X-RAY CRYSTALLOGRAPHIC STUDIES OF THERAPEUTIC ENZYMES: NITROREDUCTASE AND AKR1C3.

By
ANDREW LEE LOVERING

A thesis submitted to
The University of Birmingham
For the degree of
DOCTOR OF PHILOSOPHY

School of Biosciences
The University of Birmingham
February 2003

Acknowledgements

A very informal part this – at last after writing countless pages of (nearly) structured work. Great thanks first to Scott, it's been a pleasure. I think I appreciate you never dumbing anything down most, although it was hard going at the start! It's been nice (in terms of each thinking you were working with the other, in actuality I was in the pub) to have a co-supervisor in Eva (Peter too), who was wonderful in sorting out my lab work. Thanks to both. I won't forget the effort or friendship. That should get me a good reference.

I had a great time in collaborating with Jon and Chris, thanks also to Anna & Anya for helping whilst I worked in their lab. Last of the semi-formal thanks goes to all the technical support on data collection trips, and that at home (Rosemary, I swear I never mean to wind you up, it's just inevitable), the BIP lot too.

Labmates. What the hell, why not single a few people out. Andy Milner was superb – a big thank you and shout of "ehhh cabbage". Owen Mather and Paul Race, cheers for everything. Klaus – thanks for all the technical advice and laughter at some of my wilder theories. Vicki Howard was a great help in settling me into the lab, various lab-related thank yous to floors 3, 5, 7 and 8. If I didn't mention the teaching part of my PhD then sorry, I just ran out of space. Just then.

Obvious thanks to outside influences. That lot above would get a bit much otherwise. Firstly all of my family. Secondly Helen. And all the people who like the Birmingham Groundhog Day scenario of get drunk, eat curry. Jim & Lee – you've got nothing. Mel & Bob, the Cancer Studies lot, Uni mates and various drinkers. If you were looking for a laugh then sorry, I suggest you read some of my more outlandish theories at the end of the thesis.

Abstract

The *Escherichia coli* enzyme nitroreductase has been proposed as a candidate for the Gene-Directed Enzyme Prodrug Therapy approach in treating cancer. Structural studies on the enzyme were instigated in a first step towards improving enzyme activity. The enzyme was crystallized with the substrate analogue, nicotinic acid, and the structures of three crystal forms obtained. The fold has a mixed α/β structure, with a molecule of nicotinic acid bound next to the FMN cofactor. Several active site residues were identified as candidates for mutation. This procedure produced many mutant enzymes with increased catalytic activity. One double and four single mutants were chosen from these and crystal structures determined. The resulting information from this, and the establishment of a proof of principle, provides the basis for iterative cycles of enzyme improvement.

The human hydroxysteroid dehydrogenase AKR1C3 has been proposed to play a role in prostaglandin metabolism. Its inhibition by non-steroidal anti-inflammatory drugs may be important in a tumour differentiation strategy. AKR1C3 was crystallized, and the structure solved with bound nucleotide cofactor and several inhibitors, including the drugs indomethacin and flufenamic acid. Having obtained information on drug binding to AKR1C3, selective inhibitors can be designed, avoiding inhibition of "housekeeping" enzymes such as cyclo-oxygenases.

Table of Contents

List of Illustrations

Chapter Seven

Chapter Eight

List of Tables

Abbreviations

AKR	aldoketoreductase
Amp	ampicillin
APC	adenomatous polyposis coli
Ap-1	adenomatous polyposis coli family gene 1
ASU	asymmetric unit
ATRA	all-trans retenoic acid
Bcl-2	B-cell lymphoma family gene 2
BPB	bromophenol blue
BSA	bovine serum albumen
CB1954	5-(1-aziridinyl)-2,4-dinitrobenzamide
CCD	charge-coupled device
COX	cyclo-oxygenase
CRUK	Cancer Research United Kingdom
Da	Daltons
DHT	5α-dihydrotestosterone
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
DTT	dithiothreitol
E. cloacae	*Enterobacter cloacae*
E. coli	*Escherichia coli*
EDTA	ethylenediamine tetra-acetic acid
ESRF	European Synchrotron Radiation Facility, Grenoble, France
FAD	flavin adenine dinucleotide
FLA	flufenamic acid
FMN	flavin mononucleotide
FOM	figure of merit
FRase	flavin reductase
G/A/VDEPT	gene/antibody/virus-directed enzyme prodrug therapy
G. echinata	*Glycyrrhiza echinata*
Holo	holoenzyme
HPLC	high performance liquid chromotography
H. pylori	*Helicobacter pylori*
HSD	hydroxysteroid dehydrogenase
IMN	indomethacin
IP	inflection point wavelength
IPTG	isopropyl-β-D-thiogalactopyranoside
LB	Luria Bertani
LBHB	low-barrier hydrogen bond
M/SAD	multiple/single wavelength anomalous dispersion
M/SIR	multiple/single isomorphous replacement
M/SIRAS	multiple/single isomorphous replacement anomalous scattering
MPD	2-methyl,2-4-pentanediol
MR	molecular replacement

Mwt	molecular weight
NAD(H)	nicotinamide adenine dinucleotide (reduced form)
NADP(H)	nicotinamide adenine dinucleotide phosphate (reduced form)
NCS	non-crystallographic symmetry
NF-κB	Nuclear factor κB
nfsA	major oxygen-insensitive nitroreductase of *E. coli*
nfsB / NTR	minor oxygen-insensitive nitroreductase of *E. coli*
nfsC	unidentified second minor oxygen-insensitive nitroreductase of *E. coli*
nfsI	classical nitroreductase of *E. cloacae*
NIC	nicotinic acid
NOX	NADH oxidase of *T. thermophilus*
NQO1	quinone reductase / diaphorase
NSAID	non-steroidal anti-inflammatory drug
PAGE	polyacrylamide gel electrophoresis
PAH	polycyclic aromatic hydrocarbon
PCR	polymerase chain reaction
PDB	Protein Data Bank
PEG	polyethylene glycol
PETN	pentaerythritoltetranitrate
PK	peak wavelength
PG	prostaglandin
PMSF	phenylmethylsulphonyl fluoride
PPAR	peroxisome proliferator-activated receptor
RCSB	Research Collective for Structural Biology
RM	remote wavelength
RMSD	root mean square deviation
RNA	ribonucleic acid
RXR	retenoic acid receptor
SDR	short chain dehydrogenase / reductase
SDS	sodium dodecylsulphate
SeMet	selenomethionine
SRS	Synchrotron Radiation Source, Daresbury, UK
S. typhimurium	*Salmonella typhimurium*
TEMED	tetramethylethylenediamine
TI	therapeutic index
TIM	triosephosphate isomerase
TLS	translation, libration and screw correlation parameters
TNT	2,4,6-trinitrotoluene
T. thermophilus	*Thermus thermophilus*
VDW	van der Waals
V. fischeri	*Vibrio fischeri*
V. harveyi	*Vibrio harveyi*
V_m	Matthews volume

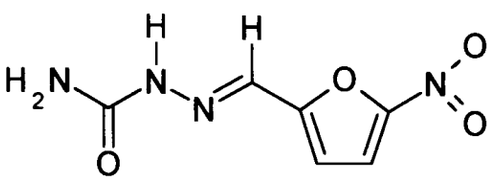
CHAPTER 1

INTRODUCTION TO NITROREDUCTASE AND ITS USE IN GENE THERAPY

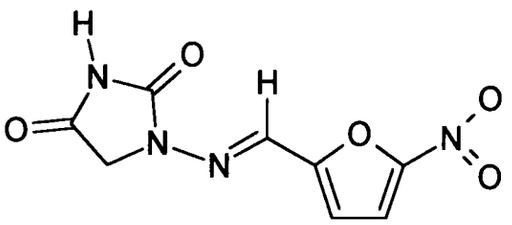
1.1 Discovery of *E. coli* Nitroreductases

Nitrofurazone (see Figure 1.1) is an antimicrobial agent (Dodd and Stillman, 1944) used to treat topical infections in burns victims. Nitrofurazone and its derivatives (e.g. nitrofurantoin) are also used in bladder and urinary tract infections. In characterising *Escherichia coli* strains that possess resistance to nitrofurazone, it was observed that these resistant strains were deficient in nitroreductase activity (McCalla et al., 1971). Upon further investigation, the nitroreductase activity of *E. coli* could be divided into an oxygen-insensitive (type I) and an oxygen-sensitive (type II) component (McCalla et al., 1975). The oxygen-sensitive enzymes are only able to catalyze the reduction of nitrocompounds under anaerobic conditions. These type II enzymes transfer two electrons to a substrate via sequential one electron steps, which are sensitive to oxygen, and can form superoxide in a futile cycle (McCalla et al., 1970). The type I enzymes are insensitive to oxygen as they transfer the two electrons to the substrate in a single step (Zenno *et al.* 1996a).

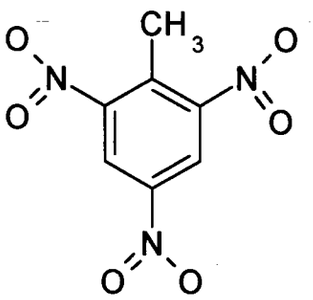
Type I nitroreductase activity was found to be responsible for nitrofurazone activation (Asnis, 1975), and can be further subdivided into one major component that uses NADPH as an electron donor (termed nfsA, (McCalla et al., 1978), and two minor components that can use either NADH or NADPH as an electron donor (termed nfsB/nfnB and nfsC, (Bryant et al., 1981). Electrons from the nucleotides are used to reduce the enzyme substrates. Later studies implicated different mutants of nfsA and nfsB in different nitrofurazone-resistant bacterial strains (Whiteway et al., 1998). Resistance to nitrofurans can often occur in a stepwise



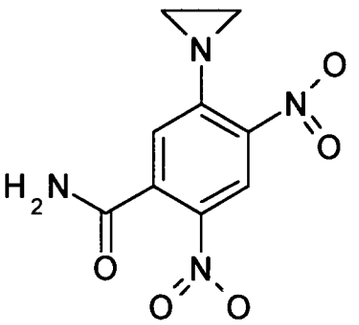
nitrofurazone



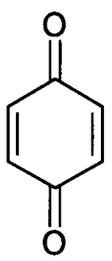
nitrofurantoin



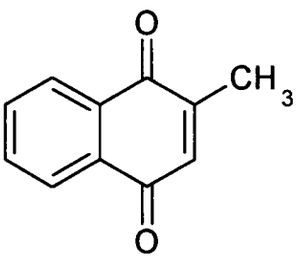
2,4,6-trinitrotoluene



CB1954



1,4-benzoquinone



menadione

Figure 1.1. Chemical structures of a variety of NTR substrates. NTR is able to reduce a broad spectrum of compounds (see text of this Chapter), usually quinones or nitroaromatics. Figure produced using ISIS draw (MDL Information Systems, San Leandro, CA).

mode of activation of nitrofurazone by nitroreductases is discussed in further detail in Chapter 7. From hereafter in the text, nfsB will be referred to as NTR.

Homologues to the type I nitroreductases of *E. coli* are found in *Salmonella typhimurium* (Watanabe *et al.* 1990) and *Enterobacter cloacae* (Bryant *et al.* 1991a), and are often referred to as "classical nitroreductases". The *S. typhimurium* enzymes are important activators of toxic compounds in the Ames test (Rosenkranz *et al.* 1981). A homologue of NTR in *Helicobacter pylori* (rdxA) is known to be mutated in strains resistant to the broad-range anti-infective drug metronidazole (Land and Johnson 1999). Metronidazole has a similar structure to nitrofurazone, possessing an aromatic five-membered ring with a nitro group.

Most of the literature concentrates on nfsA and NTR, with nfsC and the type II nitroreductases not yet mapped to the *E. coli* genome. NfsA and NTR share 7 % amino acid sequence identity, and are flavoenzymes, using FMN as the prosthetic group (Anlezark *et al.* 1992). They exhibit ping-pong kinetics (schematic in Figure 1.2), and are known to reduce a broad-range of quinones and nitroaromatics (Anlezark *et al.* 1992). These two enzymes also show homology to flavin reductases found in luminescent bacteria, nfsA with the FRP protein of *Vibrio harveyi* (Lei *et al.* 1994), and NTR with the FraseI protein of *Vibrio fischeri* (Zenno *et al.* 1994). Flavin reductases catalyze the reduction of flavins, which are then oxidised in tandem with long-chain aldehydes in bioluminescence reactions. The structure of FraseI has been determined (Koike *et al.* 1998), and shown to be structurally similar to the NADH oxidase of *Thermus thermophilus* (NOX, Hecht *et al.* 1995). A sequence alignment of this family of flavoenzymes is provided in Figure 1.3.

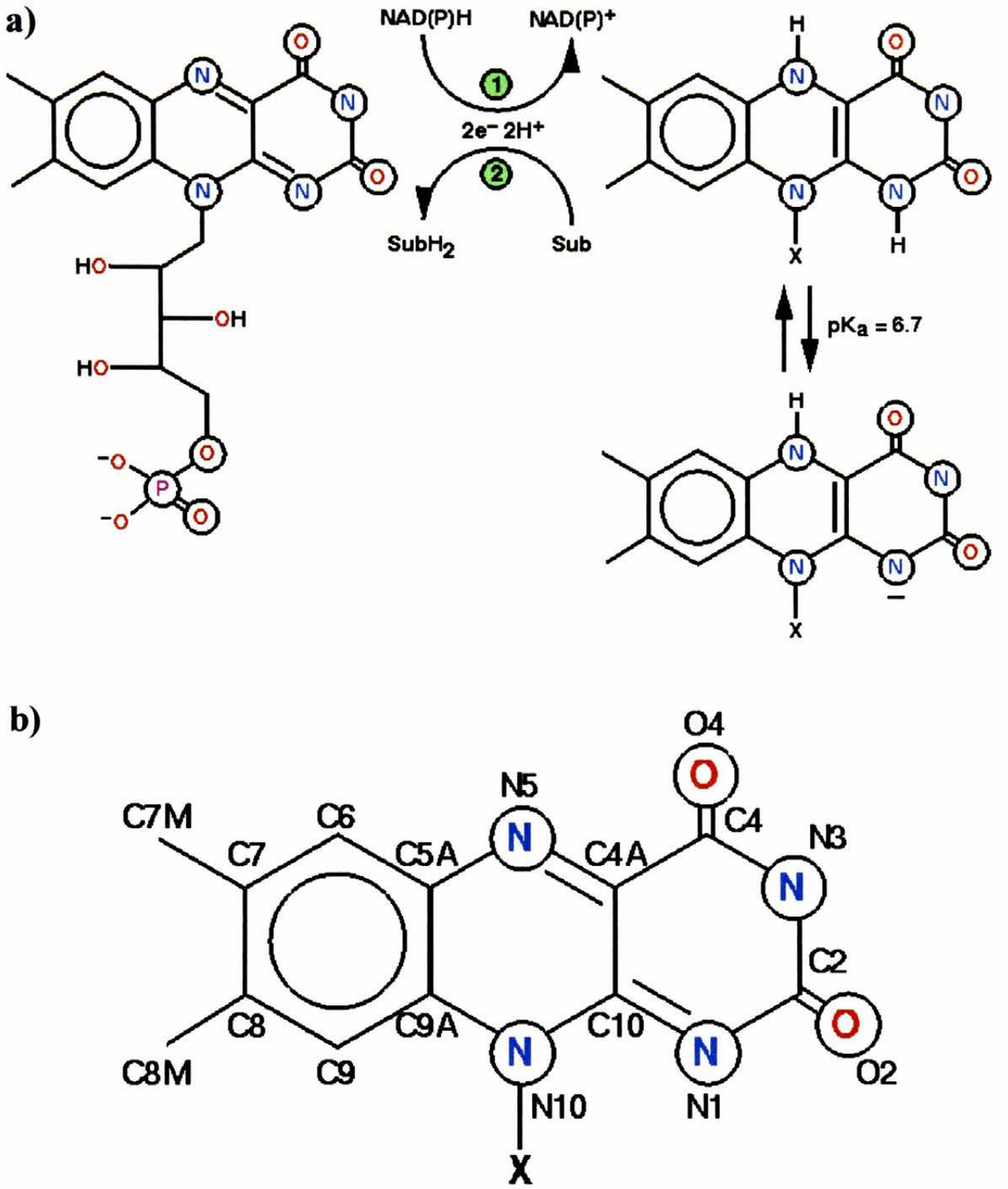


Figure 1.2. Flavin nomenclature and ping-pong mechanism. a) The reductive half (1) and oxidative (2) half-reactions of the ping-pong mechanism. In reaction 1 the flavin is reduced by NAD(P)H and alters the chemistry of the N5-C4A-C10-N1 bonding interactions. In reaction 2 the flavin then uses this reducing power to convert a substrate, Sub, into $SubH_2$, re-oxidising the flavin back to its original state. Both half-reactions involve the transfer of $2e^-$ and $2H^+$. The flavin cofactor is capable of accepting or donating one electron, and the one electron reduced species is referred to as a semiquinone. a) also shows the anionic state possible in reduced flavin. b) The numbering system of the isoalloxazine ring. X refers to the FMN ribityl or FAD dinucleotide tail.

```
        1                                                        50
entcl   ~~~~~~~MDI ISVALKRHST KAFDASKKLT AEEAEKIKTL LQYSPSSTNS
salty   ~~~~~~~MDI VSVALQRYST KAFDPSKKLT AEEADKIKTL LQYSPSSTNS
 nfsB   ~~~~~~~MDI ISVALKRHST KAFDASKKLT PEQAEQIKTL LQYSPSSTNS
frase   ~~~~~~THPI IHDLENRYTS KKYDPSKKVS QEDLAVLLEA LRLSASSINS
  nox   MEATLPVLDA KTAALKRRSI RRYRKDPVPE GLLREILEAA LR.APSAWNL

        51                                                      100
entcl   QPWHFIVAST EEGKARVAKS AAGTYVFNER KMLDASHVVV FCAKTAMDDA
salty   QPWHFIVAST EEGKARVAKS AAGNYTFNER KMLDASHVVV FCAKTAMDDA
 nfsB   QPWHFIVAST EEGKARVAKS AAGNYVFNER KMLDASHVVV FCAKTAMDDV
frase   QPWKFIVIES DAAKQRMHDS FANMHQFNQP HIKACSHVIL FANKLSYTRD
  nox   QPWRIVVVRD PATKRALREA A.....FGQA HVEEAPVVLV LYADLEDALA

        101                                                     150
entcl   WLERVVDQEE ADGRFNTPEA KAANHKGRTY FADMHRVDLK DDDQWMAKQV
salty   WLERVVDQED ADGRFATPEA KAANDKGRRF FADMHRVSLK DDHQWMAKQV
 nfsB   WLKLVVDQED ADGRFATPEA KAANDKGRKF FADMHRKDLH DDAEWMAKQV
frase   DYDVVLSKAV ADKRITEEQK EAAFASFK.. FVELNCDENG EHKAWTKPQA
  nox   HLDEVIHPGV QGER.....R EAQKQAIQRA FAAMGQEARK ...AWASGQS

        151                                                     200
entcl   YLNVGNFLLG VGAMGLDAVP IEGFDAAILD EEFGLKEKGF TSLVVVPVG.
salty   YLNVGNFLLG VAAMGLDAVP IEGFDAEVLD AEFGLKEKGY TSLVVVPVG.
 nfsB   YLNVGNFLLG VAALGLDAVP IEGFDAAILD AEFGLKEKGY TSLVVVPVG.
frase   YLALGNALHT LARLNIDSTT MEGIDPELLS EIFADELKGY ECHVALAIGY
  nox   YILLGYLLLL LEAYGLGSVP MLGFDPERVR AILGLPSR.. .AAIPALVAL

        201                   225
entcl   HHSVEDFNAT LPKSRLPLST IVTEC
salty   HHSVEDFNAG LPKSRLPLET TLTEV
 nfsB   HHSVEDFNAT LPKSRLPQNI TLTEV
frase   HHPSEDYNAS LPKSRKAFED VITIL
  nox   GYPAEE...G YPSHRLPLER VVLWR
```

Figure 1.3. Sequence alignment of NTR and homologues. Amino acids are represented by the standard single letter code. Key : entcl, classical nitroreductase of *E. cloacae*; salty, classical nitroreductase *of S. typhimurium*; nfsB (NTR), minor oxygen-insensitive nitroreductase of *E. coli*; frase, flavin reductase of *V. fischeri*; NOX, NADH oxidase of *T. thermophilus*. The sequences are numbered according to NOX. The sequence alignment was produced in GCG using PILEUP (GCG 1998).

NTR has been mapped to 13 min on the *E. coli* chromosome (Blattner *et al.* 1997), cloned and characterised (Zenno *et al.* 1996a). NTR has a molecular weight of 23.9 kDa, and can also use FAD as a cofactor, although it is associated with FMN under physiological conditions (Anlezark *et al.* 1992). The enzyme has a theoretical pI of 5.5 (calculated using protein primary sequence) and is predicted, by homology to FraseI, to be a mixed α / β structure.

1.2 The Discovery of the Anti-Cancer Drug CB1954

The drug 5-(1-aziridinyl)-2,4-dinitrobenzamide (CB1954, formulated in the Chester Beatty Laboratories, London) is rare in its ability to selectively kill tumour cells. Its anti-tumour activity was unexpected for a monofunctional alkylating agent, and had a therapeutic index (TI) of 70 against the Walker 256 carcinoma in rats (Connors *et al.* 1975). This large TI value can be compared to compounds used against cancers (cyclophosphamide TI = 22, cisplatin TI = 14). The TI is the experimental index of the safety and effectiveness of a particular compound, constituting the ratio between the toxic dose (numerator) and therapeutic dose (denominator). The cytotoxicity of CB1954 was found to be due to the formation of poorly-repaired DNA interstrand crosslinks (Roberts *et al.* 1986), which would have been expected for a difunctional alkylating agent. CB1954 could effectively eradicate the tumour cells and cure the Walker rat (Cobb *et al.* 1969).

Co-culturing of susceptible rat Walker cells with non-susceptible Chinese hamster V79 cells led to killing of both cell types after the application of CB1954 (Knox *et al.* 1988), and it was hypothesised that activation of the drug by the Walker cells was leading to its cytotoxic effects. An enzyme, upregulated in tumour cells, designated NQO1 (sometimes referred to as DT-diaphorase) was purified from Walker cells, and found to aerobically reduce CB1954

using either NADH or NADPH (Knox *et al.* 1988). The product of CB1954 reduction, a 4-hydroxylamino derivative, reacts with cellular thioesters to form the DNA reactive species (Knox *et al.* 1991). The conversion from nitro to hydroxlamine would require 4 electrons, and presumably proceed via a nitroso intermediate. Hope of using CB1954 to treat human tumours was lowered upon the discovery that the human equivalent of NQO1 had a lower activity against CB1954, with 500-5000 fold more drug needed to achieve a cytotoxic response commensurate with rat levels (Boland *et al.* 1991). To counter this, some studies are attempting to modulate CB1954 to obtain a drug optimised for human NQO1 (Knox *et al.* 1993). Another CB1954-activating enzyme is present in human tumours (termed NQO2), but requires the exogenous cofactor dihydronicotinamide riboside for catalysis (Knox *et al.* 2000).

1.3 Gene Therapy Approach to Cancer Treatment

With the failure of human NQO1 to activate CB1954, it would be necessary to introduce a foreign enzyme into tumour cells to make them susceptible to the drug. This technique is known as GDEPT (Gene-Directed Enzyme Prodrug Therapy, see Figure 1.4), and a list of GDEPT candidates proposed for use with drugs other than CB1954 is provided in Table 1.1. In a typical GDEPT approach, the DNA coding for the desired foreign enzyme is targeted to the tumour cells and integrated into the host genome. If this targeting is achieved using a retroviral vector, it is known as VDEPT (Virus-Directed Enzyme Prodrug Therapy). The tumour cell then starts expressing the foreign enzyme, and upon administration of a relatively non-toxic prodrug, converts this to a toxic species, resulting in specific tumour cell killing. In some GDEPT approaches (Nettelbeck *et al.* 2000), enzyme expression is regulated by a tumour-specific promoter.

