atom structure. (A) R-factor, and, (B) Free R value during minimisation, (C) The Patterson function, (D) The $F_o \exp(i\phi_c)$ synthesis.

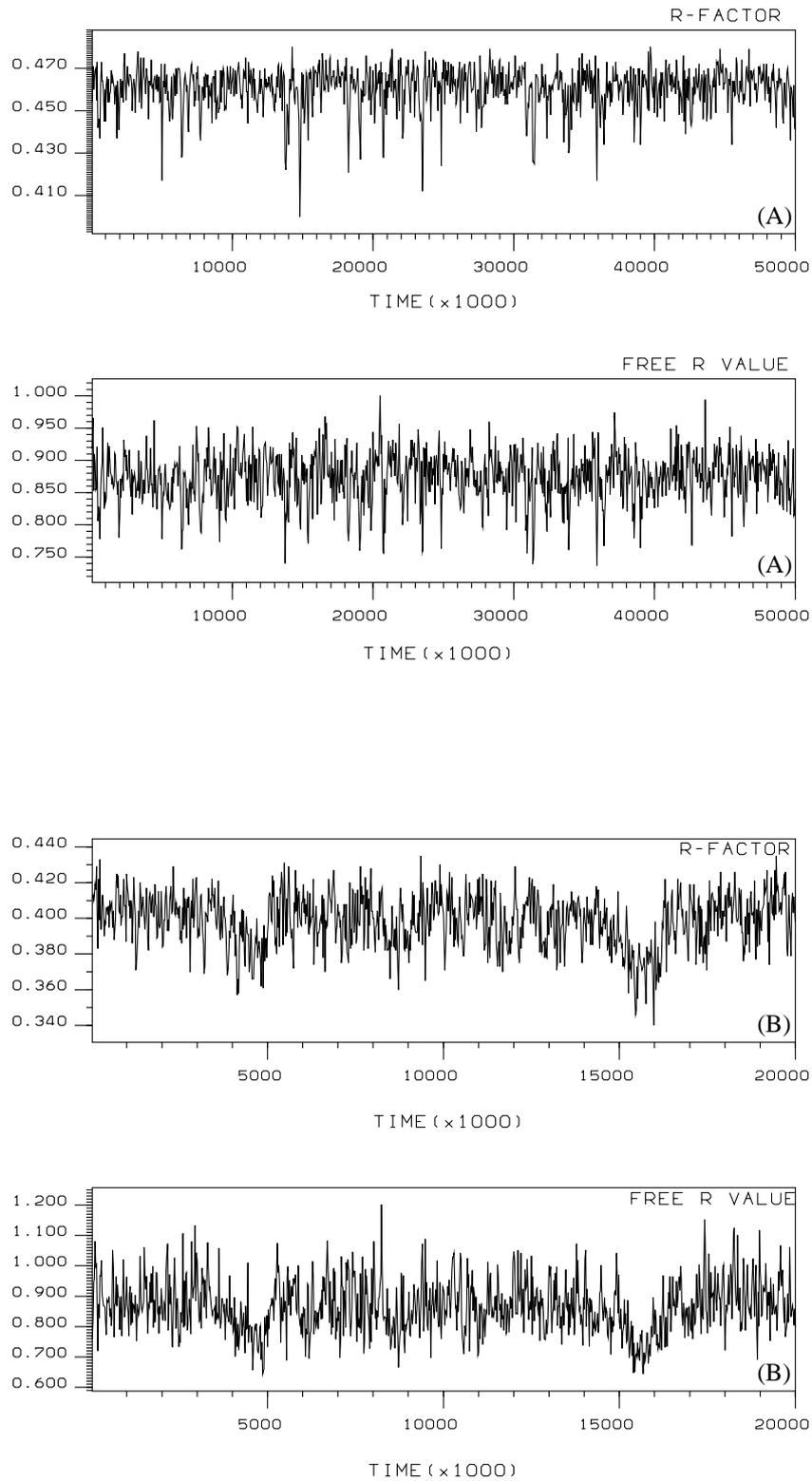


Figure 5.26: Monte Carlo : R-factor and R^{free} during two minimisations using $15\text{-}7\text{\AA}$ $h0l$ terms for the Niobium derivative, (A) : $T = 0.007$, 12 atoms, (B) : $T = 0.005$, 16 atoms.

number of observations) becomes more uniform with local minima as deep as the global minimum. The same behaviour was observed when random errors were introduced to error-free data in order to simulate the situation with ΔF data. Although the free R value can help discriminating between a local minimum and the correct solution, it can not be used to find the global minimum. Furthermore, if local minima are as deep as the minimum corresponding to the correct structure, extending the length of the chain or lowering the temperature of the system, will not help. Clearly, the best solution in such cases is to increase the number of observations by using higher resolution data, but this would give no improvement in the case of ΔF data from a heavy atom cluster.

5.4.5 An Attempt to Obtain Phase Information *ab initio*.

The knowledge of the position of the molecular centre together with a very low resolution approximation to an AhrC hexamer (a sphere of constant density), made possible the construction of a rudimentary model of the electron density for the orthorhombic form. A programme was written which given the radius and position (in the crystallographic asymmetric unit) of a number of spheres, prepared a electron density map (using the CCP4 map format) of the whole unit cell with the required number of spheres of constant density. This map was back-transformed using the programme SFC from the CCP4 suite of programmes and the resulting phases were combined with the observed amplitudes in a $F_o \exp(i\phi_{\text{spheres}})$ synthesis. Figure 5.27 shows a 7.7Å thick stack of four sections from a 115–15Å synthesis using phases calculated from a sphere of radius 32Å placed at the assumed position of the molecular centre.

Two features of this map are worth discussing : the first is the somewhat peculiar distribution of the electron density in the form of concentric shells of high density with regions of very low density between them. It is unlikely that this is due to ripples arising from series termination errors, since the majority

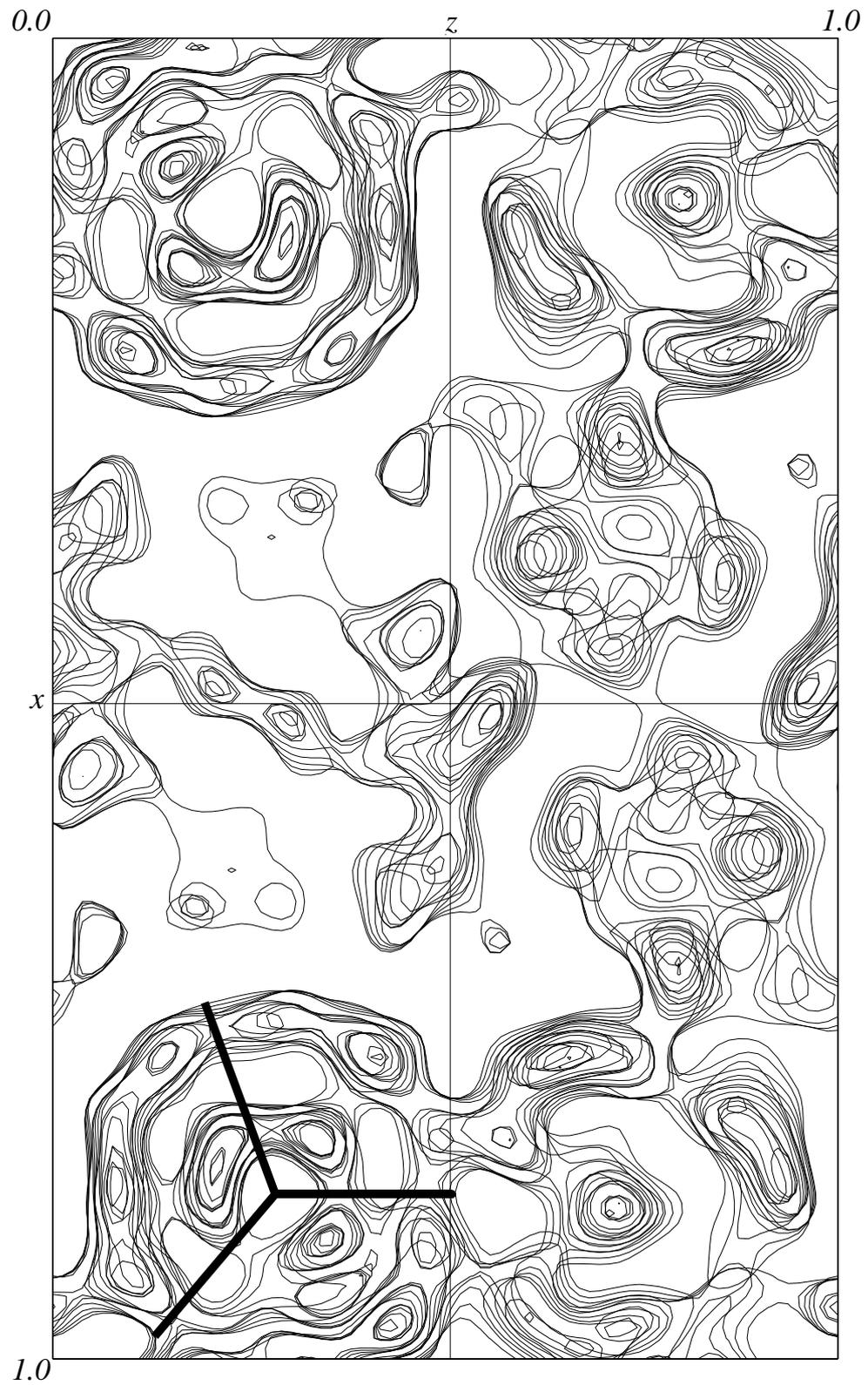


Figure 5.27: 115–15Å $F_o \exp(i\phi_{\text{spheres}})$ synthesis. Phases from a sphere of constant density at the assumed position of the molecular centre.

of the very low resolution reflections had been included in the calculation (it should be noted, however, that because these terms were estimated visually from precession photographs, their errors could be very large).

The second feature is the presence of an approximate non-crystallographic 3-fold axis parallel to y (indicated in Figure 5.27). This is consistent with the results obtained from self rotation functions calculated using low resolution data from native AhrC crystals (Figure 4.1). The appearance of a non-crystallographic 3-fold axis, which was not part of the starting model, raised hopes that phase refinement and extension through iterative real space averaging and solvent flattening, could allow us to obtain reliable phase information to a resolution sufficient for determining the heavy atom positions ($\approx 8\text{\AA}$ for the Niobium cluster) : Gerard Bricogne's suite of programmes "Joy of skewing" was used to extract a box containing a single hexamer from the $F_o \exp(i\phi_{\text{spheres}})$ map. This was averaged about the assumed non-crystallographic 3-fold axis and the density of the resulting (averaged) map was interpolated back to the crystallographic frame with a grid size identical to that of the original $F_o \exp(i\phi_{\text{spheres}})$ synthesis (Bricogne, G, 1974, 1976). A programme was written which reconstructed the whole C222₁ cell from the averaged density of a single hexamer. It was at this stage that regions outside a sphere of a given radius placed at the molecular centre (and its crystallographically equivalent positions) were set to zero. An option to truncate regions of very high negative or positive density was also available. The resulting (averaged and solvent flattened) map was back-transformed using the programme SFC, the observed and calculated amplitudes were brought to the same relative scale using the programme RSTATS (from the CCP4 suite of programmes), and the R-factor between them was calculated. The new phases could now be used to calculate a $F_o \exp(i\phi_c)$ synthesis and the procedure was iterated until convergence (as judged from the R-factor between the observed and calculated amplitudes) was achieved.

Several attempts with $F_o \exp(i\phi_{\text{spheres}})$ syntheses calculated using data from different resolution ranges failed to converge to convincingly low R-factors. Furthermore, the electron density of the $F_o \exp(i\phi_{\text{final}})$ syntheses, where ϕ_{final} are the phases from the last cycle of averaging, did not show the expected connectivity and were inconsistent with the low resolution images of the [001] and [010] projections obtained from electron microscopy.

These problems suggested that the 3-fold appearance of the map shown in Figure 5.27 is either accidental or an artifact of the procedure by which the phases were obtained : The assumption that the molecule is a sphere of constant density introduces extra (and probably unjustified) symmetry by imposing a centre of symmetry on the phases calculated from it. Furthermore, the position of the molecular centre is very close to 1/8th, 1/10th and 1/4th of the corresponding unit cell translations (if the molecular centre is at 0.13, 0.09, 0.225 the differences are $\Delta x = 1.15\text{\AA}$, $\Delta y = 0.74\text{\AA}$ and $\Delta z = 3.45\text{\AA}$). The proximity of the x and z coordinates to these “special” values, causes certain classes of acentric terms to behave as if centric (this is because the structure factors calculated from a sphere of constant density are identical to those calculated from a point atom with an unusual form factor at the centre of the sphere). For example, when $h + k = 2n$ and $l = 2n$ the real and imaginary parts of the corresponding structure factors are proportional to :

$$A = \cos(2\pi hx)\cos(2\pi ky)\cos(2\pi lz) \approx \cos(\pi h/4)\cos(\pi k/5)\cos(\pi l/2)$$

$$B = -\sin(2\pi hx)\sin(2\pi ky)\sin(2\pi lz) \approx \sin(\pi h/4)\sin(\pi k/5)\sin(\pi l/2)$$

Clearly, if $l = 2n$ then $\sin(\pi l/2) = 0$ and $B = 0$. Also, for $h = 2n+2$, $\cos(\pi h/4) = 0$ and $A = 0$. In other words, these reflections have phases 0 or π except when $h = 2n + 2$, in which case $F = 0$. Because the z coordinate deviates significantly from the value 0.25, some of the reflections for which $A = 0$ will have $B \neq 0$ and their phases will be $\pm\pi/2$.

It was not a surprise to find that a 20-12Å difference Fourier synthesis for the Niobium derivative which was calculated using phases from spheres of constant density at the assumed molecular centre, could not be interpreted in terms of a convincing heavy atom structure.

Chapter 6

Electron Microscopy of AhrC Crystals

6.1 Introduction.

Despite our best efforts, the determination of the heavy atom structures of the most promising AhrC derivatives proved impossible. Nonetheless, it was still hoped that the presence of a six-fold redundancy in the observed data would be powerful enough to allow the refinement and extension of an initial, low resolution phase set to a point where difference Fourier methods could be used to determine the positions of the heavy atoms ($\approx 8\text{\AA}$ for the $\text{Nb}_6\text{Cl}_{14}$ derivative). As already discussed (Chapter 4), the major problems in this approach are the determination of (i) the position of the molecular centre, (ii) the orientation of the non-crystallographic symmetry axes, and, (iii) an initial, low resolution, phase set. The main evidence supporting the model of the crystal packing presented in Section 4.3, comes from the pseudo-origin peaks seen in the isomorphous difference Patterson map for the Niobium derivative. Although the agreement with other, independent, methods suggested that this model was probably correct, the possibility that the pseudo-origin peaks arise from non-crystallographic symmetry

axes relating heavy atom sites bound on different (crystallographically related) molecules could not be excluded with certainty, in which case all conclusions drawn from this model would be incorrect. Clearly, an experimental demonstration of its correctness (or otherwise) would be very welcome. The determination of the orientation of the non-crystallographic symmetry axes and of a low resolution phase set can be combined. If a low resolution structure is available, the orientation of the intramolecular symmetry axes (with respect to the density of an isolated molecule) can be determined. The problem is then reduced to that of correctly orienting the low resolution model in the given crystallographic frame (assuming that the position of the molecular centre is known).

Electron microscopic image reconstruction methods are suitable both for confirming the model of crystal packing and for obtaining a low resolution image of the structure of AhrC. The theory and practice of image reconstruction is very well documented and will not be discussed in detail (Theory and Reviews : Klug, A. & Berger, J.E., 1964, Klug, A. & De Rosier, D.J. 1966, De Rosier, D.J. & Klug, A., 1968, Crowther, R.A., De Rosier, D.J. & Klug, A., 1970, Crowther, R.A., Amos, L.A., Finch, J.T., De Rosier, D.J. & Klug, A., 1970, De Rosier, D.J. & Moore, P.B., 1970, Crowther, R.A. & Amos, L.A., 1971, De Rosier, D.J., 1971, Erickson, H.P. & Klug, A., 1971, Klug, A. & Crowther, R.A., 1972, Unwin, P.N.T., 1974, Crowther, R.A. & Klug, A., 1975, Misell, D.L., 1978, Klug, A., 1983, Stewart, M., 1988a, 1988b, Morgan, D.G. & De Rosier, D., 1992, Applications : Finch, J.T., Klug, A., Stretton, A.O.W., 1964, Williams, R.C. & Fisher, H.W., 1970, Finch, J.T. & Klug, A., 1971, Mellema, J.E. & Klug, A., 1972, De Rosier, D.J. & Klug, A., 1972, Amos, L.A. & Klug, A., 1974, Unwin, P.N.T., 1975, Unwin, P.N.T. & Henderson, R., 1975, Henderson, R. & Unwin, P.N.T., 1975, Smith, P.R., Aebi, U., Josephs, R. & Kessel, M., 1976, Finch, J.T., *et al*, & Klug, A., 1977, Klug, A., Rhodes, D., Smith, J., Finch, J.T. & Thomas, J.O., 1980, Hamodrakas, S.J., Margaritis, L.H. & Nixon, P.E., 1982, Jeng, T.-W.

& Chiu, W., 1983, Brisson, A. & Unwin, P.N.T., 1985, Holzenburg, A. *et al*, & Saenger, W., 1987, Toyoshima, C. & Unwin, N., 1988, Jeng, T.-W., Crowther, R.A., Stubbs, G. & Chiu, W., 1989, Toyoshima, C. & Unwin, N., 1990, Stokes, D.L. & Green, N.M., 1990, Ceska, T.A. & Henderson, R., 1990, Valpuesta, J.M., Henderson, R. & Frey, T.G., 1990, Henderson, R., Baldwin, J.M., Ceska, T.A., Zemlin, F., Beckmann, E. & Downing, K.H., 1990, Grant, R.A., Schmid, M.F. & Chiu, W., 1991, Schultz, P. *et al*, & Oudet, P., 1993, Holzenburg, A., *et al*, & Ford, R.C., 1993, Unwin, N., 1993, Kühlbrandt, W., Wang, Da-N. & Fujiyoshi, Y., 1994, , Applications involving thin 3D crystals or a comparison with results obtained from X-ray crystallography : De Rosier, D.J. & Oliver, R.M., 1971, Cohen, C., Caspar, D.L.D., Parry, D.A.D. & Lucas, R.M., 1971, Labaw, L.W. & Davies, D.R., 1972, Finch, J.T., Gilbert, P.F.C., Klug, A. & Leberman, R., 1974, Langer, R., Poppe, C., Schramm, H.J. & Hoppe, W., 1975, McPherson, A. & Rich, A., 1973, Shelley, K. & McPherson, A., 1980, Akey, C.W. & Edelstein, S.J., 1983, Furcinitti, P.S., Oostrum, J. & Burnett, R.M., 1989, Stoops, J.K., *et al*, & Hackert, M.L., 1991, Stewart, P.L., Fuller, S.D. & Burnett, R.M., 1993, Voges, D., *et al*, & Huber, R., 1994, Cheng, R.H., *et al* & Johnson, J.E., 1994).

The two problems that we wish to tackle have different requirements and different techniques are most suitable for solving them. To determine the crystal packing we only need to identify and analyse two, preferably orthogonal, projections of thin three-dimensional AhrC crystals. Since the specimen preparation procedure will involve crushing large crystals, only very small areas will be thin enough to allow further analysis. This excludes the possibility of undertaking electron diffraction experiments for this problem. In addition, the resolution requirements are modest : even 30Å data should be adequate for confirming (or otherwise) the model present in Section 4.3. Negative staining is clearly the method of choice for this problem.

To determine the structure of AhrC to as high resolution as possible, other

methods are needed : examination using cryo-electron microscopic techniques of sufficiently large ($>1 \mu\text{m}^2$) two-dimensional crystals or of specimens exhibiting helical symmetry, can give phase information to a resolution not very much lower than X-ray crystallography (Jeng, T.-W., Crowther, R.A., Stubbs, G. & Chiu, W., 1989, Unwin, N., 1993. Kühlbrandt, W., Wang, Da-N. & Fujiyoshi, Y., 1994), although negative staining of smaller two-dimensional crystals can also prove adequate for our purposes.

Attempts at growing two-dimensional AhrC crystals suitable for electron microscopical studies have been unsuccessful. The majority of the conditions resulted in amorphous precipitation of AhrC, although in some cases, linear aggregates such as those shown in Figure 6.1 have been obtained (these may be related to the trigonal form of AhrC).

The unavailability of two-dimensional AhrC crystals does not imply that a low resolution structure can not be obtained from negatively stained thin three-dimensional crystals : In the case of the orthorhombic form, there are 34 unique reflections within the 30\AA sphere. 91% of these can be measured from only 5 projections ([010] (14 reflections), [001] (5), [100] (3), [110] (6) and [130] (3)). If the phases of these reflections can be determined experimentally, the calculation of a three-dimensional (electron density) map is feasible.

The objective of this piece of work was to identify and analyse projections of thin, negatively stained fragments of orthorhombic AhrC crystals. The phase information obtained from these projections could then be combined with the X-ray amplitudes and the resulting electron density maps could be used to determine the crystal packing and a low resolution structure of the molecule.

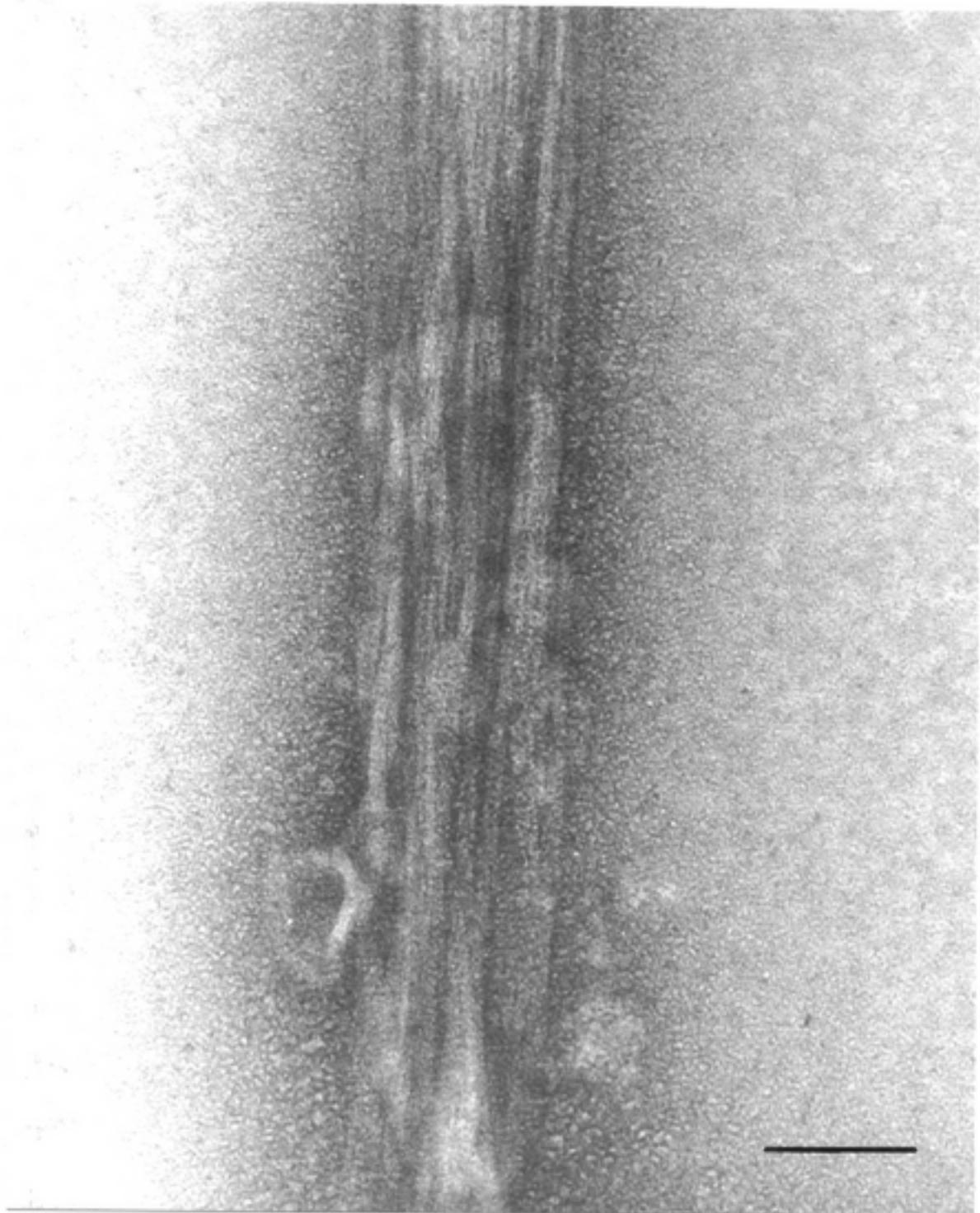


Figure 6.1: Linear aggregates of AhrC. Scale bar 500Å.

6.2 Image Reconstruction.

The discussion in this section will focus only on those aspects of image reconstruction that are relevant to the problem under examination. The fundamental principle of the method is the projection theorem : The Fourier transform of a projection of a three-dimensional object is identical to a section from the Fourier transform of the object which passes through the origin of the transform and is perpendicular to the projection axis.

If the electron microscopic images of the projections of a thin crystal were identical to those projections (that is, if the microscope was a perfect imaging instrument), the experimental procedure would be straightforward : Fourier transformation of the various projections would give both the amplitudes and phases of the corresponding central sections of the three-dimensional transform of the crystal, the complete transform could be built up section by section and the three-dimensional potential density distribution of the crystal could be reconstructed by back-transforming the complete three-dimensional transform.

Erickson, H.P. & Klug, A., 1971 showed that the effect of the various parameters that modify the image of an object (such as spherical aberration, astigmatism and defocus) are best understood in terms of their effect on the Fourier transform of the object rather than the object itself : The Fourier transform (in circular coordinates) of a two-dimensional object with a potential density distribution function $\sigma(x, y)$, is given by

$$\mathcal{T}^o(\alpha/\lambda, \phi) = \iint \sigma(x, y) e^{-2\pi i/\lambda(x\alpha \cos(\phi) + y\alpha \sin(\phi))} dx dy \quad (6.1)$$

where α is the angle of scattering, λ is the electron wavelength and ϕ is the azimuthal coordinate (in reciprocal space). The authors showed that if $\mathcal{T}^i(\alpha, \phi)$ is the Fourier transform of an image of the object, then,

$$\mathcal{T}^i(\alpha, \phi) = -\mathcal{T}^o(\alpha, \phi) f(\alpha) A(\alpha) \sin [\chi(\alpha) + \Phi(\alpha)] \quad (6.2)$$

where $f(\alpha)$ is the atomic scattering factor for electrons, $A(\alpha)$ is a function defining the shape and dimensions of the objective aperture of the microscope and $\sin[\chi(\alpha) + \Phi(\alpha)]$ is the so called contrast transfer function, with

$$\chi(\alpha) = \frac{2\pi}{\lambda} \left[\frac{-C_s \alpha^4}{4} + \frac{\Delta f \alpha^2}{2} \right] \quad (6.3)$$

In the last equation, C_s is the spherical aberration coefficient and Δf is the amount of defocusing. (This formulation of $\chi(\alpha)$ is valid only when the axial astigmatism has been corrected for). $\Phi(\alpha)$ in Equation 6.2 is related (i) to the ratio of the number of electrons scattered outside the objective aperture to those that pass through it, and, (ii) to the fraction of electrons that are inelastically scattered. It is zero if there is no objective aperture and no inelastic scattering and its value increases with decreasing radius of the aperture or increasing atomic number of the atoms that constitute the specimen.

