# Structural Studies on the Type II Dehydroquinase from *Mycobacterium tuberculosis*

**by**

**David G. Gourley**

A thesis submitted for the degree of Doctor of Philosophy
in The Faculty of Science, University of Glasgow

September 1996

Division of Biochemistry and Molecular Biology
Institute of Biomedical and Life Sciences

and

Department of Chemistry

For Fiona

## Declaration

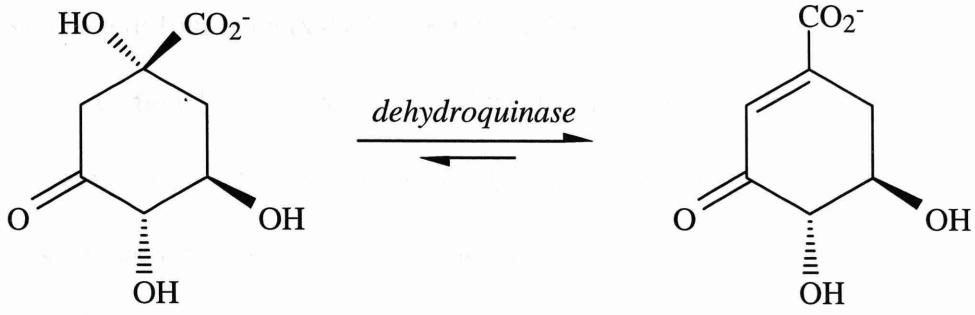
I here by declare that this thesis embodies the results of my own work, that its has been composed by myself and it does not include work forming part of thesis presented sucessfully for a degree in this or another University

Date: 30,9,96

Signature

# Summary

The enzyme 3-dehydroquinase (3-dehydroquinate dehydratase; E.C. 4.2.1.10) catalyses the dehydration of 3-dehydroquinate to 3-dehydroshikimate.




3-dehydroquinate (1) 3-dehydroshikimate (2)

This dehydration reaction is a common step in two important pathways: the biosynthetic shikimate pathway and the catabolic quinate pathway. Dehydroquinases fall into two distinct classes (type I and type II) which are distinguished by non-homologous amino-acid sequences and biophysical criteria. This thesis describes the solution of the three-dimensional structure of the type II dehydroquinase from *Mycobacterium tuberculosis* by X-ray crystallography.

A new purification protocol was designed for *M. tuberculosis* dehydroquinase to give protein pure enough for crystallisation. Crystals were grown by the sitting drop method in 12% (w/v) PEG 6000 at pH 7.1 using a protein concentration of 20 μg/ml. These crystals diffract beyond 2.0 Å. Soaking these crystals in $Hg(CN)_2$ and $K_2Pt(Cl)_4$ gave derivatives for calculating phases using multiple isomorphous replacement. These phases were then subjected to improvement using density modification techniques. A map was calculated and a model was built into the map. This model was then refined using data from 30 to 2.0 Å to give an overall R-factor of 16.6 % and an R-free of 21.5 %.

The type II dehydroquinase from *M. tuberculosis* is a dodecamer. The enzyme subunit has an ($\alpha/\beta$) type fold. It has a central twisted wall of parallel $\beta$-sheet protected on both sides by two helices. The oligomeric structure of the type II dehydroquinase is a dodecamer. The subunits combine to give an arrangement consisting of four trimers packed tetrahedrally.

A model for substrate binding to the type II dehydroquinase which is consistent with the biochemical and crystallographic data is presented and experiments to test this model are suggested.

## Acknowledgements

I would like to thank the following people for their assistance and advice. .

My supervisors Professor N. W. Isaacs and Professor J.R. Coggins for expertise guidance and advice.

The members of the molecular enzymolgy group; Tino Krell, Gillian Muir, Brian McGinn, Malcolm Horsburgh, Nel Beaumont, Graeme Thompson and especially Andrew Elwell for his computation expertise.

The members of the protein crystallography group; John Maclean, Jeremy Beauchamp, Dina Fotinu, Adrian Lapthorn, Andy Freer, Alex Rosack, and especially to Paul Emsley and Steve Prince for sharing their crystallographic expertise

Also to Keith Henderson for long and stimulating discussions on the phase problem.

# Abbreviations

| | |
|---|---|
| A | absorbance |
| amp | ampicillin |
| ATP | adenosine triphosphate |
| c.d. | circular dichroism |
| Da | Dalton |
| DAHP | 3-deoxy-D-arabino-heptulosonate-7-phosphate |
| DEAE | diethylaminoethyl |
| DEPC | diethylpyrocarbonate |
| DHQ | 3-dehydroquinase |
| DMSO | dimethylsulphoxide |
| DTT | dithiothreitol |
| EDTA | ethylene diamine tetra acetate |
| EN | energy minimisation |
| EPSP | 5-enolpyruvylshikimate 3-phosphate |
| FFT | Fast Fourier Transform |
| FPLC | fast protein liquid chromatography |
| HEPES | N-[2-Hydroxyethyl]piperazine-N'-[2-ethanesulfonic acid] |
| HPLC | high pressure liquid chromatography |
| IPTG | isopropyl-β-D-thiogalactoside |
| ITC | isothermal titration calorimetry |
| kb | kilo base pairs |
| MAD | multiple anomalous dispersion |
| MD | molecular dynamics |
| MIR | multiple isomorphous replacement |
| MIRAS | multiple isomorphous replacement with anomalous scattering |
| $M_r$ | relative molecular weight |
| NADH | reduced nicotinamide adenine dinucleotide |
| NADP | nicotinamide adenine dinucleotide phosphate |
| NADPH | nicotinamide adenine dinucleotide phosphate-reduced |
| PAGE | polyacrylamide gel electrophoresis |

| | |
|---|---|
| PEP | phosphoenol pyruvate |
| rpm | revolution per minute |
| SDH | shikimate dehydrogenase |
| SDS | sodium dodecyl sulphate |
| SIR | single isomorphous replacement |
| SIRAS | single isomorphous replacement with anomalous scattering |
| SK | shikimate kinase |
| TEMED | N, N, N', N'-tetramethyl-ethylenediamine |
| terp | Terpyridine |
| TMPA | trimethyl plumbyl acetate |
| Tris | Tris (hydroxymethyl) aminomethane |
| U | units of enzyme activity |
| u.v. | ultra violet |

# Table of Contents

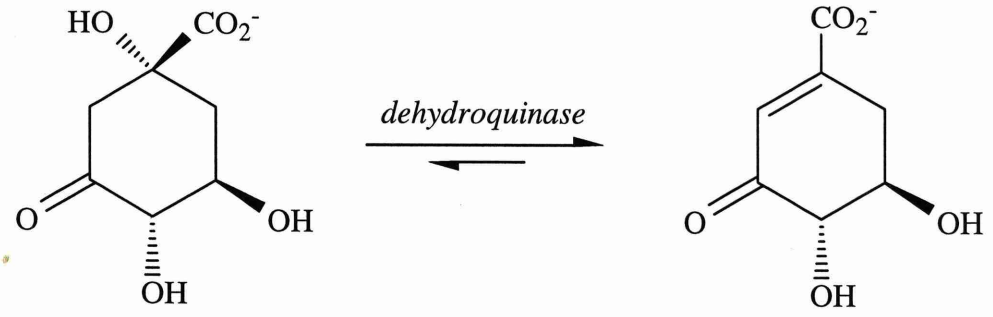
# List of Figures

# List of Tables

# 1. Dehydroquinase: A General Introduction

## 1.1 Dehydroquinase

The enzyme 3-dehydroquinase (3-dehydroquinate dehydratase; E.C. 4.2.1.10) catalyses the dehydration of 3-dehydoquinate to 3-dehydroshikimate (Fig. 1.1). This dehydration reaction is a common step in two important pathways: the shikimate pathway for aromatic biosynthesis (Haslam, 1974, 1993; Bentley, 1990) (Fig. 1.2, see Section 1.4) and the catabolic pathway for quinate metabolism (Giles *et al.*, 1967,1985) (Fig. 1.2, see section 1.5). Two classes of dehydroquinase exist with different biochemical and biophysical properties (Fig. 1.3) and no sequence similarities (Figs. 1.4 & 1.5) (Kleanthous *et al,*. 1992). The two classes of dehydroquinase are termed type I and type II. This Thesis describes the determination of the three dimensional structure of a type II dehydroquinase.




3-dehydroquinate (1) 3-dehydroshikimate (2)

**Figure 1-1 The dehydration of 3-dehydroquinate to 3-dehydroshikimate**

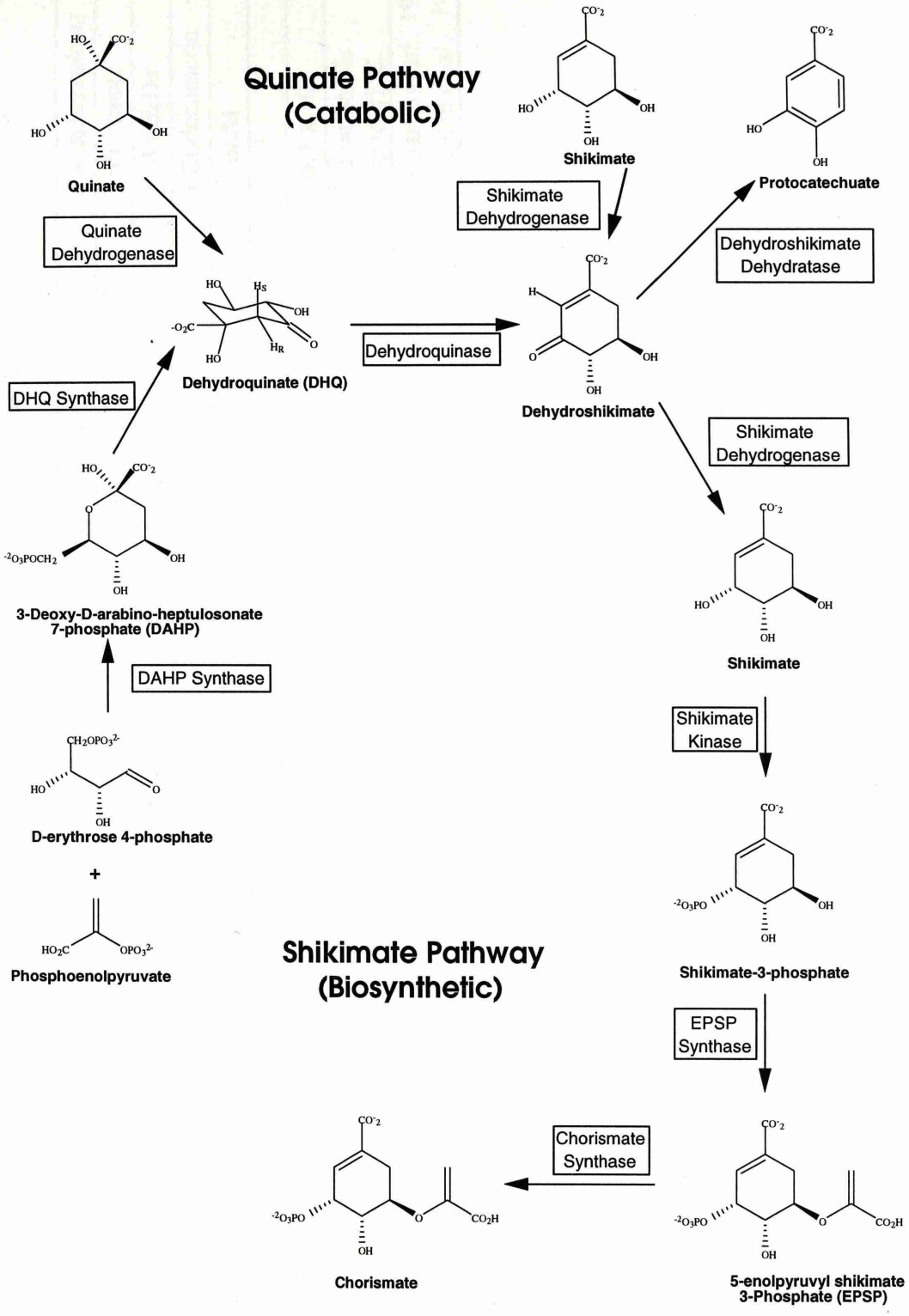



**Figure 1-2 The Shikimate and Quinate Pathways**

| | Type I | | Type II | |
|---|---|---|---|---|
| Organism | *E.coli* | *S.typi* | *A.nidulans* | *M.tuberculosis* |
| Pathway | Shikimate | Shikimate | Quinate | Shikimate |
| Sub Unit $M_r$ | 27.46 kDa | 27.70 kDa | 16.50 kDa | 14.00kDa |
| Stucture | Dimeric | Dimeric | Dodecameric | Dodecameric |
| $K_m$ for DHQ | 16μM | 18μM | 150μM | 16μM |
| $k_{cat}$ | 135 $s^{-1}$ | 200 $s^{-1}$ | 1300 $s^{-1}$ | 4$s^{-1}$ |
| $k_{cat}/K_m$ | 8.4 x $10^6$ | 1.1 x $10^7$ | 8.7 x $10^6$ | 4.4 x $10^5$ |
| Thermal Stablility | Heat Lable | Heat Labile | Heat Stable up to 87°C | Heat Stable up to 87°C |
| Reference | (Chaudhuri *et al.*, 1986; 1991) | (Servos *et al.*, 1991; Moore *et al.*, 1993) | (Da Silva *et al.*, 1986; Kleaanthous *et al*, 1992) | (Garbe *et al.*, 1991; Moore *et al.*, 1992) |

**Figure 1-3 Comparison of the Two Types Of DHQ**

# Sequence Alignment of Type I

```
            1                                                   50
S.TYPII     MKTVTVKNLI IGEGMPKIIV SLMGRDINSV KAEALAYREA TFDILEWRVD
E.COLI      MKTVTVKDLV IGTGAPKIIV SLMAKDIASV KSEALAYREA DFDILEWRVD
B.SUBTILIS  MNVLTIKGVS IGEGMPKIII PLMGKTEKQI LNEAEAVKLL NPDIVEWRVD
E.FAECALIS  MKPVIVKNVR IGEGNPKIVV PIVAPTAEDI LAEATASQTL DCDLVEWRLD

            51                                                  100
S.TYPII     HFMDIASTQS VLTAARVIRD AMPDIPLLFT FRSAKEGGEQ TITTQHYLTL
E.COLI      HYADLSNVES VMAAAKILRE TMPEKPLLFT FRSAKEGGEQ AISTEAYIAL
B.SUBTILIS  VFEKANDREA VTKLISKLRK SLEDKLFLFT FRTHKEGGSM EMDESSYLAL
E.FAECALIS  YYENVADFSD VCNLSQQVME RLGQKPLLLT FRTQKEGGEM AFSEENYFAL

            101                                                 150
S.TYPII     NRAAIDSGLV DMIDLELFTG DADVKATVDY AHAHNVYVVM SNHDFHQTPS
E.COLI      NRAAIDSGLV DMIDLELFTG DDQVKETVAY AHAHDVKVVM SNHDFHKTPE
B.SUBTILIS  LESAIQTKDI DLIDIELFSG DANVKALVSL AEENNVYVVM SNHDFEKTPV
E.FAECALIS  YHELVKKGAL DLLDIELFAN PLAADTLIHE AKKAGIKIVL CNHDFQKTPS

            151                                                 200
S.TYPII     AEEMVSRLRK MQALGADIPK IAVMPQSKHD VLTLLTATLE MQQHYADRPV
E.COLI      AEEIIARLRK MQSFDADIPK IALMPQSTSD VLTLLAATLE MQEQYADRPI
B.SUBTILIS  KDEIISRLRK MQDLGAHIPK MAVMPNDTGD LLTLLDATYT MKTIYADRPI
E.FAECALIS  QEEIVARLRQ MQMRQADICK IAVMPQDATD VLTLLSATNE MYTHYASVPI

            201                                                 250
S.TYPII     ITMSMAKEGV ISRLAGEVFG SAATFGAVKQ ASAPGQIAVN DLRSVLMILH
E.COLI      ITMSMAKTGV ISRLAGEVFG SAATFGAVKK ASAPGQISVN DLRTVLTILH
B.SUBTILIS  ITMSMAATGL ISRLSGEVFG SACTFGAGEE ASAPGQIPVS ELRSVLDILH
E.FAECALIS  VTMSMGQLGM ISRVTGQLFG SALTFGSAQQ ASAPGQLSVQ VLRNYLKTFE

            251
S.TYPII     NA...
E.COLI      QA...
B.SUBTILIS  KNTRG
E.FAECALIS  QNK..
```

**Figure 1-4 Multiple sequence alignment of the type I dehydroquinases**

# Sequence Alignment of Type II

```
                    1                                                  50
N.crassa              ASPRHIL LINGPNL.NL LGTREPQIYG STTLHDIEQA SQTLASSLGL
A.nidulans              MEKSIL LINGPNL.NL LGTREPHIYG STTLSDVEES SKGHAASLGA
A.salmonicida         MSQNHRIL LLNGPNLNNL LGKREPGIYG SKTLDEIVAD LKHNAIELGV
A.pleuropneumoniae      MKKIL LLNGPNL.NM LGKREPHIYG SQTLSDIEQH LQQSAQAQGY
Mycob.tuberculosis    MSELIVN VINGPNLGR. LGRREPAVYG GTTHDELVAL IEREAAELGL
St.coelicolor       PRSLANAPIM ILNPGNL.NL LGQRQPEIYG SDTLADVEAL CVKAAAAHGG

                    51                                                 100
N.crassa            RLTTFQSNHE GAIIDRIHQA AGFVPSPPSP SPSSAATTTE AGLGPGDKVS
A.nidulans          SLQTFQSNHE GAIVERIHAA RG..... ..... .....NTD
A.salmonicida       TLEHLQSNAE HELVSRIHQA MG..... ..... .....QVD
A.pleuropneumoniae  ELDYFQANGE ESLINRIHQA FQ..... ..... .....NTD
Mycob.tuberculosis  KAVVRQSDSE AQLLDWIHQA AD..... ..... .....AAE
St.coelicolor       TVDFRQSMHE GLEVDWIHEA R..... ..... .....LNHC

                    101                                                150
N.crassa            AIIINPGAYT HTSIGIRDAL LGT.GIPFVE VHVSNVHARE AFRHHSYLSD
A.nidulans          AIIINPGAYT HTSVAIRDAL LGV.EIPFIE LHVSNVHARE PFRHHSYFSD
A.salmonicida       YIIINPAAFT HTSVAIRDAL LGV.AIPFIE VHLSNVHARE PFRHHSYLSD
A.pleuropneumoniae  FIIINPGAFT HTSVAIRDAL LAV.SIPFIE VHLSNVHARE PFRHHSYL..
Mycob.tuberculosis  PVILNAGGLT HTSVALRDAC AEL.SAPLIE VHISNVHARE EFRRHSILS.
St.coelicolor       GIVINPAAYS HTSVAILDAL NTCDGLPVVE VHISNIHQRE PFRHHSYVSQ

                    151                                    190
N.crassa            K....AVAVI CGLGPFGYSA ALDFLGRHMK F.........
A.nidulans          K....ASGII VGLGVYGYKV AVEHVALNFK PLEKKAAL..
A.salmonicida       VAKGVAKGVI CGLGADGYQF ALTAAVHQLR AA........
A.pleuropneumoniae  ...NVAKGVI CGLGAKGYDY ALDFAISELQ KIQLGEMMNG
Mycob.tuberculosis  ...PIATGVI VGLGIQGYLL ALRYLAEHVG T.........
St.coelicolor       R....ADGVV AGCGVQGYVF GVERIAALAG AGSARA....
```

**Figure 1-5 Multiple sequence alignment of the type II dehydroquinases**

## 1.2 Type I Dehydroquinases

The type I dehydroquinases, which are exemplified by the well studied monofunctional *Escherichia coli* enzyme (Chaudhuri, *et al.*, 1986, 1991), appear only to be involved in shikimate biosynthesis. Other examples of the type I enzyme include the monofuctional enzyme from *Salmonella typhi* (Servos *et al.*, 1991; Moore *et al.*, 1992, 1993) and the dehydroquinase domains of the pentafuctional AROM polypeptide of *Neurospora crassa* (Lumsden & Coggins, 1977, 1978; Smith & Coggins, 1983; Lambert *et al.*, 1985), *Aspergillus nidulans* (Charles *et al.*, 1986; Hawkins, 1987; Hawkins *et al .*, 1993a) and *Sacharomyces cerevisiae* (Duncan *et al.*, 1987) and the bifunctional polypeptide shikimate dehydrogenase-dehydroquinase of *Pisum sativum* (Mousdale *et al .*, 1987; Deka *et al.*, 1994) and other higher plants (Koshiba, 1978; Polley, 1978). A sequence alignment of the type I dehydroquinases including the dehydroquinase domains of the pentafuctional polypeptides is shown in Fig. 1.4. Subunit sizes of the type I enzymes from various species are shown, they have a subunit size varying from about 25 kDa to 28 kDa. Estimates of the secondary structure of the type I dehydroquinase from *E.coli* are 50 % α-helix 25% β-sheet and 25% random coil (Kleanthous *et al,* 1992).

Size exclusion chromatography, sedimentation velocity and equilibrium centrifugation have indicated that the native type I enzyme from *E. coli* is a dimer (Chaudhuri *et al.*, 1986; Kleanthous *et al.*, 1992) and size exclusion chromatography has also confirmed that the type I enzyme from *S. typhi* is also a dimer (Moore *et al.*, 1993). The type I enzymes from *E. coli* and *S. typhi* unfold in the presence of guanidine hydrochloride in one continuous transition which suggests that the dimer unfolds in a single cooperative step (Kleanthous *et al.*, 1991, 1992; Moore *et al.*, 1993). The type I

enzymes are also readily thermally denatured with the dimer unfolding cooperatively (Kleanthous *et al.,* 1992; Moore *et al.,* 1993).

### 1.2.1 Mechanism of the Type I Dehydroquinases

In the dehydration of dehydroquinate to dehydroshikimate the type I enzyme removes the *pro*-R hydrogen from C-2 to give an overall *syn* elimination (Hanson *et al .,* 1963; Smith *et al.,* 1970). This is a surprising result as the axial *pro*-S hydrogen is more acidic in the preferred chair conformation and indeed the non-enzymic dehydration involves the loss of the *pro*-S hydrogen and proceeds with overall *anti*-stereochemistry (Turner *et al.,* 1975) . This can be accounted for when the enzyme mechanism is considered.

The reaction has been shown to proceed via a Schiff base mechanism by trapping the Schiff base (imine) intermediate by reduction with sodium borohydride to a stable secondary amine (Butler *et al.,* 1975; Smith & Coggins, 1983; Chaudhuri *et al.,* 1991; Moore *et al.,* 1993). This has been supported by the direct observation of the imine intermediate by electrospray mass spectrometry (Shneier *et al.,* 1991). It was proposed that the imine was formed between the substrate and an active site lysine. Experiments involving enzyme inactivation using $H^3$ labelled sodium borohydride in the presence of an equilibrium mixture of substrate and product, followed by tryptic digestion and sequencing of the labelled peptide, have indicated that in *E. coli* the lysine involved in imine formation is Lys-170 (Chaudhuri *et al.,* 1991).

The Schiff base mechanism could be responsible for the unusual stereochemistry of the elimination reaction by distorting the imine intermediate to bring the *pro*-R C-2 hydrogen coplanar with the $\pi$-acceptor orbital of the imine thus increasing its acidity

(Smith *et al .*, 1975; Vaz *et al .*, 1975) and making it more susceptible to removal by a suitably positioned basic side chain.

The type I enzyme has a simple $V_{max}$ dependence on pH consistent with the presence of a single ionising group in the active site with a $pK_a$ of 6.1 (Chaudhuri *et al.*, 1986). This led to the conclusion that there was a general base involved in the mechanism. Chemical modification work with diethyl pyrocarbonate suggested that this base was a histidine residue (Chaudhuri *et al.*, 1986). The *E. coli* enzyme contains six histidines residues, the particular histidine involved His-143, has been identified by chemical modification in the presence and absence of substrate followed by chymotryptic digestion and peptide mapping (Deka *et al.*, 1992). This has recently been confirmed by site directed mutagenesis (Leech *et al.*, 1995) which has also revealed a role for His 143 in the formation and hydrolysis of the Schiff base.

The overall mechanism proposed for the type I dehydoquinase (Fig. 1.6) involves the formation of the Schiff base between Lys-170 and the substrate followed by a base catalysed removal of the pro-R hydrogen at C-2 by His 143. The imine makes the proton abstraction more favourable by acting as an electron sink that can stabilise the resulting carbanion. The hydroxyl on C-1 is then thought to be eliminated, presumably via a protonation step on the oxygen possibly involving the protonated form of His 143 to generate the product Schiff base intermediate which is then hydrolysed and products dissociate from the enzyme.

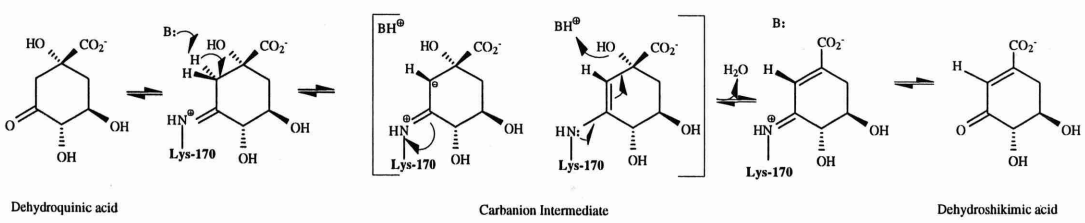



**Figure 1-6 Proposed mechanism for type I dehydroquinase**
His 143 is thought to be the general base B

## 1.3 Type II Dehydroquinases

The type II enzymes, which so far have been found only as highly multimeric monofuctional proteins, were first identified as components of the quinate catabolic pathway in *N. crassa* (Giles *et al.,* 1967; 1985) and *A. nidulans* (Hawkins *et al.,* 1982). Recently a type II enzyme has been found as a biosynthetic shikimate pathway enzyme in *Streptomyces coelicor* (White *et al.,* 1990) and in *Mycobacterium tuberculosis* (Garbe *et al.,* 1991), and as a component of both the shikimate and quinate pathways in *Amycolatopsis methanolica* (Euverink *et al.,* 1992). So unlike the type I dehydroquinases, which function only as part of the biosynthetic shikimate pathway, the type II dehydroquinases are involved in both the biosynthetic shikimate pathway and the catabolic quinate pathway.

The alignments of all the type II dehydroquinase sequences available are shown in Fig 1.5. It can be seen that there is extensive homology within each class of dehydroquinase but not between the classes (compare Figs. 1.4 & 1.5). The subunit size of the type II dehydroquinases are shown in in Fig 1.3. It can be clearly seen that the average subunit size of the type II enzyme is about 10kDa less than that for the type I enzyme. The secondary structures of the type II dehydroquinase from *A. nidulans* has been estimated by c.d. to be 75 % α-helix with the remaining structure being almost entirely β-sheet (Kleanthous *et al,..* 1992).

The type II enzyme is thought to be a dodecamer as size exclusion chromatography, sedimentation velocity and equilibrium centrifugation experiments on the type II enzyme from *A. nidulans* give a native molecular mass of approximately 12 times that of the subunit mass (Hawkins *et al.,* 1982; Kleanthous

*et al.*, 1992). Chemical denaturation of the *A. nidulans* type II enzyme requires treatment with 4M guanidine hydrochloride and unfolding does not occur in a single step but involves three intermediate states (Kleanthous *et al.*, 1992). The type II enzyme is thermally stable (Hautala *et al.*, 1975; Hawkins *et al.*, 1982; White *et al.*, 1990) and only begins unfolding at about 80°C and this again involves three intermediate states (Kleanthous *et al.*, 1992). Preliminary work done on the *Mycobacterium tuberculosis* type II enzyme indicates it is also a dodecamer and is thermally stable (Moore *et al.*, 1992).

The $K_m$ values reported for the type II enzymes are much higher than those from the type I enzymes (Fig. 1.3) with the exception of the type II enzyme from *M. tuberculosis* which has a $K_m$ of 16 μM Also importantly there is a lack of inactivation following the treatment of the type II enzyme with sodium borohydride or cyanoborohydride in the presence of a substrate/product equilibrium mixture. This combined with the fact that there is no conserved lysine in the sequence alignments makes it very unlikely that the type II enzymes use a Schiff base mechanism. Experiments using dehydroquinate differently deuteriated in the C-2 *pro*-R and *pro*-S positions have shown that the type II enzyme catalyses the reaction with the opposite stereochemistry to the type I enzyme, that is the *pro*-S hydrogen is removed in an overall *anti* elimination (Shneier *et al.*, 1993; Harris *et al.*, 1993) (Fig. 1.7).

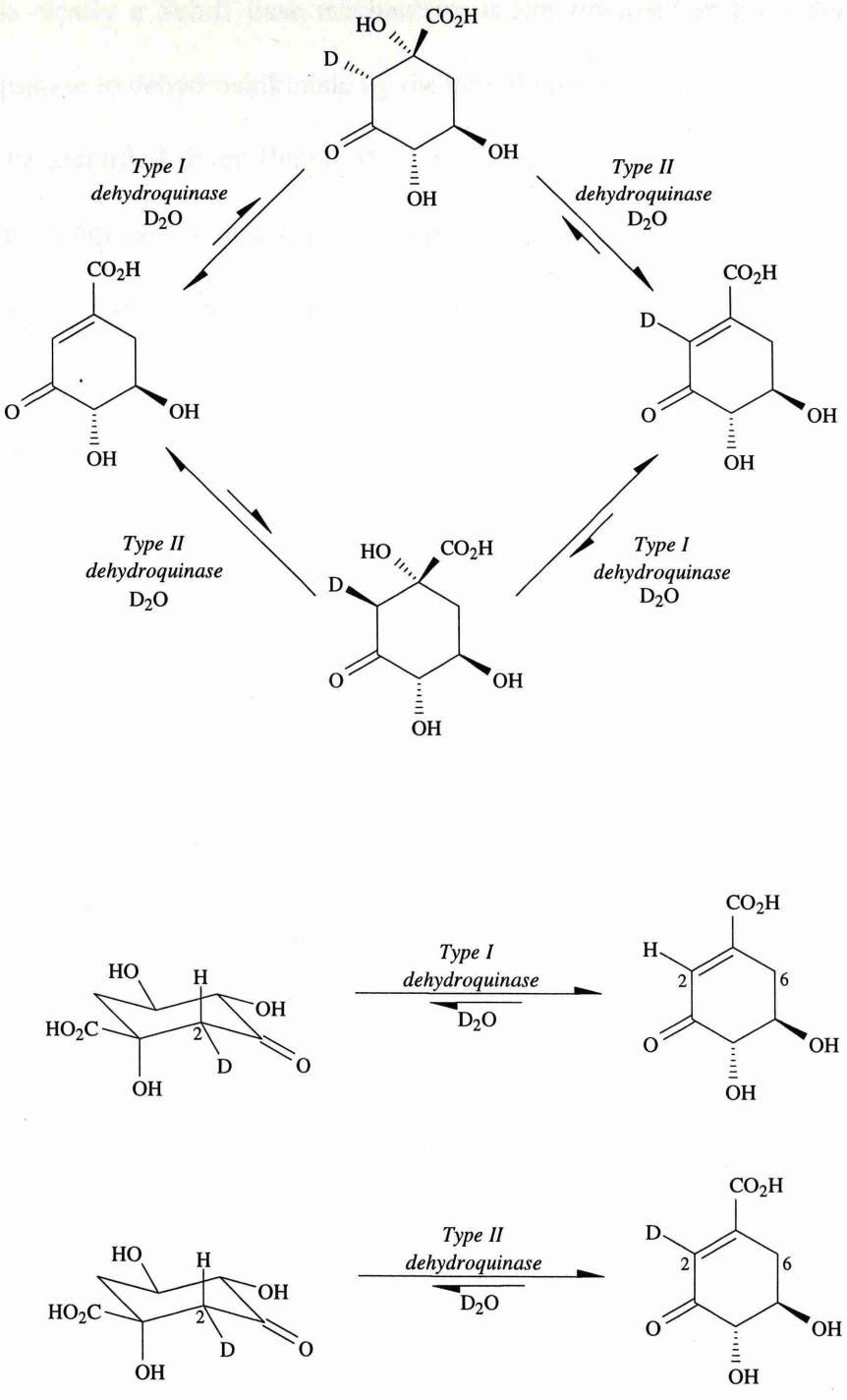



**Figure 1-7 The different stereochemistries of dehydroquinase**

As clearly a Schiff base mechanisim is not involved in the dehydration of dehydroquinase to dehydroshikimate by the type II enzymes an alternative mechanism needs to be identified. Even though there has been work done identifying active site residues by chemical modification (Kleanthous *et al.,* 1992; Krell *et al.,* 1995; 1996) and on the enzymology of the type II enzyme it has not been as comprehensive as that done on the type I enzymes and the results have so far failed to identify the enzymic mechanism.

The pH dependence of $V_{max}$ for the type II enzymes varies between species and cannot be simply interpreted unlike data for the type I enzymes (Kleanthous *et al.,* 1992; Harris *et al.,* 1996). Preliminary chemical modification work has shown that the type II enzyme from *A. nidulans* is very susceptible to inactivation by diethyl pyrocarbonate (DEP) (Kleanthous *et al.,* 1992). This suggests that there is a histidine residue in the active site of the enzyme and is supported by the protection from DEP-inactivation of the *A. nidulans* enzyme observed in the presence of an equilibrium mixture of substrate and product (Kleanthous *et al.,* 1992).

It was thought that the mechanism involved initial proton abstraction followed by the stablisation of the carbanion intermediate as an enol or enolate. The type II DHQase can be unfolded by guanidine hydrochloride in the presence of EDTA and then refolded and assayed in the presence of EDTA without significant loss of activity which suggests a metal ion is not involved in stabilisation of a carbanion intermediate (Bottomly *et al.,* 1996). Another way to stabilise this postulated carbanion would be to involve positively charged sidechains such as an arginine residue. Chemical modification, using the arginine-specific reagent phenyl glyoxal, inactivates the enzyme by modifying a hyper-reactive arginine (Krell *et al.,* 1995;

1996). This arginine has been identified as Arg-23 in the *S. coelicor* enzyme which is a residue that is conserved through out all the type II dehydroquinases (Krell *et al.*, 1995; 1996). Replacement of this arginine residue by site directed mutagenesis greatly decreased the catalytic efficiency of the enzyme (Krell *et al.*, 1996). Arg-23 is therefore likely to be involved in the catalytic mechanism but this data does not conclusively confirm that the arginine is involved in carbanion stabilisation. The pH dependence of $K_M$ for the *M. tuberculosis* enzyme is consistent with such a role for an arginine residue (Harris *et al.*, 1996). Further work will be required to confirm the mechanistic details of the type II dehydroquinase and this is one of the major reasons for determining the 3D structure of a type II dehydroquinase by X-ray crystallography.

## 1.4 The Shikimate Pathway

The shikimate pathway is the seven step biosynthetic pathway found only in plants and microorganisms which leads to chorismic acid the precursor of the vast majority of aromatic compounds (Fig. 1.2). The pathway begins with the condensation reaction of D-erythrose-4-phosphate (a product of the Calvin cycle or the pentose phosphate pathway) and phosphoenol pyruvate (a product of glycolysis) and procedes via six intermediates including 3-dehydroquinic acid (the first carbocyclic intermediate) and shikimic acid and ends with chorismic acid. Chorismic acid is the precursor of the aromatic amino acids (phenylanaline, tyrosine and tryptophan), folic acid, ubiquinone, plastiquinone, entrochelin and many other aromatic compounds (Fig. 1.8). The shikimate pathway is estimated to generate between 10 and 35 % of the dry weight of higher plants (Boudet *et al.*, 1985;

Coggins, 1986). Animals lack this pathway and so the aromatic amino acids, folic acid and a number of other aromatic compounds form an essential part of their diet.

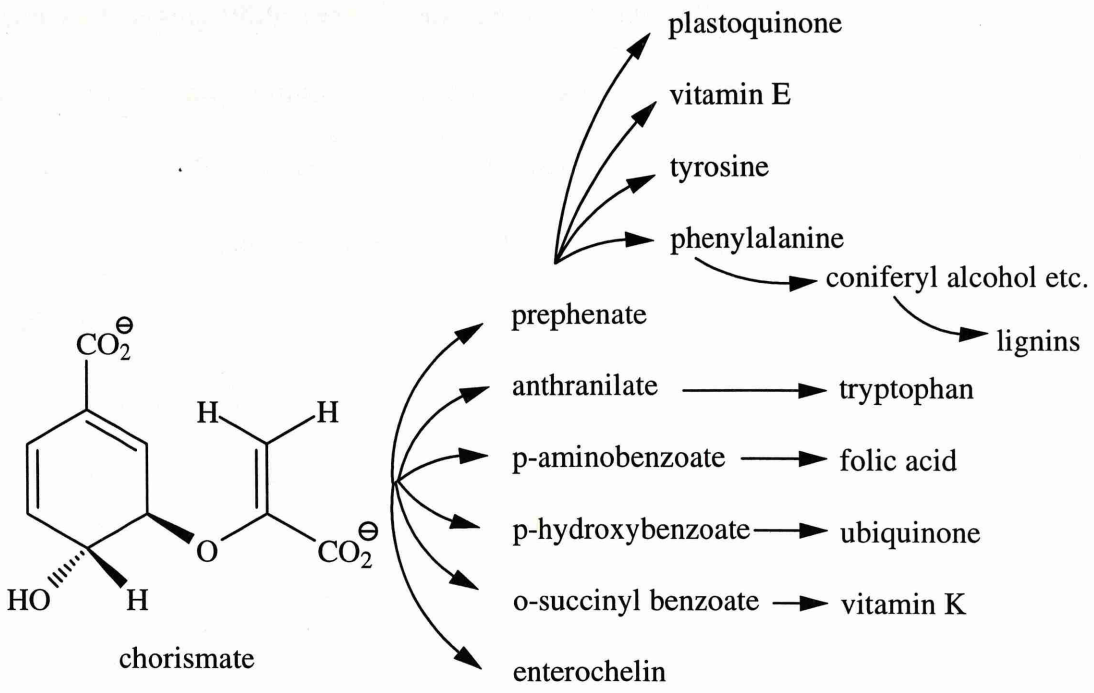



**Figure 1-8 The utilisation of chorismate**

Enzymes of this pathway are potentially very useful targets for drug and herbicide design as many of the compounds synthesised by the pathway are essential to the organism and because specific inhibitors against them should be harmless to animals (Coggins 1986; 1989; Mousdale & Coggins 1993). This is emphasised by the very commercially important, broad spectrum herbicide glyphosate (comercially known as Tumbleweed® or Roundup®) which acts by inhibiting enolpyruvylshikimate phosphate (EPSP) synthase, the sixth enzyme of the pathway (Steinrucken and Amrhein, 1980; Boocock and Coggins 1983), and by the observation that fluoroshikimate is an antibacterial agent (Davis *et al.,* 1994). This makes the study of the structure and mechanism of the enzymes in the shikimate pathway important.

The pathway gets its name from the first pathway intermediate to be isolated from a plant (Eykman, 1885; 1891). The source of the first sample of shikimic acid was the plant *Illicium religiosum* which has the Japanese name shikimi-no-ki. The name of the compound and pathway derives from this Japanese plant name. Mutant bacterial auxotrophs were used to elucidate the chemical steps in the pathway and for the isolation of the pathway intermediates (Davis, 1948, 1955; Gibson and Gibson, 1964; Gibson and Pittard, 1968). Isotopic labelling studies with $^{14}$C-labelled D-glucose indicated that the shikimate was biosynthesised from a three carbon fragment of glycolysis and a four carbon sugar (Sprinson, 1960) which were later identified as phosphoenol pyruvate and D-erythrose-4-phosphate ( Levin and Sprinson, 1964).

Since their characterisation all of the pathway intermediates have been non-enzymically synthesised (Ganem, 1978; Bentley, 1990). Initially detailed studies of the enzymes were difficult as the enzymes were only present at low levels,

particularly in plants which meant they were difficult to purify (Coggins 1989; Mousedale & Coggins 1993). Recently the advent of recombinant DNA and other molecular biological techniques has enabled the genes encoding all the shikimate pathway enzymes to be cloned and over expressed. This has led to the production of these enzymes in milligram, tens of milligrams and in some cases even hundreds of milligram quantities (Anton & Coggins, 1988; Shuttleworth *et al.*, 1992). This has facilitated detailed chemical, biochemical and biophysical studies on the pathway enzymes including site directed and group specific chemical modification, site directed mutagenesis, c.d., micro-calorimetry and crystallography. All seven enzymes from the shikimate pathway have now been crystallised: DAHP synthase (Shumilin *et al.*,1996), the dehydroquinate synthase domain of the AROM multifuctional enzyme from *A. nidulans* (K. Brown personal comunication), the type I DHQase from *S. typhi* (Boys *et al.*, 1992) , the type II dehydroquinase from *M. tuberculosis* (Gourley *et al.*, 1994; the structural studies reported in this thesis), *E. coli* shikimate dehydrogenase (S. Chackrewarthy, personal communication 1995), the shikimate kinase from *Erwinia chrysanthemi* (T. Krell personal communication), EPSP synthase from *E. coli* (Stallings *et al.*, 1991) and chorismate synthase from *E. coli* and *Stapyhloccocus aureus* (A.Lapthorn personal communication).

### 1.4.1 Organisation of the Shikimate Pathway Enzymes

The reactions and intermediates of the shikimate pathway are common to every organism in which the pathway occurs. However the structural organisation of the enzymes that catalyse the seven pathway reactions varies significantly between organisms from different taxonomic groups. In enteric bacteria, like *E. coli,* the seven reactions are catalysed by separate enzymes which can be independently purified.

In plants like *Physcomitrella patens* (Polley, 1978), *P. sativum* (Mousdale *et. al.*, 1987; Daka *et al,* 1994) and *Phaseolus mungo* (Boudet *et al.*, 1985) the dehydroquinase and shikimate dehydrogenase activities occur on the same bifunctional polypeptide while all the other reactions are catalysed by monofunctional polypeptides.

In fungi like *A. nidulans* (Charles *et al.*, 1986; Hawkins, 1987; Hawkins *et al.*, 1993a), *N. crassa* (Lumsden & Coggins, 1977, 1978; Smith & Coggins , 1983 ; Lambert *et al.*, 1985) and in the yeast *S. cerevisiae* (Duncan *et al.*, 1987) the enzyme activities catalysing the second to sixth reactions occur on a single pentafunctional polypeptide called the AROM multifunctional enzyme which is encoded by a single gene. The individual domains of this pentafunctional AROM protein are individually homologous to the five equivalent monofunctional enzymes found in bacteria like *E. coli.* This suggests that the *arom* locus is the product of multiple gene fusions (Hawkins, 1987). The enzyme organisation in *Euglena gracilis* is similar to that in fungi (Ahmed & Giles, 1969). A schematic representation of the structural organisation of the shikimate pathway enzymes in a number of organisms is shown in Fig.1.9.

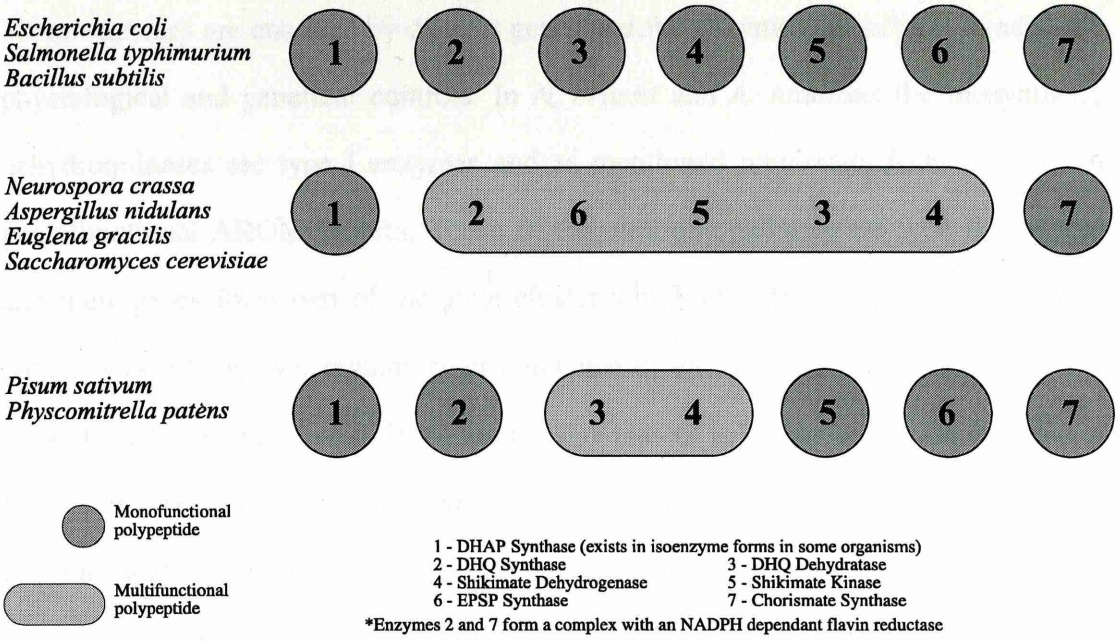



**Figure 1-9 The organisation of the shikimate pathway enzymes in different organisms**

## 1.5 The Quinate Pathway

Quinic acid a very abundant plant metabolite which forms about 2-10% of the dry weight of higher plants (Bentley, 1990). Fungi can utilise this compound as a sole carbon source via the inducible quinate pathway (Fig. 1.2). In this pathway quinic acid is first converted to dehydroquinate by quinate dehydrogenase and then follows a reaction identical to that found in the shikimate pathway: the dehydration of dehydroquinate to dehydroshikimate. The dehydroshikimate is converted by dehydroshikimate dehydratase to protocatechuate which is then used by the β-ketoadipate pathway (Giles *et al.*, 1985; Hawkins *et al.*, 1993) for catabolism.

Although the dehydroquinase reaction is common to both the biosynthetic shikimate pathway and the catabolic quinate pathway in fungi like *N. crassa* and *A. nidulans* different enzymes are involved in the biosynthetic and catabolic pathways.

These enzymes are encoded by distinct genes and the enzymes are subject to separate physiological and genetical controls. In *N. crassa* and *A. nidulans* the biosynthetic dehydroquinases are type I enzymes and as mentioned previously form part of the pentafunctional AROM protein. Those of the catabolic pathway are type II enzymes and their genes form part of the gene cluster which encodes the catabolic pathway enzymes as well as two regulatory proteins one of which is an activator protein and the other a repressor protein. In *N. crassa* this cluster is known as the *qa* cluster and in *A. nidulans* it is known as the *qut* cluster. The proposed modular structure of the enzymes and regulatory proteins of the quinate pathway is shown in Fig. 1.10. The function of the *A. nidulans qut D* gene product is thought to be a permease involved in the uptake of quinic acid while the product of the *qutG* gene is thought to be a phosphatase although its function has not been established (Lamb *et al.*, 1996).

The production of the enzymes of this pathway is regulated at the level of transcription by the two proteins that are encoded by the regulatory genes in the cluster. One of these encodes the activator [QUTA in *A. nidulans* (Lamb *et al.*, 1996) and the QA-1F in *N. crassa* (Giles *et al.*, 1985; Hawkins *et al.*, 1993)] and the other encodes the repressor [QUTR in *A. nidulans* and QA-1S in *N. crassa* (Giles *et al.*, 1985; Hawkins *et al.*, 1993)]. The activator initiates transcription by binding specific regions of DNA and the repressor is thought to have its effect by binding the activator protein directly and preventing transcription (Giles *et al.*, 1985; Beri *et al.*, 1987 Anton *et al.*, 1987; Hawkins *et al.*, 1992, 1993).

The inducers quinate, dehydroquinate and dehydroshikimate cause the repressor to no longer inhibit the activator. This allows the activator to initiate transcription of all the genes in the quinate pathway cluster including surprisingly the genes encoding the repressor. The reason for this latter event happening is probably to ensure that there is enough repressor present to switch off transcription as soon as the inducer level declines.

### 1.5.1 The Evolutionary Relationship Between the Quinate and Shikimate Pathways

An evolutionary interesting fact is that the QUTA and QA-1F proteins are homologous with the two N-terminal domains of AROM (the DHQ synthase and EPSP synthase domains) and the corresponding individual *E.coli* monofunctional enzymes and that the QUTR and QA-1S proteins are homologous to the three C-terminal domains of AROM (the shikimate kinase, dehydroquinase and shikimate dehydrogenase domains) and their equivalent monfunctional enzymes from *E. coli* (Fig. 1.10) (Anton *et al..*, 1987; Hawkins *et al.,* 1992, 1993). This homology indicates that the repressor has arisen by the duplication, adaptation and splitting of the *arom* gene or its precusor (Anton *et al..*, 1987; Hawkins *et al.,* 1992, 1993). It is thought that the shikimate dehydrogenase like domain in the repressor is the binding site for quinate and dehydroshikimate as mutations in this domain cause the repressor to be uninducible by quinate or dehydroshikimate (Huiet & Giles, 1986; Anton *et al.*, 1987) .

It therefore appears that there is a strongly conserved fold for binding quinate and/or dehydroshikimate which is common to both shikimate dehydrogenase and the repressor protein and is also found in the inducible quinate/shikimate dehydrogenase, the first enzyme of the quinate catabolic pathway, which is itself regulated by the repressor protein.

This idea has been extended to the dehydroquinase domain where dehydroquinate is thought to bind and exert its affect (Anton *et al.*, 1987). Recently biophysical techniques like far-UV c.d. spectroscopy and fluorescence emission spectroscopy have shown the secondary and tertiary structures of the *A. nidulans* repressor protein (QUTR) dehydroquinase-like domain and the *A. nidulans* biosynthetic dehydroquinase domain derived from the AROM protein to be virtually indistinguishable (Lamb *et al.,* 1996) which is consistent with them having a

common structure. The reason for the repressor protein also having a shikimate kinase-like domain is unclear though it has been suggested that because this domain contains the highly conserved type A consensus sequence which is crucial in purine nucleotide utilisation, that the domain has a kinase function and this may be involved in the interaction with the activator protein (Anton *et al.*, 1987).

The activator proteins are made up of three domains, two of which are homologous to the two domains found at the N terminus of AROM (the DHQ synthase and the EPSP synthase domains of AROM) and a third which is at the N-terminus of the activator and has a zinc binuclear cluster motif. This latter domain is thought to be involved in DNA recognition and binding. So the quinate pathway seems to have obtained its regulatory proteins by duplicating and adapting genes encoding the shikimate pathway enzymes and apparently also gained its quinate dehydrogenase by taking and altering the gene for shikimate dehydrogenase. These results highlight the complex evolutionary relationships between the proteins involved in the functioning of the shikimate and quinate pathways.

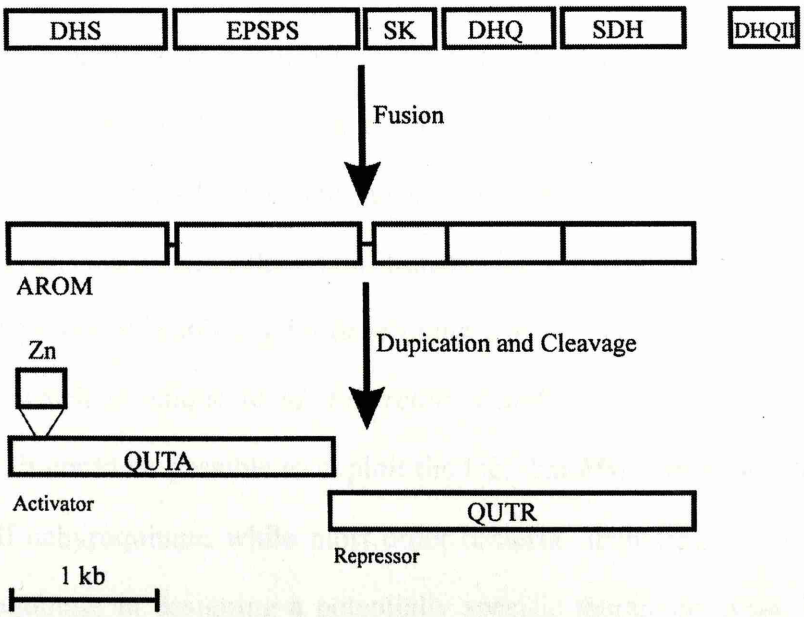



**Figure 1-10 Evolutionary relationship between the shikimate and quinate pathways**

## 1.6 The Type II Dehydroquinase of *M. tuberculosis* as a Target for Drug Design

It is estimated that between the years 1990 and 2000 90 million cases of *Mycobacterium tuberculosis* (TB) infection will occur, resulting in 30 million deaths, more deaths than caused by any other pathogen (Bloom & Murray, 1992). Even in developed countries TB has returned as a mass killer. This is partly because early diagnosis of the disease is difficult and treatment is complex involving multiple drugs which have to be administered over a period of 6 months. In addition there are increased numbers of susceptible hosts for TB because of the possibility of coinfection with human immunodeficiency virus, and new strains of *M. tuberculosis* that are resistant to the drug therapies normally used have evolved (Barnes *et al.*, 1991; Bloom & Murray, 1992; Spratt 1994; Murray 1994). This situation has been described by the World Health Organisation as a global emergency and has lead to an urgent search for novel anti-TB drugs.