Enzyme	Prodrug	Drug
Cytosine Deaminase	5-Fluorocytosine	5-Fluorouracil
Glucose Oxidase	Glucose	Hydrogen Peroxide
β-Glucosidase	Amygdalin	Cyanide
Carboxypeptidase A	Methotrexate-alanine	Methotrexate
Alkaline Phosphatase	Phenolmustard Phosphate	Phenolmustard

Table 1.1. Various proposals for GDEPT systems in cancer therapy.

Early studies attempting to localise therapeutic enzymes to tumour cells used antibodies (ADEPT, Antibody-Directed Enzyme Prodrug Therapy). ADEPT has the limitation of only being able to use enzymes that do not require endogenous cofactors, and also has the potential to trigger an immune response. The advantages of G/V/ADEPT over conventional chemotherapy are enormous. Traditional anti-cancer agents (e.g. cisplatin) are anti-proliferatives, and unlike G/V/ADEPT, can cause damage to non-cancerous highly-proliferating cell types including bone marrow and gut epithelia. The damage to bone marrow is particularly significant, as this weakens the immune system of the patient undergoing treatment. Anti-proliferatives also lose their effectiveness against solid tumours, where most of the cells in the tumour mass are not dividing rapidly.

Aside from localising enzymes to tumour cells, prodrugs can be activated in other ways. The conversion to a toxic species can be triggered by therapeutic radiation (Wilson *et al.* 1996), low extracellular pH in tumour masses (Tietze *et al.* 1989), tumour hypoxia (Denny *et al.* 1993) or upregulated endogenous enzymes (Traver *et al.* 1992). A summary of prodrug strategies is provided by Denny (2001).

1.4 The Use of NTR in a GDEPT System

Instead of using the Walker enzyme in a GDEPT approach, it was discovered that NTR can activate CB1954 more efficiently (Knox *et al.* 1992), with NTR possessing a k_{cat} ~ 100-fold

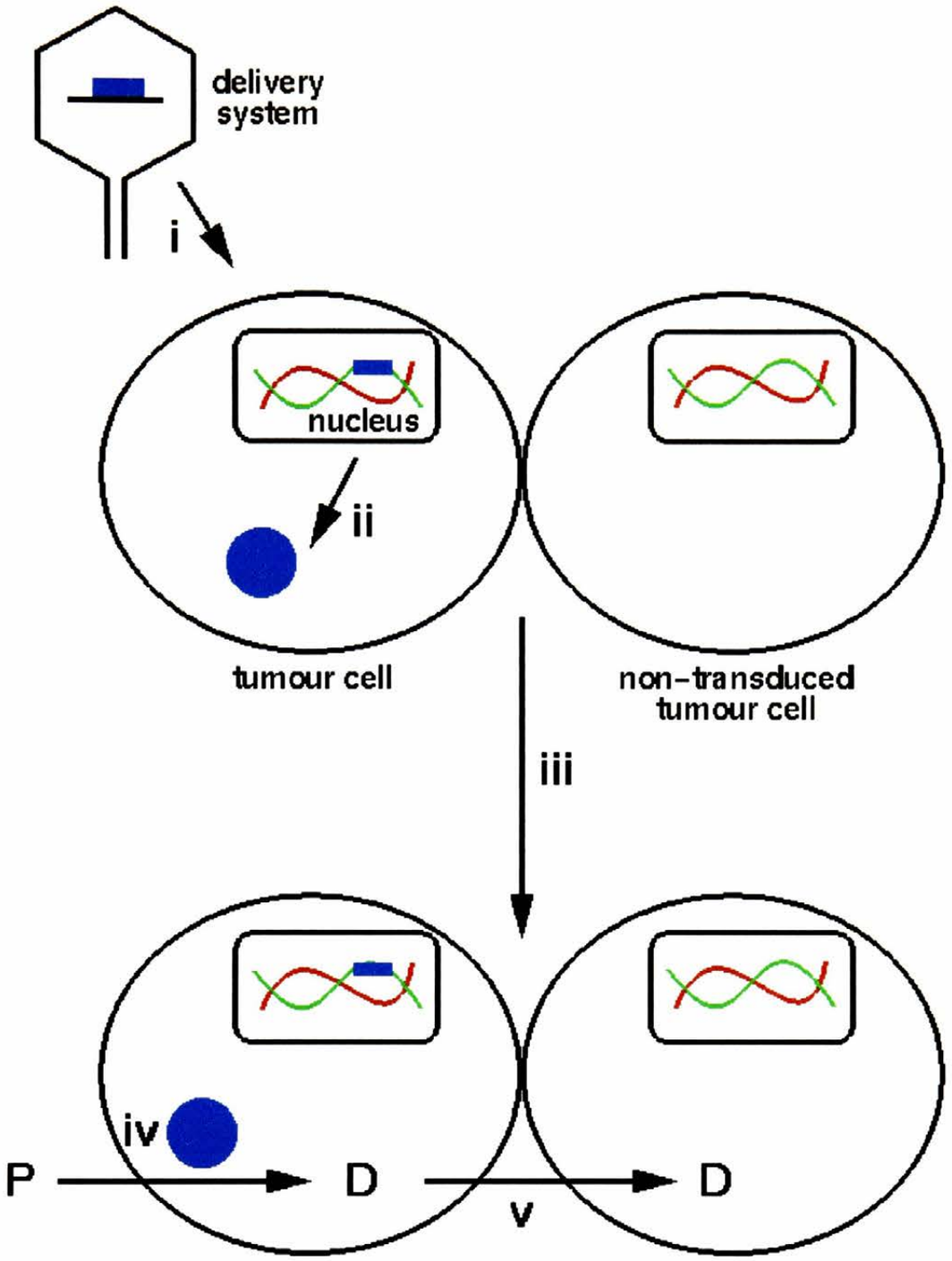


Figure 1.4. Schematic of GDEPT methodology. A gene encoding for the foreign protein is initially targeted specifically to tumour cells (i). This is most commonly achieved using an adenoviral vector. The foreign gene is then hopefully integrated into the host genome. At this stage, it is inevitable that some cells in the tumour may remain untransduced, and not contain the desirable gene. The gene is then transcribed & translated, and the transduced cells contain the foreign enzyme (ii). After successful expression of the gene, a relatively harmless prodrug (P) is introduced into circulation (iii), and is converted to a toxic drug (D) by the enzyme (iv). This will then lead to selective cell killing of the transduced tumour cells. Depending on the GDEPT system in use, the drug may diffuse over a limited distance, and kill non-transduced tumour cells (v), a process known as the "bystander effect" (Bridgewater *et al.* 1997).

higher than rat NQO1. Targeting of NTR to tumours would also allow the treatment of a wider variety of cancers than the use of an endogenous enzyme.

The kinetic advantage of NTR over rat NQO1 is offset slightly by the fact that NTR produces two products upon reduction of CB1954 (Knox *et al.* 1992, Figure 1.5). NTR generates both the toxic 4-hydroxylamine derivative, and a non-toxic 2-hydroxylamine derivative in equimolar quantities. There are three main advantages of using the NTR : CB1954 combination in a GDEPT system. Firstly, untransfected cells can be killed by a "bystander effect", where the toxic product can diffuse into the surrounding environment over a limited distance (Bridgewater *et al.* 1997, Figure 1.4 part v). A second advantage would be the limitation of damage should any enzyme reach circulation from dead cells – the lack of NAD(P)H in the blood would prevent accidental activation of bloodstream prodrug. Finally, the NTR : CB1954 GDEPT system cell killing is independent of cell cycle stage, unlike that of many other treatments e.g. the thymidine kinase activation of ganciclovir (Moolten 1986).

1.5 Aims & Objectives

Although the NTR : CB1954 combination is usable and in clinical trials (Palmer *et al.* 2000), it would be advantageous to try to improve the catalytic efficiency of the system. A relatively high K_m for CB1954 *c.f.* nitrofurazone (880 and 160 μM, respectively) and low potential maximum serum CB1954 concentration (~ 10 μM) mean that catalytic efficiency is far from optimal. It is the target of this study to obtain structural information of NTR and use this to design site-directed mutants of NTR with increased catalytic activity. If this design process proves successful, and mutants with an improved turnover of CB1954 are obtained, it would

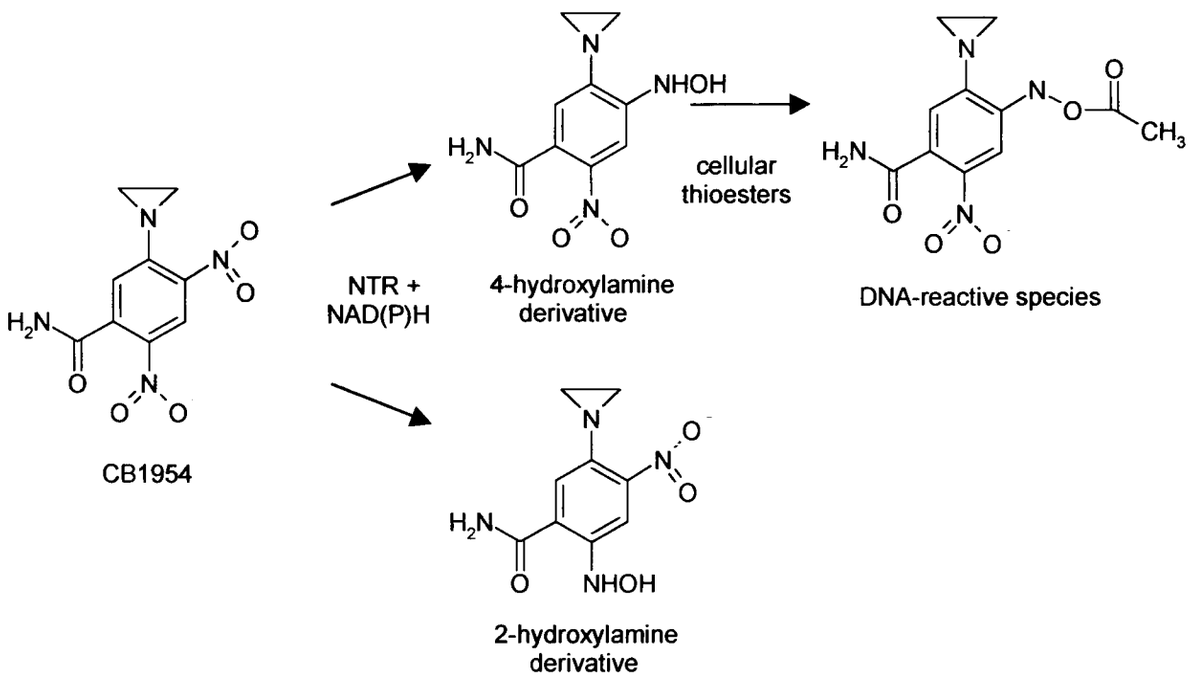


Figure 1.5. Reaction of CB1954 with NTR. Unlike NQO1, which gives only the 4-hydroxlamine derivative, NTR gives two products upon CB1954 reduction. The 2-hydroxylamine has little to no cellular toxicity, but the 4-hydroxlamine product reacts with cellular thioesters (e.g. acetyl CoenzymeA) to yield a DNA-reactive species. The reduction of CB1954 to a hydroxylamine requires 4 electrons, and is postulated to achieve this via a nitroso intermediate. It is not known whether the final 2 electrons come from enzymatic reduction or from solution reducing agents. Figure prepared using ISIS draw (MDL Information Systems).

also be desirable to possess structural information on these enzymes. This may lead to an advantageous situation of "iterative design cycles".

1.6 Physiological Role of NTR, and Use In Other Applications

Although NTR has been studied for decades, no definite knowledge of its physiological role has arisen. Recent studies have linked nfsA with redox stress (Liochev *et al.* 1999), its expression induced by the transcriptional regulator SoxR under conditions of paraquat exposure. It has been postulated that NTR may also respond to redox stress, with the discovery that NTR transcription is increased by the activator protein MarA (Barbosa and Levy 2002). MarA is known to respond to environmental hazards including antibiotics, disinfectants, organic solvents and agents of oxidative stress (Alekshun *et al.* 1997). It can be seen from the studies involving nitroreductase-deficient nitrofurazone-resistant *E. coli* that NTR is not essential for cell growth (Whiteway *et al.* 1998). There is, of course, redundancy in *E. coli* nitroreductase activity, and effects of NTR knock-outs may not be observed until all nitroreductase enzymes are disabled.

Along with the use of NTR in GDEPT, the broad substrate specificity of nitroreductases lends their use to bioremediation. Nitroaromatics have a wide distribution in the environment due to their heavy use in several industrial processes (McCartney *et al.* 1986). The NTR homologue in *E .cloacae* (nfsl) has been proposed as an enzyme to clear TNT (2,4,6-trinitrotoluene) from the environment (Bryant *et al.* 1991b). TNT-contaminated soil resulting from the manufacture, usage and disposal of explosives is both hazardous and highly toxic (Whong and Edwards 1984). TNT undergoes three sequential two electron reductions to the compound 2-amino-4-6-dinitrotoluene, which can then be further modified or conjugated, and

metabolised by other enzymes (Nishino *et al.* 2000). A test study using transgenic tobacco plants showed promising results in eliminating TNT from soil samples (Hannink *et al.* 2001). As NTR has also been shown to reduce TNT (personal communication, Dr. Graham White, Department of Biosciences, University of Cardiff), an improved NTR enzyme generated by these studies may also have relevance to the field of bioremediation.

1.7 Introduction to Flavoproteins

1.7.1 General Features of Flavoproteins

The diverse range of processes catalysed by flavoenzymes includes transfer of one or two electrons between redox centres, activation of oxygen, emission of light, and a wide variety of dehydrogenation reactions (Massey 1995). As a group of enzymes, flavoproteins are well characterised, mainly due to the accessible spectroscopic features of the flavin cofactor. Flavoproteins contain either FMN or FAD, and usually have a preference for one over the other. The protein : flavin cofactor interactions are normally non-covalent, although some enzymes modify the redox properties of the flavin via covalent linkage to the 6 or 8 position of the isoalloxazine ring (Mewies *et al.* 1998). This section will concentrate on the flavin cofactor as a redox centre.

The ubiquity of flavoproteins in redox reactions lies in their ability to transfer either one or two electrons. Oxidised flavin can accept one electron to form a semiquinone intermediate, and a further electron, in combination with two protons, to form reduced flavin (see Figure 1.2). Some flavoproteins are unable to accept or donate more than one electron e.g. ETF (electron transfer flavoprotein), the reasons why have been well documented (Scrutton *et al.* 1999). The transfer of electrons to and from the flavin can be split into a reductive

half-reaction (where the nucleotide reduces the flavin), and an oxidative half-reaction (where the flavin uses this reducing power to reduce the enzyme substrate, and regenerate oxidised flavin). The redox activity of flavins is localised in the isoalloxazine ring, the rest of the cofactor not involved in catalysis. The N1-C10-C4A-N5 locus effectively acts as an "electron sink". The redox potential of the flavin two electron reduction is approximately -200 mV in solution, but varies between –400 mV and +60 mV in flavoproteins (Ghisla and Massey 1989).

1.7.2 Flavoprotein Catalysis

The environment of the flavin cofactor is known to modulate its redox potential, and generally positive charge is thought to increase the redox potential, whereas negative charge or hydrophobic interactions are thought to decrease it (interactions occurring around the isoalloxazine ring, Massey 1995). This modulation of flavin redox potential is apparent in cases involving positively charged residues near the N1 atom, which serve to stabilise an anionic form of reduced flavin, which has a negative charge at this position (Figure 1.2, Ghisla and Massey 1989). Substrate binding to flavins may occur on either the *re* or *si* face of the isoalloxazine ring, and the other side of the cofactor is usually buried by the protein. Flavoproteins are known to have a wide variety of folds (Fraaije and Mattevi 2000), and often position the cofactor near another redox centre such as an iron-sulphur cluster (Enroth *et al.* 2000) or heme (Narhi and Fulco 1986).

The role of flavoproteins in dehydrogenation reactions has raised questions as to the nature of catalysis. This area is controversial and several competing theories exist (Ghisla and Massey 1989). The most popular theory involves direct hydride transfer between the substrate C(H)

and the flavin N5. This is supported by many crystal structures showing proximity between the N5 and C(H) (e.g. glutathione reductase, Pai and Schulz 1983). The reaction could also proceed via a radical mechanism. In a dehydrogenation reaction this would involve removal of a hydrogen atom from the substrate and the collapse of the subsequent radical pair to give a reduced flavin and oxidised product. The abstraction of the hydrogen atom would be the rate-limiting step, and this is used to validate the relevance of this particular theory on a case by case basis (Walker and Edmonson 1987). The final theory centres on the production of a carbanion by abstraction of a substrate proton from the C(H) group, as proposed by some authors for flavocytochrome b2 (Rao and Lederer 1998). This step would be catalysed by an active site base, and the carbanion would donate two electrons to the flavin, directly or via a covalent intermediate. This hypothesis is not as popular as the others mentioned above, and is favoured for "activated" substrates, such as the oxidation of acyl-CoA (Massey and Ghisla 1983). Currently, the direct hydride transfer hypothesis is supported by the majority of the academic community. A full discussion of flavoprotein mechanisms is provided in Ghisla and Massey (1988) or Fraaije and Mattevi (2000).

CHAPTER 2

INTRODUCTION TO 3α-HYDROXYSTEROID DEHYDROGENASE

2.1. The AKR Superfamily

Oxidoreductase enzymes can be placed into three superfamilies: the long-chain alcohol dehydrogenases, short-chain dehydrogenases / reductases (SDR superfamily) and the aldo-keto reductases (AKR superfamily). The long-chain alcohol dehydrogenases and SDRs are structurally distinct, the former being zinc metalloproteins, and the latter possessing a Rosmann fold for NAD(P)(H) binding. The AKRs have some catalytic residues in common with the SDRs (see end of this section), but have a totally different protein fold. SDRs are multimeric, whereas AKRs are usually monomeric (with a few exceptions : Kavanagh *et al.* 2002; Kozma *et al.* 2002), and form an $\alpha_8\beta_8$ barrel with two additional α-helices (Rondeau *et al.* 1992). AKR enzymes bind NAD(P)H in an extended conformation, using loops at the C-terminal end of the barrel (Wilson *et al.* 1992). AKRs reduce and oxidize a wide variety of compounds (Table 2.1), some of which are potential therapeutic targets.

Enzyme	AKR Identity	Substrate	Therapeutic Target
Human 3α-HSD type II	AKR1C3	Steroids	Cancer
Human aldose reductase	AKR1B1	Glucose	Diabetes
Human aldehyde reductase	AKR1A1	Neurotransmitters	Depression
Rat aflatoxin reductase	AKR7	Aflatoxins	-
G. echinata polyketide reductase	AKR4A3	Pro-antibiotics	-
Mouse vas deferens protein	AKR1B7	Androgens	-
Bovine prostaglandin F synthase	AKR1C7	Prostaglandin D_2	-

Table 2.1. Diversity of the AKR superfamily.

AKRs can be found in mammals, amphibians, plants, yeast, protozoa and bacteria (Jez *et al.* 1997). To avoid confusion, the naming of these enzymes is similar to that used for the

cytochrome P_{450} family, with the prefix AKR followed by a number to indicate the subfamily, with further letters and numbers to identify subgroups (Jez *et al.* 1996a). The subgroups are named with reference to human liver aldose reductase. All the AKR enzyme structures have the $\alpha_8\beta_8$ fold, and several groups of conserved residues have been shown to be important (Table 2.2).

Conserved Residues	Function
D50, N167, Q190, S271, R276	Nucleotide cofactor binding
D50, Y55, K84, H/E117*	Active site geometry / catalysis
G22, G45, D112, P119, G164, P186 (always conserved)	Barrel formation
G20, G62, L113, W148, G158, E192 (usually present)	

Table 2.2. Conserved residues in the AKR superfamily. *H for carbonyl reduction, E for C=C reduction.

Positioning of the conserved active site residues with respect to the cofactor is shown in Figures 2.1 & 2.2. Mutagenesis studies and kinetic data suggest Y55 to be the general acid or base (Petrash *et al.* 1994). Three possible mechanisms for catalysis exist. Using carbonyl reduction as an example, in the first hypothesis proton donation from an active site acid is the first step, generating a carbocation, followed by hydride transfer from the nucleotide to yield the hydroxy product. A second mechanism would involve hydride transfer occurring first, generating an oxyanion, followed by proton donation from the active site acid. The third hypothesis follows a concerted model, with the two steps occurring concurrently. Solvent isotope effect measurements favour an oxyanion intermediate (Penning *et al.* 2001), and the geometry of the His-Tyr-Lys active site cluster can be likened to the catalytic triad and oxyanion hole of the serine proteases (Dodson and Wlodawer 1998). The SDR family of enzymes possess a Tyr-X-X-X-Lys motif at the active site, and the identity of these with the AKR conserved residues provides an example of convergent evolution. Some members of the AKR family retain the structural features mentioned above, but are not yet shown to possess

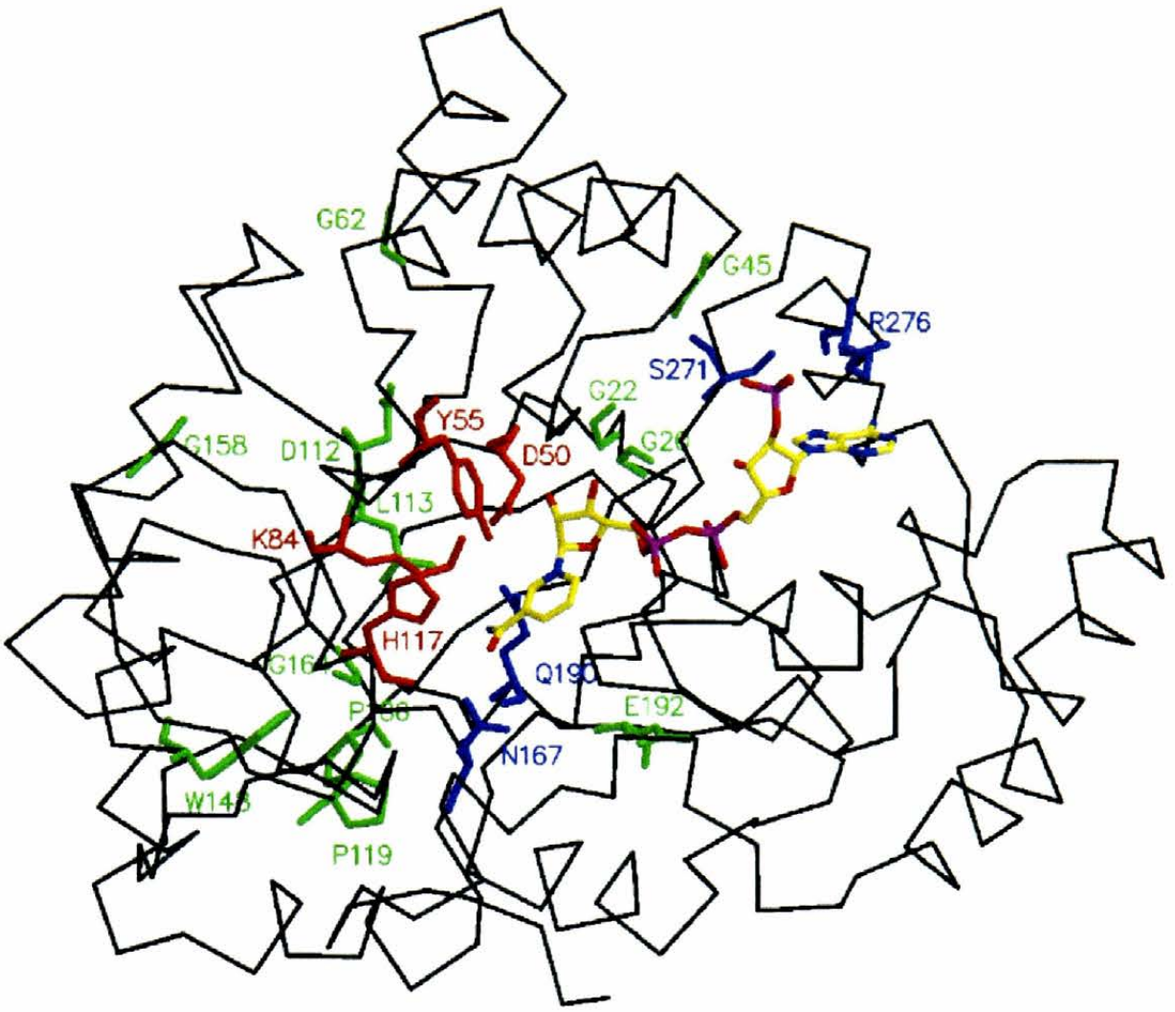


Figure 2.1. Positioning of groups of conserved AKR residues. The figure was constructed using the AKR1C3 : $NADP^+$: acetate ternary complex (see Chapter 8). Nucleotide represented as stick model, coloured according to atom type : C, yellow; P, magenta; O, red; N, blue. Residues represented as stick model, coloured according to conservation type : nucleotide binding, blue; active site geometry & catalysis, red; barrel formation, green. Residues labelled using standard single letter code and sequence residue number. Protein Cα trace in dark grey. Figure prepared using MOLSCRIPT (Kraulis 1991) and RASTER 3D (Meritt and Bacon 1997).