Equation 6.2 shows that the Fourier transform of the image is proportional to the Fourier transform of the object modulated by three factors all of which are real numbers, and thus, affect only the amplitude of the transform. Of these factors, the effects of the transfer function $\sin[\chi(\alpha) + \Phi(\alpha)]$ on the transform of the object will be discussed in some detail. Figure 6.2 shows the variation of the values of the transfer function versus reciprocal resolution (in nm^{-1}) for different values of Δf and $\Phi(\alpha)$, and assuming that $\lambda = 0.042\text{\AA}$ and $C_s = 1.3\text{mm}$. For $\Delta f = 0\text{ nm}$ and $\Phi(\alpha) = 0\text{ rad}$ (plot (A)), the transfer function has very small values for the region from 0 to 0.5 nm^{-1} , which means that all low resolution terms with $d_{min} > 20\text{\AA}$ do not contribute significantly to the image. The function reaches a maximum at about 1.8 nm^{-1} ($d_{min} = 5.5\text{\AA}$) and these terms will contribute to the image with maximum positive contrast. At about 4.6\AA resolution (2.15 nm^{-1}) the function goes to zero (and corresponding parts of the transform do not contribute to the image) and then changes sign. The sign reversal means that the phases of all terms between 4.6 and 3.9\AA resolution (when it reaches again zero) are shifted by 180 degrees. The physical interpretation of this phase shift is that corresponding

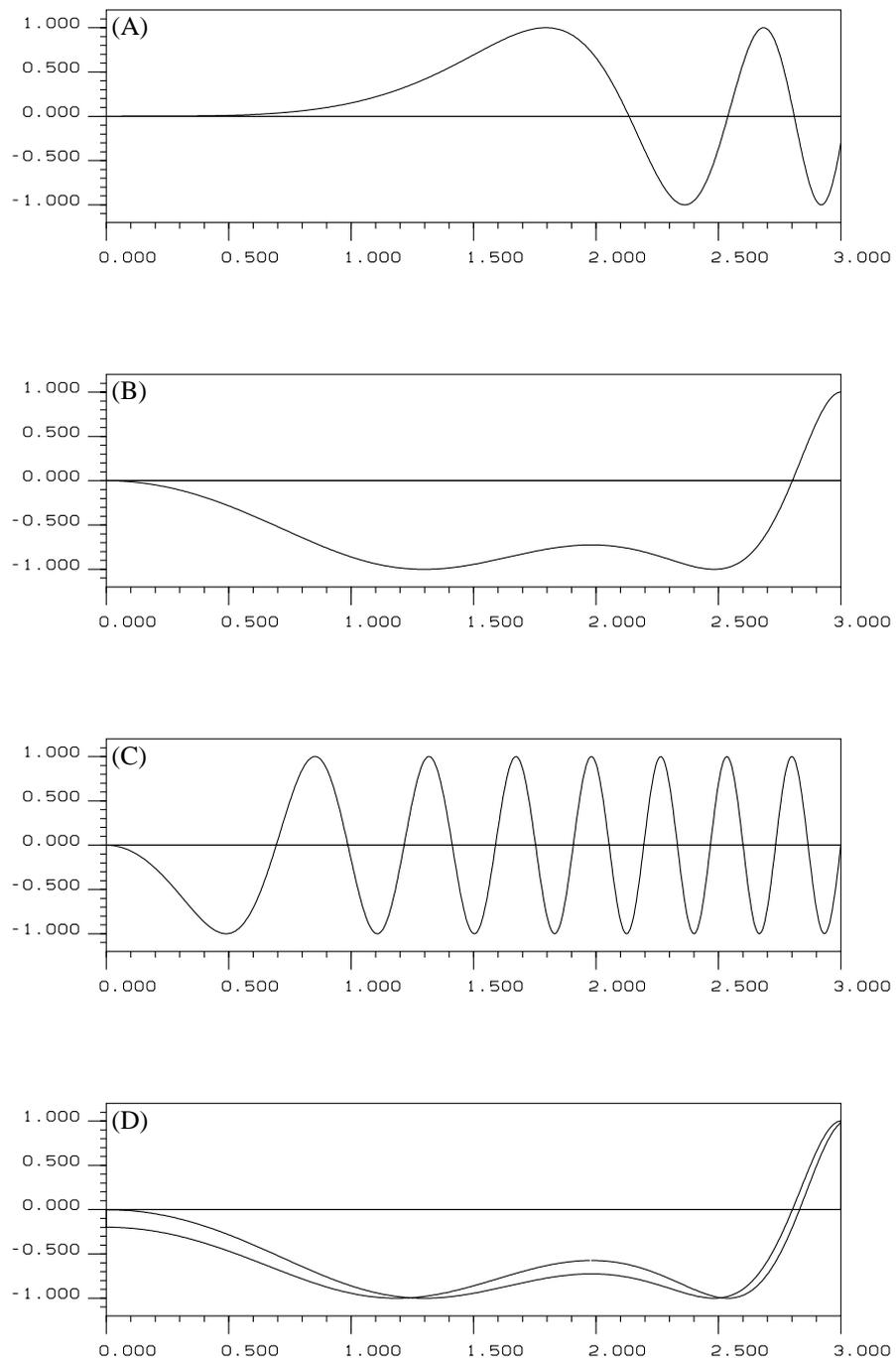


Figure 6.2: Variation of the transfer function versus resolution (in nm^{-1}) for different values of Δf and $\Phi(\alpha)$: (A) 0 nm, 0 rad (B) 90,0 (C) 500,0 and (D) Superposition of 90,0 and 90,0.2

parts of the transform of the object contribute with reversed contrast to the image (areas that should be white will appear black). For $\Delta f = 90$ nm, $\Phi(\alpha) = 0$ (plot (B)), low resolution terms again do not contribute significantly to the image, but the function remains close to its optimum value for a wide resolution range (from about 12 to 4.5Å resolution). When $\Delta f = 500$ nm and $\Phi(\alpha) = 0$ (plot (C)) low resolution terms contribute to the image, but at higher angles the function oscillates rapidly between positive and negative contrast. Finally, plot (D) shows the effect which $\Phi(\alpha)$ has on the transfer function. The two superimposed graphs correspond to $\Delta f = 90$ nm, $\Phi(\alpha) = 0.2$ rad and $\Delta f = 90$ nm and $\Phi(\alpha) = 0.0$ rad (identical to (B)). It can be seen that the most important difference is the enhanced contribution of the low resolution terms to the final image¹.

In Figure 6.3, the effects of the transfer function on the images and their transforms are illustrated through a series of images of the [001] projection of a thin, negatively stained AhrC crystal, obtained at different values of Δf . For each image, the modulus of its Fourier transform is also shown. For the in focus image (A) both the transform and the image have a smooth appearance, with most of the reflections very weak or missing. At $\Delta f \approx 640$ nm (Image (B)), all reflections out to 20Å are strong and internal detail is clearly visible in the image. As the value of Δf increases from ≈ 1200 nm (C) to 1900 nm (D) and 3250 nm (E), the phases of some of the higher resolution reflections are reversed and corresponding image details appear white instead of black, which together with the enhanced contribution from low resolution terms, reduces the amount of detail in the corresponding images. The effects of the contrast transfer function on the transform are obvious. For example, the reflection 5,1 is relatively strong in (C), it is missing in (D), and reappears in (E) but with its phase reversed. Similarly, the reflection 0,2 in images (D) and (E) has its phase reversed.

The preceding analysis shows the importance of choosing the amount of un-

¹A constant value for the function $\Phi(\alpha)$ is unrealistic, but it illustrates its effect in the resolution range of interest for this study in a simple and direct manner.

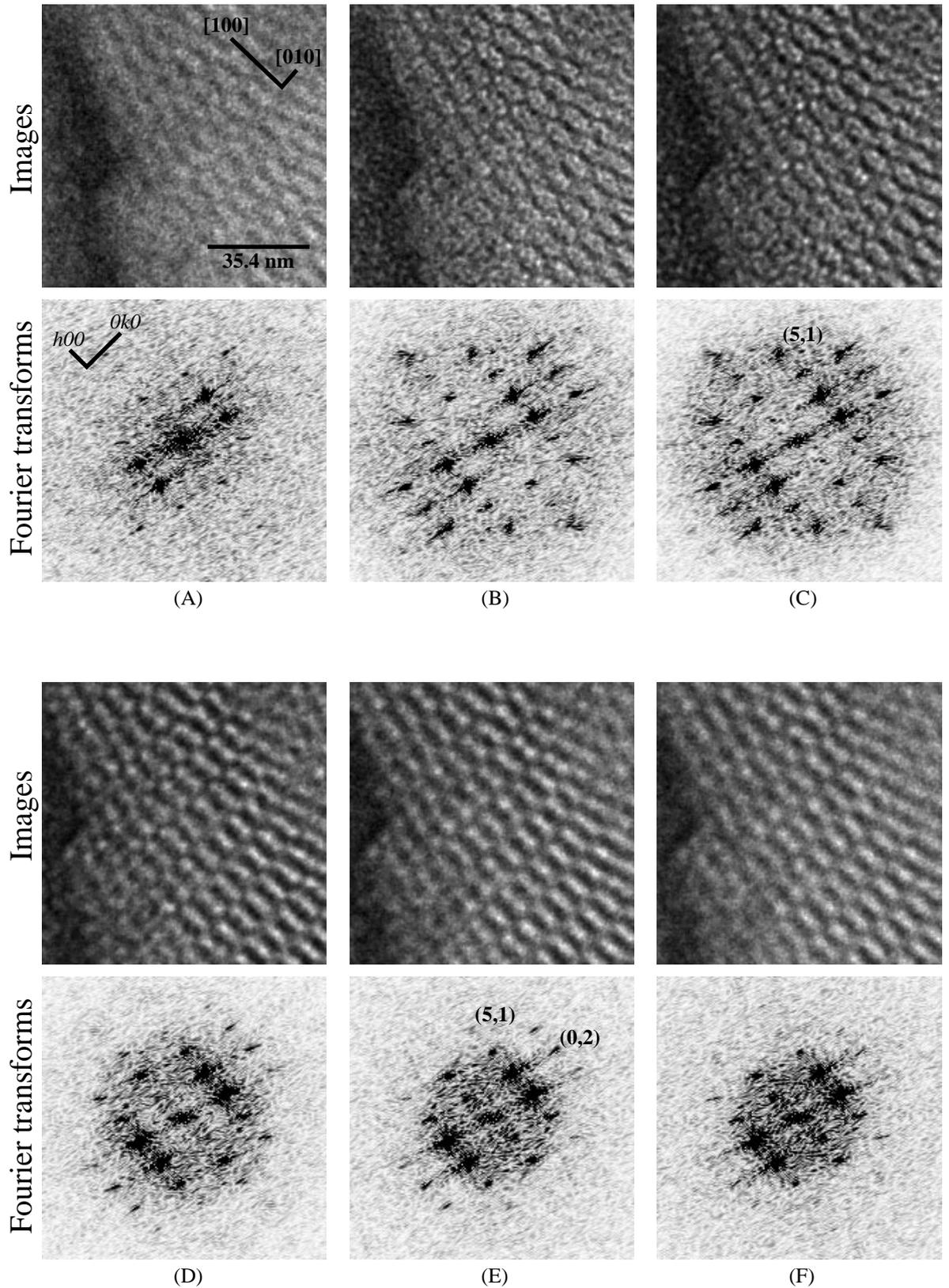


Figure 6.3: Focus series of a thin, negatively stained AhrC crystal down the $[001]$ axis. Approximate defocus values are (A) in focus, (B) 640 nm, (C) 1200 nm, (D) 1900 nm, (E) 3250 nm, (F) > 3500 nm.

derfocus that will optimise the value of the transfer function in the resolution range of interest, but without introducing sign reversals. We will now discuss what the resolution range of interest is in the case of AhrC. Negative staining techniques set an upper resolution limit of about 18Å. Since our intention is to combine the phases determined from image reconstruction with the amplitudes determined from X-ray crystallography, it is important to choose a resolution range for which the phases of all terms are determined by the contrast between protein and solvent of crystallisation and not by the internal structure of the protein (Bragg, W.L. & Perutz, M.F., 1952). A safe estimate would be a high resolution cutoff equal to the longest dimension of the largest domain of AhrC. A value of 25Å is probably an overcautious estimate, giving an optimum Δf at about 650 nm².

The availability of amplitudes determined from X-ray crystallography makes the whole procedure of image analysis much simpler. As long as the first zero of the transfer function is outside the resolution range of interest, image reconstruction can proceed as if the microscope was a perfect imaging instrument : no corrections for the effects of the transfer function are necessary.

6.3 Specimen Preparation, Data Collection and Image Processing.

AhrC crystals were crushed in a stabilising solution consisting of 10% MPD and 100 mM acetate buffer at pH=4.9. A droplet of the solution containing the crystal fragments was transferred to a carbon-coated grid and allowed to stand

²It is worth noting that the difference in the contrast between protein and solvent in the crystals and between stain and stain-excluding regions in the micrographs, affects only the amplitudes of the reflections and not their phases, and so, does not invalidate the process of combining X-ray amplitudes with electron microscopy phases. On the other hand, positive staining is a potentially serious problem. The assumption will be made that the amount of positive staining is so small that it can not reverse the sign of the strong reflections.

for approximately 1 min. The fragments were negatively stained with a 2% (w/v) solution of uranyl acetate in water.

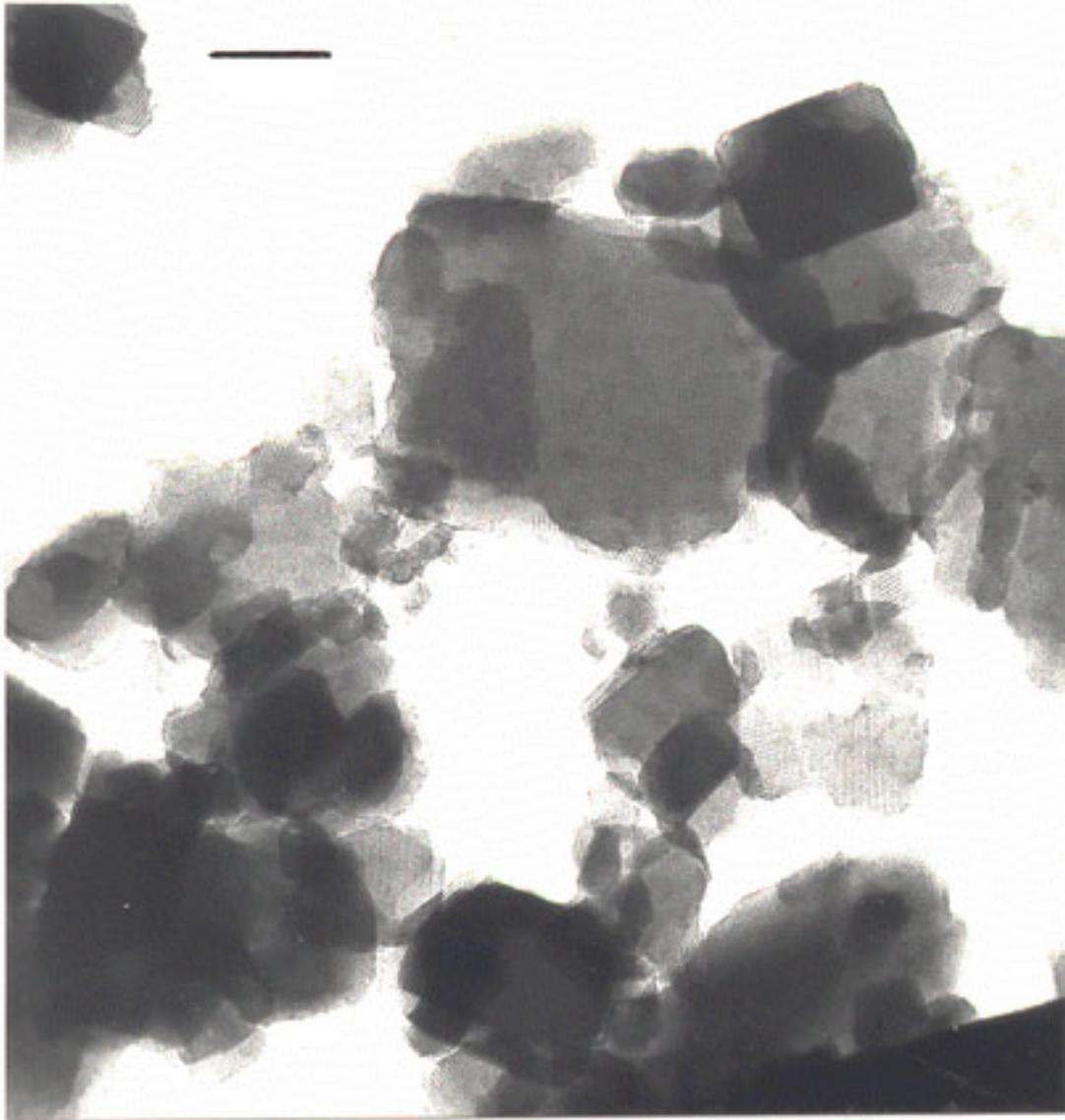


Figure 6.4: Electron micrograph of crushed, negatively stained AhrC crystals. Scale bar 2000Å.

Grids were examined in a Philips CM-10 transmission electron microscope operating at 100 kV. Images were recorded on Agfa Scientia 23 D 56 electron-image sheet film, or later, with a 512×512 Charge Coupled Device fitted on the same microscope. Low dose imaging conditions have not been used. Figure 6.4 is a rather rare example of a micrograph showing several different projections of the orthorhombic form of AhrC. The best micrographs were identified

by optical diffraction using a He/Ne LASER-powered optical diffractometer and digitised using a Joyce-Lobel microdensitometer with a scan step of 25 μm . Further processing of CCD images or digitised micrographs was performed using the MRC/CCP4 suite of programmes, as follows.

The image was displayed using the programme DSLOAD, and a well preserved area was extracted using the programme BOXIM. This was transformed using FFT and the transform was displayed and masked using DSLOAD. The masked transform was back-transformed and the filtered image examined with DSLOAD. If the projection could be identified (from the transform, the filtered image or through an automatic procedure described below), the reciprocal lattice parameters were refined (programme MMLATREF) and the amplitudes and phases of the observed reflections were extracted (MMBOX). A programme was written which determined the phase shifts that must be applied in order to move the origin of the image to a permissible for the given plane group position. The value of two useful phase residuals could now be calculated : The first residual is applicable to centrosymmetric terms only, and equals the mean phase difference between the observed phase angles and the values expected from symmetry considerations. The second residual is the mean phase difference between symmetry related reflections. In the final step, the phase angles were set to their expected values, they were combined with the corresponding X-ray amplitudes and the electron density map was calculated.

A programme was written to help with the identification of projections that could not be recognised by inspection. The programme generated the coordinates (in reciprocal space) of all reflections of the orthorhombic form within the 30 \AA sphere and calculated the agreement between the observed reciprocal lattice and each of the unique reciprocal lattice planes (within the 30 \AA sphere). At the end of the procedure, a user-defined number of "best" solutions was written out for further examination.

6.4 Characterised Projections.

6.4.1 The [001] Projection.

The great majority of well preserved areas examined, belonged to this projection. This is consistent with the presence of protein layers parallel to the xy plane (Section 4.3). Most of the [001] images analysed showed reflections out to about 25\AA , although in one case, periodicities of the order of 18\AA could be detected (Figure 6.5). A typical CCD image of this projection, recorded at ≈ 640 nm underfocus, is shown in Figure 6.6. Figure 6.7(A) is the modulus of its Fourier transform and Figure 6.7(B) shows the phases and amplitudes of individual reflections. It can be seen that most of the strongest reflections have phases very close to the expected values (0 or π), and the phases of symmetry related reflections are similar. Within the 25\AA sphere, the value of R_{symm} for six pairs of symmetry related reflections is 20.2%, the mean phase difference from the expected phase angles (0 or π) for 18 observed reflections is 23.0° and the mean phase difference between symmetry related reflections (six pairs) is 26.5° (the expected values of these two phase residuals for a random distribution of phase angles are 45° and 90° respectively). Figure 6.8 shows the variation of the mean phase difference between the observed phase angles and those expected from symmetry considerations as the origin is moved systematically over the entire unit cell. The four peaks are equivalent by crystallographic symmetry and they correspond to a mean phase difference between observed and expected phase angles of 14.6° (only the 11 strongest reflections have been used for this calculation).

Figure 6.9(A) shows the 25\AA $F_{X\text{-rays}} \exp(i\phi_{\text{EM}})$ synthesis for this projection and Figure 6.9(B) is the same map but with its contrast reversed. This should be compared with Figure 6.9(C) which is a magnified area from the filtered image. Given that map (C) is based on EM amplitudes which have not been corrected for the effect of the transfer function and EM phases that have not been set to

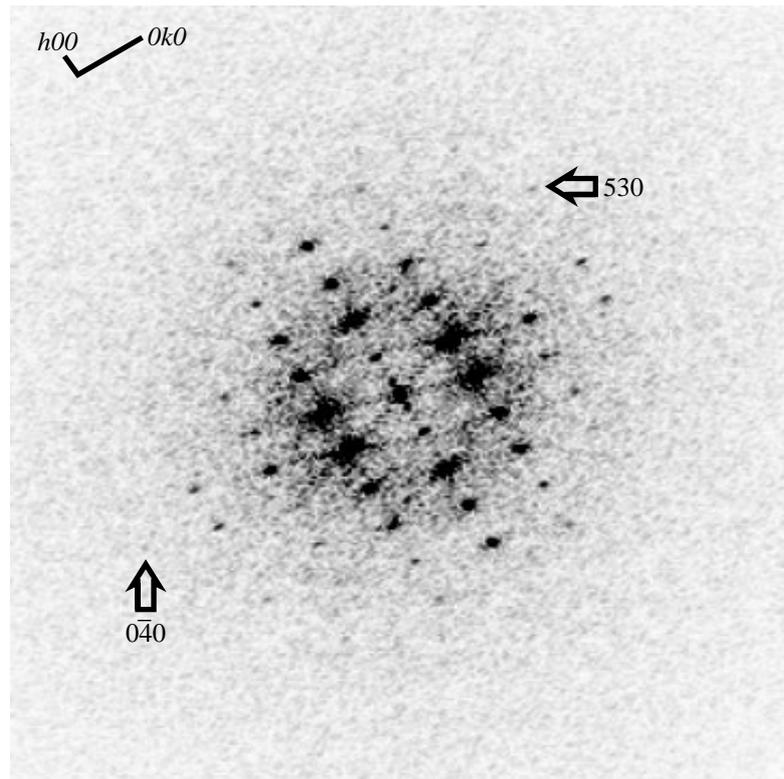


Figure 6.5: Modulus of the Fourier transform of a well preserved image of the [001] projection. The positions of two weak but observed reflections corresponding to spacings of 21.5 (530) and 18.3Å ($0\bar{4}0$) are indicated.

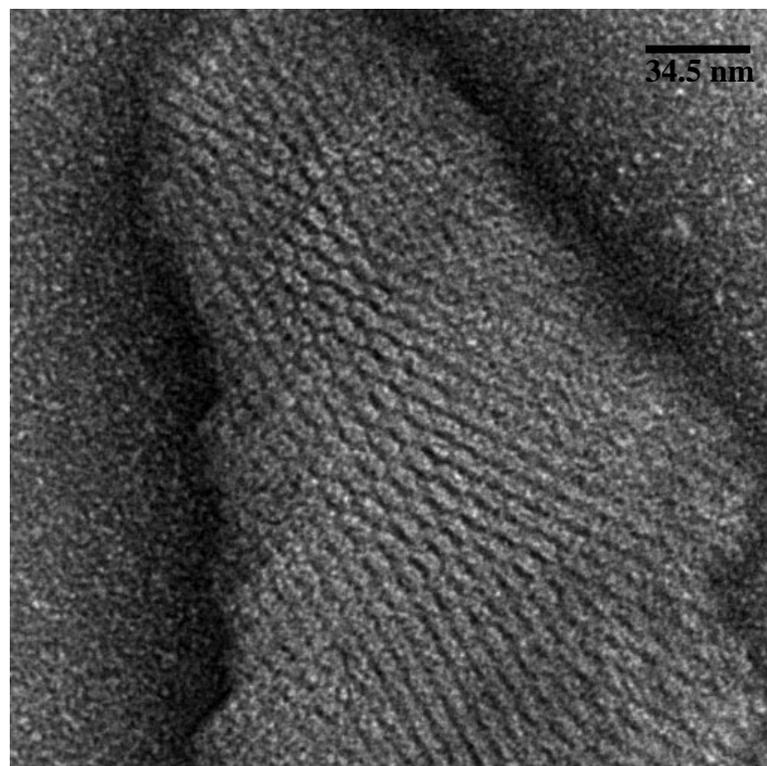
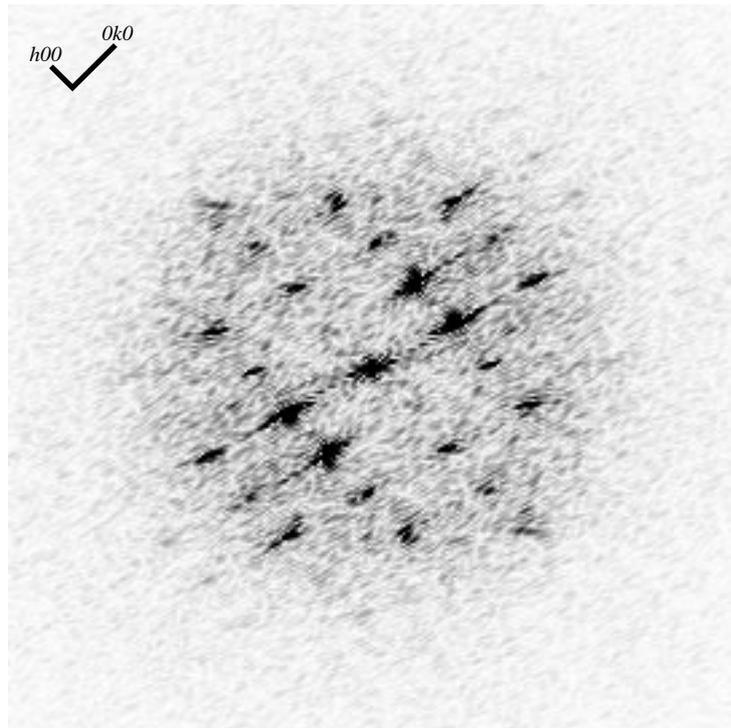
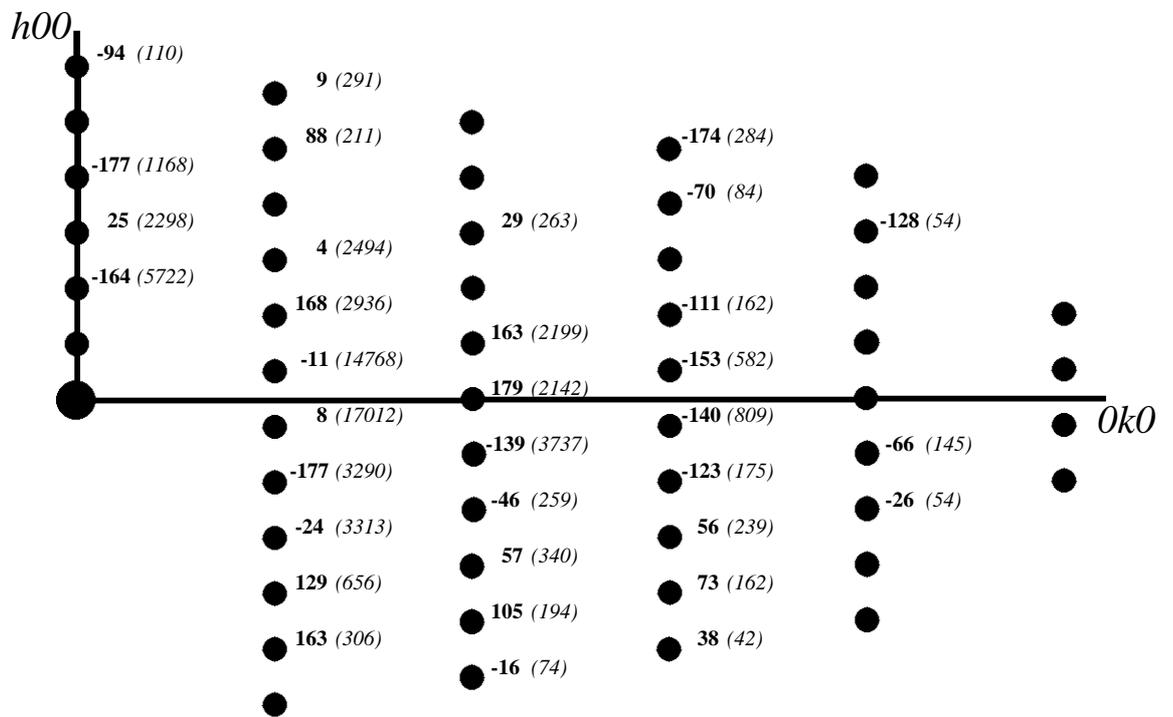


Figure 6.6: CCD image of the [001] projection recorded at $\Delta f \approx 640$ nm.