There is a need for specific anti-mycobacterial agents because of the long time period in which a therapeutic agent needs to be taken to completely clear a TB infection (Young & Duncan, 1995). A broad spectrum anti-bacterial may cause undesirable side effects when taken over this long period of time, for example disturbing the normal flora of the gut and skin. Also if the drugs were used to treat other less serious illnesses the risk of transferable drug resistance would be increased.

One possible strategy for developing a new TB drug is to identify an essential enzyme which is unique to *M. tuberculosis* and design a specific inhibitor of this enzyme. It could be possible to exploit the fact that *Mycobacterium tuberculosis* has a type II dehyroquinase while most other bacteria including *E. coli* have a type I dehydroquinase in designing a potentially specific theraputic agent. This is because the mechanisms of the two types of dehydroquinase are clearly distinct.

## 1.7 The *M. tuberculosis* Dehydroquinase

The objective of this work was to determine the 3-D of the type II dehydroquinase from *M. tuberculosis* using X-ray crystallography. The gene of this enzyme has been cloned and sequenced (Garbe *et al.,* 1991) and overexpressed in *E. coli* (Moore *et al.,* 1992). This has allowed the enzyme to be purified and partially characterised (Moore *et al.,* 1992). The enzyme functions as part of the shikimate pathway and is the only dehydroquinase found in *M. tuberculosis.* The genes in *M. tuberculosis* encoding dehydroquinate synthase and dehydroquinase are linked but transcription results in the synthesis of two separate proteins (Garbe *et al.,* 1991). The sequence of the *M. tuberculosis* dehydroquinase shows good homology with the other type II dehydroquinases (Fig. 1.5) and structural information obtained on the *M. tuberculosis* enzyme could be extended to the other type II dehydroquinases with a high degree of confidence. The *M. tuberculosis* enzyme has the lowest reported $K_m$ (Moore *et al.,* 1992) of all the type II dehydroquinases ( see Fig. 1.3). It has been suggested that this low $K_M$ reflects the role of the *M. tuberculosis* enzyme in biosythesis and is similar to the $K_M$ values of the biosynthetic type I enzymes. In contrast the catabolic enzymes have much higher $K_M$`s. The *S. coelicolor* enzyme which also has a biosythetic role does not fit this simple model as it has a very high $K_m$ of 650mM (White *et al.,* 1990).

The objective of this work is to determine the structure of this type II dehydroquinase as a means of helping elucidate the enzymic mechanism, of understanding the structural basis for the high thermal and chemical stablity of the multimeric structures and to facilitate the design of specific inhibitors which could be investigated as potential anti-tuberculosis agents.

# 2. Materials and Reagents

## 2.1 Chemicals and Biochemicals

Benzamidine, bromophenol blue, Coomasie Brilliant Blue, N,N,N`,N`-tetramethylethylene diamine (TEMED) potassium tetrachloroplaninate were obtained from Sigma Chemicàls Co., Poole, Dorset, UK.

Dithiothreitol (DDT) and Tris buffer were obtained from Boehringer Mannheim, Lewes, Sussex, UK.

Polyethylene glycol (PEG) 6000 (Biochemical grade) and 2-mercaptoethanol were obtained from BDH Chemicals, Poole, Dorset, UK.

Acrylamide, bisacrylamide and SDS, were obtained from FSA Laboratory Supplies, Loughborough, Leicestershire, UK.

Ammonium tetrachloroaurate (III), mercury (II) cyanide, trimethyl plumbyl acetate and potassium cyanoaurate were obtained from Strem Chemicals Orwell,Royston,UK.

All other chemicals were of analytical grade and were obtained from one of the following suppliers: Aldrich Chemical Co. Ltd., Poole Dorset, UK, BDH Chemicals: FSA Laboratory Supplies; Koch-Light Ltd., Colnbrook, Buckinghamshire, UK.

## 2.2 Other Materials

### Concentrators

Centricon-10 and Centricon-30 concentrators were obtained from Amicon Ltd., Stonehouse Gloucestershire, UK.

**Crystallisation Sitting Drop Trays**

Cryschem trays were used in his work. These were obtained from N.B.S. Biologicals, Huntingdon, Cambridge, UK.

**Chromatography Media**

BIO-GEL® HTP gel (Hydroxapatite) was obtained from Bio-Rad laboratores, Hercules, CA, USA.

**Prepacked Columns**

A Superose 12 HR 10/30 column was obtained from Pharmacia and used on a Pharmacia FPLC System, Pharmacia, Milton Keynes, Buckinghamshire, UK.

A Neobar $^{tm}$ $AQ^4$ ion exchange column was obtained from Flowgen Instuments Ltd. Sittingbourne, Kent,UK.

## 2.3 General Lab Methods

General methods for handling proteins and enzymes were as described in *Methods In Enzymology; Guide to protein purification, Vol.182, pp 19-83.* (Murray P. Deutsher, ed ., 1990)

### 2.3.1 pH Measurement

pH measurements were made with a Radiometer Model 26 pH meter (Copenhagen, Denmark), calibrated at room temperature.

### 2.3.2 Enzyme concentration Measurements

Protein concentration were measured using absorbency at 280 nm on a Gilford model 252 spectrophotometer. The type II dehydroquinase from *M. tuberculosis* has an extinction coefficient of 0.605 for a 1.0mg/ml solution (Gill & Von Hippel, 1989).

### 2.3.3 Polyacrylamide gel electrophoresis

Electrophoresis using polyacrylamide gels in the presence of SDS was performed by the method of Laemmli, (1970), with a 4% stacking gel and a 15% running gel. The ratio of acrylamide:bisacrylamide was 30:0.8 and 0.03% (v/v) TEMED and 0.05% (w/v) ammonium persulphate were used to catalyse polymerisation. After electrophoresis gels were stained for protein using Coomassie Blue. The Coomassie reagent was 0.1% (w/v) Coomassie Brilliant Blue R-250 in 10% (v/v) glacial acetic acid, 50 % (v/v) methanol; the destaining reagent was 10 % (v/v) acetic acid, 10% (v/v) methanol. This technique was used to check for protein contaminants and also to check for any degradation.

### 2.3.4 Enzyme Assay

Dehydroquinase activity was determined by monitoring the production of dehydroshikimate from dehydroquinate using a spectrophotometer (a Gilford model 252 spectrophotometer equipped with an external slave recorder) with the monochromator set at 234nm ($\varepsilon=12x10^3M^{-1}cm^{-1}$ for 3-dehydroshikimic) (Salmon and Davis, 1953; Mitsuhashi and Davis, 1954). The assay mixture contained (final concentrations): 50mM Tris/HCl pH 7.8 and 1mM ammonium dehydroquinate (as substrate). The substrate was prepared as described by Coggins *et al.,* (1987). This assay was used to monitor the purification procedure and particularly to identify purification steps where large amounts of protein were either lost or inactivated

### 2.3.5 Mass Spectrometric Analysis

Protein samples for mass spectrometric analyses were diluted with HPLC grade water and reconcentrated with a Centricon 10. This was done to give an overall

100 fold dilution of any salts or buffers. The protein was concentrated to a final concentration of 40 nanomoles/ml.

Mass spectrometry was performed on a VG Platform quadrupole mass spectrometer fitted with a pneumatically assisted electrospray source and controlled via VG MassLynx software (VG Biotech Ltd., Altrincham. Cheshire, UK). For experimental details see (Krell et al., 1995).

## 2.4 Crystallisation

### 2.4.1 Crystallisation Using the Sitting Drop Method

Cryschem trays are disposable plastic sitting drop trays. Each tray has 24 reservoirs that are 16mm diameter and can take 1ml of precipitant solution. There is a central column in each reservoir which holds a raised well which can accommodate a 20 µl drop. 5µl of the reservoir solution was pipetted into each of the raised wells of that reservoir. Then 5µl of the protein solution of known concentration in 50mM Tris/HCl pH 7.8 was added to the well. Then the reservoirs were sealed using clear plastic tape and the tray left in a temperature controlled room at 20°C.

### 2.4.2 Preparation of Precipitant Solutions

Precipitant solutions for use in the reservoirs for the sitting drop method were made in the following way. Solutions containing the required amount of precipitant and buffer solution were made up in 1 ml Eppendorf centrifuge tubes. These tubes were then spun in a micro-centrifuge for 5 minutes to remove any undissolved precipitants and dust. The top 750µl of the 1ml of solution were then pipetted into the reservoir of the tray.

## 2.5 Procedure For Soaking Crystals

A large number of soaks were done to find heavy atom derivatives for *M. tuberculosis* DHQ. The strategy was to do preliminary experiments using crystals that were too small for diffraction for the initial soaks. This was done to find out how reactive a particular heavy atom compound was for the crystals by looking for cracking, sinking and colour uptake. Soaks were done on diffraction quality crystals using concentrations and times of soak based on observations made from the preliminary experiments. The crystals were very susceptible to changes in their surrounding solution and were only stable in the mother liquor in which they were grown. In order to do soaks an artificial mother liquor that closely resembled this had to be prepared. This was done by setting up crystallisation trays exactly like those for growing the crystals, but a 50mm Tris/HCl solution was added instead of the protein solution and having a drop size of 50 µl instead of 10 µl. These trays were left in the temperature stabilised room for two weeks to equilibrate. All the heavy atom solutions were made up in this artificial mother liquor.

Crystals were soaked in 50 µl drops of the heavy atom solution of the required concentration made up in artificial mother liquor in Cryschem trays sealed with clear plastic tape. The crystals were transferred from their growth solution into the soaking drop. This was done using a Pasteur pipette that had been drawn out in a Bunsen burner so its tip width was about 0.7 mm. If the crystals were then back soaked this was done by transferring a crystal after soaking into a drop in another well containing only artificial mother liquor. During soaking crystals were monitored at regular intervals for cracking colour uptake and sinking.

## 2.6 Computing

### 2.6.1 CCP4

Programs from the CCP4 suite were used extensively in this project. After a data set had been reduced and scaled either by XDS and XSCALE or DENZO and SCALEPACK the resulting file was run through the ROTAPREP, SORTMTZ, ROTAVATA, AGROVATA and TRUNCATE chain.

ROTAPREP - written by Eleanor Dodson. This program was used to produce a multirecord MTZ file from either the output from XSCALE or SCALEPACK.

SORTMTZ - written by P. J. Daley . Sorting of reflections was done by this program.

ROTAVATA - was written by J.M. Smith and A J. Wonacott. This program normally calculates scale and temperature factors. The data used in this project had been previously scaled either with XSCALE or SCALEPACK so this program was run with cycles set to 0 only to prepare the ROT.SCALE file suitable for inputting into AGROVATA.

AGROVATA - was written by J.M. Smith and A J. Wonacott. As scaling had been run previously this program was run with a unit scale to produce a MTZ file suitable for TRUNCATE and to produce statistics on data completeness.

TRUNCATE - written by K.S. Wilson and S. French. This program was used to convert the MTZ file containing average intensities into one containing mean amplitudes and anomalous differences. If the data had been processed using XDS then the actual truncation procedure was not done as during integration in XDS all reflections tending towards negativity are set to zero. For the data processed in

DENZO then the data were truncated. Then the programme does a Wilson plot to calculate an approximate absolute scale to put the data on.

CAD - written by E. Dodson was used extensively in this project. It was used to accumulate both native and derivative reflection data onto one file, with each data set being contained in different columns of that file.

FFT - written by L.F. Ten Eyck. This Fast Fourier Transform program was used for the calculation of difference Patterson, Difference Fouriers and Fouriers.

MLPHARE - written by Z. Otwinowski, this program refines heavy parameters by maxium likelihood and calculates phases.

DM - written by K. Cowtan was used to perform density modification.

SFALL - written by E. Dodson and E. Baker. This program was used to calculate structure factors from models.

SIGMAA - Written by R. Read, this program was used to combine MIR phases with phases calculated from partial structures.

### 2.6.2 Model Building .

Model building was done using O ( Jones & Kjeldgarrd, 1994) and its associated programs.

# 3. Purification of "Crystallography Grade" *M. tuberculosis* Dehydroquinase

## 3.1 Introduction

One of the most important factors in a structural analysis project on a protein using X-ray crystallography is the quality and quantity of the subject protein. The protein should be free of microheterogeneity as well as not having any contamination by unrelated proteins. Microheterogeneity in a protein sample may be due to variation at all levels of protein structure. At the primary structure level microheterogeneity can be the result of genetic variation or it may be due to side chain damage (for example deamidation of glutamine and asparagine residues or oxidation of cysteine residues). It may also be due to degradation of the polypeptide chain. It can occur at the secondary structure level because of mis-folding during synthesis or to partial unfolding during isolation and purification. The existence of conformers can cause microheterogeneity at the tertiary structure level. Oligomerisation differences and dissociation of oligomers can lead to variation at the level of quaternary structure. Microheterogeneity can occur as a result of differences caused by post-transcriptional and post-translational modification, multiple aggregation states and partial binding of ligands or foreign molecules.

Contamination by unrelated proteins or, microheterogeneity can inhibit crystal nucleation or if nucleation occurs, can lead to competition for sites on growing crystals and generate lattice errors. This can cause dislocations, internal disorder, irregular faces and premature ending of crystal growth. This may cause any diffraction to be

poor. The chances of getting large single crystals diffracting to a high resolution with a small mosaic spread are greatly increased if a protein sample is free of contamination and microheterogeneity.

This chapter reports the protocol for purification of the Type II dehydroquinase from *M. tuberculosis.* This protocol was designed to remove all other proteins and any microheterogeneity from the sample. Special care was taken to avoid the introduction of microheterogeneity during purification and storage.

### 3.2 Quantity of sample

When embarking on a study of a protein by X-ray crystallography a good supply of protein is needed for the crystallisation trials (see Chapter 4). This is needed to improve crystal size and quality once initial crystallisation conditions have been found. Also if the project involves using multiple isomorphous replacement (MIR), to provide sufficient protein to grow crystals for screening for heavy atom derivatives (see Chapter 6). In this project there was a plentiful supply of protein since the gene encoding the Type II dehydroquinase from *M. tuberculosis* had been cloned and over expressed in *E.coli* (Garbe *et al.,* 1991; Moore *et al.,* 1992).

### 3.3 Quality Control

The integrity of the protein was analysed throughout the purification by SDS polyacrylamide gel electrophoresis and by enzyme assays and at the end of the purification by electrospray mass spectrometry as described in sections 2.3.3, 2.3.4 and 2.3.5 respectively.

### 3.4 The Overproducing Strain

*M. tuberculosis* dehydroquinase was purified from an *aroD* ⁻ strain of *E. coli* (SK3430) containing the recombinant plasmid pKK44 (Moore *et al.,* 1992), which places the coding sequences of the *M. tuberculosis aroQ* gene (which encodes the dehydroquinase) under the control of the inducible $P_{trc}$ promoter (Garbe *et al.,* 1991).

### 3.5 Previous Purification Protocol

The purification protocol used in this work differs from the purification protocol described previously (Moore *et al.,* 1992). This is because the initial protocol contained a heat shock step and it was thought that this might introduce microheterogeneity by causing partial unfolding and deamidation of the protein. This was of particular concern as previous crystallisation trials had produced only small microcrystals (A. Anderson personal communication). Also previously, after purification and before crystallisation trials, the protein had been lyophilised. Lyophilisation can remove the crucial bound water molecules in the solvation shell leading to unfolding and denaturation of the protein and so reducing the chances of successful crystallisation. In this work the protein was not lyophilised and crystallisations were usually set up using protein solutions that had been concentrated by ultrafiltration immediately after the final purification step. Alternatively protein solutions, which had been stored in 50% (v/v) glycerol-containing buffer at -20$^0$C (see below), were dialysed to remove the glycerol and then used to set up the crystallisations.

### 3.6 Protein Storage

For long term storage the purified protein was dialysed into 50% v/v glycerol in 50mM Tris.HCl (pH 7.8) over a period of about 48 hours. The resulting protein solution was then aliquoted into 1ml volumes and stored at -20$^0$C. When required for crystallisations the aliquots were dialysed overnight into 50mM Tris.HCl (pH 7.8).

### 3.7 Purification Protocol

#### 3.7.1 Initial Purification

This was carried out by the staff in the Laboratory of Prof. A.R. Hawkins in the Department of Biochemistry and Genetics, University of Newcastle upon Tyne.

SK3430/pK44 cells were grown at 37$^o$C to mid log phase in 7 litres of a complete medium containing 20g of bactotryptone, 10g of yeast extract and 10g of NaCl /litre, in orbital shakers at 37$^0$C. Then isopropyl-b-thiogalactoside (200 mg/ml) was added and growth continued for a further 12 hours. Bacteria were harvested by centrifugation (2500 X g) at 4°C and re-suspended in 90 ml of 50mM Tris.HCl (pH 7.8) containing 1mM EDTA, 1mM DTT, 1mM benzamidine and lysed by sonication at 0°C. The insoluble cell debris was pelleted by centrifugation (4$^o$C, 6000 X g) and the supernatant volume was made up to 170ml with 50mM Tris.HCl (pH 7.8). This was loaded onto a Q Sepharose (fast flow) column (180ml) which had been previously equilibrated with 50mM Tris.HCl (pH 7.8) . Following washing with 3 column volumes of 50mM Tris.HCl (pH 7.8) the protein was eluted with a 250mM to 750mM KCl gradient in 50mM Tris.HCl (pH 7.8) containing 1mM-EDTA, 1mM DTT, 1mM benzamidine. Fractions of 13 ml were collected and those containing greater than 15 units/ml of dehydroquinase activity were pooled.

Ammonium sulphate was added to these pooled fractions to a concentration of 1.1 M and the pH adjusted to 6.8 using $NaH_2PO_4$. The protein solution was then loaded onto a Phenyl-Sepharose (fast flow, low substitution) column (100ml) equilibrated with 50mM potassium phosphate (pH6.8) containing 1mM DTT and 1.4M ammonium sulphate. The column was washed through with three column volumes of 50mM-potassium phosphate (pH6.8) containing 1mM DTT and 1.4M ammonium sulphate and then the dehydroquinase activity was eluted with a 1.4 to 0.0M gradient of ammonium sulphate in 800ml of 50mM-potassium phosphate (pH6.8), containing 1mM DTT with fractions of 13ml being collected. Fractions containing an activity greater than 12 units/per ml were pooled and desalted on a Sephadex G-25 column (200ml) equilibrated in 20mM-potassium phosphate (pH6.8) containing 1mM DTT buffer.

### 3.7.2 Final Purification

The partially purified sample was then further purified in Glasgow to give protein suitable for crystallisation.

The desalted sample from the Phenyl-Sepharose column was loaded onto a hydroxyapatite column (50ml) which had been previously equilibrated with 20mM potassium phosphate (pH6.8) containing 1mM DTT buffer. After washing with seven column volumes to remove any unbound protein, the column was eluted using a 0 to 400mM potassium phosphate gradient in 20mM potassium phosphate (pH 6.8), containing 1mM DTT. 10ml fractions were collected and absorbance at 280nm was used to locate the dehydroquinase containing fractions. Since the enzyme is eluted in a large amount of phosphate and phosphate is a competitive inhibitor of the type II dehydroquinase (Kleanthous *et al.*, 1992) using the enzyme assay to test for

dehydroquinase containing fractions is not satisfactory in this case. Fractions which had an $A_{280}$ greater than 0.2 were pooled and dialysed against 50mM Tris.HCl (pH 7.8) overnight and then loaded onto a Neobar$^{TM}$ AQ$^{4}$ anion exchange column on a Pharmacia FPLC $^{TM}$ system which had been previously equilibrated with 50mM Tris.HCl (pH 7.8). The column was then washed through with four column volumes of 50mM Tris.HCl (pH 7.8). Then a 0.0 to 1.0M NaCl gradient in 80ml of 50mM Tris/HCl, pH 7.8 was used to elute the protein. 1ml fractions were collected and those fractions containing protein as shown by the sharp single peak in the $A_{280}$ trace from the FPLC $^{TM}$ detector were desalted by dialysis and concentrated using a Centricon 10 cell. Protein was either used immediately to set up crystallisations or it was dialysed against 50 % (v/v) glycerol-containing buffer (see section 3.7) for long term storage.

### 3.8 Purification Results

The specific activity of the protein increased significantly throughout the purification protocol, see Table 3.1.

After the final purification described above the protein was judged to be pure by SDS-PAGE stained with Coomassie blue as can be seen in Fig. 3.1. The high purity of the purified protein was seen in spectra from the electrospray mass spectrometry (Fig 3.2a,b). MaxEnt spectra deconvoluted from the raw data from electrospray mass spectometry can be used for a semi-quantitative anlysis by comparing integrated peak-areas (Ferrige, Seddon, Green, Jarvis & Skilling, 1992).The peak in this spectra with the molecular weight ($M_r$) 15661.0 corresponds to the molecular weight of DHQ from M. tuberculosis with the N-terminal methionine residue cleaved off. The small peak to the right of the major peak is a potassium adduct (mass difference 37 Da).

| | Vol (ml) | Conc (mg/ml) | Total Protein (mg) | Activity (units/ml) | Total Activity (units) | Specific Activity (units/mg) | Purification (fold) | Yield (%) |
|---|---|---|---|---|---|---|---|---|
| Newcastle | 150 | 1.34 | 201 | 43.25 | 6487.5 | 32.3 | 1 | 100.0 |
| hydroxyapatite | 70 | 2.32 | 163 | 88.95 | 6226.6 | 38.2 | 1.2 | 96.0 |
| Neobar | 12 | 11.66 | 140 | 502.80 | 6034.0 | 43.1 | 1.3 | 93.0 |
| Superose 12 | 60 | 2.20 | 132 | 95.7 | 5742.0 | 43.5 | 1.3 | 88.5 |

**Table 3-1 Purification table for M. tuberculosis Dehydroquinase (Practical details are given in section 3.8)**

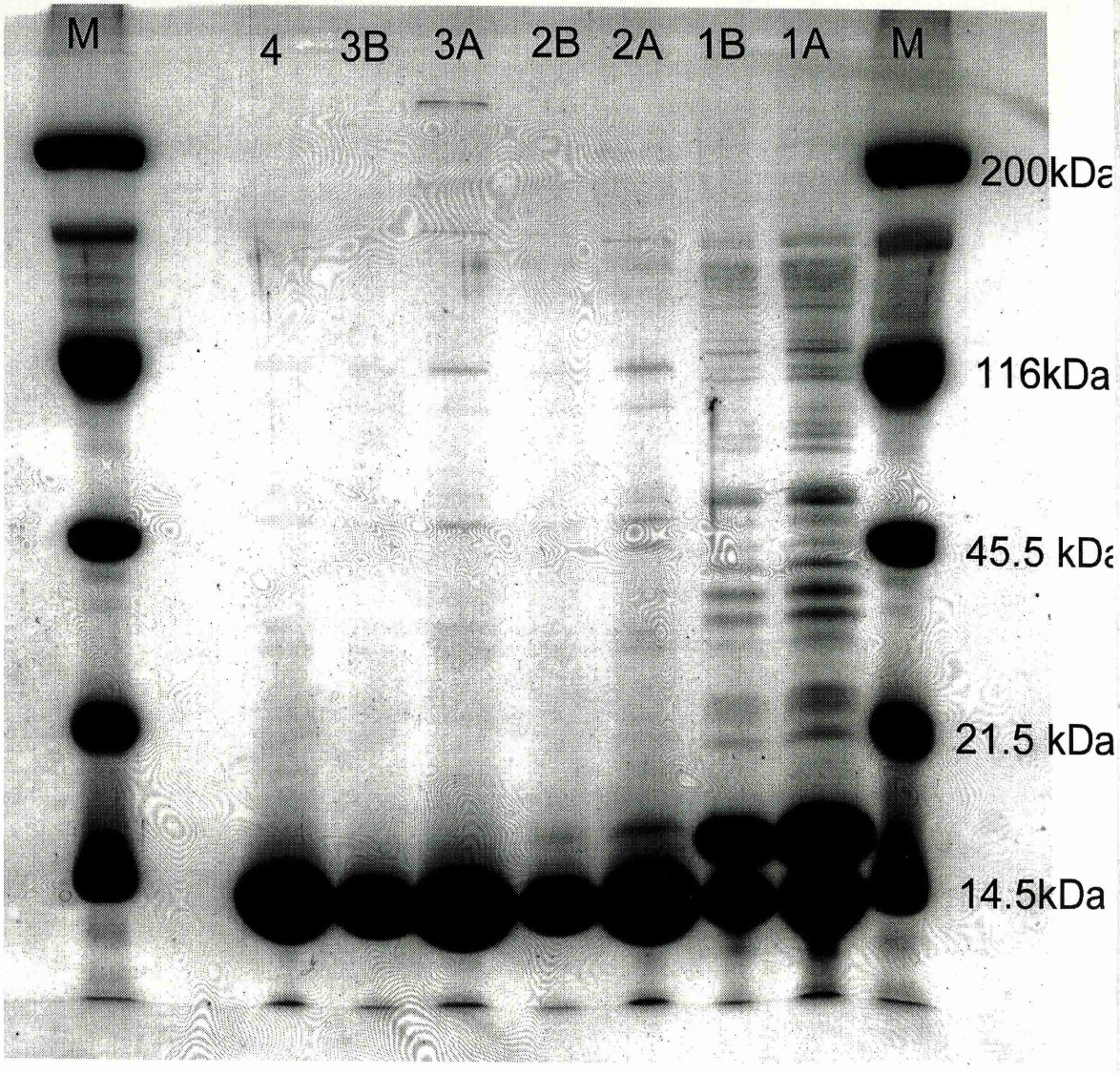
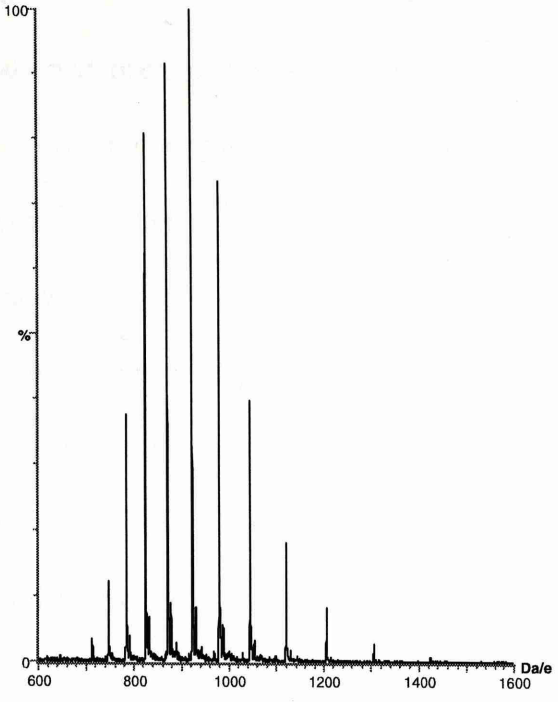



**Figure 3-1 SDS-Page Gel (15 %)**

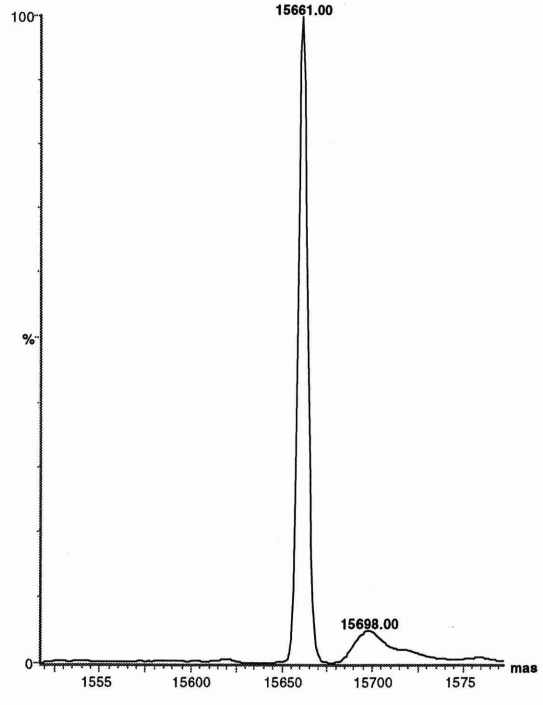
**M-Markers**

1A - 60 µg Protein from Newcastle
1B-20µg Protein from Newcastle
2A-60µg Protein from hydroxyapatite column
2B-20µg Protein from hydroxyapatite column
3A-60µg Protein from Neobar column
3B-20µg Protein from Neobar column
4-60µg Purified Protein




mass/charge

**Figure 3-2 Raw Data From Electospray Mass Spectrometry**




$M_r$

**Figure 3-3 MaxEnt Deconvoluted Mass Spectra**

### 3.9 Further Purification

After good crystals had been grown and a native data set collected (see Chapter 5) it was felt that the reproducibility and the quality of diffraction of the crystals might be improved by further purifying the protein. It had been noticed that, when all other conditions were kept the same and a Centricon 30 cell was used instead of a Centricon 10 cell, the resulting protein gave bigger and better diffracting crystal as well as a smaller percentage of mosaic crystals per batch. The ultrafiltration may have reduced any micro-heterogeneity of the protein due to degradation and would certainly have removed any traces of monomer in the normally dodecameric protein. This observation led to the introduction of gel filtration an additional purification step.

Approximately one year into the project the purification procedure was modified to include a gel filtration step. Protein fractions from the Neobar column step were concentrated to about 20 mg/ml by ultrafiltration in a Centricon cell and 0.5 ml samples were loaded on to a Pharmacia Superose-12 column which had previously been washed with 50mM Tris.HCl (pH 7.8). The column was run at a flow rate of 0.5 ml/min and 0.5 ml fractions were collected. A single large peak was seen in the $A_{280}$ trace after about 32 minutes and the fractions containing this peak were found to have high dehydroquinase activity. Protein purified with this extra purification step was again concentrated to the required concentration by ultrafiltration and used for crystallisation trials (see Chapter 4).

# 4. Crystallisation of Dehydroquinase from M. tuberculosis

## 4.1 Crystallisation of Proteins

When determining the structure of a protein using X-ray diffraction, single, well ordered crystals have to be grown. These have to be large, usually between 0.2mm-1.0mm in all dimensions, although smaller crystals of size less than 10μm in two dimensions have been used successfully for structure determination when very powerful synchrotron sources and flash freezing techniques are used (Xiao *et al.*, 1996).

Crystallisation of macromolecules is usually accomplished by a trial and error approach using a large body of empirically determined procedures and conditions (MacPherson, 1982).

When proteins are in an aqueous environment they minimise their free energy by becoming fully solvated. If the properties of the solvent or solute (protein) are altered so there is insufficient water to maintain solvation a thermodynamically unstable supersaturated state is formed. When the system returns to equilibrium the protein molecules may aggregate to form an amorphous precipitate or, if the molecules have enough time to form regular arrays, they can crystallise. Supersaturation can be achieved by lowering the protein's solubility, for example by altering the temperature, the pH and the ionic strength of the medium.

Decreasing the amount of water available to solvate the protein can also lead to supersaturation. This can be done by adding inorganic salts which compete with the protein for water of solvation. The concentration of water can also be effectively reduced by adding organic solvents like ethanol or 2-methyl-3,4-pentanediol (MPD)

which also compete for water. Polymers such as polyethylene glycol also serve to dehydrate proteins in solution as well as altering the solvents dielectric constant.

Crystallisation occurs in three phases, nucleation, growth and cessation of growth. Nucleation is the first stage of crystallisation and the amount of nucleation increases with the degree of supersaturation. There can be too many nucleation sites resulting in showers of small crystals unsuitable for X-ray study if there is a too high a degree of supersaturation. So supersaturation must be achieved slowly and when a low degree of superstaturation has been obtained it should not be allowed to increase. The growth phase is when molecules form layers parallel to crystal planes and cessation of crystal growth occurs either because the supersaturation level of the protein falls below that required for crystal growth or the faces of the crystal become defective or poisoned.

## 4.2 Vapour Diffusion

In this project the vapour diffusion method was used to achieve supersaturation. Vapour diffusion occurs in a sealed environment where there is a reservoir of solution containing a high concentration of precipitant and a drop, not in liquid contact with the reservoir, containing a protein solution. This drop has a significantly lower volume and concentration of precipitant than the reservoir. The concentration of precipitant in the drop equilibrates through the vapour phase with that in the reservoir. This not only gives a higher precipitant concentration but also increases the protein concentration in solution. This leads to a deficit of water for full solvation of the protein and the desired state of supersaturation is achieved. The

sitting drop method which uses the technique of vapour diffusion was used in this project to slowly approach a low degree of supersaturation.

### 4.3 Preliminary Crystallisation Trials

Work previously done on the Type II dehydroquinase from *M. tuberculosis* had produced micro-crystals (unsuitable for X-ray diffraction purposes) by varying concentrations of PEG 4000 and 6000, in 50mM-200mM Tris/HCl buffer in a pH range from 6.8 to 7.8 and using various concentration of enzyme (A. Anderson personal communication). It was decided to set up systematic crystallisation trials using the sitting drop method and to vary the 3 parameters of precipitant concentration, temperature and solution pH. The concentration of PEG 6000 varied from 5% (w/v) to 25% (w/v) in steps of 5% (w/v). The buffer used was 200mM Tris/HCl and the pH of the solutions was varied from 6.8 to 8.0 in steps of 0.4 of a pH unit. Each of the conditions was tried in duplicate with one at 20°C and the other at 4°C.

### 4.4 Results

The resulting crystal growths from these trials are shown in Tables 4.1, and 4.2. Two minutes after crystallisation were set up many small micro-crystals appeared in all the wells at room temperature which had a PEG 6000 concentration of 25%. A photograph of these micro crystals is shown in Fig 4.1. Within two days all the wells with a PEG 6000 concentration of 20% had slightly larger micro-crystals in them. Over the next two weeks micro-crystals appeared in the well containing 15% PEG at pH 8.0 and pH 7.6. Showers of small micro crystals appeared in the wells containing

15% PEG at pH 6.8 and pH 7.2 about a month from when they were set up. Their dimensions were of 0.1 mm x 0.1 mm x 0.1 mm and diffraction from them was observed to 8 Å on the Siemens area detector. In the trials set up at 4°C only micro crystals or dendritic poly-morphic crystals were observed.

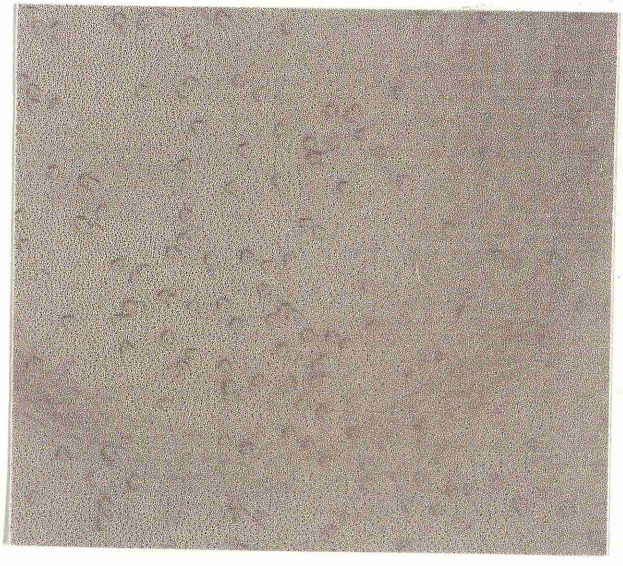


**Figure 4-1 Microcrystals Of Type II DHQ**

| pH | % PEG | Crystal Formation | Time Taken |
|---|---|---|---|
| 6.8-8.0 | 20-25 | micro crystals | 1min - 2days |
| 6.8-8.0 | 15 | small crystals | 2-4 weeks |
| 6.8-8.0 | 10 | large crystals | 4 months |
| 6.8-8.0 | 5 | none | |

**Table 4-1Crystallisation at 20°C**

| pH | % PEG | Crystal Formation | Time Taken |
|---|---|---|---|
| 6.8-8.0 | 20-25 | micro crystals | 2days |
| 6.8-8.0 | 15 | small crystals | 1 month |
| 6.8-8.0 | 5-10 | none | |

**Table 4-2 Crystallisations At 4 °C**

In the wells in which microcrystals appeared dehydration of the protein solution had occurred too quickly and the protein had become supersaturated to too high a degree so too many sites of nucleation had occurred. The same thing to a lesser degree had happened in the wells containing the poorly diffracting small crystals. It was thought that if dehydration could be achieved more slowly then bigger crystals might be obtained. It was decided to try to improve the size of the crystals that were grown at 20°C in 15 % PEG at pH 6.8 and pH 7.2. This was done by setting up trials where the % PEG was reduced in small steps of 1% between 15 % (w/v) and 10 % (w/v)

Between one and two months after the second crystallisation trials were set up crystals appeared in all the wells. A summary of the results is given in Table 4.3.

| pH | % PEG | Crystal Size (mm) | Time Taken |
|---|---|---|---|
| 6.8-7.2 | 14-15 | 0.1 x 0.1 x 0.1 | 6-8 weeks |
| 6.8 | 12-13 | 0.15 x 0.15 x 0.15 | 6-8 weeks |
| 6.8 | 11 | 0.2 x 0.2 x 0.2 | 12 weeks |
| 6.8-7.2 | 10 | 0.5 x 0.5 x 0.5 | 6 months |
| 7.2 | 13 | 0.15 x 0.15 x 0.5 | 6 weeks |
| 7.2 | 11.-12 | 0.2 x 0.2 x 0.2 | 8 weeks |

**Table 4-3 Crystallisation At 20°C**

After eight weeks crystals of size 0.2mm x 0.2mm x 0.2mm were obtained (Fig 4.2, Table 4.3). These crystals diffracted to a resolution of 4.0 Å. Data were collected from this crystal to characterise it (see Chap. 5).

Crystallisations were set up varying the pH of the solution in fine steps and by altering the protein concentration. This trial was set up with the pH varying from 6.8 to 7.2 in steps of 0.05 pH units, with protein concentrations of 5 $mgml^{-1}$, 10 $mgml^{-1}$, 15 $mgml^{-1}$, 20 $mgml^{-1}$ and 25$mgml^{-1}$.

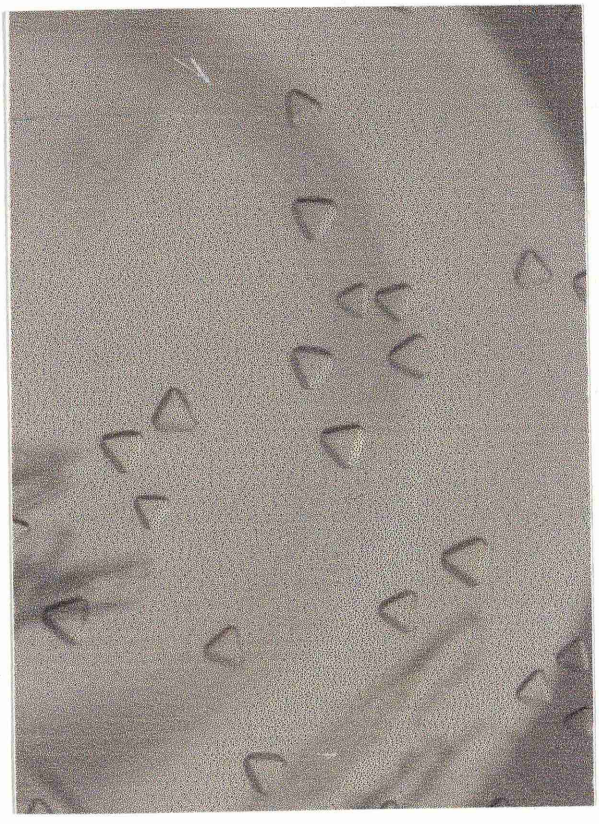


**Figure 4-2 Small Crystals of Type II DHQ**

| pH | Protein conc. (mg/ml) | Crystal Size | Crystal morpholgy | Time Taken |
|---|---|---|---|---|
| 6.80-6.95 | 25 | 0.3 x 0.3 x 0.3 | good | 7 weeks |
| 7.00-7.20 | 25 | 0.4 x 0.4 x 0.4 | twinned | 8 weeks |
| 6.80-6.95 | 20 | 0.4 x 0.4 x 0.4 | good | 8 weeks |
| 7.00-7.20 | 20 | 0.5 x 0.5 x 0.5 | good | 8 weeks |

**Table 4-4 Crystallisation At 20 °C**

| pH | Protein conc. (mg/ml) | Crystal Size | Crystal morpholgy | Time Taken |
|---|---|---|---|---|
| 6.80-6.85 | 15 | 0.4 x 0.4 x 0.4 | good | 8 weeks |
| 6.90-7.05 | 15 | 0.3 x 0.3 x 0.3 | twinned | 8 weeks |
| 7.10-7.20 | 15 | 0.4 x 0.4 x 0.4 | good | 8 weeks |
| 6.80-7.20 | 10 | 0.2 x 0.2 x 0.2 | good | 8 weeks |
| 6.80-7.20 | 15 | 0.1 x 0.1 x 0.1 | good | 8 weeks |

**Table 4-5 Crystallisation At 20 °C**

The results of this trial are shown in Tables 4.4 and 4.5. It can be seen that small changes in pH ( of the order 0.05 of a pH unit) do not have a significant affect on crystal size and quality. The best conditions for crystal growth were 12% (w/v) PEG 6000 at pH 7.1. These conditions gave crystals that were 0.5 x 0.5 x 0.5 mm in size (see Fig 4.2) and when mounted on the Siemens Area detector were observed to diffract to a resolution of 2.2Å. A full data set was collected to this resolution ( Gourley *et al.,* 1994, see chap. 5) on the Siemens area detector.

**Figure 4-3 Crystals Of Type II Dehydroquinase**

## 4.5 Conclusion

The conditions for reproducible growth of well diffracting crystals had been found. These crystals were considered good enough to used in a full structural study using X-ray crystallography by the multiple isomorphous replacement method.

# 5. Data Collection and Processing

## 5.1 Introduction

Data collection on native crystals of the Type II dehydroquinase from *M. tuberculosis* was initially done in house on the Siemens area detector. Later when a high resolution native set was required data were collected at station 9.5 at S.R.S., Daresbury.

## 5.2 Data Collection Equipment

During the last decade there has been a revolution in X-ray crystallographic data collection technology resulting in an enormous increase in data acquisition rates and in the range of macromolecules which can be investigated. This revolution has involved the development of more powerful and reliable rotating anode generators. However the most important advances have been in the development of X-ray area detectors and the development and increased use of dedicated synchrotron beam lines for macromolecular crystallography.

### 5.2.1 Rotating Anodes

Modern rotating anode generators are very powerful with regard to the number of X-ray photons/unit area/second. They consist of an anode, usually a drum made of copper, rotating at high speed whilst being bombarded by a beam of focussed electrons emitted from a heated tungsten filament. The whole assembly is contained in a high vacum (less than $10^{-5}$ Pa). When the electrons hit the copper anode a beam of X-rays are produced. Then either a graphite monochromator crystal is used to

select for a single wavelength of X-ray (usually $Cu_{K\alpha}$, $\lambda$= 1.542Å) or mirrors are used to focus and monochromate the beam. In this project a Siemens high power rotating anode generator was used.

### 5.2.2 Synchrotron Radiation

Synchrotron storage rings represent unique facilities for the determination of macromolecular structures. X-rays from a synchrotron can be thousands of times more intense than those from a rotating anode source. This means that crystals usually diffract to higher resolution at a synchrotron and data can be collected much more rapidly. Over the last decade, synchrotron radiation has been exploited in macromolecular crystallography with impressive results.

Synchrotrons are storage rings where accelerated charged particles (electrons or positrons) circulate in bunches. The paths of these particles are defined by bending magnets and the bunch shape is defined by quadrapole or sextapole magnets. X-rays and other radiation is emitted when these charged particles change velocity. This happens when they curve through a bending magnet or pass through a wiggler or undulator in a straight section of the storage ring. The curvature of the bending magnets is the major factor in determining the spectrum of radiation released. Wigglers are composed of a series of magnets that cause the beam particle to undergo multiple transverse excursions from the normal pathway. The small radius of curvature of the wiggler magnets causes the energy of the synchrotron radiation to be increased. An undulator consists of a periodic arrangement of magnets that provides constructive interference of radiation emitted from individual poles. Undulators produce the most intense X-rays but only in narrow energy ranges.

Synchrotrons produce a continuous X-ray spectrum from a few tenths of an angstrom to a few angstroms in wavelength. Consequently synchrotron sources can be used to collect data at selected wavelengths. This wavelength tuning can be used to help maximise the anomalous signal for MIRAS, SIRAS and MAD experiments. Maximising the anomalous signal has led to a reduction in the number of heavy atom derivatives that are needed to obtain adequate phasing.

### 5.2.3 X-ray Detectors

Collection of X-ray data on film with the time consuming wet development and digital scanning has been superseded by area detectors that give quick digital output of the diffraction. Collection of data and processing of data is much quicker and simpler. Two types of area detector were used in this work: a multi-wire proportional counter made by Siemens and Image Plate detectors made by Mar.

### 5.2.4 The Siemens Area Detector

The Siemens area detector (Durbin *et al.,* 1986; Howard *et al.,* 1987) has a sealed chamber containing a gas mixture (mostly composed of xenon) at four atmospheres pressure. Also in the chamber there are three wire planes. Two of these are the front and back cathode grids, which are at ground potetinal, have a wire spacing of 0.3mm and are perpendicular to one another. The third is the anode grid that is centrally placed, has a wire spacing of 1.0 mm and is held at a potential of 5kV. When X-rays hit the xenon, the xenon ionises and the ions pairs are accelerated towards the wires of the anode and cathodes, ionising more xenon on the way. This has the effect of multiplying the original signal by a factor of $10^5$ which allows it to be detected electronically. A capacitative readout is used to determine positional

information for the incident X-rays. The active area of the detector is 11.5cm in diameter and the X-rays gain entry through a 1mm thick beryllium window. The positional resolution of the detector is 200 μm in both the x and y directions. The array of pixels is 512 x 512, with each pixel representing 200 x 200 $\mu m^2$. The output of the detector is the x-y co-ordinate of the incident X-ray, and the appropriate pixel is then incremented by one count in the computer memory.

### 5.2.5 Mar Image Plates

Mar image plates ( Hendrix *et al.,* 1988) come in two sizes the small-Mar, which is 18cm and the big-Mar which is 30 cm. Image plates are made from a 150 μm thick layer of barium halide phosphor (BaFBr) doped with europium ($Eu^{2+}$), mounted on a flexible plastic sheet and sealed in a light tight box with a front X-ray transparent window. The image plate works in the following way. An image is recorded by incident X-rays creating colour centres, which are metastable states of trapped electrons in the BaFBr:Eu-X-phosphor. An excitation energy of 6eV is required to create a colour centre. The image is then read by illuminating the plate with a red He-Ne laser ($\lambda$=633nm). Upon absorption of a photon of this red He-Ne light, the meta stable electron returns to its ground state under emission of a violet light (4eV, $\lambda$=390mm). The intensity of this blue stimulated luminescence is proportional to the intensity of absorbed X-rays. The luminescence emitted is collected by a ellipsoidal collector. A $160^0$ solid angle of luminescence is reflected on to a photomultiplier through an appropriate set of optical filters. The illuminating of the plate and the measuring of the emitted light is done on a pixel by pixel basis using a combined laser and reading head. The intensity of the spot can be calculated from the reading on the photomuliplier.

Mar image plates have a spiral readout method where the image plate rotates at increasing rate while the scanning head moves inwards along a radius at constant speed. After data are read out they are transformed into Cartesian co-ordinates with a pixel size of 150 x 150 $\mu m^2$. Reading and erasing takes about 2 minutes for the small-Mar and about 4 minutes for the big-Mar.

## 5.3 Data Processing

Once data have been collected they have to be processed to reduce the raw data to a list of reflection indices and corresponding measured intensities. Data reduction generally involves the following steps:

- Identification of intense well defined spots over a specific rotation range.
- Autoindexing of these spots to determine orientation of the cell and to determine the space group and cell dimensions.
- Predicition of the reflection intensities from the refined parameters and the space group.
- The integration of the reflection intensities either by summation or profile fitting.
- Reduction of the reflection list to a unique and ordered reflection list for the specified space group.
- Scaling of the data to take account of the Lorentz factor, air and crystal absorption due to asymmetry in the crystal shape, and the crystal decay during data collection
- Output of final data statistics and a list of reflections with measured intensities and associated statisitcal errors.
- Post-refinement of the unit cell, detector parameters and central beam position using all the reflections collected over the whole data set.

### 5.3.1 XDS

The XDS package (Kabsch, 1988a,b) was used to process data collected on the Siemens area detector. XDS contains eight programs; XYCOOR, INIT, COLSPOT, IDXREF, COLPROF, PROFIT, CORRECT and GLOREF. XYCOOR processes and uses a brass plate frame to give a pixel to centimetre calibration for the detector. INIT determines the initial background, the anode wire modulation correction factors. The identification of bright well defined spots in the first few rotation frames that cover a total rotation range of $5^{o}$ is done in the program called COLSPOT which makes a SPOT.XDS file. IDXREF uses this file to determine the orientation matrix for the crystal. This auto-indexing procedure (Kabsch, 1993) involves performing a search of difference vectors between reciprocal lattice points, after the vectors have been histogramed, filtered and refined. This procedure can find both the orientation matrix and the cell dimensions. IDXREF lists all 14 lattice type with a rating of fit for inspection of the crystallographer. This list indicates lattice types of the most probable space group.

Profile fitting (Diamond 1969) in three dimensions on a background is used to integrate the reflection intensities. An empirically derived model reflection shape is scaled to the data and then integrated. This procedure assumes strong and weak reflections have the same normalised profile. The observed profiles vary over the detector face, so nine model profiles are usually used and the intensity for each reflection is calculated using the nearest reference profile. COLPROF collects three-dimensional reflection profiles which are used by PROFIT to estimate reflection intensities. CORRECT corrects the intensities for decay , absorption and detector surface sensitivity. GLOREF uses all the observed reflections and refines all the

diffraction parameters. An unknown space group can be determined by repeating CORRECT and GLOREF for each of the most probable space groups. This is done using the appropriate reindexing transformation and conventional cell constants found in the table printed in the IDXREF step. CORRECT prints out the R-factor statistics between symmetry-related reflections, and the intensities of all reflections with indices of type *h*00, 0*k*0 and 00*l*. Comparison of these for values for the most probable space groups should indicate for most cases the correct space group.

XSCALE, also written by Dr. W. Kabsch, was used to scale data and merge symmetry equivalent reflections. It calculates and applies scale factors and adds together partially recorded reflections, monitors and rejects bad agreements between repeated and symmetry equivalent measurements and then averages them for output.

### 5.3.2 DENZO and XDISPLAYF

The program DENZO (Otwinowski 1993) in combination with the program XDISPLAYF was used to process all the data collected on the Mar image plates. This process follows the general steps for data reduction as mentioned in section 5.3.1 There are however several important differences in the way DENZO and XDS process data. A major difference is that DENZO is run interactively using XDISPLAYF to display a diffraction image which is used for preliminary analysis, indexing and then refining the crystal and detector parameters. This enables the user to monitor and control the initial selection of intense spots to be used by the autoindexing subroutine.

Another important difference between DENZO and XDS is in the autoindexing. In DENZO the strong reflections that are selected to be used in autoindexing come from a single frame. The autoindexing in DENZO is based on a

novel algorithm which involves a complete search of all the possible indices of all reflections selected by the peak searching routine.

A Fast Fourier transform (FFT) algorithm is used to do this search and it takes advantage of the fact that finding all values of one index (e.g. $h$) is independent of finding all values of another index (e.g. $k$). When the program finds values for one index for all the reflections, this is equivalent to finding the real space direction of the crystal. This real space indexing method assigns indices to all reflections simultaneously. This means that a small percentage of incorrectly identified diffraction maxima will not affect this method.

The real space indexing method, unlike reciprocal space methods, is also insensitive to how many short difference vector can be created from the list generated by the peak searching routine. DENZO does a complete search of all the possible real space vectors within a reasonable length range.

When the search for real space vectors is completed, the program finds the three best linearly independent vectors, with minimal unit cell volume that would index all of the observed peaks. Cell reduction is then done to find a standard basis for a description of the unit cell. The best cells are found for all 14 Bravais lattices. DENZO calculates an index of distortion required to fit the best triclinic lattice into the higher symmetry lattices. The user is then left to decide which is the correct lattice and space group.

In DENZO, after autoindexing, the initial crystal and detector orientation parameters are refined for each processed image. The user defines the resolution limits and order and number of parameters to be fitted. Fitting of the parameters is done by the least squares method and is fast converging.

Integration of reflection intensities in DENZO, like XDS, is done using profile fitting. However, while in XDS profiles are fitted over three-dimensions, in DENZO profile fitting is done on reflections on individual images and so is only done in two-dimensions.

SCALEPACK was used to scale and merge the measurements from each image to a unique and ordered reflection list for the specified space group. It also globally refines the crystal parameters.