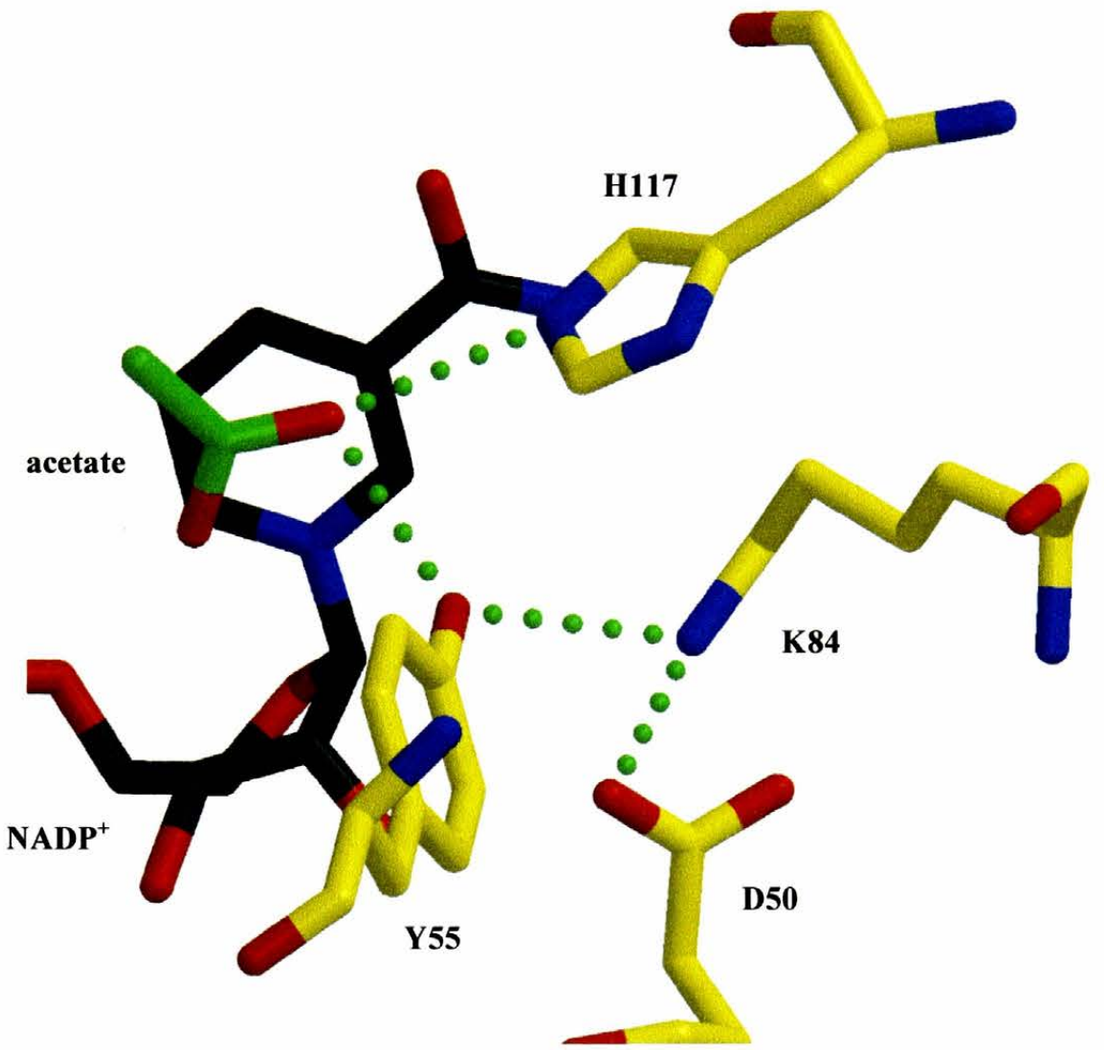


Figure 2.2. Arrangement of AKR active site residues. The figure was constructed using the AKR1C3 : $NADP^+$: acetate ternary complex (see Chapter 8). Atoms coloured according to residue and atom type : $NADP^+$ C, grey ; acetate C, green ; protein C, yellow ; O, red ; N, blue. Hydrogen bonding shown as dotted green lines. Figure prepared using MOLSCRIPT (Kraulis 1991) and RASTER 3D (Meritt and Bacon 1997).

catalytic activity, e.g. frog ρ-crystallin (Tomarev *et al.* 1984) and the β-subunit of the rat K^+ shaker channel (McCormack and McCormack 1994).

2.2 Hydroxysteroid Dehydrogenases of the AKR Superfamily

In humans, there are four AKR 3α-hydroxysteroid dehydrogenases (HSDs), named AKR1C1-4. These enzymes share a high degree of homology (Table 2.3, Figure 2.3) and have functional plasticity, using 3α, 17β and 20α-hydroxysteroids as substrates (Figure 2.4). AKR1C4 is expressed solely in the liver, but AKR1C1-3 have a wide tissue distribution (Penning *et al.* 2000). AKR1C1 is primarily a 20α-HSD and converts progesterone to the inactive steroid 20α-hydroxyprogesterone. AKR1C2 is distinguished from the other 3α-HSDs by its high affinity for bile acids and may be involved in bile acid transport. Most of the literature focuses on AKR1C2 involvement in 5α-dihydrotestosterone (5α-DHT) production and abnormal prostate growth (Gormley 1991). AKR1C3 expression is highest in the prostate and mammary gland, and has a largely reductive role, converting androgens to inactive metabolites (Penning *et al.* 2001). AKR1C3 also reduces estrone to the more active 17β-estradiol, and AKR1C3 inhibitors have been proposed for the treatment of hormonally-dependent breast cancer (Penning *et al.* 2001). The 3α-HSD enzymes are also known to activate polycyclic aromatic hydrocarbon (PAH) proximate carcinogens, and have been shown to be overexpressed in lung carcinomas (Palackal *et al.* 2002). Some role for 3α-HSDs may also exist in neurosteroid synthesis (Usami *et al.* 2002).

```
          1                                                            50
akr1c1    MDSKYQCVKL NDGHFMPVLG FGTYAPAEVP KSKALEATKL AIEAGFRHID
akr1c2    MDSKYQCVKL NDGHFMPVLG FGTYAPAEVP KSKALEAVKL AIEAGFHHID
akr1c3    MDSKQQCVKL NDGHFMPVLG FGTYAPPEVP RSKALEVSKL AIEAGFRHID
akr1c4    MDPKYQRVEL NDGHFMPVLG FGTYAPPEVP RNRAVEVTKL AIEAGFRHID
akr1a1    ~~MAASCVLL HTGQKMPLIG LGTW.KSEPG QVKA..AVKY ALSVGYRHID

          51                                                          100
akr1c1    SAHLYNNEEQ VGLAIRSKIA DG.SVKREDI FYTSKLWCNS HRPELVRPAL
akr1c2    SAHVYNNEEQ VGLAIRSKIA DG.SVKREDI FYTSKLWSNS HRPELVRPAL
akr1c3    SAHLYNNEEQ VGLAIRSKIA DG.SVKREDI FYTSKLWSTS HRPELVRPAL
akr1c4    SAYLYNNEEQ VGLAIRSKIA DG.SVKREDI FYTSKLWCTF FQPQMVQPAL
akr1a1    CAAIYGNEPE IGEALKEDVG PGKAVPREEL FVTSKLWNTK HHPEDVEPAL

          101                                                         150
akr1c1    ERSLKNLQLD YVDLYLIHFP VSVKPGEEVI PKDENGKILF DTVDLCATWE
akr1c2    ERSLKNLQLD YVDLYLIHFP VSVKPGEEVI PKDENGKILF DTVDLCATWE
akr1c3    ENSLKKAQLD YVDLYLIHSP MSLKPGEELS PTDENGKVIF DIVDLCTTWE
akr1c4    ESSLKKLQLD YVDLYLLHFP MALKPGETPL PKDENGKVIF DTVDLSATWE
akr1a1    RKTLADLQLE YLDLYLMHWP YAFERGDNPF PKNADGTICY DSTHYKETWK

          151                                                         200
akr1c1    AVEKCKDAGL AKSIGVSNFN RRQLEMILNK PGLKYKPVCN QVECHPYFNQ
akr1c2    AMEKCKDAGL AKSIGVSNFN HRLLEMILNK PGLKYKPVCN QVECHPYFNQ
akr1c3    AMEKCKDAGL AKSIGVSNFN RRQLEMILNK PGLKYKPVCN QVECHPYFNR
akr1c4    VMEKCKDAGL AKSIGVSNFN CRQLEMILNK PGLKYKPVCN QVECHPYLNQ
akr1a1    ALEALVAKGL VQALGLSNFN SRQIDDILSV ASV..RPAVL QVECHPYLAQ

          201                                                         250
akr1c1    RKLLDFCKSK DIVLVAYSAL GSHREEPWVD PNSPVLLEDP VLCALAKKHK
akr1c2    RKLLDFCKSK DIVLVAYSAL GSHREEPWVD PNSPVLLEDP VLCALAKKHK
akr1c3    SKLLDFCKSK DIVLVAYSAL GSQRDKRWVD PNSPVLLEDP VLCALAKKHK
akr1c4    SKLLDFCKSK DIVLVAHSAL GTQRHKLWVD PNSPVLLEDP VLCALAKKHK
akr1a1    NELIAHCQAR GLEVTAYSPL GSS.DRAWRD PDEPVLLEEP VVLALAEKYG
          251                                                         300
akr1c1    RTPALIALRY QLQRGVVVLA KSYNEQRIRQ NVQVFEFQLT SEEMKAIDGL
akr1c2    RTPALIALRY QLQRGVVVLA KSYNEQRIRQ NVQVFEFQLT SEEMKAIDGL
akr1c3    RTPALIALRY QLQRGVVVLA RSYNEQRIRQ NVQVFEFQLT AEDMKAIDGL
akr1c4    RTPALIALRY QLQRGVVVLA KSYNEQRIRE NIQVFEFQLT SEDMKVLDGL
akr1a1    RSPAQILLRW QVQRKVICIP KSITPSRILQ NIKVFDFTFS PEEMKQLNAL

          301                                  333
akr1c1    NRNVRY.... LTLDIF.... .AGPPNYPFS DEY
akr1c2    NRNVRY.... LTLDIF.... .AGPPNYPFS DEY
akr1c3    DRNLHY.... FNSDSF.... .ASHPNYPYS DEY
akr1c4    NRNYRY.... VVMDFL.... .MDHPDYPFS DEY
akr1a1    NKNWRYIVPM LTVDGKRVPR DAGHPLYPFN DPY
```

Figure 2.3. Sequence alignment of AKR human HSD enzymes (AKR1C family) and human aldehyde reductase (AKR1A1). Amino acids are represented by the standard single letter code, numbering is with respect to AKR1C3. Gaps in alignment indicated by full stops. Alignment produced using PILEUP feature of GCG package (GCG 1998).

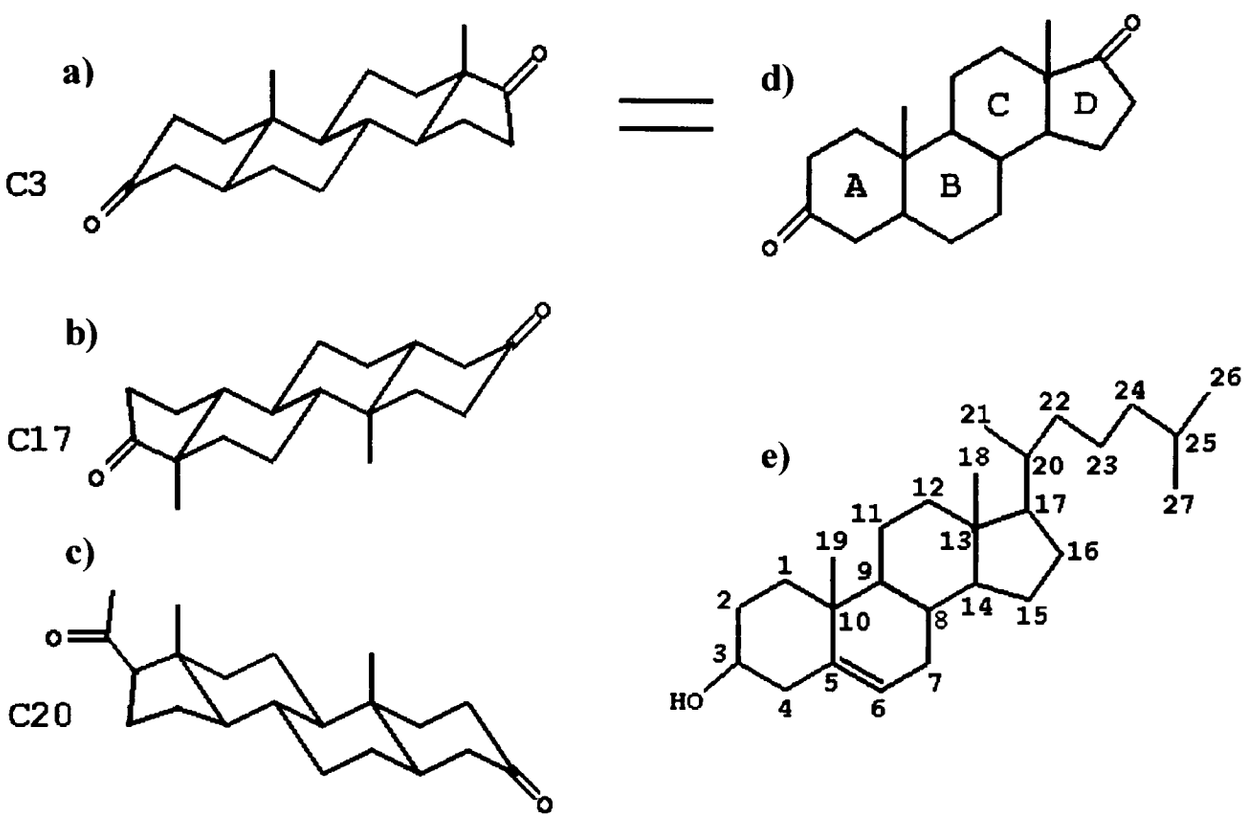


Figure 2.4. Steroid numbering system and positions for reduction / oxidation by the HSD enzymes. a), b) and c) : the relative steroidal orientations for C3, C17 and C20 oxidoreduction. The C3 and C20 orientations are flipped relative to one another 180° around the vertical axis, whereas the C3 and C17 orientations are flipped 180° in both the horizontal and vertical axes. For enzymes possessing more than one type of HSD activity, some plasticity in steroid binding is required. d) the ring classification of steroids. e) standard steroid numbering system. a) to d) show alternative representations of the molecule testosterone, and e) shows cholesterol.

Protein	% Identity to AKR1C3	Residue (in steroid binding pocket)													
		52	54	55	86	117	118	120	128	129	226	227	306	308	310
AKR1C3	100	**A**	L	**Y**	**W**	**H**	S	M	L	S	R	**W**	F	S	S
AKR1C1	86	**A**	L	**Y**	**W**	**H**	F	V	V	I	P	**W**	L	L	I
AKR1C2	88	**A**	V	**Y**	**W**	**H**	F	V	V	I	P	**W**	L	L	I
AKR1C4	84	**A**	L	**Y**	**W**	**H**	F	M	P	L	L	**W**	V	M	F

Table 2.3. Variability in AKR1C subfamily. Residues in bold are invariant.

2.3 HSDs and Prostaglandins

Prostaglandins are short-lived signalling molecules generated from the fatty acid arachidonate by cyclo-oxygenases, and are implicated in such diverse processes as inflammation, thrombosis, blood pressure regulation and cancer (Versteeg *et al.* 1999). The chemical structure of a variety of prostaglandins and related compounds is given in Figure 2.5. Synthesis of prostaglandins is dependent on the release of arachidonic acid from membranes by the action of phospholipase A_2 / C, which is then converted to prostaglandin H_2 by cyclo-oxygenases (see Section 2.4). Prostaglandins belong to a group of compounds known as the eicosanoid hormones, all of which are oxygenated 20 carbon fatty acids. Prostaglandins (PGs) have a cyclopentane ring, prostacyclins have two fused, five-membered rings, thromboxanes have a cyclohexane ring, and leukotrienes have three conjugated double bonds.

PGF_2 has a variety of functions including muscle contraction and pain transmission (Shimizu and Wolfe 1990). PGF_2 is derived from PGD_2, and upon the cloning of PGF synthase from bovine lung, the enzyme was revealed to be part of the AKR superfamily (Watanabe *et al.* 1988). In an attempt to clone the human counterpart of PGF synthase, it was discovered to be identical to the HSD enzyme AKR1C3 (Suzuki-Yamamoto *et al.* 1999). This raised important questions as to the primary physiological role of AKR1C3, the k_{cat}/K_m (the specificity constant) for its "traditional" steroidal substrate 5β-androsetene-3,17-dione being ~3.5 % of

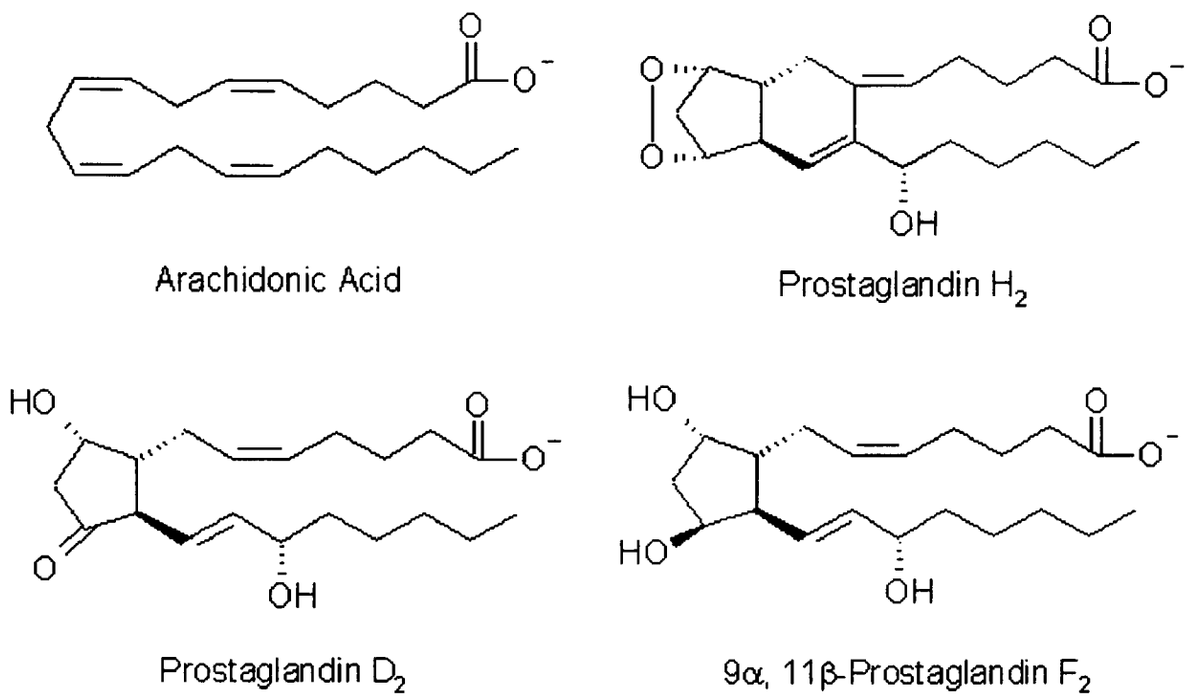


Figure 2.5. Chemical structures of the prostaglandin precursor, arachidonic acid, and various prostaglandins. All contain 20 carbon atoms, and the stereospecificity of the OH substituents and certain C-C bonds is particularly important in prostaglandin function and recognition. Figure produced using ISIS draw (MDL Information Systems).

that for PGD_2 (Suzuki-Yamamoto *et al.* 1999). The association between AKR enzymes and PGs had also been seen with rat liver 3α-HSD (AKR1C9), where PGs and arachidonic acid inhibited the oxidation of androsterone (Penning and Talalay 1983).

2.4 Cyclo-oxygenases and Nonsteroidal Anti-Inflammatory Drugs

As mentioned earlier, the primary synthesis of prostaglandins depends on cyclo-oxygenase (COX) activity. There are two enzyme isoforms (COX-1 and COX-2, 60 % amino acid identity), which show different modes of regulation. COX-1 is constitutively expressed, and is thought of as a "housekeeping" enzyme, involved in platelet aggregation and protection of the gastric mucosa (Oneill and Fordhutchinson 1993). COX-2 is induced by cytokines, oncogenes and growth factors, and is associated with inflammation (Dewitt and Smith 1995). The two isoforms also exhibit differential expression, COX-1 localised to the endoplasmic reticulum and plasma membrane, COX-2 localised to the nuclear envelope (Morita *et al.* 1995). COX-1 and 2 are found in most, if not all mammalian tissues.