(A)



(B)

Figure 6.7: (A) Modulus of the Fourier transform of an image of the [001] projection, and, (B) The **phases** (in degrees) and *amplitudes* (arbitrary units) of all observed reflections.

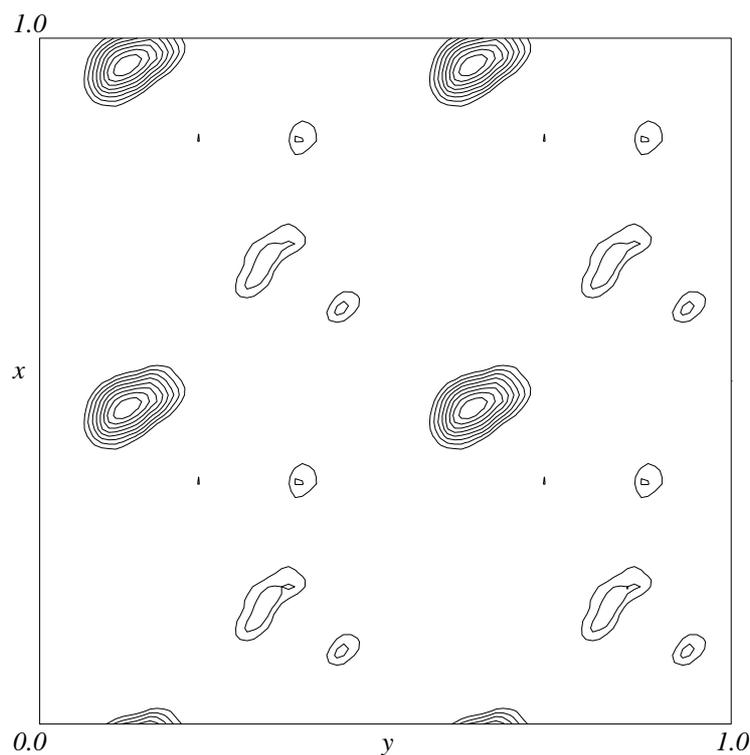


Figure 6.8: Results from the origin search using the 11 strongest $hk0$ reflections. Contours every 0.5σ with first contour at 0.5σ above the mean (45°).

their expected values, the agreement between these two maps is unexpectedly good.

Two points about this projection are worth discussing. The first is its consistency with the model of the crystal packing : The density corresponding to two overlapping AhrC hexamers is organised in three interconnected high-density areas (marked as I, II and III in Figure 6.9(B)), which surround a central, low density region. The x coordinate of the centre of gravity of these three domains is 0.13, in very good agreement with the previously obtained value. Due to the overlap, the y coordinate can not be determined accurately, but, the diameter of the solvent channels suggests that its value must be close to zero, as expected. The second point to note is that the distribution of density gives no hints about the orientation of the non-crystallographic symmetry axes. This is not surprising, given that this is a projection down a 138\AA long axis.

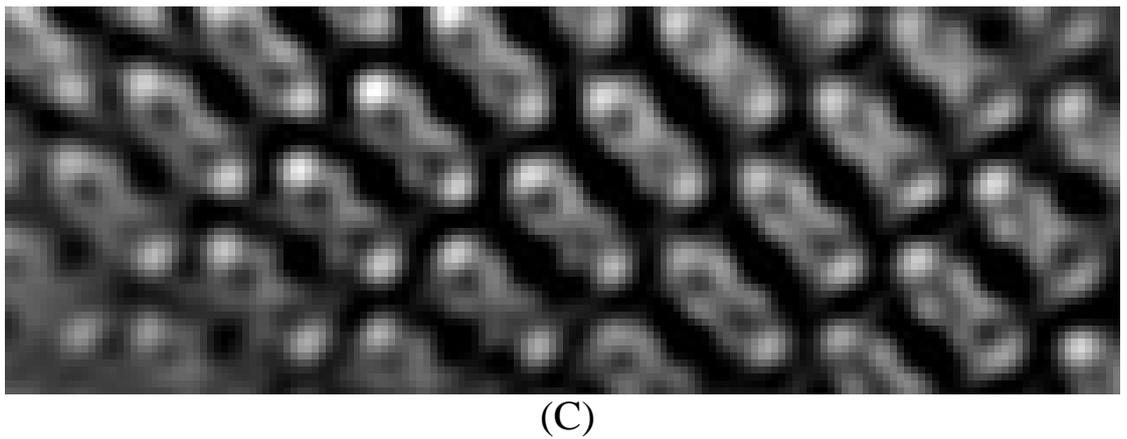
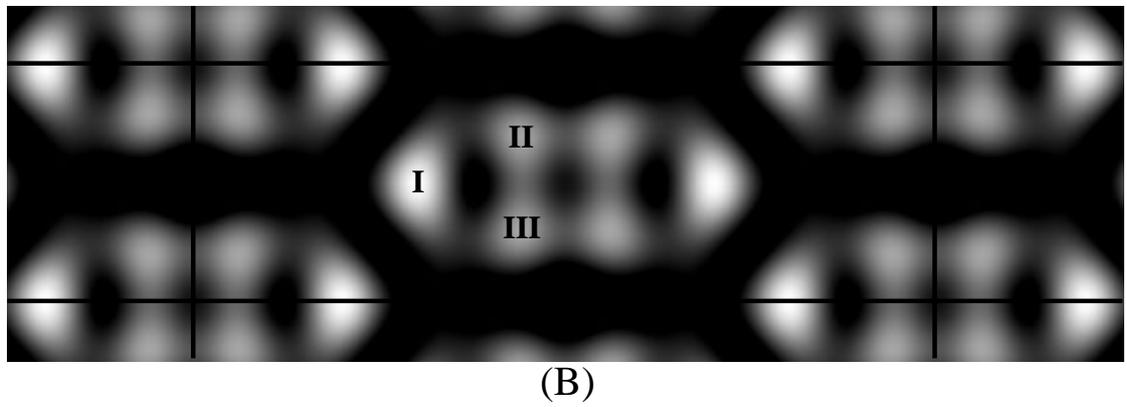
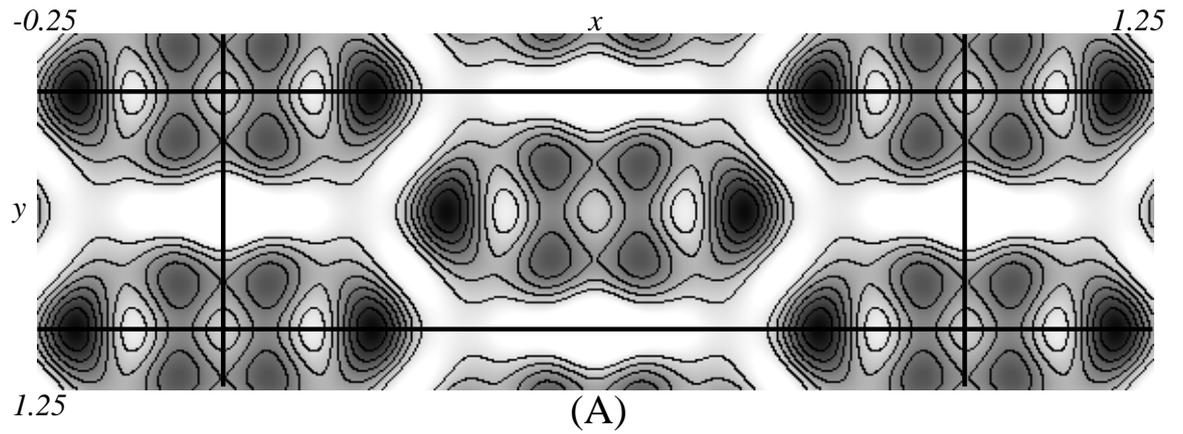


Figure 6.9: (A) and (B) : The [001] projection at 25Å resolution using X-ray amplitudes and EM phases, (C) : Magnified area from the filtered image.

6.4.2 The [010] Projection.

This projection is not as common as the [001] and only a handful of images have been obtained. Of these, only one image gave a Fourier transform which showed reflections extending beyond the first or second order. The image is shown in Figure 6.10. Figure 6.11(A) is the modulus of its Fourier transform, and in (B), the phases and amplitudes of individual reflections are shown. The phase relationships for this projection are more complex : For $l = 2n + 1$ the phases should be $\pm\pi/2$ with $F_{hol} = -F_{h0\bar{l}}$. When $l = 2n$ the phases are expected to be 0 or π with $F_{hol} = F_{h0\bar{l}}$. Within the 25\AA sphere, the R_{symm} for 13 pairs of reflections is 14.0%, the mean phase difference between the observed phase angles and the values expected from symmetry considerations is 21.5° (31 reflections) and, finally, the mean phase difference for 13 pairs of symmetry related reflections is 35.7° .

Figure 6.12(A) is the 25\AA $F_{X\text{-rays}} \exp(i\phi_{EM})$ synthesis for the [010] projection and Figure 6.12(B) is the same map but with its contrast reversed. Figure 6.12(C) is a magnified area from the filtered image. Maps (B) and (C) are again very similar. The agreement between map (B) and the distribution of density in the two-cell thick area of the original image (Figure 6.10), is worth noting.

The experimentally determined electron density of the [010] projection is again consistent with the model of the crystal packing : The density corresponding to a single AhrC hexamer is organised in three high-density areas, marked as I, II and III in Figure 6.12(B) and a low density extension, marked as IV. Domains I, II and IV are well connected, but the connectivity of area III is not obvious. The z coordinate of the centre of gravity of domains I, II and III is very close to the expected value of 0.22. It should be noted, however, that the position of the geometrical centre is significantly different with $z \approx 0.17$. Although this projection is only one molecule thick, it is again difficult to interpret this electron density map in terms of a projection of a hexamer with 32 point group symmetry.

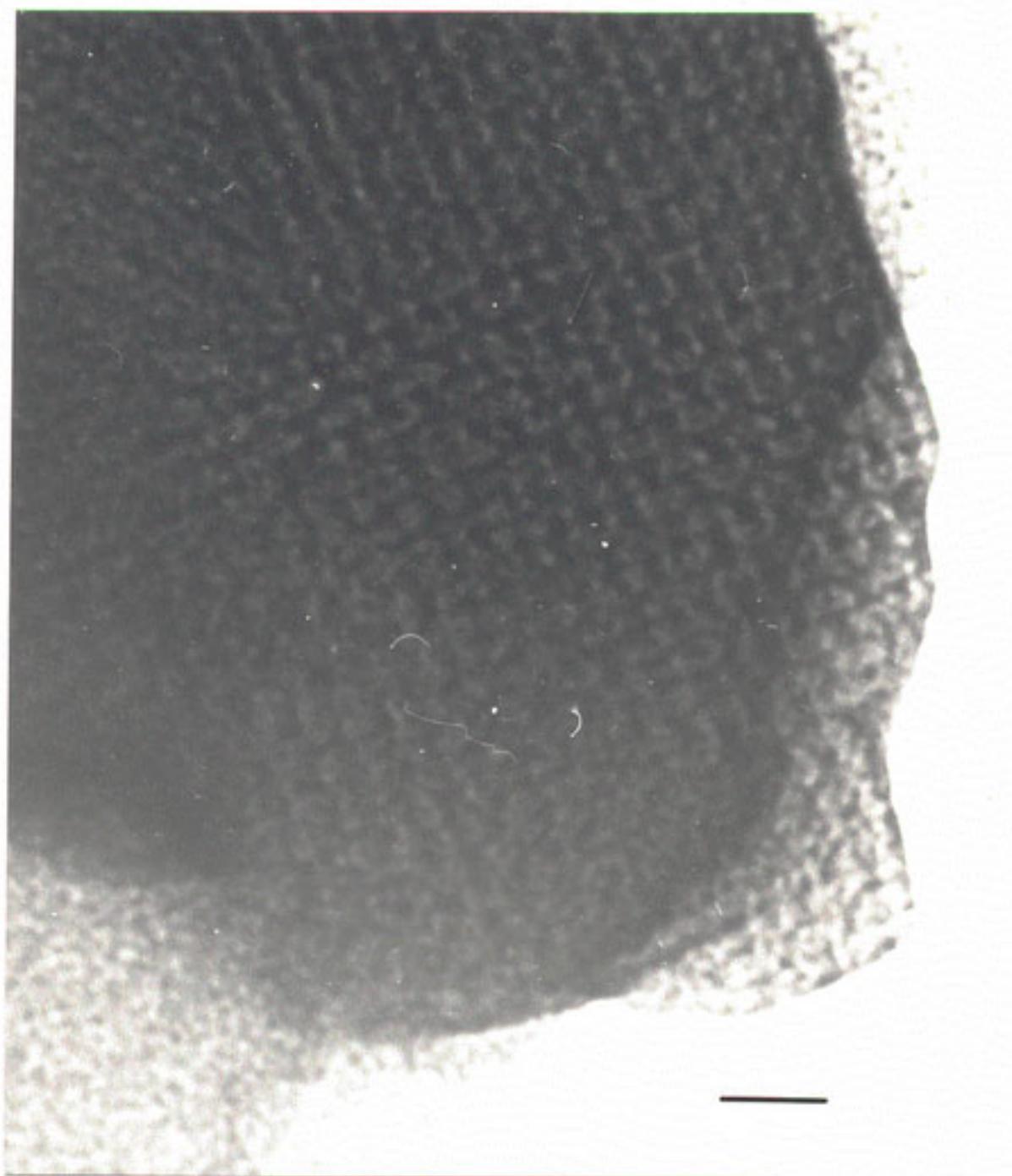
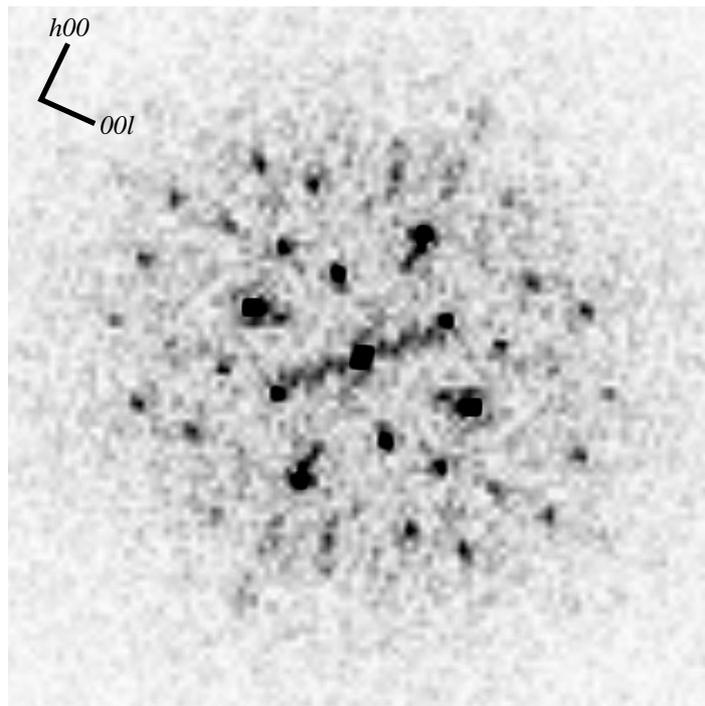
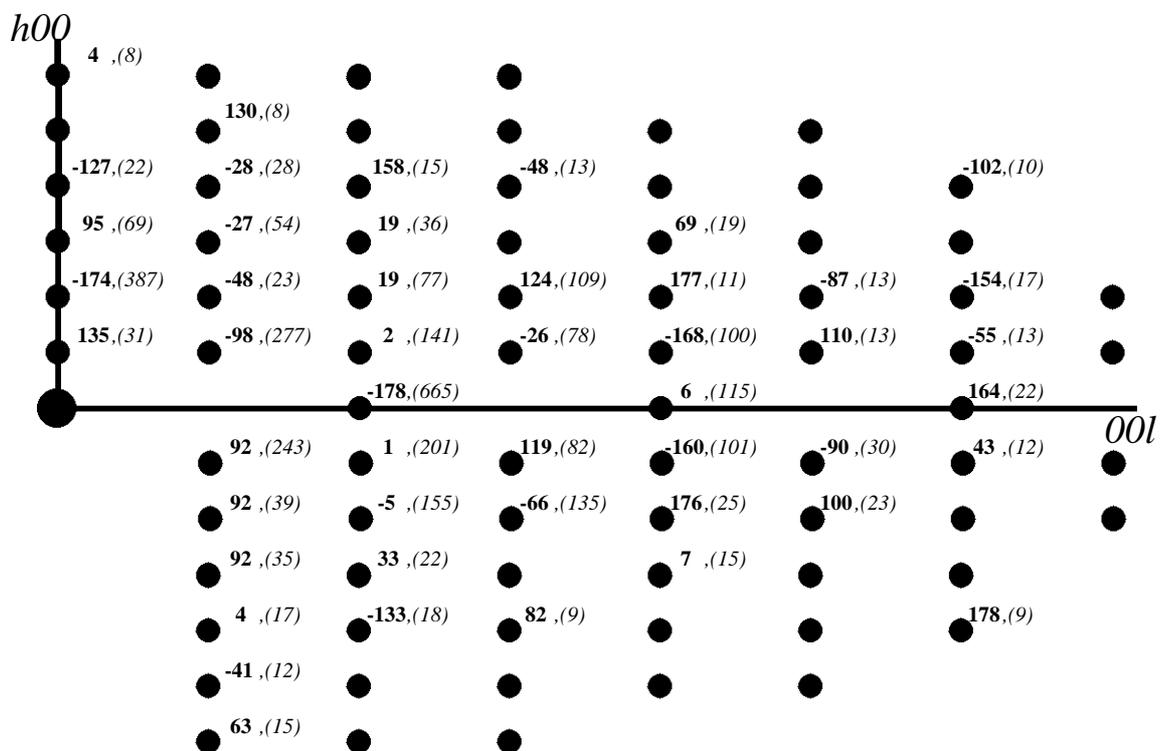


Figure 6.10: Electron micrograph of the [010] projection. Note the presence of one- and two-cell thick areas near the bottom right-hand corner of the crystal. Scale bar 200Å.



(A)



(B)

Figure 6.11: (A) Modulus of the Fourier transform of the image shown in Figure 6.10, and, (B) The phases and amplitudes of all observed reflections.

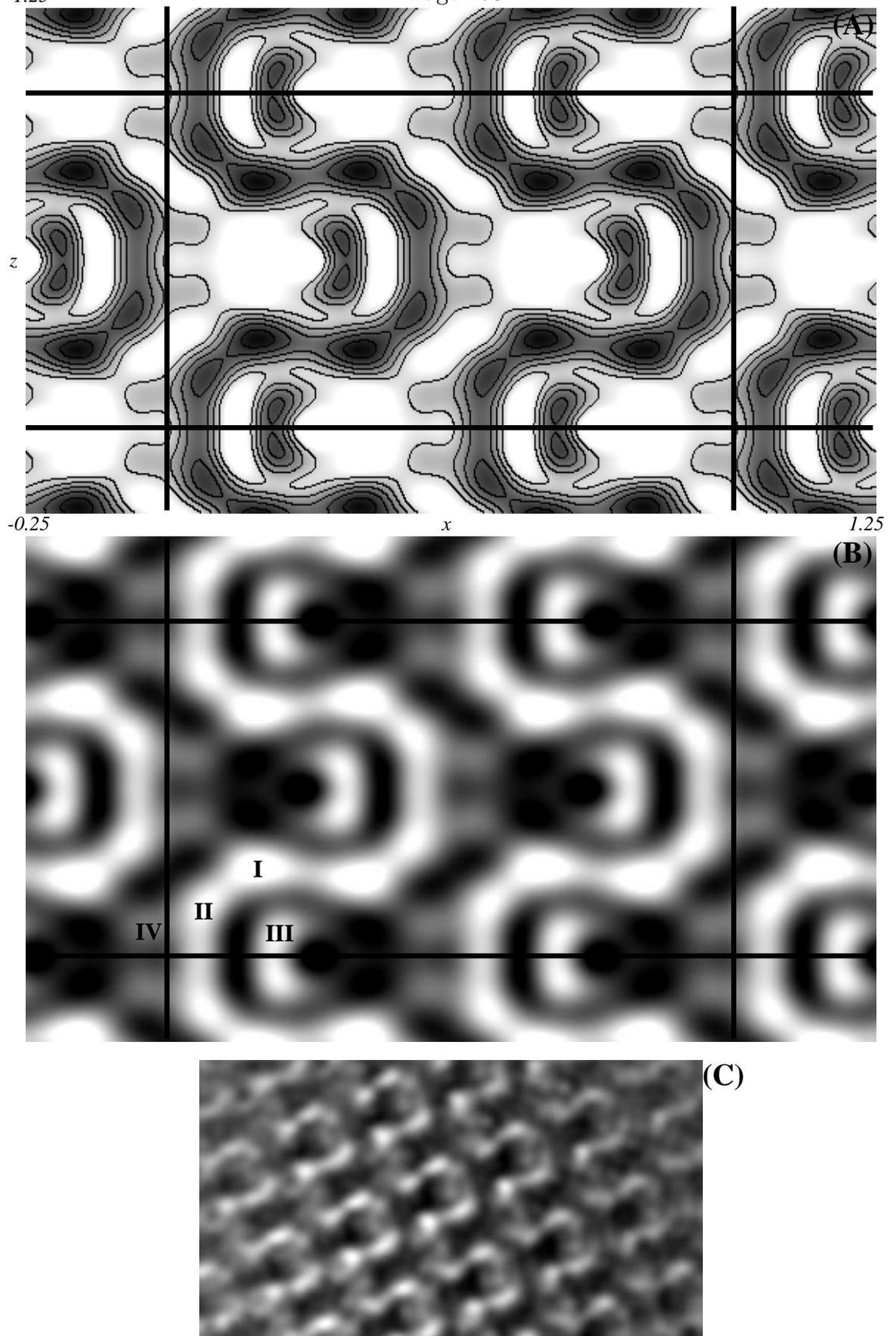


Figure 6.12: (A) and (B) : The [010] projection at 25Å resolution using X-ray amplitudes and EM phases, (C) : Magnified area from the filtered image.

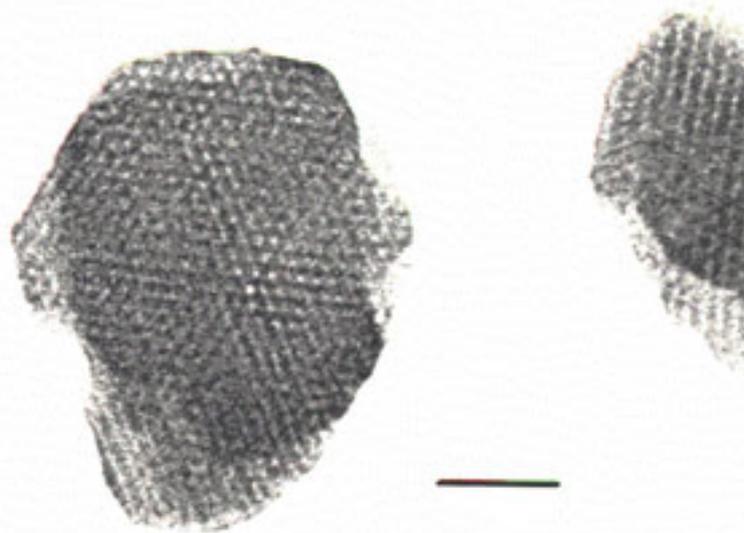


Figure 6.13: Electron micrograph of the [101] projection. Scale bar 500Å.

6.4.3 The [101] Projection.

The image is shown in Figure 6.13. The presence of an approximately regular hexagonal lattice is obvious. Figure 6.14(A) is the modulus of its Fourier transform and (B) is the filtered image. Unfortunately, the order is preserved to only 7 nm resolution and little further information can be obtained. This projection could not be identified by inspection and the automatic procedure described in Section 6.3 was employed. The reciprocal lattice plane corresponding to the [101] projection gave a relatively good agreement with the observed reciprocal lattice and further investigation suggested that this assignment is probably correct : Figure 6.14(C) shows the projection of the model of the crystal packing down the [101] axis (each sphere in this diagram represents an AhrC hexamer). The molecules form columns with their axes parallel to the direction of viewing. The approximately hexagonal packing of those columns is indicated. A 4° precession photograph of that zone is shown in (D), and (E) is the corresponding intensity-weighted reciprocal lattice representation. The arrangement and intensity of the 202 and 111 reflections is consistent with the proposed assignment.