## 5.4 Collection of Initial Native Data Set

Data were collected on the first crystal of DHQ that was grown that diffracted far enough to allow for a full characterisation. This crystal diffracted to a resolution of 4 Å. As the space group and cell dimensions were as yet unknown, data were collected with the crystal to detector distance at 160mm doing a $\phi$ scan with a $\Delta\phi$= 0.25° as when this was done in a test frame the reflections were spaced far enough apart to allow for processing. The $2\theta$ angle was slightly off-set at 358° to allow all data to be collected to the limit of diffraction of the crystal. The exposure time for each frame was 300 sec and 360 frames were collected.

For this data the IDXREF output showed the Bravais lattice with highest symmetry and the best fit was face centred cubic, which gave cell dimensions of $a=b=c$=127.95 Å. CORRECT and GLOREF were rerun as described above for the three space groups with a face centred cubic Bravais lattice F23 F432 and F4(1)32. It was found that the space group was F23 rather than F432 or F4(1)32 as CORRECT rejected as misfit 915 reflections from 6037 when run in F432 or F4(1)32 as opposed

to only 16 when run in F23. Also by comparing the merging R factors on intensity the $R_{merge}$;

$$R_{merge} = \frac{\sum_h \sum_i |(I_h - I_{hi})|}{\sum_h \sum_i I_{hi}}$$

where $I_h$ is the weighted mean measured intensities of the observations $I_{hi}$ in which the intensities of the symmetry related reflections, which ideally should be the same, are compared. $R_{merge}$ ($I$) gives an estimate of their disagreement.

When CORRECT and GLOREF was run with the data reindexed as F23 the $R_{merge}$ ($I$) was 6.3 % after scaling as opposed to 21.6 % when CORRECT and GLOREF were repeated with the data reindexed as F432 or F4(1) 32. Hence it was concluded that the space group of these crystals of *M. tuberculosis* DHQ was F23 with $a = b = c = 127.95$ Å.

Scaling was done using XSCALE and statistics from this scaling and merging are shown in Table 5.1.

| Dmax (Å) | Dmin (Å) | $N_{obs}$ | $N_{uniq}$ | Completeness % | Multi | $R_{merge}$ | $\frac{\bar{I}}{\sigma(\bar{I})}$ |
|---|---|---|---|---|---|---|---|
| 31.94 | 11.84 | 353 | 69 | 100 | 5.1 | 0.057 | 9.4 |
| 11.84 | 8.68 | 548 | 96 | 99.5 | 5.7 | 0.053 | 12.6 |
| 8.68 | 7.17 | 655 | 117 | 100 | 5.6 | 0.054 | 12.2 |
| 7.17 | 6.25 | 667 | 134 | 96.3 | 5.0 | 0.078 | 9.1 |
| 6.25 | 5.61 | 716 | 155 | 100 | 4.6 | 0.080 | 8.9 |
| 5.61 | 5.14 | 722 | 170 | 100 | 4.2 | 0.078 | 9.3 |
| 5.14 | 4.76 | 752 | 175 | 97.0 | 4.3 | 0.069 | 10.2 |
| 4.76 | 4.46 | 821 | 198 | 100 | 4.1 | 0.073 | 9.5 |
| 4.46 | 4.21 | 871 | 210 | 100 | 4.1 | 0.081 | 8.5 |
| 4.21 | 3.98 | 822 | 215 | 100 | 3.8 | 0.091 | 8.0 |
|  | total | 6927 | 1539 | 99.7 | 4.5 | 0.071 | 9.43 |

**Table 5-1 Merging Statistics For The 4 Å Native Data Set**

| | |
|---|---|
| Dmin | maximum resolution |
| Dmax | minimum resolution |
| $N_{obs}$ | number of observed reflections |
| $N_{uniq}$ | number of unique reflections |
| Multi | multiplicity of data |

The resulting file from XSCALE containing a list of indices with measured intensities was converted into mtz format using ROTAPREP, SORTMTZ was then used to sort the reflections and then ROTAVATA and AGROVATA were used with unity scaling to give a file suitable for input into TRUNCATE which was run without truncating to convert the intensities to amplitudes.

The Wilson plot option was used to calculate approximate absolute scale and temperature factor from the observed intensities. This uses the theory of Wilson (1948) which states that if the atoms are randomly distributed through the unit cell then;

$$<< ff >>= scale < f_{obs} >^2 \exp\left(-2B\sin^2\theta/\lambda^2\right)$$

Where $<< ff >>$ is the mean squared atomic form factor summed over the entire cell and $<f_{obs}>$ is the observed amplitude. If a least squares line is fitted throught the plot:

$$\ln\frac{\langle ff\rangle}{\langle f_{obs}^2\rangle} = -2B\frac{\sin^2\theta}{\lambda^2}$$

then the temperature factor B and the scale factor can be found as for this line

B=-(gradient)

scale=exp(-intercept).

The Wilson plot for this data set is shown in Fig. 5.1.

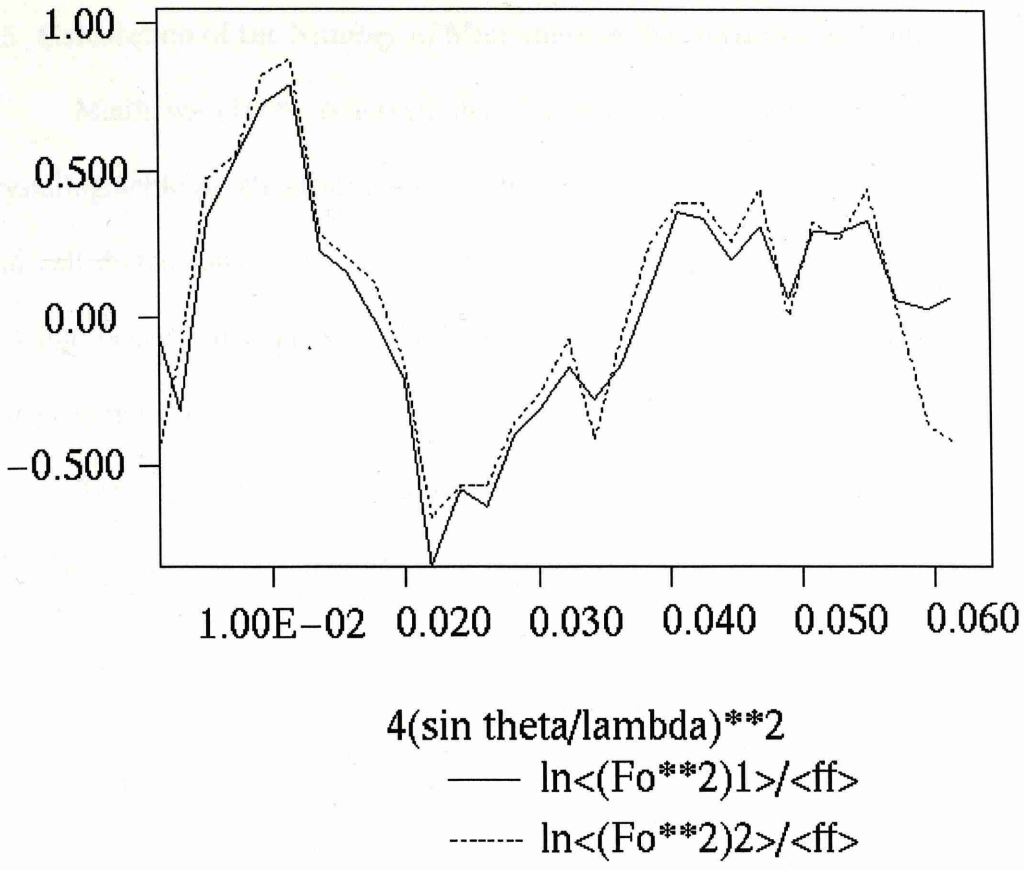



**Figure 5-1 Wilson Plot For The 4 Å Native Data Set**

## 5.5 Calculation of the Number of Monomers in the Asymmetric Unit

Matthews (1968) observed that the probable number of molecules in the crystallographic asymmetric unit could be determined directly from the space group and cell dimensions of the crystal and the molecular weight of the protein. This is possible because the fraction of the crystal volume occupied by solvent lies in the fairly restricted range, from about 27% to 60%.

The crystal volume per unit volume molecular weight ($V_m$ ) was calculated assuming there was one monomer of DHQ in the asymmetric unit;

$$V_m = \frac{\text{volume of asymmetric unit}}{\text{molecular weight of protein in asymmetric unit}}$$

$$\text{Volume of asymmetric unit} = \frac{\text{volume of unit cell}}{\text{number of asymmetric units in unit cell}}$$

$$\text{Volume of asymmetric unit} = \frac{(127.95)^3}{48} \qquad \text{for F23}$$

$$\text{Volume of asymetric unit} = 43639.5 \text{ Å}^3$$

The molecular weight of protein in the asymmetric unit, assuming one monomer of DHQ in asymmetric unit = 15 739 Da

$$V_m = 2.77 \text{ Å}^3/\text{Da}$$

From this, using the approximation given by Matthews, the fraction of volume occupied by proteins ($V_{prot}$) can be calculated;

$$V_{prot} = \frac{1.23}{V_m}$$

$$V_{prot} = 0.445$$

This would then give a solvent content of 55.5% which is in the range of observed solvent contents of proteins.

When we assume there are 2 monomers in the asymmetric unit and repeat the calculation we get $V_m$=1.38 $Å^3$/Da, $V_{prot}$ = 0.891 and a solvent content of 10.9% which is well outside the range Matthews observed for protein crystals.

These calculations suggest that there is one monomer of DHQ in the asymmetric unit and 48 monomers in the unit cell. This supports the biochemical evidence that type II dehydroquinase exists as a dodecamer (Hawkins *et al.,* 1982; Kleanthous *et al.,* 1992) and suggests there are four dodecamers in the unit cell.

## 5.6 Data Collection on a Better Diffracting Crystal

One of the later larger crystals (size 0.5mm x 0.5mm x 0.5mm) was mounted and diffraction was observed using the Siemens area detector to about 2.2 Å. A full data set was collected from this crystal. This data were processed in the same way as the 4Å data set. Details about the data set are shown in Table 5.2. The Wilson plot for this data set is shown in Fig 5.2. The data set was found to be 95.8 % complete to 2.2 Å with an $R_{merge}$ of 5.5 %.

| Dmax (Å) | Dmin (Å) | Nobs | Nuniq | Completeness % | Multi | $R_{merge}$ | $\frac{\bar{I}}{\sigma(\bar{I})}$ |
|---|---|---|---|---|---|---|---|
| 31.94 | 11.0 | 390 | 81 | 100 | 4.8 | 3.1 | 23.8 |
| 11.0 | 7.8 | 806 | 141 | 100 | 5.7 | 3.3 | 22.7 |
| 7.8 | 6.2 | 1243 | 207 | 100 | 6.0 | 3.7 | 21.7 |
| 6.2 | 5.0 | 2151 | 368 | 100 | 5.8 | 3.8 | 20.8 |
| 5.0 | 4.2 | 2999 | 526 | 100 | 5.7 | 3.8 | 20.6 |
| 4.2 | 3.5 | 4938 | 932 | 100 | 5.3 | 4.5 | 18.0 |
| 3.5 | 3.1 | 4494 | 962 | 100 | 4.7 | 5.6 | 14.5 |
| 3.1 | 2.8 | 4845 | 1136 | 100 | 4.3 | 8.3 | 9.7 |
| 2.8 | 2.6 | 4139 | 1055 | 100 | 3.9 | 11.6 | 7.1 |
| 2.6 | 2.4 | 4819 | 1445 | 100 | 3.3 | 14.5 | 5.4 |
| 2.4 | 2.2 | 4572 | 1944 | 95.8 | 2.4 | 21.4 | 3.4 |
|  | total | 35396 | 8797 | 99.31 | 4.0 | 5.5 | 14.3 |

**Table 5-2 Merging Statistics for the 2.2 Å Native Data Set**

Dmin = maximum resolution

Dmax = minimum resolution

Nobs = number of observed reflections

Nuniq = number of unique reflections

Multi = multiplicity of data

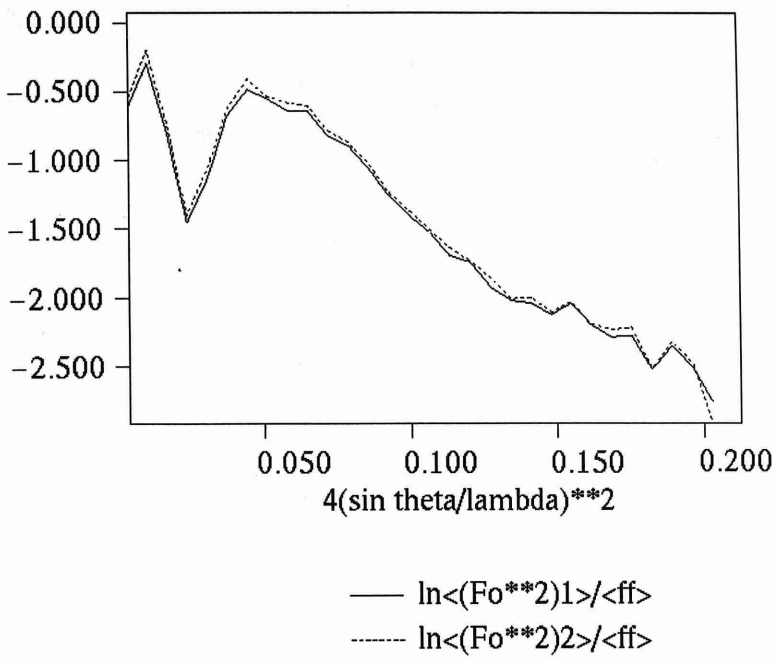



**Figure 5-2 Wilson Plot for 2.2 Å Native Data Set**

## 5.7 Comparison of Native Data Sets

When the two data sets that were collected in-house were merged the statistics on merging were very poor as illustrated in Table 5.3. This was very worrying and when the reflection file of the 2.2Å data set was checked it was found that it exhibited pseudo symmetry where $I_{hkl}=I_{hlk}$. It was found that the 2.2 Å data set could be indexed as F432 with an $R_{merge}$ of 14 % and with few reflections being rejected as misfits (3.4%). It is very unlikely that the crystal from which the 2.2 Å data was collected was F432 as it would have given values of $V_m$=1.38 $Å^3$/da , $V_{prot}$ = 0.891 and a solvent content of 10.9 % which are values regarded as unlikely to be found in protein crystals. It was thought that this pseudo symmetry was the result of systematic twinning of the crystal.

| Dmax (Å) | Dmin (Å) | Nobs | Nuniq | Completeness % | Multi | $R_{merge}$ |
|---|---|---|---|---|---|---|
| 31.94 | 11.0 | 1087 | 81 | 100 | 13.5 | 10.1 |
| 11.0 | 7.8 | 2238 | 141 | 100 | 15.8 | 12.4 |
| 7.8 | 6.2 | 2643 | 207 | 100 | 12.8 | 10.3 |
| 6.2 | 5.0 | 2787 | 368 | 100 | 7.6 | 15.7 |
| 5.0 | 4.2 | 3651 | 526 | 100 | 6.9 | 13.4 |
| 4.2 | 3.5 | 6430 | 932 | 100 | 6.9 | 12.4 |
| 3.5 | 3.1 | 6638 | 962 | 100 | 6.9 | 10.6 |
| 3.1 | 2.8 | 8494 | 1136 | 100 | 4.3 | 10.3 |
| 2.8 | 2.6 | 4139 | 1055 | 100 | 3.9 | 11.6 |
| 2.6 | 2.4 | 4819 | 1445 | 100 | 3.3 | 15.2 |
| 2.4 | 2.2 | 4572 | 1944 | 95.8 | 2.4 | 21.9 |
| | total | 41628 | 8797 | 99.31 | 5.0 | 13.5 |

**Table 5-3 Merging Statistics for the 4Å and 2.2 Å Native Data Sets**

Dmin = maximum resolution

Dmax = minimum resolution

Nobs = number of observed reflections

Nuniq = number of unique reflections

Multi = multiplicity of data

It has been observed that sometimes when a crystal is twinned the observed distribution of acentric intensities do not follow the exponential curve for the theoretical distribution but instead have a sigmoidal curve (Gomis-Rü *et al.*,1995). The cumulative distribution function of the 4Å data set is shown in Fig 5.3. It can be seen from this data set where twinning is not suspected that the observed acentric reflections follow an exponential curve and do not deviate significantly from the curve for the theoretical distribution. However this is not the case for the 2.2 Å data as can be seen in Fig 5.4, where the observed distribution for acentric reflections does not follow the theoretical distribution but instead gives a sigmoidal curve. For this reason the 2.2 Å data set was suspected of being twinned and was not used further.

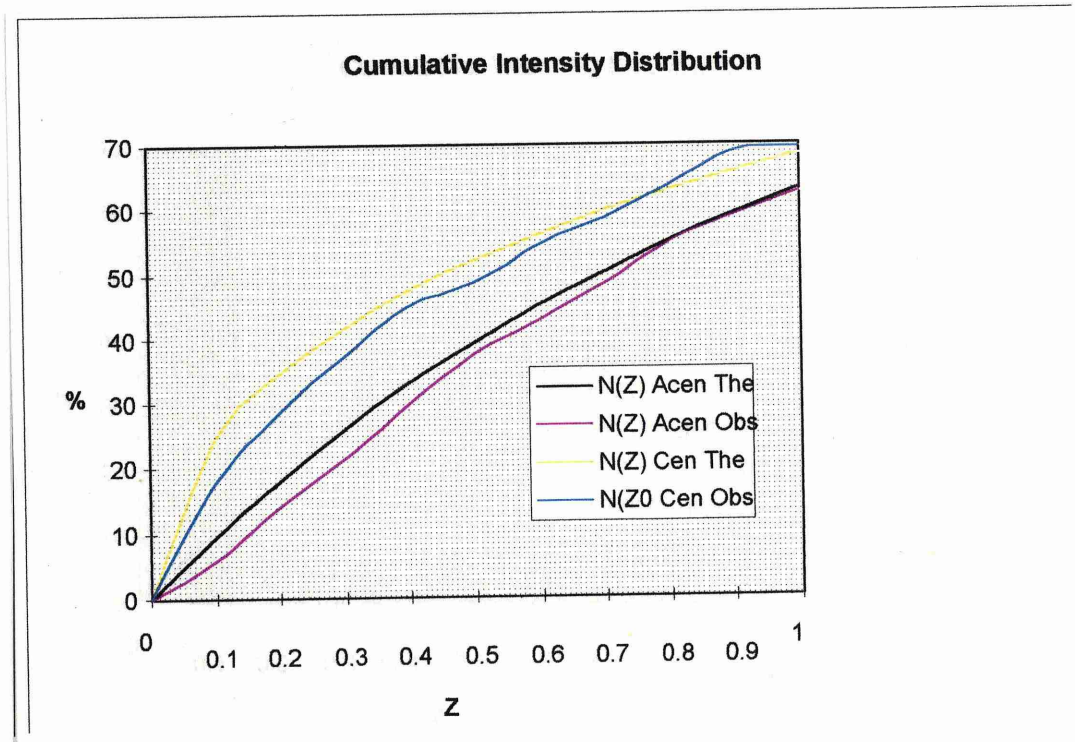



**Figure 5-3 Cumulative Intensity Distribution For Centric And Acentric Reflections For 4 Å Native**

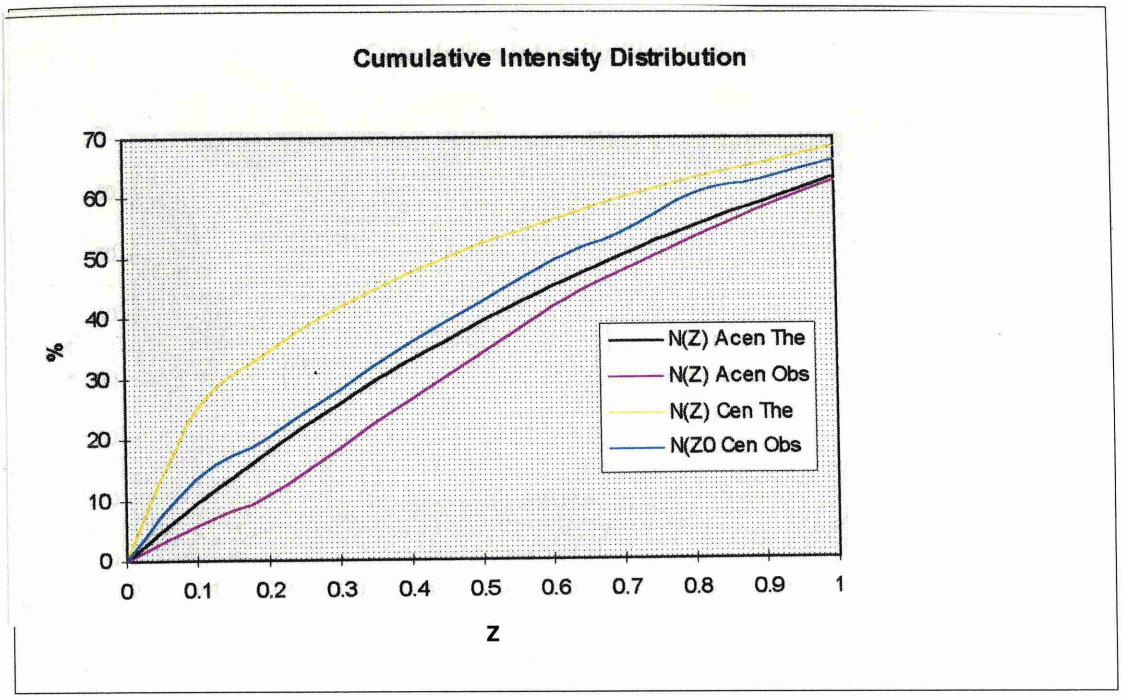



**Figure 5-4 Cumulative Intensity Distribution For Centric And Acentric Reflections For 2.2 Å Native Data Set**

## 5.8 Native Data Collection at S.R.S Daresbury Station 9.5

A high resolution native data set was collected at station 9.5 at the S.R.S Daresbury. This station has a tuneable monochromator and the X-ray beam was tuned to a wavelength of 0.89Å. Data were collected on this station using a 30cm Mar imageplate, an oscillation range of 1°, and using a crystal to detector distance of 310mm which gave no overlapping reflections. Each image was exposed for six minutes. 21 frames were collected giving a total of 21° of data.

These data were processed using DENZO and XDISPLAYF followed by scaling in SCALEPACK. It had an overall completeness of 96.0 % and gave an overall $R_{merge}$ of 4.3 %. Merging statistics are shown in Table 5.4. The Wilson plot for this data set is shown in Fig 5.5 and the cumulative intensity distribution shown in Fig 5.6 shows this crystal was not twinned.

| Dmax (Å) | Dmin (Å) | Nobs | Nuniq | Completeness % | Multi | $R_{merge}$ | $\frac{\bar{I}}{\sigma(\bar{I})}$ |
|---|---|---|---|---|---|---|---|
| 19.46 | 5.90 | 1091 | 380 | 85.3 | 2.9 | 0.029 | 18.3 |
| 5.90 | 4.33. | 1845 | 644 | 91.3 | 2.9 | 0.030 | 18.0 |
| 4.33. | 3.58 | 2343 | 849 | 95.7 | 2.8 | 0.034 | 16.6 |
| 3.58 | 3.12 | 2728 | 1005 | 96.9 | 2.7 | 0.037 | 16.0 |
| 3.12 | 2.81 | 3080 | 1147 | 98.5 | 2.7 | 0.043 | 15.0 |
| 2.81 | 2.57 | 2728 | 1269 | 98.8 | 2.7 | 0.053 | 12.1 |
| 2.57 | 2.38 | 3080 | 1375 | 98.8 | 2.7 | 0.057 | 11.1 |
| 2.38 | 2.23 | 3397 | 1432 | 96.0 | 2.7 | 0.066 | 10.3 |
| 2.23 | 2.11 | 3697 | 1525 | 96.4 | 2.7 | 0.077 | 9.2 |
| 2.11 | 2.00 | 4346 | 1581 | 94.8 | 2.7 | 0.102 | 7.0 |
|  | total | 37418 | 11412 | 96.0 | 2.7 | 0.043 | 11.6 |

**Table 5-4 Merging Statistics for 2.0 Å Daresbury Native Data Set**

Dmin = maximum resolution

Dmax = minimum resolution

Nobs = number of observed reflections

Nuniq = number of unique reflections

Multi = multiplicity of data

## 5.9 The Merging of Native Data

The merging statistics of the 4.0Å native data set collected in-house and the 2.0 Å native data set collected on station 9.5 at Daresbury are shown in Table 5.5. The overall R_merge was 6.2 %. The Wilson plot for this data set is shown in Fig 5.7 and the cumulative intensity distribution is shown in Fig 5.8. This combined data set was used as the native data set.

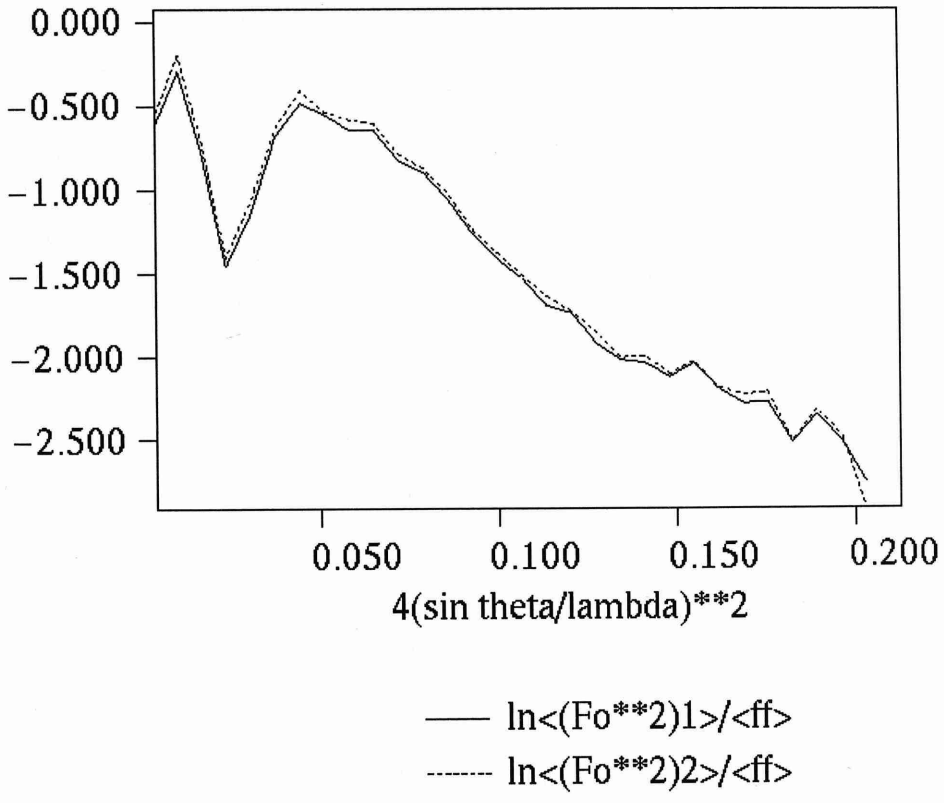



**Figure 5-5 Wilson Plot for 2.0 Å Native Data Set**

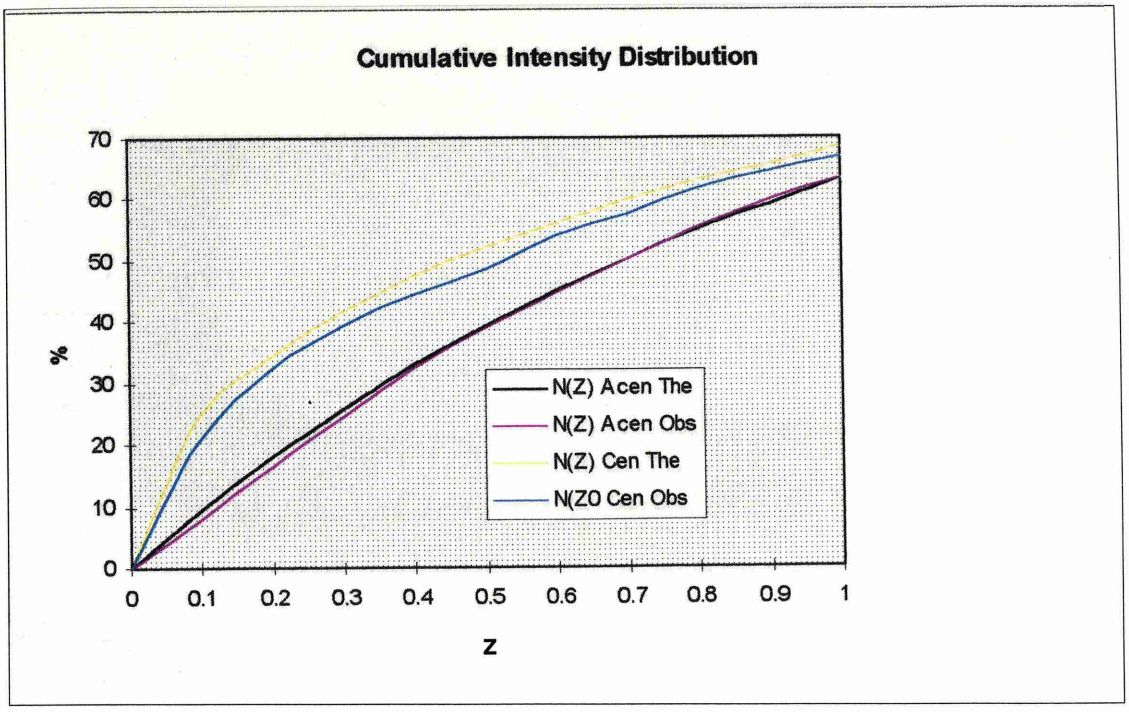



**Figure 5-6 Cumulative Intensity Distribution For Centric And Acentric Reflections**

| Dmin (Å) | Nobs | Nuniq | Completeness % | Multi | $R_{merge}$ | $\frac{\bar{I}}{\sigma(\bar{I})}$ |
|---|---|---|---|---|---|---|
| 6.22 | 3147 | 429 | 100 | 7.3 | 0.061 | 9.9 |
| 4.44. | 4757 | 701 | 99.5 | 6.8 | 0.070 | 9.3 |
| 3.63 | 3938 | 859 | 96,9 | 4.6 | 0.065 | 9.6 |
| 3.15 | 2762 | 1019 | 98.2 | 2.7 | 0.036 | 16.5 |
| 2.82 | 3077 | 1145 | 97.8 | 2.7 | 0.043 | 14.9 |
| 2.58 | 3425 | 1279 | 98.1 | 2.7 | 0.053 | 12.3 |
| 2.39 | 3755 | 1393 | 99.1 | 2.7 | 0.057 | 11.3 |
| 2.23 | 3947 | 1453 | 96.5 | 2.7 | 0.067 | 10.1 |
| 2.11 | 4227 | 1537 | 96.1 | 2.8 | 0.077 | 9.2 |
| 2.00 | 4383 | 1597 | 94.6 | 2.7 | 0.102 | 7.0 |
| total | 37418 | 11412 | 97.4 | 3.3 | 0.062 | 7.44 |

**Table 5-5 Statistics for Merging 4.0 Å Native Data and 2.0 Å Daresbury Native Data Set**

Dmin = maximum resolution

Dmax = minimum resolution

Nobs = number of observed reflections

Nuniq = number of unique reflections

Multi = multiplicity of data

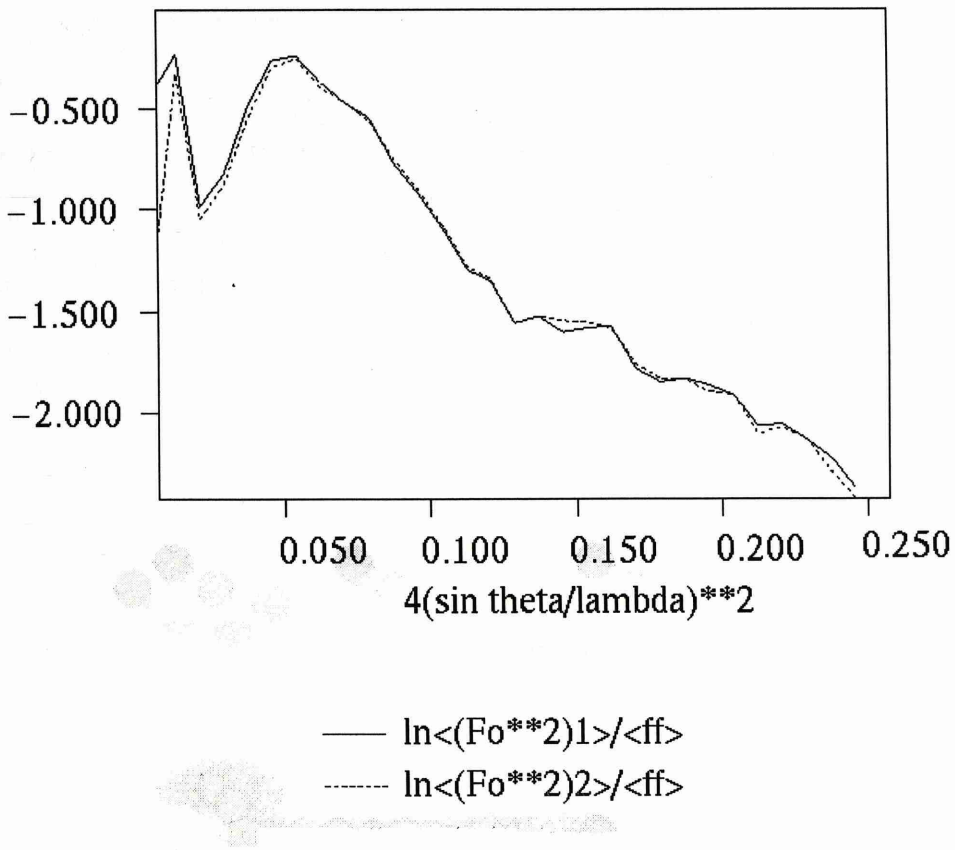



**Figure 5-7 Wilson Plot for Merged Native Data Set**

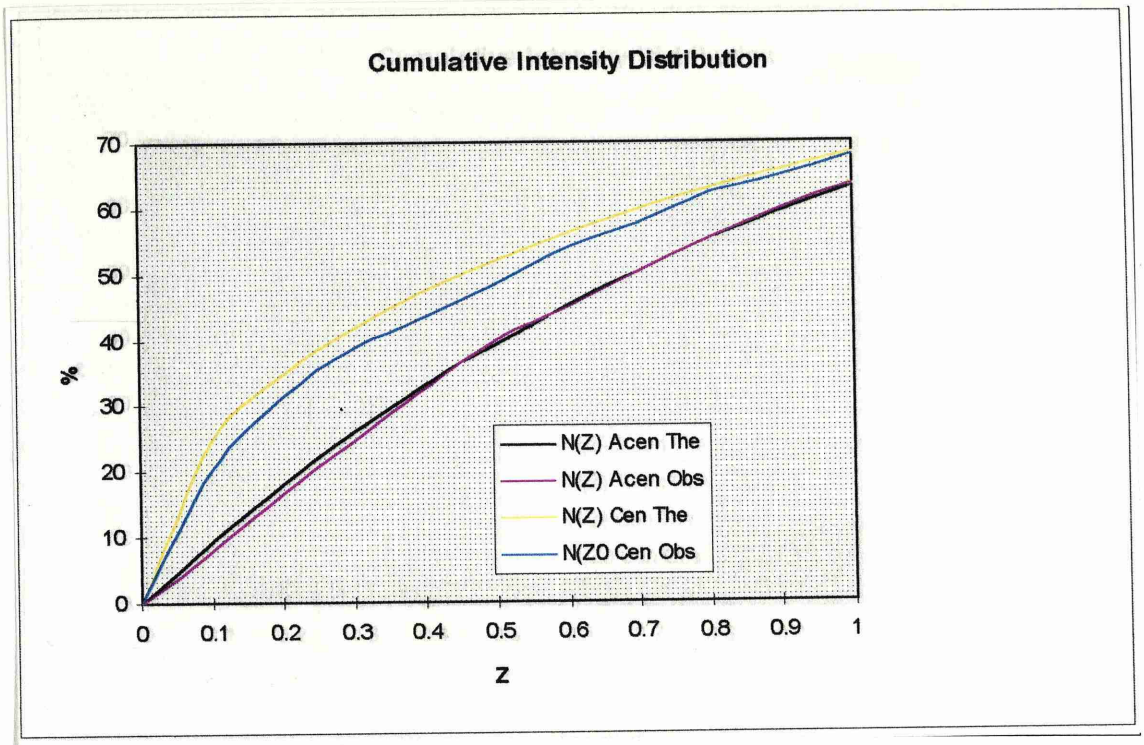



**Figure 5-8 Cumulative Intensity Distribution For Centric And Acentric Reflections For Merged Native Data Set**

## 5.10 Summary

For this project data were collected in Glasgow on a Siemens area detector with a rotating anode source and at the Daresbury Synchrotron Station 9.5 using a Mar image plate. The 2.2 Å data set collected in Glasgow was suspected of being twinned and was not used in the structure determination. A 4.0Å data set collected on the Siemens area detector and the 2.0 Å data set from Daresbury were merged and used as the native for structure determination.

# 6. Phase Determination By Isomorphous Replacement

## 6.1 Introduction

The diffraction pattern seen when a crystal diffracts a collimated beam of X-rays is the Fourier transform of the electron density of the atomic structure in the asymmetric unit of the crystal. By calculating an inverse Fourier transform from the diffraction data it is possible to make a reconstruction of the electron density of the scattering atoms in the unit cell. The electron density ρ at position (xyz) in a unit cell of volume V is related to the structure factor $F_{hkl}$ by the equation:

$$\rho(xyz) = \frac{1}{V} \sum_{h} \sum_{k} \sum_{l} F_{hkl} \exp[-2\pi i(hx + ky + lz)]$$

The structure factor $F_{hkl}$ has an amplitude component $|F_{hkl}|$ and a phase component α which has the relationship:

$F_{hkl} = |F_{hkl}| \exp[i\alpha]$

The two equations can be combined to give the equation:

$$\rho(xyz) = \frac{1}{V} \sum_{h} \sum_{k} \sum_{l} |F_{hkl}| \exp[-2\pi i(hx + ky + lz) + i\alpha(hkl)]$$

This means in order to compute $\rho(xyz)$ both the intensities and phases from reflections need to be known. The phase of $F_{hkl}$ is not directly obtainable from single measurements of reflection intensities and must be obtained by other methods.

There are two main methods of obtaining phases for protein structure determination; molecular replacement, which is used when the protein under study has sequence homology to a known structure and isomorphous replacement, which is the method that has to be used when there is no model structure available. As DHQ

from *M. tuberculosis* has no sequence homology to any known structure it was not possible to use the molecular replacement method to determine phases. Instead for this project the isomorphous replacement method was used to determine the phases.

Robertson first utilised isomorphous replacement for the structure determination of phthalocyanines in 1937 (Robertson & Woodward, 1937) and the technique was first applied to protein structure determination for haemoglobin (Green *et al.*, 1954). Isomorphous replacement is done for small molecules by replacing an atom with a similar heavier one. This is difficult to achieve specifically for a protein unless there is a metal atom like zinc or magnesium that can be replaced by a heavier metal or there is a methionine which can be replaced with seleno-methionine by using recombinant DNA technology. It is more common to introduce heavy atoms into protein crystals by attempting to bind them to sites on the protein surface. This is possible because protein crystals contain a large percentage of solvent and have large channels containing mother liquor and the heavy atoms can be diffused into these channels. This gives suitable derivatives if a high percentage of the heavy atoms bind at specific sites and if that binding does not cause disturbance of the molecular or crystal structure which would cause non-isomorphism.

## 6.2 Phase Determination By Isomorphous Replacement

$\overline{F}_P$ the structure factor for the native protein is a vector and can be described in terms of a structure magnitude $|F_p|$ and phase $\alpha_p$. The structure factor $\overline{F}_{PH}$ of the heavy atom derivative is also a vector with magnitude $|F_{PH}|$ and phase $\alpha_{PH}$. For isomorphous structures $\overline{F}_{PH}$ can be derived from $\overline{F}_P$ by the vector addition of $\overline{F}_H$,

which is the contribution of the heavy atoms to the structure factors of the derivatives. The relationship is ;

$$\overline{F}_{PH} = \overline{F}_P + \overline{F}_H \text{ (Fig 6.1)}$$

The magnitudes of structure factors $|F_{PH}|$ and $|F_P|$ can be measured experimentally. The positions of the heavy atoms in the cell can be found either using the Patterson method, by direct methods, or from difference Fouriers (if some phase information is already known). Using the cosine law the following equation gives the phase of the protein $\alpha_p$

$$\alpha_p = \alpha_H + \cos^{-1}\left(\frac{\left(F_{PH}^{\ 2} - F_P^{\ 2} - F_H^{\ 2}\right)}{2F_P F_H}\right) = \alpha_H \mp \alpha'$$

This equation shows that there are two possible values for $\alpha_p$. A Harker construction in Fig. 6.2 illustrates the ambiguity. This ambiguity in phase determination can be overcome by either using a second derivative or by using anomalous scattering if there is a large enough anomalous scattering from the heavy atom to produce significant differences in the reflections forming Freidel pairs. The Harker construction shown in Fig. 6.3 illustrates the way a second isomorphous derivative can overcome this phase ambiguity. It can be seen from this construction that if $F_{H1}$ and $F_{H2}$ are co-linear then the derivative circles would superimpose and there would be two equally possible values of the phase. This is why the second derivative has to have a different site of occupancy from the first.

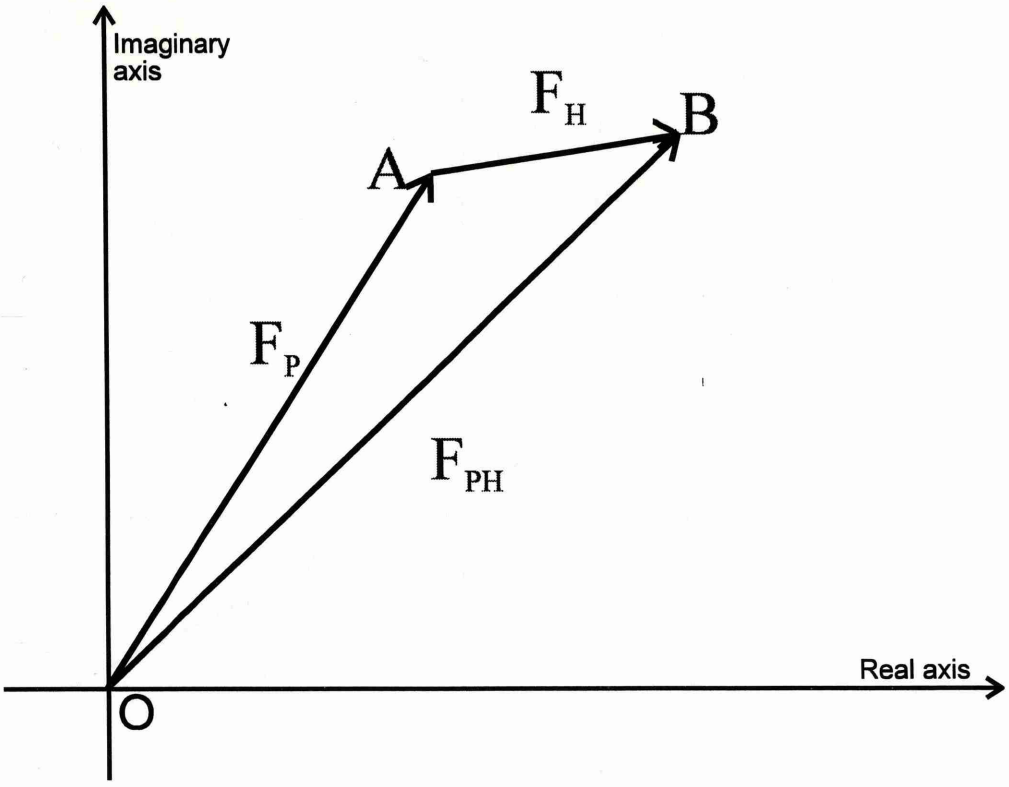



**Figure 6-1 The Argand Diagram**

This illustrates the native protein ($F_p$) and heavy atom ($F_H$) contributions to the structure factor ($F_{PH}$) for the heavy atom derivative of the protein.

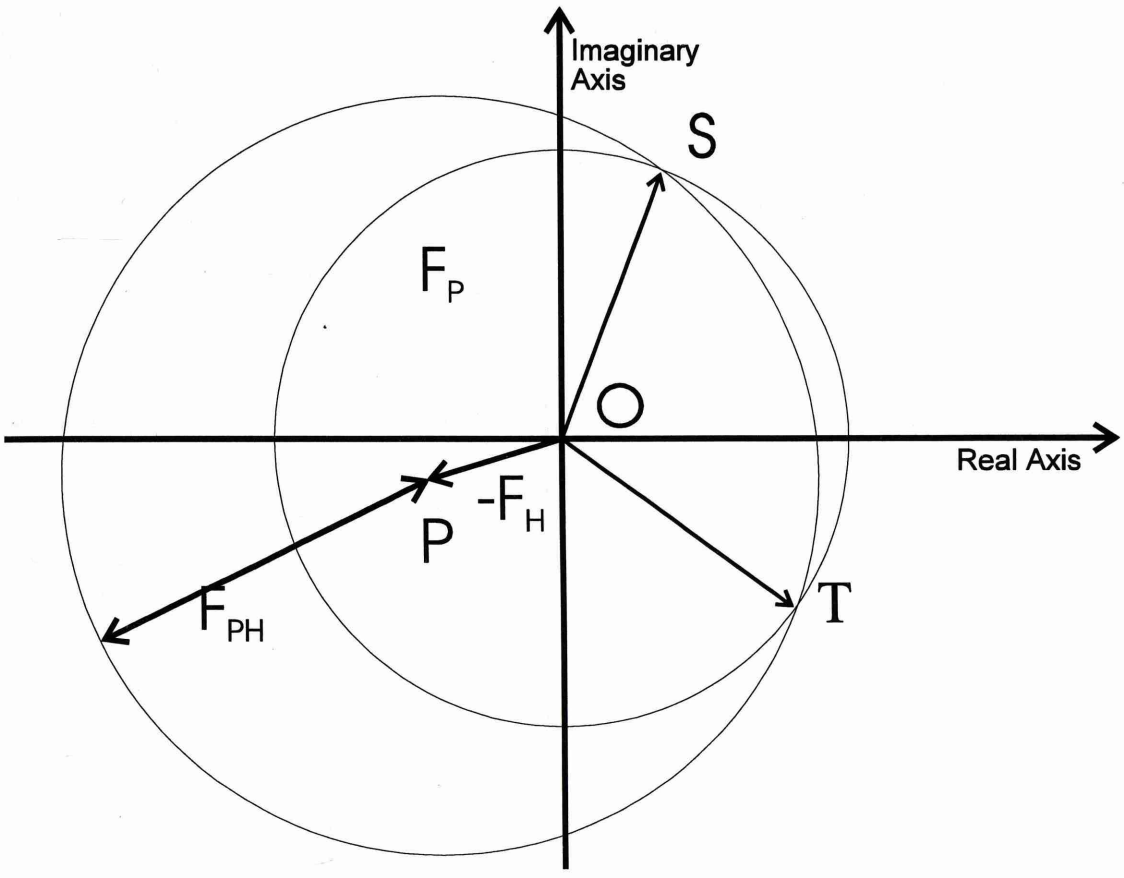



**Figure 6-2 Harker Construction for S.I.R. Method**

The vector $-\overline{F}_H$ is drawn from the centre O. A circle of radius $F_{PH}$ is drawn centred on the end of the vetor at P. A second circle of radius $F_P$ is drawn with its centre at O. There are two general points of intersection of the circle, S and T. The vectors OT and OS represent the possibilities for $\overline{F}_P$.

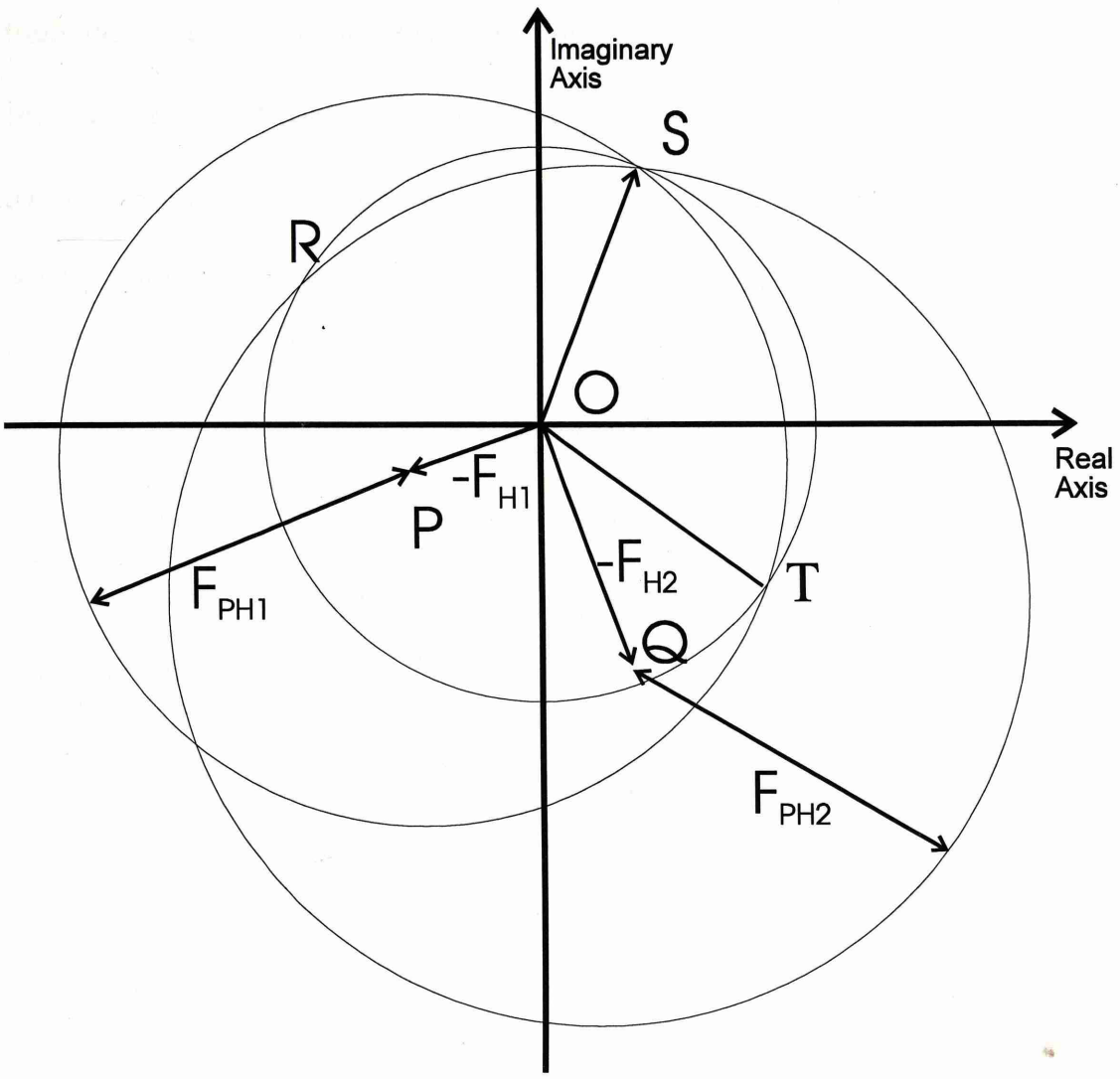



**Figure 6-3 The Harker Construction For The MIR Method**

The vector for the heavy atom contribution from the first derivative, $-\overline{F}_{H1}$ is drawn from the centre O. A circle of radius $F_{PH1}$ is drawn centred on the end of the vetor at P. A second circle of radius $F_P$ is drawn with its centre at O. There are two general points of intersection of the circle, S and T. The vector $-\overline{F}_{H2}$ for the contribution of the second heavy atom derivative is drawn as $-F_{H2}$ from the centre O. A circle of radius $F_{PH2}$ is drawn centred on the end of the vector at Q. As S is the point of intersection of all three circles the vector OS must be the correct choice for $\overline{F}_P$.

In practice the phase triangles determined by the phasing process are distorted from their true and unknown shape because of errors in measurement and scaling of $|F_{PH}|$ and $|F_H|$ and because non-isomorphism causes errors in the estimate of $|F_H|$. It has been shown that an approximation for this lack of closure can be made by assuming the error is due only to $F_{PH}$ as illustrated in Fig 6.4 and this then gives the equation

$$\varepsilon = \left(F_{PH}(obs) - F_{PH}(calc)\right)$$

If a Gaussian distribution of errors is assumed, the probability that the phase angle $\alpha$ for the structure factor is correct is related to this lack of closure. For the $j^{th}$ heavy atom the relationship is;

$$P_j(\alpha) = N \exp\left(\frac{-\varepsilon_j(\alpha)^2}{2E_j^2}\right)$$

Where $N$ is the normalisation factor, $\varepsilon(\alpha)$ is the lack of closure of the phase triangle along $F_{PH}$ at the protein phase angle $\alpha$, and $E_j$ is the standard deviation of the distribution.