Both COX enzymes contain a heme cofactor and catalyse two reactions : a cyclo-oxygenase step that uses arachidonic acid and two O_2 molecules to form PGG_2; and a second peroxidase step that reduces PGG_2 to PGH_2. Important active site residues have been identified from structures of COX-1 (Picot *et al.* 1994), COX-2 (Luong *et al.* 1996) and an arachidonic acid complex of COX-1 (Malkowski *et al.* 2000). The active site geometry is radically different to that of the AKRs, as would be expected given the sequence and cofactor differences. A comparison of the AKR, SDR and COX protein folds is provided in Figure 2.6. With their central role in inflammation and signalling, efforts have been made to inhibit the COX enzymes for therapeutic gain in pain relief and arthritis. As COX-1 is somewhat beneficial in

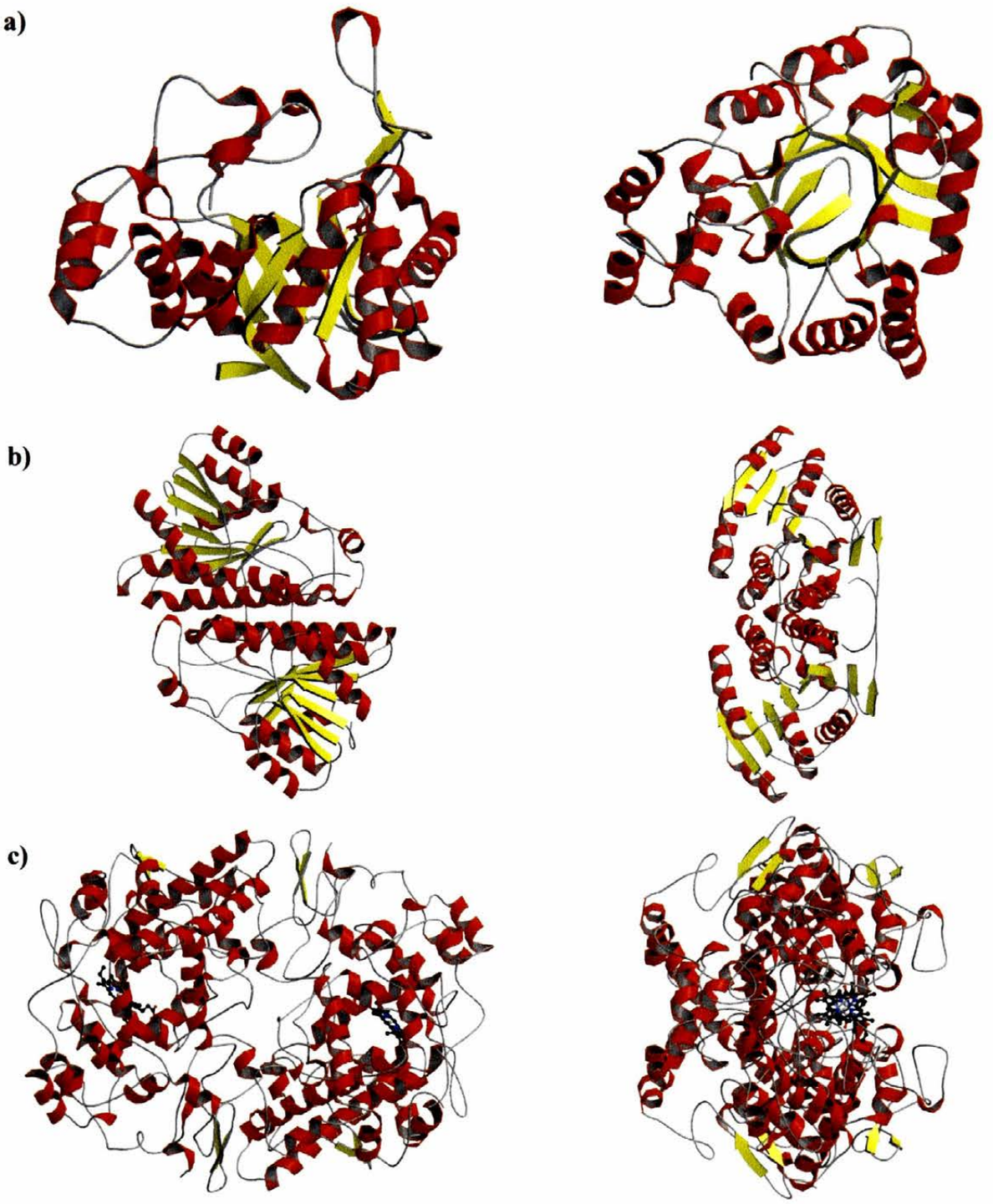


Figure 2.6. Differing folds of the AKR, SDR and COX protein families. Left hand column view is at 90° to right hand column view. Diagram is coloured according to secondary structure : α-helices in red and β-strands in yellow. a) AKR pig aldose reductase (PDB code 1AH4, Urzhumtsev *et al.* 1997), b) SDR *Drosophila lebanonesis* alcohol dehydrogenase (PDB code 1B16, Benach *et al.* 1999), c) the COX-2 isoform from humans (PDB code 6COX, Kurumbail *et al.* 1996). The SDR and COX examples shown are dimeric, and the AKR example monomeric. Figure prepared using MOLSCRIPT (Kraulis 1991) and RASTER3D (Meritt and Bacon. 1997).

the prevention of gastric ulcers (Robert 1983), and COX-2 is labelled as the main protagonist due to its inducible expression, COX-2 specific inhibitors have also been identified (Prescott and Fitzpatrick 2000).

Nonsteroidal anti-inflammatory drugs (NSAIDs) have long been used in pain alleviation and the treatment of inflammation, the most well-documented of which is aspirin. The chemical structure of a selection of NSAIDs is given in Figure 2.7. In 1982, the Nobel prize was awarded to Bergstrom, Samuelson and Vane for the elucidation of the target of NSAID action – discovered to be the step of prostaglandin formation. The particular action of aspirin is to acetylate and inactivate COX-1 and COX-2 (Vane 1971). The effect of other NSAIDs is also to inhibit COX activity, and further COX : NSAID structures have been solved (Kurumbail *et al.* 1996; Loll *et al.* 1996).

2.5 Cancers & The Protective Effect of NSAIDs

Colorectal cancer is the third most common of all cancer deaths, and the incidence of it is increasing (WHO 1997). Most colorectal cancers involve a progression of steps (see Figure 2.8). The tumour mass initially forms an adenoma, eventually becoming metastatic whereupon it is classified as an adenocarcinoma. The individual steps in this process involve the mutation of several key "gate-keeper" genes including p53, Ki-ras, DNA mismatch repair enzymes and the adenomatous polyposis coli (APC) gene. Many studies have established that NSAIDs have a protective effect against colon cancer (Shiff and Rigas 1997), and sulindac was the first drug demonstrated to regress colonic polyps in familial adenomatous polyposis (Takayama *et al.* 1998). Celecoxib, a selective COX-2 inhibitor, is given as a cancer preventative agent to individuals known to possess a faulty APC gene (Tegeder *et al.* 2001).

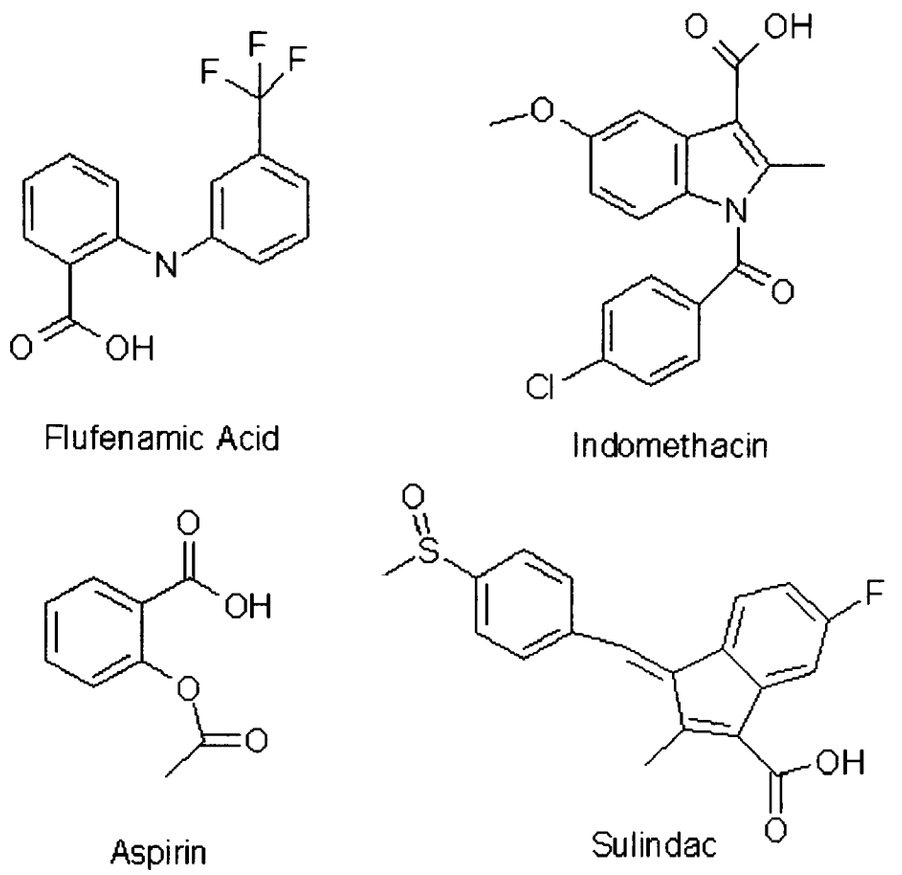


Figure 2.7. Chemical structures of various NSAIDs. Figure produced using ISIS draw (MDL Information Systems).

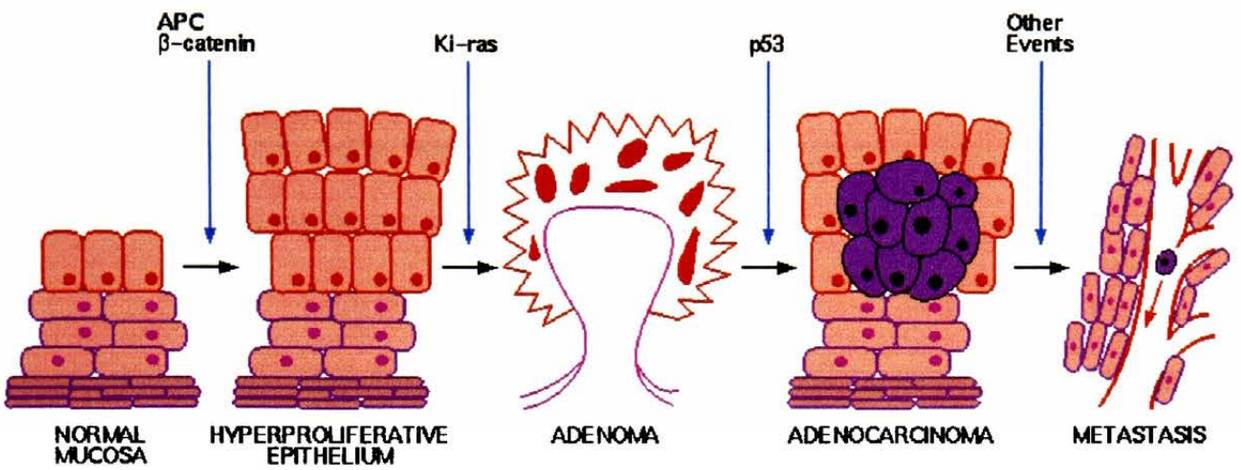


Figure 2.8. Progression of colorectal cancer. Individual "gate-keeper" mutations in protective genes labelled by blue arrows, progressive stages labelled by black arrows. Figure adapted from Shiff and Rigas 1999.

The anti-tumour activity of NSAIDs is not limited to colon cancer, with chemopreventative activity shown against both lung and breast cancer (Fosslien 2000), and NSAID-induced apoptosis of human myeloid leukaemia has been demonstrated *in vitro* (Klampfer *et al.* 1999). This is particularly relevant as tumours of myelopoietic bone marrow show a similar stepwise progression to those seen in the colon.

The majority of studies link the anti-oncogenic effect of NSAIDs with their ability to inhibit the COX enzymes (Vane 1971). Studies of colon cancer do indeed show increased levels of COX-2 and PGE_2 (Shiff *et al.* 1997). Aside from this finding, several other factors encourage the link between NSAID protection and COXs : 1) PGs are known to reduce apoptosis (PGE_2 induces Bcl-2), and overexpression of COX-2 in colon cells increases their life span (DuBois *et al.* 1996) ; 2) in APC knockout mice the COX-2 $^{-/-}$ mice develop fewer tumours than COX-2 $^{-/+}$ mice, which have fewer still than COX-2 $^{+/+}$ individuals (Oshima *et al.* 1996) ; 3) COX-2 induces angiogenesis (Tsujii *et al.* 1998) ; 4) free arachidonic acid has been shown to promote apoptosis (Surette *et al.* 1996), and NSAID inhibition of COX would elevate cellular levels ; and 5) inflammation is a known risk factor for several types of cancer (Prescott *et al.* 2000).

There is, however, a growing body of evidence that the chemoprotective effects of NSAIDs are independent of COX activity. The most prominent of these was the realisation that much higher doses of NSAIDs were required to show an anti-oncogenic effect than that needed to inhibit PG synthesis by COX (Flower *et al.* 1980). Other studies also showed that NSAIDs could inhibit tumour formation in COX-deficient cell lines (Hanif *et al.* 1996; Zhang *et al.* 1999). Some NSAIDs are not COX inhibitors (e.g. sulindac sulphide), yet induce the cellular

apoptotic response (Thompson *et al.* 1997). These COX-independent activities of NSAIDs have been suggested to elicit their effects via the transcription factors NF-κB and AP-1 (Tegeder *et al.* 2001), but as the proven chemoprotective NSAID indomethacin is unable to bind these, some doubts remain. To complicate the situation further, some NSAIDs (indomethacin included) can activate the peroxisome proliferator-activated receptor γ ($PPAR_\gamma$), and others such as salicylate cannot (Schoonjans *et al.* 1996). Loss of $PPAR_\gamma$ function is associated with some human colon cancers (Sarraf *et al.* 1999). $PPAR_\gamma$ is known to form complexes with retinoid X receptors (RXR), and have an influence over cellular proliferation and differentiation (Clay *et al.* 2000). Evidence for $PPAR_\gamma$ being an effector of NSAID-induced chemoprotection is given by Wick (Wick *et al.* 2002). The interplay of COX-dependent and independent effects and the different NSAID types in cancer prevention is obviously highly complex, and may involve a combination of a variety of the proposed pathways.

2.6 AKR1C3 as a Candidate for NSAID Effects

The HL60 cell line (derived from human myeloid leukaemia) is a well-studied system used to investigate cellular differentiation. The sensitivity of HL60 cell cultures to the differentiation agent all-trans retinoic acid (ATRA) was increased with the addition of indomethacin and the steroid hormone dexamethasone (Bunce *et al.* 1994). These, and other compounds that increased differentiation, were all found to be inhibitors of 3α-HSD enzymes (Bunce *et al.* 1996). It was concluded that inhibition of a 3α-HSD was leading to the differentiation effects, and this 3α-HSD was shown to be AKR1C3 (Bunce *et al.* 1994). Figure 2.9 details the proposed mechanism by which NSAIDs and other inhibitory compounds effect AKR1C3

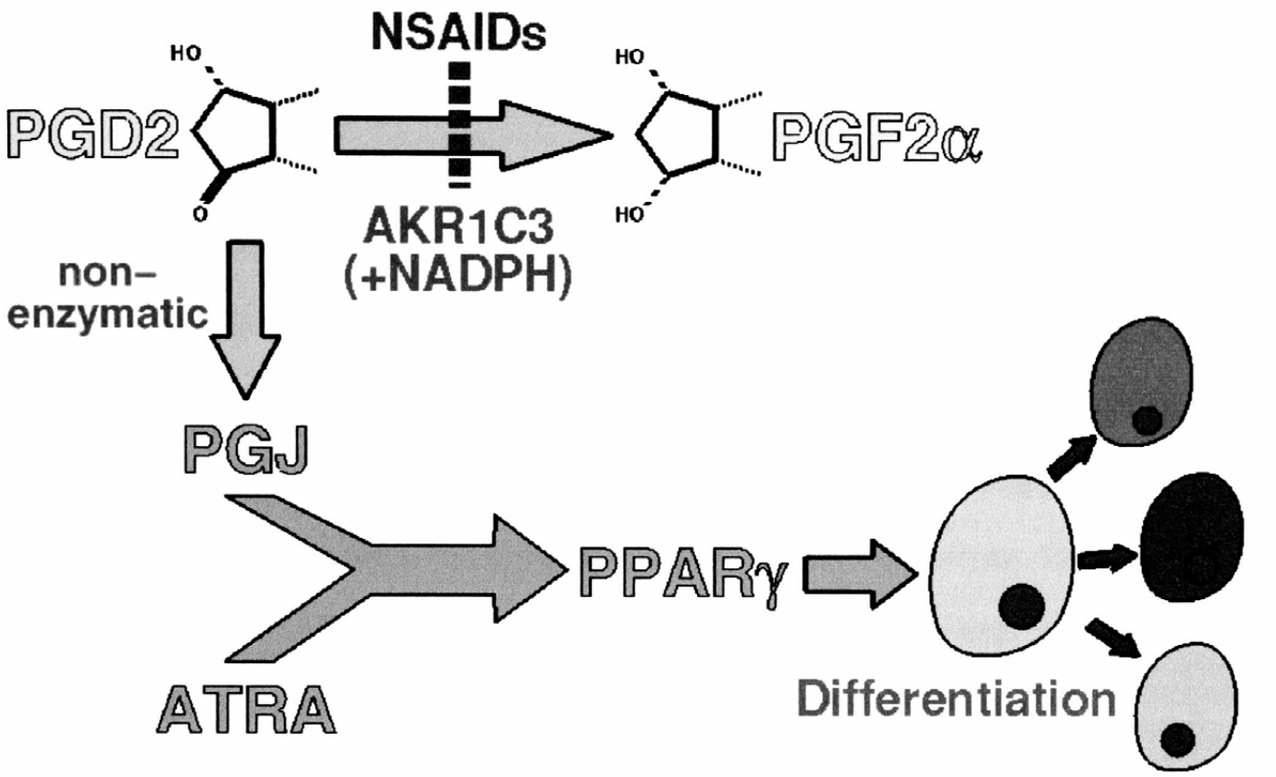


Figure 2.9. Proposed mechanism for AKR1C3 involvement in NSAID chemoprotection. Inhibition of AKR1C3 reductase activity by NSAIDs leads to an increase in cellular levels of prostaglandin D_2. Non-enzymatic conversion of prostaglandin D_2 to prostaglandin J occurs, and this can activate PPARγ In tandem with differentiation agents like all-trans retenoic acid. Downstream events of PPARγ activation (such as cadherin-E expression, Ohta *et al.* 2002) then lead to the desired chemoprotective effects, including differentiation of tumour cells. The scenario between this proposed COX-independent mechanism, and other COX-independent mechanisms remains complex however, an example being the direct binding of PPARγ by some NSAIDs (see main body of text).

activity and promote differentiation (personal communication, C.M. Bunce, School of Biosciences, University of Birmingham).

2.7 Aims of the Project

It is the purpose of this study to obtain structural information on AKR1C3 and use this to examine the functional plasticity between the AKR1C1-4 subfamily members. If possible, it would be desirable to look at NSAID binding and relate this to the COX studies, with an eventual target of acquiring enough information to allow the design of AKR1C3-specific NSAIDs. Such compounds would hopefully be useful in differentiation therapy, treating some hormonally-dependent cancers, and assist mechanistic studies on the enzyme. Furthermore, AKR1C3-specific NSAIDs could also provide a means for determining what extent of the chemoprotective activity of NSAIDs may be attributed to AKR1C3 inhibition.

CHAPTER 3

INTRODUCTION TO X-RAY CRYSTALLOGRAPHY

3.1 Foreword

In tandem with mechanistic studies on proteins, it is desirable to find out about the precise distribution of the amino acids in the protein structure and use this information to explain how that protein functions. Since the small size of proteins limits studies using longer wavelength light (microscopy of large complexes excluded), a wavelength appropriate to the features of the molecule studied would need to be used. Due to the distance between atoms being on the Ångstrom scale, x-ray wavelengths (0.1 to 100 Å) are used for visualising protein structure.

With proteins as small as they are, a single protein molecule would not be able to scatter a detectable amount of x-rays. It had been previously observed that many different types of crystals diffracted x-rays well (the idea being that their extraordinary external appearance reflected some internal regularity being a correct assumption). This led to the hypothesis that a protein crystal would prove similarly useful. Protein crystals were known to exist, with haemoglobin crystallizing directly from earthworm blood (studies by Hunefeld in 1840), and excelsin storage protein from brazil nuts also able to crystallize (studies by Hartig in 1855). Studies on early crystalline samples proved disappointing as the proteins were treated in the same manner as small molecules, and the drying out process destroyed the crystal order. However, in 1934, Bernal & Crowfoot (Bernal *et al.* 1934), kept their protein crystal hydrated whilst collecting data and the first protein crystal diffraction pattern was observed. From this starting point, it took many years before the diffraction patterns could be analysed and used in structure determination.

3.2 Theory of Diffraction

With a crystal consisting of repeating units arranged in an essentially infinite array (truncated by an exhaustion of supply or poisoning of the crystal growth faces), it can be thought of as a series of regular lattice planes that dissect the unit cell. The unit cell is defined as the smallest portion of the crystal that can recreate the crystal itself when translational symmetry is applied in three dimensions. The interplay of crystal lattice planes, x-rays and the resultant diffraction pattern is explained by Bragg's Law, a pictorial description of which is given in Figure 3.1.

The diffraction pattern consists of an array of reflections arising from the fulfillment of Bragg's Law, with their position relative to the scattering angle 2θ (as this increases so does the resolution, as the spacing between the planes d decreases), and the intensity of the diffraction spot is proportional to the electron density of the scattering atoms of that particular Bragg plane. The crystal and its contents are said to be in real space, and the diffraction pattern in reciprocal space. It is useful to separate the two this way since the real space electron density is derived from the reciprocal space diffraction pattern. As Bragg's Law is rearranged to $\sin\theta = n\lambda/2 \times 1/d$, it can be seen that d is inversely proportional to $\sin\theta$. A decrease in d would lead to an increase in θ (seen from the existence of high resolution spots at larger scattering angles), and also the shortening of distance between lattice points in reciprocal space when dealing with larger unit cells in real space. In describing vectors in the unit cell, real space uses a, b, c and reciprocal space uses a*, b*, c*.

Reconstruction of the diffracted x-rays and visualization of the structure of the crystal contents remains elusive, as information on the phases of the diffracted x-rays is lost. The phase of a scattered x-ray wave determines its periodicity with respect to the direct beam, and

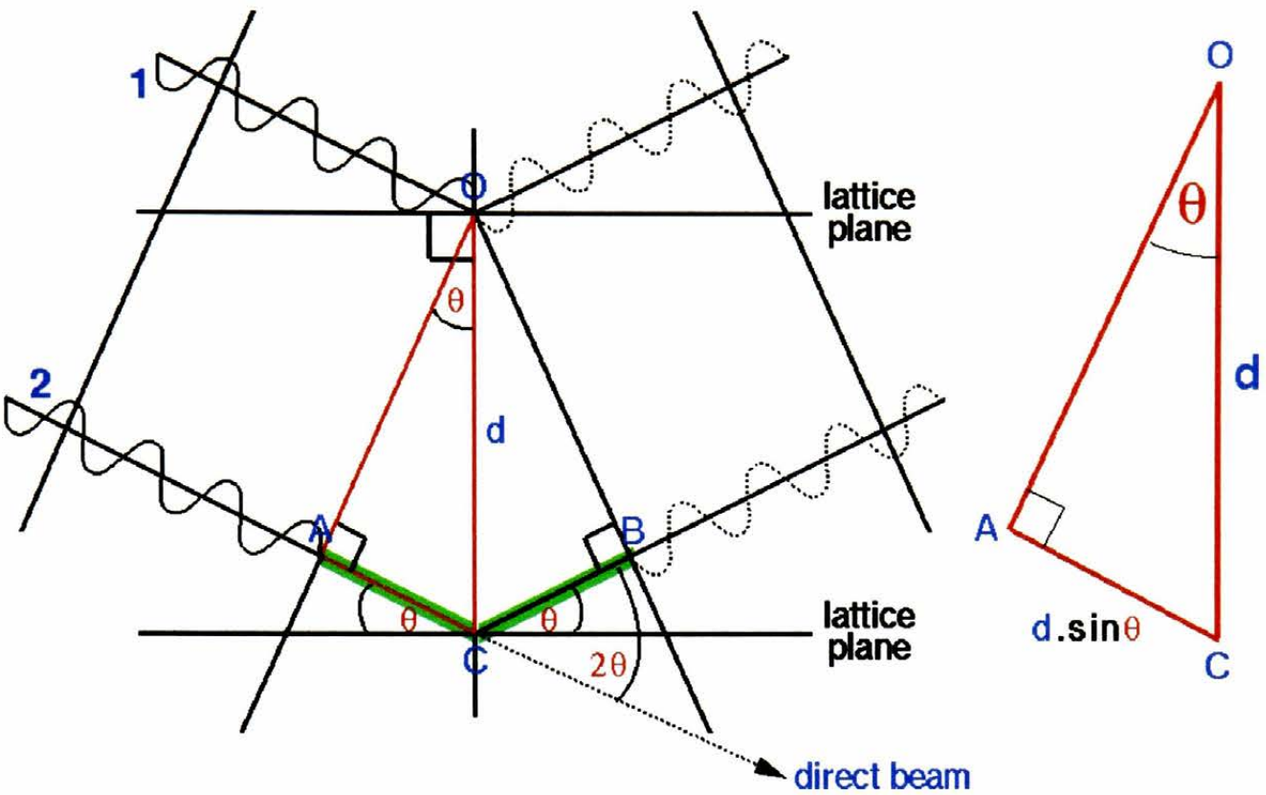


Figure 3.1. Pictorial representation of Bragg's Law. For diffraction to be obtained, the diffracted x-rays have to reinforce each other, i.e. be in phase. The incoming x-rays (beams 1 and 2) strike the crystal lattice planes at points O and C, at an angle of θ. The planes are spaced apart by a distance d. If the diffracted x-rays are to emerge in phase (the extra distance travelled by x-ray 2 *c.f.* x-ray 1 must be equivalent to an integral number of wavelengths [nλ], *i.e.* AC = CB = dsinθ). Hence, Bragg's Law nλ=2dsinθ, with the diffracted x-ray leaving the crystal at an angle of θ to the lattice planes, and an angle of 2θ to the direct beam.

this information is needed for every diffraction spot. In very small unit cells (mainly those of small molecules or very small proteins / peptides), statistical methods exist for the estimation of the phase values, and the structure of the cell contents can be phased *de-novo* using so-called direct methods. With the larger cell volumes found in protein crystallography, some method has to be employed to estimate the phase values from experimental observations (see section 3.3).