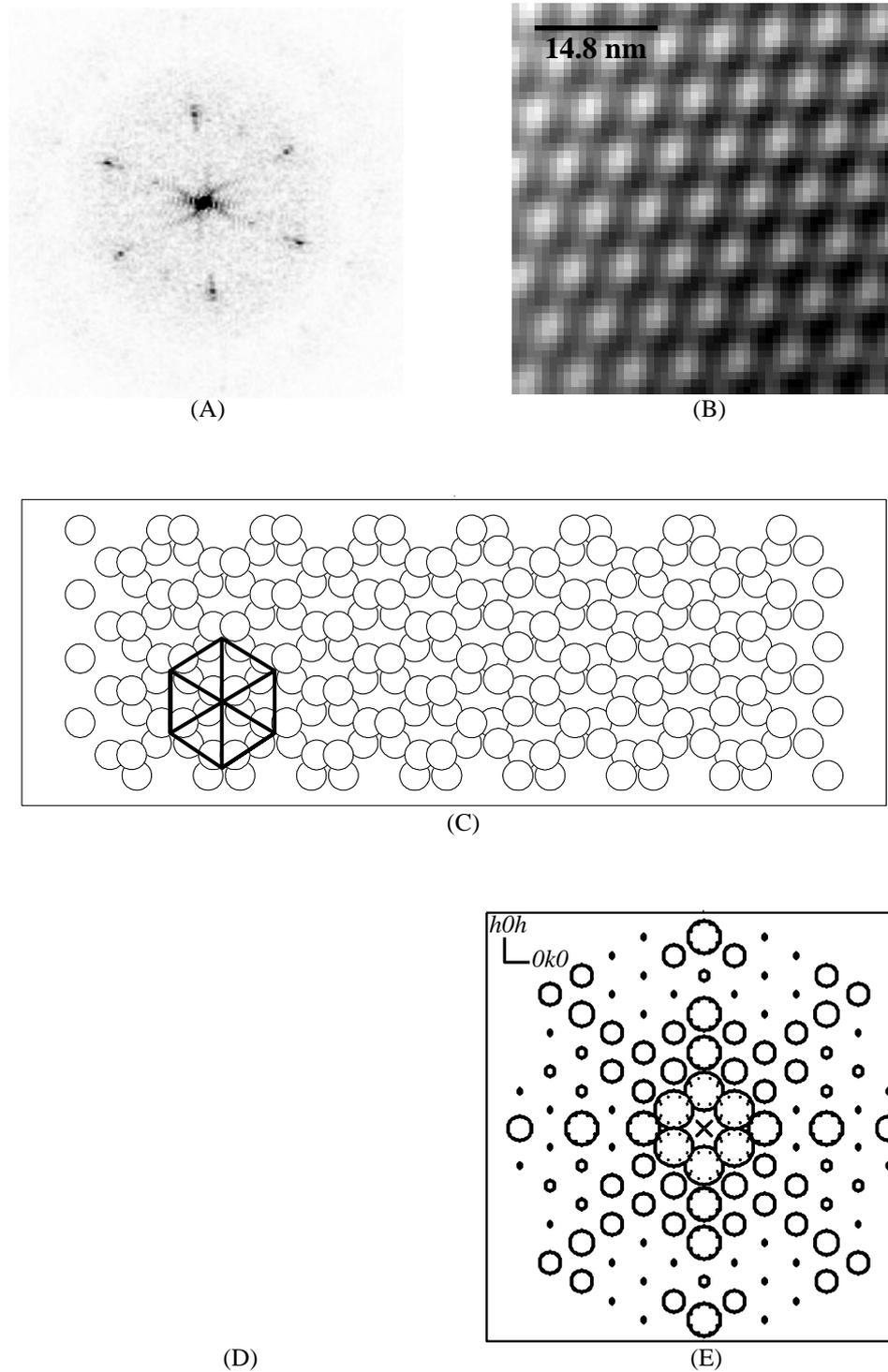


Figure 6.14: [101] projection : (A) Modulus of the Fourier transform, (B) Filtered image, (C) View of the model of the crystal packing, (D) 4° precession photograph of the hkh plane, and, (E) The intensity-weighted reciprocal lattice.

6.4.4 The [130] Projection.

This projection was again identified through the automatic procedure described in Section 6.3. Figure 6.15(A) shows the modulus of the Fourier transform of a selected area and (B) is the corresponding filtered image. The order is not well preserved : the highest resolution reflection is 311, corresponding to $d_{min} = 49\text{\AA}$. Figure 6.15(C) is a view of the model of the crystal packing down the [130] axis and, (D) is a tone representation of the 50\AA $F_{X\text{-rays}} \exp(i\phi_{EM})$ synthesis for this projection.

6.5 Other Projections.

It has not been possible to index the Fourier transforms of all relatively well preserved images. Figure 6.16 shows two examples of transforms that resisted all our attempts to index them, together with the corresponding filtered images. One characteristic common to all these projections is that each has been observed only once, which suggests that they may correspond to “rare events” of the specimen preparation procedure. The electron micrograph in Figure 6.17 provides an example of such an event : different areas of a seemingly uniform crystal have striking differences both in their appearance and in the observed periodicities. The simplest explanation is that the thickness of the corresponding areas is not the same and that their difference is less than one unit cell along the projection axis (adding one unit cell-thick layers should leave the projection unchanged).

Other possible problems arising from the preparation procedure are : (i) specimens that are less than one unit cell thick along the projection axis, (ii) uniformly distorted crystalline areas (usually through bending), (iii) one unit cell-thick crystals that are tilted with respect to the electron beam, etc. All images whose transform could not be indexed in terms of a projection of the three-dimensional crystals, have been discarded.

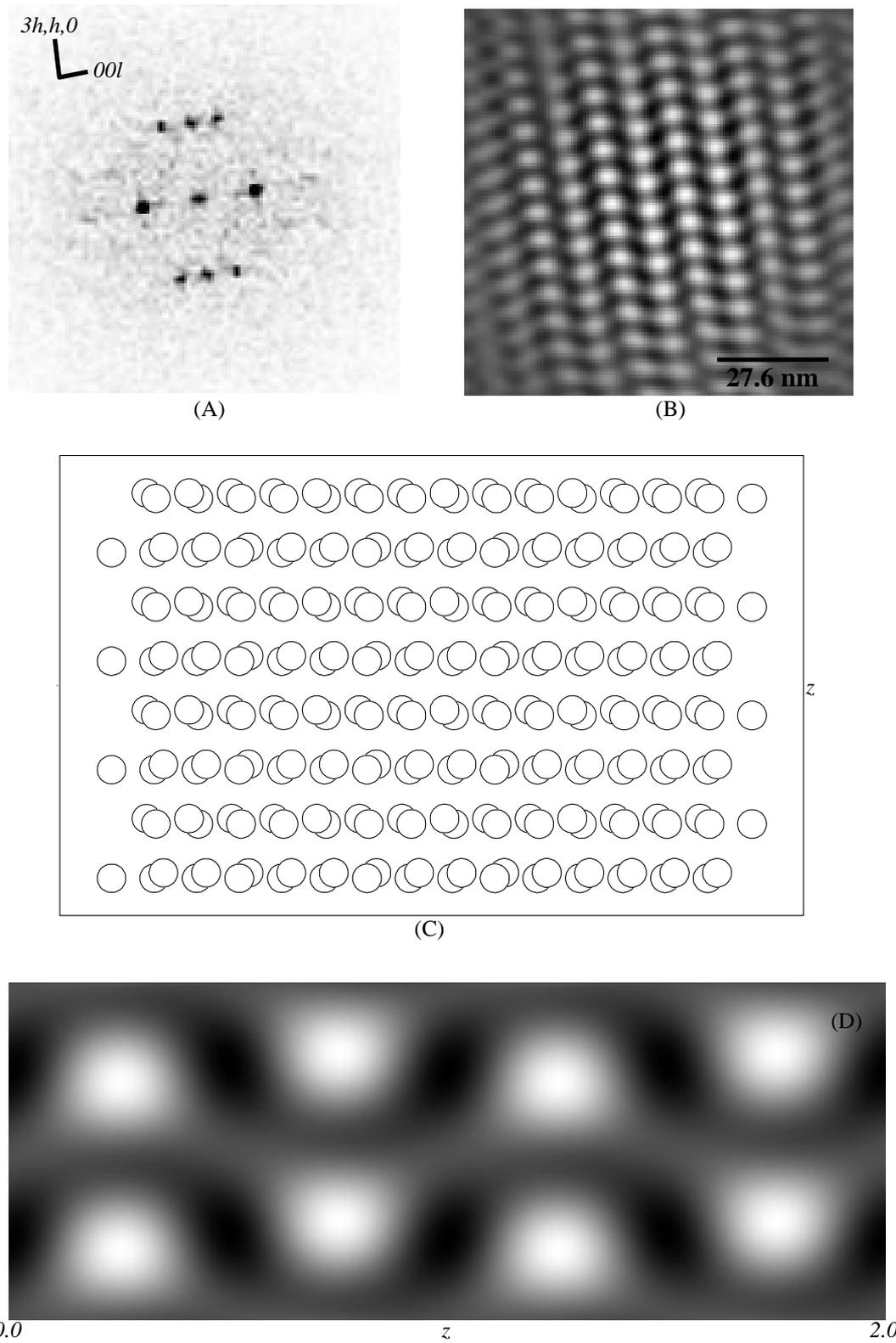


Figure 6.15: [130] projection : (A) Modulus of the Fourier transform, (B) Filtered image, (C) View of the model of the crystal packing, (D) 50 \AA synthesis using X-ray amplitudes and EM phases.

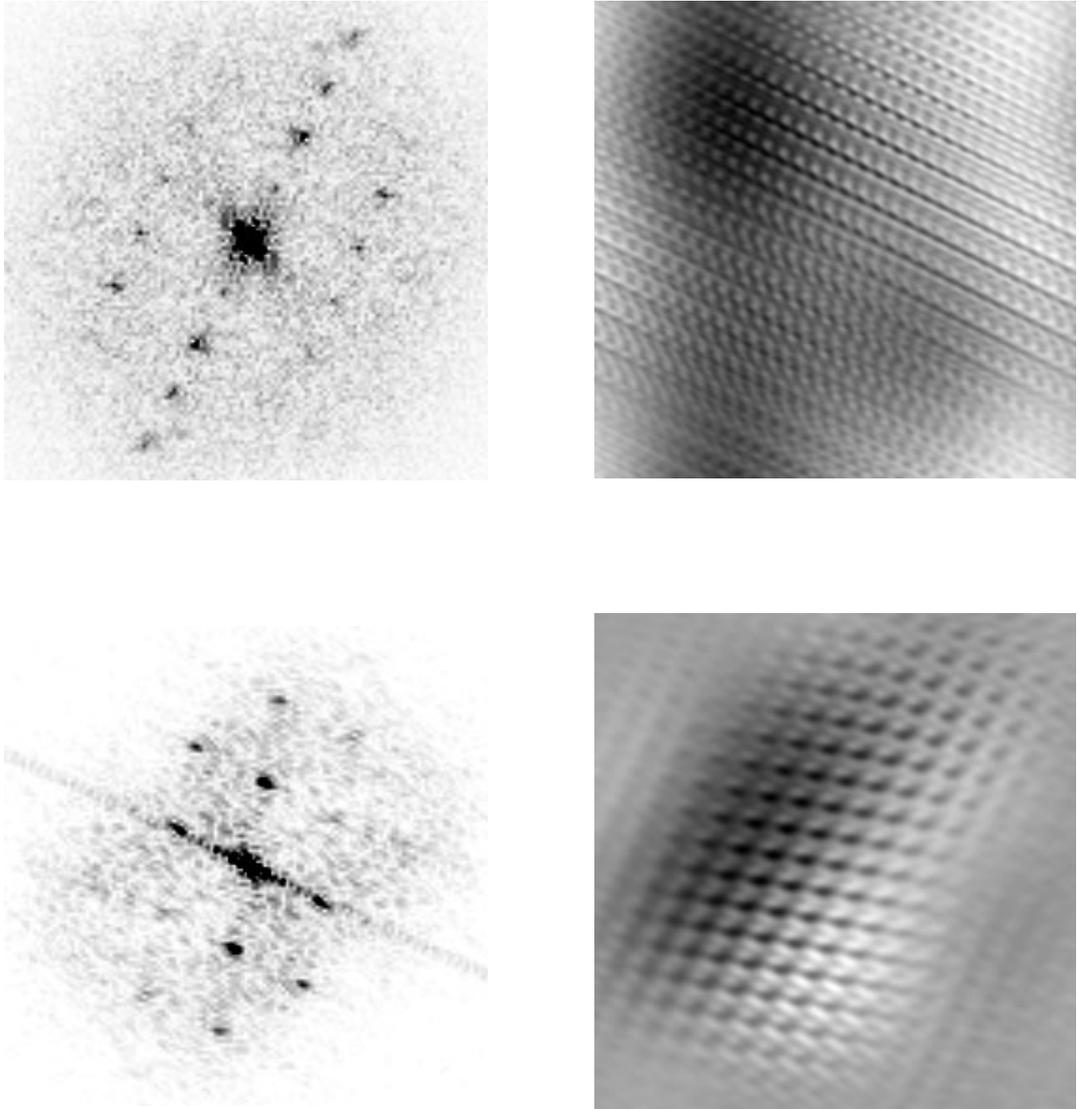


Figure 6.16: Moduli of the Fourier transforms (first column) and filtered images (second column) of two projections which could not be characterised.

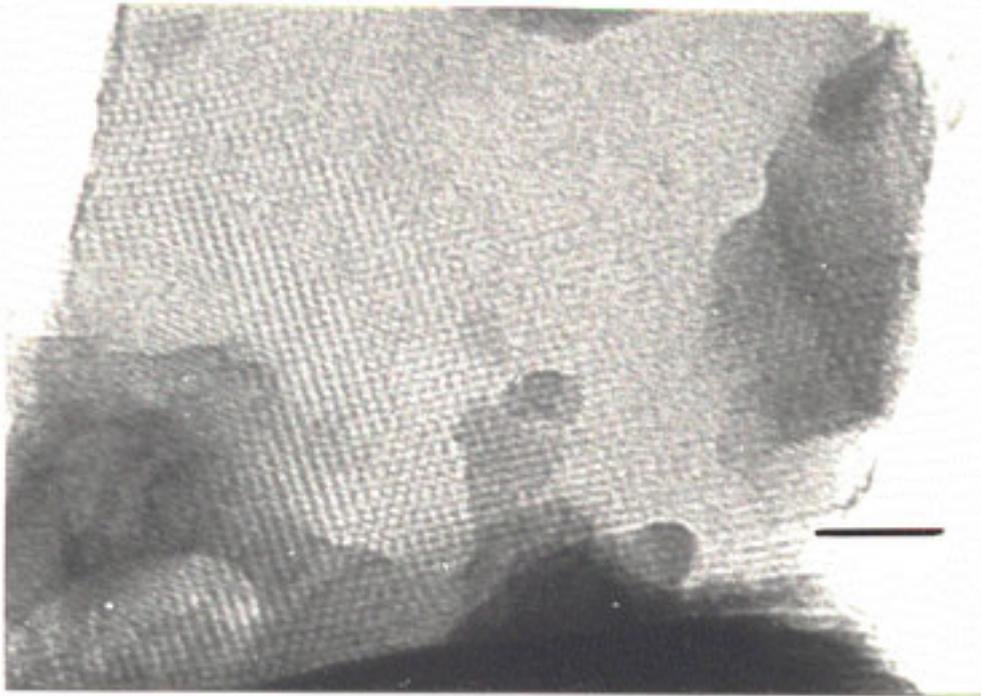


Figure 6.17: Electron micrograph showing different periodicities within the same area. Scale bar 500Å.

6.6 The Calculation of a 30Å Electron Density Map.

We saw in Section 6.1 that for the calculation of an experimentally determined 30Å electron density map, five projections must be identified and analysed : [010], [001], [100], [110] and [130]. The analysis of the [010], [001] and [130] projections has already been described (Section 6.4). Despite the large number of crystalline areas examined, no well preserved images of the two missing projections could be identified.

There are two obvious experimental solutions to this problem : The first is to rotate and then tilt a thin crystal (whose orientation is known) in such a way that the required projection axis is parallel to the electron beam. If, for example, a thin crystal is found with its [101] direction parallel to the electron beam, then the [100] projection can be reached through the following operations : (i) Rotate

the crystal about the [101] axis in such a way that the [010] axis is parallel to the tilt axis of the microscope, and, (ii) Tilt the crystal by 30.8° or -30.8° depending on the orientation of the axes. This will be the [100] projection. The [110] projection can now be reached by aligning the [001] direction with the tilt axis and then rotating about it by 17.8° . The two most serious problem with this procedure are : (i) the great majority of well preserved areas belong to the [001] and [010] projections, both of which are at right angles to the missing projections, and, (ii) Because tilting increases the apparent thickness of the specimen, very thin and relatively large crystals are needed. Attempts at finding crystals suitable for this type of experiment have been unsuccessful.

The second solution would be to fix the crystals with, say, glutaraldehyde, embed them in a polymer and cut sections perpendicular to the required axes in an ultramicrotome. (as described, for example, in Labaw, L.W. & Davies, D.R., 1972 or Langer, R., Poppe, C., Schramm, H.J. & Hoppe, W., 1975). AhrC crystals proved very sensitive to this treatment : fixing them for 2 hrs in 0.5% (v/v) glutaraldehyde resulted in a complete loss of the diffraction pattern. No further consideration was given to this method.

Within the 30\AA sphere, there are 12 non-centrosymmetric terms whose phases are unknown and 22 centrosymmetric terms whose signs are all known with the exception of three $0kl$ terms. It was decided to calculate phases for all non-centrosymmetric terms from a model of the electron density of the crystal consisting of spheres of constant density at all crystallographically equivalent positions of the molecular centre. These would be combined with the experimentally determined phases and a 30\AA three-dimensional electron density map could be calculated.

The amount of error introduced into the final three-dimensional map from this procedure depends on how good a model a sphere of constant density is for the given structure. Figures 6.18 and 6.19 illustrate this : Column (A) in

Figure 6.18 shows six sections through the density corresponding to a hypothetical 32 hexamer at 30Å resolution. Column (B) shows equivalent sections through the density of a sphere of constant density from which phases for all non-centrosymmetric terms were calculated. Column (C) is the map obtained from the mixed phase set and (D) is the difference between (C) and (A). Clearly, map (C) is a good approximation to the true density. Figure 6.19 shows the same calculations but for a hypothetical molecule (Column (A)) for which a sphere of constant density (Column (B)) is not a good model. The map calculated from the mixed phase set (Column (C)) is now a bad approximation to the true density, although, some correct features (such as the lack of density on section $y = 10/16$) are still present.

These model calculations suggest that (i) the final three-dimensional map can be a good approximation to the true structure only in favorable cases, (ii) the errors introduced in the final model can not be predicted, and, (iii) if a sphere of constant density is a bad approximation to the density of one molecule in the final map, then, serious errors are almost certainly present.

Figure 6.20(A) is a surface representation (drawn at $\approx 1.5\sigma$ above mean) of the 30Å electron density map corresponding to one hexamer. The most striking feature of this map is its lack of symmetry. The density is organised in six domains (marked I to VI in Figure 6.20) : Domains I, II and III are well connected and the same is true for domains IV, V and VI. Both trimers are irregular in shape and their relative orientation deviates significantly from that expected from a regular (32 or 6) hexamer. Figure 6.20(B) is a one molecule thick stack of sections from the same map, but this time the view is down the [010] axis. This should be compared with (C) which is the equivalent part of the experimentally determined [010] projection. The three high-density areas in the [010] projection can be identified with projections of the domains I, II and V; III and VI; and IV.

Although several features of this map might prove to be correct, the fact that

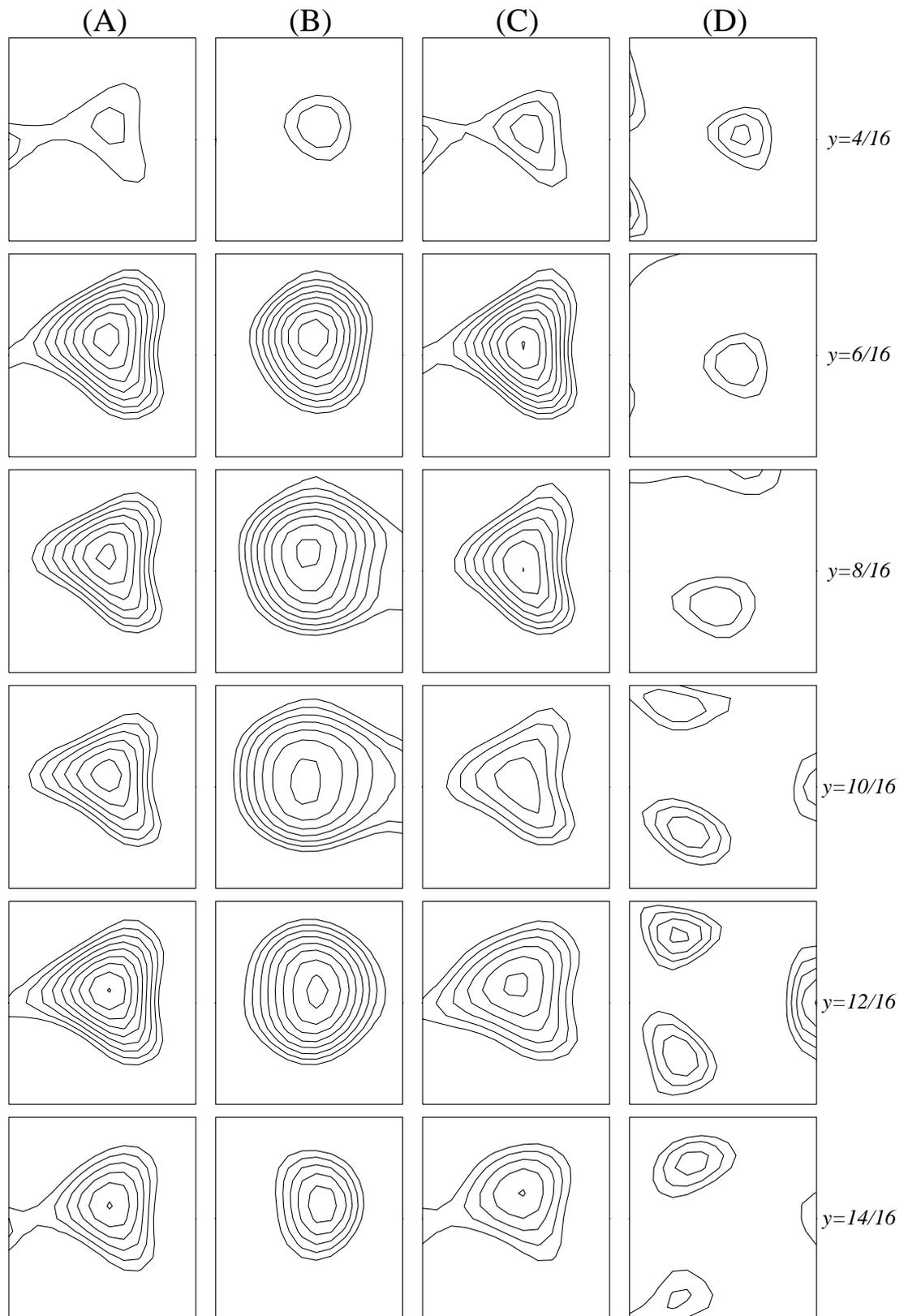


Figure 6.18: Sections through (A) : the density of a hypothetical 32 hexamer at 30Å resolution, (B) : a sphere of constant density, (C) : the map calculated with the mixed phase set, (D) : the difference map (C)-(A).

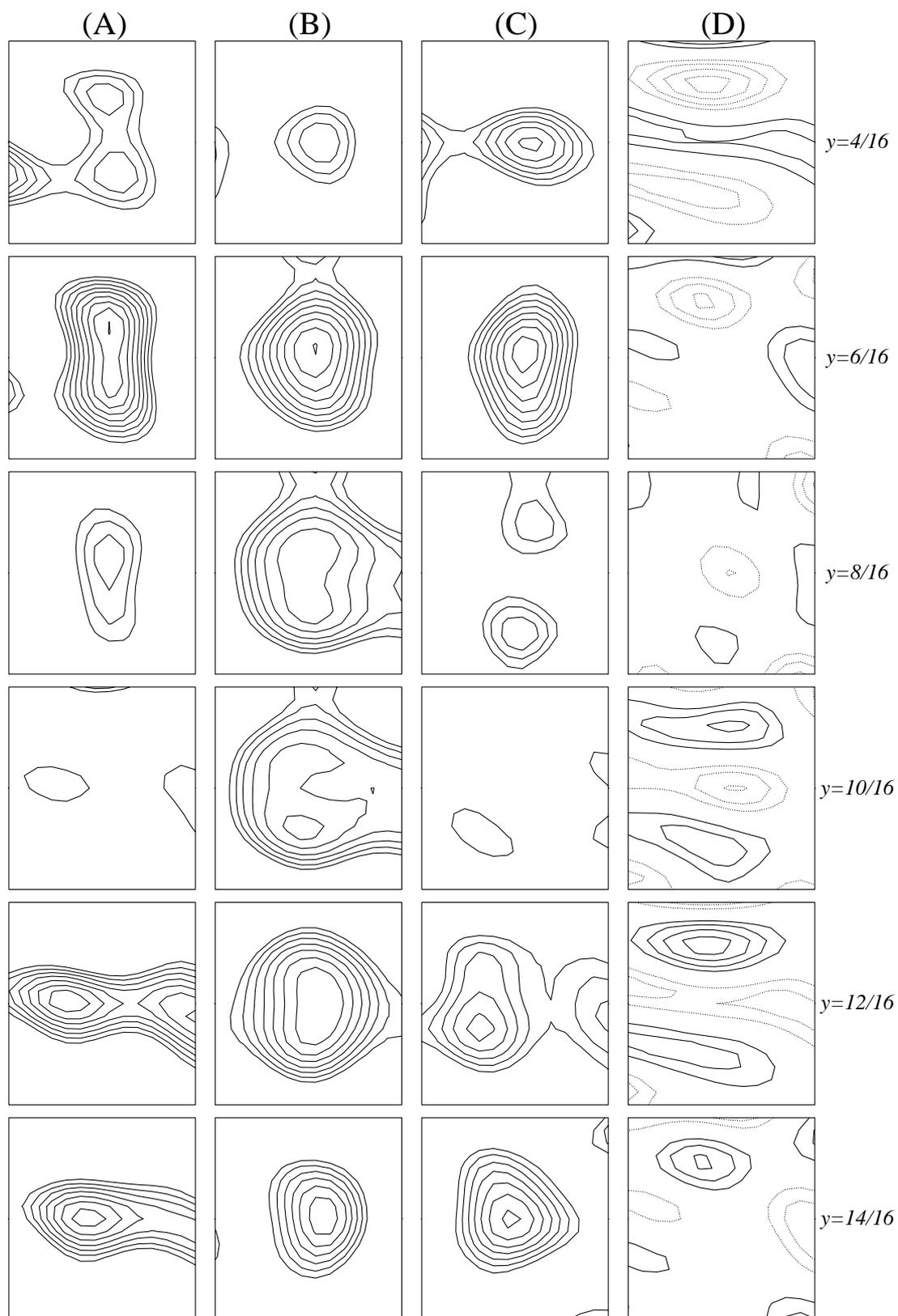
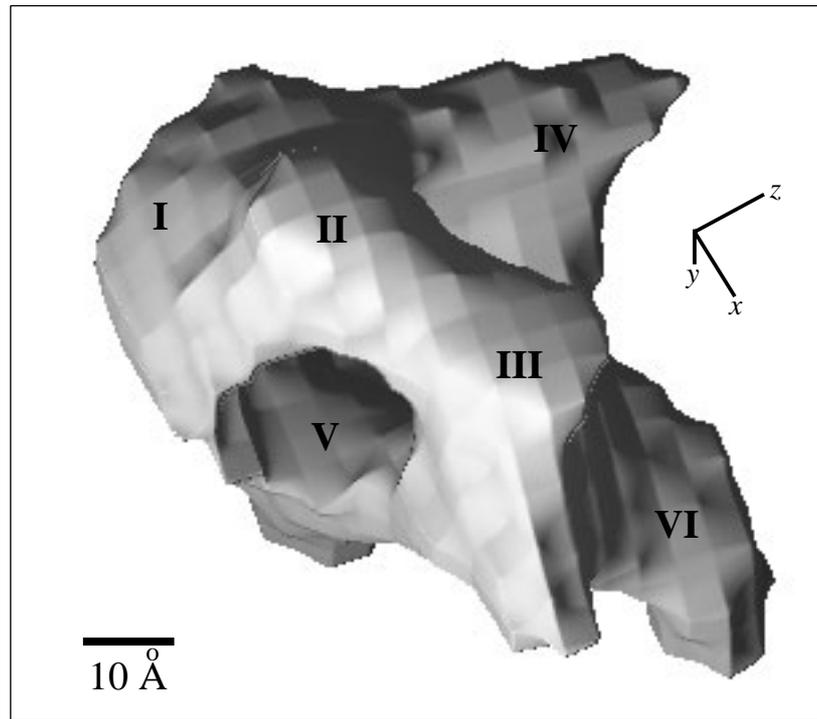
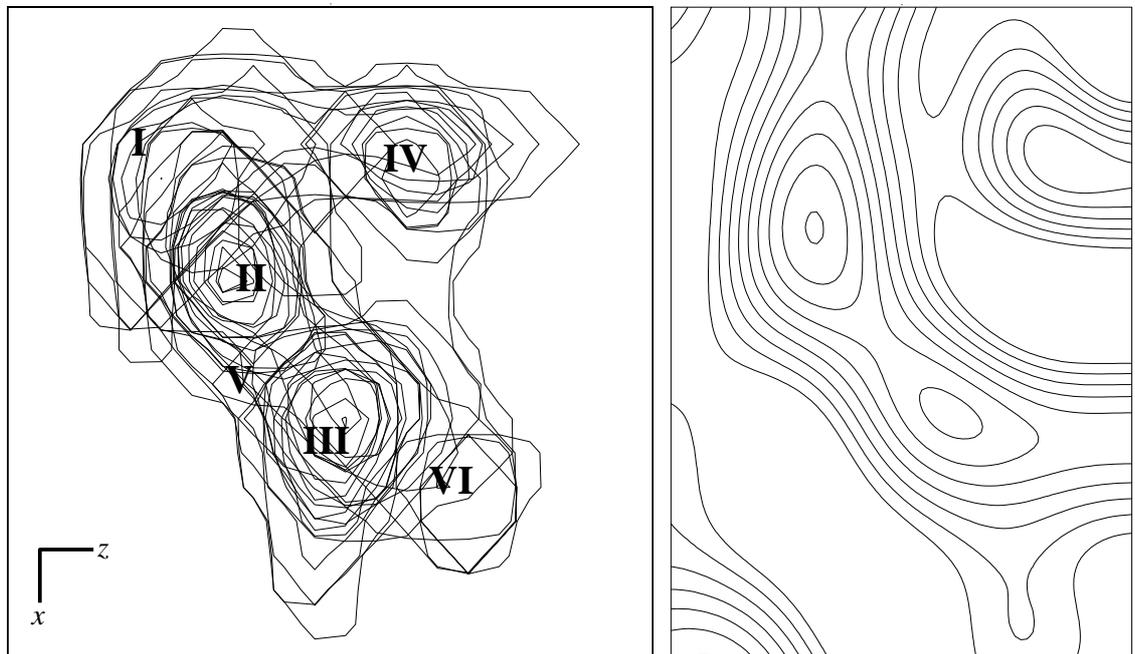


Figure 6.19: Sections through (A) : the density of a hypothetical molecule at 30Å resolution, (B) : a sphere of constant density, (C) : the map calculated with the mixed phase set, (D) : the difference map (C)-(A).