When several derivatives are used in phasing, there are two possible phase angles derived from $P(\alpha)$, the most probable phase which corresponds to the maximum value for $P(\alpha)$ and the best phase, which is given by the centroid of the probability distribution (Blow & Crick 1959). Using this best phase gives an electron density map with the least error and this is shown by the equation for the best structure factor:

$$F_{best} = m|F_p|e^{i\alpha_{best}}$$

Where $m$ is a figure of merit dependent on the sharpness of the phase probability and is equivalent to a weighting function and is defined by:

$$m = \frac{\int_{\alpha=0}^{2\pi} e^{i\alpha} P(\alpha) d\alpha}{\int_{\alpha=0}^{2\pi} (\alpha) d\alpha}$$

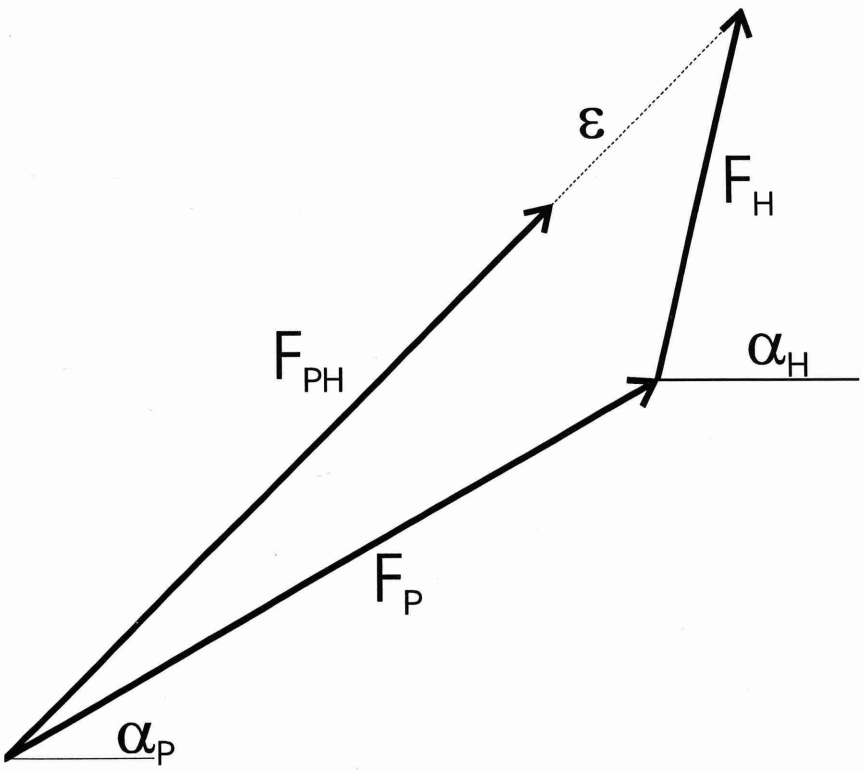



**Figure 6-4 Errors Causing Lack of Closure on Phase Triangles**

Where $m$ is a figure of merit dependent on the sharpness of the phase probability and is equivalent to a weighting function and is defined by:

$$m = \frac{\int_{\alpha=0}^{2\pi} e^{i\alpha} P(\alpha) d\alpha}{\int_{\alpha=0}^{2\pi} (\alpha) d\alpha}$$

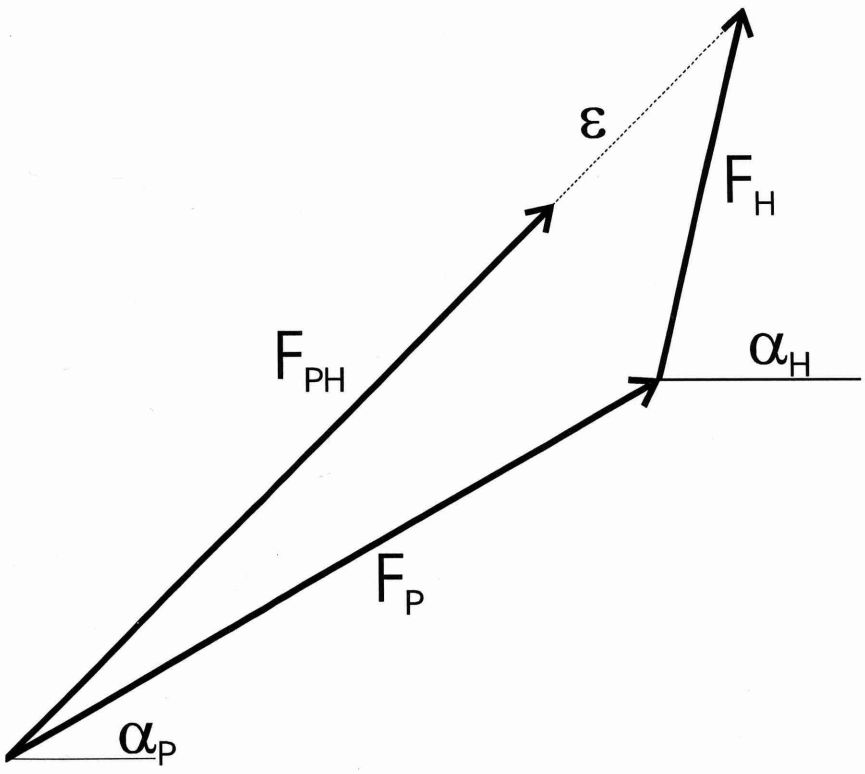



**Figure 6-4 Errors Causing Lack of Closure on Phase Triangles**

## 6.3 Isomorphous Replacement Procedure

Isomorphous replacement has five stages:

1. The preparation of derivatives
2. The collection and processing of native and derivative data
3. The scaling together of native and derivative data
4. The determination of heavy atom positions
5. The refinement of heavy atom positions and the calculation of phases.

### 6.3.1 Preparation of Derivatives

Derivatives can be prepared by either soaking the crystals in solutions containing heavy atom salts or by co-crystallising the protein in the presence of heavy atom containing salts. Lack of isomorphism is more likely to occur with co-crystallisation as there is the possibility of heavy atoms binding where the protein makes crystal contacts. Soaking for heavy atoms derivatives is still very much a trial and error process. The primary sequence of a protein can give clues as to what heavy atom salts to use in a trial as certain reagents bind to particular residues, but the design of a completely rational heavy atom soak experiment is not straightforward as some residues might be buried in the protein core or be involved in crystal contact and thus be unavailable to bind heavy atoms to give suitable derivatives.

The pH at which a soaking experiment is done is also important in deciding on which heavy atom reagents should be tried. The binding properties of a residue for heavy atoms are dependent on pH. Histidine is better at binding ligands at higher pH values where it is neutral and cysteine is more reactive about pH 8.3 when it is ionised. Some heavy atom salts are easily hydrolysed and cannot be used in the

alkaline pH region. Examples of these are $UO_2^{2+}$ and $Sm^{3+}$ salts. As the crystals of DHQ were grown at pH 7.15 the heavy atom reagents that were chosen for the initial trials were ones that would work best at this pH.

If the mother liquor of the crystal contains high concentrations of salts or phosphate buffers, these can complex or precipitate heavy atoms and stop them binding to the protein. This was not considered a problem with this project as the crystals had been grown using PEG 6000 as the precipitant in Tris/HCl buffer and Tris and PEG are fairly inert to heavy atoms. Problems are caused by some heavy atoms like silver (as silver nitrate) reacting with the chloride ions in the buffer and giving a precipitate, so with silver it would have been necessary to use Tris acetate not Tris chloride.

### 6.3.2 Data Collection and Processing

The method of collection and processing of data from derivatised crystals was the same as that described for the native data set in Chapter 5.

### 6.3.3 Scaling of Native Data

Once the native and derivative data sets have been collected, before heavy atom positions can be located the derivative data has to be scaled to the native data. Scaling of derivative data to the native data was done using the CCP4 program SCALEIT. This calculates a derivative to native scaling function and applies it to the derivative data. This program determines scales from the squared amplitudes. Isotropic scaling was used and had a scaling function of the form:

$$K = C \exp(-B\sin\theta/\lambda)$$

A qualitative test for the degree of heavy atom substitution is the calculation of an agreement index $R_{iso}$ between the native ($F_P$ ) and derivative ($F_{PH}$) structure factors;

$$R_{iso} = \frac{\sum |F_{PH} - F_P|}{\sum |F_P|}$$

Experience indicates that a value from 0.15 to 0.25 can be considered as a reasonable agreement for a good derivative. If the value is lower than this it can be indicative of low or no substitution and values higher than this indicate that non isomorphism may be a problem. In the ideal case with a perfectly isomorphous derivative and no experimental errors, when $R_{iso}$ is evaluated for equivalent resolution bins, it should not increase with resolution. In practice $R_{iso}$ will increase with resolution as a result of experimental errors and non-isomorphism.

### 6.3.4 Determination Of Heavy Atom Positions

#### *6.3.4.1 The Patterson Method*

A brief description of the Patterson method for location of heavy atom will be given in this section, a more detailed discription can be found in Blundell and Johnson (1976). An isomorphous difference Patterson can be obtained using the coefficients $(F_{PH}\text{-}F_P)^2$ (Perutz, 1956; Blow,1958) The expanded expression for these coefficients is

$$\left(F_{PH} - F\right)^2 = 4F_P^2 \sin^4\left(\frac{\alpha_P - \alpha_{PH}}{2}\right) \qquad (1)$$

$$+ 4F_H^2 \cos^2\left(\alpha_{PH} - \alpha_H\right) \qquad (2)$$

$$- 4F_P F_H \sin^2\left(\frac{\alpha_P - \alpha_{PH}}{2}\right)\cos\left(\alpha_{PH} - \alpha_H\right) \qquad (3)$$

If values of $F_H$ are small then the angles ($\alpha_P$-$\alpha_{PH}$) will tend to be small and the terms in (1) which give protein- protein interactions will be of low weight. However if $F_H << F_P$ then ($\alpha_P$-$\alpha_{PH}$) becomes random resulting in noise and the term (2) will give heavy atom vectors with half the expected peak size.

As $F_H$ becomes larger the term (1) for protein-protein interactions increases but the contribution of the heavy atom vector map will increase as ($\alpha_P$-$\alpha_{PH}$) also will tend to decrease.

Noise in a difference Patterson can be caused by the magnitude of difference between the native and derivative amplitudes being not exactly equal to the heavy atom structure factors. This may be due to errors in measurement and scaling as well as to a lack of isomorphism. When there are several sites of low occupancy or the protein size is large the noise peaks due to terms (1) and (3) can be larger than the heavy atom vector peaks and this results in difficulties in deconvoluting the Patterson. The proper difference Patterson would have coefficients $\left(F_{PH}^2 - F_P^2\right)$. This would give a map with weak peaks corresponding not only to heavy atom-heavy atom vectors but also to heavy atom-protein vectors which would make the maps very difficult to interpret. Also as the difference Patterson depends on the difference of measurement rather than the difference of the square roots it is more sensitive to scaling and measurement errors.

If anomalous differences have been accurately measured an anomalous difference Patterson with coefficients of $(F_{PH}(+) - F_{PH}(-))^2$ can be calculated. These coefficients are approximately equivalent to the coefficients of $F_H^2 \sin^2(\alpha_{PH}-\alpha_H)$ The resulting maps contain peaks of half weight because of the $\sin^2$ term multiplying

the $F_{H.}^2$ and can be noisy depending on the accuracy of measurement of the anomalous differences.

#### *6.3.4.2 Pattersons in Spacegroup F23*

The space group F23 has 48 symmetry operators and because of this Pattersons can be noisy and difficult to interpret. The Harker planes in F23 are (u,v,0), (u,v,1/2), (u,v,1) (u,0,w), (u,1/2,w), (u,1,w), (0,v,w), (1/2,v,w) and (1,v,w). Patterson maps were calculated using the CCP4 program FFT. The Patterson space group for F23 is Fm3. In this project attempts were made to solve Pattersons both by hand and using the SHELXS-90 (Sheldrick 1991) which uses a Patterson vector superposition minium function.

#### *6.3.4.3 Difference Fouriers*

After the major site of a heavy atom derivative has been found difference Fouriers using the SIR phases can be used to find other sites or positions of heavy atoms in other derivatives. These difference Fouriers are calculated with coefficients:

$$m(F_{PH} - F_P)\exp(i\alpha_P)$$

These maps may show confusing ghost peaks corresponding to the site from which the SIR phases have been generated. This is because the Fourier synthesis is dominated by phases more than amplitudes.

### 6.3.5 Refinement Of Phases

When the heavy atom positions have been determined, the heavy atom parameters are optimised using maximum likelihood refinement. In this method the phase probability is calculated by a simple likelihood function defined by Blow and

Crick (1959). The assumption made in this function is that the errors in measurements of native amplitudes are smaller than the combination of errors in lack of isomorphism and in the measurement of heavy atom amplitudes. The probability of the phase being correct is then used to weight the lack of closure estimate.

## 6.4 Experimental

During this screening for heavy atom derivatives numerous crystals were soaked in many different heavy atom solutions and then if the crystals still diffracted data was collected from them. The crystals were very susceptible to some heavy atom salts, and even soaking in sub-millimolar solutions for seconds caused complete loss of diffraction. This was the case for mercury acetate and gold chloride.

The crystals were however unreactive with respect to some other heavy atoms salts like potassium gold cyanide and trimethyl plumbyl acetate, where no loss of diffraction was observed even after soaking in 20 mM solutions for several days. This was observed when there was no specfic interaction between the heavy atom and the protein

Derivatives were eventually found using two reagents $Hg(CN)_2$ and $K_2Pt(Cl)_4$.

## 6.5 The $Hg(CN)_2$ Derivative.

$Hg(CN)_2$ was very reactive, as shown by the fact that soaking the crystal in 1mM Hg(CN) for 1 minute with a five minute back soak gave the best phasing. If the salt concentration or the time of soaking were increased, the crystal did not diffract. Even under these soaking conditions some reduction in the quality of diffraction was observed and for the best crystals the data set had a $D_{min}$ of 4.0 Å. This data set was collected in

house on the Siemens area detector and its completeness and merging statistics are given in Table 6.1. This data set scales with the native to give an $R_{iso}$ of 15.8 %

| Dmax (Å) | Dmin (Å) | Nobs | Nuniq | Completeness | Multi | $R_{merge}$ | $\frac{\bar{I}}{\sigma(\bar{I})}$ |
|---|---|---|---|---|---|---|---|
| 31.94 | 11.84 | 294 | 68 | 100 | 4.3 | 0.045 | 13.9 |
| 11.84 | 8.68 | 470 | 96 | 99.5 | 4.9 | 0.049 | 13.7 |
| 8.68 | 7.17 | 540 | 117 | 100 | 4.6 | 0.050 | 12.8 |
| 7.17 | 6.25 | 534 | 132 | 95.1 | 4.0 | 0.071 | 10.2 |
| 6.25 | 5.61 | 564 | 155 | 100 | 3.6 | 0.068 | 10.4 |
| 5.61 | 5.14 | 555 | 169 | 100 | 3.3 | 0.068 | 10.6 |
| 5.14 | 4.76 | 585 | 174 | 96.7 | 3.4 | 0.058 | 12.2 |
| 4.76 | 4.46 | 627 | 197 | 100 | 3.2 | 0.063 | 10.9 |
| 4.46 | 4.21 | 664 | 205 | 100 | 3.2 | 0.070 | 10.0 |
| 4.21 | 3.98 | 673 | 214 | 100 | 3.1 | 0.083 | 8.7 |
| | total | 5505 | 1539 | 99.7 | 3.6 | 0.062 | 10.9 |

**Table 6-6 Merging Statistics For The 4 Å Hg(CN)2 Data Set**

Dmin = maximum resolution

Dmax = minimum resolution

Nobs = number of observed reflections

Nuniq = number of unique reflections

Multi = multiplicity of data

### 6.5.1 Solution Of Patterson For $Hg(CN)_2$ Derivative

The initial Patterson was difficult to interpret manually due to noise in the map, a big origin ripple and the high symmetry of the space group. However SHELXS-90 gave a clear solution for a single site at 0.144, 0.065, 0.716

### 6.5.2 Calculation of SIR Phases

MLPHARE was used to refine the position and occupancy of this Hg site and then to calulate phases calculated from it. The occupancy of the site was set to 0.1 and refinement was done for initially for 10 cycles of refinement using only the centric reflections followed by 10 cycles using all reflections. This resulted in an overall figures of merit of 0.32. The final occupancy of this site was 0.265. These computed phases were used to calculate a difference Fourier for all the other data sets from soaked crystals.

## 6.6 The $K_2Pt(Cl)_4$ Derivative

The site of the platinum binding in a crystal soaked in 1mM $K_2Pt(Cl)_4$ was found using a difference Fourier. Data were collected from this crystal at S.R.S Daresbury station 7.2 using a 18 cm Mar image plate. This crystal diffracted to 4 Å and data was collected at a crystal to detector distance of 115mm and using an oscillation range of $2^o$. Unfortunately radiation damage to the crystal was great and after 12 frames had been collected the data collection was stopped. The completeness and merging statistics are shown in Table 6.2. This data was scaled against the native and gave an $R_{iso}$ of 23.9 %. This derivative was found to have a phasing power of 0.6 which was considered adequate for a second derivative and was used for the calculation of MIR phases.

| Dmax (Å) | Dmin (Å) | Nobs | Nuniq | Completeness % | Multi | $R_{merge}$ | $\frac{\bar{I}}{\sigma(\bar{I})}$ |
|---|---|---|---|---|---|---|---|
| 29.49 | 11.54 | 350 | 133 | 100 | 2.6 | 4.6 | 13.5 |
| 11.54 | 8.56 | 343 | 131 | 100 | 2.6 | 6.4 | 12.6 |
| 8.56 | 6.82 | 356 | 135 | 100 | 2.6 | 11.2 | 10.4 |
| 6.82 | 5.96 | 376 | 144 | 100 | 2.6 | 12.2 | 6.5 |
| 5.96 | 5.42 | 325 | 134 | 100 | 2.4 | 11.9 | 6.4 |
| 5.42 | 5.04 | 329 | 137 | 100 | 2.4 | 10.9 | 6.2 |
| 5.04 | 4.74 | 346 | 143 | 100 | 2.4 | 14.5 | 5.9 |
| 4.74 | 4.50 | 335 | 138 | 100 | 2.4 | 15.3 | 4.9 |
| 4.50 | 4.31 | 350 | 144 | 98 | 2.4 | 16.2 | 4.7 |
| 4.31 | 4.20 | 358 | 148 | 96 | 2.4 | 23.5 | 4.5 |
| | total | 3449 | 1398 | 92.2 | 2.5 | 12.5 | 5.1 |

**Table 6-7 Merging Statistics for K2PtCl4 Derivative Collected at Daresbury**

Dmin = maximum resolution

Dmax = minimum resolution

Nobs = number of observed reflections

Nuniq = number of unique reflections

Multi = multiplicity of data

## 6.7 Collection Of A High Resolution $Hg(CN)_2$ Data Set

As maps calculated using the two derivatives mentioned previously in this section were difficult to interpret it was decided to collect another data set from a crystal soaked in $Hg(CN)_2$, but with a reduced concentration and time in order to eliminate the reduction in the quality of diffraction experienced after soaking. A crystal soaked in 0.1mM $Hg(CN)_2$ for 1 minute diffracted beyond 2.2 Å. A data set was collected at S.R.S Daresbury on station 9.5 where the wavelength was tuned to 0.99 Å to try and maximise the anomalous signal. The completeness and merging statistics for this derivative are given in table 6.3. When scaled against the native gave an $R_{iso}$ of 10.6 %. This derivative had a phasing power of 0.8 and an occupancy of 0.143.

| Dmax (Å) | Dmin (Å) | Nobs | Nuniq | Completeness % | Multi | $R_{merge}$ | $\frac{\bar{I}}{\sigma(\bar{I})}$ |
|---|---|---|---|---|---|---|---|
| 19.49 | 4.62 | 3985 | 972 | 95.9 | 4.1 | 3.6 | 19.6 |
| 4.62 | 3.67 | 4062 | 967 | 99.5 | 4.2 | 4.3 | 18.8 |
| 3.67 | 3.21 | 3844 | 961 | 99.8 | 4.0 | 5.3 | 16.4 |
| 3.21 | 2.91 | 3707 | 950 | 99.9 | 3.9 | 7.3 | 14.0 |
| 2.91 | 2.70 | 3771 | 967 | 100 | 3.9 | 9.7 | 9.6 |
| 2.70 | 2.54 | 3628 | 954 | 99.9 | 3.8 | 11.2 | 7.6 |
| 2.54 | 2.42 | 3719 | 944 | 100 | 3.9 | 12.7 | 6.6 |
| 2.42 | 2.31 | 3709 | 966 | 100 | 3.8 | 15.1 | 5.4 |
| 2.31 | 2.22 | 3527 | 943 | 100 | 3.7 | 16.7 | 4.9 |
| 2.22 | 2.15 | 2675 | 836 | 89.7 | 3.2 | 28.6 | 3.6 |
| | Total | 36624 | 9461 | 98.9 | 3.8 | 6.6 | 11.4 |

**Table 6-8 Merging Statistics for High Resolution Hg(CN)2 Derivative Collected at Daresbury**

Dmin = maximum resolution

Dmax = minimum resolution

Nobs = number of observed reflections

Nuniq = number of unique reflections

Multi = multiplicity of data

## 6.8 Calculation of MIR phases

MLPHARE was used to refine the heavy atom parameters of the $Hg(CN)_2$ and $PtCl_4$ derivatives. The refinement of heavy atoms and calculation of phases was done using both $Hg(CN)_2$ data sets. Even though this risks biasing the Hg occupancies or error estimates, the maps using phases calculated in this way were better than those computed with phases calculated using one of either of the two $Hg(CN)_2$ data sets. The results of heavy atom refinement are given in Table 6.3. The anomalous signal from the high resolution $Hg(CN)_2$ data was used also in the heavy atom refinement and phase calculation. It gave an anomalous occupancy to 4Å of 0.105 and a graph of the Cullis_R_anomalous against resolution is shown in fig 6.8. The overall Figures of Merit were 0.40 for acentric reflections and 0.60 for centric reflections. The analysis of Figure of Merit against resolution is given in fig 6.9.

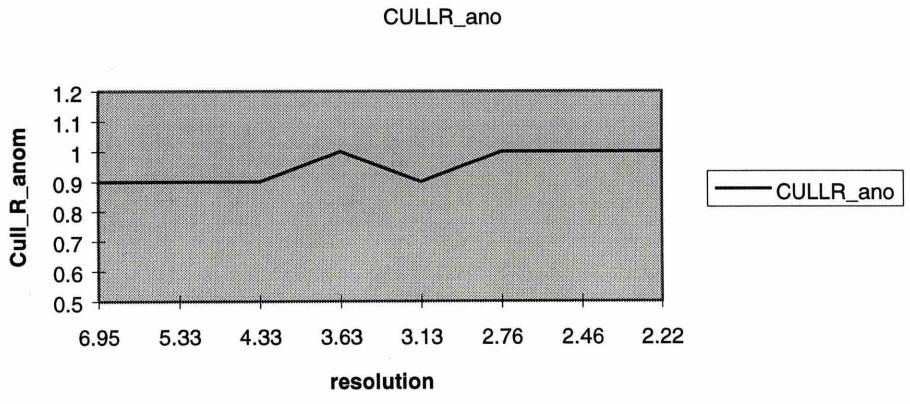



**Figure 6-5 Cullis R anomalous versus resolution**

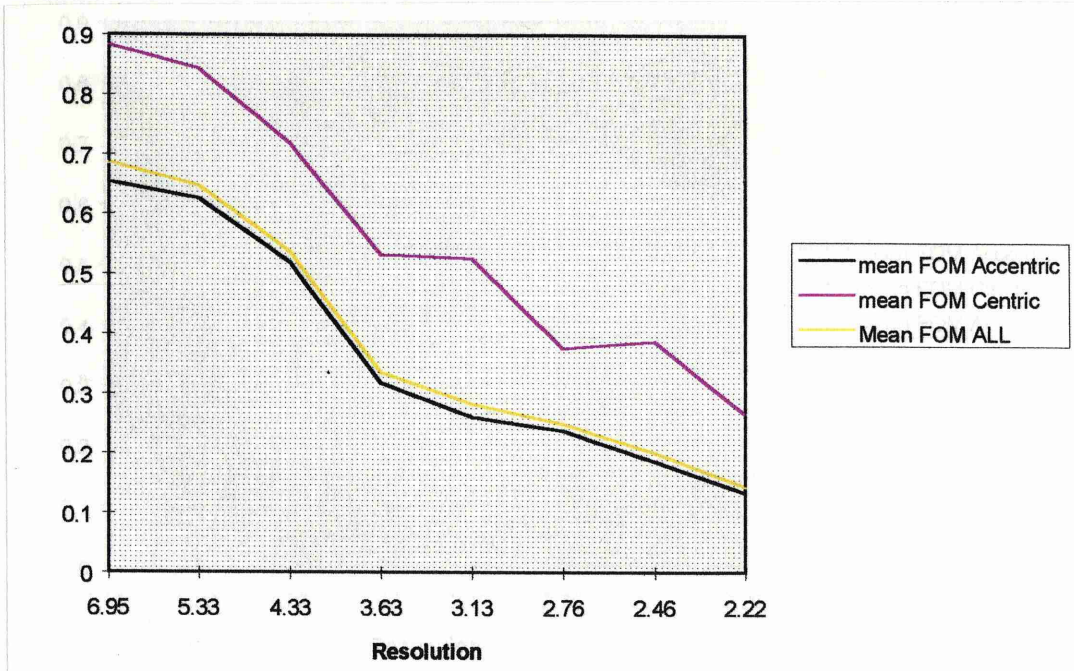



**Figure 6-6 MIR Figures of Merit**

| | No of sites | $D_{min}$ (Å) | Co-ordinate | Isomorphous occupancy | $R_{cullis}$ | Phasing Power |
|---|---|---|---|---|---|---|
| mercury | 1 | 3.96 | (0.114,0.065,0.717) | 0.266 | 0.75 | 1.6 |
| mercury | 1 | 2.2 | (0.131,0.053,0.741) | 0.143 | 0.84 | 0.8 |
| platinum | 1 | 4.38 | (0.080,0.169,0.218) | 0.218 | 0.95 | 0.7 |

**Table 6-9 Results of Heavy Atom Refinement**

$$R_{cullis} = \frac{\sum |F_{PH} \pm F_P - F_H|}{\sum |F_{PH} - F_P|}$$

Phasing Power is the root mean square of $\frac{|F_H|}{\sum |F_{PH} \pm F_P - F_H|}$

## 6.9 Summary

This chapter describes how a good derivative was found by soaking a crystal in 1mM $Hg(CN)_2$ for one minute followed by back soaking for five minutes. The phases from this good derivative enabled the site of the $K_2PtCl_4$ derivative to be

found. This derivative was poorer in phasing power. A high resolution data set was collected from a crystal soaked in 0.1mM $Hg(CN)_2$ for 1 minute. All three data sets were used for determining phases. The next chapter describes the use of density modification techniques to improve and extend these phases and also the calculation and interpretation of electron density maps

# 7. Interpretation of Electron Density Map for *M. tuberculosis* DHQ

## 7.1 Making Electron Density Maps

Electron density maps were calculated using the program FFT and converted from CCP4 format to O format using the program MAPPAGE. Maps were displayed using the program O (Jones *et al.,* 1991) on an Evans and Sutherland ESV10 graphics machine.

## 7.2 MIR Map

An MIR map was calculated using the phases calculated by MLPHARE as described in section 6.7. This map was calculated to a resolution of 2.6 Å and as can be seen in Fig 7.1, the map is of reasonable quality. The protein solvent boundary and some secondary structural elements can clearly be seen. In order to improve the quality of this map density modification was performed on the phases using the CCP4 program DM.

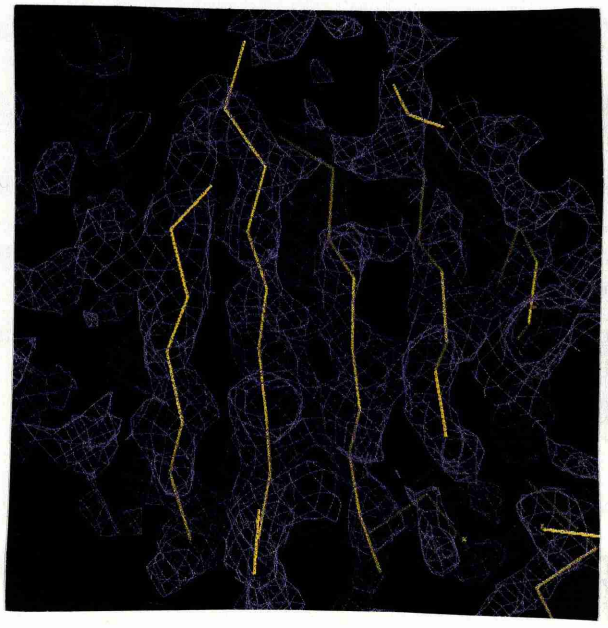


**Figure 7-1 MIR map**

## 7.3 Density Modification

The program DM was used to improve and extend the phases. DM was used in this work to apply the techniques of solvent flattening and histogram matching. Solvent flattening assumes that any density in the solvent region of a protein arises from noise fluctuations and that solvent density should be flat. The technique involves calculating a mask that defines the boundary between solvent and protein (Wang, B. C 1985). This mask is then used in a procedure that involves density in the solvent region being flattened to a constant low value. The flattened map is then inverted through a reverse Fourier transform to produce new structure factors that are then combined with the original MIR phases.

The density histogram of an electron density map is the probability distribution of the electron component of the protein structure. Histogram matching exploits the fact that proteins containing similar amounts of solvent have almost identical density histograms at the same resolution even if they do not have the same fold (Zhang & Main, 1990). The method involves binning the experimental electron density values according to size. The densities from each bin are then scaled so that the final distribution is the same as that for a well-refined standard protein.

The use of DM to perform solvent flattening and histogram matching options made the electron density maps more interpretable. A comparison of the electron density of the MIR map and the map resulting from DM is shown in Fig. 7.2a & b.

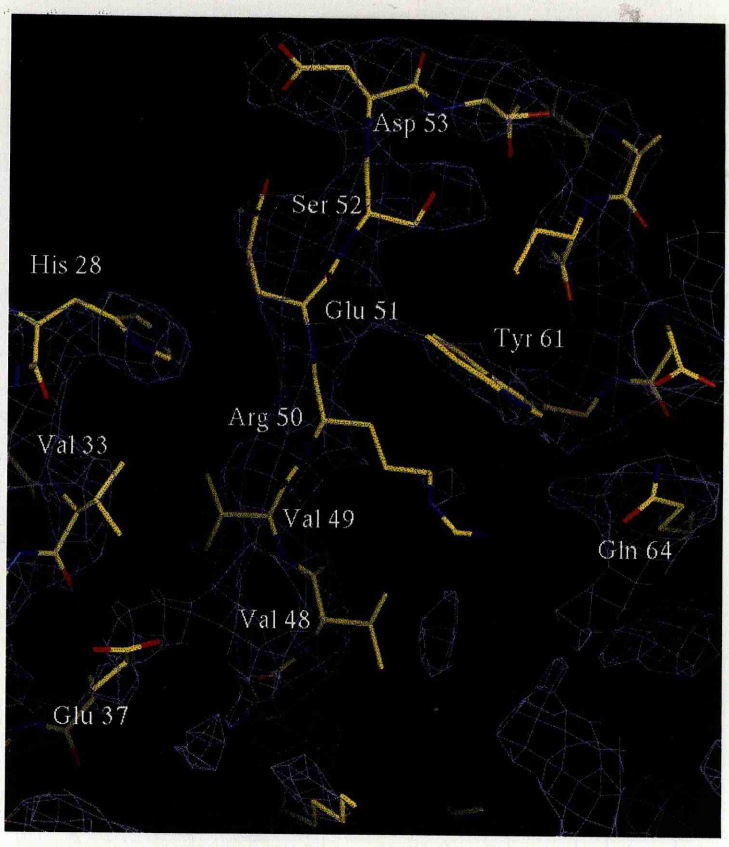



**Figure 7-2a Comparison of DM with MIR map**

This diagram shows a portion of the MIR map and the refined structure.

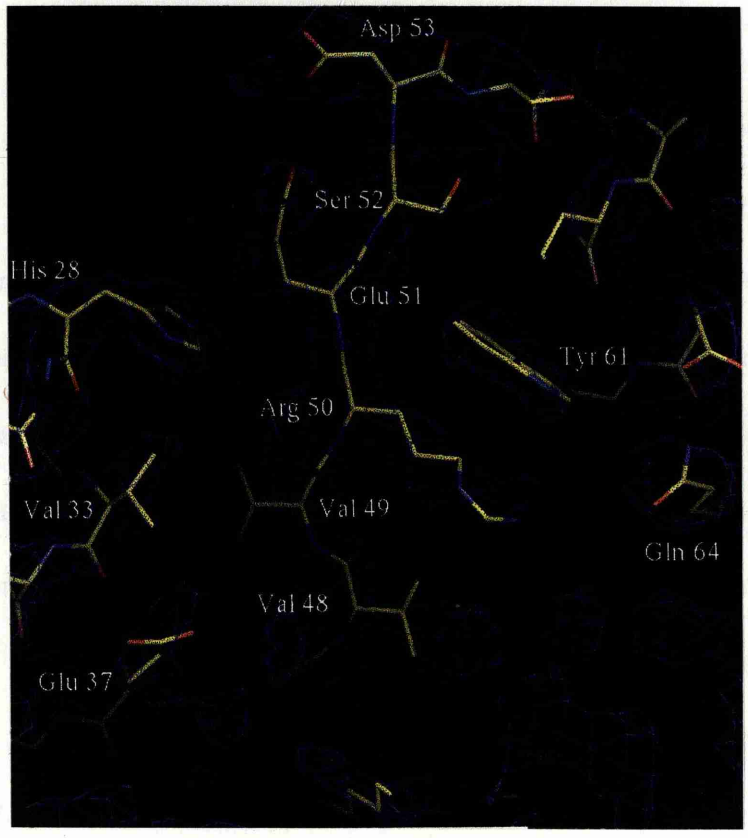



**Figure 7-2b Comparison of DM with MIR map**

This diagram shows a portion of the DM map and the refined structure. In order to show the improvement in quality with DM it is in the same orientation as the MIR map shown in Figure 7-2a

## 7.4 Interpretation of the Maps

### 7.4.1 Making Skeletons of Electron Density

The program Bones ( Jones & Kjeldgaard 1994) was used to skeletonise the maps which were subsequently read into O. Skeletonisation is done by Bones to automatically reduce the 3 D electron density map to lines following the connectivity ( Greer, 1985). It works by reducing the electron density to a series of integers that correspond to contour levels above a user selected base threshold. These values are skeletonised and then these fragments are classified by a user- defined length minimum to determine whether they are side chain or main chain.

Bones was run several times to find the optimum input values for base level and contour interval and minimum main chain length that gave skeletons that showed the clearest true features. The best skeleton showed quite clearly where there was β-sheet and less evidently where there was α-helix in the protein.

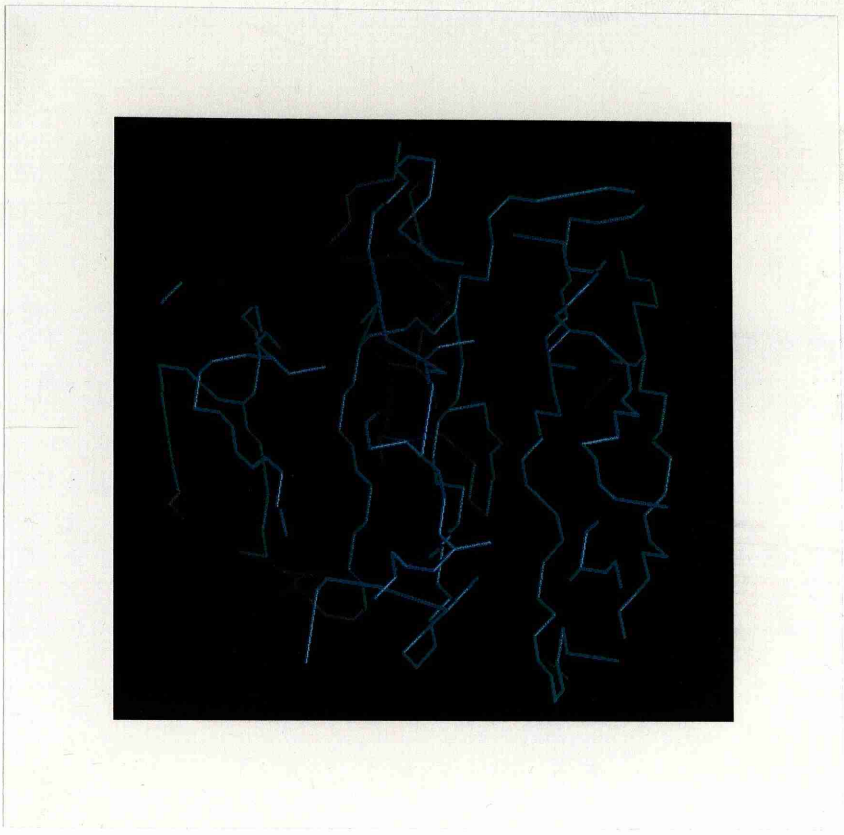


**Figure 7-3 Bones from 2.6 DM map**

This diagram shows the mainchain bones made from the DM map.

## 7.5 Initial Tracing

### 7.5.1 Tracing Of Main Chain

A model was fitted into the DM map using baton command in O to assign Cα positions (using the edited bones described in section 7.4.1 and shown in Fig 7.3 as a guide). The initial model was a polyalanine trace consisting of three polypeptide segments of 118 residues in total. This model had four helices and five strands and is shown in Fig 7.4. The overall fold was identified as being an α/β type.

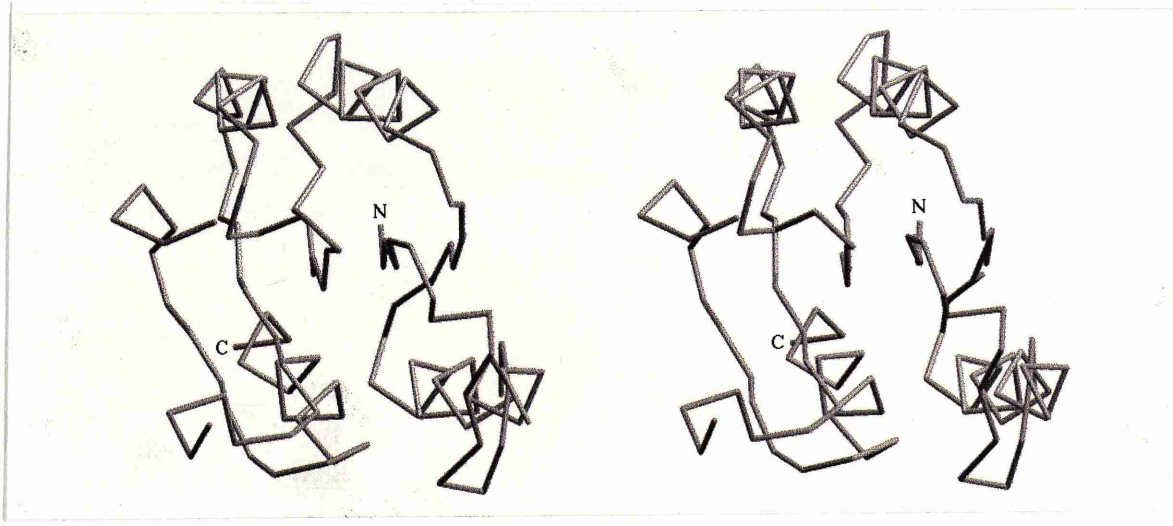



**Figure 7-4 Stereo Diagram Showing The Intial Backbone Trace of DHQ**

The intial backbone trace showed 4 helicies and five strands.

### 7.5.2 Assignment of Sequence to the Model

There was side chain density clearly visible for one of the strands in the backbone tracing. This strand was thought to be near the N terminal of the protein. The residues in this strand gave a good fit of sequence to electron density and this is shown in Fig 7.5. There was a reasonable fit of sequence in the following loop but

after that there was a gap in the density and the sequence fit after the gap was not clear.

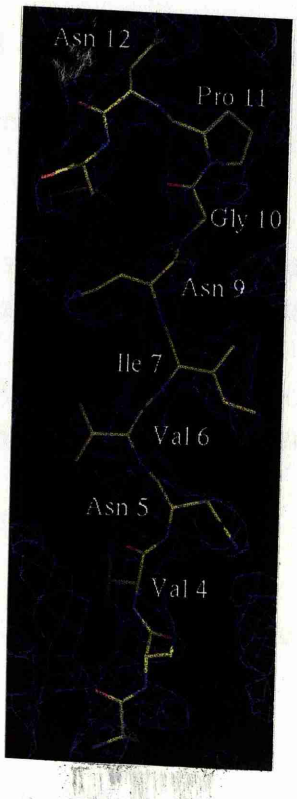



**Figure 7-5 The Fitting of the First Strand Into the DM Map**

The sidechain density for Asn 6, Ile 8, Asn 9 and Asn 12 is good. The map density is contoured at 1σ.

In order to assist the assignment of sequence to the model a secondary structure prediction was run using the program PHD (Rost *et al.*, 1994). PHD predicted the structure to have five strands and five helices and a diagram of this structure predication is shown is fig 7.6. In the secondary structure prediction the one and only tryptophan was predicted to be in the second helix. The density for what was assumed to be the second helix from the back bone tracing was examined and density for a very large residue was found. The tryptophan fitted well into this density (Fig. 7.7) and the sequence was traced forward from this to Arg 108 where there was

another gap in the density. The validity of this tracing was supported by the fact that the mercury heavy atom was bound, in this tracing, to the one and only cysteine residue, as shown in fig 7.8. There was good density for the cysteine and when the map was contoured at a level of five sigma the only density that was observed was that for the sulphur of the cysteine.

```
MSELIVNVINGPNLGRLGRREPAVYGGTTHDELVALIEREAAELGLKAVVRQSDSEAQLL
     EEEEE              EE    HHHHHHHHHHHHHHH   EEEE   HHHHHH

DWIHQAADAAEPVILNAGGLTHTSVALRDACAELSAPLIEVHISNVHAREEFRRHSILSP
HHHHHHH    EEEE         HHHHHHHHHH    EEEEEE    HHHHHHHH

IATGVIVGLGIQGYLLALRYLAEHVGT
  EEEEE   HHHHHHHHHHHHHH
```

**Figure 7-6 The Secondary Structure Predicion for DHQ**

H Helix

E Strand

This prediciton was made using the program PHD

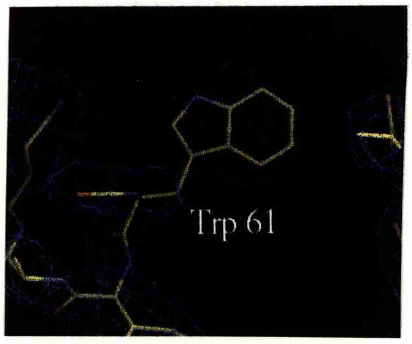



**Figure 7-7 The Fitting of Trp 61 into the DM density**

This diagram illustrates the quality of the density for the single Trp residue in DHQ. The model in the density is that of the refined structure. The density in contoured at 1$\sigma$.

This tracing was also supported by the fact the Pt appeared to be binding between His 88 and His 108 (see Fig 7.9). The sequence was then fitted backwards from the Trp 61 to residue 28 that was just after the gap in density mentioned above. Although there was a second gap in density after Arg108 there was clear sidechain density for residues 116 to 142. At this point 118 of the 146 residues had been found and a model of the monomer is shown in Fig 7.10. A Ramachandran plot of the trace is shown in Fig. 7.11. This model was then subject to refinement.

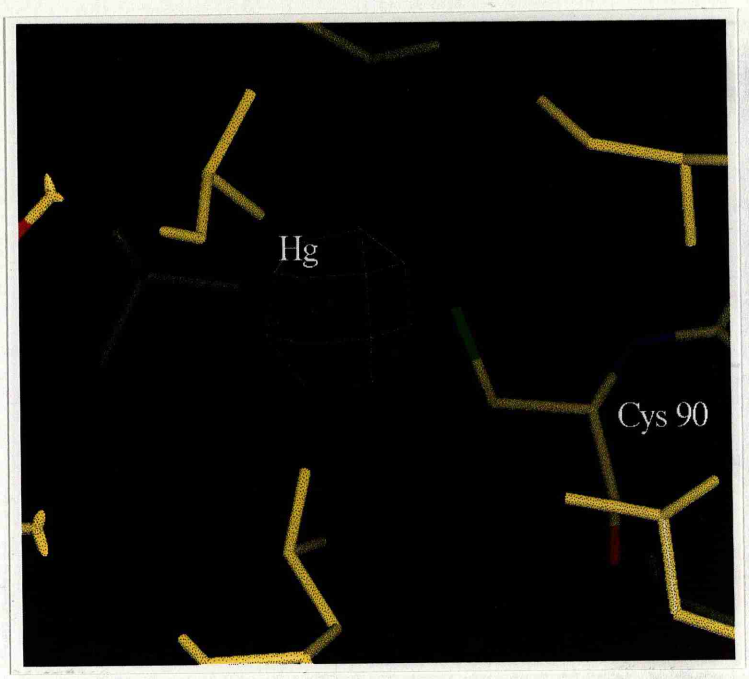



**Figure 7-8 The Mercury Binding Site**

The sulphur of Cys 90 is shown in green. The white density is the difference density for the mercury displayed. The centre purple cross is the position of the mercury after refinement in MLPHARE.

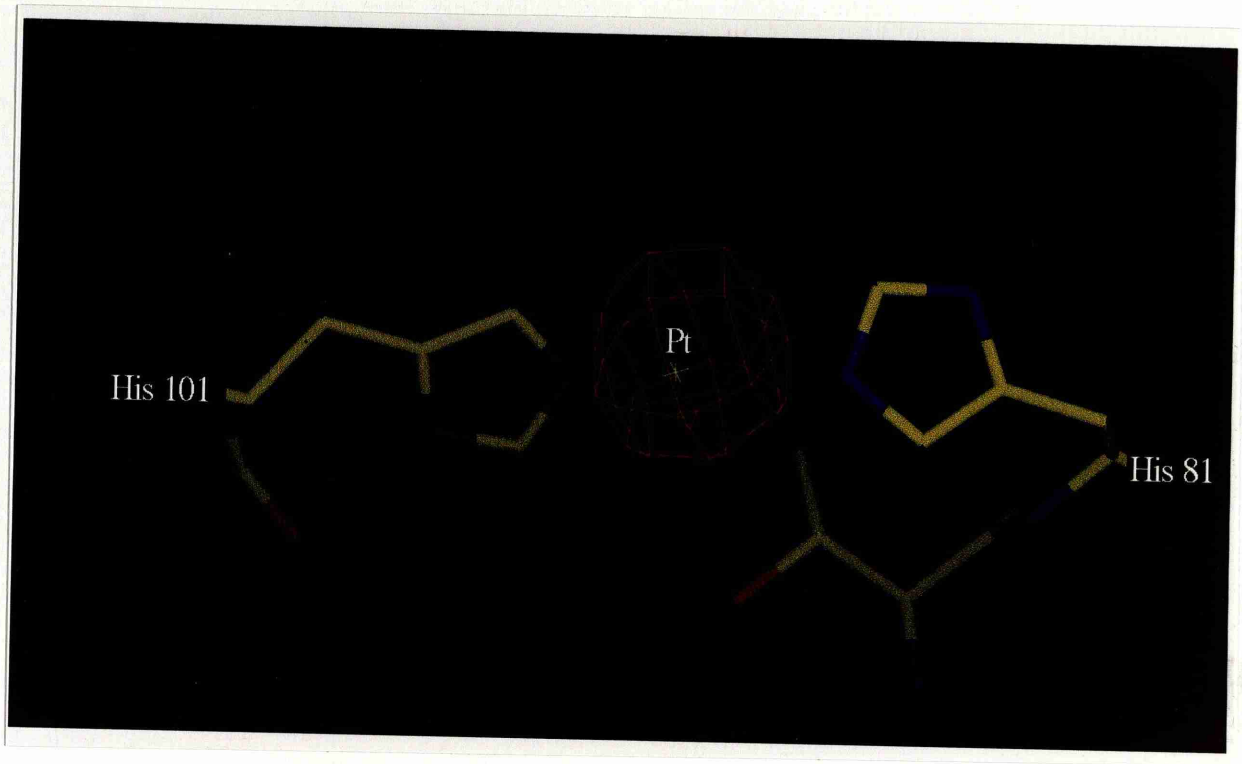



**Figure 7-9 The Platinum Binding Site**

The platinum difference density is shown in red.

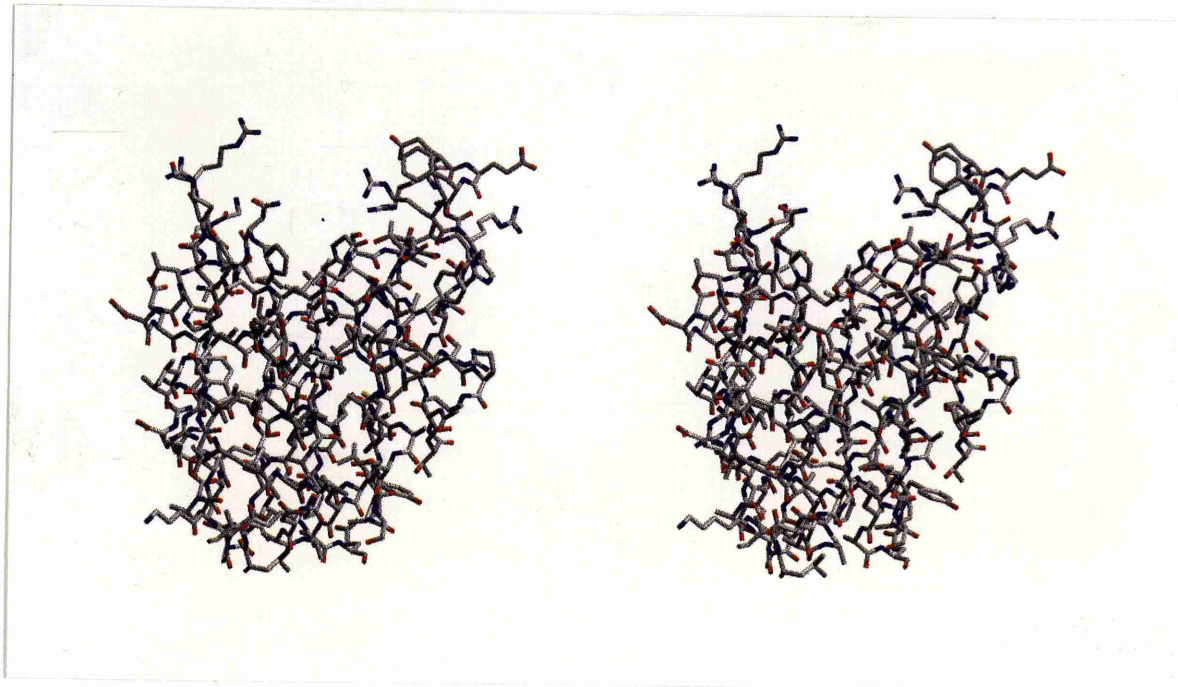
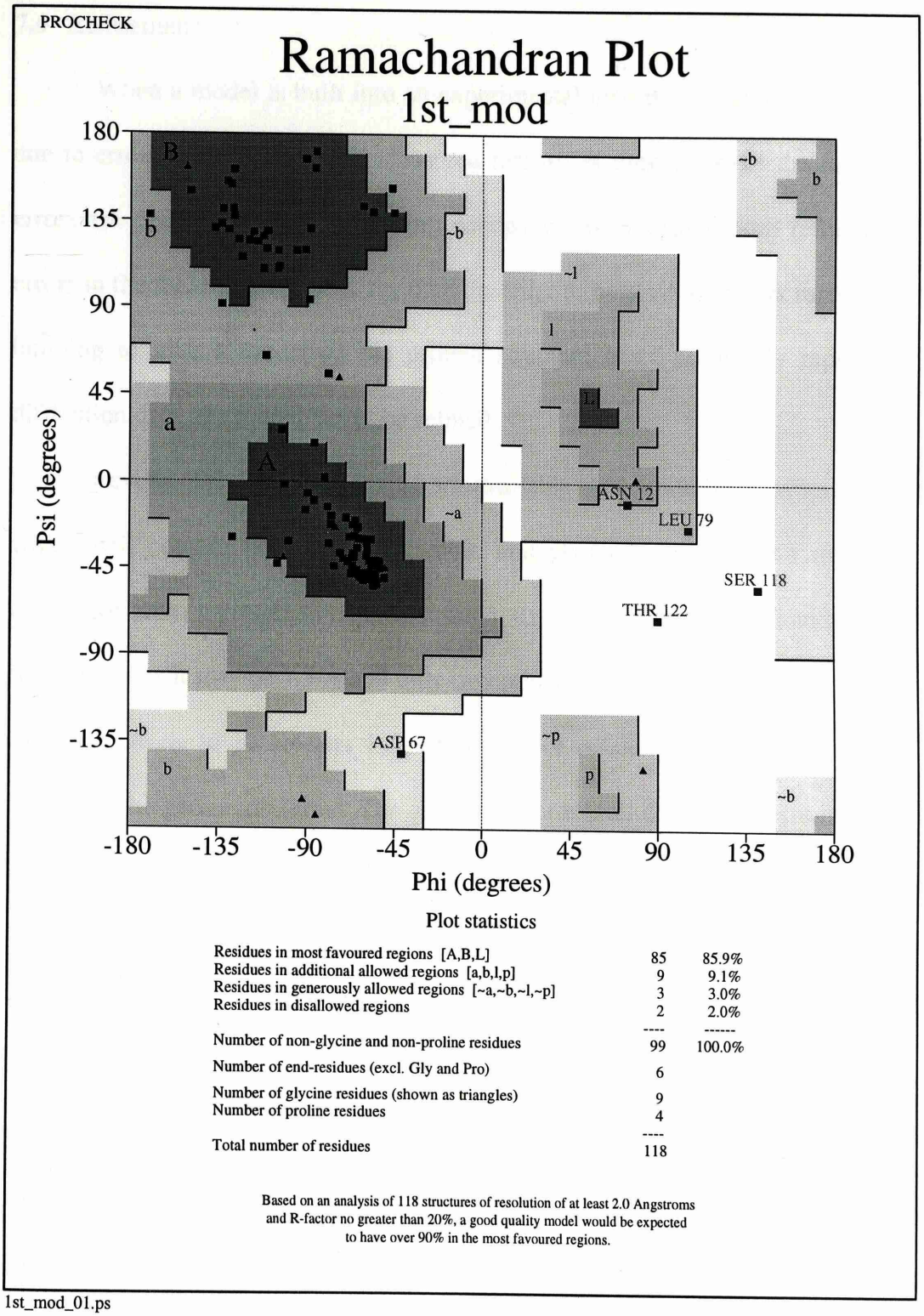


**Figure 7-10 Stereo Diagram of the Initial Model of DHQ**

All the atoms in the initial model are shown in standard colours.