The diffracted x-ray contains contributions from all atoms in the unit cell, and may be described as a structure factor F_{hkl}. Since this structure factor is a series (containing contributions from atom A, atom B.....atom N), a fourier transform can be applied to it, generating another series that essentially describes the electron density of the unit cell. The equation connecting the structure factor (F_{hkl}, reciprocal space) and the electron density in the unit cell (ρ [x,y,z], real space) is represented in Equation 3.1:

$$\rho(x,y,z) = \frac{1}{V} \sum_h \sum_k \sum_l F_{hkl}\, e^{-2\pi i(hx+ky+lz-\alpha'_{hkl})}$$

Equation 3.1. Equation linking real and reciprocal space, x, y, and z are points in the unit cell (ρ the value of the electron density at a particular point), h, k and l are indices of the reflection hkl, V is the volume of the unit cell and i is $\sqrt{-1}$. The phase contribution α of reflection hkl is represented by α' ($\alpha=2\pi\alpha'$).

Crystals are not perfect and are subject to disorder. This disorder limits the effective resolution of structural studies, with the diffraction of many protein crystals only observable to a medium resolution (around 2 Å). Crystals show varying resolution limits, and a fair correlation can be observed between effective resolution and crystal solvent content, with resolution decreasing as the solvent content increases. This correlation acts merely as a guide, however, and crystals of high solvent have been shown to diffract to high resolution and *vice*

versa. Disorder of the crystal lattice is termed mosaicity and high values indicate increased disorder, with a greater number of partial and a lower number of full reflections on an oscillation image for the same relative oscillation range.

3.3 Phasing Protein Structures

3.3.1 Overview of Techniques Used

To estimate phases of the diffracted x-rays using experimental methods, two techniques are used that alter the intensities of the spots in the diffraction pattern, and a third technique compares the collected data with that from a previously determined homologue. In isomorphous replacement, the intensity difference is achieved by the inclusion of a heavy atom in the protein crystal (an atom at least as heavy as those in the first row of the transition elements, third row transition elements are most commonly used), but with anomalous scattering the intensity differences come from interactions of the heavy atom with the x-rays at a particular wavelength. A third technique for estimating phases, molecular replacement, can be utilised when the structure of a homologous protein to the one studied is known, no intensity differences are measured.

3.3.2 Single or Multiple Isomorphous Replacement

With this technique, measurable changes in diffraction spot intensity are made by the introduction of a heavy atom into the crystal lattice, binding to the protein in an ordered manner and altering the unit cell dimensions as little as possible (being isomorphous with the "native" crystal form). This is traditionally achieved by long periods of soaking the crystal in a mother liquor solution, supplemented with a low concentration of heavy atom compound e.g. Hg and Pt salts. These may react covalently (Hg bonding to cysteine S), ionically ($PtCl_4^{2-}$

interacting with histidine NH^+) or hydrophobically (I binding to hydrophobic patches) with the protein. An excellent review of heavy atom derivitisation of protein crystals is provided in (Blundell and Johnson 1976). Heavy atoms are required due to their large number of electrons, which make a larger and more measurable intensity change in the diffraction pattern.

To estimate phases using the intensity differences achieved upon heavy atom derivitisation of the protein crystal, it is convenient to represent the structure factors of the reflections as vectors on an Argand diagram, with the real axis horizontal, and the imaginary axis vertical. The intensity is represented as the length of the vector, and its phase as the angle between the vector and the real axis. The protein (F_P) and the heavy atom (F_H) structure factors combine to give an additive contribution (F_{PH}). Computationally the phasing comes from finding the position of the simple heavy atom substructure by looking at the difference between F_{PH} and F_P. An assumption is made that $F_{PH} - F_P = F_H$. This is shown diagrammatically in Figure 3.2.

It can be seen that this single derivative provides two potential solutions for the phase angle. If the two are sufficiently close, such that their average gives a good estimate of the true solution, then the structure may be solved with this single derivative (known as SIR, single isomorphous replacement). More commonly, other derivatives are needed to remove this phase ambiguity (MIR, multiple isomorphous replacement). Further derivatives must bind at different sites to previously obtained derivatives, to ensure that the ambiguity can be resolved.

The actual information in positioning the heavy atom in the unit cell to allow use of the aforementioned procedure comes from a calculation known as a Patterson function. The

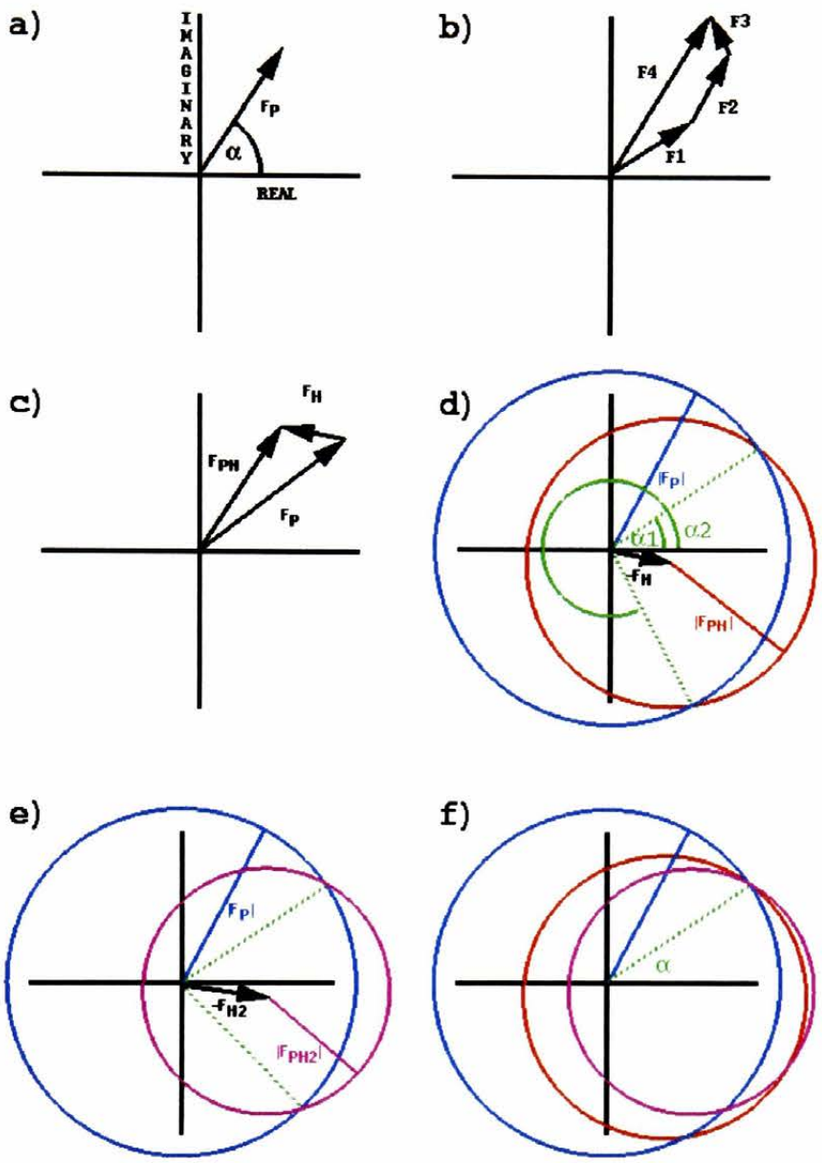


Figure 3.2. Geometric representation of MIR phasing. Having found the positions of the heavy atoms in the unit cell, the phases for the protein : heavy atom structure factors can be calculated. A) Argand style diagram representing structure factors as vectors. The magnitude of the structure factor is represented by the vector length, and the phase is represented by the angle made with the real axis (the vector has 2 perpendicular components, one along the real axis, and one along the imaginary axis). b) Individual structure factors (F1 to F3) can be added into a single contribution (F4). Part (c) shows that because the diffractive contributions from atoms are additive, $F_{PH} = F_P + F_H$. In part (d) the known quantities $|F_{HP}|$ and $|F_P|$ (calculated from measured spot intensities), the calculated $|F_H|$ and phase angle of F_H (from the heavy atom substructure) are used to solve the equation $F_P = F_{PH} - F_H$. This construction provides two equally valid solutions for the phase angle of F_P (represented as dotted green lines, angles $\alpha 1$ and $\alpha 2$). Part (e) repeats construction (d), using a different derivative. Information from (d) and (e) indicate the true solution of the phase angle in (f). In practice, MIR solutions are calculated numerically rather than geometrically, by obtaining a solution to the equation $F^2_{HP} = F^2_P + F^2_H + 2F_PF_H \cos(\alpha_P + \alpha_H)$.

Patterson function is computed from data without the use of phases (it uses $|F_{PH}-F_P|^2$), and gives a contour map showing vectors between atoms. A difference Patterson map (calculated using the difference between F_P and F_{PH}) shows vectors between heavy atoms of the derivative, which due to the small number of heavy atoms present allows us to reconstruct their substructure (the atomic placement of the heavy atoms with respect to a fixed point, the origin). This search for heavy atoms can be simplified using knowledge of the crystal symmetry and looking for vectors between related atoms in areas of the map known as Harker sections.

3.3.3 Anomalous Scattering

Friedel discovered that reflections that come from the lattice planes hkl and –h-k-l are of equal intensity, and are known as Friedel pairs. However, the inclusion of certain atoms in the crystal lattice (either by soaking, inclusion from the start of crystal growth or as a natural component of the protein e.g. metalloproteins), at certain x-ray wavelengths can absorb the x-rays and re-emit them out of phase. As a result of this, the Friedel pairs are no longer equivalent and this is known as anomalous dispersion. The wavelength at which this phenomenon is maximal is called the absorption edge of the atom (Figure 3.3). The anomalous dispersion of atoms can be used in phasing and anomalous scatterers most commonly used are Hg, Fe and Se. Selenium offers a distinct advantage, since it can be incorporated into the protein with 100 % occupancy as a selenomethionine derivative (see Chapter 5).

The anomalous scattering contribution of an atom can be described as a vector, having two perpendicular contributions. The vector is represented in the Argand diagram format, on a

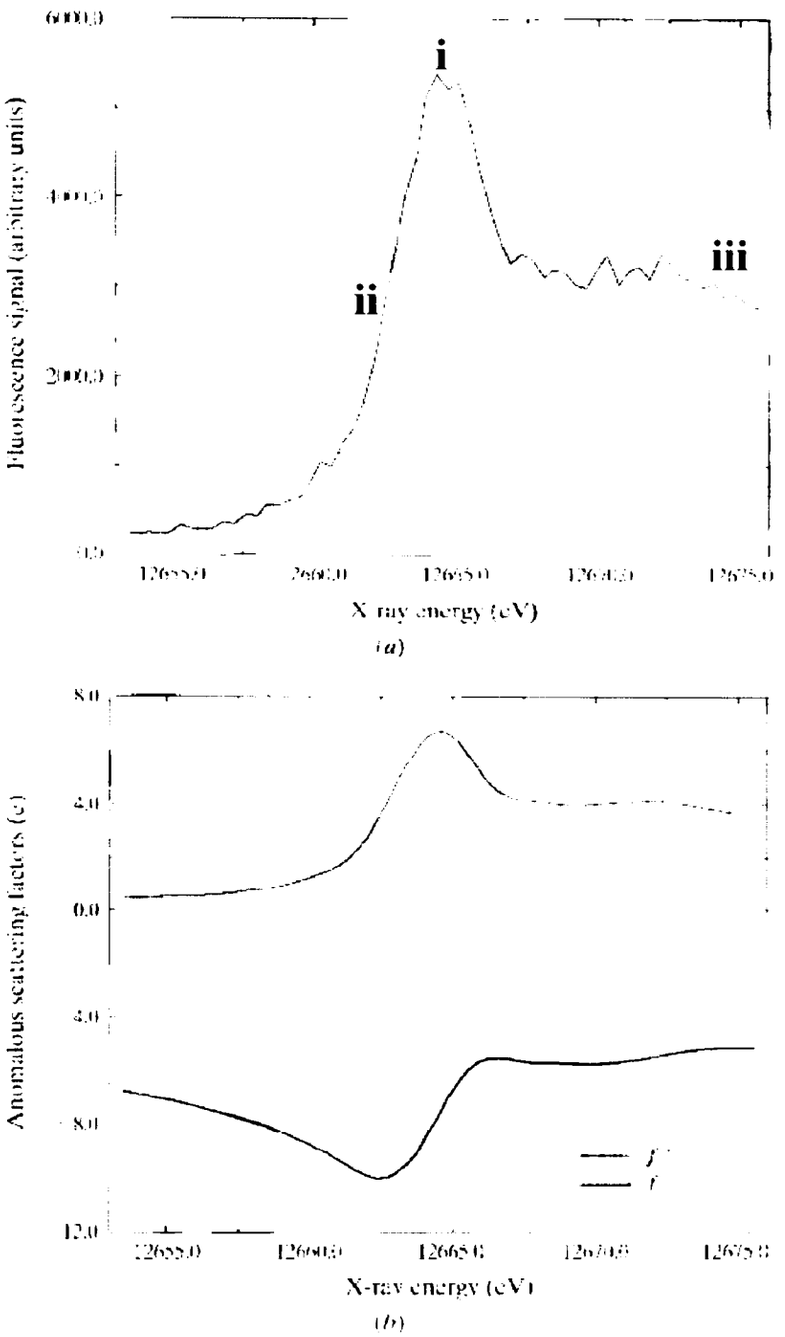


Figure 3.3. X-ray absorption profile of Selenium. The peak (i), inflection point (ii) and remote (iii) energies are shown in section (a). The imaginary component f'' (maximum agreeing with i) and real component f' (minimum agreeing with ii) of the anomalous scattering are given in section (b). Figure adapted from Walsh *et al.* 1999.

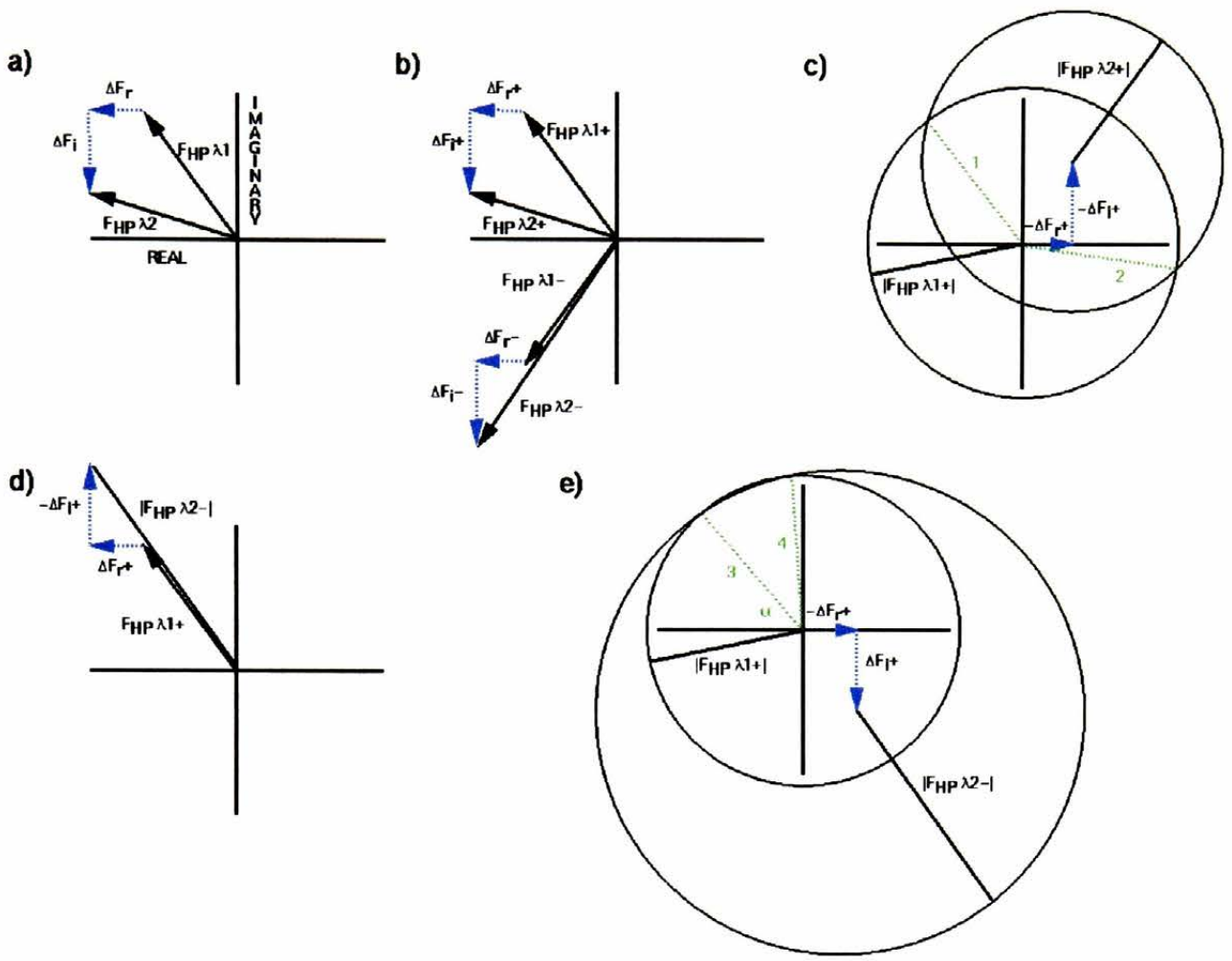


Figure 3.4. Geometric representation of MAD phasing. (a) structure factors for F_{HP} at $\lambda 1$ (no anomalous scattering) and F_{HP} at $\lambda 2$ (anomalous scattering). The effect is seen as 2 components, one real (ΔF_r) and the other imaginary (Δf_i). Due to the anomalous scattering, there is a breakdown in Friedel's Law (part (b)). $F_{HP}\lambda 1+$ and $F_{HP}\lambda 1-$ are equivalent, but $F_{HP}\lambda 2+$ and $F_{HP}\lambda 2-$ are not. After location of the anomalous scatterers using Patterson methods, and knowledge of ΔF_r and Δf_i from standard tables, a construction like that in part (c) can be used to estimate phases for $F_{HP}\lambda 1$. The construction is a solution to $F_{HP}\lambda 1+ = F_{HP}\lambda 2+ - \Delta F_r+ - \Delta f_i+$. At this point the phase has two solutions (green dotted lines 1 + 2), but can be solved unambiguously using the Friedel pair of this reflection, and a construction reflecting $F_{HP}\lambda 2-$ across the real axis (parts (d) and (e)). The two phase solutions (green dotted lines 3 + 4) are obtained, with the true phase, 3, agreeing with phase 1 obtained in part (c).

horizontal axis (real axis) and a vertical axis (imaginary axis). It is because of these two contributions that F_{khl} is no longer the same as F_{-h-k-l} (Figure 3.4). At the x-ray absorption maximum of the anomalous scattering atom, the imaginary contribution is at a maximum. At the inflection point of the absorption edge, the real component reaches its minimum value. An absorption profile of selenium is shown in Figure 3.3. Using this peak, inflection point, and a third, remote wavelength where the anomalous contribution is negligible, the intensity differences can be maximised and used in phasing (MAD, multiwavelength anomalous dispersion, Figure 3.4).

Occasionally the peak wavelength contains sufficient anomalous signal to locate the scatterer with the use of just this one dataset (SAD, singlewavelength anomalous dispersion). However, like SIR, this gives two possible solutions to the phase determination and is referred to as the hand problem. Electron density maps are calculated for both possibilities (in practice using both enantiomers of the anomalous scattering atom positions), and the most interpretable map used. Density modification (where the protein electron density and solvent electron density boundaries are "flattened out", giving a clearer map and improved phases) is especially useful in making the correct solution stand out.

3.3.4 Advances In Anomalous Scattering

The MAD technique is very popular since a single crystal is often sufficient, and avoids all problems of non-isomorphism (all the data from a single crystal, no unit cell changes as observed in the soaking of crystals needed in SIR/MIR). MAD requires the use of tunable wavelength x-rays, obtainable at synchrotrons (see section 3.6). With synchrotron beamtime at a premium, it is a distinct advantage to use SAD rather than MAD. Several excellent

examples of the use of SAD are provided in Dauter *et al.* 2002. It is becoming possible to use SAD phasing on Cu K_α home sources, with an excellent anomalous contribution from atoms such as Iodine (Evans *et al.* 2002). Anomalous scattering can also be used in combination with isomorphous replacement (M/SIRAS, multiple/single isomorphous replacement anomalous scattering) to further reduce phase ambiguity.

Determination of anomalous scattering positions becomes more complex as the number of scatterers increases, but the successful use of programs such as Shake & Bake (Weeks and Miller 1999) allow the location of larger numbers of heavy atom sites. This employs a direct methods procedure to locate the anomalous scattering substructure.

3.3.5 Molecular Replacement

If the protein being studied has reasonable homology to a protein whose structure has previously been determined (this is very case dependent, but the usual limit of success requires > 30 % sequence identity), then the structure can be solved without the need for derivitisation or anomalous scattering experiments. The phases from the known homologue are used in calculations estimating the phases of the unknown structure dataset.

If the structure is isomorphous (same unit cell, same protein, most commonly encountered when attempting substrate or inhibitor soaks on a crystal, or when dealing with site-directed mutants), then the phases from the known structure are sufficient to phase that of the unknown. Typically, a few cycles of rigid-body refinement will be used to allow for slight crystal non-isomorphism (changes of less than 2 % in cell axes). If the known structure (hereafter referred to as the search model) is non-isomorphous with the unknown x-ray data,

then a six-dimensional search must be undertaken – the orientation of the search model in the unknown structure unit cell must be established. Newer molecular replacement programs (Kissinger *et al.* 1999) are able to search in all six dimensions simultaneously, but the search is more traditionally split into a 3D rotation step and a 3D translation step, executed in that order.