(A)



(B)

(C)

Figure 6.20: (A) : Surface representation of the 30Å electron density map, (B) : a stack of sections from the same map, and, (C) the corresponding area from the [010] projection.

a sphere of constant density is a bad approximation to its density and the absence of the expected symmetry elements, suggested that this model contains serious errors. Attempts to refine the phases (at constant resolution) though real space averaging have been unsuccessful.

Chapter 7

Molecular Replacement

7.1 The Model.

Greg van Duyne and Paul Sigler (Yale University) determined the crystal structure of the hexameric core (residues 80–152) of the Arginine Repressor from *Escherichia coli* (ArgR) at 2.5Å resolution. They generously made the atomic coordinates of their model available to us before publication. Their help is gratefully acknowledged.

Richardson type diagrams (Richardson, J.S., 1985, drawn using MOLSCRIPT, Kraulis, P.J., 1991) of the structure of the protomer, trimer and hexamer are shown in Figure 7.1 : (A) and (B) are orthogonal views of the structure of the protomer (the rotation axis is parallel to the longest dimension of the page). It consists of two adjacent β -hairpins and two α -helices which run approximately parallel to each of the hairpins, with an inter-helix angle of $\approx 90^\circ$. In (C) the orientation of the protomer is identical to the one seen in the trimer when viewed along the 3-fold axis (shown in (D)). Views (B) and (C) are approximately related by a 70° rotation about an axis parallel to the shortest dimension of the page. (E) and (F) are views of the hexameric core along the 3-fold and one of the 2-fold axes respectively. The hexamer is organised as a dimer of trimers in an approximately eclipsed conformation. Small deviations from exact 3_2 symmetry

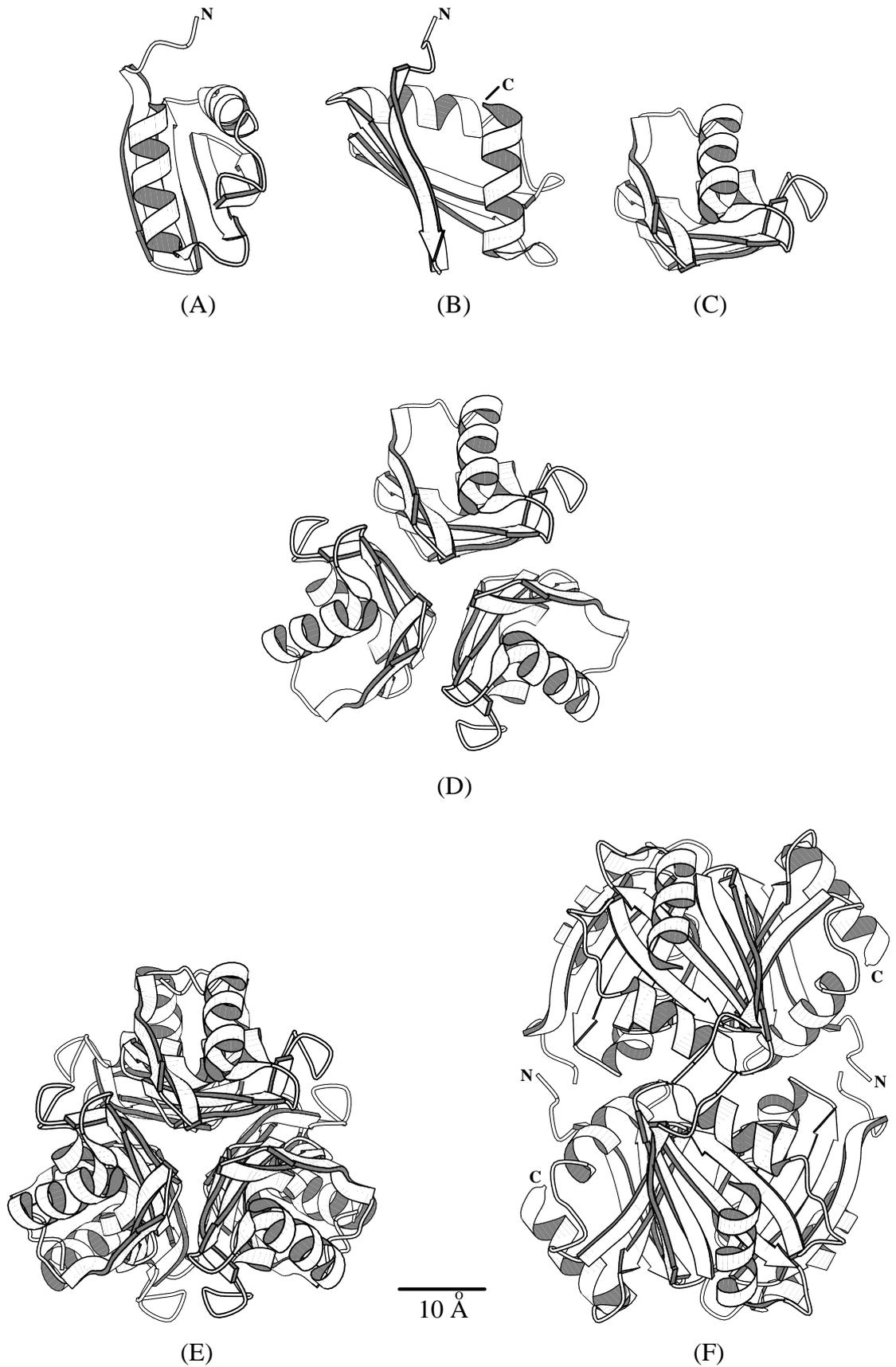


Figure 7.1: Schematic diagrams of the structure of the hexameric core of ArgR (Greg van Duyne and Paul Sigler, personal communication).

are present.

A sequence alignment of AhrC and ArgR in the region of interest is shown in Figure 7.2. From a total of 73 amino acids, 34% are identical and another 23% are similar. Following Chothia, C. & Lesk, A.M., 1986, the root mean square deviation in the positions of main chain atoms is expected to be $\approx 1.37\text{\AA}$.

```

ArgR  80- SPLKNLVLDIDYNDVAVVVIHTSPGAAQLIARLLDSLKAE
AhrC  78- RALMDAFVKIDSASHMIVLKTMPGNAQAIGALMDNLDWDE
          . * . . * *      . . * . * * * * * . * . * . * . *
          . . * * * * * * .      . . . . * * * .

ArgR      GILGTIAGDDTIFTTPANGFTVKDLYEAILELFDQEL      -156
AhrC      -MMGTICGDDTILIIICRTPEDTEGVKNRLELL      -149
          . . * * * * * * * .      . . . . * * * .

```

Figure 7.2: Pairwise alignment of the C-terminal region of ArgR and AhrC.

The structure of the hexameric core of ArgR accounts for less than half of the scattering material present in the asymmetric unit of the orthorhombic form of AhrC. Determination of the structure of AhrC from the ArgR model alone, although possible, would not be trivial. It was hoped that the phase information obtained from even half of the structure would allow the determination of the positions of the heavy atoms. The MIRAS phases could, then, be combined with the phases from the correctly oriented and positioned model, the non-crystallographic symmetry operators and the molecular envelope(s) determined from the (hopefully) improved MIRAS-MR map and the procedure of phase refinement and extension could be initiated.

The following sections describe the results from our attempts to solve the molecular replacement problem for both the orthorhombic and monoclinic forms of AhrC.

7.2 Orthorhombic Form.

7.2.1 The Programmes ALMN-TFFC and AMoRe.

Our first attempts at determining the correct orientation of the hexameric core of ArgR were based on the fast rotation function (Section 4.2) as implemented in the programmes POLARRFN and ALMN from the CCP4 suite of programmes. The result from numerous attempts with different resolution ranges and integration radii was an inconsistent collection of uniform distributions of peaks. Some of the orientations obtained were further tested using the Crowther & Blow, 1967, T_2 translation function as formulated by Harada, Y., Lifchitz, A., Berthou, J. & Jolles, 1981, and modified by Tickle, I.J., 1985, 1992 to allow for subtraction of all intramolecular vectors (programme TFFC from the CCP4 suite) :

$$T_2(\mathbf{t}) = \int_V (P_o(\mathbf{u}) - P_{11}(\mathbf{u}))(P_c(\mathbf{u}, \mathbf{t}) - P_{11}(\mathbf{u}))d\mathbf{u} \quad (7.1)$$

where, \mathbf{u} is a vector in Patterson space, P_o is the Patterson function of the unknown crystal, $P_c(\mathbf{t})$ is the Patterson function calculated from a model structure (in the space group of the unknown crystal) which is obtained by translating the model by \mathbf{t} before applying the crystallographic symmetry and P_{11} is the Patterson function corresponding to the set of intramolecular vectors of the model in the space group of the unknown crystal¹. All translation functions examined showed, as before, a more or less uniform distribution of peaks, none of which was consistent with the assumed position of the molecular centre. Model calculations with data calculated from hypothetical structures in space group C222₁ showed that these problems were not due to a space group specific bug in the programmes.

Our next attempt at solving the molecular replacement problem was based

¹Equation 7.1 represents the convolution of two Patterson functions, and by the convolution theorem, the Fourier transform of T_2 is the product of the Fourier transforms of these functions. T_2 can, thus, be obtained by Fast Fourier Transformation of a set of coefficients corresponding to the product of the Fourier coefficients of the Patterson functions.

on AMoRe, a package of programmes developed by Navaza, J., 1992 (and references therein), and modified by Dodson, E., for the CCP4 suite of programmes. AMoRe uses a modified fast rotation function (Crowther, R.A., 1972) which is made more accurate by (i) omitting low (angular) resolution spherical harmonic coefficients, and, (ii) by using numerical integration instead of Fourier-Bessel expansions (Navaza, J, 1987). Because AMoRe calculates and stores the whole molecular transform, both the fast rotation function and the Crowther & Blow translation function are made even faster, thus allowing an automatic exploration of a large number of rotation function solutions (typically the best 100 solutions).

Again, the great majority of cross rotation functions examined showed a more or less uniform distribution of peaks, with no identifiable consistency between results obtained from different calculations. It should be noted, however, that when low resolution data were used (12-6Å), the translation functions corresponding to some of the best orientations, gave top peaks at the assumed position of the molecular centre. Difference Fourier syntheses for the Niobium derivative which were calculated using phases from these solutions were uninterpretable. No progress could be made with any of these solutions.

7.2.2 X-PLOR.

Brünger, A.T., 1990, 1992b, proposed a new strategy for solving the molecular replacement problem and incorporated his ideas in the programme X-PLOR (Brünger, A.T., 1992c). The first step in this approach, is the calculation of a rotation function based on Huber's real-space Patterson search algorithm (Huber, R., 1985) : the Patterson function of the search model placed in a P1 cell is calculated, the vectors corresponding to a user-defined number of the highest peaks are saved and, then, for every unique rotation matrix Ω , the value of

$$RF(\Omega) = \langle P_{obs}P_{model}(\Omega) \rangle \quad (7.2)$$

is calculated. In this equation, P_{obs} is the Patterson function of the unknown crystal and $P_{model}(\Omega)$ is the rotated Patterson function of the search model consisting of only the selected peaks.

In the second step, a large number of the best solutions from the rotation function are examined using the ‘‘Patterson Correlation refinement’’ procedure (PC-refinement). This is a conjugate gradient minimisation of the target function $(1 - PC(\Omega, \Omega_i, t_i))$, where Ω is the rotation matrix defining the orientation of the molecule, Ω_i are rotation matrices of individual domains or other collections of atoms that are treated as rigid bodies, t_i are the corresponding translations, and

$$PC(\Omega, \Omega_i, t_i) = \frac{\langle E_o^2 E_c^2(\Omega, \Omega_i, t_i) - \langle E_o^2 \rangle \langle E_c^2(\Omega, \Omega_i, t_i) \rangle \rangle}{\sqrt{\langle E_o^4 - \langle E_o^2 \rangle^2 \rangle - \langle E_c^4(\Omega, \Omega_i, t_i) - \langle E_c^2(\Omega, \Omega_i, t_i) \rangle^2 \rangle}} \quad (7.3)$$

The right-hand side of Equation 7.3 is the linear correlation coefficient between the squared normalised amplitudes of the structure factors of the unknown crystal (E_o^2) and of the model structure placed in a P1 cell identical in geometry with the unknown crystal ($E_c^2(\Omega, \Omega_i, t_i)$). This is equivalent to calculating the linear correlation coefficient between the (sharpened) observed Patterson function and the Patterson function of the model structure in a P1 cell. Clearly, the aim of PC-refinement is to maximise the agreement between the predicted and observed intramolecular vectors, by making the tacit assumption that cross vectors between crystallographically related molecules (in the unknown crystal) can be treated as noise. If the procedure is successful, PC-refinement of the correct orientation will converge to a much higher value of the correlation coefficient than any of the false solutions. An additional advantage of this approach, is that the availability of a refined model can offer a significant improvement of the signal to noise ratio in a subsequent calculation of a translation function.

The translation function in X-PLOR is identical to the one used in BRUTE (discussed in Section 5.4.2), the only difference being that X-PLOR uses E -values instead of F s (although it is possible to do the calculation on F s).

Several attempts with different resolution ranges, integration radii, number of vectors and PC-refinement protocols, have all resulted in a more or less uniform distribution of the PC metric. When low resolution data were used, some promising solutions appeared. Figure 7.3 shows the results from PC-refinement of the best 298 orientations from a rotation function calculated with 13-5Å data, integration radius 25Å and 1000 vectors : (A) is the value of the Patterson Correlation coefficient before refinement of the orientation of the hexamer and (B) is after 15 cycles of conjugate gradient minimisation of the target function.

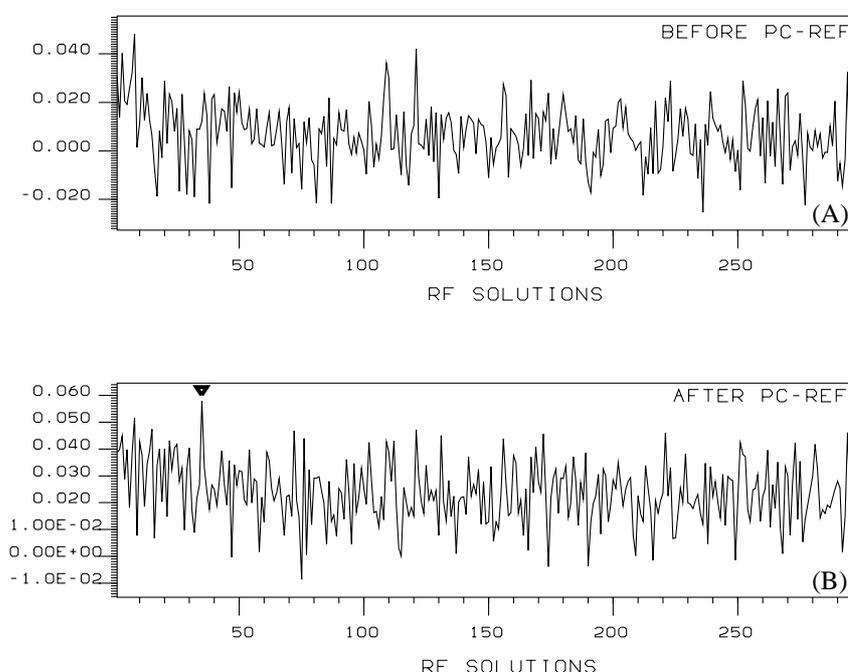


Figure 7.3: The value of the PC metric, (A) before, and, (B) after PC refinement of the orientation of the hexamer for the best 298 solutions from a rotation function calculated with low resolution data.

A translation function which was calculated using the orientation corresponding to the highest peak from this search (indicated by an arrow in (B)), gave a uniform distribution of peaks none of which was close to the expected position of the molecular centre. Similar results have been obtained from all orientations

examined.

The organisation of the hexamer as a dimer of trimers and the not very extensive interface between the two trimers suggested that the relative orientation of the trimers about the 3-fold might not be the same for AhrC and ArgR. Figure 7.4 shows the results from PC-refinement using as a search model only the ArgR trimer. The lack of signal is again evident.

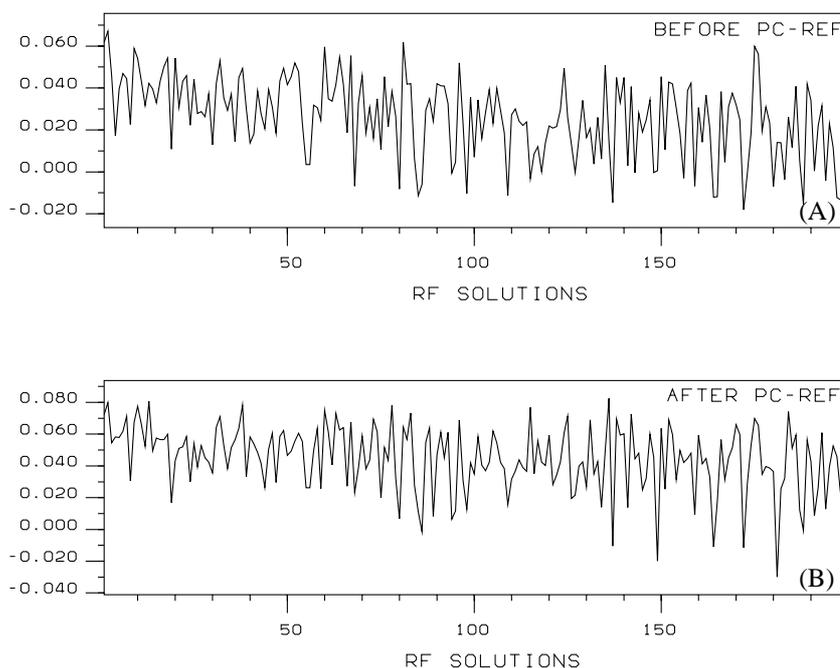


Figure 7.4: The value of the PC metric, (A) before, and, (B) after PC refinement of the orientation of the ArgR trimer for the best 198 orientations from the rotation function.

Similar calculations (using X-PLOR and AMoRe) have been carried out with models of the hexameric core in which the two trimers have been rotated with respect to each other by -15 , -10 , -5 , 5 , 10 and 15° (the rotation is about the common 3-fold axis). Again, no convincing solutions could be identified.

Brünger, A.T., 1992c, suggested a new “Direct Rotation Function” algorithm. This is a calculation of the PC metric (without refinement) for all unique orien-

tations of the molecule. The best solutions from this search are then subjected to the usual PC-refinement. Figure 7.5 shows the results from this search.

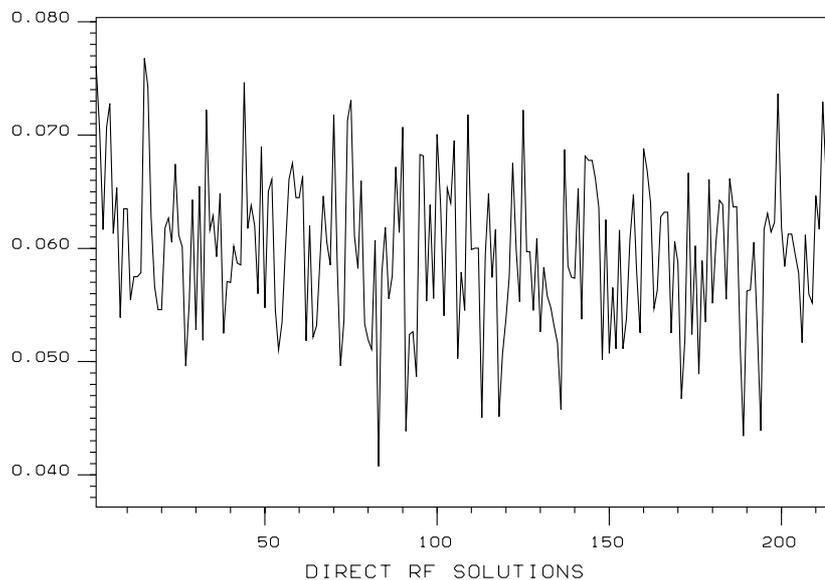


Figure 7.5: The value of the PC metric after PC refinement of the orientation of the ArgR hexamer for the best 216 orientations from the “Direct Rotation Function”.

Our final attempt at solving the molecular replacement problem was a brute force approach based on the knowledge of the position of the molecular centre : for every unique orientation of the search model, the translation corresponding to one of the equivalent positions of the molecular centre was applied, the crystallographically related molecules (in the space group of the unknown crystals, $C222_1$) were generated, and 12 cycles of rigid-body refinement of the orientation of the hexamer were carried out (target function was the correlation coefficient between the observed and calculated amplitudes of the structure factors). Since both intra- and inter-molecular vectors are used, this search is expected to be more powerful than PC-refinement (assuming that the position of the molecular centre is known with sufficient accuracy). Figure 7.6 shows the values of the correlation coefficient (after rigid-body refinement) for 7030 unique orientations

and for two searches conducted using data from different resolution ranges (a 10° step size of the three Eulerian angles (α, β, γ) was used for both calculations). The search using 7-5Å data has also been performed with the molecular centre at $x = -0.13$, $y = -0.09$ and $z = -0.225$ but with similar (negative) results.

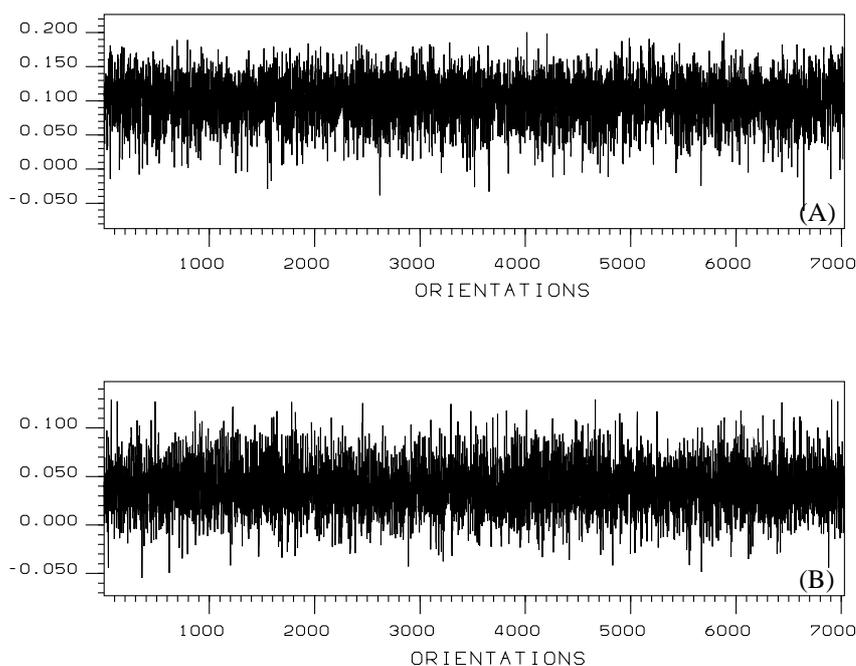


Figure 7.6: Results from a brute force search using X-PLOR. (A) 9-7Å, (B) 7-5Å.

7.3 Monoclinic Form.

7.3.1 A Model of the Crystal Packing.

Analysis of the [001] and [010] projections of monoclinic AhrC crystals allowed us to obtain a model of the crystal packing for this form. Because this model was subsequently used to judge the results obtained from the translation function calculations, the procedure by which it was determined will be discussed in this Section.

The presence of a centered superlattice in the [001] projection was inferred from a precession photograph of the $hk0$ level (Figure 2.8 in Chapter 2). This pattern of systematically weak and strong reflections generates a large peak at $u = 0.5, v = 0.25$ in the low resolution [001] native Patterson projection (shown in Figure 7.7).