1st_mod_01.ps

**Figure 7-11 The Ramachandran Plot for the Initial Model of DHQ**

## 7.6 Refinement

When a model is built into an experimental map it will usually contain errors due to errors in the phases and measured intensities used to make the maps. These errors are caused by low heavy atom occupancy, poor heavy atom refinement and errors in the measurement of X-ray data. In order to remove the errors made in model building to give a model of the protein structure that accurately represents the diffraction data, the model has to be refined.

Several refinement programs are available that minimise a residual function which will contain both crystallographic and geometric terms. The resolution of diffraction from a protein crystal is limited, usually being between 2 and 3 Å, while in better cases it may be 1.5 Å and only exceptionally 1.0 Å. This means that the ratio of observations to parameters is inadequate for refinement. This is overcome by including the geometric terms. The geometric terms can consist of restraints on bond distances, bond angles and dihedrals and of planar contraints placed on certain groups of atoms. These are all based on tables of standard geometry derived from hundreds of structures from the Cambridge Structural Data base. There are two different methods of including geometric terms in refinement. One of these methods is to express the geometrical terms as stereochemistry restraints (Konnert, 1976). The other method is to express it as energetic restraints. The programs PROLSQ and REFMAC incorporates stereochemical restraints method. The program X-PLOR implements the energetic restraint approach.

The crystallographic terms in all programs are based on a sum of observed structure factor amplitude minus calculated structure factor amplitude over all diffraction space. A refinement cycle usually consists of computational refinement

followed by manual rebuilding of the model. Manual rebuilding generally involves correction of major errors too large for rapid automatic correction by computational refinement, or addition or removal of atoms. This is done using a three-dimensional molecular graphics workstation with programs like O or FRODO. There are three main methods used by computer programs that refine protein structures: the conventional least squares method which is implemented in programs like PROLSQ, molecular dynamics which is used in X-PLOR and refinement using maximum likelihood which is implemented in REFMAC.

Standard refinement programs like PROLSQ and X-PLOR work by minimising a residual that is the weighted sum of squared deviations between the observed ($\mathbf{F_o}$) and calculated ($\mathbf{F_c}$) structure factors amplitudes, including a relative scale factor k:

$$\sum w(F_o - kF_c)^2 + \text{geometric terms}$$

The main difference in these programs is the minimisation methods. PROLSQ uses a method introduced by Konnert and Hendrickson (1976,1980). It involves a procedure where the Cartesian co-ordinates of each atom are allowed to vary independently, with the stereochemistry being restrained by forces between the atoms. This involves minimisation of a composite observational function of diffraction data and sterochemical restraints

X-PLOR (Brunger *et al.*, 1987; 1990; 1993) performs computation refinement using molecular dynamics. The program works by exploring ranges of conformational space through molecular dynamics and analyses the energy function of the structure ($E_{tot}$), which contains a combination of the empirical ($E_{emprical}$) and effective energy ($E_{effective}$) terms: $E_{tot} = E_{emprical} + E_{effective}$

$E_{emprical}$ describes the energy of the molecule as a function of the atomic coordinates (Lifson and Stern 1982; Burkert and Allinger 1982; Brooks et al. 1983; Némethy *et al.*, 1983; Karplus & Petsko 1990). The empirical energy functions contain conformational and nonbonded interaction energy terms involving sets of two, three, and four atoms. A harmonic approximation is used to account for deformations in bond length and angles. Four-atom terms are used for torsion potentials. Two-atom terms are employed for the nonbonded interactions. In X-plor the empirical energy function has the general form:

$$E_{emprical} = \sum_{p=1}^{N} [w_{BOND}^{P} E_{BOND} + w_{ANGLE}^{P} E_{ANGLE} + w_{DIHE}^{P} E_{DIHE} + w_{IMPR}^{P} E_{IMPR} + w_{VDW}^{P} E_{VDW} + w_{ELEC}^{P} E_{ELEC} + w_{PVDW}^{P} E_{PVDW} + w_{PELEC}^{P} E_{PELEC}]$$

here

$E_{BOND}$ describes the bond energy with $w_{BOND}^{P}$ being its weight.

$E_{ANGLE}$ describes the bond angle energy with $w_{ANGLE}^{P}$ being its weight.

$E_{DIHE}$ describes the dihedral angle energy with $w_{DIHE}^{P}$ being its weight.

$E_{IMPR}$ describes the improper angle energy with $w_{IMPR}^{P}$ being its weight.

$E_{VDW}$ describes the non-symmetry-related van der Waals energy with $w_{VDW}^{P}$ being its weight

$E_{ELEC}$ describes the non-symmetry-related electrostatic energy with $w_{ELEC}^{P}$ being its weight.

$E_{PVDW}$ describes the van der Waals energy between symmetry-related atoms with $w_{PVDW}^{P}$ being its weight.

$E_{PELEC}$ describes the symmetry-related electrostatic energy with $w_{PELEC}^{P}$ being its weight.

The effective energy term confines the exploration to regions allowed by the X-ray data and error estimates. The use of molecular dynamics can help to avoid local minima (Brunger *et al.*,1987) using a procedure called simulated annealing. By "heating" to high temperatures and slowly cooling the molecule using molecular dynamics, the program searches for a lower energy structure. This procedure is based on the principle that if an increase in kinetic energy is modelled into the system then local energy minima can be overcome to find global ones. The simulated slow cooling allows the molecule to find the lowest energy state. An analogous physical example is when a substance in a glass state is heated to melt it and then allowed by cool slowly to achieve an energetically lower state.

## 7.7 Refinement By Maxium Likelihood

As mentioned in the previous section PROLSQ and X-PLOR both refine by using a least squares residual ( $\sum w(F_o - kF_c)^2$ + geometric terms). This is not an ideal minimization target as what in effect is minimised (ignoring errors) is the rms deviation between the model electron density and the density computed using Fourier coefficients $F_o\exp(i\alpha_c)$ (Silva & Rossman 1985). Minimisation of this deviation can be achieved by either improving the model or by introducing systematic errors that remove the difference between the model and the $F_o\exp(i\alpha_c)$ map. This is a real problem in macromolecular refinement where the ratio of parameters to observations is low and overfitting of data with the introduction of systematic errors is common.

Refining by maximising the log-likelihood gain rather than minimising the conventional least squares residual helps reduce this problem (Read ,1990; Bricogne 1991,1993,1996; Pannu & Read 1996; Murshudov *et al.*, 1996). This is the method

that is applied in maximium likelihood refinement. The principle behind maximum likelihood is that the quality of a model can be judged by its consistency with the experimental observations. This means that if the model is correct then there would be a reasonably high probability of making an observation with that value. If all relevant observations are taken as a set then the probability of making the entire set of observations is an excellent measure of the quality of the model. If the assumption is made that all the observations are independent then the joint probability of making the entire set of observations is the product of the probabilities of making each independent observation. This joint probability is the likelihood function (*L*).

$$L = \prod_{hkl} p(F_o; F_c)$$

If the logarithm of the likelihood function is taken and multiplied by negative one with $\Lambda$=-ln(L). We have the minimisation problem

$$\Lambda = -\sum_{hkl} \ln\left(p(F_o; F_c)\right)$$

## 7.8 Maximum Likelihood Refinement Using Refmac

In this project refinement was done using the maximium likelihood method. This method was applied using the program REFMAC (Murshudov *et al.*, 1996,1997) and had to be adapted to work in cubic space groups by the authors before this work could be performed.

Once the overall parameters of likelihood have been worked out REFMAC performs one cycle of atomic parameter refinement. It does this by calculating the gradient and the diagonal terms of the second derivative matrix. The gradient

calculation is done by convoluting the atomic density with the difference density as suggested by Agarwal (Agarwal, R.C. 1978; Agarwal *et al.,* 1980)..

At the end of a cycle the program writes out map coefficients for SIGMAA ( Read 1994) style $(m|F_o| - D|F_c|)$ and $(2m|F_o| - D|F_c|)$ maps.

### 7.8.1 Correctness of the Refined Structure

The progress of a crystallographic refinement is normally monitored using the residual index or R-factor:

$$R = \frac{\sum \| F_{obs} | - | F_{calc} \|}{\sum | F_{obs} |}$$

This is an indication of the relative error of the model, and as the fit of the model to the data improves the R-factor will decrease, because the agreement between $F_{obs}$ and $F_{calc}$ will improve. This is generally accepted as an approximate guide-line to the progress of the refinement procedure. Care is necessary however as the R-factor can be reduced by fitting noise in the data, or be pushed down artificially by applying resolution cut-offs, by the improper addition of water molecules or by the individual refinement of temperature factors when the resolution of the data does not warrant it. Doing such things can mask gross errors and lead to wrong structures that are not discovered till well into refinement or even after publication. For this reason a cross-validation called the Free R-Factor is also used.

### 7.8.2 The Free R-factor

The Free-R factor for cross-validation was first used by Axel Brunger (1992). It involves setting aside a small fraction of data into a test set which is not used in the refinement, but for which an R-factor is nevertheless calculated all the time. The test

set should contain at least 500 reflections. This Free R-factor is not susceptible to being reduced by overfitting because the reflections it is calculated from are not used in the refinement. Comparing the values of the conventional and free R-factor indicates the extent to which one has over-fitted the data as well as about the quality of model and data. As mentioned before the refinement program will use any degree of freedom it is given to reduce the discrepancy between observed and calculated structure factors. However, since the data is affected by error, and since the model is not an exhaustive description of all scattering matter in the crystal (space- and time-averaged), this easily leads to a situation in which the errors are compensated by erroneous changes to the model. The Free R-factor gives an indication of whether the addition of more variables (eg the addition of extra atoms or the use of individual temperature factors) to the refinement proceedure is valid. If a refinement step in which more variables are introduced than in the previous cycle does not lead to a significant drop in R-free, any drop in R-factor is likely to have occurred from over-modelling the data.

### 7.8.3 Refinement Of Solvent Structure

The program ARP (Lamzin &Wilson 1993) was used for the addition of water molecules to the model and the subsequent refinement of their positions. ARP updates the model by identifying and removing poorly defined atoms and adding new atoms. An atom is rejected if it is located in density in the REFMAC 2Fo-Fc map below a certain cutoff (in this case below 0.5 x rms density). The addition of atoms is performed by analysis of the difference density and picking up possible new atoms

under hydrogen bond or other distance constraints that are defined by the user (usually about 2.2Å).

## 7.9 Results Of Refinement Using Refmac

The set of reflections that were used throughtout the refinement procedure to calculate the Free R-factor were assigned using the program FREERFLAG. There was 1100 reflections in this set and these reflections were never used in the refinement procedure. Refinement was done using the remaining 10138 reflections.

The strategy for refinement was to refine till the Free R-factor was no longer decreasing. REFMAC was initially run for 28 cycles as after 28 cycles the R-free was no longer decreasing. The R-factor decreased from 48.3 % to 30.0 % and the R-free decreased from 46.88 % to 34.8 %. The decrease in both these values as a function of cycle number is illustrated in fig 7.12. The model from this refinement was then rebuilt using the 2Fo-Fc map from REFMAC.

The main part of rebuilding involved tracing the missing residues from 108 to 116 into density that had become clear in this map. The side chain torsion angles of Arg 38 and Glu 55 were altered to fit the density better. Thr 80, Ser 83, Glu 110, Phe 111 and His 114 were assigned different rotamers that fitted the density better. Rebuilding of the mainchain was done for Thr 28, His 29 and Ser 54.

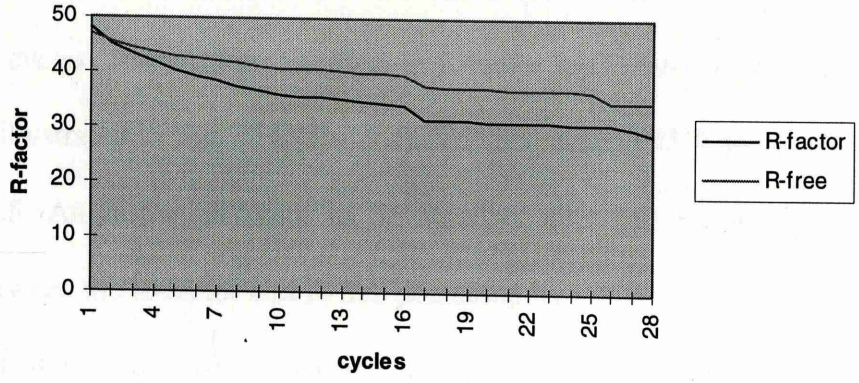



**Figure 7-12 The Initial Refinement Using REFMAC**

The rebuilt model was then subject to a further 180 cycles of REFMAC refinement. The resulting decreases in R-factor and R-free as a function of cycle number is illustrated in Fig 7.13. The R-factor decreased from 32.6 to 24.9 and the R-free decreased from 36.3 to 30.8.

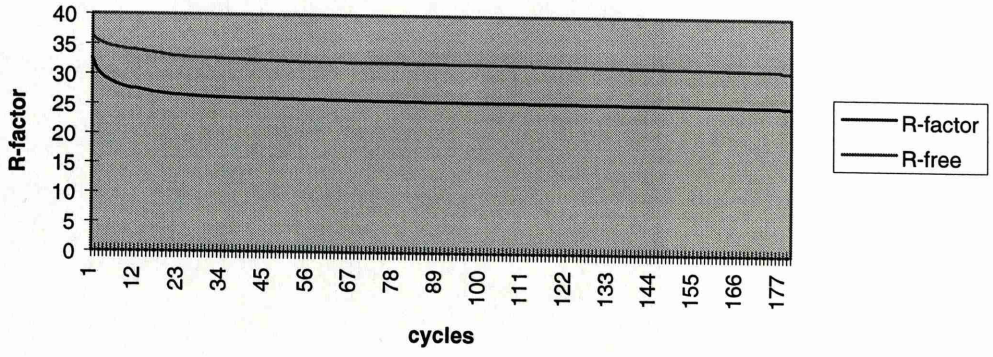



**Figure 7-13 Reduction Of R-factor and R-free in the Second Run of Refmac**

The resulting model from this refinement run was then rebuilt to remove an error in the model where the sequence from residue 67 to residue 79 was out of register with the density. This is illustrated in Fig 7.14 . This part of the model was rebuilt and the correct tracing is shown in Fig 7.15.

Other rebuilding involved adjusting the sidechain torsion angles for Glu 31, Arg50 and Ser 54 so their sidechains fitted better into density. Leu 74 and Arg 113 were assigned different rotamers. The main chain positioning of Leu 43 and Arg 108

was rebuilt to fit the density better. REFMAC refinement was carried out for a further 28 cycles. The resulting decrease in R-factor and R-free as a function of cycle number is illustrated in Fig 7.16. The R-factor decreased to 23.5 and the R-free decreased to 26.5. An important point to be noted is after this refinement run the difference between the R-factor and R-free decreased by ~ 1 % as a result of the removal of the out of register error. When the new 2Fo-Fc map was analysed the only rebuilding required was for the main chain of residues 117 to 122 and the repositioning of their sidechains. After this rebuilding ARP was used to add water molecules to the model.

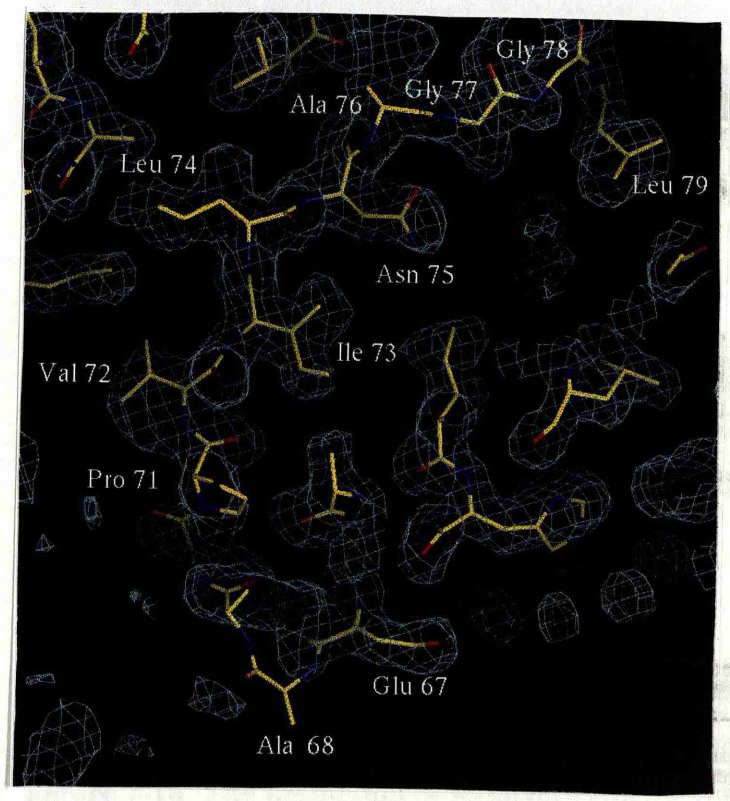



**Figure 7-14 The Mistracing Of the Map**

The mistracing of the chain from residue 67 to residue 79 is clearly revealed by the REFMAC 2Fo-Fc map contoured at 1σ.

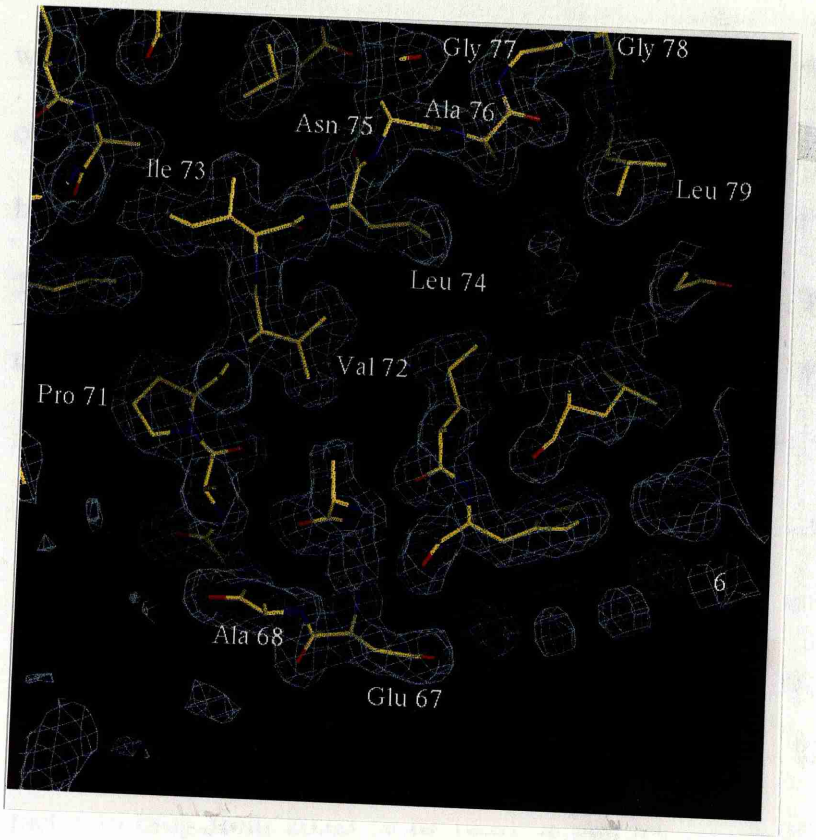



**Figure 7-15 The Corrected Tracing**

The residues from 67 to 79 are shown correctly traced in the REFMAC 2Fo-Fc map contoured at 1σ.

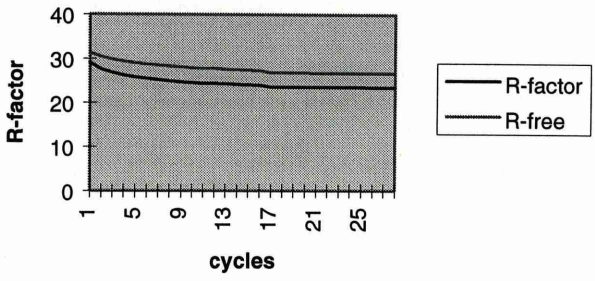



**Figure 7-16 The Reduction of R-factors During the Third Refinement Run**

ARP was run placing water molecules in the density of the Fo-Fc map. Before water molecules were placed certain conditions had to be met; the density had to be above a threshold which was defined automatically by the program using statistical analysis of the difference density distribution, water molecules were not postioned closer than 2.20 to existing atoms, water molecules had to have an N or O atom

within 3.3 Å and the water molecules would not be put closer than 2.50 Å to each other. The oxygen atoms of the new water molecules had B-factors assigned on the basis of the Fo-Fc map density height (taking into account the resolution range 20.00-2.0 Å of the map). Water molecules were removed if they were below 0.5 sigma in the 2Fo-Fc map from REFMAC or if they were located closer than 1.80 Å to protein atoms. Waters located closer than 1.80 Å to each other were merged by rejecting one atom and assigning the second to the weighted average xyz and 1/B parameters.

A round of refinement was done where ARP was run followed by 5 cycles of refinement using REFMAC followed again by ARP. This cycle was repeated 10 times. The addition of waters followed by refinement in REFMAC caused the R-factor to drop from 20.63 % to 16.61 % and the R-free dropped from 24.54 % to 21.50 %. A graph showing the progress of refinement is shown in Fig 7.17.

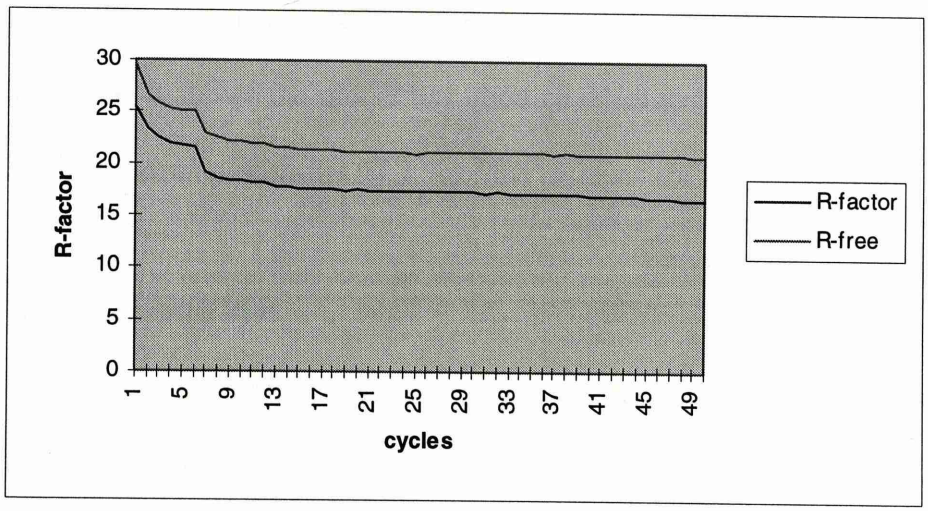



**Figure 7-17 Progress Of the Refinement Using ARP and Refmac**

## 7.10 Validating Structure Using PROCHECK

The PROCHECK suite of programs (Laskowski et al., 1993) was used for assessing the "stereochemical quality" of the refined protein structure. PROCHECK assesses the overall stereochemical quality of the given protein structure, as compared with well-refined structures at the same resolution, and gives an indication of its local, residue-by-residue reliability. To assess a structure, the program makes use of a number of parameters that have been found to be good indicators of stereochemical quality, described in detail in Morris *et al.* (1992).The checks also make use of "ideal" bond lengths and bond angles, as derived from a recent and comprehensive analysis (Engh & Huber, 1991) of small molecule structures in the Cambridge Structural Database.

The output of PROCHECK for the refined DHQ structure is given in the appendix. It shows the model has good covalent geometry and stereochemistry. The Ramachandran plot shows 97.3 % of the residues are in the most favourable region.

## 7.11 Luzzattii Plot

A Luzzatii plot ( Luzzatii, 1952) was done using both the R-factor and the R-free and is give in ( Fig 7.17). Luzzati theory applies to incomplete refinements and estimates the r.m.s. shifts needed to reach R=0. It has been stated by Cruickshank(1996) that a Luzatti plot does not provide a satisfactory method for estimating errors at the end of refinement. This is because in the least-squares method the equations for shifts are quite different from the equations to estimate variance in completed refinements. Luzzatti's theory assumes there are no errors in the diffration

data and that the only errors in the model are coordinate ones. It also assumes that the Gaussian probabilty distribution for these coordinate errors is the same for all atoms, and there is no dependence of these errors on temperature factors. In proteins however errors do increase with temperature factors. The Luzzati plot for DHQ based on the conventional R-factor indicates an r.m.s shift of 0.14 Å is required to reach R=0. The R-free based Luzzatii which has been suggested by Brunger (1992) to give a more accurate estimate of the r.m.s. coordinate error of the model gives a value close to 0.23 Å.

Cruickshank (1996) has suggested that his Diffraction-data Precision Indicator should be used as an indication of the precision of a stucture as an alternative to using a Luzatti plot. This indicates the diffraction-data-only error contribution for an atom with temperature factor equal ot the Wilson B-factor. The program REFMAC calculates this value which for the DHQ structure at the end of refinement was 14.5 %.

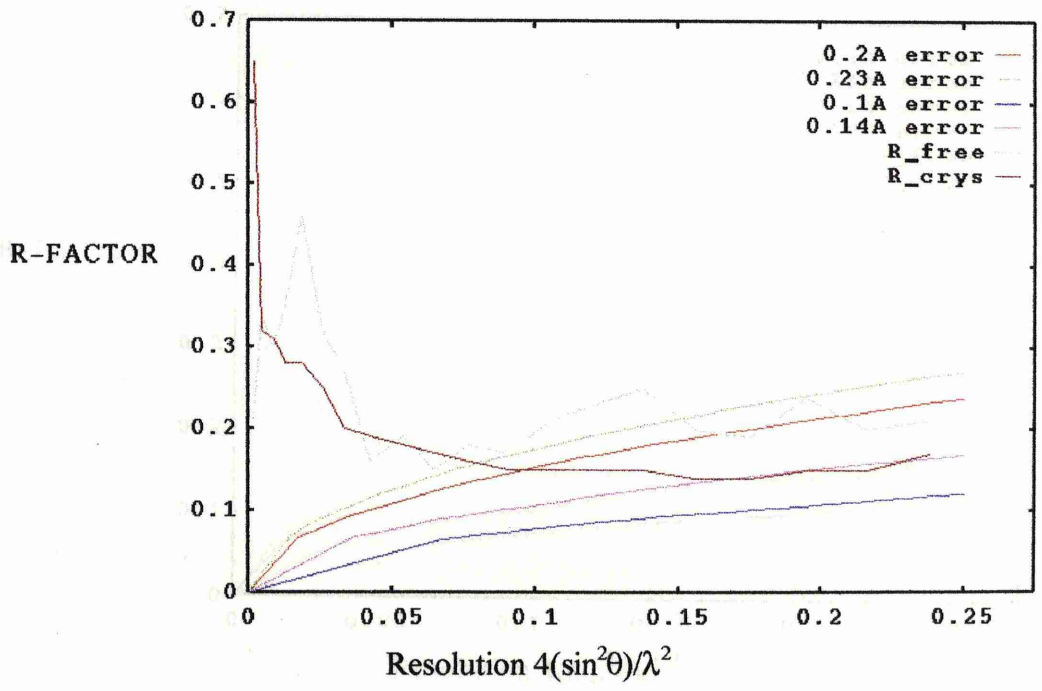



**Figure 7-18 Luzzatti Plot**

A Luzzati plot for the refined structure of DHQ showing the R-factor and R-free as a function of $(\sin^2\theta)/\lambda^2$ is shown in Fig. 7.18.

## 7.12 Conclusion

The phases calculated using multiple isomorphous replacement were modified using DM to give an improved map. A model containing most of the DHQ residues was built into this map. This model was subject to cycles of computional refinement followed by rebuilding on the graphics terminal which reduced the R-factor from 48.3 % to 23.3 % and more importantly the R-free from 46.9 % to 27.4 %. By the end of this refinement the model contained 128 of the 146 residues of DHQ. Solvent molecules were then added to the model using cycles of ARP and REFMAC which caused the R-factor to fall to 16.6 % and the R-free to falls to 21.6 %. Two residues are missing at the N terminus, four residues are missing at the C terminus and 12 residues are missing from a flexible loop between the first strand and the first helix (from Arg 18 to Gly 25). This missing loop had been determined to be highly flexible by biochemical methods and this flexibility is thought to be functionally important in the enzyme.

It can be concluded that final model is correct since the R-factor and R-free are low and the Ramachandran plot and other stereochemistry checks using PROCHECK are good. Also the structure made sense as far as having sidechains in sensible enviroments with polar sidechains exposed and hydrophobic sidechains buried. A description of the 2.0 Å structure of the type two dehydroquinase from *Mycobacterium tuberculosis* is discussed in the next chapter.

# 8. General Discussion

## 8.1 Monomer Structure Of Type II DHQ

The enzyme subunit is of a β/α type (Fig 8.1). A stereo diagram of the backbone fold is given in Fig 8.1. Secondary structure assignments have been made with a combination of the facilities in the programs PROCHECK and SETOR (Evans 1993) and by visual inspection. Secondary structural elements are mapped onto the amino acid sequence in Table 8.1. Approximately 47 % of the residues in the model are in helices and 25.3 % are in β-strands, A schematic diagram showing the arrangement of strands and helices is given in Fig. 8.3. The fold consists of a five stranded parallel β-sheet core flanked by four α-helices. The β-sheet, β2-β1-β3-β4-β5, is the same as found in flavodoxin (Burnett *et al.,* 1974). When the connectivity of the strands is classified according to the convention of Richardson (1977) the topology can be specified as -1x,+2x,+1x,+1x. These connections of successive β/α units follow the standard right handed rule which result in the second and third helices being on one side of the β-sheet and α-helices one and four being on the other (Fig 8.4). The four α-helices flank the hydrophobic core on both sides of the β-sheet and also protect from water the extensive network of hydrogen bonds.

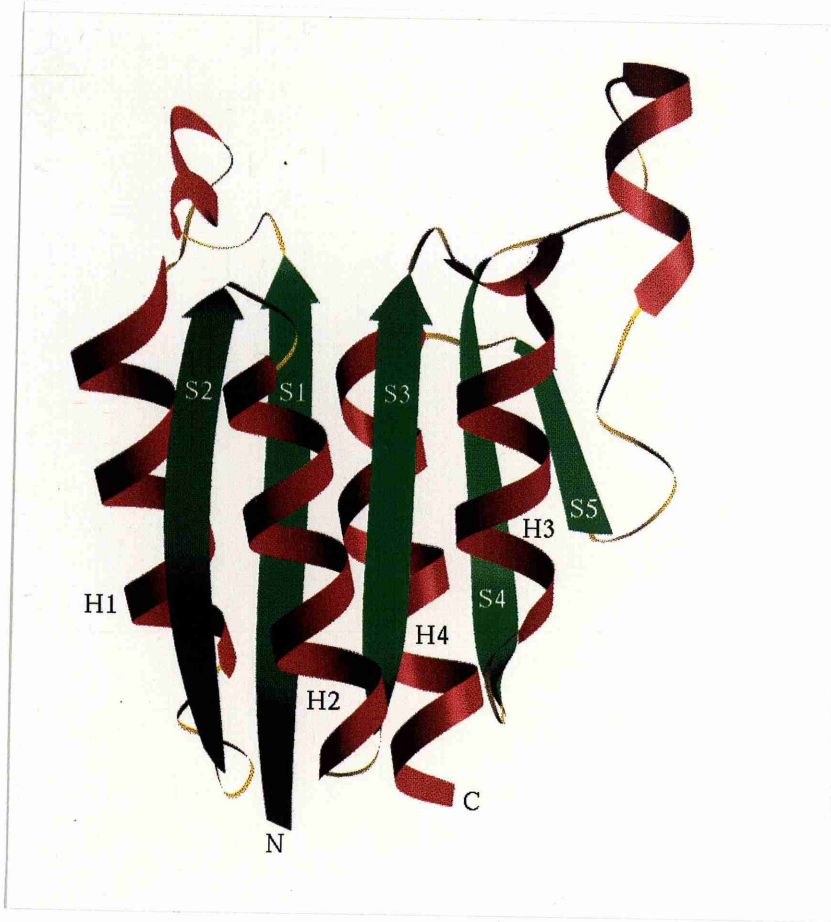



**Figure 8-1 The Subunit Fold of Type II DHQ**

Ribbon diagram showing secondary structural elements with the helicies in red and the strands in green.

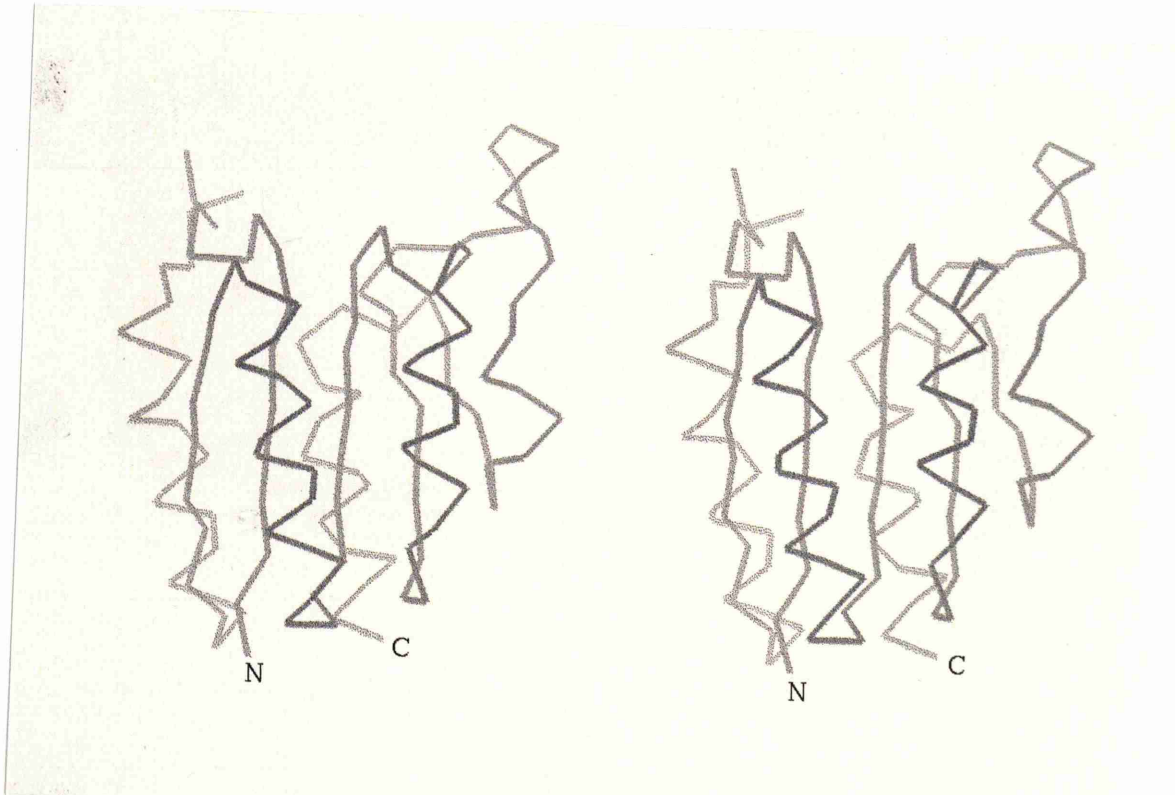



**Figure 8-2 A Stereo Diagram Showing the Backbone Fold of Type II DHQ**

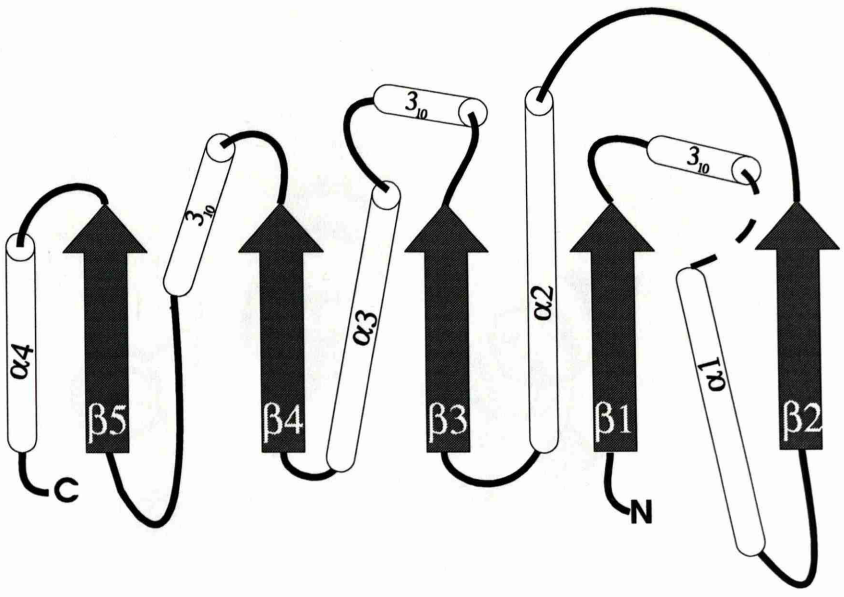



**Figure 8-3 A Schematic Diagram of the Arrangement of Strands and Helicies in Type II DHQ**

The missing loop is shown as a dashed line. Not to scale.

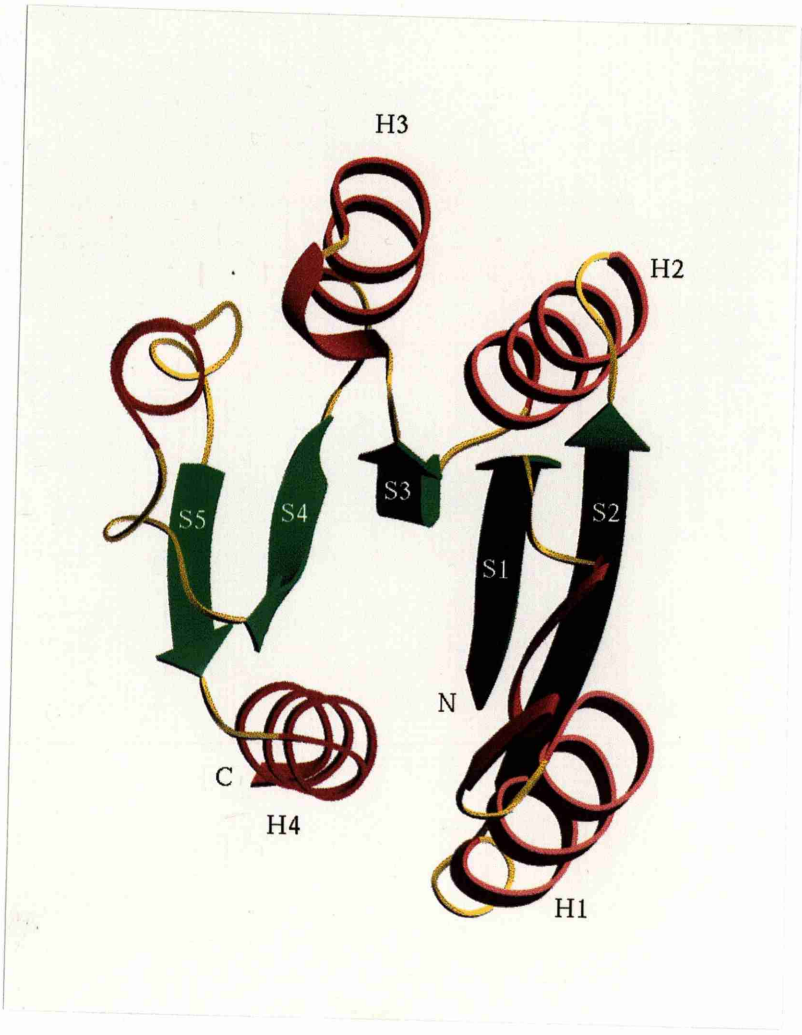



**Figure 8-4 The Subunit Fold Of Type II DHQ**

In an orientation showing how the helicies flank and protect the strands.

The first strand begins at Leu 3 and ends at Asn 9. This strand is connected to a short stretch of $3_{10}$ helix by a type II tight turn containing residues Gly 10, Pro 11, Asn 12 and Leu 13. Asn 12 has φ/ψ angles of 79 ° and -4°. A stereo diagram is given of this turn in Fig 8.5. The following $3_{10}$ helix is made up of Gly 14, Arg 15 and Leu 16. After this small turn of helix there is a missing region for residues 17 to 27 for which there was no convincing electron density.

```
SELIVNVINGPNLGRLGRREPAVYGGTTHDELVALIEREAAELGLKAVVRQSDSEAQLL
  SSSSSSSS HHHH MMMMMMMMMMMHHHHHHHHHHHHHHH   SSSSSSSS  HHHHH

DWIHQAADAAEPVILNAGGLTHTSVALRDACAELSAPLIEVHISNVHAREEFRRHSILSP
HHHHHHHHH SSSSSSS HHHHH HHHHHHHHHH SSSSSSS        HHHHHHHHH

IATGVIVGLGIQGYLLALRYLAEHVGT
 SSSSSSS  HHHHHHHHHHHHHHMMMM
```

**Table 8-1 The Secondary Structure of Type II DHQ**

S=Strand
H=Helix
M=Missing

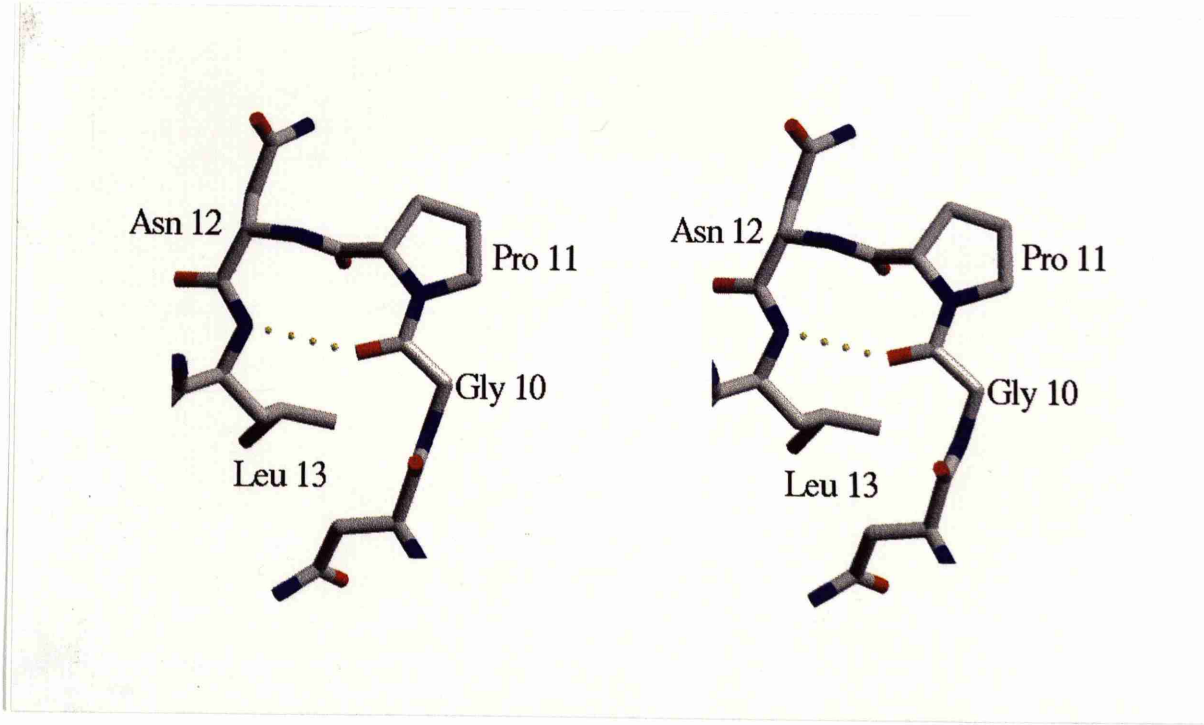



**Figure 8-5 Stereo Diagram of the Type II Turn**

Showing residues 10 to 13. With atoms shown with standard colours.

After this missing region the first α-helix begins at Thr 28. This helix ends with two sets of bifurcated hydrogen bonds; from the mainchain N of Glu 42 to the mainchain O of Arg 38 and Glu 39, from the mainchain N of Leu 43 to the mainchain O of Glu39 and Ala 40. Hydrogen bonds from the mainchain O of Ala 40 to the mainchain N of Gly 44 and Leu 45 stabilise a loop between this helix and the second strand. The second strand begins at Leu 45 and ends at Ser 52. The second helix is connected to this strand via a loop made up of Asp 53, Ser 54 and Glu 55. This loop is stabilised by hydrogen bonds from the mainchain O of Ser 54 to the mainchain N of Gln 57 and Leu 58 and from the mainchain O of Glu 55 to the mainchain N Leu 58 and Leu 59. This helix also ends with bifurcated hydrogen bonds from the mainchain N Ala 69 to the mainchain O in Ala 65 and Ala 66.

The third strand begins at Pro 71 and ends at Asn 75. After this strand there are two turns of $3_{10}$ helix starting at Ala 76 and ending at His 81. The third α-helix then starts at Ser 83 and ends at Leu 93 in a turn of $3_{10}$ helix. The fourth strand starts at Pro 96 and ends at Ile 102. There is loop from Ser 103 to Arg 108. A $3_{10}$ helix begins at Glu 109 and ends at His 114. This is followed by a region of random coil from Ser 115 to Ala 121. The fifth and final strand starts at Gly 123 and ends at Gly 127. The final α-helix begins at Leu 128 and ends at Glu 142.

A stereo diagram of the monomer showing all the non-hydrogen atoms is given in Fig. 8.6.

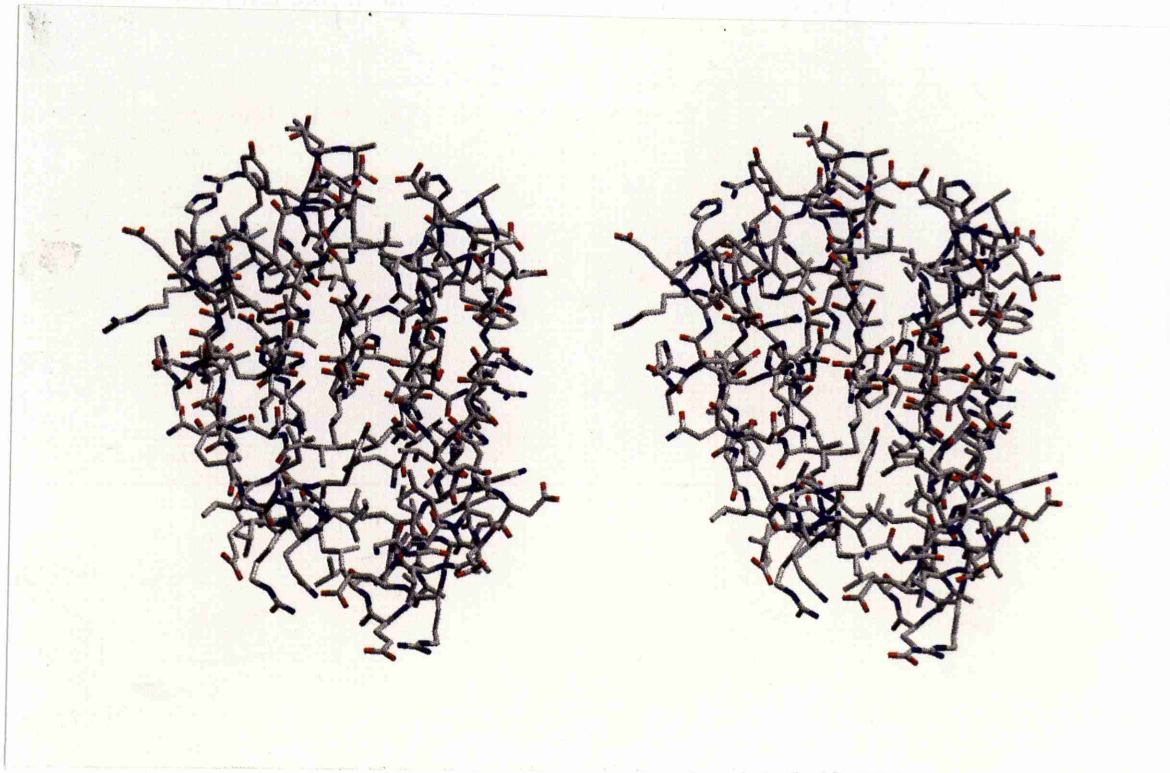


**Figure 8-6 The Stereo Diagram of Type II DHQ Monomer**
Showing all the residues with their atoms having standard colours.

## 8.2 The Dodecameric Structure

The oligomeric structure of the type II dehydroquinase is a dodecamer. The subunits combine to give an arrangement consisting of four trimers packed tetrahedrally (Fig 8.7 a-f). On forming the dodecamer each monomer buries 2226 $Å^2$ of surface area. To form the dodecamer each subunit makes two-fold contacts (Fig 8.8) and three-fold contacts (8.9).

The two-fold interface is an intersubunit anti-parallel β-sheet between two equivalent strands related by a two-fold axis (Fig 8.8). Hydrogen bond formation between pairs of subunits is facilitated as the two fold axis is perpendicular to the β-sheet and the two equivalent strands (β5) are antiparallel to each other. The β5 strands that form this interaction have a palindromic sequence GVIVG. Hydrogen bonds are formed between the Val 124 of one subunit and Val 126 of another subunit (Table 8.2, Fig 8.10). This interaction is also stabilised by hydrogen bonds from Asn 104 ND2 to the main-chain oxygens of Ser 118 and Ala 119 (Table 8.2, Fig 8.11). There is also a hydrogen bond interaction between the NH2 of Arg 113 and the main chain O of Ser 115 (Fig 8.12). This arrangements of pairs of two subunits gives an extended β-sheet of ten strands (Fig 8.8).

| Molecule 1 | Molecule 2 | Distance (Å) |
|---|---|---|
| Val 124 N | Val 126 O | 2.93 |
| Val 126 N | Val 124 O | 3.03 |
| Val 126 O | Val 124 N | 2.93 |
| Val 124 O | Val 126 N | 3.03 |
| Asn 104 ND2 | Ser 118 O | 2.86 |
| Ser 118 O | Asn 104 ND2 | 2.86 |
| Asn 104 ND2 | Ala 121 O | 2.93 |
| Ala 121 O | Asn 104 ND2 | 2.93 |
| Arg 113 NH2 | Ser 115 O | 3.25 |
| Ser 115 O | Arg 113 NH2 | 3.25 |

**Table 8-2 H-Bond Interactions Involved in Making the 2-fold contacts**

The three-fold contacts that contribute to the formation of the dodecamer are between heterologous faces. Each subunit is in contact with two others via intermolecular hydrogen bonds and salt bridges. There is a salt bridge between Arg 15

from one subunit to Asp 67 of another subunit (see Table 8.3, Fig 8.13). There are also salt bridges between Glu 109 of one subunit with Arg 87 of another subunit and between Arg 112 of one subunit and Glu 88 of another (see Table 8.3 , Fig 8.15). Hydrogen bonds that contribute to the three-fold contact are from NE of Arg 15 to OE1 and OE2 of Glu 92 of the other subunit (see Table 8.4, Fig 8.14) and between the ND2 of Asn 12 and the main chain O of Ala 85 of a second subunit (Fig 8.16) and between Asp 53 OD2 of one subunit with His 63 NE2 of another (Fig 8.14).