The Patterson map, as discussed earlier, gives an identical set of vectors wherever the search model is positioned. It is only dependent on the model orientation. In the rotation function step, Patterson maps of the unknown data (using F_P^2) and those computed from a set of orientations of the search model are compared. If the unknown structure contains more than one molecule per ASU (as described in section 3.4), several correct solutions are obtained. The solutions are ranked on their agreement with the data, and the top rotation trials used in a subsequent translation step search. Here the search models are positioned in the unit cell in discrete translational steps, and their calculated structure factor amplitudes compared with those from the unknown structure diffraction data.

The separation of the 6D search into two 3D searches lowers computational time considerably. Positioned search molecules are then analysed with regard to their crystal packing interactions. Electron density maps are calculated to verify that a correct solution has been found. In molecular replacement searches, a truncated or poly-alanine search model is often preferred to reduce existing phase bias in the initial model.

3.4 Crystal Symmetry

The unit cell can be translated in steps equivalent to the length of cell axes to recreate the entire crystal. The asymmetric unit (ASU) is defined as the smallest part of the crystal that can recreate the unit cell using a series of rotations and translations. Any symmetry between multiple protein copies in the ASU is designated non-crystallographic symmetry (NCS). The symmetry of the unit cell and the relationship between the asymmetric units that comprise it fall into seven crystal systems. These are (in order of increasing symmetry) triclinic, monoclinic, orthorhombic, tetragonal, rhombohedral, hexagonal and cubic. These crystal systems may have further lattice symmetry known as point groups. A full description of these symmetry elements (hereafter referred to as spacegroups) can be found in Stout and Jensen (1968).

3.5 Model Building & Refinement

With a set of phases and a subsequently calculated electron density map, the protein model can then be built. If molecular replacement was used to phase the structure, model building is a simple case of replacing the non-conserved residues between the model and the target, and looking for any shifts in the protein structure. With other cases, knowledge of the protein primary sequence is used, possibly with other biochemical data to identify where the residues fit into the electron density. With some derivatives (especially selenomethionine), it is easy to associate the protein sequence with the located heavy atom positions. Several aids to building exist. With high resolution data (usually 1.2 Å or better, cases with very accurate phases will tolerate up to 2.5 Å), programs such as ArpWarp (Morris *et al.* 2002) can trace the polypeptide automatically, provided that the initial phase estimates are sufficiently accurate.

With noisier, low-resolution maps, programs can build a skeleton of "bones" between peaks in the electron density as a guide (Kleywegt and Jones 1994).

The built model is then refined to ensure a good fit to the collected diffraction data. The refinement protocol may use different targets to ensure a better fit (e.g. conjugent gradient or maximum liklihood targets - Murshudov *et al.* 1997), and use information from the phase estimates or a target just using the structure factor amplitudes alone. Refinement using experimental phase information is more accurate and faster than refinement against amplitudes alone. Restraints are used in refinement (geometry libraries), which can be weighted according to the resolution of the structure – high resolution structures are refined with more "freedom". The residual index (Rfactor) is used to measure convergence:

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

Equation 3.2. Definition of the residual index (Rfactor, represented above by R, F_{obs} relates to observed *i.e.* measured amplitudes, and F_{calc} to amplitudes calculated from the model)

It is considered best practice to exclude around 5 % of the data from this refinement to avoid over-fitting of the structure (the 5 % is a random selection taken over the resolution range, and is not used in refinement of the structure, but is included in a cross-validation calculation known as Rfree). Values of Rfree obtained at the end of model refinement are highly dependent on the resolution of the data, the quality and completeness of the data, but most importantly, the accuracy of the final model. The model must also have good geometry, usually compared to values derived from small molecule structures (Engh and Huber 1991), and make biochemical sense.

3.6 Production of X-rays

X-rays (in laboratory use) are produced by accelerating electrons towards a metal target, usually Cu or Mo. The incoming electron displaces a lower orbital electron in the metal atom, which is replaced by a higher orbital electron, emitting the excess energy as an x-ray. Older laboratory sources use a sealed tube, whilst newer sources prefer a rotating anode target which can dissipate the heat produced more efficiently and give more powerful x-rays. The electrons are produced using a large charge difference between a cathode filament and the target anode. The x-rays are then filtered to give a monochromatic wavelength, and focussed on the crystal sample.

Higher energy x-rays are produced in particle accelerators, where electrons circulate at near to the speed of light. The electrons are guided using magnetisation, and emit highly parallel, high intensity x-rays at a tangent. The resulting x-ray beam (synchrotron radiation) is far more powerful than that from home sources, and datasets may be collected in minutes as opposed to days. The data are usually of a far higher quality and there is also the capability of tuning the precise x-ray wavelength used, ideal for the MAD & SAD phasing techniques. A diagram of the ESRF (European Synchrotron Radiation Facility) is provided in Figure 3.5.

3.7 Protein Crystallization

Crystallization of macromolecules is often viewed as an enigma by those who do not practice it, and almost always so by those that do. Many indicators to the ability of a protein to crystallize exist, most importantly purity (both bulk sample purity and also sample micro heterogeneity) and lack of polydispersity in light scattering (Ferredamare *et al.* 1994). It does however remain largely a case of empirical trial and error.

Figure 3.5. The ESRF, Grenoble, France. Steep overhead view given in inset. The circular storage ring of the electrons can clearly be seen. The strip extending into the storage ring is the linear accelerator (linac), used to initially boost the electrons up to speed before circulation.

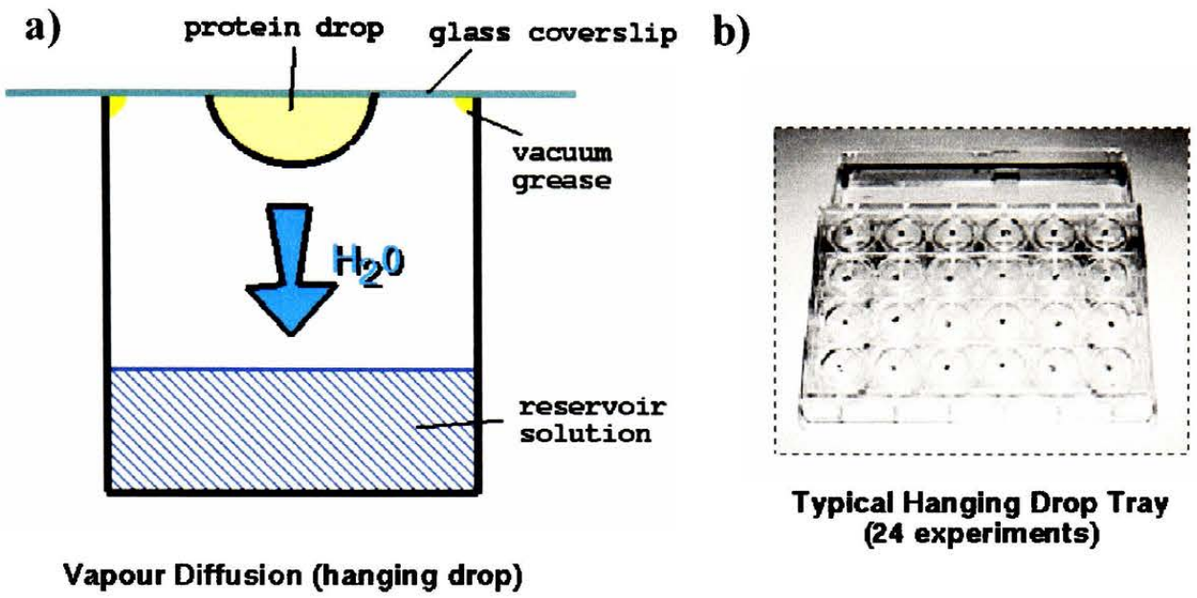


Figure 3.6. Vapour-diffusion method for protein crystallization. Section (a) provides a schematic of a vapour-diffusion experiment, the hydrated protein drop losing water to the reservoir solution over a period of time. The drop has a typical volume of 5-10 µl, and the reservoir solution 0.5-1 ml. Section (b) shows a VDX plate (Hampton Research, Laguna Niguel, CA), used to house such experiments.

The sample to be crystallized is usually filtered to remove both proteolytic bacteria and any aggregate that may hinder crystal growth. It is then concentrated to approximately 10 mg/ml, or even higher dependent on solubility.

The most popular method for protein crystallization is the vapour diffusion experiment (Figure 3.6), which can be either hanging drop (suspended from a glass coverslip), sitting drop (sat on a pillar) or sandwich drop (lying between two glass coverslips). The protein solution is mixed with a crystallization solution which contains a pH buffer, a precipitant and possibly extra salt or additives (see McPherson 1999). The mixed drop is then placed in a sealed environment and equilibrates slowly with the crystallization solution, usually losing water over time. The resultant concentrating of the protein and precipitant then forces molecules out of solution, hopefully into a crystalline phase. A protein solubility curve (also called a phase diagram) is shown in Figure 3.7.

The protein moves into the supersaturation zone, leading to a crystal nucleus. This is defined as a stable, regular array of protein molecules. As protein molecules add to this growing lattice, the solution protein concentration decreases, and the equilibrium lies in the metastable zone. It is desirable to avoid phase separation, but in some cases protein crystal growth in this region can be seen to occur.

Other techniques for crystal growth include dialysis (where the protein is slowly equilibrated against a crystallization solution in a capillary or dialysis button) or batch method (the mixing of the protein and crystallization solutions, often under oil, results in direct crystal growth). A shift in temperature can also alter protein solubility and lead to direct crystallization.

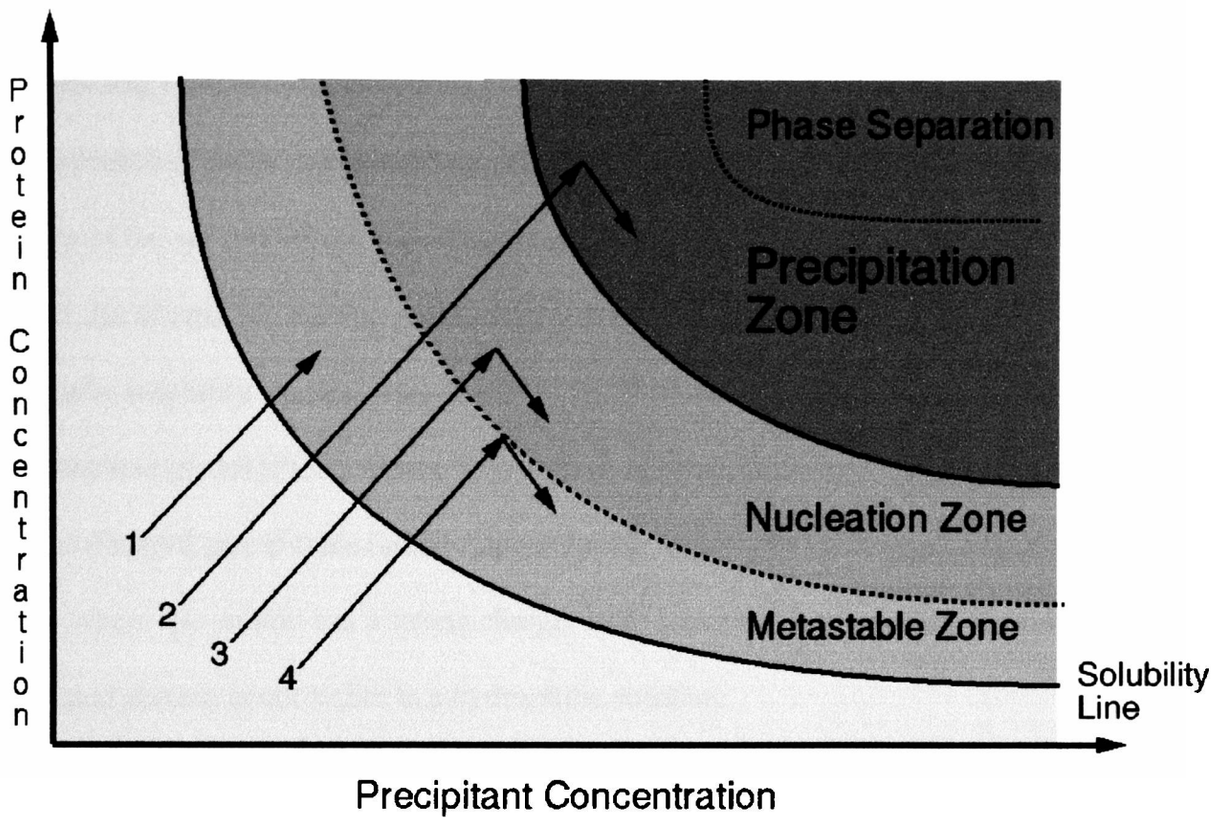


Figure 3.7. A typical protein solubility diagram. The protein & precipitant concentrations typically start off low (at a point in the bottom left of the diagram) and increase as the drop dehydrates. Example 1 would lead to a clear drop upon equilibration, example 2 to a protein precipitate, and example 3 to a shower of microcrystals. Example 4 shows an idealised case, where nucleation is achieved upon touching the nucleation zone, and then a few single crystals grow to a large size in the metastable zone.

With initial crysallisation trials, it is commonplace to use a set of standard crystallization solutions that have proved successful in crystallizing other proteins (Jancarik and Kim 1991). This approach remains moderately successful. If some crystalline material is observed in one or more of the set conditions, these can then be altered in an attempt to optimise crystal growth. An alternative starting point may be to use knowledge of the protein pI in a series of systematic solubility studies (Ries-Kautt 2001). It is assumed that neutral precipitants (polyethylene glycol [PEG], alcohols) will have most success with a neutral protein (pH at the pI), and charged precipitants (ammonium sulphate, other salts) will give a result away from the pI, where the protein has a strong charge. Both scenarios lead to crystallization when the uncharged protein is not stable in a hydrophilic solution.

Many other techniques may be used to assist the macromolecular crystallographer in growing diffraction-quality crystals. Seeding artificially inserts a foreign nucleus into the protein drop and initiates crystal growth. The seed may be of a previously grown crystal, or be epitaxial – from a related protein or just another growth surface e.g. clothes fibre. An excellent source of information covering all details of protein crystallization is given by McPherson (1999).

3.8 Sample Mounting, Data Collection & Processing

As mentioned earlier (section 3.1), protein crystals need to remain hydrated during data collection or they lose their internal order. A room temperature dataset can be collected by taking the crystal and sealing it in a glass capillary with beeswax, leaving a small buffer region of mother liquor to keep the crystal hydrated. A disadvantage of this technique is that radiation damage to the crystal at room temperature is higher than at lower temperatures

(primary damage from the direct beam is unavoidable, but secondary damage from free radicals can spread throughout the lattice). This will eventually affect the electron density map and the effective resolution of the data. To prevent or minimise such damage, most data is collected at 100 K. The crystal is transferred to a solution containing a cryoprotectant (unless the mother liquor will already freeze without forming ice), and flash frozen in a loop, bathed in a cone of dry nitrogen gas. Crystals become essentially immortal on home x-ray sources and last far longer at synchrotrons. However, some crystals are too fragile to survive such soaking, and many suffer with increased mosaicity (lattice disorder). See Figure 3.8 for an explanation of both styles of crystal mounting.

The diffraction pattern itself was traditionally collected on x-ray sensitive film, but modern techniques use an image plate system, multiwire area detector or CCD (charge-coupled device). Image plate detectors use a phosphor that is stimulated by x-rays, which is then recorded after further laser stimulation using a photocell. Multiwire area detectors have a series of wire grids that produce a signal when x-rays polarise gas molecules in their vicinity. CCDs are different in that they have no "dead time", responding directly to the x-rays that strike them. They do, however, use phosphors, which transfer light to a fibre optical array, the signal of which is amplified and read. Image plates remain the most popular detector type at home sources due to the cheaper cost, greater dynamic range and lack of urgency, whilst CCDs are more popular at synchrotrons with their high spatial resolution and rapid read out time. CCDs are prone to zingers (spots from background cosmic radiation), and so use on a home source with long exposures would be problematic, requiring use of a fine ϕ slicing approach where data is collected in smaller segments.

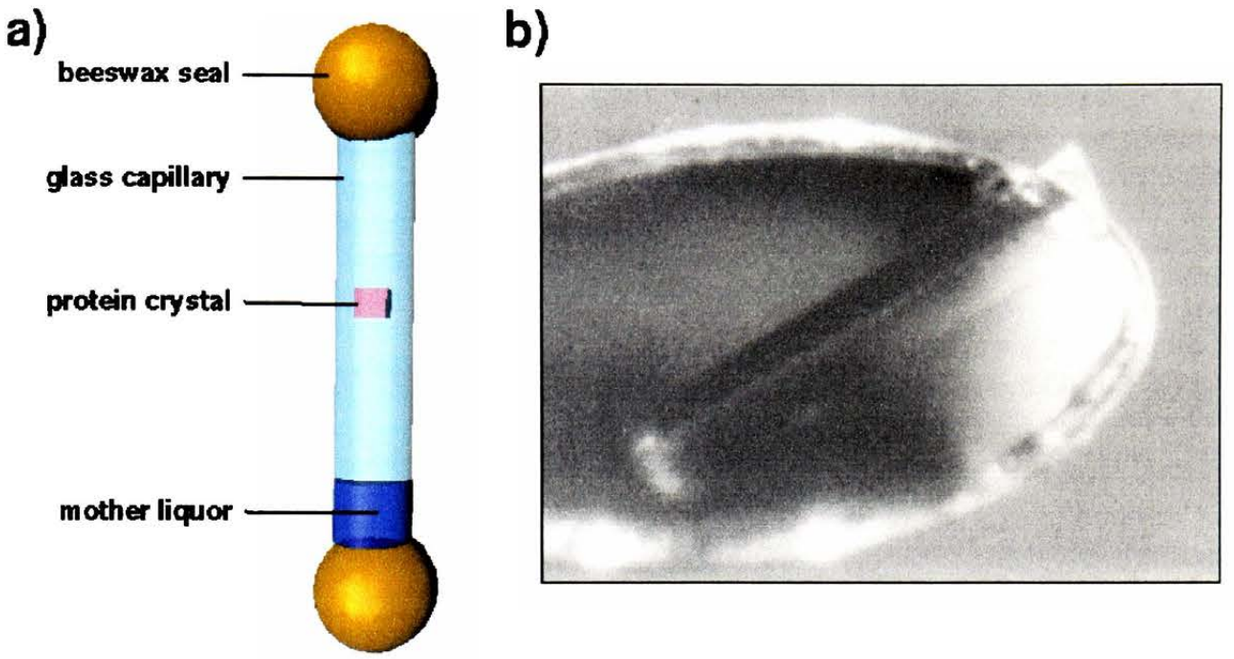


Figure 3.8. Sample mounting for data collection. Section (a) shows a schematic of a typical room temperature crystal mount. The crystal is held by surface tension against the wall of the capillary, and is kept hydrated by a small buffer region of mother liquor. Beeswax is used to seal both ends of the capillary. Section (b) shows a cryo-mounted crystal (of 3α-hydroxysteroid dehydrogenase, see Chapter 8) flash-frozen in a loop, held in a stream of dry gaseous nitrogen at 100 K. The loop itself is usually made from a low-diffracting material such as rayon or mylar. Crystal size approximately 40 x 40 x 300 μm.

The mounted crystal is placed on a goniometer head, on the spindle axis, which usually rotates parallel to the face of the detector (perpendicular to the x-ray source), and allows a section of the cell volume to be sampled. A beamstop prevents the direct x-ray beam from hitting the detector. The exact rotation range to be sampled depends on the crystal alignment and spacegroup symmetry.

Each spot in the diffraction data set is assigned a value of hkl, and the intensity of the spot measured via integration. The square root of this intensity is proportional to the amplitude of the structure factor it represents. From the positioning of the spots relative to one another and the direct beam, and the distance of the detector relative to the crystal, unit cell axes and angles are calculated and a putative spacegroup assigned. The true spacegroup may only become apparent after data reduction or phasing, and is sometimes ambiguous even upon final refinement of a structure. The intensities are scaled relative to one another, and data reduction commenced (symmetry-related reflections are merged and the dataset is reduced to the asymmetric unit, where a set of unique reflections correspond to a section of the reciprocal unit cell). The finalised dataset is then given a Rfree set (a random 5 % of the data tagged as a cross-validation set, these data are not included in refinement) and used in refinement of the protein structure of interest.

CHAPTER 4

OBTAINING CRYSTALS OF NITROREDUCTASE

4.1 Introduction

It is generally agreed that the single most important determinant in obtaining protein crystals is a high level of protein purity and homogeneity. At the outset of this investigation large quantities of active NTR, purified to > 95 % purity (as assayed by SDS-PAGE) were desired. Upon achieving this, the protein stock could be utilised in a series of crystallization trials to screen for a set of conditions that would yield large, reproducible crystals suitable for diffraction studies.

4.2 Materials & Methods

4.2.1 General Procedures

Unless stated otherwise, all chemicals and reagents were purchased from Sigma-Aldrich (Poole, UK). The spectrophotometer used in all O.D. determinations was a Kontron Uvikon 923 model (Bio-Tek Kontron Instruments, Milton Keynes, UK).

4.2.1.1 LB Media

Luria-Bertani (LB) media was made to the following recipe : 10 g tryptone digest (from casein), 10 g NaCl and 5 g yeast extract, made up to a final volume of 1 litre with de-ionised water and autoclaved. "Half-strength" LB for selenomethionine derivative starter culture media was made using 50 % v/v LB and 50 % v/v sterile water.

4.2.1.2 SDS-PAGE

All gels were made using the Biorad mini-protean apparatus (Biorad, Hemel Hempstead, UK). The resolving gel was made using 3.3 ml of 30 % w/v acrylamide stock solution, 2.5 ml of 4 X lower Tris (1.5 M Tris, pH 8.8, 0.4 % w/v SDS), 4 ml water, 150 µl of 10 % w/v $(NH_4)_2S_2O_8$ and 6 µl of TEMED. The gel was levelled with the use of a few drops of water-saturated n-butanol. The stacking gel was made using 0.5 ml of 30 % w/v acrylamide stock solution, 1.25 ml of 4 X upper Tris (0.5 M Tris, pH 6.8, 0.4 % w/v SDS), 3.2 ml water, 20 µl 10 % w/v $(NH_4)_2S_2O_8$ and 6 µl TEMED. The finished gels were ran at 100 V in a buffer containing 12 g of Tris base, 66 g L-glycine and 22.5 ml 10 % w/v SDS, made up to a final volume of 2.25 litres. Gel samples were loaded in 20 µl aliquots, after heating for 3 minutes at 100°C in the presence of 5 µl gel loading buffer (0.2 M Tris, pH 6.8, 0.52 M DTT, 0.35 M SDS, 50 % w/v glycerol and 3.7 mM BPB). All gels were stained and destained using a solution containing 40 % v/v methanol, 10 % v/v glacial acetic acid, 50 % v/v water and the presence/absence of 2.5 g Coomassie Brilliant Blue, respectively. After reaching the desired level of staining, gels were dried using a Biorad 583 gel-drier and vacuum pump.

4.2.1.3 Dialysis Tubing

All dialysis tubing was boiled for 10 minutes in a litre of water containing 0.1 mM EDTA, 0.1 % v/v β-mercaptoethanol and 10-20 mg of $NaHCO_3$. Boiled tubing was rinsed with, and kept in, sterile water at 4°C.

4.2.1.4 Protein Concentration Determination

For rapid protein concentration determination during the purification protocol, a solution with an absorbance of 1.0 was estimated to have a protein concentration of 0.5 mg/ml. For more

sensitive measurements, the Bradford protein assay (Bradford 1976) was preferred, with calibration against a set of BSA standards of known concentration.