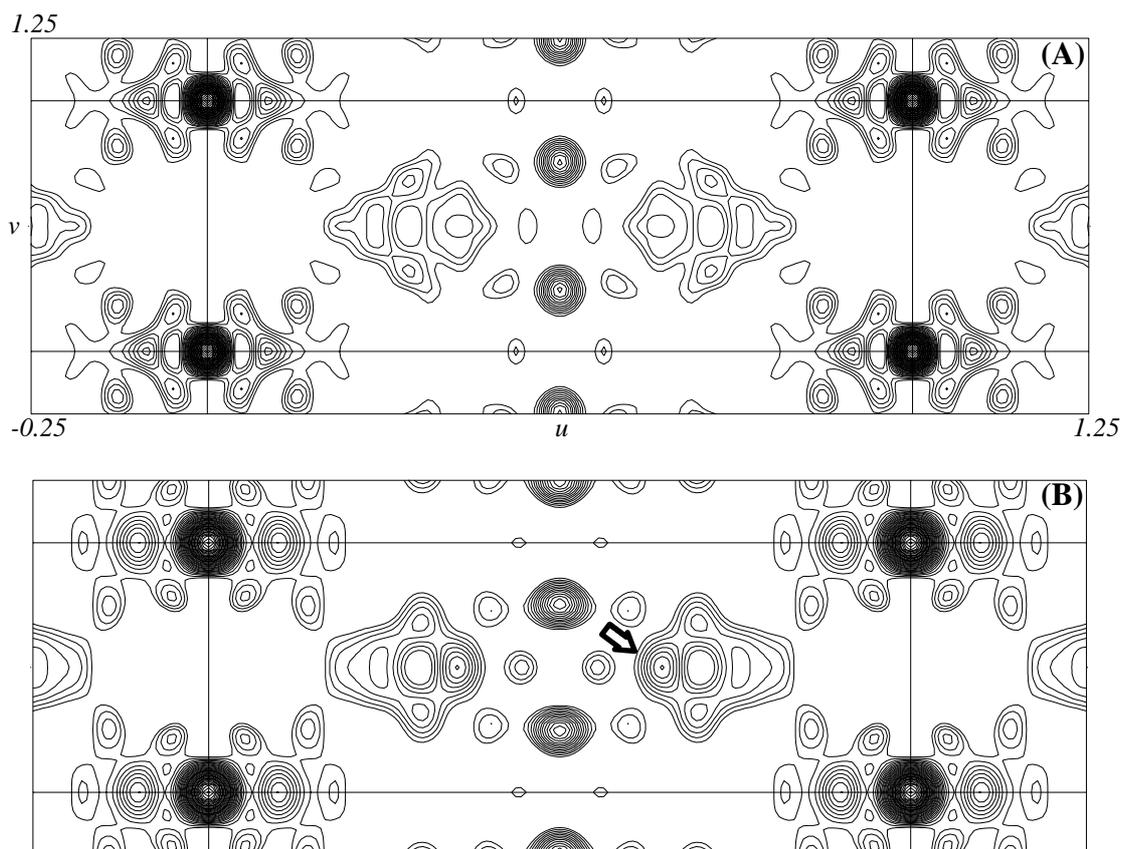


Figure 7.7: Monoclinic form : (A) 202-12, and, (B) 202-15Å [001] native Patterson projection. Contours every 3% of the origin peak.

The height of this peak is $\approx 37\%$ of the origin peak, fairly close to the value expected for a Patterson peak arising from the presence in the unit cell of two crystallographically independent molecules which are related by a simple translation. Given that both the b ($=72.6\text{\AA}$) and c ($=73.0\text{\AA}$) translations are too short to allow significant overlap of two hexamers, we conclude that the four molecules in the unit cell are resolved in this projection. Since the large peak is at $u = 0.5$,

the two molecules whose projections are related by a simple translation must belong to different asymmetric units, that is, the translation $u = 0.5$, $v = 0.25$ must relate the projection of one of the molecules in the asymmetric unit with a crystallographically equivalent copy of the second. This immediately suggests that the sum of the x coordinates of the two molecules in the asymmetric unit is ≈ 0.5 . Furthermore, since we require that after application of the crystallographic symmetry to the second molecule, its projection will be identical to that of the first, we can conclude that the two molecules in the asymmetric unit are related by a non-crystallographic “glide” line parallel to y with a translational component of ≈ 0.25 . Because the plane group of this projection (pg) is polar, the y coordinate of one molecule can be set to 0.0 to fix the origin, in which case the second molecule is at $y \approx 0.25$. This leaves only one parameter to be determined : the x coordinate of only one of the molecules in the asymmetric unit, or equivalently, the separation between them.

We have determined this coordinate by three different methods : (i) examination of very low resolution permutation syntheses, (ii) analysis of the native Patterson projection, and, (iii) through a systematic correlation search. Figure 7.8 shows the two unique permutation maps for the two strongest low resolution [001] reflections (110 and 210)². Clearly, the first synthesis is the only reasonable solution. The two molecules in the asymmetric unit are separated by 0.25 along y as expected. Their x coordinates are $x_1 = 0.149$ and $x_2 = 0.351$ ($=0.5-0.149$).

The x coordinates of the projection of the two molecules in the asymmetric unit can be determined more accurately from the [001] native Patterson projection : because the non-crystallographic symmetry (a “glide” line with a translational component of 0.25 parallel to y) is parallel to the crystallographic glide

²Because this projection is non-centrosymmetric, four possible phase angles were examined for each reflection ($\pm\pi/4$, $\pm 3\pi/4$). Although the total number of phase combinations is 16, fixing the origin and the enantiomorph leaves only 4 unique phase sets. The electron density maps corresponding to these four sets are two syntheses plus their Babinet opposites. Because only two reflections are used, the electron density maps of each pair of Babinet opposites are equivalent (related by a simple translation), leaving only two unique syntheses.

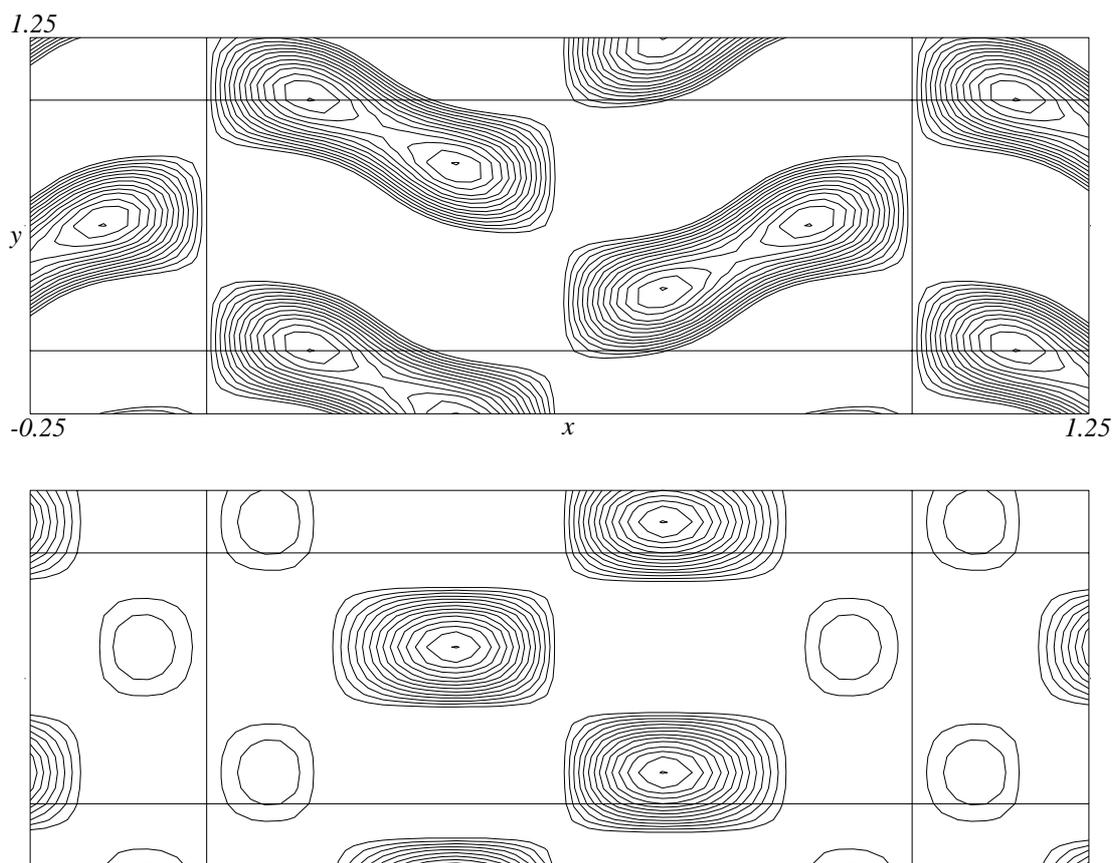


Figure 7.8: Monoclinic form : [001] Permutation syntheses.

line, cross vectors between molecules related by both crystallographic and non-crystallographic symmetry will coincide at $u = 2x_1$, $v = 0.5$ and at $u = 2x_2$, $v = 0.5$, thus generating significant Patterson peaks at those positions. The expected peaks are obvious in the $202\text{-}15\text{\AA}$ Patterson projection (marked by an arrow in Figure 7.7(B)) but not in the $202\text{-}12\text{\AA}$ map, suggesting again that the non-crystallographic symmetry is not exact. Their positions are $u = 0.33$ and $u = 0.67$, giving $x_1 = 0.165$ and $x_2 = 0.335$. The difference from the values determined from the permutation synthesis is 3.2\AA .

Finally, an attempt was made to determine the x coordinates through a systematic search : A programme was written which for every position (along x) of a disc of constant density calculated the correlation coefficient and R -factor between the observed and calculated amplitudes of the structure factors. Figure 7.9 shows the results obtained from a search conducted with a disc of radius 28\AA and

all data between 202 and 22Å resolution (18 reflections). The highest peak is at $x \approx 0.10$ but the function has high values for the whole range from $x=0.09$ to $x=0.20$.

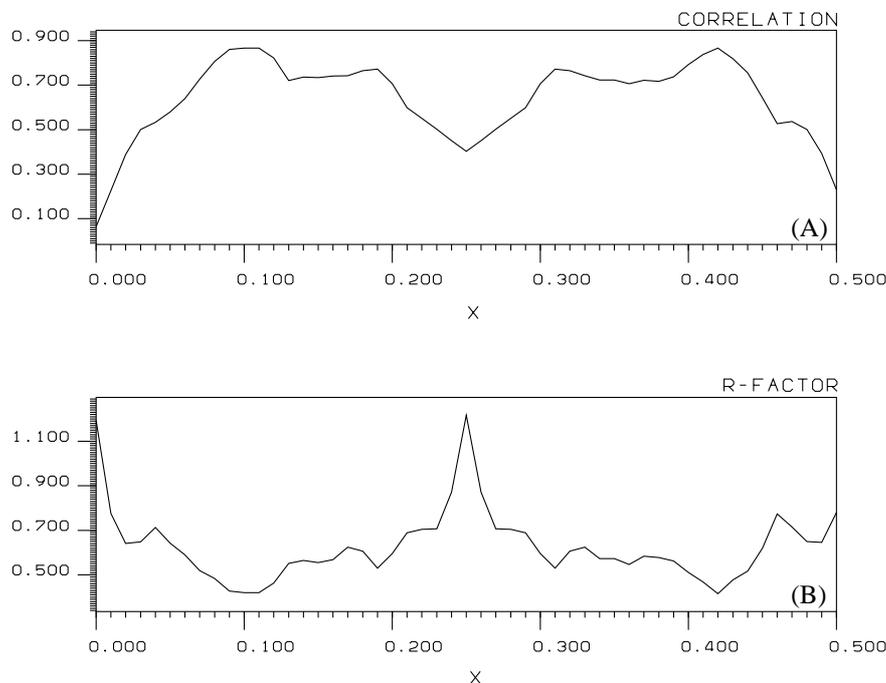


Figure 7.9: Results from a systematic search along x using discs of constant density, (A) Correlation coefficient, (B) R -factor.

The analysis of the [010] projection was again based on examination of low resolution permutation syntheses and on a systematic search using discs of constant density. Figure 7.10 shows the 60Å permutation map that gave the best agreement with the previously established x coordinates of the two molecules. Their z coordinates (as determined from this map) are 0.138 and 0.550.

A programme was written which for every crystallographically unique combination of the coordinates of two discs of constant density, calculated the R -factor and linear correlation coefficient between the observed and calculated amplitudes of the structure factors. The best solution from this search (with $C = 0.79$ and $R = 0.38$) placed the centre of the two discs at $x_1 = 0.10$, $z_1 = 0.15$ and $x_2 = 0.42$, $z_2 = 0.55$.

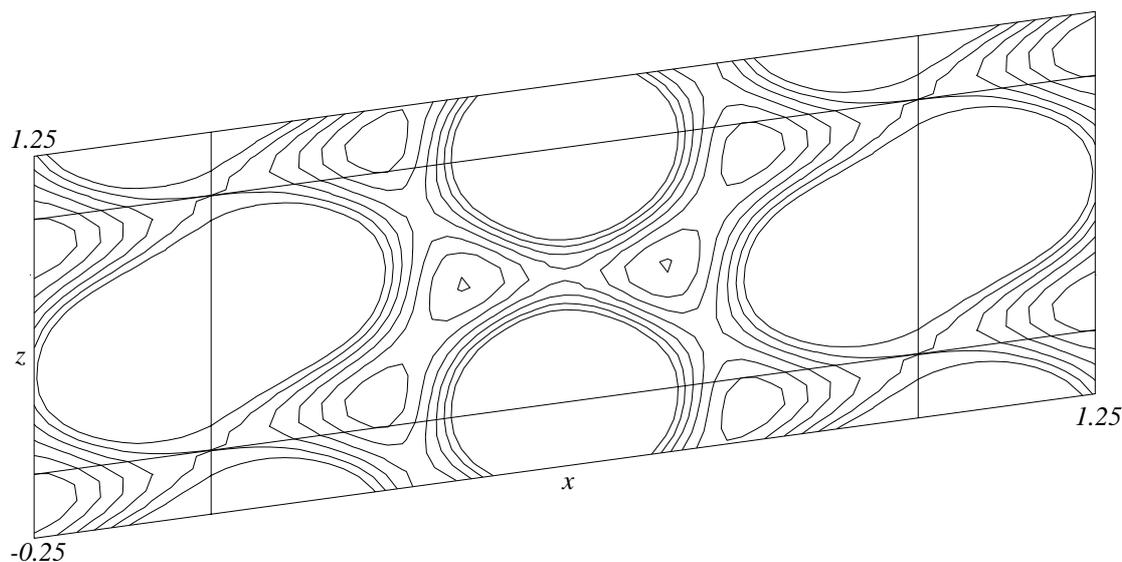


Figure 7.10: [010] Projection : A 60Å permutation synthesis (6 reflections : $\bar{1}01$, 100, 101, 200, 201 and 300).

In conclusion, the positions of the molecular centres of the two hexamers in the asymmetric unit of the monoclinic form have been determined as $x_1 \approx 0.16$, $y_1 = 0.0$, $z_1 \approx 0.15$ and $x_2 \approx 0.34$, $y_2 \approx 0.25$, $z_2 \approx 0.55$.

7.3.2 Molecular Replacement.

Several attempts (using AMoRe and X-PLOR) to solve the molecular replacement problem for this crystal form have again been unsuccessful. When low resolution data were used, some promising solutions appeared, but no progress could be made with any of them.

Figure 7.11(C) and (D) shows results obtained from PC-refinement of the best 121 orientations from a rotation function calculated using 8-5Å data. The uniform distribution of peaks in these graphs is typical of the calculations performed. Figure 7.11(A) and (B) shows results obtained from PC-refinement of the best 302 orientations from a rotation function calculated using very low resolution data (12-7Å). Comparison of Figures 7.11(A) and (B), shows that the two peaks

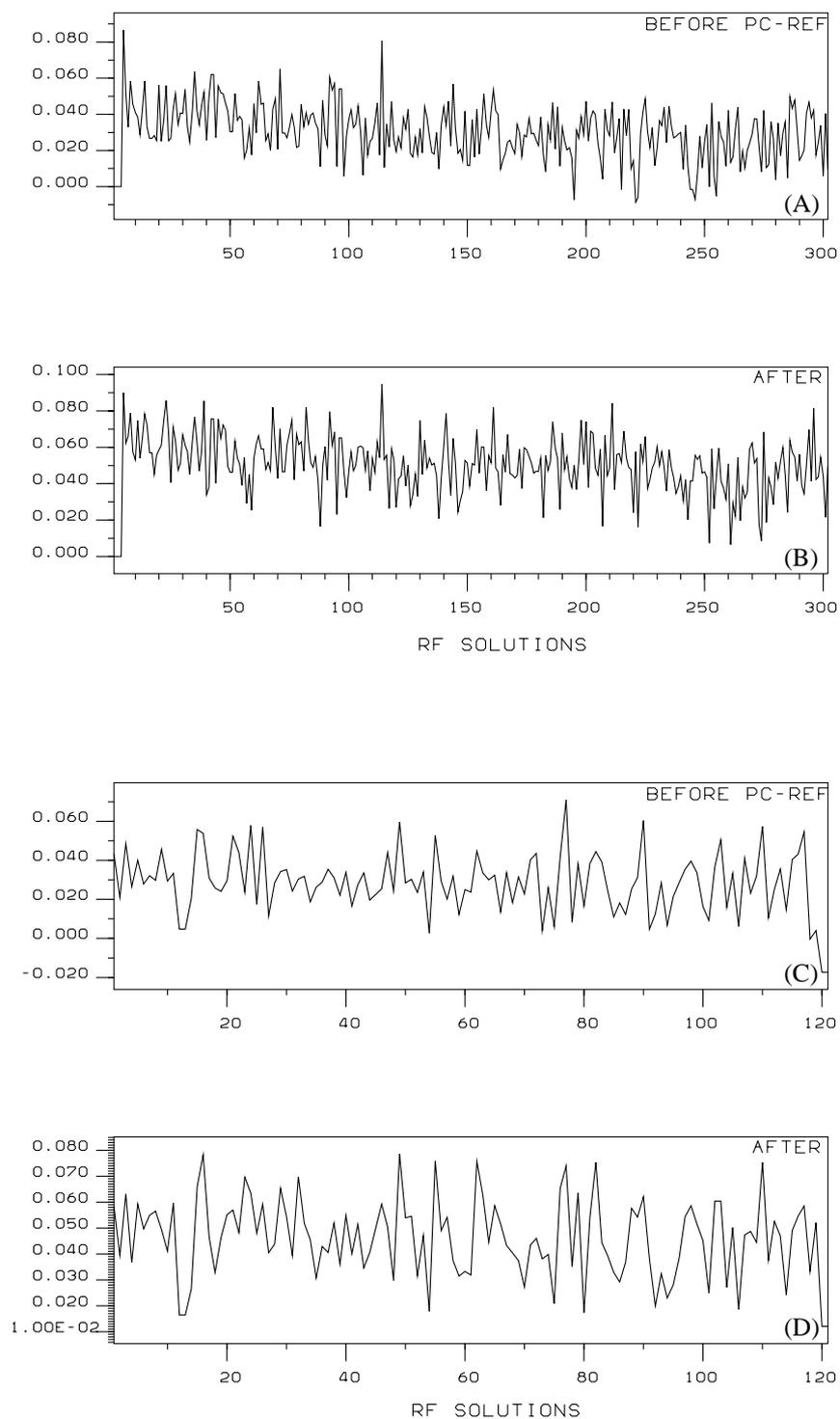


Figure 7.11: Values of the PC metric before (A and C) and after (B and D) PC-refinement of the best solutions from two rotation functions calculated using 12-7 and 8-5Å data.

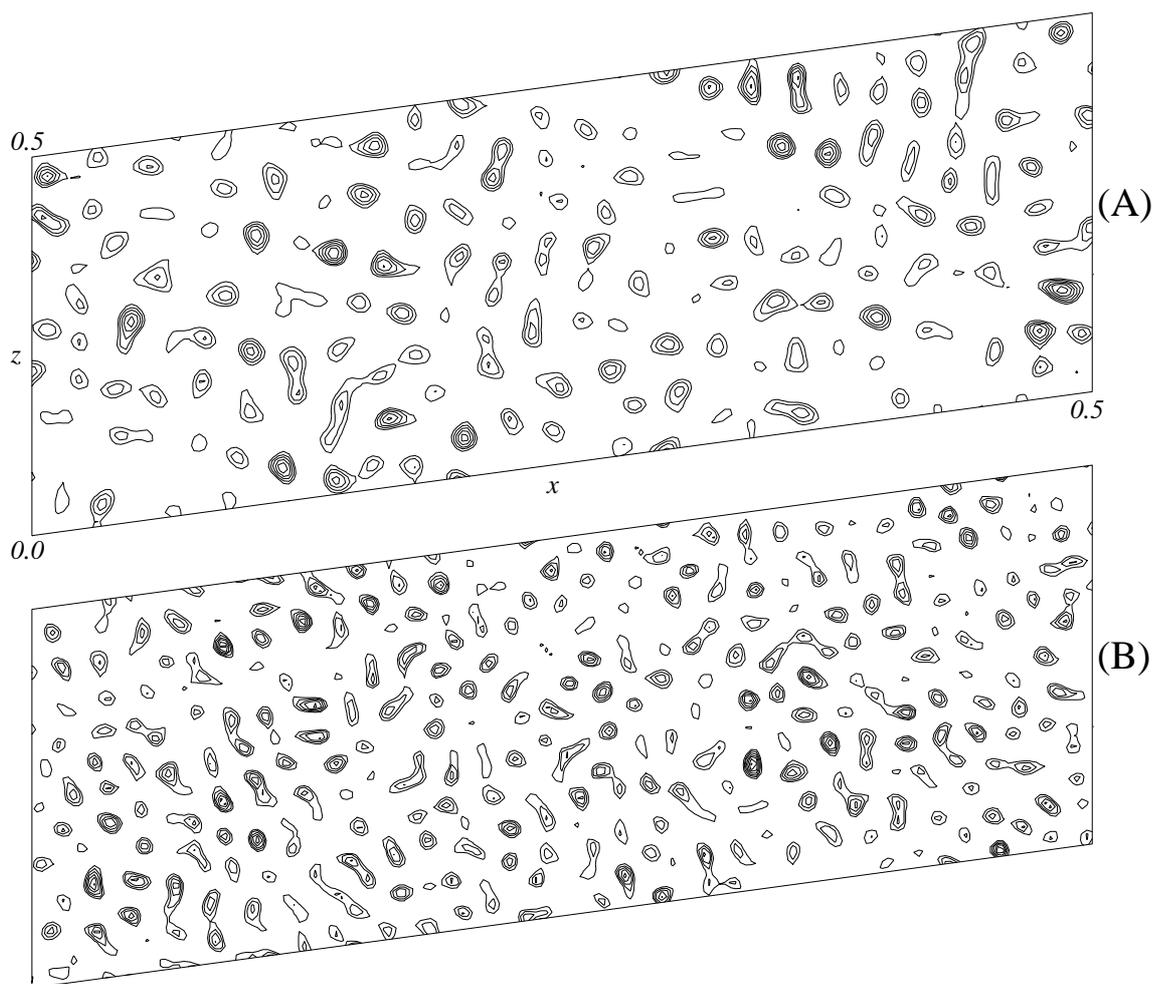


Figure 7.12: Monoclinic form : Results from two translation function calculations in X-PLOR (A) 12-7, (B) 8-5Å. Contours every 0.5σ above mean.

that are clearly above the noise level in (A), are still the highest peaks in (B), but the distribution of the Patterson correlation coefficient after PC-refinement is more uniform.

Figure 7.12 shows two translation functions which were calculated in X-PLOR using the same orientation (the one corresponding to the highest peak from PC-refinement) but different resolution ranges (12-7Å for (A) and 8-5Å for (B)). The more or less uniform distribution of peaks and the inconsistency of the two functions is obvious. Similar results have been obtained from all translation functions examined.

Discussion

Before discussing the future prospects of this project, the results from some experiments which have not been mentioned in the preceding Chapters will be presented.

We saw in Chapter 3 that most of the heavy atom containing compounds damage the crystals even at very low concentrations. It was thought that these problems were due to the presence of two cysteine residues in AhrC. Coleen Miller (Department of Genetics, University of Leeds) prepared (using PCR-based site-directed mutagenesis) three mutant forms of the *ahrC* gene encoding for proteins in which either or both of the cysteine residues were substituted for serine. Attempts to purify the products of these genes showed that their solubility properties were totally different from that of the native AhrC; all three mutants are insoluble in Arg buffer (Section 2.1.1) containing NaCl with concentrations ranging from 100 mM to 3 M. A possible explanation of this result is that the mutants can not fold properly and have, thus, formed inclusion bodies.

Carol Holtham (Department of Genetics, University of Leeds) prepared a gene construct encoding for a chimaeric molecule consisting of the N-terminal (DNA-binding) domain of AhrC (residues 1-87) and the C-terminal domain of ArgR

(residues 91-156). The product of this gene has solubility properties very similar to that of the native AhrC and it was purified as described in Chapter 2. Attempts to crystallise this AhrC-ArgR chimaera under conditions similar to those used for AhrC have been unsuccessful. Further attempts using other, unrelated crystallisation conditions (including a commercially available wide screen) have also been unsuccessful.

The availability of a model for the hexameric core fragment of ArgR and the problems encountered with obtaining a Molecular Replacement solution for the AhrC crystals (Chapter 7) suggested that an attempt to prepare and crystallise the core fragment of AhrC would be a worthwhile exercise : Its structure could be determined by Molecular Replacement using the hexameric core fragment of ArgR as a search model and the thus determined structure could, in turn, be used as a search model for the native AhrC crystals. Due to the unavailability of a gene construct encoding for the C-terminal domain of AhrC, it was decided to prepare the core fragment through proteolytic cleavage of intact AhrC. Figure D.1 shows the results from the attempt to cleave AhrC using the V8 endopeptidase.

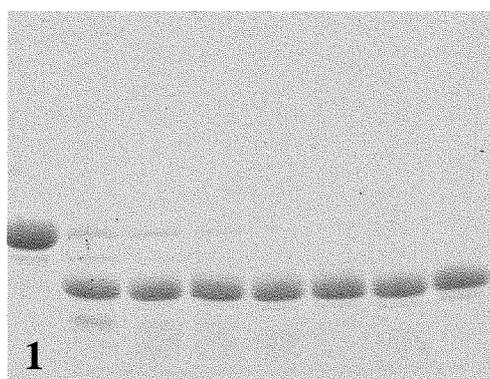


Figure D.1: Lane 1 : Intact AhrC, Lanes 2, 3, 4, ..., samples taken after 2, 4, 8, ..., minutes of incubation in the presence of 2% (w/w) V8 at 38°C.

The molecular weight of the major band seen in Lanes 2 to 7 (as judged from its electrophoretic mobility, data not shown) is 8 kDa and probably corresponds to the hexameric core fragment of AhrC. The reaction was stopped at $t = 8$ min by the addition of 3 mM PMSF and the product was stored at 4°C for further use. Attempts to crystallise the AhrC core from this protein preparation have all been unsuccessful.



In summary, our attempts to determine the structure of AhrC using Multiple Isomorphous Replacement, Electron Microscopy or Molecular Replacement, have all been unsuccessful.

The major problem with the MIR procedure is clearly the determination of the positions of the heavy atom sites. Although it is not unlikely that the heavy atom positions can still be determined with the existing data and methods, the best way forward appears to be the preparation of other derivatives with, hopefully, fewer sites: the analysis of crystal packing of the orthorhombic form showed that large solvent channels run parallel to the [001] direction. It should be possible to soak these crystals in solutions containing large heavy atom clusters such as undecagold or tetrairidium which, if bound to AhrC, should give derivatives with fewer sites³. Another, more thoughtful way for preparing a derivative with a small number of substitution sites would be to synthesise a heavy atom containing analogue of L-Arginine which could then be used in soaking or co-crystallisation experiments.