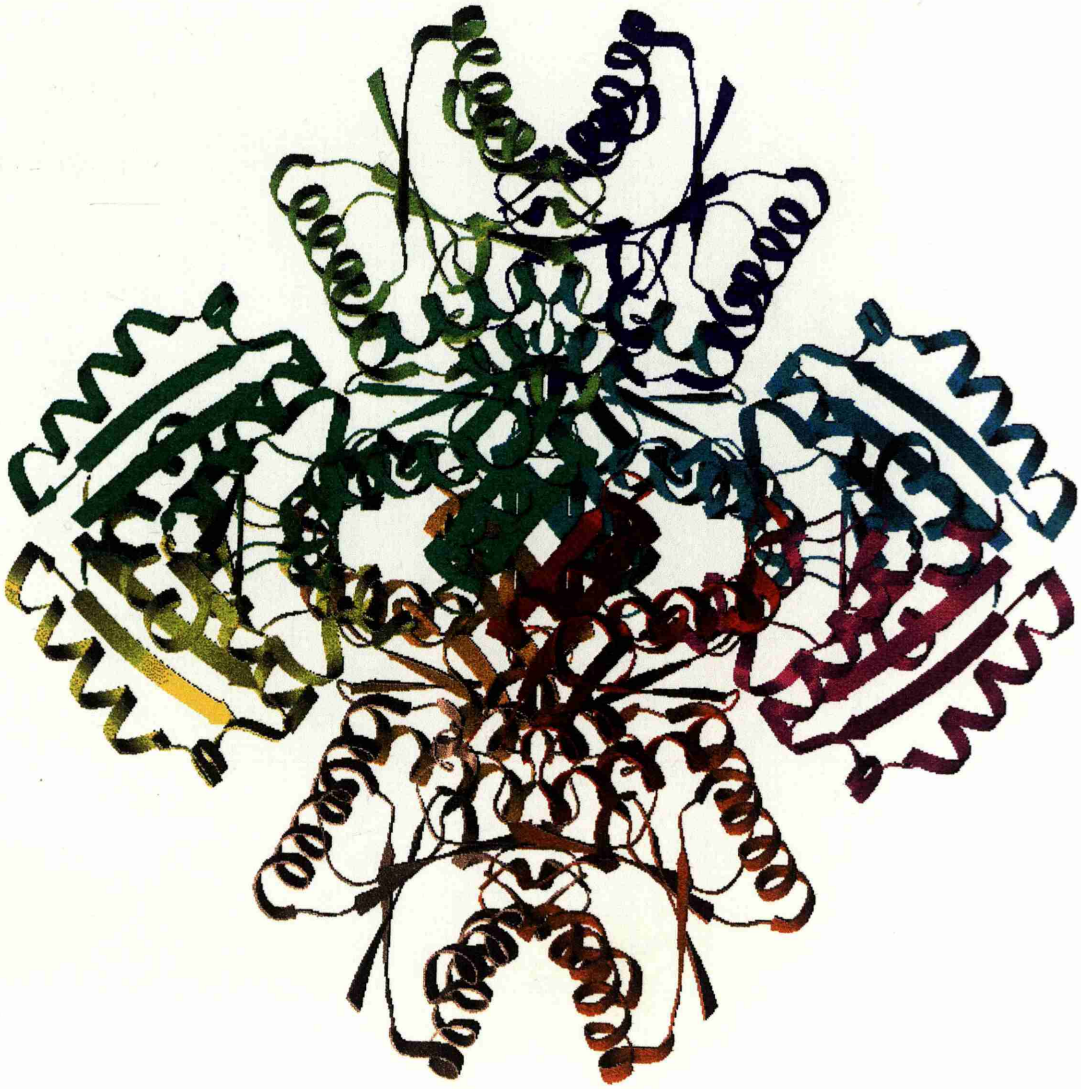


**Figure 8-7a Type II DHQ Dodecamer**

Looking down 2-fold axis and showing secondary structural elements. Each monomer has a different colour.

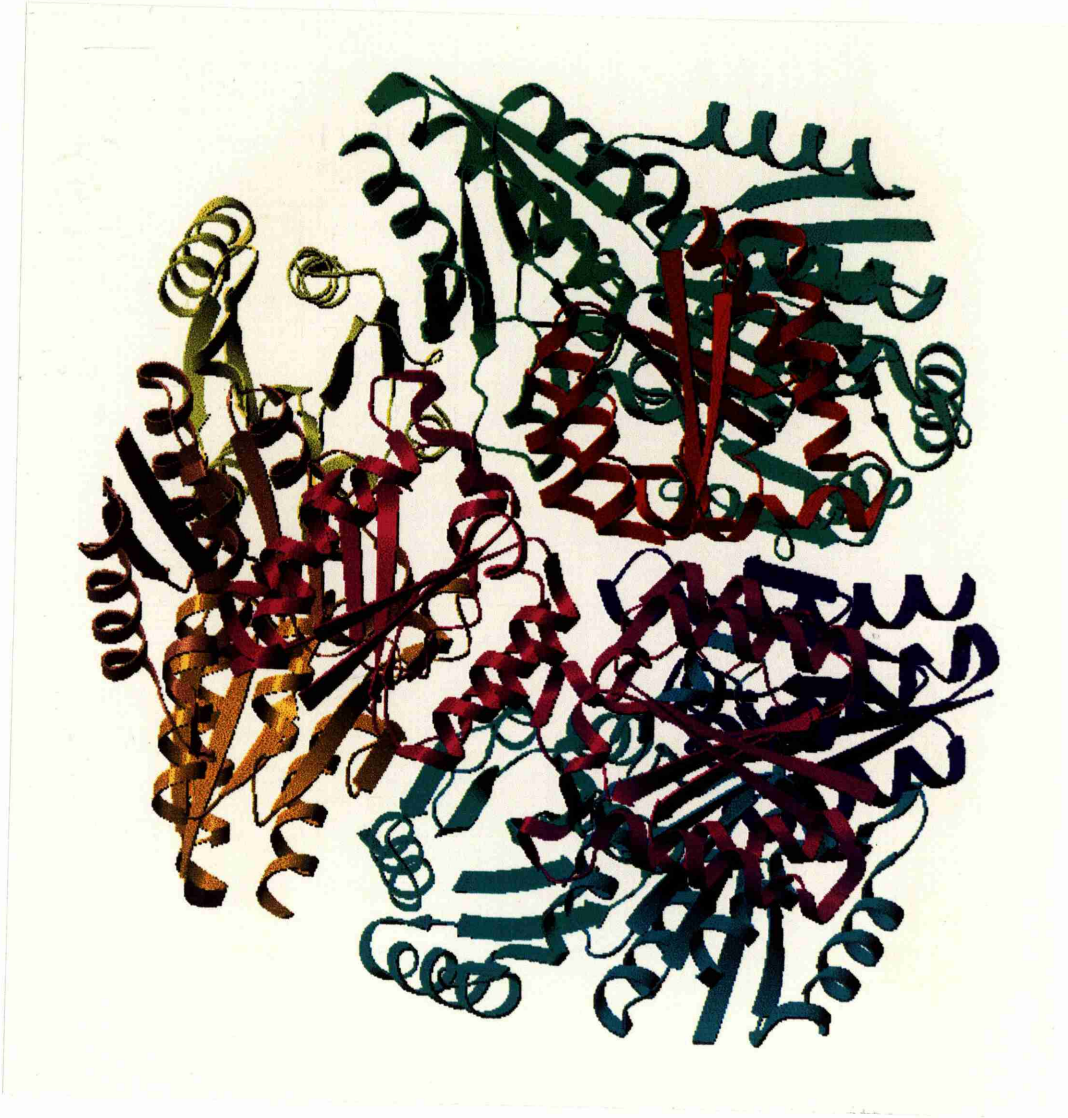


**Figure 8-7b Type II DHQ Dodecamer**

looking down 3-fold axis and showing secondary structural elements. Each monomer is shown with a different colour.

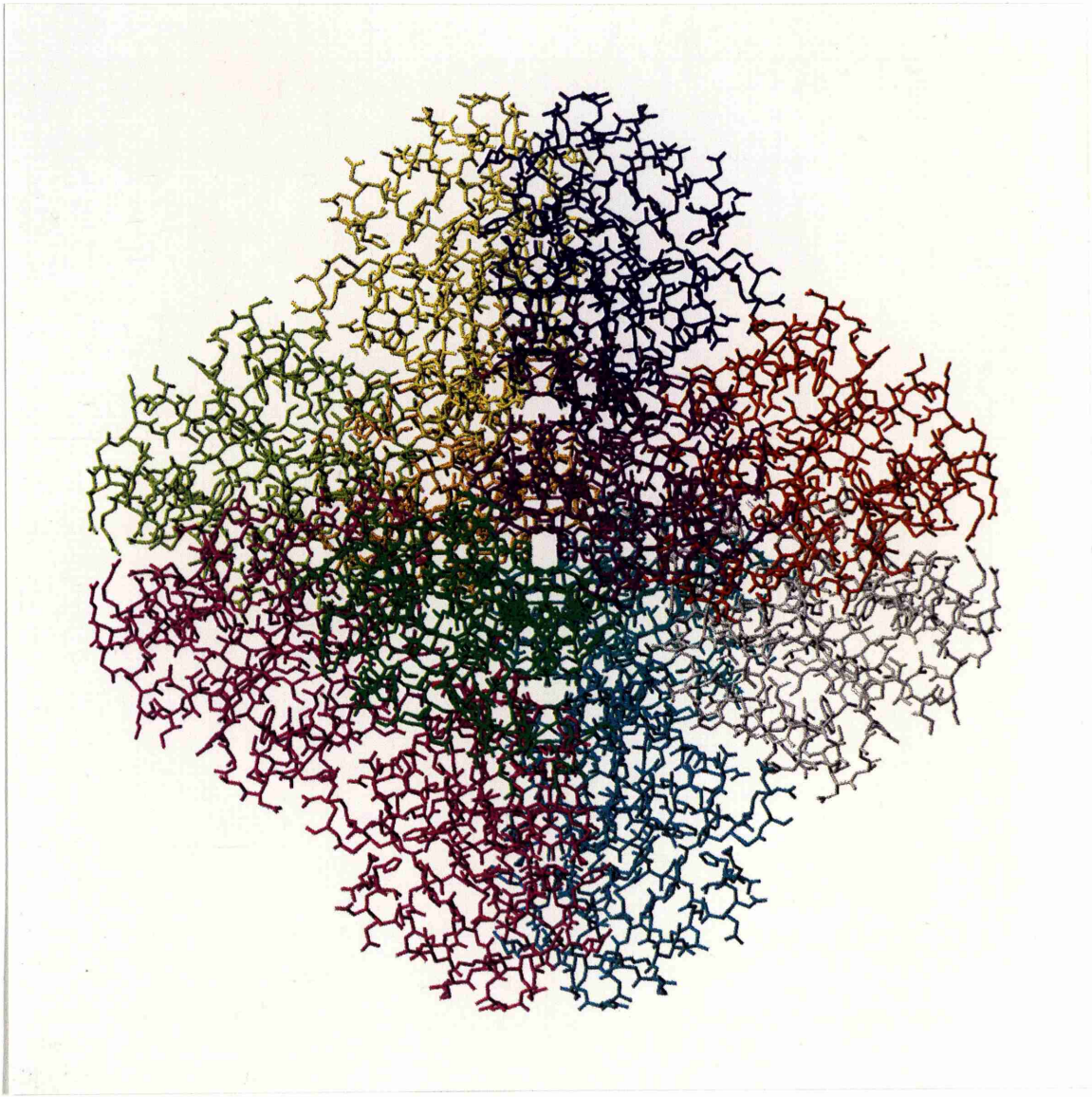


**Figure 8-7c Type II DHQ Dodecamer**

Looking down 2-fold axis and showing all residues. Each monomer is shown in a different colour.

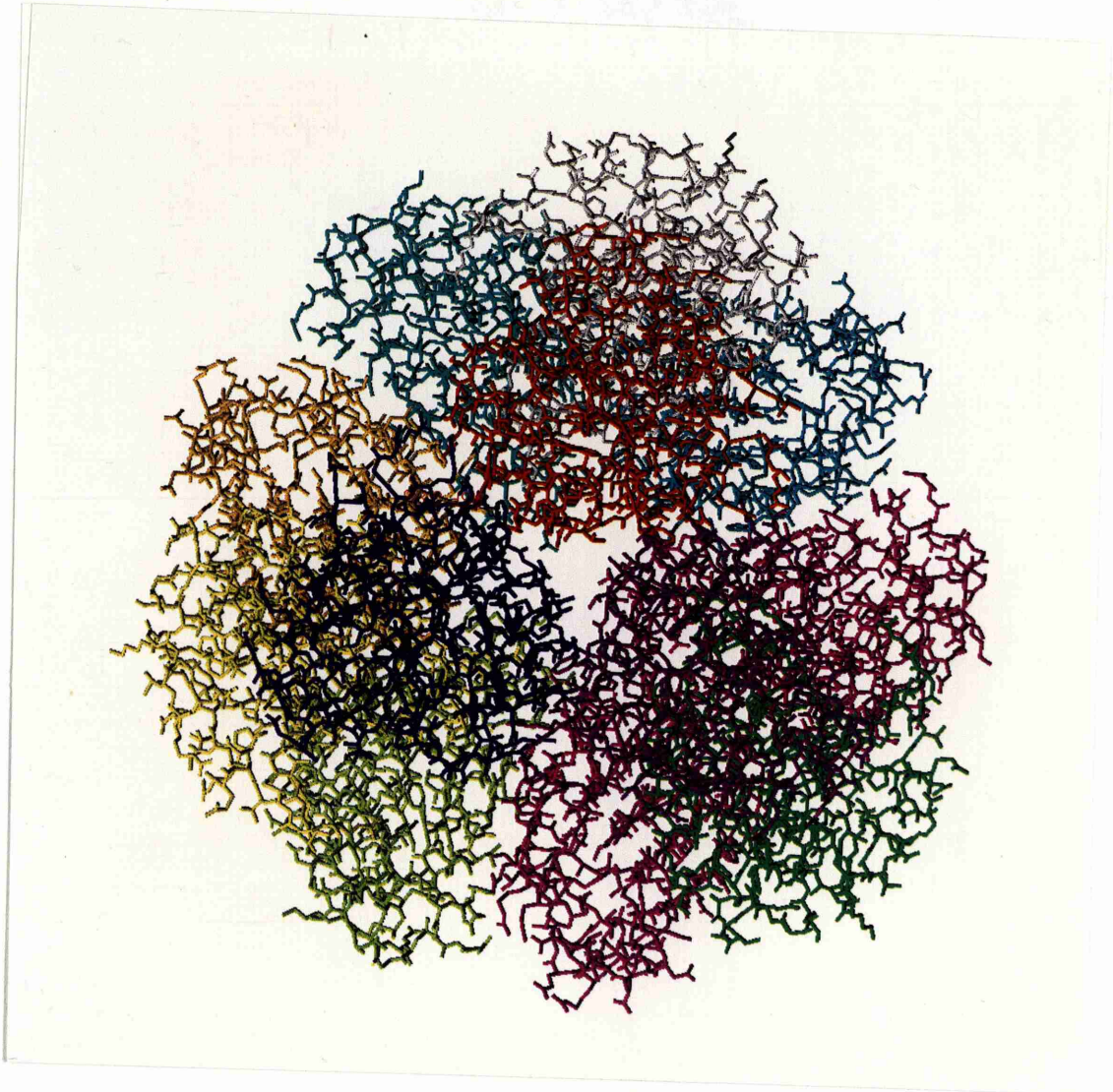


**Figure 8-7d Type II DHQ Dodecamer**

Looking down 3-fold axis and showing all residues. Each monomer is shown in a different colour.

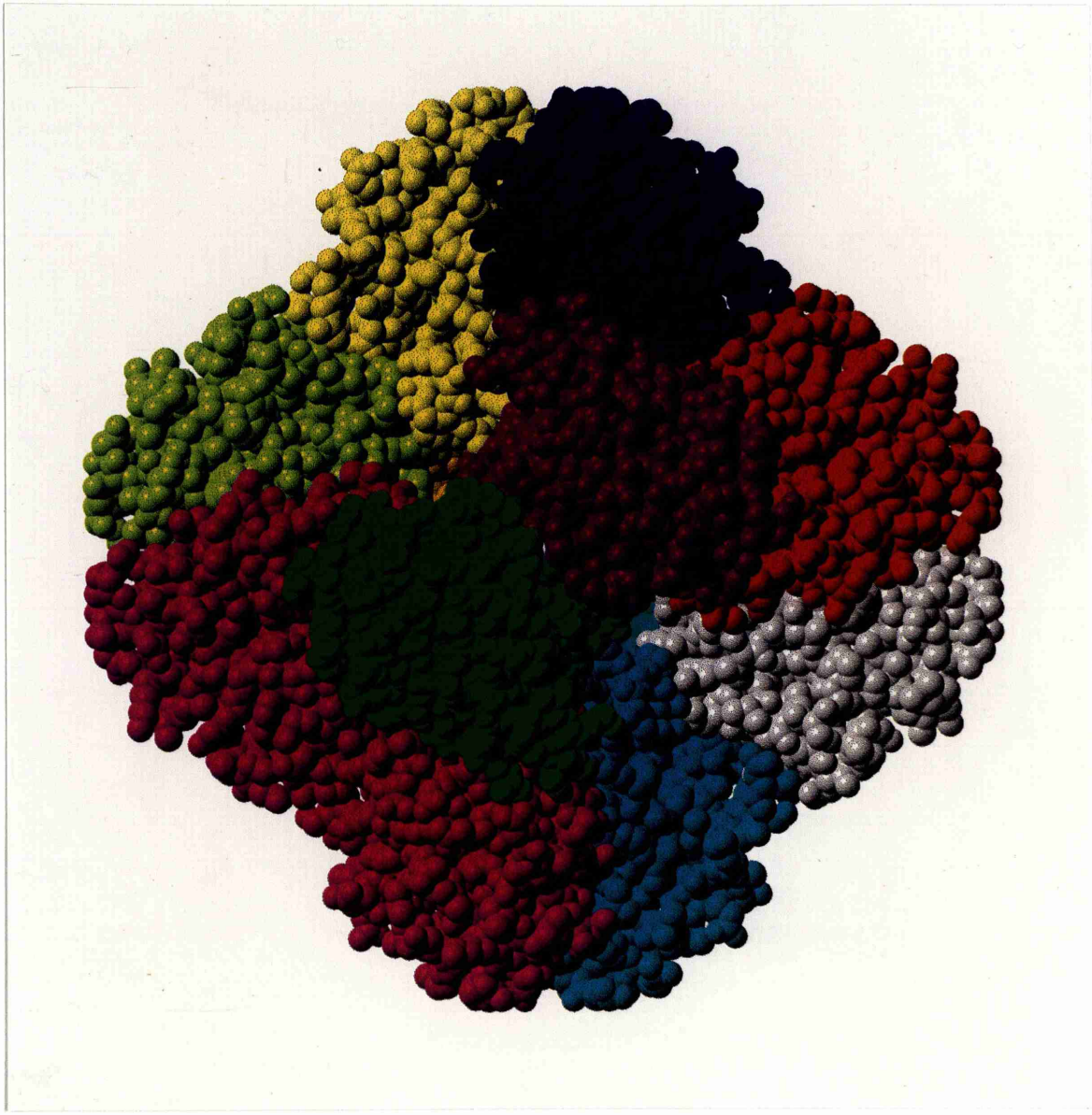


**Figure 8-7e Type II DHQ Dodecamer**

Looking down 2-fold axis and with all atoms shown as Van der Waals spheres. Each monomer is shown in a different colour.

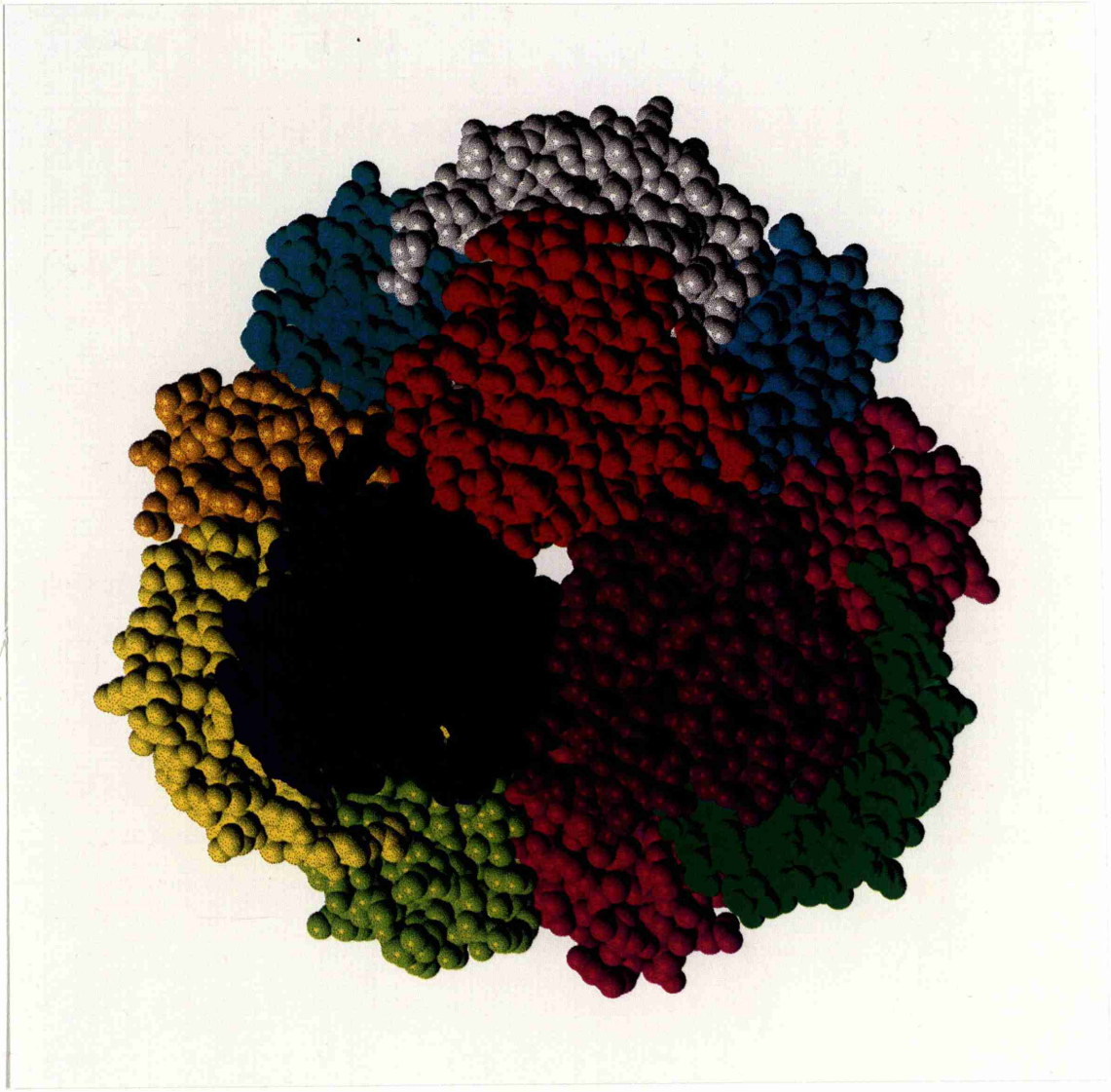


**Figure 8-7f Type II DHQ Dodecamer**

Looking down 3-fold axis and with all atoms shown as Van der Waals spheres. Each monomer is shown in a different colour

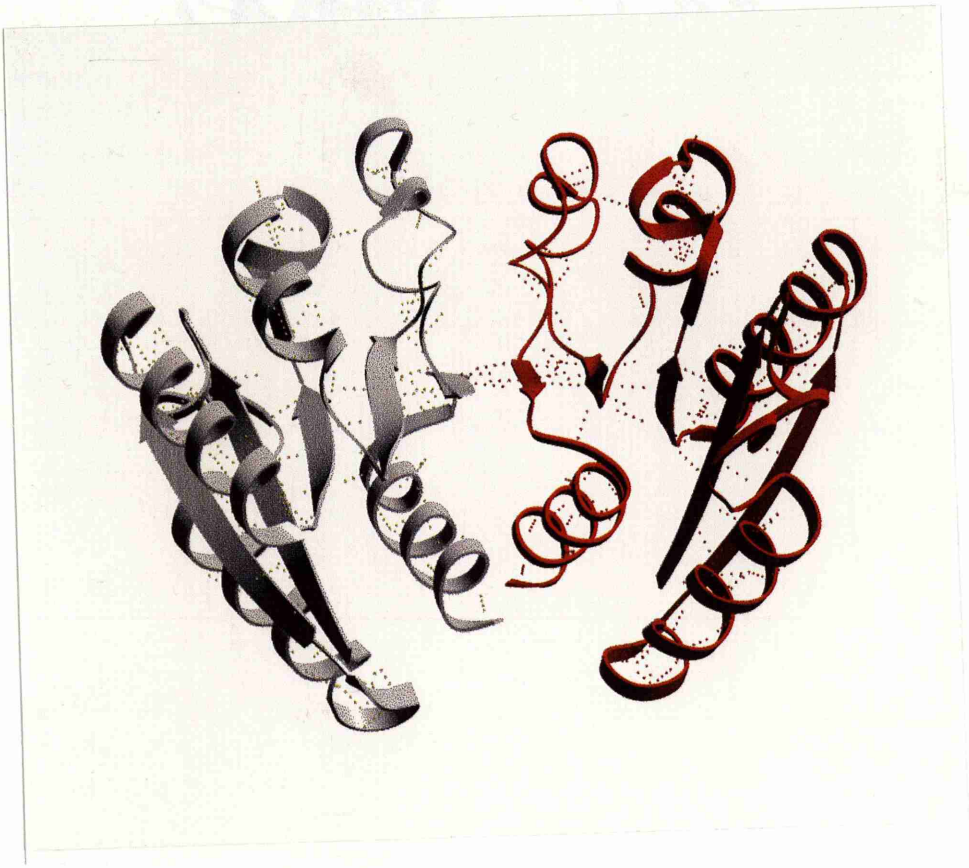


**Figure 8-8 2-fold Interaction in Oligomerisation**

The different monomers are shown as red and white. The parallel β-sheets of the two monomers interact anti-parallel β-sheets giving an extended 10 strand β-Sheet.

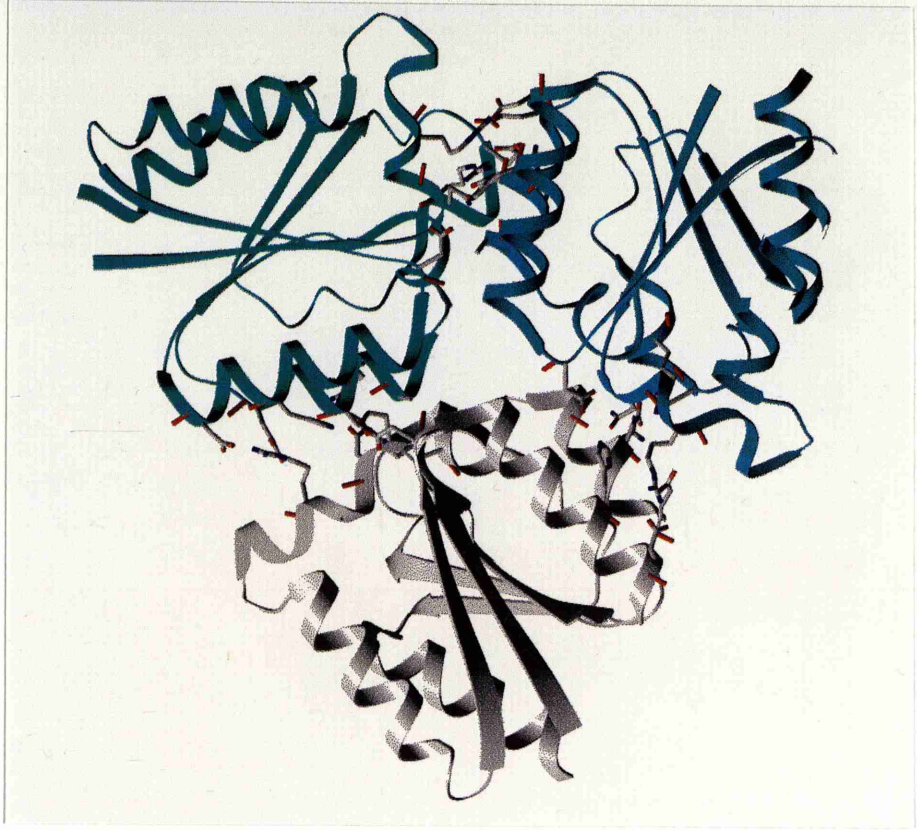


**Figure 8-9 3-fold Interaction in Oligomerisation**

Each monomer is shown in a different colour. Sidechains involved in making this 3-fold interaction are shown.

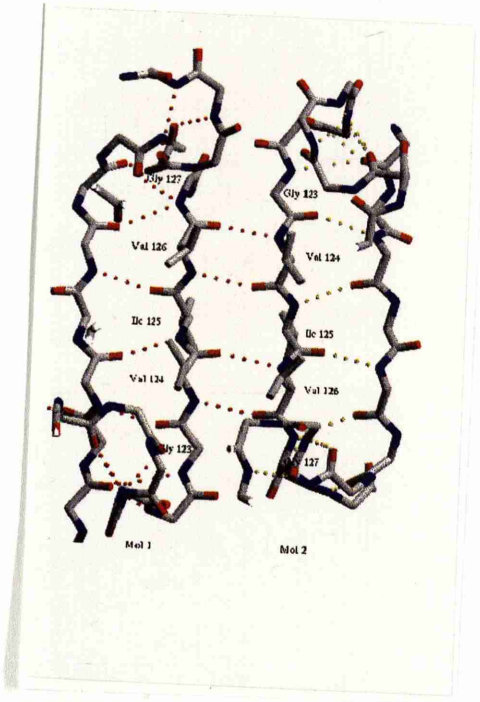



**Figure 8-10 2-fold Interaction with Palindromic Sequence**

The two equivalent strands from each monomer are shown. The hydrogen bonds involved in making this interaction are shown as spheres.

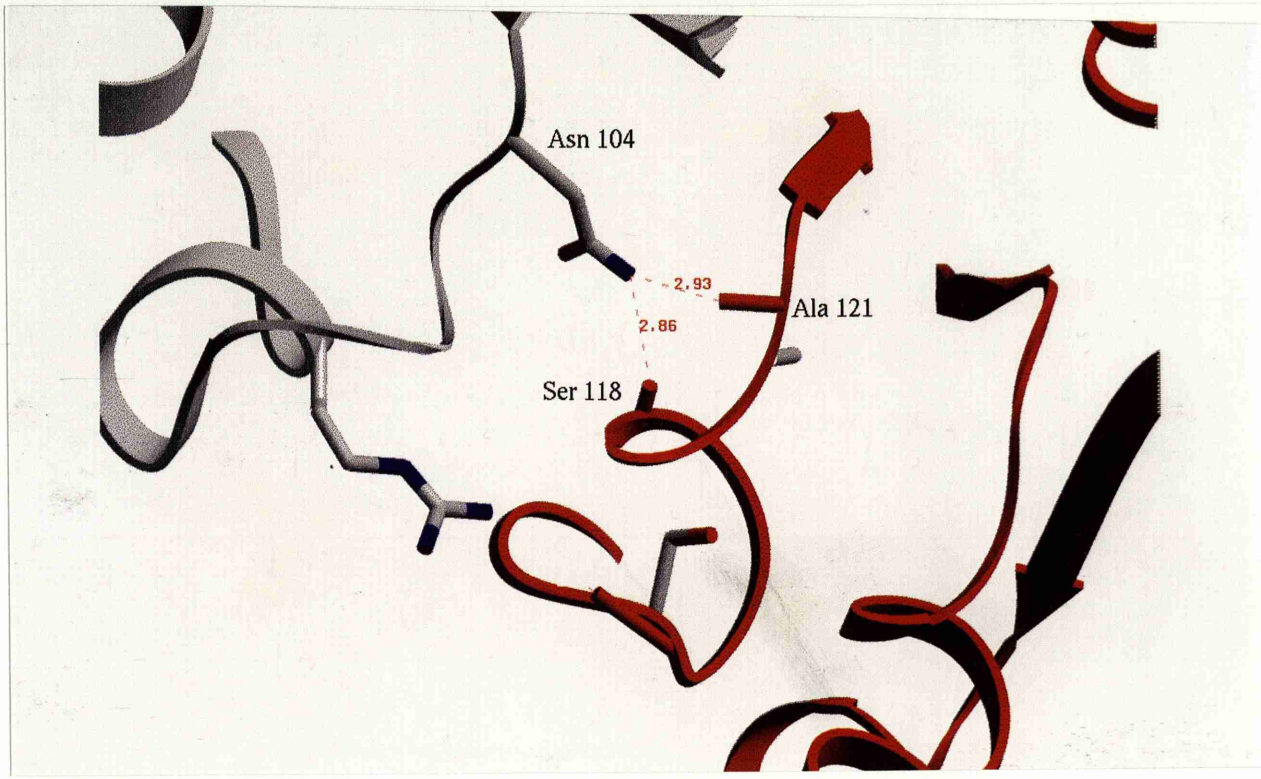



**Figure 8-11 Hydrogen Bonds Between Asn 104 of one Molecule and the Mainchain O of Ser 118 and Ala 121 of Another Molecule. This interaction is important in making the 2-fold contacts.**

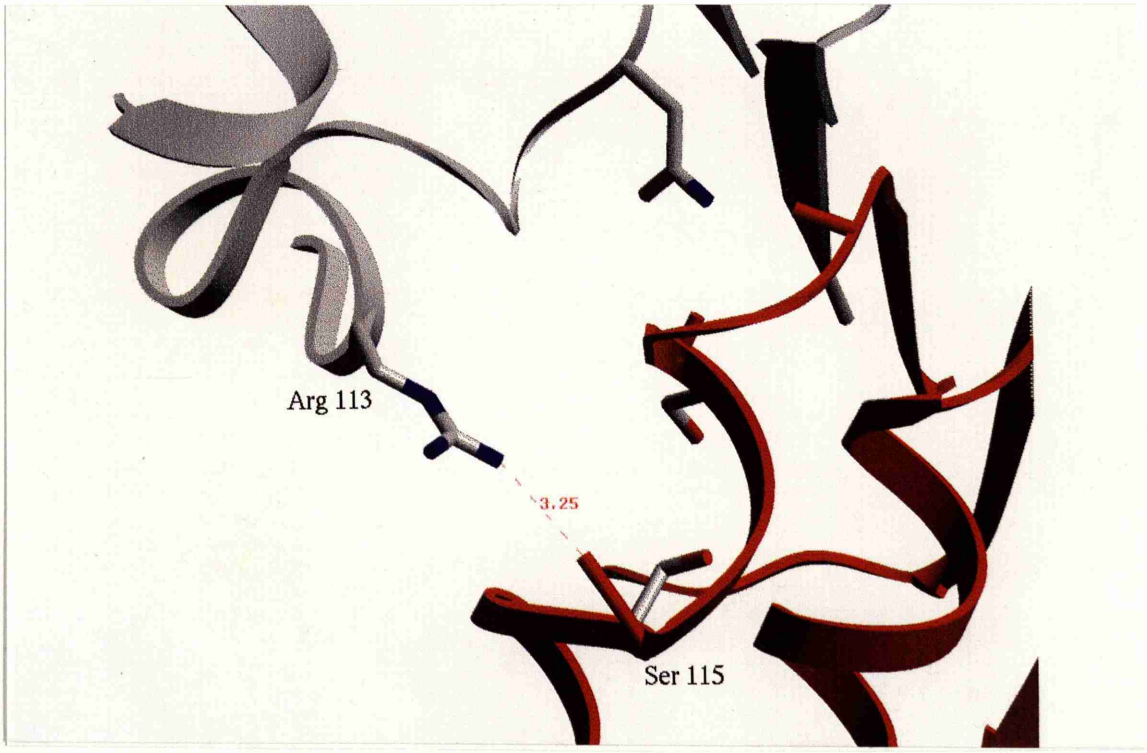



**Figure 8-12 Hydrogen Bonds Between Arg 113 of one Molecule and the Mainchain O of Ser 115 of Another Molecule. This interaction is important in making the 2-fold contacts.**

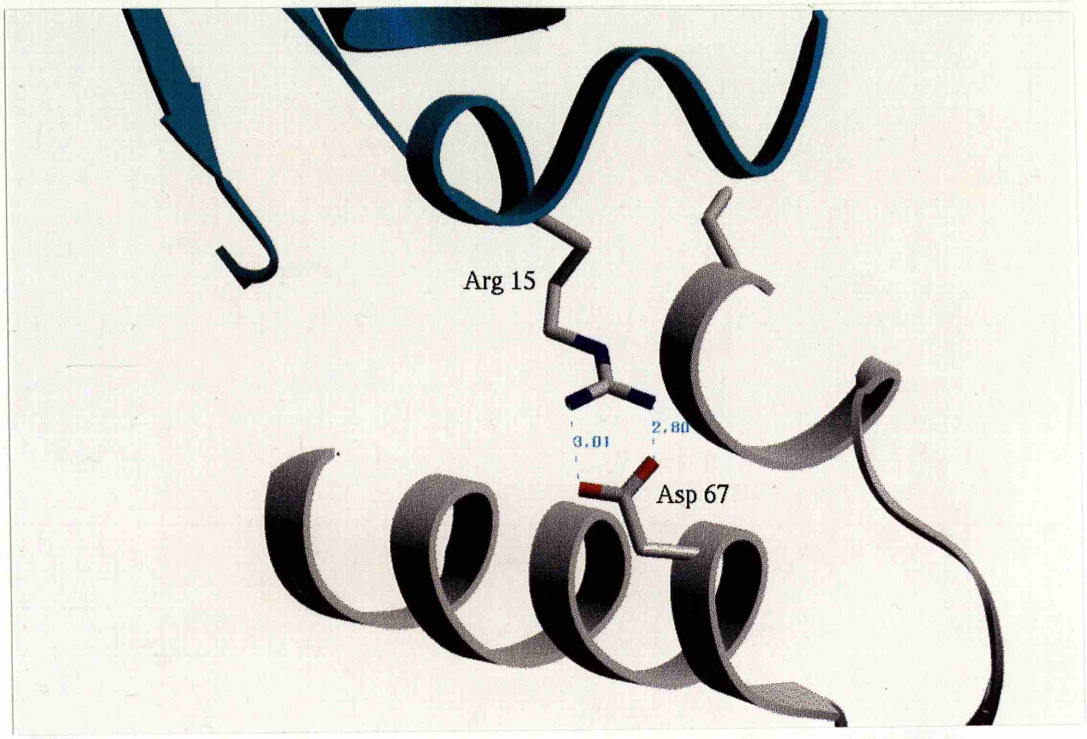



**Figure 8-13 Salt Bridge Between Arg 15 from one Subunit to Asp 67 of Another Subunit. This interaction is important in making the 3-fold contacts.**

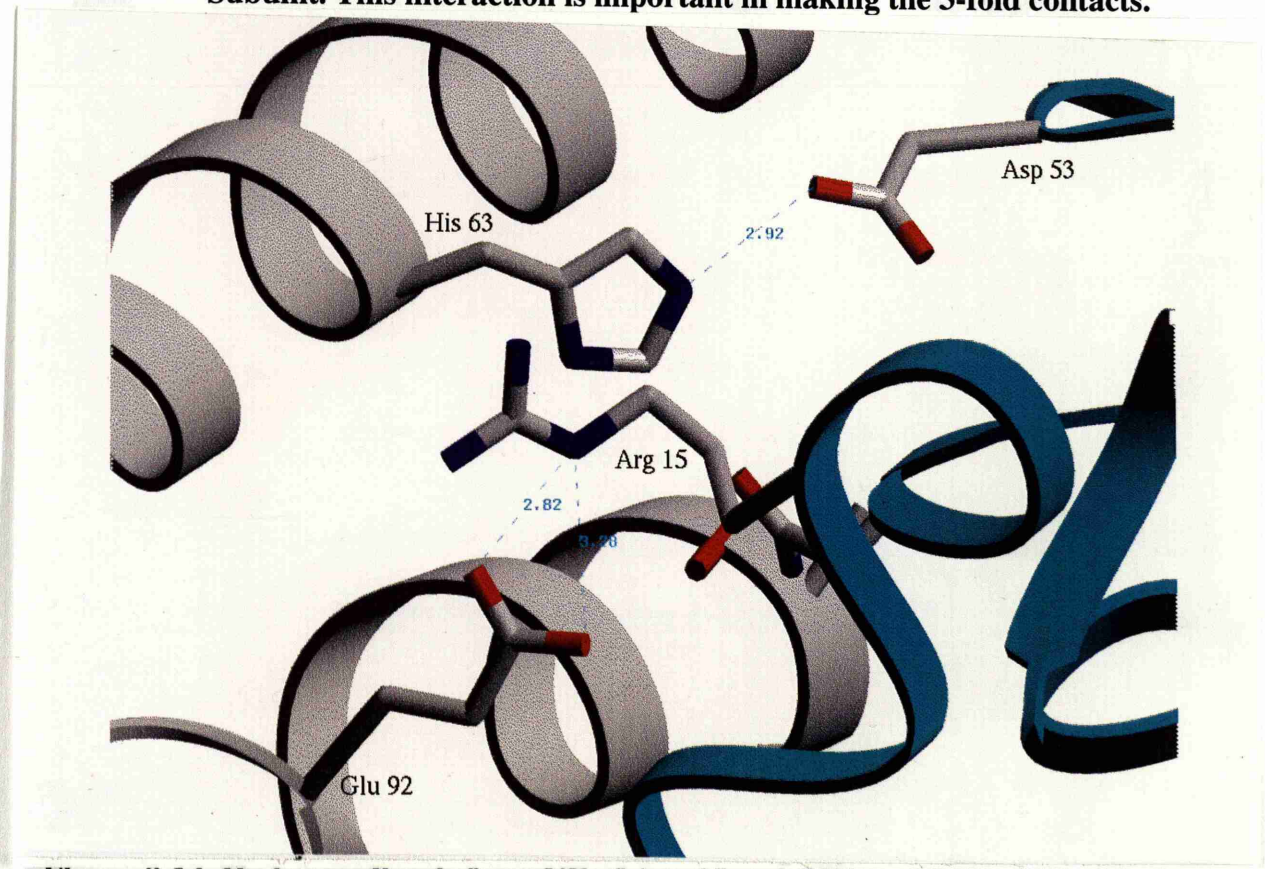



**Figure 8-14 Hydrogen Bonds from NE of Arg 15 and OE1 and OE2 of Glu 92 of the Other Subunit and from Asp 53 OD2 of one Subunit with His 63 NE2 of Another. This interaction is important in making the 3-fold contacts.**

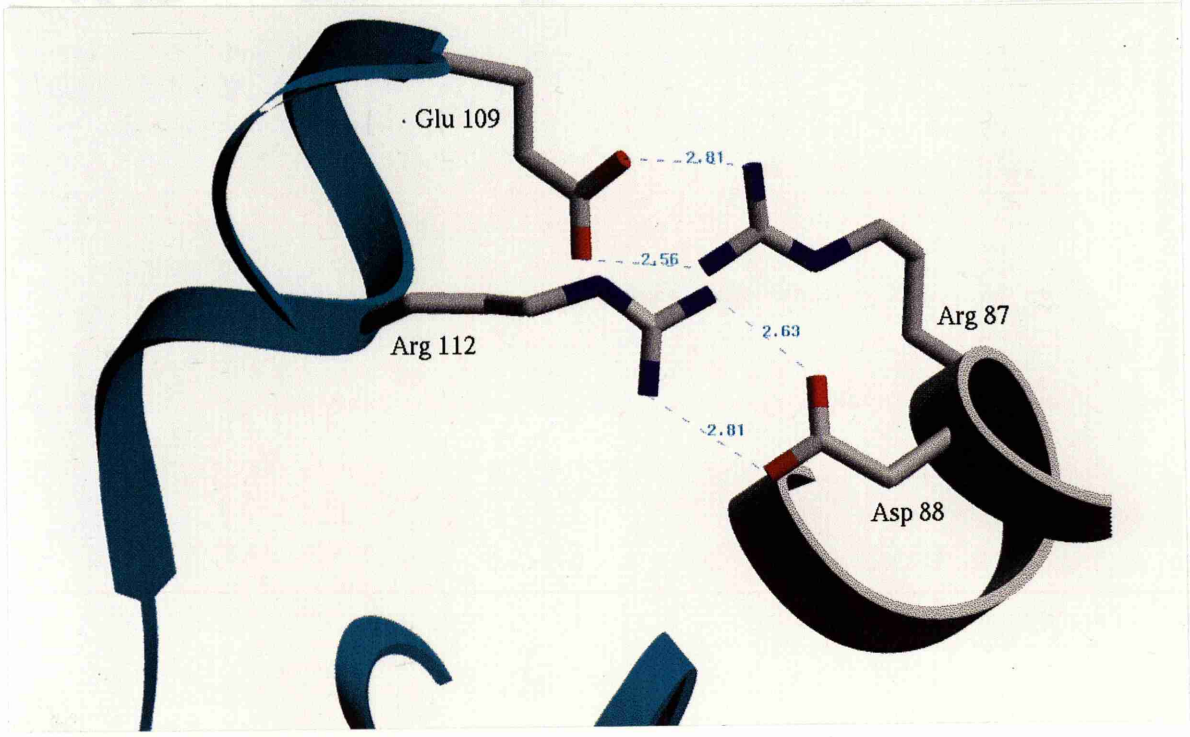



**Figure 8-15 Salt Bridges between Arg 87 from one Subunit to Glu 109 of Another Subunit and between Asp 88 of one subunit and Arg 112 of the Other Subunit. This interaction is important in making the 3-fold contacts.**

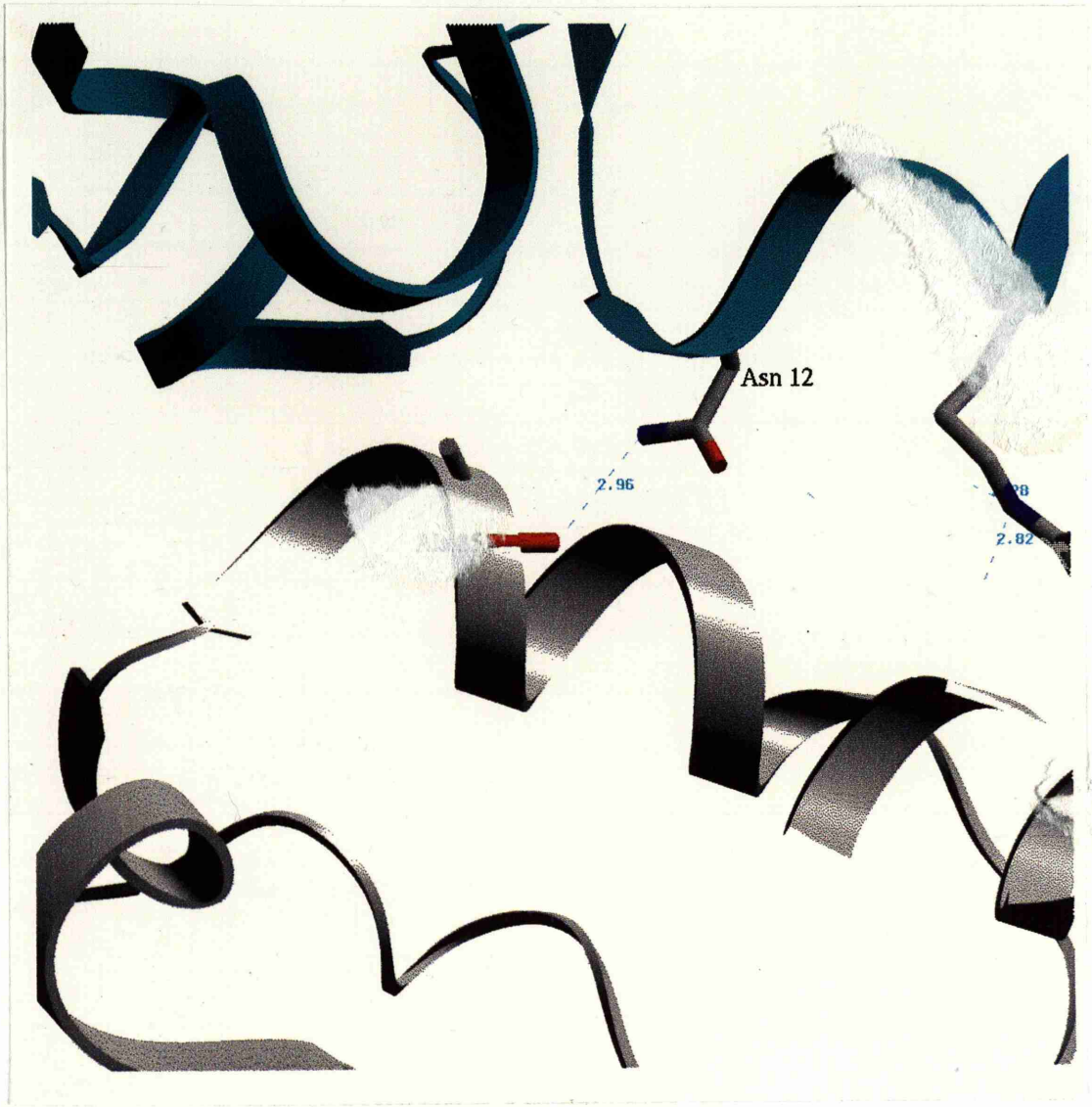



**Figure 8-16 Hydrogen Bonds from Asn 12 of one Subunit to the Mainchain O of Ala 85. This interaction is important in making the 3-fold contacts.**

| Molecule 1 | Molecule 2 | Molecule 3 | Distance (Å) |
|---|---|---|---|
| Arg 15 NH1 | Asp 67 OD1 | | 2.80 |
| Arg 15 NH2 | Asp 67 OD2 | | 3.01 |
| Asp 67 OD1 | | Arg 15 NH1 | 2.80 |
| Asp 67 OD2 | | Arg 15 NH2 | 3.01 |
| Arg 87 NH1 | | Glu 109 OE2 | 2.56 |
| Arg 87 NH2 | | Glu109 OE1 | 2.80 |
| Asp 88 OD1 | | Arg 112 NH2 | 2.63 |
| Asp 88 OD2 | | Arg 112 NH1 | 2.81 |
| Glu 109 OE2 | Arg 87 NH1 | | 2.56 |
| Glu109 OE1 | Arg 87 NH2 | | 2.80 |
| Arg 112 NH2 | Asp 88 OD1 | | 2.63 |
| Arg 112 NH1 | Asp 88 OD2 | | 2.81 |
| | Arg 15 NH1 | Asp 67 OD1 | 2.80 |
| | Arg 15 NH2 | Asp 67 OD2 | 3.01 |
| | Glu 109 OE2 | Arg 87 NH1 | 2.56 |
| | Glu109 OE1 | Arg 87 NH2 | 2.80 |
| | Arg 112 NH2 | Asp 88 OD1 | 2.63 |
| | Arg 112 NH1 | Asp 88 OD2 | 2.81 |

**Table 8-3 Salt Linkages involved in Making the 3-fold contacts**

| Molecule 1 | Molecule 2 | Molecule 3 | Distance (Å) |
|---|---|---|---|
| Asn 12 ND2 | Ala 85 O | | 2.96 |
| Arg 15 NE | Glu 92 OE2 | | 3.28 |
| Asp 53 OD2 | His 63 NE2 | | 2.92 |
| His 63 | | Asp 53 | 2.92 |
| Ala 85 0 | | Asn 12 ND2 | 2.96 |
| Glu 92 OE2 | | Arg 15 NE | 3.28 |
| | Asn 12 ND2 | Ala 85 0 | 2.96 |
| | Arg 15 NE | Glu 92 OE2 | 3.28 |
| | Asp 53 OD2 | His 63 NE2 | 2.92 |

**Table 8-4 H-Bond Interactions Involved in Making the 3-fold contacts**

## 8.3 Location of the Substrate Binding Site

The only biochemical knowledge about the location of the active site was that Arg-19 was thought to be involved in the enzyme mechanism (Krell *et al.*, 1995; 1996: Bottomley *et al.*, 1996a). In this structure this arginine is in the missing loop from residues 17 to 25 that occurs after the first strand and before the first helix. This missing loop had been determined to be highly flexible by biochemical methods (Bottomley *et al.*, 1996a). The flexibility is thought to be involved in a ligand induced conformational change in the enzyme and may be functionally important.

As the structure of type II DHQ was the apo form without substrate bound there was no direct structural evidence for the location of the active site. To try to locate the active site an analysis was made of the location in the structure of the residues that are conserved amongst all the species in which DHQ type II enzymes are found. It was observed that the majority of these conserved residues were clustered in the same part of the structure, as illustrated in Fig 8.17.