4.2.2 Protein Expression

4.2.2.1 Transformation of Cells

The plasmid DNA construct pJG6A1b used for NTR over-expression was a generous gift from Dr Peter Searle (Institute of Cancer Studies, University of Birmingham). The plasmid is a derivative of pET11c (Novagen, Madison, WI), with a NTR insert that was originally amplified from *E. coli* DH5α using PCR. The DNA sequence was found to be identical to that previously reported (Anlezark *et al.* 1992). The plasmid was added to a 100 µl aliquot of competent BL21 cells (Phillips *et al.* 1984), which were then gently mixed and left on ice for 30 minutes. The cells were subjected to a heat shock step at 42°C for 90 seconds and then mixed with 1 ml sterile LB media and left at 37°C for 1 hour before being plated out onto LB agar plates containing ampicillin (hereafter referred to as +amp, and unless stated otherwise at a concentration of 100 µg/ml).

4.2.2.2 Bacterial Growth & Protein Induction

Successful transformants, determined by growth on a +amp plate, were picked off using a sterile loop and placed into conical flasks containing 50 ml LB media +amp and left to shake overnight at 37°C. Five ml of this starter culture was used to inoculate 500 ml of LB +amp and left to shake at 37°C until the OD at 560 nm reached ~0.8. Expression of NTR was induced by the addition of 5 ml of 100 mg/ml IPTG. The cells were then grown for a further 4-6 hours and harvested by centrifugation at 8000 X g for 10 minutes. Pelleted cells were frozen at -20°C to facilitate cell wall weakening and aid lysis.

4.2.3 Preparation of a Crude Extract

The cells and all cell suspensions were kept on ice during the preparation of the extract. A volume of 4 ml of lysis buffer (10 mM KH_2PO_4, pH 7.0, 10 μM FMN, 10 μg/ml PMSF) was added per gram of dry cell weight. The cells were brought gently into suspension using a rubber teat placed over the end of a glass rod. The cellular suspension was then broken open using 2 cycles of sonification (Heat Systems XL2020 sonicator [Heat Systems, Farmingdale, NY], level 5 strength, 4 X 15 second bursts), and DNAse (1 μg/ml) and RNAse (1 μg/ml) added to hydrolyse polynucleotides for reduced sample viscosity. The lysed cells were centrifuged at 8000 X g for 10 minutes and the supernatant collected. Ammonium sulphate was added to the supernatant until a final concentration of 0.5 M was reached, upon which the solution was centrifuged at 10,000 X g for 10 minutes. The resultant supernatant was filtered through a 0.45 μm filter and used as the starting material for the purification procedure.

4.2.4 Purification of the Crude Extract

The extract was applied onto a column of phenyl sepharose media (approx. 50 ml in volume), pre-equilibrated with 20 mM KH_2PO_4, pH 7.0, 0.5 M $(NH_4)SO_4$ and 10 μM FMN. The column was then washed with this equilibration buffer until the OD at 280 nm of the eluate was close to zero. Fractions were eluted with a stepwise gradient using a buffer containing 10 mM Tris, pH 7.6 and 10 μM FMN. NTR fractions were pooled and dialysed against the elution buffer at 4°C overnight.

Dialysed protein was loaded onto a column (approx. 40 ml in volume) of Q-sepharose (Pharmacia, Milton Keynes, UK), pre-equilibrated with 20 mM Tris, pH 7.6. Protein fractions were eluted using a linear salt gradient from 0.1 to 0.4 M NaCl, made with a gradient maker

and two solutions of 20 mM Tris, pH 7.6 (75 ml each). NTR fractions were pooled and dialysed against 10 mM KH_2PO_4, pH 7.0, 10 μM FMN at 4°C overnight.

The dialysed NTR was applied to a small column (approx. 10 ml in volume) of hydroxylapatite equilibrated in 10 mM KH_2PO_4, pH 7.0, 10 μM FMN. NTR was found not to bind the column and pass-through fractions were collected and pooled.

4.2.5 Crystallization Sample Preparation

Purified NTR was dialysed at 4°C for 4 hours against a buffer containing 5 mM Tris, pH 7.0 and 10 μM FMN. This dialysed sample was concentrated to a volume of approximately 2-5 ml using an AMICON concentrator (Millipore, Watertown, MA), and then concentrated further to a protein content of ~20 mg/ml using a CENTRICON 10,000 Mwt cut-off spin filter (Millipore) spun at 5000 X g. The final protein stock was stored at 4°C after the addition of 0.01 % w/v sodium azide.

4.2.6 Crystallization

4.2.6.1 Coverslip Preparation

Glass coverslips were immersed in a bath of Aquasil siliconising solution (Pierce, Rockford, IL), washed briefly in a bath of methanol and then allowed to dry for several minutes in a heating cabinet at 60°C.

4.2.6.2 Initial Sparse Matrix Sampling & General Crystal Tray Preparation

For initial crystallization trials, a vapour diffusion experiment was preferred, using Hampton Research sparse matrix screens I and II (Hampton Research, Laguna Niguel, CA). VDX

plates (Hampton Research) were used to house 600 µl of reservoir solution per well, with 4 trays of 24 wells each providing a basis to screen 96 conditions, 1 to 48 from each screen. Wells were greased with Dow Corning high vacuum silicon grease to provide a tight seal for the vapour-diffusion experiment. The protein sample (volume of 5µl, at 10 mg/ml) was mixed with 5 µl of the reservoir solution on a pre-siliconised coverslip, inverted and placed over the well to give a hanging-drop equilibration experiment.

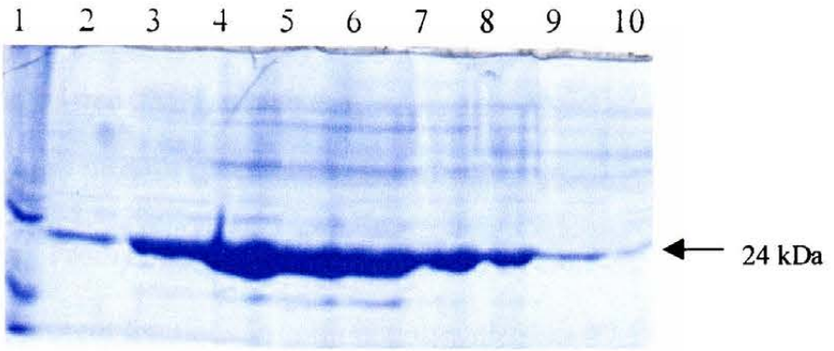
4.2.6.3 Incomplete Factorial Screening

To facilitate controlled screening of any promising crystallization conditions, an incomplete factorial sampling technique was used to explore experimental variables. This method is fully described by Carter and Carter (1979).

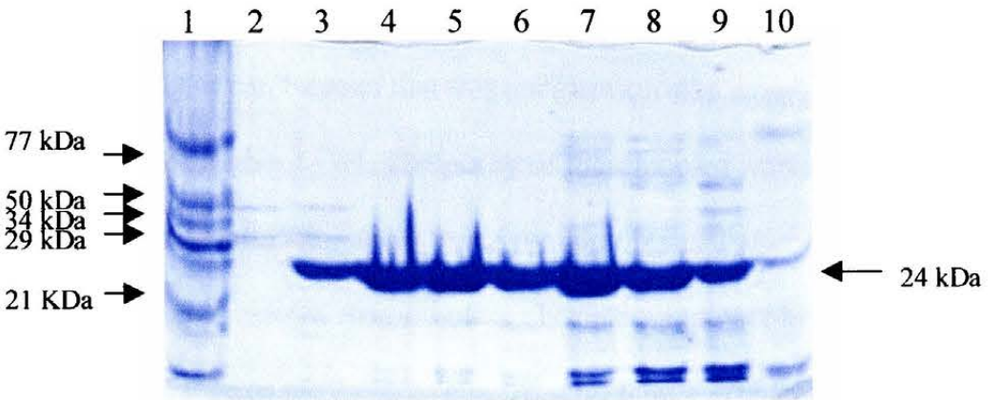
4.3 Results & Discussion

4.3.1 Protein Expression & Purification

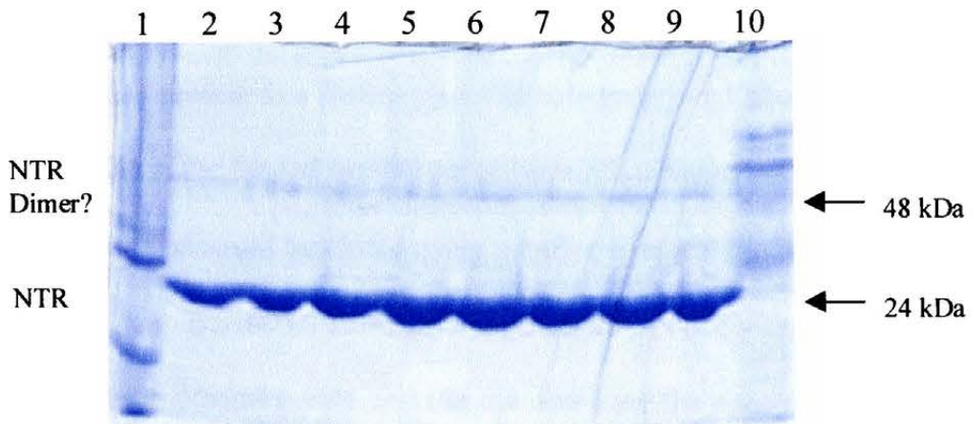
The NTR construct expressed well in the BL21 cells, giving an average of 30 mg of purified protein per litre of LB culture. A bright yellow band could be observed upon loading the crude extract onto the phenyl sepharose column. The yellow nature of the protein solution was subsequently witnessed in all further purification steps and was often used as a crude guide to the NTR content of the samples. The lengthy washing step of the phenyl sepharose column allowed a large proportion of contaminating proteins to be removed from the lysate solution (results not shown). Elution of protein from the column and subsequent SDS-PAGE analysis of the fractions obtained showed a clear, major band at the expected NTR Mwt. of 24,000 Da (Figure 4.1). From the OD 280 nm measurements and gel lane observations, it was decided to pool fractions that possessed at least half the OD 280 nm value of that of the most



Gel from phenyl sepharose column. Lane 1 = size markers, lanes 2-10 = fractions from Tris elution.



Gel from Q-sepharose column. Lane 1 = size markers, Lanes 2-10 = fractions from NaCl gradient wash (Increasing salt concentration left to right).



Gel from hydroxylapatite column. Lane 1 = size markers, Lanes 2-9 = run through fractions, lane 10 = 0.2 M phosphate wash fraction.

Figure 4.1. SDS-PAGE gels from NTR purification steps. Size markers identical on all three gels.

concentrated fraction. After running this pooled sample on the Q-sepharose column, two yellow bands were seen when eluting with the salt gradient. This was attributed to both the NTR enzyme and a band of free FMN, which originated from FMN added to the running buffers, and can presumably interact with the positively charged matrix via its negatively charged phosphate group. From conductivity measurements both of the gradient starter solutions and the eluted protein fractions in conjunction with SDS-PAGE, it could be seen that NTR eluted at a conductivity of 10 mMHO, commensurate with a value of ~200 mM NaCl. Fractions with half-maximal OD 280 nm were pooled, and from SDS-PAGE analysis of the eluant (Figure 4.1), it can be seen that this purification step removes many of the contaminating proteins of higher MWt. The purity of these pooled fractions was deemed not quite ideal, and indeed crystallization trials using protein obtained after just these two column steps yielded no appreciable results. Several other chromatography columns were tried in an attempt to increase the purity of the sample (e.g. trisacryl blue – affinity matrix for nucleotide-binding proteins), but had little effect.

Application of the pooled protein to a hydroxylapatite column proved advantageous. It could be clearly seen that NTR in the loaded sample passed straight through without binding to the matrix. Other bound proteins could be eluted using a buffer comprised of 0.2 M NaH_2PO_4, pH 7.0, and these were ran on a SDS-PAGE gel alongside the NTR fractions (Figure 4.1). The NTR run-through shows improved purity, and the fraction from the phosphate wash shows the contaminants bound by the column. The second band, at a MWt. of ~48,000 Da in the NTR fractions has been attributed to a dimeric form of the enzyme, even though reducing SDS-PAGE was used. All of the "run-through" NTR fractions were pooled and their purity was found to be of sufficient quality for use in crystallographic studies.

4.3.2 Crystallization

From the sparse matrix trials, several conditions were observed that gave crystalline needles, but one well (0.2 M sodium acetate, pH 4.6 and 8 % w/v PEG 4000) gave promising nucleated flat plates, approximately 0.4 mm in the longest dimension (Figure 4.2). Initial incomplete factorial sampling was not found to improve the crystals significantly to a point where they could be used in diffraction experiments. It was decided to investigate the effects of additives, and after an exhaustive screening process (48 small molecule compounds investigated), the inclusion of 15 mM nicotinic acid yielded large, single crystals (up to a size of 1 x 0.5 x 0.4 mm, Figure 4.3). The success of nicotinic acid was assumed to be related to its similarity to the nicotinamide headgroup of NAD(P)H. All crystals obtained had a yellow colour and were found to be sensitive to mechanical disruption, favouring the chance that they were protein crystals. From observation of the crystals alone, it was obvious that two forms could be grown in the same drop - one square and plate-like, and the other elongated and lozenge-shaped (Figure 4.3). Typical "final" conditions for producing these crystals were : 10-15 mg/ml protein equilibrated against a reservoir solution containing 0.2 M sodium acetate, pH 4.6, 8-12 % w/v PEG 4000 and 15 mM nicotinic acid. In the interests of cryo-protection for sample mounting and data collection, it was discovered that the crystals could be grown with 25 % v/v ethylene glycol incorporated into the reservoir solution, allowing direct freezing of crystals in mother liquor and eliminating any need for post-growth soaking.

4.4 Summary

NTR was over-expressed successfully and a purification protocol established that resulted in a NTR solution of > 95 % purity. Purified NTR was used in a series of crystallization trials,

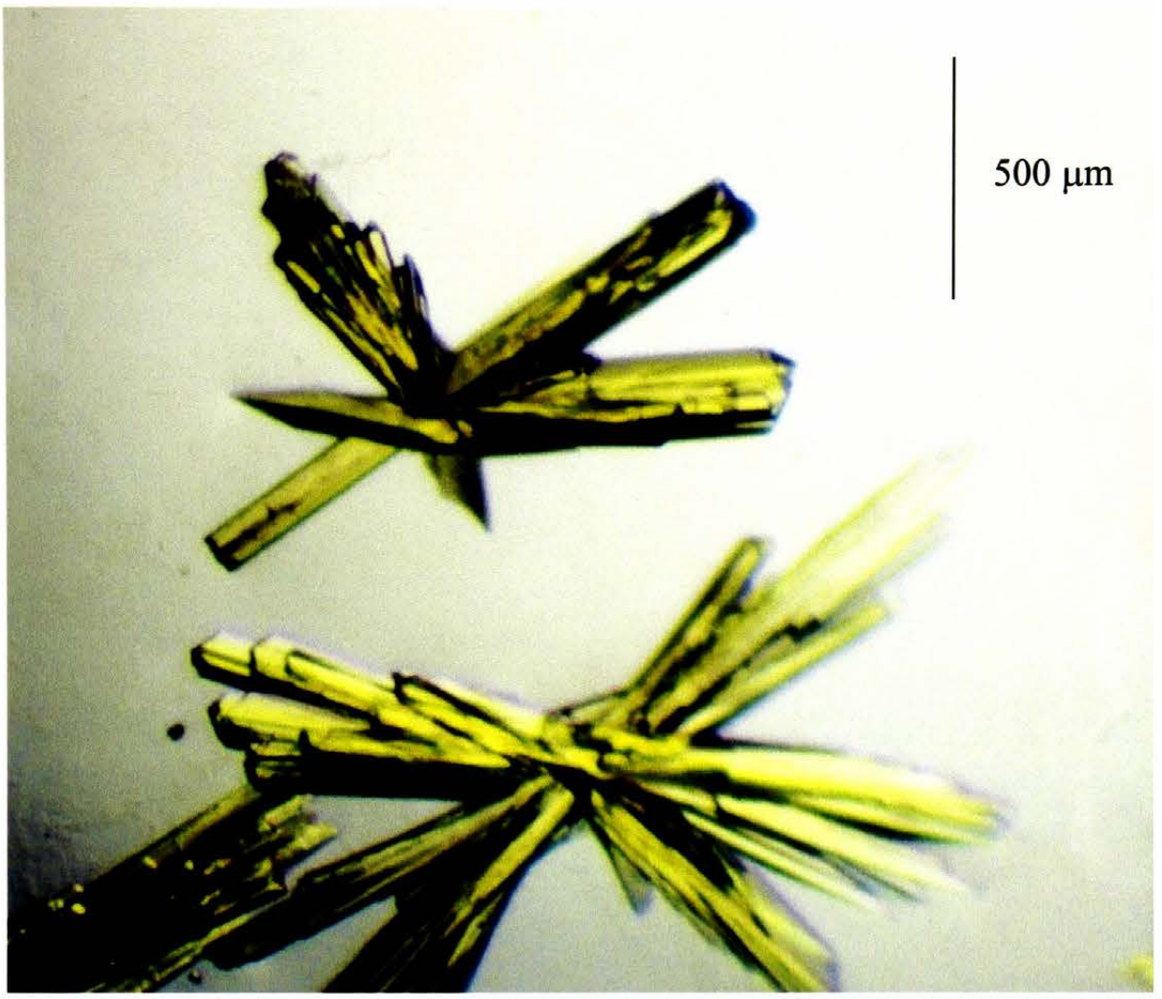


Figure 4.2. Crystals from initial sparse matrix screening experiments. Crystals grown using vapour diffusion experiments and Hampton Research Crystal Screen I solution 37 (0.2 M sodium acetate, pH 4.6, 8 % w/v PEG 4000).

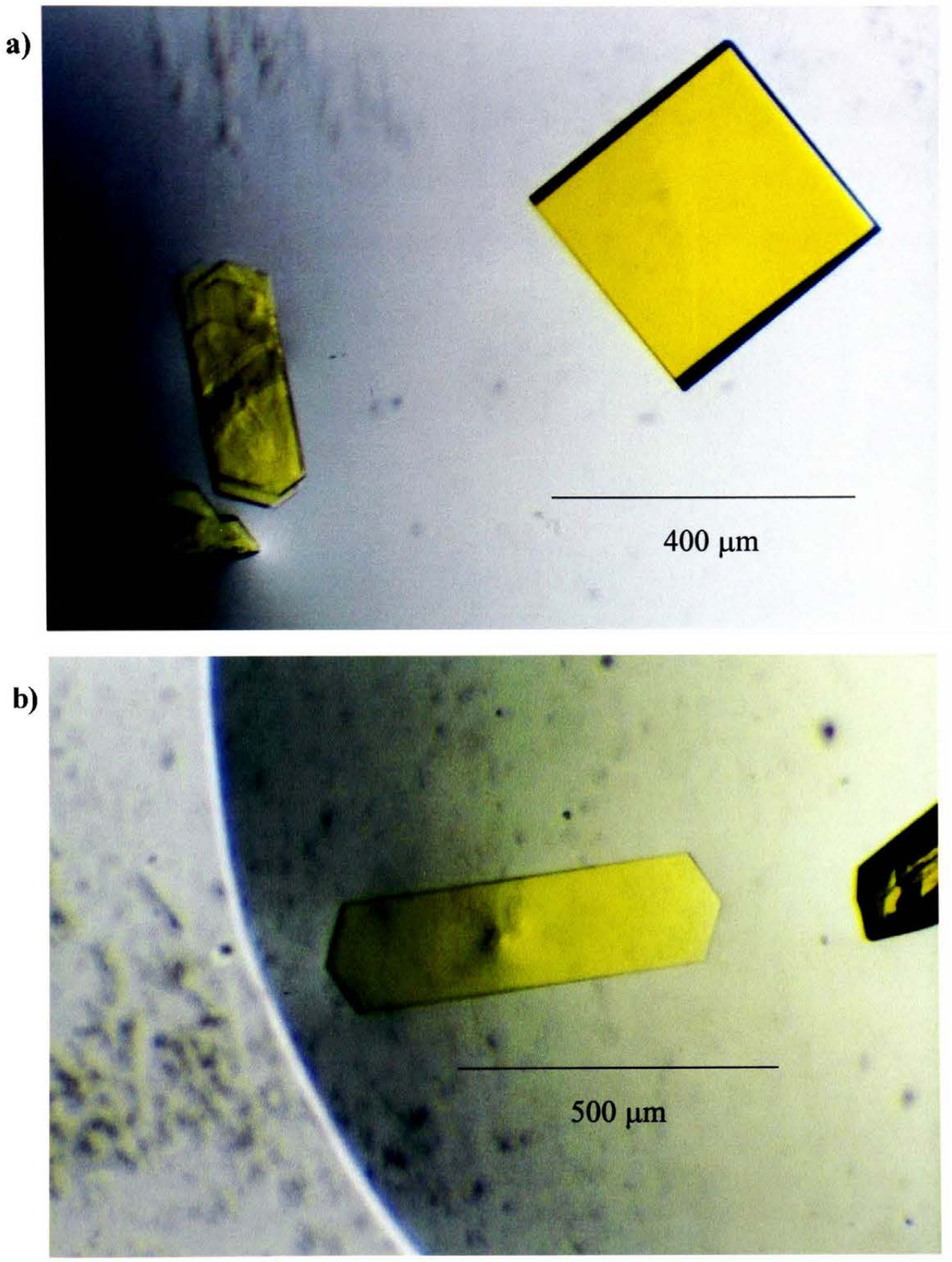


Figure 4.3. Single NTR crystals from optimised growth conditions: a) tetragonal form; b) monoclinic form. Crystals grow from identical conditions of 100 mM sodium acetate, pH 4.6, 8-12 % w/v PEG 4000, 15 mM nicotinic acid and 25 % v/v ethylene glycol. Protein used at a concentration of 12 mg/ml, 3 μl protein + 3 μl reservoir solution drops, equilibrated over a reservoir solution of 600 μl. Crystals appear within 1-2 days and reach maximum size after a period of a week. The difference in shape between the two isoforms can be easily seen.

culminating in the growth of single crystals of a suitable size for X-ray diffraction studies. Multiple crystal forms of NTR were observed. Owing to an increase in the power of phasing programs in the past few years (e.g. DM Multi, Cowtan 1994), multiple crystal forms are now seen as advantageous and no longer such a detriment to structure determination via their lack of isomorphism. It will also be of use to attempt to phase and solve all forms, reducing bias and error with several independent determinations of the NTR structure.

CHAPTER 5

SOLVING THE STRUCTURE OF NTR

5.1 Introduction

Having obtained protein crystals (Chapter 4), the first step was to test their diffraction qualities, and if suitable, propose a method for estimating the phases of the diffracted X-rays. With a successful derivative (as judged by the criteria provided in Chapter 3), map calculation and model building can be commenced, hopefully leading to a situation where most of the protein of interest can be visualised in the electron density map.

5.2 Materials & Methods

5.2.1 Native Crystal & General Data Collection

5.2.1.1 General Data Collection

Crystals were taken direct from the growth drop using a Hampton cryo-loop and flash-cooled to 100 K in a stream of nitrogen gas using an Oxford cryostream apparatus (Oxford Cryosystems, Oxford, UK). Unless stated otherwise, all diffraction data were indexed using DENZO, and integrated and scaled using SCALEPACK (Otwinowski *et al.* 1997). All subsequent data manipulations were accomplished using the CCP4 suite of programs (CCP4 1994).

5.2.1.2 Native Crystal Data Collection

Initial diffraction images of the native crystals were collected at: i) the Synchrotron Radiation Source (SRS), Daresbury on beamline 9.6 and ii) on the *in-house* RAXIS system. Data

collection for the published monoclinic and native tetragonal crystal forms were collected at the European Synchrotron Radiation Facility (ESRF), Grenoble, France on beamline ID14-1.