Electron microscopy (or crystallography) of two-dimensional AhrC crystals should give the structure of AhrC to a resolution sufficient for initiating the

³It is interesting to note that the accessibility of the gold cluster can be varied by attaching it to reactive groups through straight chain molecules of different lengths (Weinstein S., Jahn, W., Hansen, H., Wittmann, H.G. & Yonath, A., 1989). This makes possible the preparation of useful heavy atom derivatives in a controlled and systematic way.

process of phase refinement and extension. Although our attempts at growing two-dimensional crystals have not been successful, the range of conditions tried is by no means exhaustive and further attempts may be in order.

One other promising way for determining the structure of AhrC is to pursue the crystallographic studies of the AhrC-ArgR chimaera or of individual domains of AhrC, especially the hexameric core fragment. Although proteolytic cleavage of intact AhrC gave a relatively pure preparation of this fragment, a serious crystallographic study would probably require the production of a gene construct encoding for the C-terminal domain of AhrC.

Finally, crystallisation of an AhrC-DNA complex would not only be exciting in itself, but it might also make the preparation of useful heavy atom derivatives with few sites easier (through chemical modification of the DNA).

Appendix A

Trigonal Form :

A Redetermination

Very recently, a precession photograph of a trigonal AhrC crystal taken in a direction perpendicular to the 3-fold was obtained. This is shown in Figure A.1.

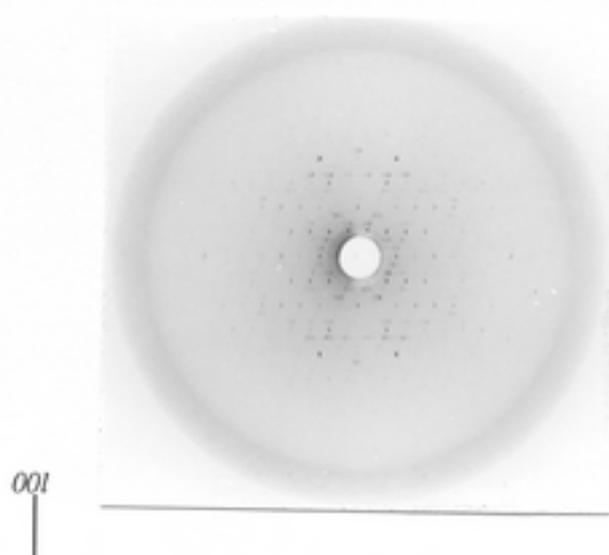


Figure A.1: 8° precession photograph of a trigonal AhrC crystal taken in a direction perpendicular to the crystallographic 3-fold axis.

Figure A.2 shows a magnified area around the origin of the reciprocal lattice. Due to the very high mosaic spread of this crystal, the possibility that this pho-

tograph corresponds to a superposition of more than one reciprocal lattice plane can not be excluded with certainty. For this reason a detailed interpretation will not be attempted. One conclusion that can safely be drawn, however, is that the lattice is rhombohedral and not primitive as was incorrectly deduced from examination of the $hk0$ and $hk1$ precession photographs (Figures 2.10 and 2.11 in Chapter 2). The only possible enantiomorphic space groups are R3 and R32 and since the $hk0$ level has symmetry 6mm (Figure 2.10), the space group is R32 and not P312 or P321 as discussed in Section 2.2.3.

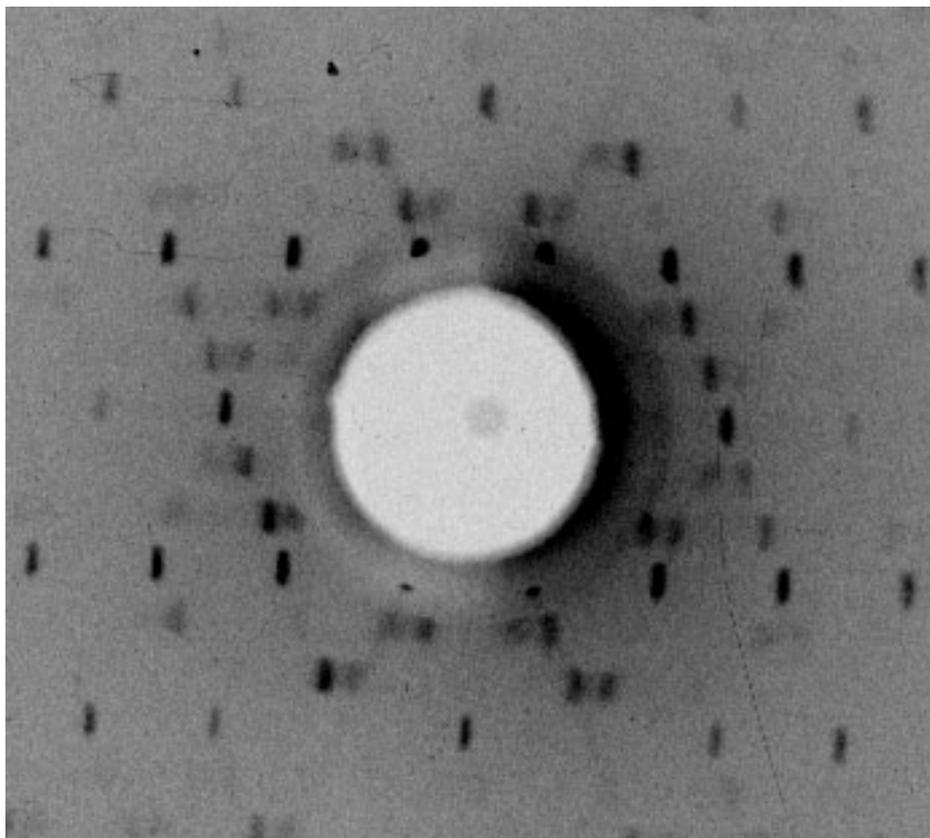


Figure A.2: Magnified area of the precession photograph shown in Figure A.1.

References

Acharya, R., Fry, E., Stuart, D., Fox, G., Rowlands, D. & Brown, F. (1989), “The Three-Dimensional Structure of Foot-and-Mouth Disease Virus at 2.9Å Resolution”, *Nature*, **337**, 709–716.

Åkervall, K., *et al*, & Moring, I. (1971), “X-ray Diffraction Studies of the Structure of Satellite Tobacco Necrosis Virus”, *Cold Spring Harbor Symp. Quant. Biol.*, **36**, 469–483.

Akey, C.W. & Edelstein, S.J. (1983), “Equivalence of the Projected Structure of Thin Catalase Crystals Preserved for Electron Microscopy by Negative Stain, Glucose or Embedding in the Presence of Tannic Acid”, *J. Mol. Biol.*, **163**, 575–612.

Alder, B.J. & Wainwright, T.E. (1959), “Studies in Molecular Dynamics. I. General Method.”, *J. Chem. Phys.*, **31**, 459–466.

Almassy, R.J., Janson, C.A., Hamlin, R., Xuong, N-H. & Eisenberg, D. (1986), “Novel subunit-subunit interactions in the structure of glutamine synthetase.”, *Nature*, **323**, 304–309.

Amos, L.A. & Klug, A. (1974), "Arrangement of subunits in flagellar microtubules", *J. Cell Sci.*, **14**, 523–549.

Argos, P., Ford, G.C. & Rossmann, M.G. (1975), "An Application of the Molecular Replacement Technique in Direct Space to a Known Protein Structure", *Acta Cryst.*, **A31**, 499–506.

Argos, P. & Rossmann, M.G. (1974), "Determining Heavy-Atom Positions using Non-Crystallographic Symmetry", *Acta Cryst.*, **A30**, 672–677.

Argos, P. & Rossmann, M.G. (1976), "A Method to Determine Heavy-Atom Positions for Virus Structures", *Acta Cryst.*, **B32**, 2975–2979.

Arndt, W. & Wonacott, A.J. (1977), "The Rotation Method in Crystallography", North Holland, Amsterdam.

Arnold, E., *et al*, & Rossmann, M.G. (1987), "The Structure Determination of a Common Cold Virus, Human Rhinovirus 14", *Acta Cryst.*, **A43**, 346–361.

Baumberg, S. & Harwood, C.R. (1979), "Carbon and Nitrogen Repression of Arginine Catabolic Enzymes in *Bacillus subtilis*", *J. Bacteriol.*, **137**, 189–196.

Baumberg, S. & Klingel, U. (1993), "Biosynthesis of Arginine, Proline, and Related Compounds" in "*Bacillus subtilis* and Other Gram-Positive Bacteria. Biochemistry, Physiology, and Molecular Genetics" edited by Sonenshein, A.L., Hoch, J.A. & Losick, R., Washington D.C. : American Society for Microbiology, 299–306.

Baumberg, S. & Mountain, A. (1984), "*Bacillus subtilis* 168 Mutants Resistant to Arginine Hydroxamate in the Presence of Ornithine or Citrulline", *J. Gen. Microbiol.*, **130**, 1247–1252.

Berg, J.M. (1993), "Zinc-finger proteins", *Curr. Opin. Struct. Biol.*, **3**, 11–16.

Bloomer, A.C., Champness, J.N., Bricogne, G., Staden, R. & Klug, A. (1978), "Protein Disk of Tobacco Mosaic Virus at 2.8Å Resolution Showing the Interactions Within and Between Subunits", *Nature*, **276**, 362–368.

Blow, D.M. & Crick, F.H.C. (1959), "The Treatment of Errors in the Isomorphous Replacement Method", *Acta Cryst.*, **12**, 794–802.

Blundell, T.L. & Johnson, L.N. (1976), "Protein Crystallography", Academic Press, London.

Boyes-Watson, J. & Perutz, M.F. (1943), "X-ray Analysis of Haemoglobin", *Nature*, **151**, 714–716.

Boys, C.W.G, Czaplowski, L.G., Phillips, S.E.V., Baumberg, S. & Stockley, P.G. (1990), "Crystallization of the Arginine-dependent Repressor/Activator AhrC from *Bacillus subtilis*", *J. Mol. Biol.*, **213**, 227–228.

Bragg, W.L. & Perutz, M.F. (1952), "The External Form of the Haemoglobin Molecule. I", *Acta Cryst.*, **5**, 277–283.

Bricogne, G. (1974), "Geometric Sources of Redundancy in Intensity Data and Their Use for Phase Determination", *Acta Cryst.*, **A30**, 395–405.

Bricogne, G. (1976), "Methods and Programs for Direct-Space Exploitation of Geometric Redundancies", *Acta Cryst.*, **A32**, 832–847.

Brisson, A. & Unwin, P.N.T. (1985), "Quaternary structure of the acetylcholine receptor", *Nature*, **315**, 474–477.

Brünger, A.T. (1990), “Extension of Molecular Replacement : A New Search Strategy based on Patterson Correlation Refinement”, *Acta Cryst.*, **A46**, 46–57.

Brünger, A.T. (1992a), “Free R value : a novel statistical quantity for assessing the accuracy of crystal structures”, *Nature*, **355**, 472–475.

Brünger, A.T. (1992b), “Molecular Replacement with X-PLOR : PC-refinement and free R value”, in the Proceedings of the CCP4 Study Weekend, January 31st, February 1st, 1992, pp 49–61.

Brünger, A.T. (1992c), “X-PLOR Version 3.1. A System for X-ray Crystallography and NMR”, Yale University Press, New Haven and London.

Brünger, A.T. & Nilges, M. (1993), “Computational Challenges for Macromolecular Structure Determination by X-ray Crystallography and Solution NMR-Spectroscopy”, *Quart. Rev. Biophys.*, **26**, 49–125.

Buerger, M.J. (1951), “A New Approach to Crystal-Structure Analysis”, *Acta Cryst.*, **4**, 531–544.

Buerger, M.J. (1959), “Vector Space and its Application in Crystal-Structure Investigation”, John Wiley & Sons, New York.

Buerger, M.J. (1964), “Image Methods in Crystal-Structure Analysis” in “Advanced Methods of Crystallography” edited by Ramachandran, G.N., Academic Press, London and New York.

Burke, M., Merican, A.F. & Sherratt, D.J. (1994), “Mutant *Escherichia coli* arginine repressor proteins that fail to bind L-Arginine, yet retain the ability to bind their normal DNA-binding sites”, *Mol. Microbiol.*, **13**, 609–618.

Burley, S.K. (1994), "DNA-binding motifs from eukaryotic transcription factors", *Curr. Opin. Struct. Biol.*, **4**, 3–11.

Ceska, T.A. & Henderson, R. (1990), "Analysis of High-resolution Electron Diffraction Patterns from Purple Membrane Labelled with Heavy-atoms", *J. Mol. Biol.*, **213**, 539–560.

Champness, J.N., Bloomer, A.C., Bricogne, G., Butler, P.J.G. & Klug, A. (1976), "The Structure of the Protein Disk of Tobacco Mosaic Virus to 5Å Resolution", *Nature*, **259**, 20–24.

Cheng, R.H., *et al.* & Johnson, J.E. (1994), "Functional implications of quasi-equivalence in a T=3 icosahedral animal virus established by cryo-electron microscopy and X-ray crystallography", *Structure*, **2**, 271–282.

Chothia, C. & Lesk, A.M. (1986), "The relation between the divergence of sequence and structure in proteins", *EMBO J.*, **5**, 823–826.

Cochran, W. (1952), "A Relation between the Signs of Structure Factors", *Acta Cryst.*, **5**, 65–67.

Cochran, W. (1954), "The Determination of Signs of Structure Factors from the Intensities", *Acta Cryst.*, **7**, 581–583.

Cochran, W. (1955), "Relations between the Phases of Structure Factors", *Acta Cryst.*, **8**, 473–478.

Cochran, W. & Douglas, A.S. (1953), "A New Application of EDSAC to Crystal Structure Analysis", *Nature*, **171**, 1112–1113.

Cochran, W. & Douglas, A.S. (1955), "The Use of a High-Speed Digital Com-

puter for the Direct Determination of Crystal Structures. I", *Proc. Roy. Soc. Lond.*, **A227**, 486–500.

Cochran, W. & Woolfson, M.M. (1954), "Have Hauptman & Karle solved the phase problem ?", *Acta Cryst.*, **7**, 450–451.

Cochran, W. & Woolfson, M.M. (1955), "The Theory of Sign Relations Between Structure Factors", *Acta Cryst.*, **8**, 1–12.

Cohen, C., Caspar, D.L.D., Parry, D.A.D. & Lucas, R.M. (1971), "Tropomyosin Crystal Dynamics", *Cold Spring Harbor Symp. Quant. Biol.*, **36**, 205–216.

Collaborative Computational Project, Number 4 (1994), "The CCP4 Suite : Programs for Protein Crystallography", *Acta Cryst.*, **D50**, 760–763.

Corey, R.B., Stanford, R.H., Marsh, R.E. ,Leung, Y.C. & Kay, L.M. (1962), "An X-ray Investigation of Wet Lysozyme Chloride Crystals. Preliminary Report on Crystals Containing Complex Ions of Niobium and Tantalum.", *Acta Cryst.*, **15**, 1157–1163.

Crick, F.H.C. & Magdoff, B.S. (1956), "The Theory of the Method of Isomorphous Replacement for Protein Crystals. I", *Acta Cryst.*, **9**, 901–908.

Crowther, R.A. (1967), "A Linear Analysis of the Non-Crystallographic Symmetry Problem", *Acta Cryst.*, **22**, 758–764.

Crowther, R.A. (1969), "The Use of Non-Crystallographic Symmetry for Phase Determination", *Acta Cryst.*, **B25**, 2571–2580.

Crowther, R.A. (1972), "The Fast Rotation Function" in "The Molecular Replacement Method" Edited by Rossmann, M.G., Gordon & Breach, New York.

Crowther, R.A. & Amos, L.A. (1971), “Harmonic Analysis of Electron Microscope Images with Rotational Symmetry”, *J. Mol. Biol.*, **60**, 123–130.

Crowther, R.A., Amos, L.A., Finch, J.T., De Rosier, D.J. & Klug, A. (1970), “Three Dimensional Reconstructions of Spherical Viruses by Fourier Synthesis from Electron Micrographs”, *Nature*, **226**, 421–425.

Crowther, R.A. & Blow, D.M. (1967), “A Method for Positioning a Known Molecule in an Unknown Crystal Structure”, *Acta Cryst.*, **23**, 544–548.

Crowther, R.A., De Rosier, D.J. & Klug, A. (1970), “The reconstruction of a three-dimensional structure from projections and its application to electron microscopy”, *Proc. Roy. Soc. Lond.*, **A317**, 319–340.

Crowther, R.A. & Klug, A. (1975), “Structural Analysis of Macromolecular Assemblies by Image Reconstruction from Electron Micrographs”, *Ann. Rev. Biochem.*, **44**, 161–182.

Cullis, A.F., Muirhead, H., Perutz, M.F., Rossmann, M.G. & North, A.C.T. (1961), “The Structure of Haemoglobin : VIII. A Three-Dimensional Fourier Synthesis at 5.5Å Resolution : Determination of the Phase Angles”, *Proc. Roy. Soc. Lond.*, **A265**, 15–38.

Czaplewski, L.G., North, A.K., Smith, M.C.M., Baumberg, S. & Stockley, P.G. (1992), “Purification and initial characterization of AhrC : the regulator of arginine metabolism genes in *Bacillus subtilis*”, *Mol. Microbiol.*, **6**, 267–275.

David, P.R. & Subbiah, S. (1994), “Low-resolution Real-Space Envelopes : the Application of the Condensing-Protocol Approach to the *ab initio* Macromolecular Phase Problem of a Variety of Examples”, *Acta Cryst.*, **D50**, 132–138.

De Rosier, D.J. (1971), "Three-dimensional image reconstruction of helical structures", *Phil. Trans. Roy. Soc. Lond.*, **B261**, 209–210.

De Rosier, D.J. & Klug, A. (1968), "Reconstruction of Three Dimensional Structures from Electron Micrographs", *Nature*, **217**, 130–134.

De Rosier, D.J. & Klug, A. (1971), "Structure of the Tubular Variants of the Head of Bacteriophage T4 (Polyheads)", *J. Mol. Biol.*, **65**, 469–488.

De Rosier, D.J. & Moore, P.B. (1970), "Reconstruction of Three-dimensional Images from Electron Micrographs of Structures with Helical Symmetry", *J. Mol. Biol.*, **52**, 355–369.

De Rosier, D.J. & Oliver, R.M. (1971), "A Low Resolution Electron-Density Map of Lipoyl Transsuccinylase, the Core of the α -Ketoglutarate Dehydrogenase Complex", *Cold Spring Harbor Symp. Quant. Biol.*, **36**, 199–203.

Dickerson, R.E., Kopka, M.L., Varnum, J.C. & Weinzierl, J.E. (1967), "Bias, Feedback and Reliability in Isomorphous Phase Analysis", *Acta Cryst.*, **23**, 511–522.

Donohue, J. & Bryden, J.H. (1955), "The Interpretation of Patterson Functions : An Application of the Superposition Method", *Acta Cryst.*, **8**, 314–316.

Ducruix, A. & Giegé, R. (1992), "Crystallization of Nucleic Acids & Proteins. A Practical Approach.", Oxford University Press, Oxford.

Eliopoulos, E., Geddes, A.J., Brett, M., Pappin, D.J.C., Findlay, J.B.C. (1982), "A structural model for the chromophore-binding domain of ovine rhodopsin", *Int. J. Biol. Macromol.*, **4**, 263–268.

Ellenberger, T. (1994), "Getting a grip on DNA recognition : structures of

the basic region leucine zipper, and the basic region helix-loop-helix DNA-binding domains”, *Curr. Opin. Struct. Biol.*, **4**, 12-21.

Erickson, H.P. & Klug, A. (1971), “Measurement and compensation of defocusing and aberrations by Fourier processing of electron micrographs”, *Phil. Trans. Roy. Soc. Lond.*, **B261**, 105–118.

Finch, J.T., Gilbert, P.F.C, Klug, A. & Leberman, R. (1974), “X-ray Analysis of the Disk of Tobacco Mosaic Virus Protein”, *J. Mol. Biol.*, **86**, 183–192.

Finch, J.T. & Klug, A. (1971), “Three-dimensional reconstruction of the stacked-disk aggregate of tobacco mosaic virus protein from electron micrographs”, *Phil. Trans. Roy. Soc. Lond.*, **B261**, 211–219.

Finch, J.T., Klug, A. & Stretton, A.O.W. (1964), “The structure of the “Polyheads” of T4 Bacteriophage”, *J. Mol. Biol.*, **10**, 570–575.

Finch, J.T., Lutter, L.C., Rhodes, D., Brown, R.S., Rushton, B., Levitt, M. & Klug, A. (1977), “Structure of nucleosome core particles of chromatin”, *Nature*, **269**, 29–36.

Freemont, P.S., Lane, A.N. & Sanderson, M.R. (1991), “Structural aspects of protein-DNA recognition”, *Biochem. J.*, **278**, 1–23.

Fridrichsons, J. & Mathieson, A.McL. (1962), “Image-seeking. A Brief Study of Its Scope and Comments on Certain Limitations”, *Acta Cryst.*, **15**, 1065–1074.

Fry, E., Acharya, R. & Stuart, D. (1993), “Methods Used in the Structure Determination of Foot-and-Mouth Disease Virus”, *Acta Cryst.*, **A49**, 45–55.

Fujinaga, M. & Read, R.J. (1987), “Experiences with a New Translation-

Function Program”, *J. Appl. Cryst.*, **20**, 517–521.

Furcinitti, P.S., Oostrum, J. & Burnett, R.M. (1989), “Adenovirus polypeptide IX revealed as capsid cement by difference images from electron microscopy and crystallography”, *EMBO J.*, **8**, 3563–3570.

Gaykema, W.P.J., *et al.* & Beintema, J.J. (1984), “3.2Å Structure of the Copper-containing Oxygen-carrying Protein *Panulirus interruptus* haemocyanin”, *Nature*, **309**, 23–29.

Germain, G., Main, P. & Woolfson, M.M. (1970), “On the Application of Phase Relationships to Complex Structures II. Getting a Good Start”, *Acta Cryst.*, **B26**, 274–285.

Germain, G. & Woolfson, M.M. (1966), “Some Ideas on the Deconvolution of the Patterson Function”, *Acta Cryst.*, **21**, 845–848.

Germain, G. & Woolfson, M.M. (1968), “On the Application of Phase Relationships to Complex Structures”, *Acta Cryst.*, **B24**, 91–96.

Goodsell, D.S., Kaczor-Grzeskowiak, M. & Dickerson, R.E. (1994), “The Crystal Structure of C-C-A-T-T-A-A-T-G-G. Implications for Bending of B-DNA at T-A Steps”, *J. Mol. Biol.*, **239**, 79–96.

Grant, R.A., Schmid, M.F. & Chiu, W. (1991), “Analysis of Symmetry and Three-dimensional Reconstruction of Thin gp32*I Crystals”, *J. Mol. Biol.*, **217**, 551–562.

Hai-Fu, F., Jia-Xing, Main, P. & Woolfson, M.M. (1983), “On the Application of Phase Relationships to Complex Structures. XXIII. Automatic Determination of Crystal Structures having Pseudo-Translational Symmetry by a Modified

MULTAN Procedure”, *Acta Cryst.*, **A39**, 566–569.

Hammersley, J.M. & Handscomb, D.C. (1964), “Monte Carlo Methods”, London : Methuen & Co Ltd, New York : John Wiley & Sons, Inc.

Hamodrakas, S.J., Margaritis, L.H. & Nixon, P.E. (1982), “Crystalline layer in *Drosophila melanogaster* eggshell : arrangement of components as revealed by negative staining and reconstruction”, *Int. J. Biol. Macromol.*, **4**, 25–31.

Harada, Y., Lifchitz, A., Berthou, J. & Jolles, P. (1981), “A Translation Function Combining Packing and Diffraction Information : An Application to Lysozyme (High-Temperature Form)”, *Acta Cryst.*, **A37**, 398–406.

Harrison, S.C. (1991), “A structural taxonomy of DNA-binding domains”, *Nature*, **353**, 715–719.

Harrison, S.C. & Aggarwal, A.K. (1990), “DNA Recognition by Proteins with the Helix-Turn-Helix Motif”, *Annu. Rev. Biochem.*, **59**, 933–969.

Harrison, S.C., Olson, A.J., Schutt, C.E. & Winkler, F.K. & Bricogne, G. (1978), “Tomato Bushy Stunt Virus at 2.9Å Resolution”, *Nature*, **276**, 368–373.

Hart, R.G. (1961), “Refinement of heavy atom parameters other than relative y 's”, *Acta Cryst.*, **14**, 1194–1195.

Harwood, C.R. & Baumberg, S. (1977), “Arginine Hydroxamate-resistant Mutants of *Bacillus subtilis* with Altered Control of Arginine Metabolism”, *J. Gen. Microbiol.*, **100**, 177–188.

Hauptman, H. (1975), “A New Method in the Probabilistic Theory of the Structure Invariants”, *Acta Cryst.*, **A31**, 680–687.

Hauptman, H. & Karle, J. (1953), "The Solution of the Phase Problem : I. The Centrosymmetric Crystal", *Am. Crystallogr. Assoc. Monogr.*, No.3, Wilmington : The Letter Shop.

Henderson, R., Baldwin, J.M., Ceska, T.A., Zemlin, F., Beckmann, E. & Downing, K.H. (1990), "Model for the Structure of Bacteriorhodopsin Based on High-resolution Electron Cryo-microscopy", *J. Mol. Biol.*, **213**, 899–929.

Henderson, R. & Unwin, P.N.T (1975), "Three-dimensional model of purple membrane obtained by electron microscopy", *Nature*, **257**, 28–32.

Higgins, D.G., Bleasby, A.J. & Fuchs, R. (1992), "CLUSTAL-V : Improved Software for Multiple Sequence Alignment", *CABIOS*, **8**, 189–191.

Hodgkin, D.C. (1987), "Patterson and Pattersons" in "Patterson and Pattersons. Fifty years of the Patterson Function", Edited by Glusker, J.P., Patterson, B.K., & Rossi, M., Oxford University Press, Oxford.

Holzenburg, A., *et al*, & Ford, R.C. (1993), "Three-dimensional structure of photosystem II", *Nature*, **363**, 470–472.

Holzenburg, A., *et al*, & Saenger, W. (1987), "Structure of D-ribulose-1,5-bisphosphate carboxylase/oxygenase from *Alcaligenes eutrophus* H16", *Nature*, **325**, 730–732.

Hope, H. (1988), "Cryocrystallography of Biological Macromolecules : a Generally Applicable Method", *Acta Cryst.*, **B44**, 22–26.

Huber, R. (1985), "Experience with the Application of Patterson Search Techniques", in the Proceedings of the CCP4 Study Weekend 15–16 February, 1985, pp 58–61.

International Tables for X-ray Crystallography, Vol. I, (1952), Edited by Henry, N.F.M. & Lonsdale, K., Kynoch Press, Birmingham, England.

International Tables for X-ray Crystallography, Vol. IV, (1974), Edited by Ibers, J.A. & Hamilton, W.C., Kynoch Press, Birmingham, England.

Ito, N. (1991), "X-ray Crystallographic Studies on Galactose Oxidase", Ph. D. thesis, University of Leeds.

Jack, A. (1973), "Direct Determination of X-ray Phases for Tobacco Mosaic Virus Protein using Non-crystallographic Symmetry", *Acta Cryst.*, **A29**, 545–554.