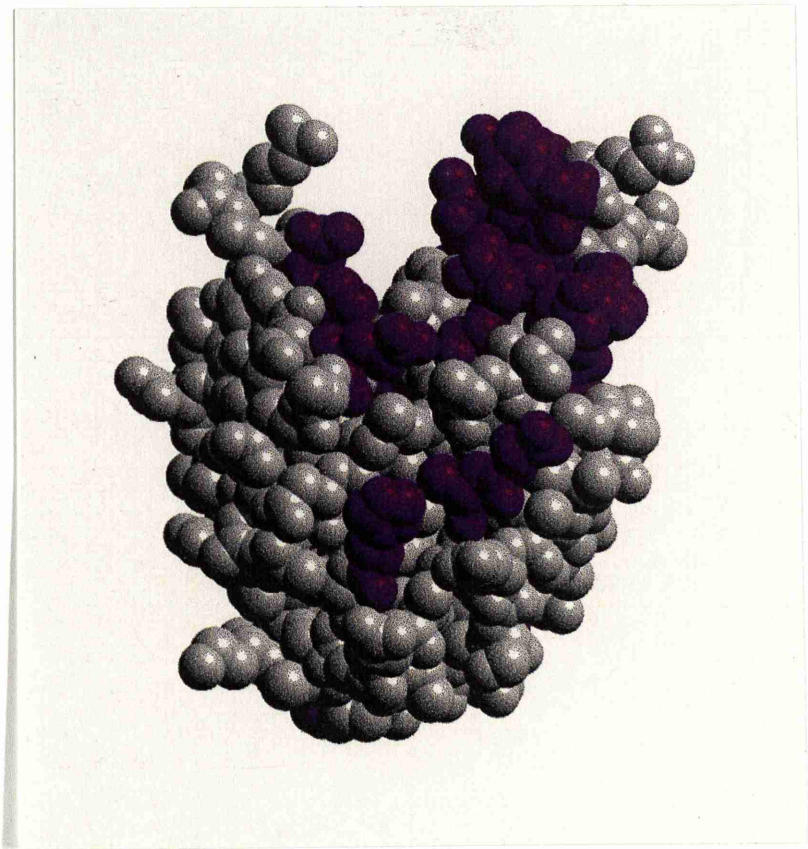


**Figure 8-17 The Structure of Type II DHQ with Van Der Waals Representation showing the location of residues totally conserved in sequence alignments of type II DHQ.**

White residues non conserved residues.

Purple residues totally conserved residues.

There was a cleft in this area that contained several residues functionally suitable for substrate binding and for catalysis of the dehydration reaction. The

residues Asn 75, Thr 80, His 81, Glu 99, His 101 and Arg 108 were all lining the cleft and had sidechains pointing into it. An attempt was made to model the substrate into this cleft (Fig 8.18). This modelling in of substrate was done without altering the position of any atoms in the protein structure.

This model has the carboxylate group of dehydroquinic acid binding to Arg-108. The C-5 hydroxyl group of the substrate has been shown by substrate analogue experiments to make an important interaction with the enzyme. The removal of this interaction results in a decrease in $k_{cat}/K_m$ of the order of $10^5$ (J. Harris 1994; Harris *et al.,* 1996). In the model the C-5 hydroxyl makes a hydrogen bond with Glu-99. The C-4 hydroxyl group of substrate is in a reasonable position to H-bond to Thr-80.

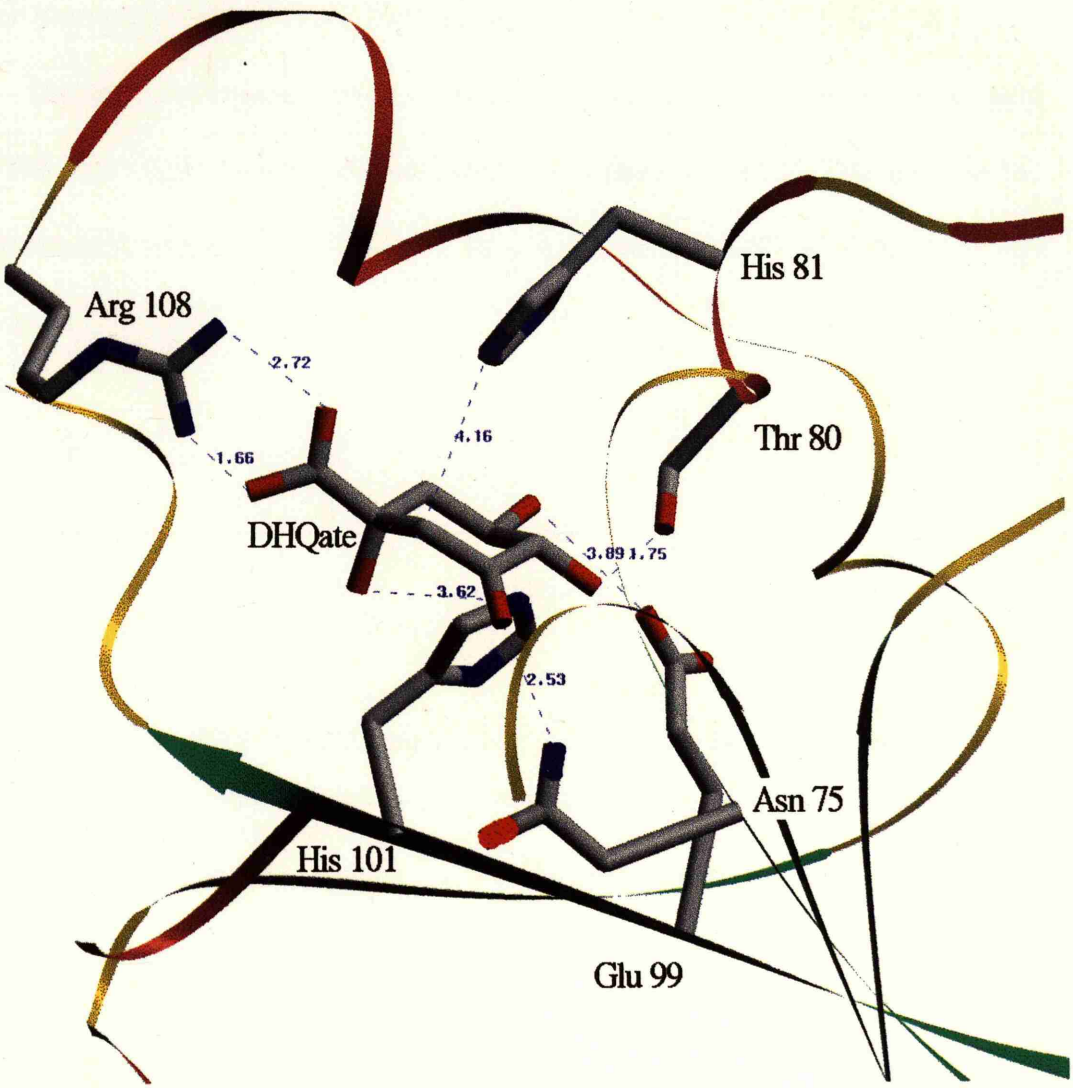



**Figure 8-18 Proposed Binding Of Substrate To Type II DHQ**

Stick diagram showing the proposed interaction of substrate with acive site residues. Only residues involved interaction with substrate are shown. Secondary structural elements are shown in the same way as Fig. 8.1.

## 8.4 A Putative Mechanism For The Type II Dehydroquinase

The trans-elimination of water from dehydroquinate to give dehydroshikimate as catalysed by the Type II dehydroquinase is a 1,2-(β) elimination. Three different mechanisms can be envisaged for dehydration by 1,2-elimination, differing from each other in timing of H-C and HO-C bond breaking.

The dehydration could be a concerted reaction, going via a one step transition state. This mechanism is referred to as an E2 mechanism (Fig 8.19). This involves the simultaneous removal of the hydrogen by a base and elimination of the hydroxyl group by acid catalysis.

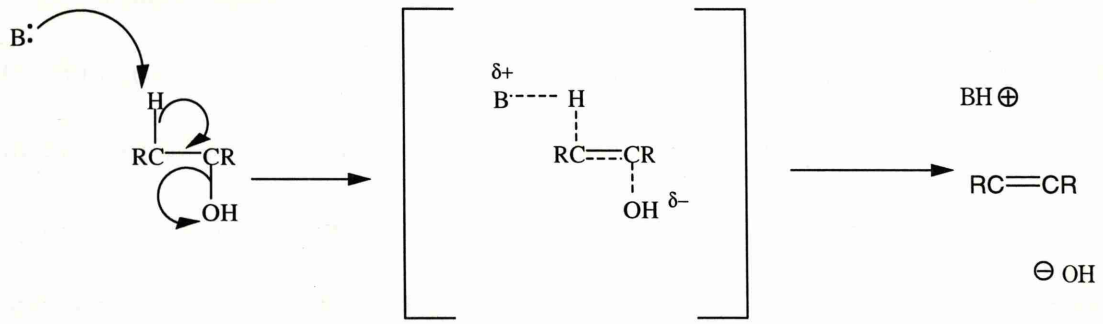



**Figure 8-19 The E2 Mechanism For Elimination of Water**

Alternatively, the H-C and C-OH bonds can be broken separately in a two step process. If the C-OH bond is broken first a carbocationic intermediate is involved in a reaction that is referred to an E1 mechanism. Which is shown in Fig 8.20.

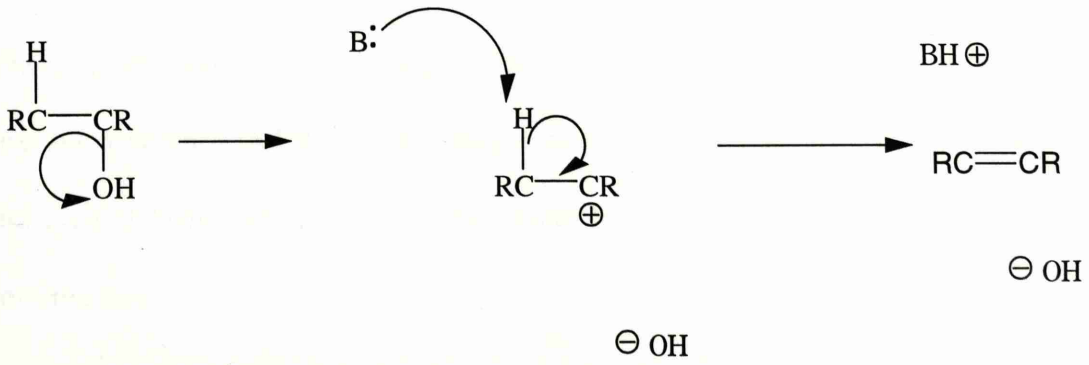



**Figure 8-20 The E1 Mechanism for Elimination of Water**

The other possible mechanism is that the H-C bond could be broken first, involving a carbanion intermediate. This is referred to as the E1cB mechanism (Fig 8.21). In the case of the dehydroquinase reaction the carbanion intermediate could be an enol or an enolate.

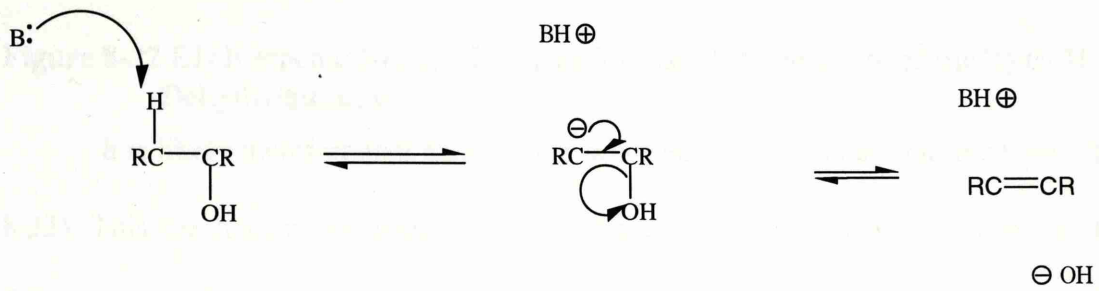



**Figure 8-21 The E1cB Mechanism for the Elimination of Water**

Kinetic studies done by the group of Chris Abell in the University of Cambridge Chemical Laboratory have identified which elimination mechanism occurs in the type II dehydroquinases (Harris *et al.*, 1996). It has been found by kinetic isotope studies that proton abstraction in the type II DHQ is partially rate determining and occurs at or before the highest transition state. This is inconsistent with the idea of the dehydration occurring via an E1 mechanism. There is also an absence of a solvent isotope effect on $V_{max}/K_m$ for the *M. Tuberculosis* DHQ. This is evidence against a concerted E2 mechanism and against formation of an enol intermediate. Both of these mechanisms would be expected to involve a solvent derived proton. There is also an increase in $k_{cat}$ with pH that is not expected for a mechanism involving an enol. All the kinetic data is however consistant with the dehydration in the type II DHQ occuring via a step wise E1cB mechanism involving an enolate intermediate.

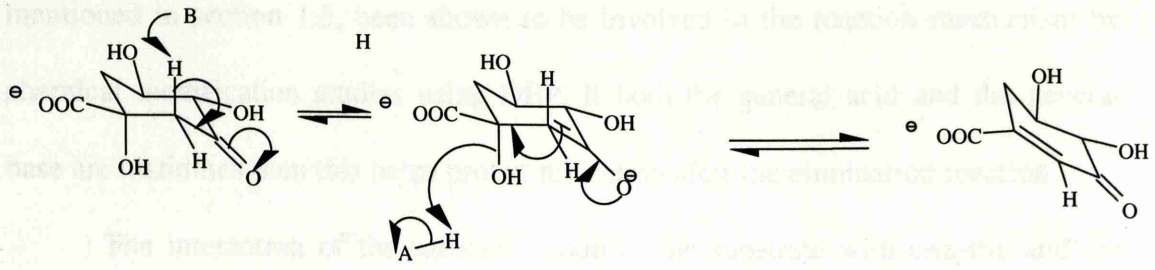



**Figure 8-22 E1cB mechanism Involving an Enolate Intermediate of the Type II Dehydroquinase**

It is likely therefore that dehydration proceeds via an enolate intermediate (Fig 8.22). This mechanism involves an initial base-catalysed abstraction of a proton to give a enolate intermediate. The hydroxyl can then be readily eliminated by acid catalysed protonation followed by loss of water.

The cleft of the active site proposed in section 8.3 was analysed to identify possible candidates for residues functionally important in the mechanism. This would involve the identification of the acid and base, as well as the location of the residues important for substrate recognition and binding. The method the enzyme uses to stabilise the enolate intermediate was also considered. The positions of the residues in this cleft were analysed and a model of the substrate was fitted into likely positions, that satisfied the limited biochemical knowledge and made chemical sense with regard to hydrogen bonding, charge pairing and sterochemical constraints. The fit of the modelled substrate is not perfect but this is because some degree of conformational change in the enzyme is expected on binding substrate.

The most convincing positioning of substrate is illustrated in Fig 8.18. This gives His 81 as the most likely candidate for the base involved in the initial step of proton removal. His-101 is then suitably positioned to act as the acid to donate a proton to the hydroxyl so it could leave as a water molecule. Histidines have, as

mentioned in section 1.3, been shown to be involved in the reaction mechanism by chemical modification studies using DEP. If both the general acid and the general base are histidines then this helps proton relocation after the elimination reaction.

The interaction of the carbonyl group of the substrate with enzyme and the method of stabilisation of the enolate intermediate is still unclear. In other enzyme systems stabilisation of enolate intermediates are done in various ways that are discussed below.

Fructose-1,6-bisphosphate adolases catalyse the reversible aldol condensation of dihydroxyacetone phosphate and glyceraldhyde 3-phosphate to fructose-1,6-bisphosphate (Fig.8.23). Two classes of aldolase exist with distinct differences in their molecular and catalytic properties.

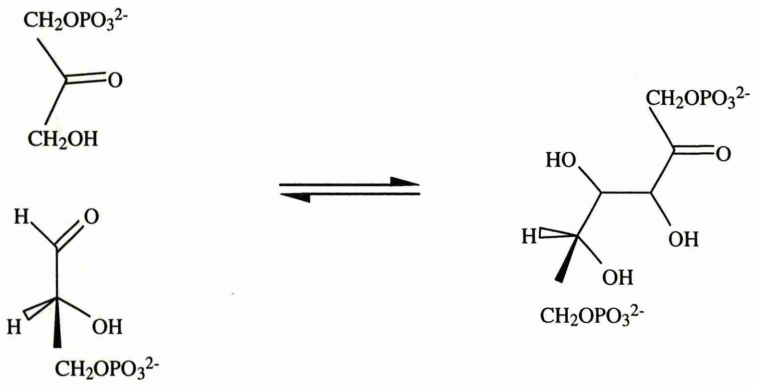



**Figure 8-23 The Reversible Aldol Condensation of Dihydroxyacetone Phosphate and Gyceraldhyde 3-phosphate to Fructose-1,6-Bisphosphate by Aldolase**

The mechanism of aldol condensation in the class I aldolases is similar to the mechanism of action of the type I dehydroquinase and proceeds via a Schiff base mechanism (Lai *et al.*,1972). The mechanism of the class II aldolase catalysed reaction involves initial formation of an enzyme substrate complex with dihydroxyacetone phosphate. It is believed that the metal ion forms a partial bond with the carbonyl group of the substrate. This acts as an electron sink to stabilise the

enolate intermediate and facilitate the nucleophilic attack on the aldehyde substrate (see Fig 8.24).

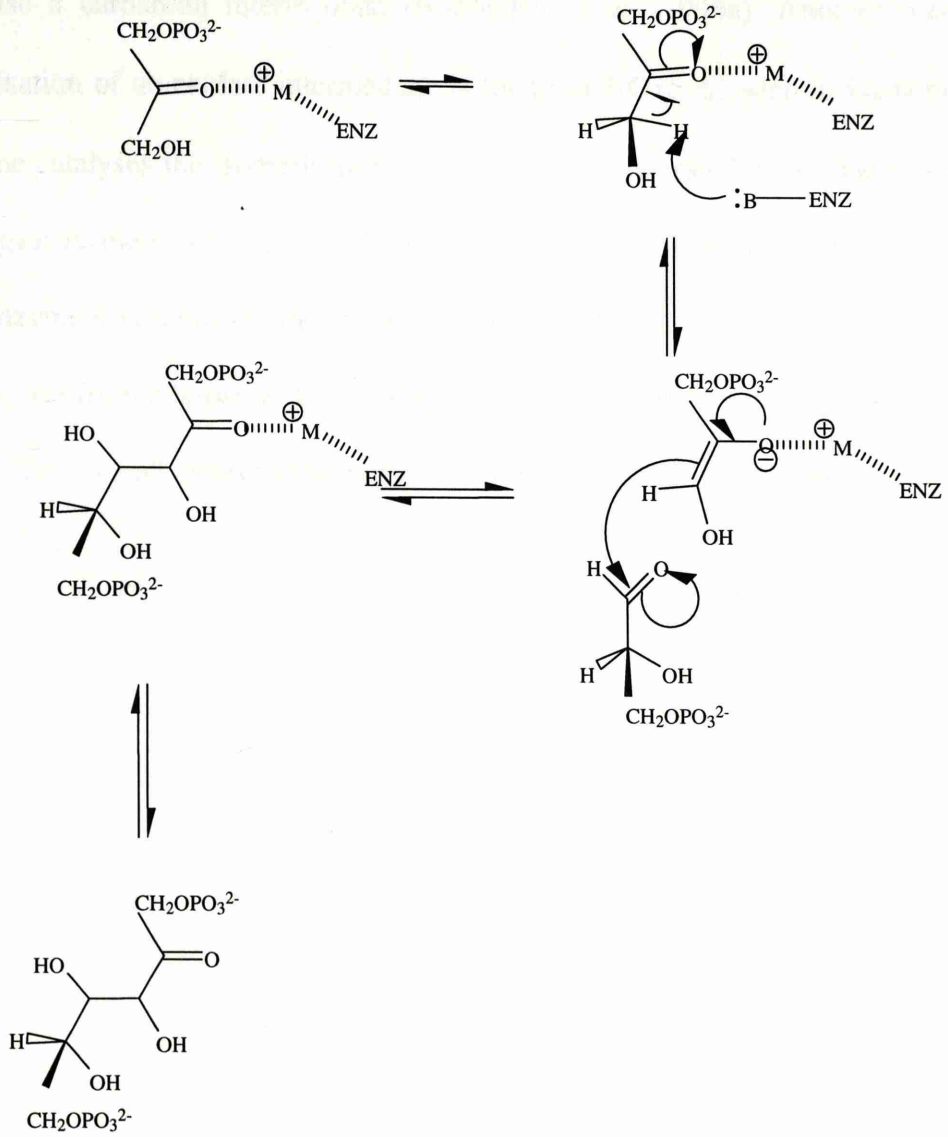



**Figure 8-24 The Mechanism Of Class II Aldolase Involving a Metal Ion to Stabilise the Enolate Intermediate**

M is the metal ion

No metal has been found in the DHQ structure and there is biochemical evidence indicating the type II dehydroquinase does not use a divalent metal ion to stabilise a carbanion intermediate (Bottomley *et al.,* 1996a). Another method of stabilisation of an enolate intermediate is found in 3-Oxo-$\Delta^5$ -steroid isomerase, this enzyme catalyses the isomerisation of various β,γ-unsaturated 3-oxo-steriods to their conjugate isomers at nearly a diffusion controlled rate (Hawkinson *et al.,* 1991). For this enzyme it is proposed that a dienolate intermediate is stabilised by two hydrogen bonds, one from a tyrosine -OH and another from an aspartic acid -COOH (Wu et al., 1997). The overall catalytic mechanism for this reaction is shown in Fig 8.25.

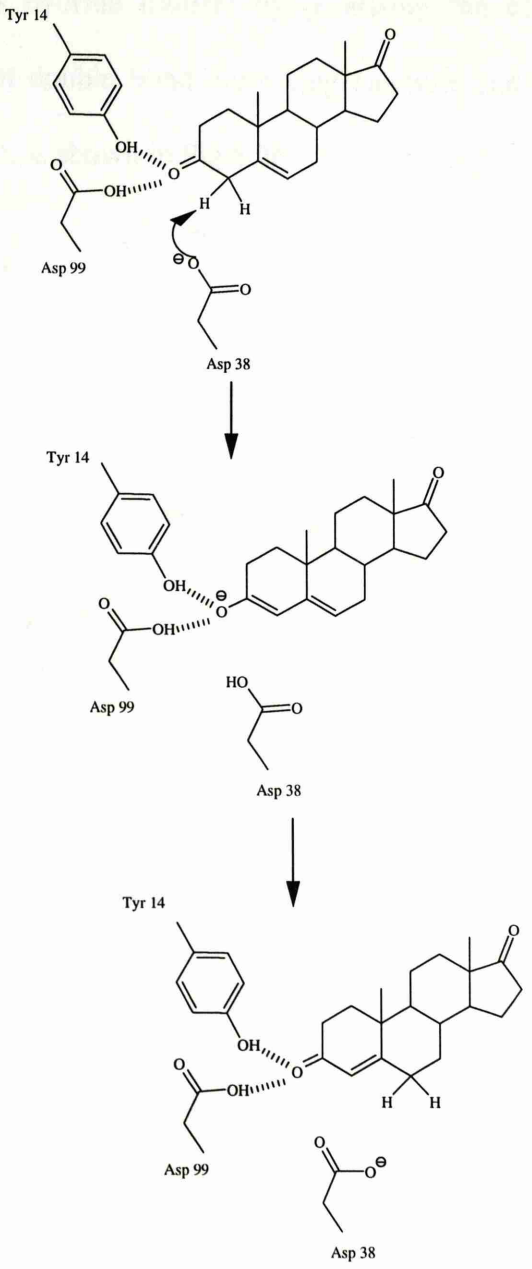



**Figure 8-25 The Isomerisation of β,γ-Unsaturated 3-Oxo-Steriods to their Conjugate Isomers By 3-Oxo-$\Delta^5$ -Steroid Isomerase**

This illustrates the use of hydrogen bonds to stabilise an enolate intermediate.

A similar mechanism of stabilisation is found in the *M. tuberculosis* 2-trans-enoyl acyl carrier protein reductase (Quemard *et al.,* 1995, Dessen *et al.,* 1995). This enzymatic reaction involves a hydrogen bond forming between the carbonyl bond of

the fatty acid of the 2-trans-enoyl acyl carrier protein and an active site lysine. This hydrogen bond facilitates hydride transfer by polarising the carbonyl group and conjugated carbon-carbon double bond promoting catalysis and stabilising enolate formation. This mechanism is shown in Fig 8.26.

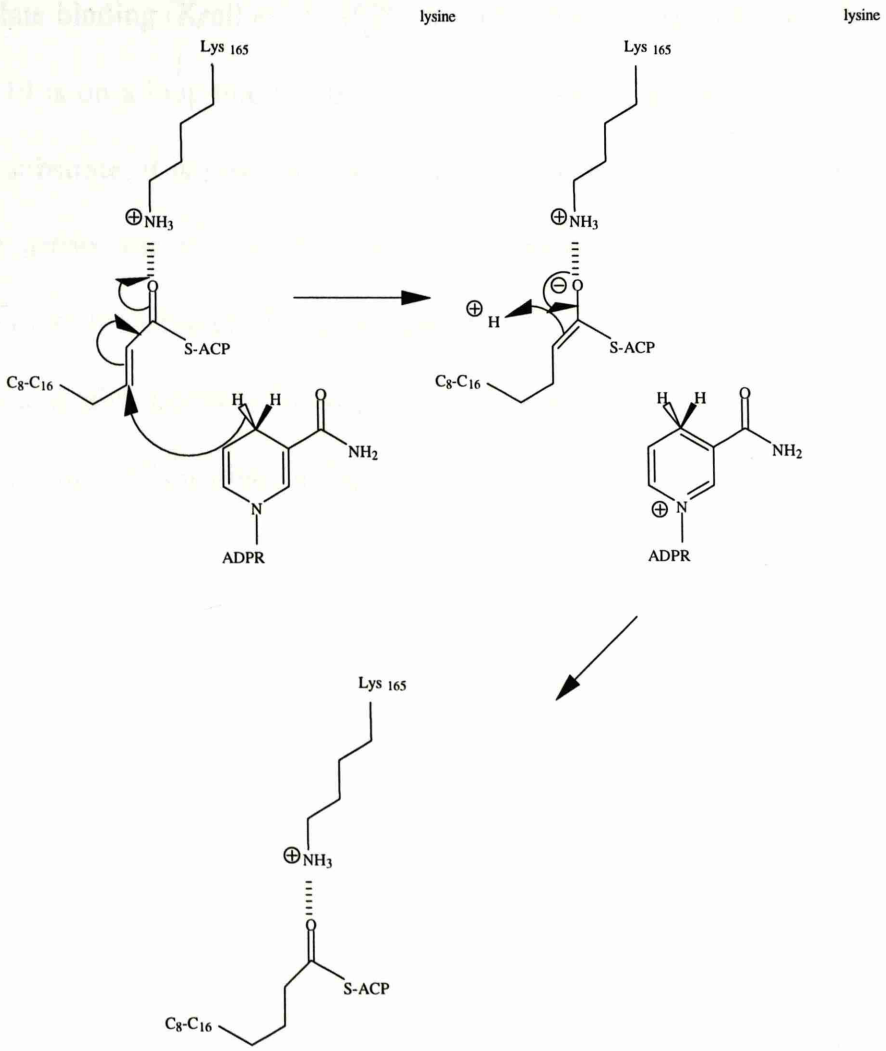



**Figure 8-26 The Reaction Mechanism of M. tuberculosis 2-Trans-Enoyl Acyl Carrier Protein Reductase**

ACP = Acyl Carrier Protein

As discussed in the previous section the normal methods of activating a carbonyl group to allow it to form an enolate are: the use of bivalent metal such as $Zn^{2+}$, the use of positively charge residue such as arginine or the stabilisation via hydrogen bonds. There is however no bivalent metal or positively charged residue

suitably located in the active site of this apo-enzyme structure. As mentioned in section 1.3 Arg 19 has been proven by the chemical modification and site directed mutagenesis to have an important functional role in the enzyme mechanism of type II DHQ. Site directed mutagenesis work indicates this Arg is not involved in carboxylate binding (Krell *et al.,* 1996). Another role for Arg 19 has to be identified. As Arg 19 is on a loop that has been proven to undergo a conformational change on binding substrate, it is possible that when the enzyme binds substrate Arg 19 moves into the active site to stabilise the enolate intermediate. The missing loop was modelled into the structure using the lego-loop function in O. This modelling proves it is geometrically possible for Arg 19 to be in a suitable position to perform this role and a diagram of this is given in Fig 8.27.

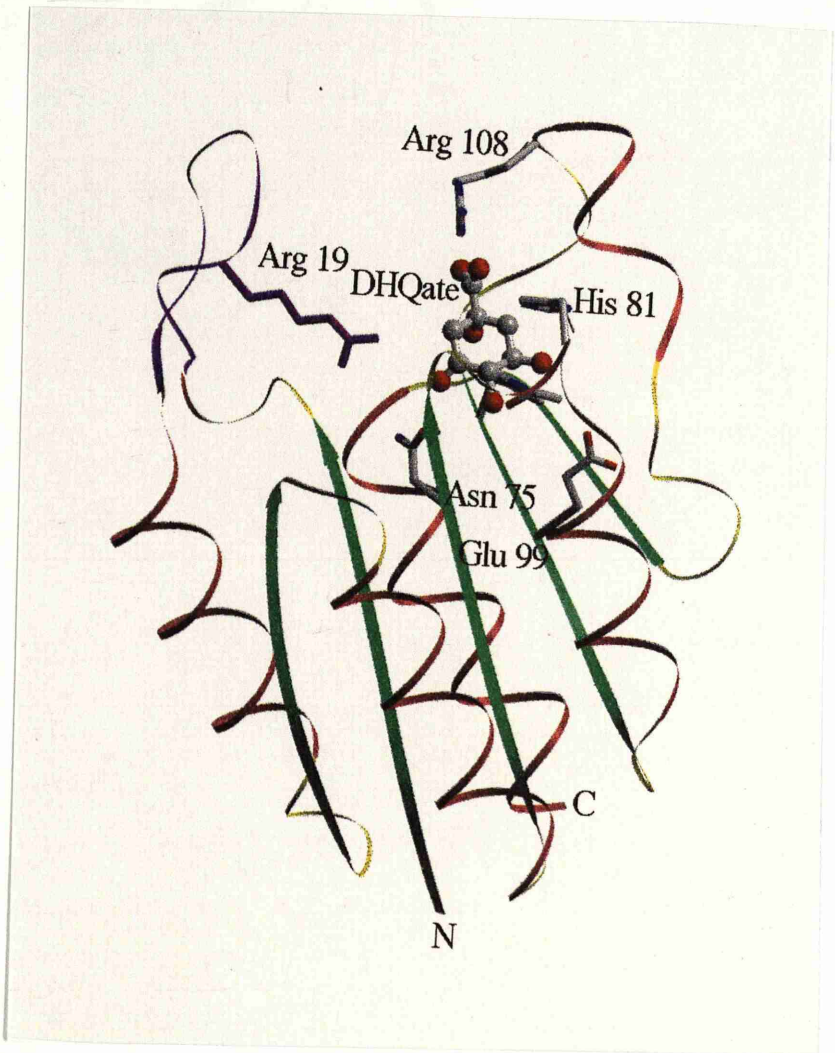



**Figure 8-27a Modelling of Missing Loop from DHQ**

Fig 8.27a has the monomer in the same orientation as Fig. 8.1. The missing loop and Arg 19 are shown in purple.

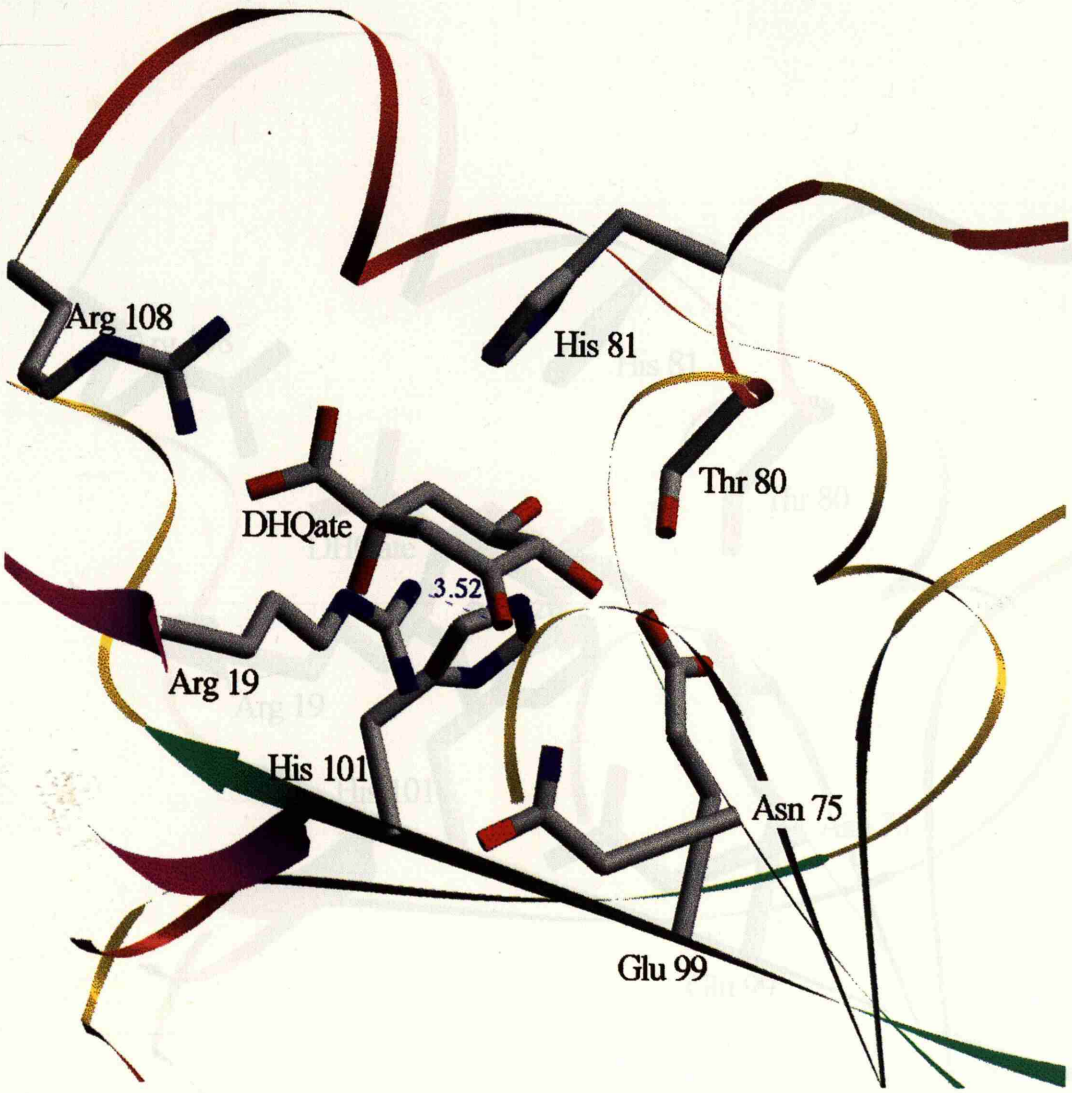



**Figure 8-27b Modelling of Missing Loop from DHQ**

Fig 8.27b shows a close up of the active site in the same orientation as Fig. 8.18 with the modelled in missing loop coloured purple and the atoms of Arg 19 shown in standard colours.

If Arg 19 does not perform this stabilisation role there are other possible ways the enzyme could stabilise the enolate. There could be a hydrogen bond interaction with the backbone at Gly-78 and Gly-79 or there could be a hydrogen bond with Asn-75. The putative enolate mechanism for the type II dehydroquinase is given in Fig 8.28. The residue involved in stabilisation of the enolate is given a label X though as mentioned above it is probably Arg 19. This mechanism is only a hypothesis based on the structure and taking into account the known biochemical information. In the next section I will suggest further experiments to test this hypothesis.

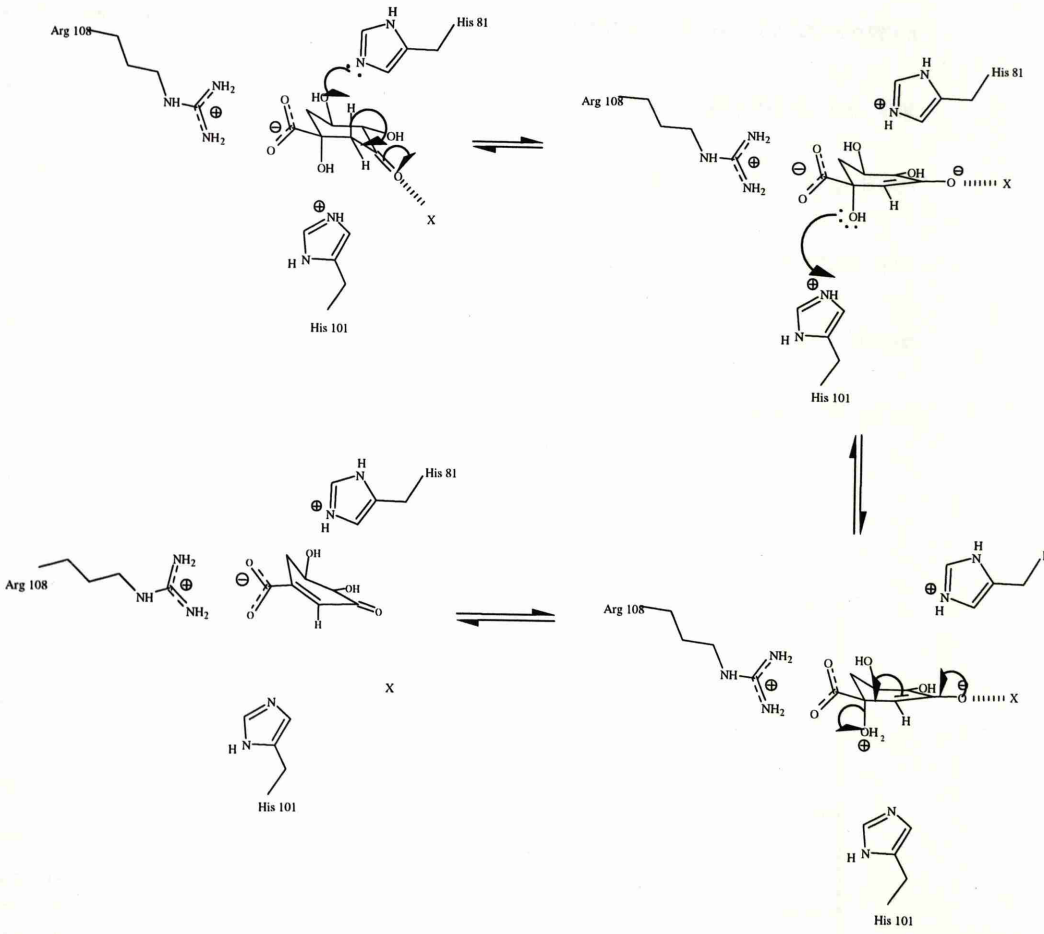



**Figure 8-28 The Proposed Mechanism of Type II DHQ**

Where X is the residue involved in enolate stabilisation.

## 8.5 Further work:

Direct experimental evidence of binding of dehydroquinate is required and the following experiments should be done:

- A crystal of Type II dehydroquinase from *Mycobacterium tuberculosis* should be soaked in substrate and then frozen using cry-cooling flash-freezing techniques to try to trap either product or substrate in the active site. It would be hoped that data collected on a frozen soaked crystal would produce direct evidence for the location and method of substrate binding.

- Crystals of a mutant type II Dehydroquinase from *Streptomyces coelicolor* have been grown (see Fig 8.29). This mutant has the arginine on the missing flexible loop mutated to an alanine. These mutants have a lower $k_{cat}$ than either the *S. coelicolor* native enzyme or the *M. tuberculosis* native enzyme see (Table. 8.15). This fact might make it easier for substrate to be trapped into these crystals by freezing. Data collection on these crystals soaked in substrate would give information about the precise position of the dehydroquinate.

| | $k_{cat}$ [$s^{-1}$] | $K_m$ [μM] | $K_{cat}/K_m$ [$s^{-1}M^{-1}$] |
|---|---|---|---|
| wildtype *S. coelicor* | 960 | 1100 | $8.7 \times 10^5$ |
| wildtype *M. tuberculosis* | 4 | 16 | $4.4 \times 10^5$ |
| Arginine to Alanine Mutant | 0.032 | 170 | 188 |

**Table 8-5 Kinetic Data For DHQ From M. tuberculosis, S. coelicolor and S. coelicolor mutant.**

- Site directed mutagenesis of the proposed acid His-101 and the proposed base His-88 will allow their roles in catalysis to be evaluated. If the hypothesis about

their roles in catalysis is correct mutating either of them to a phenylalanine would result in a decrease in $k_{cat}$. There might also be an effect on $K_m$ because of the introduction of a more hydrophobic residue into the active site.

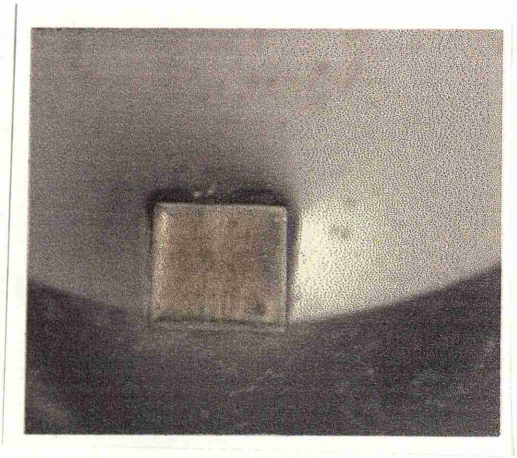


**Figure 8-29 Crystal Of mutant of S. coelicor DHQ**

Crystal size 1.0 x 0.4 x 0.3 mm

- Glu-99 could be mutated to an Ala as this residue is proposed to be involved in substrate binding. It would be interesting to see if this mutation had the same effect on $k_{cat}/K_m$ as having a substrate analogue with no hydroxyl group on the

C-5 carbon. If the model for substrate binding is correct this mutation would eliminate the same substrate enzyme interaction as happens with the substrate analogue.

- Mutation of the Thr-81 to eliminate the proposed H-bond between it and the hydroxyl group on C-4 would also test whether this residue is involved in substrate binding. If the hypothesis is correct, one would expect an increase in $K_m$ and possibly a decrease in $k_{cat}/K_m$

- Mutating Arg108 to a neutral residue like glutamine and alanine would be expected to result in a greatly increased $K_m$.

- It would also be interesting to mutate Asn-75 to a leucine to see what effect if any it had on catalysis and enzyme binding. For if this residue is involved in forming hydrogen bonds elimination of its potential to do this by mutating it to a leucine would cause a decrease in $k_{cat}$ and an increase in $K_m$.

- There are certain biochemical differences between the *M. tuberculosis* and *S. coelicolor* dehydroquinases. In order to explain these differences the crystallisation of native *S. coelicolor* enzyme is being attempted and crystals have been grown see Fig 8.13. These crystals do not yet give useful data but efforts are being made to improve their quality so that a full structure determination may be attempted.

# 9. References


Abrahms, J.P. and Leslie, A.G.W. (1996) Methods used in the stucture determination of bovine mitrochondrial $F_1$ ATPase Acta Cryst. **D52,** 30-42.

Agarwal, R.C. (1978) A new least squares refinement technique based on the fast fourier transform alogrithim. Acta Cryst. **A34** 791-809.

Agarwal, R.C., Lifchitz, A. and Dodson, E. (1980) in the Refinement of Protein structures . Proceedings of Daresbury study weekend. 36-39.

Anton, I.A., Duncan, K. and Coggins, J.R. (1987) A eukaryotic repressor protein, the *qa-1s* product of *Neurospora crassa*, is homologous to part of the *arom* multifunctional enzyme. J. Mol. Biol. **197**, 367-371.

Barnes, P.F., Bloch, A.B., Davidson, P.T. and Snider, D. (1991) Tuberculosis in patients with human immunidefiency virus infection. N. Engl. J. Med. **324**, 1644-50.

Bentley, R. (1990). The shikimate pathway - a metabolic tree with branches. Critical Reviews in Biochemistry and Molecular biology **25,** 307-384.

Beri, R.K., Whittington, H., Roberts, C.F. and Hawkins, A.R.(1987) Isolation and charactierization of the positively acting regulatory gene qutA from Aspergillus nidulans. Nucleic Acids Res. **19**, 7991-8001.

Bloom, B.R. and Murray, C.J. (1992) Tuberculosis: commentary on a re-emergent killer. Science **257**, 1055-64.

Blow, D.M, and Crick, F.H.C. (1959) The treatment of errors in the Isomorphous Replacement Method Acta Cryst. **12**, 794-802

Blundell, T.L. and Johnson, L.M. (1976). Protein Crystallography. Academic Press, London.

Boocock, M.R. and Coggins, J.R. (1983) Kinetics of 5-enolpyruvylshikimate 3-phosphate synthase by glyphosate. FEBS Letters **154**, 127-133.

Bottomley, J.R., Clayton, C.L., Chalk, P.A. and Kleanthous, C. Cloning, sequencing, expression, purification and preliminary characterisation of a type II dehydrogenase from Helicobacter pylori. (1996b) Biochem. J. 318, in press.

Bottomley, J.R., Hawkins, A.R. and Kleanthous, C. (1996a) Conformational changes and the role of metals in the mechanism of type II dehydroquinase from Asergillus nidulans. Biochem. J. **319**, 269-278.

Boudet, A.M., Graziana, A. and Ranjeva, R. (1985) In "Annual Proceedings of the Phytochemical Society of Europe" (C.F. Van Sumere and P.J. Lea, eds), Vol.25, pp. 135-159. Clarendon Press, Oxford.

Boys, C.W.G., Fawcett, S.M., Sawyer, L., Moore, J.D., Charles, I.G., Hawkins, A.R., Deka, R., Kleanthous, C. and Coggins, J.R. (1992) The crystallisation of a Type I 3-dehydroquinase from *Salmonella typhi.* J. Mol. Biol. **227**, 352-355.

Burnett, R.M., Darling, G.D., Kenall, D.S., Le Quesene, M.E., Mayhew, S.G., Smith, W.W., Ludwig, M. (1974) J.Biol Chem. **249**, 4383-4392.

Bricogne, G. and Irwin, J., (1996) Maxium-likelihood refinement of incomplete models with Buster+ TNT, Proceedings of Daresbury study weekend 85-92.

Bricogne, G. (1993). Direct phase determination by entropy maximisation and likelihood ranking. Status reports and perspectives. Acta Cryst. **D49,** 37-60.

Brooks, B., Bruccoleri, R., Olafson, B., States, D., Swaminathan, S., and Karplus, M., (1983)."CHARMM: A program for macromolecular energy minimisation and molecular dynamics calculations" , J.Comp. Chem. ***4,*** 187-217.

Brünger, A. T., Krukowski, A., and Erickson, J. (1990). "Slow-Cooling Protocols for Crystallographic Refinement by Simulated Annealing", Acta Cryst. **A46**, 585--593.

Brunger, A.T. (1993) XPLOR: a system for X-ray crystallography and NMR . Yale University press, New Haven, London.

Brunger, A.T. (1992) Free *R* value: a novel statisical quantity for assessing the accuracy of crystal structures, Nature **355**, 472-475.

Brünger, A.T. and Karplus, M. (1988). "Polar hydrogen positions in proteins: Empirical energy function placement and neutron diffraction comparison", Proteins **4**,148--156.

Brunger, A.T., Kuruyan, J. and Karplus, M., (1987) Crystallographic R factor refinement by molecular dynamics. Science. **235** 458-460

A.T. Brunger (1995)"The free R Value: A More Objective Statistic for Crystallography", Methods in Enzym. in press.

Butler, J.R., Alworth, W.L. & Nugent, M.J. (1974) Mechanism of dehydroquinase catalysed dehydration. I. Formation of a Schiff base intermediate. J.Am. Chem. Soc. **96**, 1617-1618.

Brukert, U. and Allinger, N.L. (1982). In: Molecular Mechanics, ACS Monograph **177**, American Chemical Society, Washington.

Charles, I.J., Keyte, J.W., Brammar, W.J. & Hawkins, A.R. (1985) Nucleotide sequence encoding the biosythetic dehydroquinase function of the pentafunctional *AROM* locus of *Aspergillus nidulans.* Nucleic Acids Res. **13**, 8119–8128.

Charles, I.J., Keyte, J.W., Brammar, W.J. & Hawkins, A.R. (1986) The isolation and nucleotide sequence of the complex *AROM* locus of *Aspergillus nidulans.* Nucleic Acids Res. **14**, 2201–2213.

Chaudhuri, S., Duncan, K., Graham, L.D. and Coggins, J.R. (1991) Identification of the active-site lysine residues of two biosynthetic 3-dehydroquinases. Biochem. J. **275**, 1-6.

Chaudhuri, S., Lambert, J.M., McColl, L. and Coggins, J.R. (1986) Purification and characterisation of 3-dehydroquinase from *Escherichia coli.* Biochem. J. **239**, 699-704.

Coggins, J.R. (1986) Enzymology and Molecular Biology as aids for the invention and improvement of herbicides. In "*Biotechnology and Crop Improvement and Crop Protection*" (British Crop Protection Council Monograph No. 34.) (ed. Day, P.R.) pp.101-110, British Crop Protection Council, London.

Coggins, J.R. (1989) The shikimate pathway as a target for herbicides. In "*Herbicides and Plant Metabolism*" (ed. Dodge, A.) pp.97-112. Cambridge University Press, Cambridge.

Coggins, J.R., Boocock, M.R., Chaudhuri, S., Lambert, J.M., Lumsden, J., Nimmo,G.A. and Smith, D.D.S. (1987) The *arom* multifunctional enzyme (*Neurospora crassa*). Methods in Enzymology **142**, 325-335.

Cowtan, K.D., and Main, P., (1993) Improvement of macromolecular electron density maps by the simultaneous application of real and reciprical space constraints. Acta Cryst. **D49,** 148-157.

Davies, G.M., Barrett-Bee, K.J., Jude, D.A., Lehan, M., Nichols, W.W., Pinder, P.E., Thain, J.L., Watkins, W.J. & Wilson, R.G. (1994). (6S)-6-fluoroshikimic acid, an antibacterial agent acting on the aromatic biosynthetic pathway. Antimicrob Agents Chemother **38**, 403-406.

Davis, B.D. (1955) Intermediates in amino acid biosynthesis. Advances in Enzymology **16**, 247-312.

Dessen, A., Quemard, A., Blanchard, J.S. Jacobs, W.R., Sacchettini, J.C. (1995) Crystal structure and function of the isoniazid target of Mycobacterium tuberculosis. Science **267,** 1638-1645.

Deka, R., Anton, I.A., Dunbar, B. and Coggins, J.R. (1994) The characterisation of the shikimate pathway enzyme dehydroquinase from *Pisum sativum*. FEBS Lett. **349**, 397-402.

Deka, R., Kleanthous, C. and Coggins, J.R. (1992) Identification of the essential histidine residue at the active site of *Escherichia coli* dehydroquinase J. Biol. Chem. **267**, 22237-22242.

Diamond, R., (1969) Profile analysis in single crystal diffractometry. Acta Cryst. **A25**, 43-55.

Duncan, K., Chaudhuri, S., Campbell, M.S. and Coggins, J.R. (1986) The overexpression and complete amino acid sequence of the 3-dehydroquinase of *Escherichia coli.* Biochem. J. **238**, 475-483.

Duncan, K., Edwards, R.M. and Coggins, J.R. (1987) The pentafunctional *arom* enzyme of *S. cerevisiae* is a mosaic of monofunctional domains. Biochem. J. **246**, 375-386.

Durbin, R.M., Burns , R., Moulai, J., Metcalf, P., Freymann., Blum. M., Anderson , J.E., Harrison, S.C., and Wiley, D.C., (1986) Protein, DNA and virus crystallography with a focused imaging proportional counter. Science 232, 1127-1132.

Engh R A & Huber R (1991). Accurate bond and angle parameters for X-ray protein structure refinement. Acta Cryst., **A47**, 392-400

Euverink, G.J.W., Hessels, G.I., Vrijbloed, J.W., Coggins, J.R. & Dijkhuizen, L. (1992) Purification and characterisation of a dual function 3-dehydroquinate dehydratase from *Amycolatopsis methanolica*. J. Gen. Microbiol. **138**, 2449-2457.

Evans, S.V., (1993) SETOR:hardware lighted three-dimensional solid model representations of macromolecules, J. Mol. Graphics, 11, 134-138.

Ganem, B. (1978) From glucose to aromatics: recent developments in natural products of the shikimate pathway. Tetrahedron **34**, 3353-83.

Garbe, T., Servos, S., Hawkins, A.R., Dimitriadis, G., Young, D., Dougan, G. & Charles, I.G. (1990) The *Mycobacterium tuberculosis* shikimate pathway genes: Evolutionary relationship between biosynthetic and catabolic 3-dehydroquinases. Molec. Gen. Genet. **228**, 385–392.

Gibson, F. & Pittard, J. (1968) Pathways of biosynthesis of aromatic amino acids and vitamins and their control in microrganisms. Bacteriol. Rev. **32**, 465–492.

Gibson, M.I. and Gibson, F. (1964) Preliminary studies on the isolation and metabolism of an intermediate in aromatic biosynthesis: chorismic acid. Biochem. J. **90**, 248-61.

Giles, N.H., Case, M.E., Baum, J., Geever, R., Huiet, L., Patel, V. & Tyler, B. (1985) Gene organisation and regulation in the QA (quinic acid) cluster of *Neurospora crassa*. Microbiol. Rev. **49**, 338–358.

Giles, N.H., Partridge, C.W.H., Ahmed, S.I., Case, M.E. (1967) The occurrence of two dehydroquinases in *Neurospora crassa*, one constitutive and one inducible. Proc. Natl. Acad. Sci. USA **58**, 1930-1937

Gomis-Rüth, F.X., Fita, I., Kiefersauer, R. Huber, Availes and Navaza, (1995) determination of hemihedral twinning and intial stuctural analysis of crystals if the Procarbopeptidase A ternary complex. Acta Cryst. **D51** 819-823.

Gourley, D.G., Coggins, J.R., Isaacs, N.W., Moore, J.D., Charles, I.G. and Hawkins, A.R. (1994) Crystallisation of a Type II Dehydroquinase from *Mycobacterium tuberculosis* . J. Mol. Biol. **241**, 488-491.

Green, D.W., Ingram, V.M. and Perutz, M.F.(1954). The stucture of haemoglobin . Proc.Roy. Soc. **A225**, 287-295.

Greer, J. (1985) Computer skeletonization and automatic electron-density map analysis. Methods In Enzymology, Vol.**115**, pp.206-224.

Hanson, K.R. & Rose, I.A. (1963) The absolute stereochemical course of citric acid biosynthesis. Proc. Natl. Acad. Sci. U.S.A. **50**, 981-988.

Harris, J., Concepcion-Gonzalez B, Kleanthous, C., Hawkins, A.R., Coggins, J.R. and Abell, C. (1996b) Mechanistic studies on type II dehydroquinases. Biochem. J. in the press

Harris, J., Kleanthous, C., Coggins, J.R., Hawkins, A.R. & Abell, C. (1993) Different mechanistic and stereochemical courses for the reactions catalysed by type I and type II dehydroquinases. J. Chem. Soc. Chem. Comm. 1080-1081.

Harris, J.M, (1994) Ph.D. Thesis, University of Cambridge.

Harris, J.M., Watkins, W.J., Hawkins, A.R., Coggins, J.R. & Abell, C. (1996a) Comparison of the substrate specificity of type I and type II dehydroquinases with 5-deoxy and 4,5-dideoxydehydroquinate. J.Chem. Soc. Chem. Comm in the press.

Haslam (1974) *The Shikimate Pathway*, Butterworths, London.

Haslam, E., (1993) *Shikimic Acid : Metabolism and Metabolites*, J. Wiley and Sons.,Chichester.

Hautala, J.A., Jacobson, J.W., Case, M.E. and Giles, N.H. (1975) Purification and characterization of catabolic dehydroquinase an enzyme of the inducible quinic acid catabolic pathway of *Neurospora crassa*. J. Biol. Chem. **250**, 6008-14.

Hawkins, A.R.(1987). The complex *arom* locus of *Aspergillus nidulans.* Evidence for multiple gene fusions and convergent evolution Curr Genet. **11**, 291-298.

Hawkins, A.R., Giles. N. H. & Kinghorn. J.R. (1982). Genetical and Biochemical aspects of quinate breakdown in the filamentous fungus *Aspergillus nidulans.* Biochem Genet. **20**, 271-286.

Hawkins, A.R., Lamb, H.K.& Roberts, C.F. (1992) Structure of the *Aspergillus nidulans qut* repressor-encoding gene: implications for the regulation of transcription initiation. Gene **110**, 109-14.

Hawkins, A.R., Lamb, H.K., Moore, J.D.& Roberts, C.F. (1993a) Genesis of eukaryotic transcriptional activator and repressor proteins by splitting a multidomain anabolic enzyme. Gene **136**, 49-54.

Hawkins, A.R., Lamb, H.K., Moore, J.D., Charles, I.G. & Roberts, C.F. (1993b) The pre-chorismate (shikimate) and quinate pathways in filamentous fungi: theoretical and practical aspects. J. Gen. Microbiol. **139**, 2891-99.

Hawkinson, D.C, Eames, T.C.M., Pollack, R.M. (1991). Energetics of 3-Oxo-delta-5-steroid isomerase source of the catalytic power of the enzyme. Biochemistry. **30,** No.45,.10849-10858

Hendrix, J., and Lentfer, A., (1988) An Image Plate Scanner. EMBL Research Reports. 170-171.

Howard, A.J., Gilliland, G.L., Finzel, B.C., Poulos,T.L., Ohlendorf, D.H., Salemme, F.R., (1987). The use of an imaging proportional counter in macromolecular crystallography. J.Appll Crys. **20,** 383-387.

Huiet, L. and Giles, N.H. (1986) The qa repressor gene of Neurospora crassa: wild type and mutant nucleotide sequences. Proc. Natl. Acad. Sci. USA **83**, 3381-85.

Jones, T A. and Kjeldgaard, M. (1994) Making the First Trace with O.*"From first map to final model."* (ed. Bailey, S. Hubbard, R. Waller, D.). Daresbury Laboratory, Warrington, England, 1-13

Jones, T. A., Zou, J. Y., Cowan, S. W. and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps in these models. *Acta Cryst.* **47** 110-119

Kabsch , W. (1988) Automatic indexing of rotation diffraction patterns. J Appl.Crys . **21**, 67-71.

Kabsch , W. (1988) Evalutation of single-crystal X-ray diffraction data from a position sensitive detector. J Appl.Crys . **21**, 916-924.

Kabsch , W. (1993) Automatic processing of rotation diffraction data from crystals of intially unknown cell constants. J Appl.Crys. **26**, 795-800.

Karplus, M. and Petsko, G.A. (1990). ``Molecular-Dynamics Simulations in Biology", Nature **347**,631--39.

Kleanthous, C., Deka, R., Davies, K., Kelly, S., Cooper, A., Harding, S.E. Price, N.C., Hawkins, A.R. and Coggins, J.R. (1992) A comparison of the enzymological and biophysical properties of two distinct classes of dehydroquinase enzymes. Biochem. J. **282**, 687-695.

Kleanthous, C., Reilly, M., Cooper, A., Kelly, S., Price, N.C. and Coggins, J.R. (1991) Stabilisation of the shikimate pathway enzyme dehydroquinase by covalently bound ligand. J. Biol. Chem. **266**, 10893-10898.

Konnert J.H., Acta Cryst., A32, (1976) 614-617

Konnert, J.H.and Hendrickson, W.A., (1980) A restrained-parameter thermal factor refinement procedure. Acta Cryst. **A36** 344-350.

Koshiba, T. (1978) Purification of two forms of the associated 3-dehydroquinate hydro-lyase and shikimate:NADPD oxidireductase in *Phaseolus mungo* seedlings. Biochim. Biophys. Acta. **522**, 10-18.

Krell, T., Horsburgh, M.J., Cooper, A., Kelly, S.M. & Coggins, J.R. (1996) Localisation of the active site of the type II dehydroquinases. J. Biol. Chem. **271** 24492-97.

Krell, T., Pitt, A.R. and Coggins, J.R. (1995) The use of electrospray mass spectrometry to identify an essential arginine residue in type II dehydroquinases. FEBS Letters **360**, 93-96.

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature **227**, 680-685.

Lamb, H.K., Moore, J.D., Lakey, J.H., Levett, L., Wheeler, K.A., Lago, H., Coggins, J.R., and Hawkins, A.R. (1996) Comparative analysis of the QUTR transcription repressor protein and the 3 C-terminal domains of the pentafunctional *AROM* enzyme. Biochem. J. **313**, 941-50.

Lambert, J.M., Boocock, M.R. and Coggins, J.R. (1985) The 3-dehydroquinate synthase activity of the pentafunctional *arom* complex of *N. crassa* is zinc dependent. Biochem. J. **226**, 817-829.

Lai, C.Y and Horecker, B.L. (1972) Aldolase: A model for enzyme structure-function relationships. Essays in Biochemistry **8**, 149-178.

Lamzin, V.S. and Wilson, K.S., (1993) Acta Cryst., **D49**, 129-147.

Laskowski R A, MacArthur M W, Moss D S & Thornton J M (1993). PROCHECK:a program to check the stereochemical quality of protein structures. J. Appl. Cryst., **26**, 283-291

Leech, A.P., James, R., Coggins, J.R. and Kleanthous, C. (1995) Mutagenesis of active site residues in type I dehydroquinase from *E. coli*; stalled catalysis in a histidine to alanine mutant. J. Biol. Chem. **270**, 25827-836.

Levin, J.G. & Sprinson, D.B. (1964) The enymatic formation and isolation of 3-enolpyruvyl-shikimate-5-phosphate. J. Biol. Chem. **239**, 1142-50.

Lifson, S. and Stern, P.S. (1982). ``Born-Oppenheimer Energy Surfaces of SimilarMolecules Interrelations Between Bond Lengths, Bond Angles, and Frequencies of Normal Vibrations in Alkanes", J.Chem. Phys. **77**, 4542--4550.

Luzzati, P.V. (1952). ``Traitement Statistique des Erreurs dans la Determination des Structures Cristallines",Acta Cryst **5**, 802--810.

Lumsden, J. and Coggins, J.R. (1977) The subunit structure of the *arom* multienzyme complex of *Neurospora crassa*: a possible pentafunctional polypeptide chain. Biochem. J. **161**, 599-607.

Lumsden, J. and Coggins, J.R. (1978) The subunit structure of the *arom* multienzyme complex of *Neurospora crassa*: evidence from peptide maps for the identity of the subunits. Biochem. J. **169**, 441-444.

Macpherson, A.(1982). The preparation and analysis of protein crystals Chapter 3, John Wiley & Sons, New York.

Mathews, B.W. (1968). Solvent content of protein crystals. J. Mol. Biol. **33,** 491-497.

Mitsuhashi, S., Davis, B.D. (1954) Aromatic biosynthesis. XII. Conversion of 5-dehydroquinic acid to 5-dehydroshikimic acid by 5-dehydroquinase. Biochim. Biophys. Acta **15**, 54-61.

Moore, J.D., Hawkins, A.R., Charles, I.G., Deka, R., Coggins, J.R., Cooper, A., Kelly, S.M. and Price, N.C. (1993) Characterisation of the type I dehydroquinase from *Salmonella typhi.* Biochem. J. **295**, 277-285.

Moore, J.D., Lamb, H.K., Garbe, T., Servos, S., Dougan, G., Charles, I.G., & Hawkins, A.R., (1992). Inducible overproduction of the *Aspergillus nidulans* pentafunctional *AROM* protein and the type -I and -II 3-dehydroquinases from *Salmonella typhi* and *Mycobacterium tuberculosis*. Biochem. J. **287,** 173-181.

Morris A L, MacArthur M W, Hutchinson E G & Thornton J M (1992). Stereochemical quality of protein structure coordinates. Proteins, **12**, 345-364

Mousdale, D.M. and Coggins, J.R. (1991) Amino acid synthesis. In "*Target Sites for Herbicide Action*" (ed. Kirkwood, R.C.) pp.29-56. Plenum Press, New York.

Mousdale, D.M. and Coggins, J.R. (1993) The Shikimate pathway. In. "*Methods in Plant Biochemistry, Volume 9*" (ed. Lea, P.J.) pp 1-23. Academic Press, London.

Mousdale, D.M., Campbell, M.S. and Coggins, J.R. (1987) Purification and characterisation of the bifunctional dehydroquinase-shikimate: NADP oxidoreductase from pea seedlings. Phytochemistry **26**, 2665-2670.

Murshudov, N.G., Dodson, E.J., Vagin, A.A. (1996). Application of Maxium-likelihood methods for macromolecular refinement, Proceedings of Daresbury study weekend 93-104

Murshudov, N.G., Vagin, A.A., Dodson, E.J. (1997). Refinement of macromolecular structures by the Maxium-likelihood method, Acta Cryst. (1997) **D53,** 240-255.

Otwinowski, Z. (1993). Oscillation data reduction program.Application of Maxium-likelihood methods for macromolecular refinement, Proceedings of Daresbury study weekend 56-62.

Némethy, G., Pottie, M.S. and Scheraga, H.A. (1983) ``Energy Parameters in Polypeptides. 9. Updating of Geometrical Parameters, Nonbonded Interactions, and Hydrogen Bond Interactions for the Naturally Occurring Amino Acids", J. Phys. Chem. **87**, 1883--1887.

Pannu, N.S., Read, R.J. (1996). Improved stucture refinement through maxium-likelihood, Proceedings of Daresbury study weekend 75-84.

Patel, V.B. & Giles, N.H. (1979) Purification of the *arom* mutienzyme aggregate from *Euglena gracilis*. Biochim. Biophys. Acta. **567**, 24–34.

Perutz, M.F. (1956) . Acta Cryst., **9**, 867.

Polley, L.D. (1978) Purification and characterisation of 3-dehydroquinate hydrolyase and shikimate oxidoreductase: evidence for a bifunctional enzyme. Biochim. Biophys. Acta. **526**, 259–266.

Quemard, A., Sacchettini, J.C., Dessen, A., Vilcheze, C., Bittman, R., Jacobs, W.R., Blanchard., J.S. (1995). Enzymic characterization of the target for isoniazid in Mycobacterium tuberculosis. Biochemistry **34,** 8235-8241.

Read, R.J., (1990). Structure factor probabilities for related structures. Acta Cryst., **A46** 900-912.

Richardson, J.S. (1977). β-sheet topology and the relatednedd of proteins. Nature **268,** 495-500

Robertson, J.M. & Woodward, I. (1937) An X-ray study of the pthalocyanines . Part III, Quantitative stucture determination of nickel pthalocyanine. J.Chem. Soc. **219,** 219-230.

Rost, Burkhard; Sander, Chris (1993) Prediction of protein structure at better than 70% accuracy. J. Mol. Biol.. **232**, 584-599.

Rost, Burkhard; Sander, Chris (1994) Combining evolutionary information and neural networks to predict protein secondary structure. Proteins, **19**, 55-72.

Salamon, I.I. and Davis, B.D. (1953) Aromatic biosynthesis. IX. The isolation of a precursor of shikimic acid. J. Am. Chem. Soc. **75**, 5567-5571.

Sato, M., Yamamoto, M., Imada, K., Katsube, Y., Tanaka, N., Higashi, T. (1992) A High-Speed Data-Collection System for Large-Unit-Cell Crystals using an Image Plate as a Detector. J.Appl. Cryst. **25**, 348-357.

Sayre, D. (1974) . Least-squares refinement of high resolution phasing of small proteins. Acta Cryst. **A30**. 180-184

Servos, S., Chatfield, S., Hone, D., Levine, M., Dimitriadis, G., Pickard, D., Dougan, G., Fairweather, N., & Charles, I.G., (1991). Molecular cloning and characterisation of the *aroD* gene encoding 3-dehydroquinase from *Salmonella typhi* . J. Gen . Microbiol. **137**, 147-152.

Sheldrick, G.M. (1991). Heavy atom location using SHELXS-90, Proceedings of Daresbury study weekend 23-38

Shneier, A., Harris, J., Kleanthous, C., Coggins, J.R., Hawkins, A.R. & Abell, C. (1993) Evidence for different stereochemical courses in the reaction catalysed by the the type I and type II dehydroquinases. Bioorg. Med. Chem. Lett. **3**, 1399-1402

Shneier, A., Kleanthous, C., Deka, R., Coggins, J.R. and Abell, C. (1991) Observation of an imine intermediate on dehydroquinase by electrospray mass spectrometry. J. Am. Chem. Soc. **113**, 9416-9418.

Shumilin, I.A., Kretsinger, R.H. & Bauerle, R. (1996) Purification, crystallisation and preliminary crystallographic analysis of 3-deoxy-D-*arabino*-heptulosonate 7-phosphate synthase from *Escherichia coli*. Proteins **24**, 404-6.

Shuttleworth, W.A., Hough, C.D., Bertrand, K.P. and Evans, J.N.S. (1992) Over-production of 5-enolpyruvylshikimate-3-phosphate synthase in *Escherichia coli*: use of the T7 promoter. Protein Engineering **5**, 461-6.

Silva, A.M. and Rossman, M.G. (1985). The refinement of Southern bean Mosaic virus in recipricol space. Acta Cryst., **B41,** 147-157.

Smith, D.D.S. and Coggins, J.R. (1983) Limited proteolysis of the *arom* enzyme complex of *N. crassa*: evidence for the isolation of a bifunctional domain of the pentafunctional polypeptide. Biochem. J. **213**, 405-415.

Spratt, B.G. (1994) Resistance to antibiotics mediated by target alterations. Science **264**, 388-93.

Stallings, W.C., Abdel-Meguid, S.S., Lim, L.W., Shieh, H.S., Dayringer, H.E., Leimgruber, N.K., Stegeman, R.A., Anderson, K.S., Sikorski, J.A., Padgette, S.R. and Kishore, G.M. (1991) Structure and topological symmetry of the glyphosate target 5-*enol*-pyruvylshikimate-3-phosphate synthase: a distinctive protein fold. Proc. Natl. Acad. Sci. U.S.A. **88**, 5046-5050.

Steinrucken, H.C. & Amrhein, N. (1980) The herbicide glyphosate is a potent inhibitor of 5-enolpyruvylshikimate acid 3-phosphate synthase. Biochem. Biophys. Res. Commun. **94**, 1207-12.

Swanson, S.M.,(1994) Core tracing; depicting conections between features in electron density. Acta Cryst. **A50** 695-708.

Turner, M.J., Smith, B.W. and Haslam, E. (1975) The shikimate pathway. Part IV. The stereochemistry of the 3-dehydroquinate dehydratase reaction and observations on 3-dehydroquinate synthetase. J. Chem. Soc. Perkin Trans. 52-55.

Vaz, A.D.N., Butler, J.R. and Nugent, M.J. (1975) Dehydroquinase catalyzed dehydration. II. Identification of the reactive conformation of the substrate responsible for the *syn* elimination. J.Am. Chem. Soc. **97**, 5914-5

Walker, J.E., Saraste, M., Runswick, M.J. and Gay, N.J. (1982) Distantly related sequence in the a and b subunits of ATP synthase, myosin, kinases and other ATP requiring enzymes and a common nucleotide binding fold. EMBO J. **1**, 945-51.

Wang, B.C. (1985) Resolution of phase ambinguity in macromolecular crystallography. Methods in Enzymolgy, **115** 90-112

Weiss, U. & Edwards, J.M. (1980) *The Biosynthesis of Aromatic Compounds.* Wiley, New York.

Weiss, U., Davis, B.D. and Mingioli, E.S. (1953) Aromatic biosynthesis. X. Identification of an early precursor as 5-dehydroquinic acid. J. Am. Chem. Soc. **75**, 5572-76

White, P.J., Young, J., Hunter, I.S., Nimmo, H.G. and Coggins, J.R. (1990) The purification and characterisation of 3-dehydroquinase from *Streptomyces coelicolor.* Biochem. J. **265**, 735-738.

Wu, Z.R., Ebrahimian, S., Zawrotny, M.E., Thornburg, L.D., Perez-Alvarado, G.C., Brothers, P., Pollack, R.M., Summers, M.F. (1997) Solution Structure of 3-Oxo-$\Delta^5$ -steroid isomerase. Science. **276** 415-418.

Xiao, B., Jones, D., Madrazo, J., Soneji, Y., Aitken, A., and Gablin, S. (1996) Crystallization of a 14-3-3 protein . Acta Cryst. **D52** 203-206.

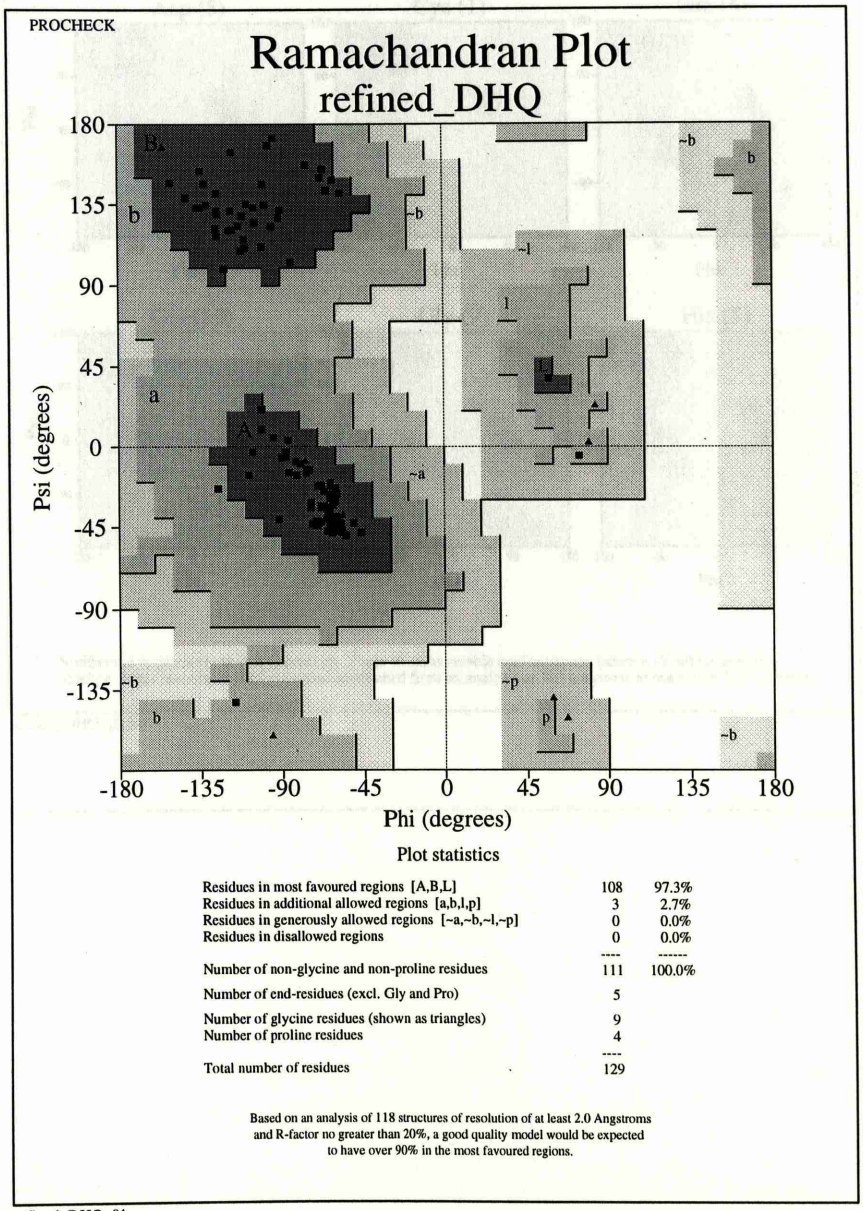
Young, D.B. and Duncan, K. (1995) Prospects for new interventions in the treatment and prevention of Mycobacterial disease. Ann. Rev. Microbiol. **49**, 641-73.

Zang, K.Y and Main, P., (1990) Histogram matching as a new density modification technique for phase refinement and extension of protein molecules. Acta Cryst. **A49** 41-46.

Zang, K.Y and Main, P., (1990) The use of Sayre`s Equation with solvent flattening and histogram matching for phase extension and refinement of protein structures. Acta Cryst. **A49** 377-381

# Appendix

## Results Of Analysis of Final Refined DHQ Model using PROCHECK




Plot statistics

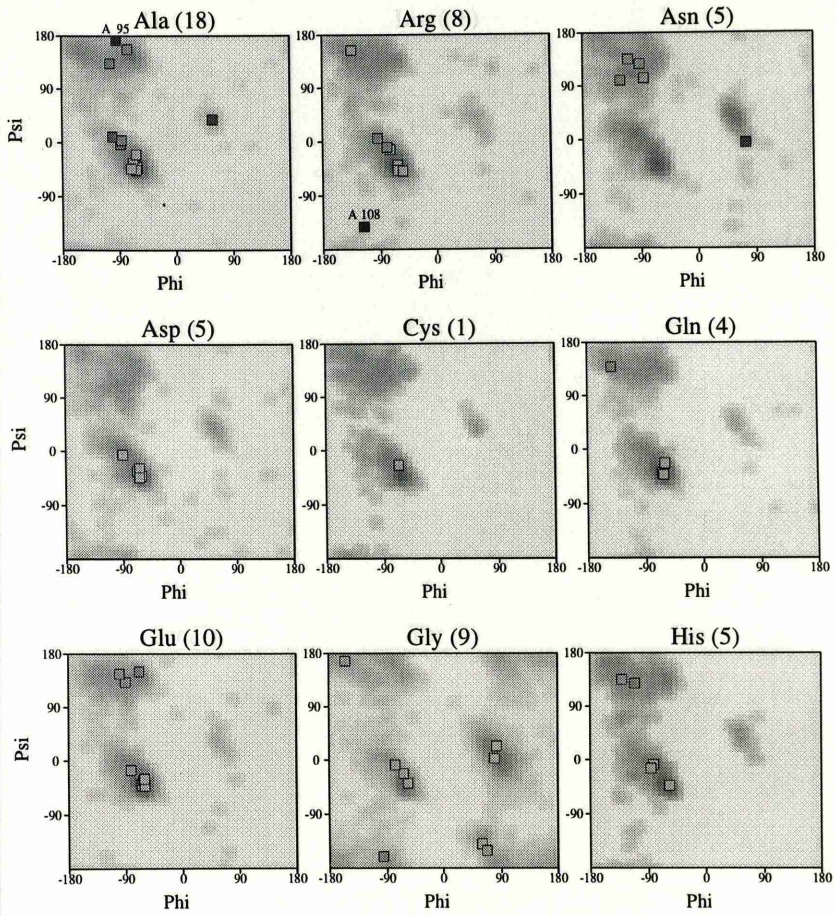
| | | |
|---|---|---|
| Residues in most favoured regions [A,B,L] | 108 | 97.3% |
| Residues in additional allowed regions [a,b,l,p] | 3 | 2.7% |
| Residues in generously allowed regions [~a,~b,~l,~p] | 0 | 0.0% |
| Residues in disallowed regions | 0 | 0.0% |
| Number of non-glycine and non-proline residues | 111 | 100.0% |
| Number of end-residues (excl. Gly and Pro) | 5 | |
| Number of glycine residues (shown as triangles) | 9 | |
| Number of proline residues | 4 | |
| Total number of residues | 129 | |

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

refined_DHQ_01.ps

# Ramachandran plots for all residue types

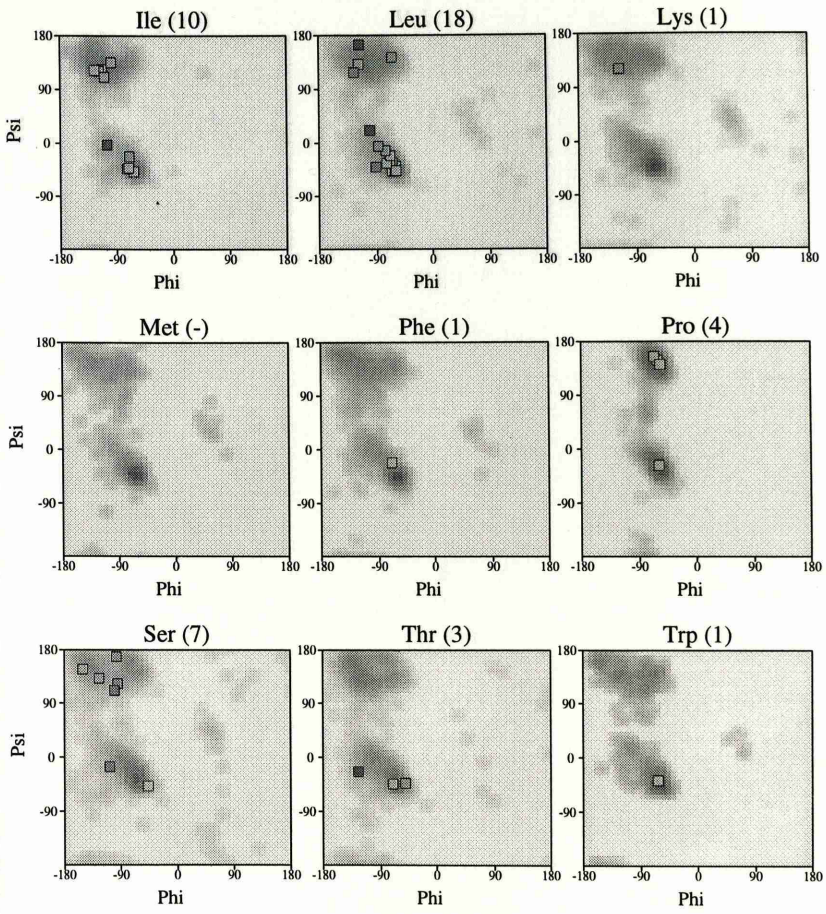
## refined_DHQ




Numbers of residues are shown in brackets. Those in unfavourable conformations (score < -2.00) are labelled.
Shading shows favourable conformations as obtained from an analysis of 163 structures at resolution 2.0A or better.

# Ramachandran plots for all residue types

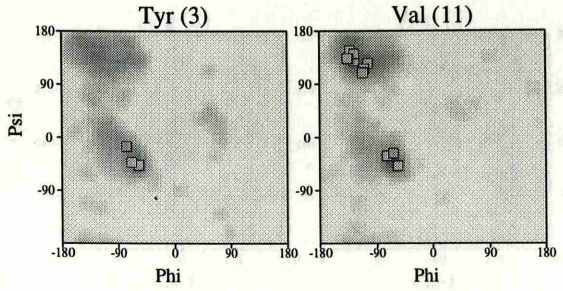
## refined_DHQ




Numbers of residues are shown in brackets. Those in unfavourable conformations (score < -2.00) are labelled.
Shading shows favourable conformations as obtained from an analysis of 163 structures at resolution 2.0A or better.

# Ramachandran plots for all residue types

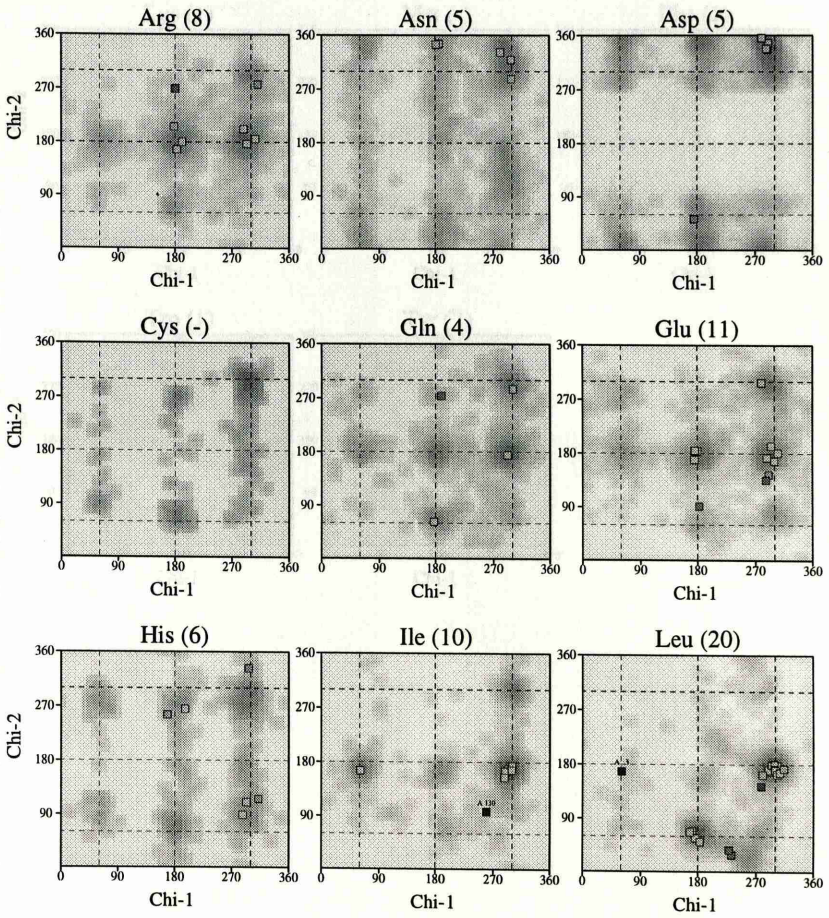
## refined_DHQ




Numbers of residues are shown in brackets. Those in unfavourable conformations (score < -2.00) are labelled.
Shading shows favourable conformations as obtained from an analysis of 163 structures at resolution 2.0A or better.

# Chi1-Chi2 plots

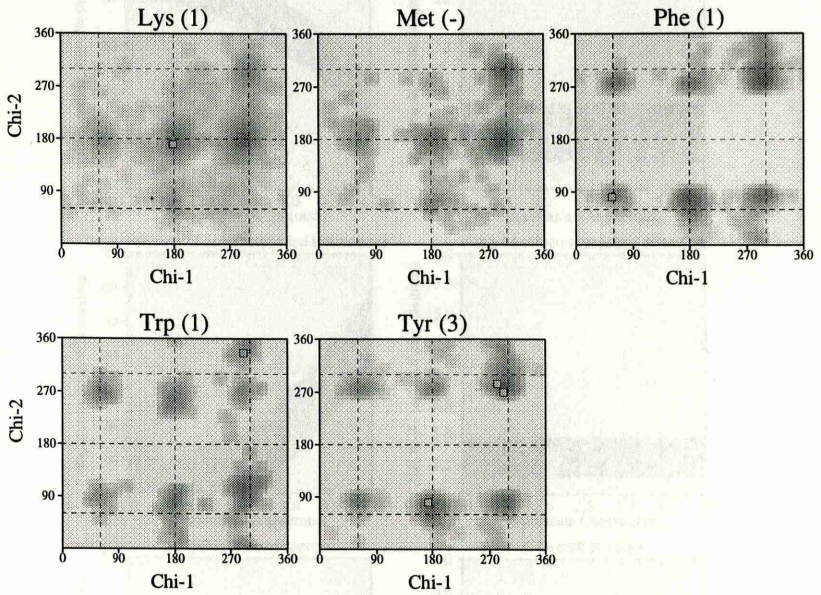
## refined_DHQ




Numbers of residues are shown in brackets. Those in unfavourable conformations (score < -2.00) are labelled.
Shading shows favourable conformations as obtained from an analysis of 163 structures at resolution 2.0A or better.

# Chi1-Chi2 plots

## refined_DHQ

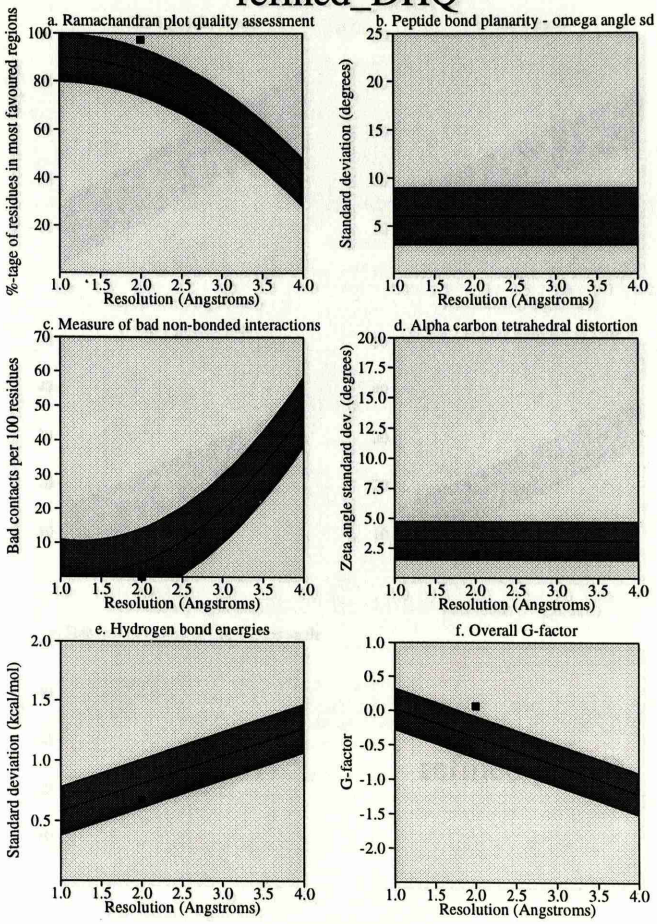



Numbers of residues are shown in brackets. Those in unfavourable conformations (score < -2.00) are labelled.
Shading shows favourable conformations as obtained from an analysis of 163 structures at resolution 2.0A or better.

PROCHECK

# Main-chain parameters

## refined_DHQ




### Plot statistics

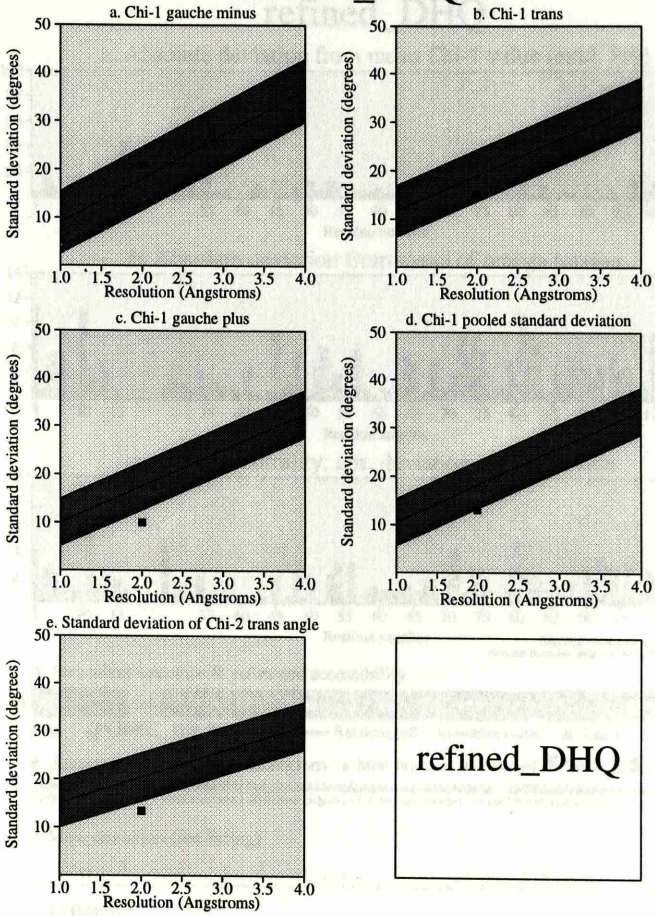
| Stereochemical parameter | No. of data pts | Parameter value | Comparison values Typical value | Band width | No. of band widths from mean | |
|---|---|---|---|---|---|---|
| a. %-tage residues in A, B, L | 111 | 97.3 | 83.8 | 10.0 | 1.3 | BETTER |
| b. Omega angle st dev | 126 | 3.7 | 6.0 | 3.0 | -0.8 | Inside |
| c. Bad contacts / 100 residues | 0 | 0.0 | 4.2 | 10.0 | -0.4 | Inside |
| d. Zeta angle st dev | 119 | 2.0 | 3.1 | 1.6 | -0.7 | Inside |
| e. H-bond energy st dev | 91 | 0.7 | 0.8 | 0.2 | -0.7 | Inside |
| f. Overall G-factor | 129 | 0.1 | -0.4 | 0.3 | 1.5 | BETTER |

refined_DHQ_04.ps

PROCHECK

# Side-chain parameters

## refined_DHQ




refined_DHQ

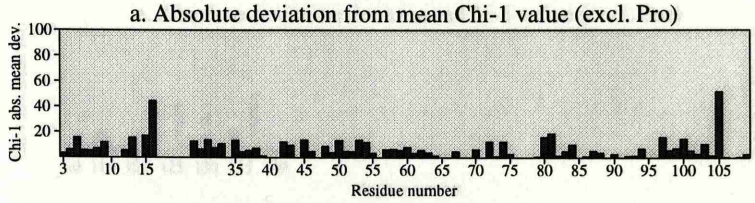
### Plot statistics

| Stereochemical parameter | No. of data pts | Parameter value | Comparison values Typical value | Band width | No. of band widths from mean | |
|---|---|---|---|---|---|---|
| a. Chi-1 gauche minus st dev | 7 | 21.0 | 18.1 | 6.5 | 0.5 | Inside |
| b. Chi-1 trans st dev | 37 | 15.2 | 19.0 | 5.3 | -0.7 | Inside |
| c. Chi-1 gauche plus st dev | 53 | 10.0 | 17.5 | 4.9 | -1.5 | BETTER |
| d. Chi-1 pooled st dev | 97 | 13.1 | 18.2 | 4.8 | -1.0 | BETTER |
| e. Chi-2 trans st dev | 40 | 13.4 | 20.4 | 5.0 | -1.4 | BETTER |

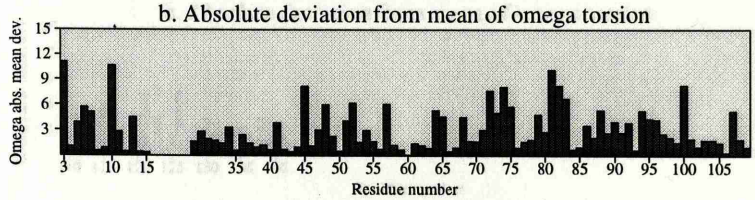
refined_DHQ_05.ps

# Residue properties
## refined_DHQ

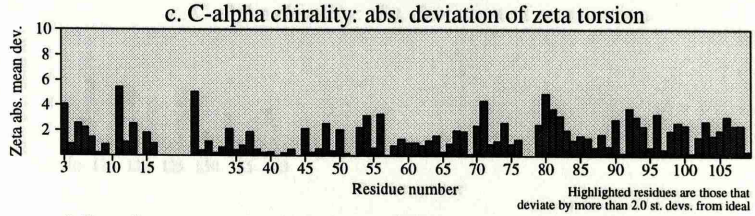
a. Absolute deviation from mean Chi-1 value (excl. Pro)




b. Absolute deviation from mean of omega torsion

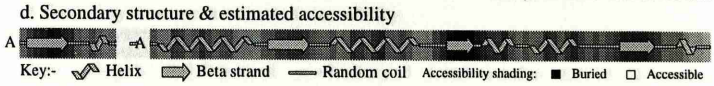



c. C-alpha chirality: abs. deviation of zeta torsion



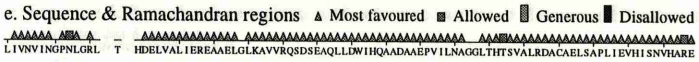

Highlighted residues are those that deviate by more than 2.0 st. devs. from ideal

d. Secondary structure & estimated accessibility



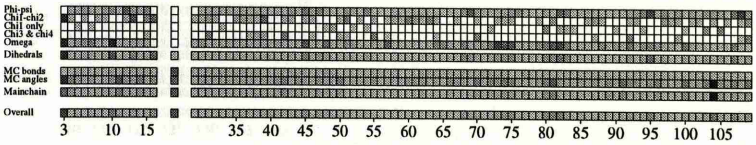

Key:- Helix Beta strand Random coil Accessibility shading: Buried Accessible

e. Sequence & Ramachandran regions Most favoured Allowed Generous Disallowed



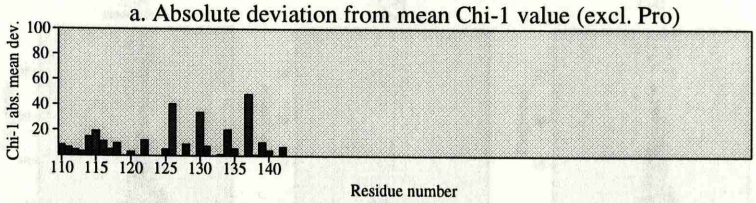

f. Max. deviation (see listing)

g. G-factors

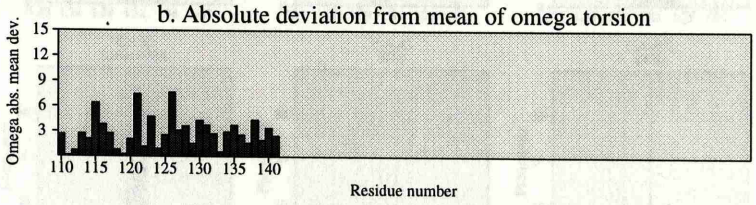


# Residue properties
## refined_DHQ

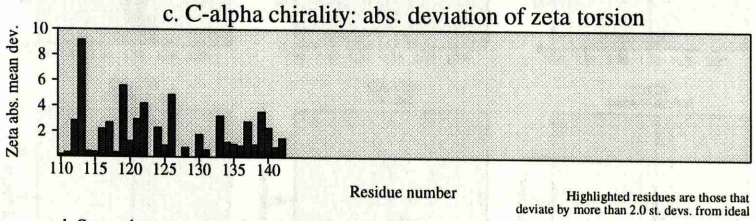
a. Absolute deviation from mean Chi-1 value (excl. Pro)




b. Absolute deviation from mean of omega torsion

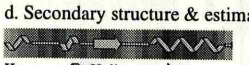



c. C-alpha chirality: abs. deviation of zeta torsion



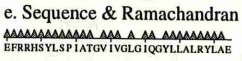

Highlighted residues are those that deviate by more than 2.0 st. devs. from ideal

d. Secondary structure & estimated accessibility

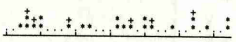


Key:- Helix Beta strand Random coil Accessibility shading: Buried Accessible

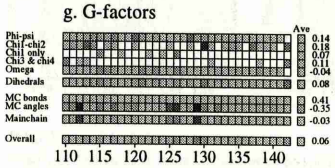
e. Sequence & Ramachandran regions Most favoured Allowed Generous Disallowed




f. Max. deviation (see listing)

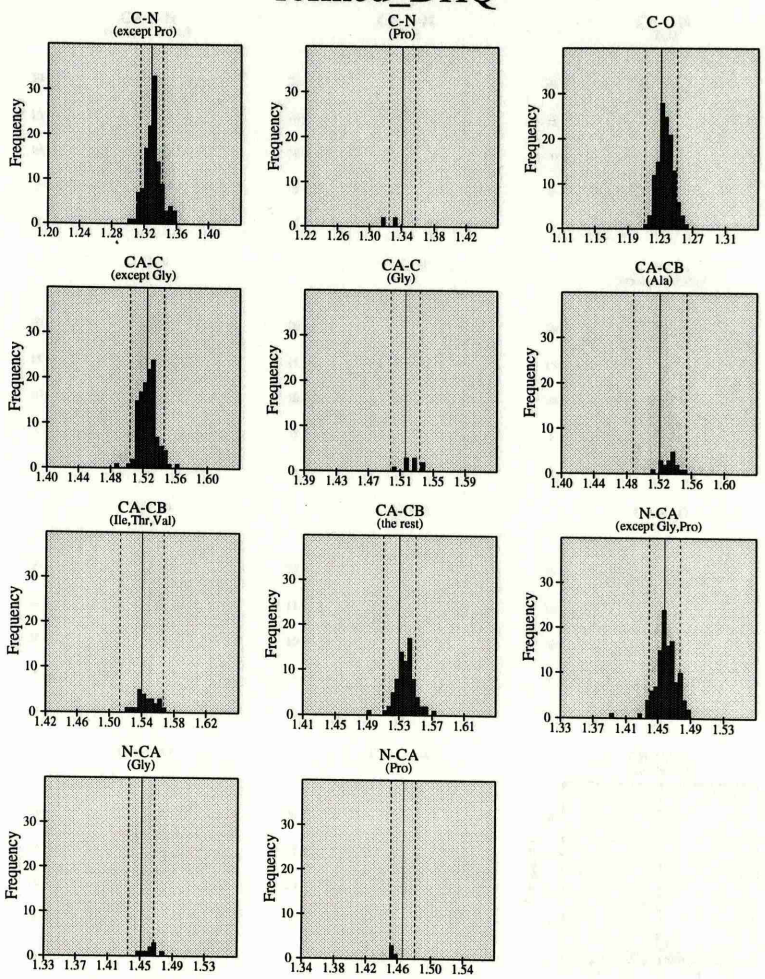


g. G-factors



# Main-chain bond lengths

## refined_DHQ

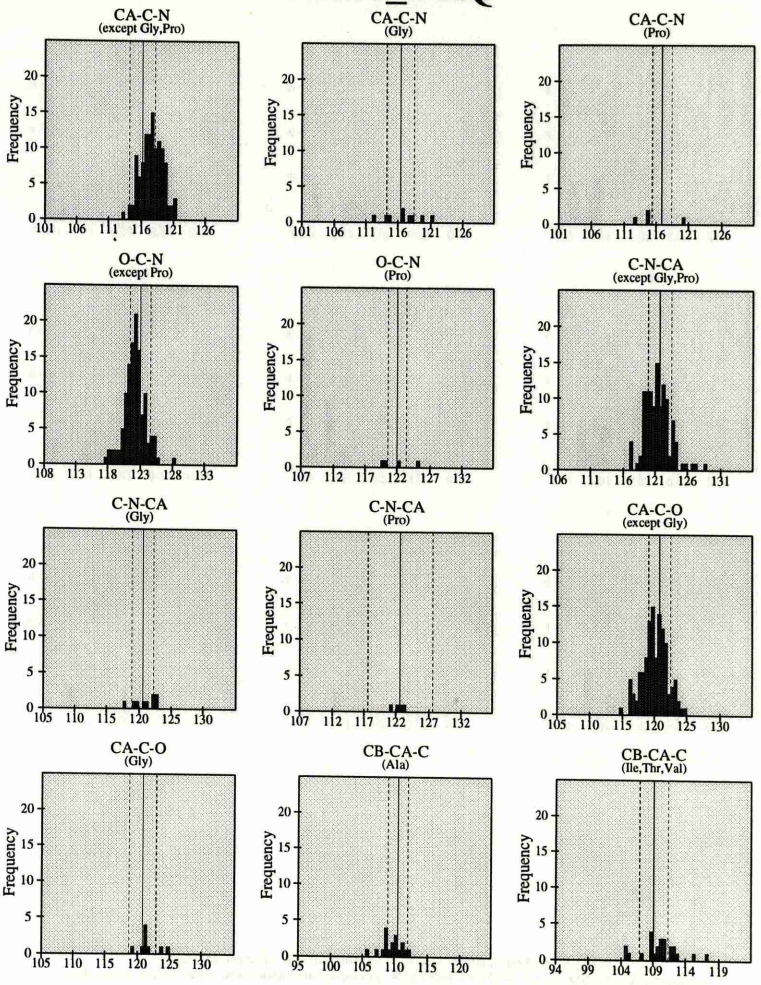



Black bars > 2.0 st. devs. from mean.

Solid and dashed lines represent the mean and standard deviation values as per Engh & Huber small-molecule data.

# Main-chain bond angles

## refined_DHQ

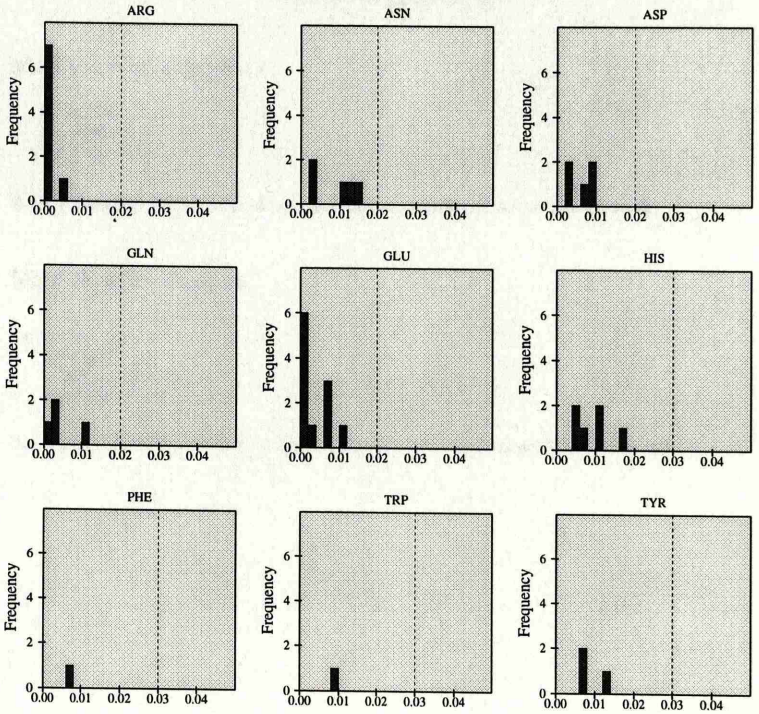



Black bars > 2.0 st. devs. from mean.

Solid and dashed lines represent the mean and standard deviation values as per Engh & Huber small-molecule data.

# RMS distances from planarity

## refined_DHQ




Histograms showing RMS distances of planar atoms from best-fit plane.
Black bars indicate large deviations from planarity: RMS dist > 0.03 for rings, and > 0.02 otherwise.

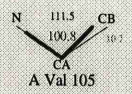
# Distorted geometry
## refined_DHQ

### Main-chain bond lengths

A His 29

Bonds differing by > 0.05A from small-molecule values. Values shown: "ideal", difference, actual

### Main-chain bond angles



A Val 105

Bond angles differing by > 10.0 degrees from small-molec values. Values shown: "ideal", actual, diff.