Jeng, T.-W. & Chiu, W. (1983), "Low Dose Electron Microscopy of the Crotoxin Complex Thin Crystals", *J. Mol. Biol.*, **164**, 329–346.

Jeng, T.-W., Crowther, R.A., Stubbs, G. & Chiu, W. (1989), "Visualization of α -helices in Tobacco Mosaic Virus by Cryo-electron Microscopy", *J. Mol. Biol.*, **205**, 251–257.

Jones, Y. & Stuart, D. (1991), "Locating Heavy Atom Sites by Automatic Patterson Search — GROPAT" in the Proceedings of the CCP4 Study Weekend 25–26 January, 1991, pp 39–48.

Jones, Y.E., Walker, N.P.C. & Stuart, D.I. (1991), "Methodology Employed for the Structure Determination of Tumour Necrosis Factor, a Case of High Non-Crystallographic Symmetry", *Acta Cryst.*, **A47**, 753–770.

Kabsch, W. (1988), "Automatic Indexing of Rotation Diffraction Patterns", *J. Appl. Cryst.*, **21**, 67–71.

Kabsch, W. (1993), "Automatic Processing of Rotation Diffraction Data from Crystals of Initially Unknown Symmetry and Cell Constants", *J. Appl. Cryst.*, **26**, 795–800.

Karle, J. & Hauptman, H. (1950), "The Phases and Magnitudes of the Structure Factors", *Acta Cryst.*, **3**, 181–187.

Karle, J. & Hauptman, H. (1956), "A Theory of Phase Determination for the Four Types on Non-Centrosymmetric Space Groups $1P222$, $2P22$, $3P_12$, $3P_22$ ", *Acta Cryst.*, **9**, 635–651.

Karle, I.L. & Karle, J. (1963), "An Application of a New Phase Determination Procedure to the Structure of Cyclo(hexaglycyl) Hemihydrate", *Acta Cryst.*, **16**, 969–975.

Karle, I.L. & Karle, J. (1964), "An Application of the Symbolic Addition Method to the Structure of L-Arginine Dihydrate", *Acta Cryst.*, **17**, 835–841.

Karle, J. & Karle, I.L. (1966), "The Symbolic Addition Procedure for Phase Determination for Centrosymmetric and Noncentrosymmetric Crystals", *Acta Cryst.*, **21**, 849–859.

Keen, D.A. & McGreevy, R.L. (1990), "Structural modelling of glasses using reverse Monte Carlo simulation", *Nature*, **344**, 423–425.

Khachatryan, A., Semenovskaya, S. & Vainshtein, B. (1981), "The Thermodynamic Approach to the Structure Analysis of Crystals", *Acta Cryst.*, **A37**, 742–754.

Klug, A. (1958), "Joint Probability Distributions of Structure Factors and the Phase Problem", *Acta Cryst.*, **11**, 515–543.

Klug, A. (1971), "Interpretation of the Rotation Function Map of Satellite Tobacco Necrosis Virus : Octahedral Packing of Icosahedral Particles", *Cold Spring Harbor Symp. Quant. Biol.*, **36**, 483–486.

Klug, A. (1983), "From Macromolecules to Biological Assemblies (Nobel Lecture)", *Angew. Chem. Int. Ed. Engl.*, **22**, 565–582.

Klug, A. & Berger, J.E. (1964), "An Optical Method for the Analysis of Periodicities in Electron Micrographs, and some Observations on the Mechanism of Negative Staining", *J. Mol. Biol.*, **10**, 565–569.

Klug, A. & Crowther, R.A. (1972), "Three-dimensional Image Reconstruction from the Viewpoint of Information Theory", *Nature*, **238**, 435–440.

Klug, A. & De Rosier, D.J. (1966), "Optical Filtering of Electron Micrographs : Reconstruction of One-Sided Images", *Nature*, **212**, 29–32.

Klug, A., Rhodes, D., Smith, J., Finch, J.T. & Thomas, J.O. (1980), "A low resolution structure for the histone core of the nucleosome", *Nature*, **287**, 509–516.

Kraulis, P.J. (1991), "MOLSCRIPT : A Program to Produce Both Detailed and Schematic Plots of Protein Structures", *J. Appl. Cryst.*, **24**, 946–950.

Krieger, W. & Schenk, H. (1973), "On the Reliability of the Σ_2 Relation. II. Artificial Structures in $P2_1/c$ ", *Acta Cryst.*, **A29**, 720–723.

Kühlbrandt, W., Wang, Da-N. & Fujiyoshi, Y. (1994), "Atomic model of plant light-harvesting complex by electron crystallography", *Nature*, **367**, 614–621.

Labaw, L.W. & Davies, D.R. (1972), "The Molecular Outline of Human

γ GI Immunoglobulin from an EM Study of Crystals”, *J. Ultrastruct. Res.*, **40**, 349–365.

Langer, R., Poppe, C., Schramm, H.J. & Hoppe, W. (1975), “Electron Microscopy of Thin Protein Crystal Sections”, *J. Mol. Biol.*, **93**, 159–165.

Lawrence, M.C. (1991), “The Application of the Molecular Replacement Method to the *de novo* Determination of Protein Structure”, *Quart. Rev. Biophys.*, **24**, 399–424.

Leslie, A.G.W. (1991), “Heavy Atom Derivative Screening” in the Proceedings of the CCP4 Study Weekend, 25–26 January, 1991, pp 9–22.

Lim, D., Oppenheim, J.D., Eckhardt, T. & Maas, W.K. (1987), “Nucleotide sequence of the *argR* gene of *Escherichia coli* K-12 and isolation of its product, the arginine repressor”, *Proc. Natl. Acad. Sci. USA*, **84**, 6697–6701.

Luo, M., *et al.*, & Palmenberg, A.C. (1987), “The Atomic Structure of Mengo Virus at 3.0Å Resolution”, *Science*, **235**, 182–191.

Luo, M., Vriend, G., Kamer, G. & Rossmann, M.G. (1989), “Structure Determination of Mengo Virus”, *Acta Cryst.*, **B45**, 85–92.

Main, P. (1967), “Phase Determination Using Non-Crystallographic Symmetry”, *Acta Cryst.*, **23**, 50–54.

Main, P. (1985), “MULTAN — A Program for the Determination of Crystal Structures” in “Crystallographic Computing 3 : Data Collection, Structure Determination, Proteins, and Databases”, Edited by Sheldrick, G.M., Krüger, C. & Goddard, R., Oxford University Press, Oxford.

Main, P. & Rossmann, M.G. (1966), "Relationships among Structure Factors due to Identical Molecules in Different Crystallographic Environments", *Acta Cryst.*, **21**, 67–72.

Marsh, D.J. & Petsko, G.A. (1973), "A Low-Temperature Device for Protein Crystallography", *J. Appl. Cryst.*, **6**, 76–80.

McPherson, A. (1982), "Preparation and analysis of protein crystals", John Wiley & Sons, New York.

McPherson, A. & Rich, A. (1973), "Preliminary Study of *B. subtilis* α -Amylase Crystals by Electron Microscopy and Optical Diffraction", *J. Ultrastruct. Res.*, **44**, 75–84.

Mellema, J.E. & Klug, A. (1972), "Quaternary Structure of Gastropod Haemocyanin", *Nature*, **239**, 146–150.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. & Teller, E. (1953), "Equation of State Calculations by Fast Computing Machines", *J. Chem. Phys.*, **21**, 1087–1092.

Misell, D.L. (1978), "Image Analysis, Enhancement and Interpretation", Vol. 7 in "Practical Methods in Electron Microscopy" edited by Glauert, A.M., North-Holland Publishing Company, Oxford.

Morgan, D.G. & De Rosier, D. (1992), "Processing images of helical structures : a new twist", *Ultramicroscopy*, **46**, 263–285.

Mountain, A. & Baumberg, S. (1980), "Map Locations of Some Mutations Conferring Resistance to Arginine Hydroxamate in *Bacillus subtilis* 168", *Mol. Gen. Genet.*, **178**, 691–701.

Mountain, A., Mann, N.H., Munton, R.N. & Baumberg, S. (1984), "Cloning of a *Bacillus subtilis* restriction fragment complementing auxotrophic mutants of eight *Escherichia coli* genes of arginine biosynthesis", *Mol. Gen. Genet.*, **197**, 82–89.

Mountain, A., McChesney, J., Smith, M.C.M. & Baumberg, S. (1986), "Gene Sequence Encoding Early Enzymes of Arginine Synthesis within a Cluster in *Bacillus subtilis*, as Revealed by Cloning in *Escherichia coli*", *J. Bacteriol.*, **165**, 1026–1028.

Navaza, J. (1987), "On the Fast Rotation Function", *Acta Cryst.*, **A43**, 645–653.

Navaza, J. (1992), "AMoRe : A New Package for Molecular Replacement", in the Proceedings of the CCP4 Study Weekend, January 31st, February 1st, 1992, pp 20–32.

Navia, M.A. & Sigler, P.B. (1974), "The Application of Direct Methods to the Analysis of Heavy-Atom Derivatives", *Acta Cryst.*, **A30**, 706–712.

Neidle, S. (1973), "On the Determination of Heavy-Atom Positions in various Elastase Derivatives", *Acta Cryst.*, **B29**, 2645–2647.

Nixon, P.E. (1978), "Overlapping Patterson peaks and Direct Methods : The Structure of Prostratin", *Acta Cryst.*, **A34**, 450–453.

Nixon, P.E. & North, A.C.T. (1976), "Crystallographic Relationship between Human and Hen-Egg Lysozyme. I. Methods for the Establishment of Molecular Orientational and Positional Parameters", *Acta Cryst.*, **A32**, 320–325.

Nordman, C.E. (1980), "Procedures for Detection and Idealization of Non-

crystallographic Symmetry with Application to Phase Refinement of the Satellite Tobacco Necrosis Virus Structure”, *Acta Cryst.*, **A36**, 747–754.

North, A.K., Smith, M.C.M. & Baumberg, S. (1989), “Nucleotide sequence of a *Bacillus subtilis* arginine regulatory gene and homology of its product to the *Escherichia coli* arginine repressor”, *Gene*, **80**, 29–38.

Pabo, C.O. & Sauer, R.T. (1992), “Transcription Factors : Structural Families and Principles of DNA recognition”, *Annu. Rev. Biochem.*, **61**, 1053–1095.

Parsons, M.R. (1988), “Crystallographic Studies of L-Asparaginase and the Structure of Turkey Egg Lysozyme”, Ph. D. thesis, University of Leeds.

Patterson, A.L. (1935), “A Direct Method for the Determination of the Components of Interatomic Distances in Crystals”, *Z. Krist.*, **A90**, 517–542.

Petsko, G.A. (1975), “Protein Crystallography at Sub-zero Temperatures : Cryo-protective Mother Liquors for Protein Crystals”, *J. Mol. Biol.*, **96**, 381–392.

Phillips, S.E.V. (1994), “The β -Ribbon DNA Recognition Motif”, *Annu. Rev. Biophys. Biomol. Struct.*, **23**, 671–701.

Ramachandran, G.N. & Srinivasan, R. (1970), “Fourier Methods in Crystallography”, John Wiley & Sons, New York.

Rayment, I. (1983), “Molecular Replacement Method at Low Resolution : Optimum Strategy and Intrinsic Limitations as Determined by Calculations on Icosahedral Virus Models”, *Acta Cryst.*, **A39**, 102–116.

Rayment, I., Baker, T.S. & Caspar, D.L.D. (1983), “A Description of the Techniques and Application of Molecular Replacement Used to Determine the

Structure of Polyoma Virus Capsid at 22.5Å Resolution”, *Acta Cryst.*, **B39**, 505–516.

Rayment, I., Baker, T.S. & Caspar, D.L.D. & Murakami, W.T. (1982), “Polyoma Virus Capsid Structure at 22.5Å Resolution”, *Nature*, **295**, 110–115.

Rayment, I., *et al.*, & Johnson, J.E. (1978), “An 11Å Resolution Electron Density Map of Southern Bean Mosaic Virus”, *Acta Cryst.*, **B34**, 567–578.

Richardson, J.S. (1985), “Schematic Drawings of Protein Structures”, in “Methods in Enzymology”, Vol. 115, edited by Wyckoff, H.W., Hirs, C.H.W. & Timashoff, S.N., Academic Press, London.

Richardson, J.W., Jr. & Jacobson, R.A. (1987), “Computer-aided Analysis of Multi-solution Patterson Superpositions” in “Patterson and Pattersons. Fifty years of the Patterson Function”, Edited by Glusker, J.P., Patterson, B.K., & Rossi, M., Oxford University Press, Oxford.

Richmond, T.J., Finch, J.T., Rushton, B., Rhodes, D. & Klug, A. (1984), “Structure of the nucleosome core particle at 7Å resolution”, *Nature*, **311**, 532–537.

Robinson, W. & Sheldrick, G.M. (1988), “SHELX” in “Crystallographic Computing 4 : Techniques and New Technologies”, Edited by Isaacs, N.W. & Taylor, M.R., Oxford University Press, Oxford.

Rosenbluth, M.N. & Rosenbluth, A.W. (1954), “Further results on Monte Carlo Equations of State”, *J. Chem. Phys.*, **22**, 881–884.

Rossmann, M.G. (1972), “The Molecular Replacement Method”, Gordon & Breach, New York.

Rossmann, M.G. (1990), "The Molecular Replacement Method", *Acta Cryst.*, **A46**, 73–82.

Rossmann, M.G. & Blow, D.M. (1962), "The Detection of Sub-Units Within the Crystallographic Asymmetric Unit", *Acta Cryst.*, **15**, 24–31.

Rossmann, M.G. & Blow, D.M. (1963), "Determination of Phases by the Conditions of Non-Crystallographic Symmetry", *Acta Cryst.*, **16**, 39–45.

Rossmann, M.G. & Blow, D.M. (1964), "Solution of the Phase Equations Representing Non-Crystallographic Symmetry", *Acta Cryst.*, **17**, 1474–1475.

Rossmann, M.G., *et al.*, & Choi, H.-K. & Lynch, R.E. (1992), "Molecular Replacement Real-Space Averaging", *J. Appl. Cryst.*, **25**, 166–180.

Rossmann, M.G., *et al.*, & Vriend, G. (1985), "Structure of a Human Common Cold Virus and Functional Relationship to other Picornaviruses", *Nature*, **317**, 145–153.

Sayre, D. (1952), "The Squaring Method : A New Method for Phase Determination", *Acta Cryst.*, **5**, 60–65.

Schägger, H. & Jagow, G. (1987), "Tricine-Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis for the Separation of Proteins in the Range from 1 to 100 kDa", *Anal. Biochem.*, **166**, 368–379.

Schenk, H. (1973), "On the Reliability of the Σ_2 Relation. I. Real Structures in $P2_1/c$ ", *Acta Cryst.*, **A29**, 503–507.

Schneider, G. & Lindqvist, Y. (1994), "Ta₆Br₁₄ is a Useful Cluster Compound for Isomorphous Replacement in Protein Crystallography", *Acta Cryst.*, **D50**, 186–191.

Schultz, P., Célia, H., Riva, M., Sentenac, A. & Oudet, P. (1993), "Three-dimensional model of yeast RNA polymerase I determined by electron microscopy of two-dimensional crystals", *EMBO J.*, **12**, 2601–2607.

Seeman, N.C., Rosenberg, J.M. & Rich, A. (1976), "Sequence-specific recognition of double helical nucleic acids by proteins", *Proc. Natl. Acad. Sci. USA*, **73**, 804–808.

Semenovskaya, S.V., Khachatryan, K.A. & Khachatryan, A.G. (1985), "Statistical Mechanics Approach to the Structure Determination of a Crystal", *Acta Cryst.*, **A41**, 268–273.

Sheldrick, G.M. (1986), Manual for SHELX–86.

Sheldrick, G.M. (1985), "SHELX-84" in "Crystallographic Computing 3", Edited by Sheldrick, G.M., Krüger, C. & Goddard, R., Oxford University Press, Oxford.

Sheldrick, G.M. (1991), "Heavy Atom Location using SHELXS-90" in the Proceedings of the CCP4 Study Weekend 25–26 January, 1991, pp 23–38.

Shelley, K. & McPherson, A. (1980), "Spatially filtered images of *B. subtilis* α -amylase crystals", *J. Microscopy*, **121**, 201–210.

Shoemaker, D.P., Barieau, R.E., Donohue, J. & Lu, C.-S. (1953), "The Crystal Structure of DL-Serine", *Acta Cryst.*, **6**, 241–256.

Simpson, P.G., Dobrott, R.D. & Lipscomb, W.N. (1965), "The Symmetry Minimum Function : High Order Image Seeking Functions in X-ray Crystallography", *Acta Cryst.*, **18**, 169–179.

Singh, T.P., Bode, W. & Huber, R. (1980), "Low-Temperature Protein Crystallography. Effect on Flexibility, Temperature Factor, Mosaic Spread, Extinction and Diffuse Scattering in Two Examples : Bovine Trypsinogen and Fc Fragment", *Acta Cryst.*, **B36**, 621–627.

Smith, M.C.M., Czaplowski, L., North, A.K., Baumberg, S. & Stockley, P.G. (1989), "Sequences required for regulation of arginine biosynthesis promoters are conserved between *Bacillus subtilis* and *Escherichia coli*", *Mol. Microbiol.*, **3**, 23–28.

Smith, M.C.M., Mountain, A. & Baumberg, S. (1986a), "Cloning in *Escherichia coli* of a *Bacillus subtilis* arginine repressor gene through its ability to confer structural stability on a fragment carrying genes of arginine biosynthesis", *Mol. Gen. Genet.*, **205**, 176–182.

Smith, M.C.M., Mountain, A. & Baumberg, S. (1986b), "Sequence analysis of the *Bacillus subtilis* *argC* promoter region", *Gene*, **49**, 53–60.

Smith, P.R., Aebi, U., Josephs, R. & Kessel, M. (1976), "Studies of the Structure of the T4 Bacteriophage Tail Sheath", *J. Mol. Biol.*, **106**, 243–275.

Stanford, R.H., Marsh, R.E. & Corey, R.B. (1962), "An X-Ray Investigation of Lysozyme Chloride Crystals Containing Complex Ions of Niobium and Tantalum : Three-dimensional Fourier Plot Obtained from Data Extending to a Minimum Spacing of 5Å.", *Nature*, **196**, 1176–1178.

Stanley, E. (1979), "The Direct Determination of Centrosymmetric Electron

Density Distributions by Maximizing the Integrated Cube of the Electron Density”, *Acta Cryst.*, **A35**, 966–970.

Stanley, E. (1986), “‘Peakiness’ Test Functions”, *Acta Cryst.*, **A42**, 297–299.

Steitz, T.A. (1968), “A New Method of Locating Heavy Atoms Bound to Protein Crystals”, *Acta Cryst.*, **B24**, 504–507.

Steitz, T.A. (1990), “Structural studies of protein-nucleic acid interaction : the sources of sequence-specific binding”, *Q. Rev. Biophys.*, **23**, 205–280.

Stewart, M. (1988a), “Introduction to the Computer Image Processing of Electron Micrographs of Two-Dimensionally Ordered Biological Structures”, *J. Electron Microsc. Tech.*, **9**, 301–324.

Stewart, M. (1988b), “Computer Image Processing of Electron Micrographs of Biological Structures with Helical Symmetry”, *J. Electron Microsc. Tech.*, **9**, 325–358.

Stewart, P.L., Fuller, S.D. & Burnett, R.M. (1993), “Difference imaging of adenovirus : bridging the resolution gap between X-ray crystallography and electron microscopy”, *EMBO J.*, **12**, 2589–2599.

Stirling, C.J., Szatmari, G., Stewart, G., Smith, M.C.M. & Sherratt, D.J. (1988), “The arginine repressor is essential for plasmid-stabilizing site-specific recombination at the ColE1 *cer* locus”, *EMBO J.*, **7**, 4389–4395.

Stokes, D.L. & Green, N.M. (1990), “Structure of CaATPase : Electron Microscopy of Frozen-hydrated Crystals at 6Å Resolution in Projection”, *J. Mol. Biol.*, **213**, 529–538.

Stoops, J.K. *et al*, & Hackert, M.L. (1991), “Comparisons of the Low-Resolution Structures of Ornithine Decarboxylase by Electron Microscopy and X-ray Crystallography : The Utility of Methylamine Tungstate Stain and Butvar Support Film in the Study of Macromolecules by Transmission Electron Microscopy”, *J. Electron Microsc. Tech.*, **18**, 157–166.

Subbiah, S. (1991), “Low-Resolution Real-Space Envelopes : An Approach to the Ab Initio Macromolecular Phase Problem”, *Science*, **252**, 128–133.

Subbiah, S. (1993), “Low-resolution Real-Space Envelopes : Improvements to the Condensing Protocol Approach and a New Method to Fix the Sign of Such Envelopes”, *Acta Cryst.*, **D49**, 108–119.

Terwilliger, T.C., Kim, S.-H. & Eisenberg, D. (1987), “Generalized Method of Determining Heavy-Atom Positions Using the Difference Patterson Function”, *Acta Cryst.*, **A43**, 1–5.

Tête-Favier, F., Rondeau, J.-M., Podjarny, A. & Moras, D. (1993), “Structure Determination of Aldose Reductase : Joys and Traps of Local Symmetry Averaging”, *Acta Cryst.*, **D49**, 246–256.

Tian, G.L. & Maas, W.K. (1994), “Mutational Analysis of the Arginine Repressor of *Escherichia coli*”, *Mol. Microbiol.*, **13**, 599–608.

Tickle, I.J. (1985), “Review of Space Group General Translation Functions that Make Use of Known Structure Information and Can Be Expanded as Fourier Series”, in the Proceedings of the CCP4 Study Weekend 15–16 February, 1985, pp 22–26.

Tickle, I.J. (1992), “Fast Fourier Translation Functions”, in the Proceedings of the CCP4 Study Weekend, January 31st, February 1st, 1992, pp 20–32.

Tong, L. & Rossmann, M.G. (1993), "Patterson-Map Interpretation with Non-crystallographic Symmetry", *J. Appl. Cryst.*, **26**, 15–21.

Toyoshima, C. & Unwin, N. (1988), "Ion channel of acetylcholine receptor reconstructed from images of postsynaptic membranes", *Nature*, **336**, 247–250.

Toyoshima, C. & Unwin, N. (1990), "Three-dimensional Structure of the Acetylcholine Receptor by Cryoelectron Microscopy and Helical Image Reconstruction", *J. Cell Biol.*, **111**, 2623–2635.

Unwin, N. (1993), "Nicotinic Acetylcholine Receptor at 9Å Resolution", *J. Mol. Biol.*, **229**, 1101–1124.

Unwin, P.N.T. (1974), "Electron Microscopy of the Stacked Disk Aggregate of Tobacco Mosaic Virus Protein", *J. Mol. Biol.*, **87**, 657–670.

Unwin, P.N.T. (1975), "Beef Liver Catalase Structure : Interpretation of Electron Micrographs", *J. Mol. Biol.*, **98**, 235–242.

Unwin, P.N.T. & Henderson, R. (1975), "Molecular Structure Determination by Electron Microscopy of Unstained Crystalline Specimens", *J. Mol. Biol.*, **94**, 425–440.

Valpuesta, J.M., Henderson, R. & Frey, T.G. (1990), "Electron Cryo-microscopic Analysis of Crystalline Cytochrome Oxidase", *J. Mol. Biol.*, **214**, 237–251.

Vand, V., Niggli, A. & Pepinsky, R. (1960), "The use of a Monte Carlo method for obtaining trial and error structures.", *Acta Cryst.*, **13**, 965–1164.

Vand, V. & Pepinsky, R. (1954), "The statistical approach of Hauptman & Karle to the phase problem", *Acta Cryst.*, **7**, 451–452.

Vaughan, P.A., Sturdivant, J.H. & Pauling, L. (1950), "The Determination of the Structures of Complex Molecules and Ions from X-ray Diffraction by Their Solutions : The Structures of the Groups PtBr_6^{--} , PtCl_6^{--} , $\text{Nb}_6\text{Cl}_{12}^{++}$, $\text{Ta}_6\text{Br}_{12}^{++}$, and $\text{Ta}_6\text{Cl}_{12}^{++}$.", *J. Am. Chem. Soc.*, **72**, 5477–5486.

Voges, D., *et al*, & Huber, R. (1994), "Three-dimensional Structure of Membrane-bound Annexin V : A Correlative Electron Microscopy-X-ray Crystallography Study", *J. Mol. Biol.*, **238**, 199–213.

Wand, X. & Janin, J. (1993), "Orientation of Non-Crystallographic Symmetry Axes in Protein Crystals", *Acta Cryst.*, **D49**, 505–512.

Weinstein, S., Jahn, W., Hansen, H., Wittmann, H.G. & Yonath, A. (1989), "Novel Procedures for Derivatization of Ribosomes for Crystallographic Studies", *J. Biol. Chem.*, **264**, 19138–19142.

Williams, R.C. & Fisher, H.W. (1970), "Electron Microscopy of Tobacco Mosaic Virus under Conditions of Minimal Beam Exposure", *J. Mol. Biol.*, **52**, 121–123.

Wilson, K.S. (1978), "The Application of MULTAN to the Analysis of Isomorphous Derivatives in Protein Crystallography", *Acta Cryst.*, **B34**, 1599–1608.

Winkler, F.K., Schutt, C.E. & Harrison, S.C. & Bricogne, G. (1977), "Tomato Bushy Stunt Virus at 5.5Å Resolution", *Nature*, **265**, 509–513.

Wood, W.W. & Parker, F.R. (1957), "Monte Carlo Equation of State of Molecules Interacting with the Lennard-Jones Potential. I. A Supercritical Isotherm at about Twice the Critical Temperature", *J. Chem. Phys.*, **27**, 720–733.

Woolfson, M.M. (1954), "Structure Determination by the Method of Permutation Syntheses", *Acta Cryst.*, **7**, 65–67.

Woolfson, M.M. (1961), "Direct Methods in Crystallography", Oxford University Press, Oxford.

Woolfson, M.M. (1987), "Direct Methods — from Birth to Maturity", *Acta Cryst.*, **A43**, 593–612.

Wright, P.E. (1994), "POU domains and homeodomains", *Curr. Opin. Struct. Biol.*, **4**, 22–27.

Wright, W.B. (1958a), "The Crystal Structure of Glutathione", *Acta Cryst.*, **11**, 632–642.

Wright, W.B. (1958b), "A Comparison of the Methods Used in the Attempt to Determine the Crystal Structure of Glutathione", *Acta Cryst.*, **11**, 642–653.

Wu, H., Keller, W. & Rossmann, M.G. (1993), "Determination and Refinement of the Canine Parvovirus Empty-Capsid Structure", *Acta Cryst.*, **D49**, 572–579.

Young, A.C.M. & Dewan, J.C. (1990), "Enhancement in resolution and lack of radiation damage in a rapidly frozen lysozyme crystal subjected to high-intensity synchrotron radiation", *J. Appl. Cryst.*, **23**, 215–218.

Yvonne, J. & Stuart, D. (1991), "Locating heavy atom sites by automatic Patterson search - GROPAT" in the Proceedings of the CCP4 Study Weekend, 25–26 January, 1991, pp 39–48.

Zachariasen, W.H. (1952), "A New Analytical Method for Solving Complex Crystal Structures", *Acta Cryst.*, **5**, 68-73.

Zaloga, G. & Sarma, R. (1974), "New method for extending the diffraction pattern from protein crystals and preventing their radiation damage", *Nature*, **251**, 551-552.