5.2.2 Preparation of a Selenomethionine Derivative

5.2.2.1 Selenomethionyl Protein Preparation

The pJG6A1b plasmid was used to transform the methionine auxotroph strain B834 D3 and grow a starter culture in 50 ml "half-strength" LB +amp. Ten ml of this starter culture were used to inoculate 500 ml of pre-warmed modified M9 minimal media (containing 0.4 % w/v glucose, 42 mM Na_2HPO_4, 22 mM KH_2PO_4, 20 mM NH_4Cl, 8 mM NaCl, 4 mM nicotinamide, 8 μM thiamine, 10 μM biotin, 6 mM uracil, 2 mM guanosine, 5 mM adenine, 5 mM thymine, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 10 μM $FeSO_4$, 40 mg/ml selenomethionine, 100 μg/ml ampicillin, 0.4 mM H_3BO_3, 30 μM $CoCl_2$, 10 μM $CuSO_4$ and 80 μM $MnCl_2$). The culture was left to shake at 37°C until an OD 595 nm of 0.55 was reached. NTR expression was then induced by the addition of 1 mM IPTG. Subsequent steps are the same as those for the native protein protocol, with the exception that all buffers contained 10 mM DTT to prevent oxidation of the selenomethionine residues.

5.2.2.2 Crystallization of the derivative

Crystals were grown using the same protocol as that provided for the native protein, except that 10 mM DTT was added to the protein crystallization stock.

5.2.3 Data Collection Using Derivative Crystals

5.2.3.1 Crystal Characterisation

From crystal appearance alone, it was presumed that two crystal isoforms had grown using the same conditions. This was confirmed during initial diffraction experiments, with the square-shaped crystals belonging to the tetragonal $P4_12_12$ spacegroup, and the lozenge-shaped crystals belonging to the monoclinic $P2_1$ spacegroup. Upon crystallizing and obtaining diffraction images of the selenomethionine derivatives, it was apparent that the tetragonal crystals were isomorphous with the native tetragonal form, but the monoclinic crystals had been "replaced" by a third, orthorhombic $P2_12_12_1$ crystal form that looked identical in appearance to the monoclinic form. Images of the crystal forms are shown in Figure 5.1.

5.2.3.2 MAD data collection on Orthorhombic SeMet NTR

Data were collected at the ESRF on beamline ID14-4. Crystals of a poorer quality were used for a fluorescence spectrum taken using a RonTec detector and multi-channel analyser to obtain the optimal wavelengths at which to collect the peak and inflection point datasets. A crystal of higher quality was used for data collection. The order of the three wavelengths collected was peak (pk – 0.9795 Å), inflection point (ip – 0.9793 Å) and remote (rm – 0.9393 Å), in order to salvage the experiment should radiation damage occur (*i.e.* use the pk dataset for SAD phasing). Complete datasets were collected at all three wavelengths.

5.2.3.3 SAD data collection on Tetragonal SeMet NTR

Following the MAD data collection on the orthorhombic form, it was decided to collect some data on the tetragonal selenomethionyl NTR crystal form. In contrast to the orthorhombic phasing, and in tandem with some radiation damage experiments (collaboration with Drs

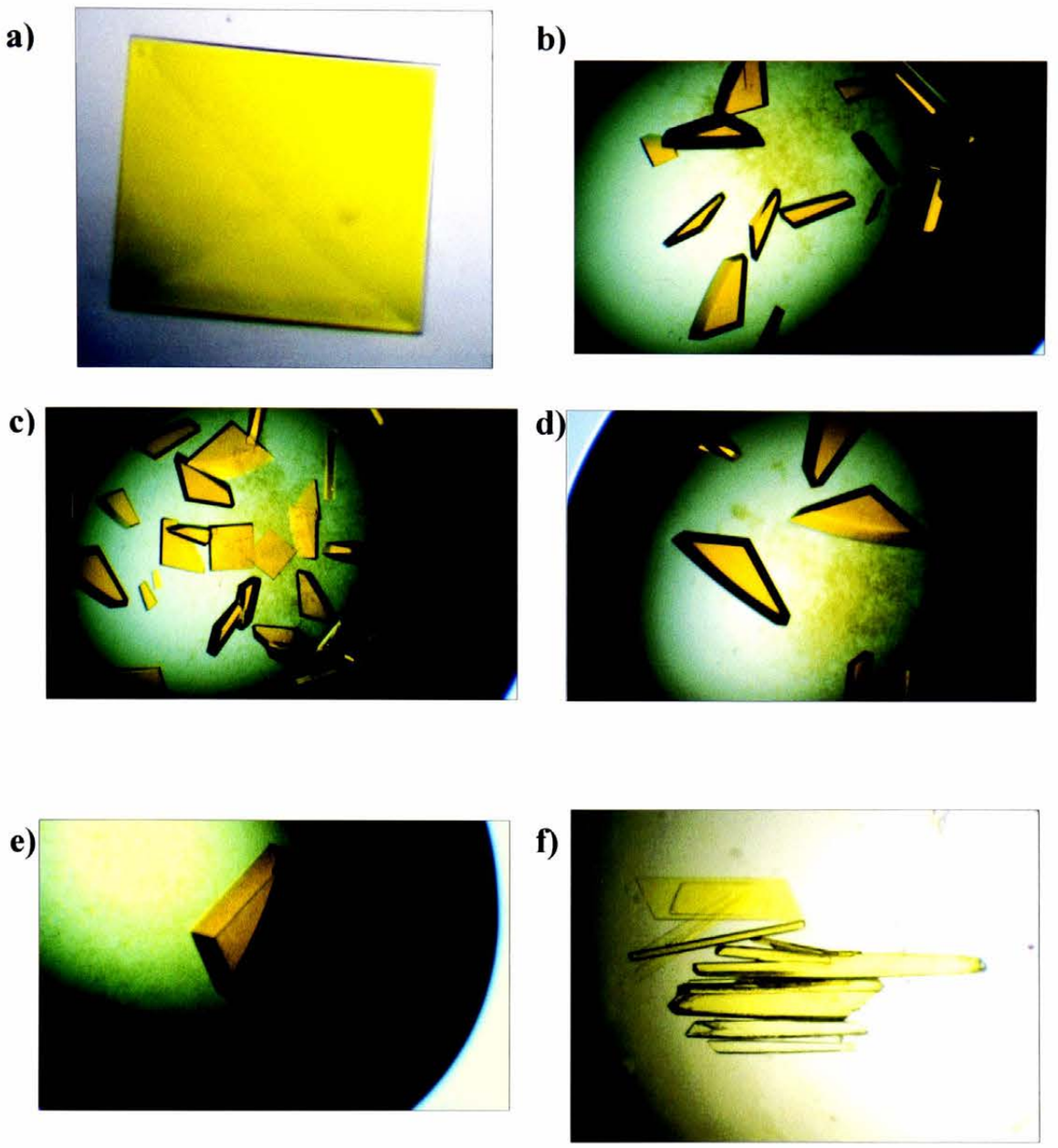


Figure 5.1. Observation of different crystal morphologies grown using optimised conditions. Photos a) to e) tetragonal crystals, note the square / half-square appearance and shortest physical dimension (c axis). Crystals all roughly on same scale, longest dimension is 400 μm. Photo f) monoclinic crystals, note the plate / rod appearance and tendency to cluster from nucleation. Longest dimension is around 800 μm.

Raimond Ravelli and Sean McSweeney, ESRF, Grenoble, France), a SAD experiment was chosen in order to compare the phases obtained with those from a more traditional MAD experiment. Data were collected (at the pk wavelength of 0.979 Å) at the ESRF on beamline ID14-4. From the fluorescence spectrum of the crystals it was apparent that the selenomethionine residues were partially oxidised (results not shown). This would be deleterious as the mixed oxidation state of the residues would severely lower the anomalous signal from the experiment. It could be seen that the crystals could be returned to a fully reduced state by the addition of 1 µl β-mercaptoethanol to the reservoir solution in the well, and allowing vapour diffusion of the reductant to occur over a period of time. As a large crystal was used for this process (400 x 400 x 100 µm), sufficiently sized pieces could be broken off for monitoring of the oxidation state using the Rontec detector. After 6 hours the crystal could be shown to be fully reduced, and was deemed usable for data collection.

5.2.4 Phasing

5.2.4.1 Phasing of Selenomethionyl Crystals

5.2.4.1.1 MAD Phasing of Orthorhombic Crystals

The three scaled wavelengths (pk, ip and rm) were read into SOLVE (Terwilliger and Berendzen 1999), and the number of anomalous scatterers set to 16, even though 20 were expected (4 monomers from a V_m calculation, 5 Met per monomer). It was preferred not to count the N-terminal methionine, assuming that flexibility of the protein chain would result in the residue not having a defined orientation, or that the residue had possibly been cleaved off during protein production. A resolution cutoff of 30-2.5 Å was used in phasing calculations.

5.2.4.1.2 SAD Phasing of Tetragonal Crystals

The scaled data were entered into the CNS suite heavy-search protocol (Grosse-Kunstleve and Brunger 1999), with an expected 8 scatterers and a resolution cutoff of 15-3 Å used in phasing calculations.

5.2.4.2 MR Phasing of Monoclinic Crystals

The scaled data were entered into the CNS suite molecular replacement protocol (Brunger 1990), using the fastdirect option for the rotation function search. The solved orthorhombic NTR co-ordinates were used as the search model (just the AB dimer). A resolution cutoff of 15-4 Å was applied in both the rotation and translation functions. An initial search was used to locate one dimer, which was then fixed and used in locating the second dimer.

5.2.5 Model Building

5.2.5.1 Initial Model Building With the MAD Data

Located selenium sites were used to initially place methionine residues. The electron density was of sufficient quality to trace the polypeptide chain using knowledge of the protein primary sequence. All model building was done using the molecular graphics package TURBO-FRODO (Roussel *et al.* 1991).

5.2.5.2 Further Model Building With Other Crystal Forms

The located selenium sites from the SAD phasing results were used as a guide to move a pre-built dimer from the refined orthorhombic structure determination into the tetragonal form unit cell (based on interatomic vectors between the heavy atom positions matching vectors calculated from the Met positions of the dimer). For the monoclinic molecular replacement

solution, the fit of the rigid-body refined model was good enough, large scale model movement was not required.

5.2.6 Refinement Protocols

Each of the three structures was refined using the CNS suite of programs (Brunger *et al.* 1998). A maximum likelihood target was used with cycles of rigid body refinement, simulated annealing, powell minimization and individual B-factor refinement, interspersed with manual rebuilding in TURBO-FRODO (Roussel and Cambillau 1991) using $2mF_o$-DF_c SIGMAA-weighted (Read 1986) electron density maps as a guide. The experimental phases from the MAD and SAD datasets were included when refining the orthorhombic and tetragonal structures, respectively. Non-crystallographic symmetry constraints or restraints were not applied at any time.

5.3 Results & Discussion

5.3.1 Data Collection

5.3.1.1 Native Crystals

The first native crystal form to be used in a diffraction experiment was the tetragonal isoform. Due to a long c axis of ~260 Å, crystal mounting (and subsequent mechanical disruption) led to disorder along this axis and "smearing" of the diffraction image (Figure 5.2). With fine phi-slicing and a detector with excellent spatial resolution, it would have been possible to collect to the crystal diffraction limit of ~1 Å. In practice, with both the native and derivative tetragonal crystals, a diffraction limit of ~1.7 Å could be collected. Full data collection statistics for the native dataset used to refine the tetragonal NTR structure are shown in Table 5.1.

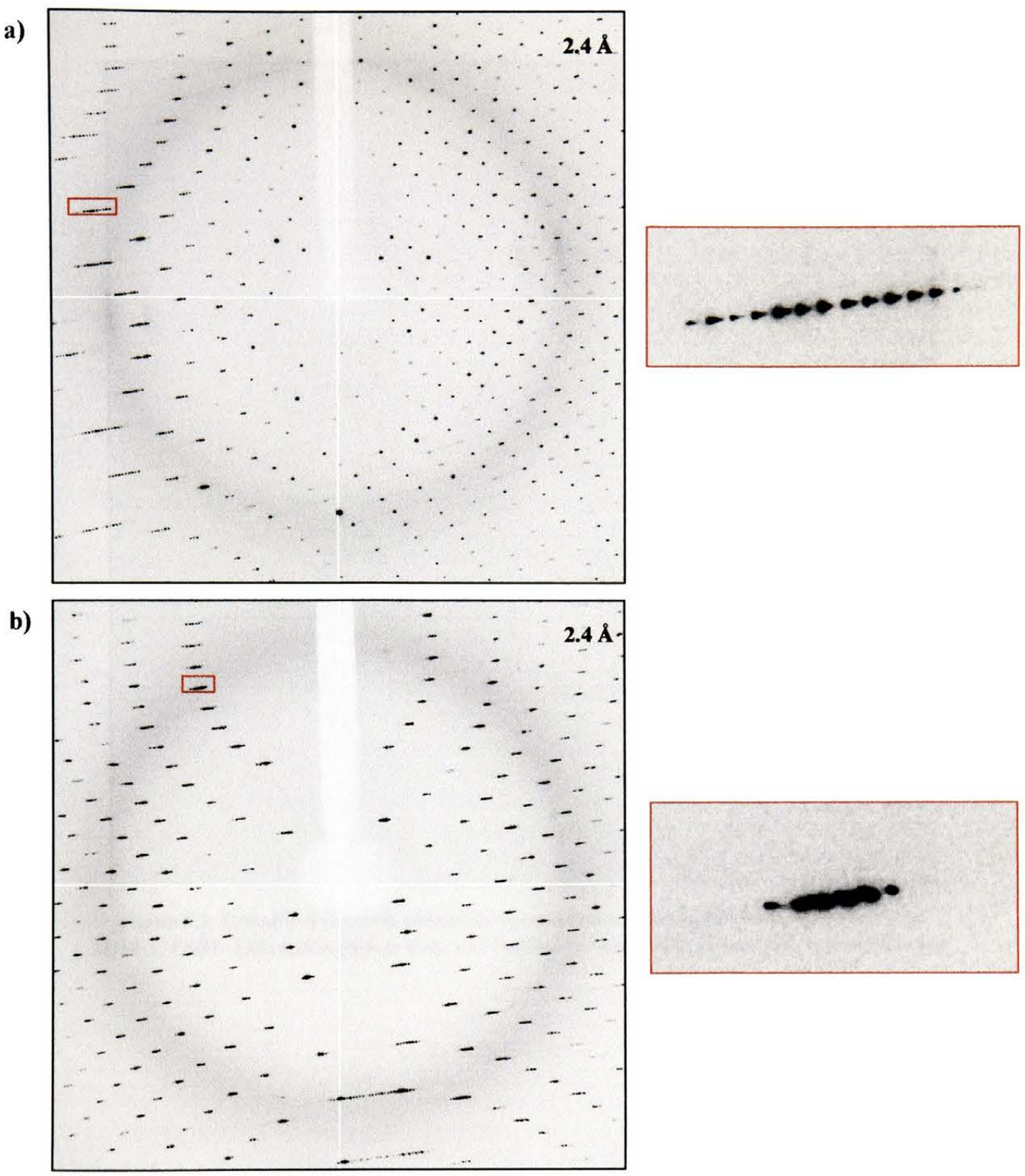


Figure 5.2. The figure highlights some of the problems facing tetragonal crystal form data collection. Data collected from a sample of tetragonal crystals, beamline ID14-1, ESRF. Images from a 0.5° oscillation, X-ray exposure identical for both diffraction patterns. a) "good" diffraction image with resolvable c-axis spots (close up on right). b) "poor" diffraction image with smeared, non-resolvable c-axis spots (close up on right). Resolution indicated in corner of main images.

Figure 5.3. Example diffraction pattern from monoclinic crystal form, collected at beamline ID14-1, ESRF. Diffraction pattern from a 1° oscillation. Resolution indicated in corner of image.

Native protein crystallization also yields a monoclinic crystal form. This has far smaller lattice parameters than the tetragonal form (longest axis, c ~120 Å), and diffracts to ~1.6 - 1.7 Å resolution. The crystals are very plate-like and this can lead to anisotropy regarding exposure levels during data collection. Full data collection statistics for the monoclinic dataset used in MR and monoclinic form refinement are shown in Table 5.1. A typical diffraction image from these crystals is shown in Figure 5.3.

5.3.1.2 MAD Data Collection

From the fluorescence spectra results (data not shown), the wavelengths used in data collection were pk = 0.9785 Å, ip = 0.9793 Å and rm = 0.9393 Å. No significant loss in diffraction spot intensity was observed during the data collection. Full statistics for all three wavelength datasets are shown in Table 5.1.

5.3.1.3 SAD Data Collection

Following the six hour equilibration with β-mercaptoethanol (see section 5.2.3.3), the diffraction pattern of the crystal looked unaffected by this procedure. A single pk dataset was collected using a wavelength of 0.979 Å. Full statistics for this dataset are shown in Table 5.1.

Crystal Properties	**Orthorhombic**			**Tetragonal**		**Monoclinic**
Spacegroup	$P2_12_12_1$			$P4_12_12$		$P2_1$
Cell Dimensions (Å)	58 x 120 x 144			57 x 57 x 262		72 x 58 x 117, β=104°
Solvent Content	50.5 %			43.1 %		48.0 %
Copies in ASU	4			2		4
Data Collection[a]		**MAD**		**SAD**	**Nat**	**Nat**
	IP	PK	RM	PK		
Wavelength (Å)	0.9793	0.9795	0.9393	0.979	0.934	0.934
No. Observations	118,238	118,089	123,859	103,813	364,515	611,309
No. Unique Reflections	38,896	38,851	39,289	29,994	48,108	85,582
Completeness (%)[b]	98.1 (91.5)	98.0 (90.9)	99.1 (98.5)	98.4 (99.1)	96.9 (94.7)	98.9 (97.0)
Redundancy	3.0 (2.2)	3.0 (2.2)	3.2 (2.7)	3.5 (3.2)	7.6 (5.6)	7.1 (6.1)
$I/\sigma(I)$	8.5 (2.8)	8.9 (2.5)	9.8 (3.2)	9.1 (8.0)	8.8 (3.6)	8.4 (1.3)
R_{sym} (%)[c]	5.5 (15.3)	5.4 (15.2)	4.9 (12.3)	6.6 (8.0)	5.6 (19.8)	6.7 (54.9)
Resolution (Å)	30-2.4	30-2.4	30-2.4	30-2.0	30-1.7	30-1.8
Structure Determination						
Figure of Merit		0.75			0.52	N/A
No. of Se Atoms Found		16 / 20			8 / 10	N/A

[a]Data sets are labelled: Nat = native; PK, IP and RM are the wavelengths corresponding to the peak, inflection point and remote energies

[b]Values in parentheses are for data in the highest resolution shell.

[c]$R_{sym} = \Sigma_j|\langle I\rangle - I_j|/\Sigma\langle I\rangle$, where I_j is the intensity of the j^{th} reflection and $\langle I\rangle$ is the average intensity

Table 5.1. Crystal properties and data collection statistics of derivative and native NTR crystal forms

5.3.2 Phasing

5.3.2.1 MAD Phasing

The program SOLVE (Terwilliger and Berendzen 1999) was used to find 16 Se sites (from a requested 16), with an overall Z-score value of 59.8 (breakdown of individual Z-score values in Table 5.2), and a mean figure of merit (FOM) of 0.75. The sites (Table 5.3) had an average height above sigma of 20.6, with no site possessing a value less than 14. The selenium positions showed non-crystallographic symmetry (NCS) commensurate with the expected 4 protein monomers per ASU.

Criteria	VALUE	Z-SCORE
Pattersons	12.0	8.20
Cross-Validation	176	39.4
NatFourier CCx100	44.4	18.2
Mean FOM x 100	54.8	11.0
Correction for Z-scores		-17.0
Overall Z-score value		59.8
After running SOLVE with NOBAYES keyword to obtain "traditional" phasing statistics :		
Phasing Power	1.11	

Table 5.2. Scoring results from MAD Phasing using SOLVE (Terwilliger and Berendzen 1999). Value column indicates score given to that criteria by SOLVE (on an arbitrary scale), Z-score column alters this value to put all criteria on the same scale. For structures incorporating this amount of heavy atoms, overall Z-scores above 20 are significant.

Site	X	Y	Z	Occupancy	B factor (Å^2)	Peak Height
1	0.693	0.646	0.161	0.959	23.218	24.82
2	0.854	0.829	0.139	0.760	15.000	24.45
3	0.985	0.764	0.215	0.688	15.035	21.77
4	0.621	0.629	0.147	0.860	43.271	14.39
5	0.083	0.662	0.181	0.745	15.000	22.87
6	0.454	0.293	0.241	0.664	31.358	16.48
7	0.341	0.086	0.069	0.794	15.000	25.24
8	0.155	0.684	0.181	0.855	24.193	17.38
9	0.947	0.510	0.087	1.060	31.956	22.35
10	0.140	0.191	0.188	0.789	15.000	25.63
11	0.524	0.209	0.176	0.399	15.000	21.92
12	0.070	0.175	0.180	0.900	26.191	21.19
13	0.284	0.378	0.194	0.885	21.895	22.06
14	0.227	0.086	0.164	0.821	25.982	18.58
15	0.588	0.233	0.177	0.773	22.141	15.04
16	0.778	0.539	0.171	0.987	49.429	17.08

Table 5.3. Heavy atom (Se) positions from MAD phasing. Co-ordinates expressed as fractions of cell axes.

5.3.2.2 SAD Phasing

The heavy atom search protocol in the CNS package (Grosse-Kunstleve and Brunger 1999) found 8 sites (from a requested 8), with an average height above sigma of 15.8 and no site possessing a value less than 9. The mean FOM was 0.52, and the correct hand for the sites was determined using the MAD selenium positions as a guide (comparing interatomic vectors from both possible hands of the SAD heavy atoms with those obtained from the MAD heavy atoms). The sites showed a clear relation to those obtained from the MAD phasing, and NCS was observed accounting for the expected 2 monomers per ASU.

5.3.2.3 MR Phasing

The cross-rotation list of peaks (top 10 trials, based on rotation function Rfactor) is shown in Table 5.4, with 4 very clear solutions, of which there are 2 pairs related by 180° in Theta3. Following a translation search to place the first dimer, a monitor value of 0.4 and a packing

value of 0.3 can be seen for the correct solution. Upon fixing this dimer and placing the second dimer, a final monitor value of 0.75 and packing value of 0.6 are observed.

Trial Number	Theta1	Theta2	Theta3	RF-function
1	179.211	68.365	92.000	0.2303
2	183.742	65.577	270.954	0.2273
4	182.348	60.000	99.319	0.2249
9	176.413	58.941	282.472	0.2187
18	180.605	57.212	90.606	0.0998
20	182.348	57.212	265.377	0.0954
25	185.136	68.365	103.501	0.0826
26	184.755	73.941	279.402	0.0823
27	183.316	51.635	273.316	0.0815
35	178.501	65.577	290.137	0.0717

Table 5.4. Top peaks for monoclinic form cross-rotation search. Rotation angle given in Eulerian format.

5.3.3 Model Building

Upon determination of the selenium atom substructure from the MAD orthorhombic dataset, it was apparent that the 16 sites corresponded to four monomers in the ASU. These monomers were assigned chain identities A-D. No density modification or NCS averaging was necessary to improve the map as the electron density was of excellent quality and sufficient to start tracing the polypeptide backbone. Initial unrefined electron density for the orthorhombic form is shown in Figure 5.4. It was decided to build one monomer completely and then use this and the selenium sites to replicate all four monomers.

Building of all secondary structure elements and loops proceeded without trouble, and problems were only encountered when trying to establish the presence or absence of the N-terminal methionine, which was built in some monomers and left out in others. The flavin cofactor and nicotinic acid ligand (denoted NIC in figures) could also be seen in these starting

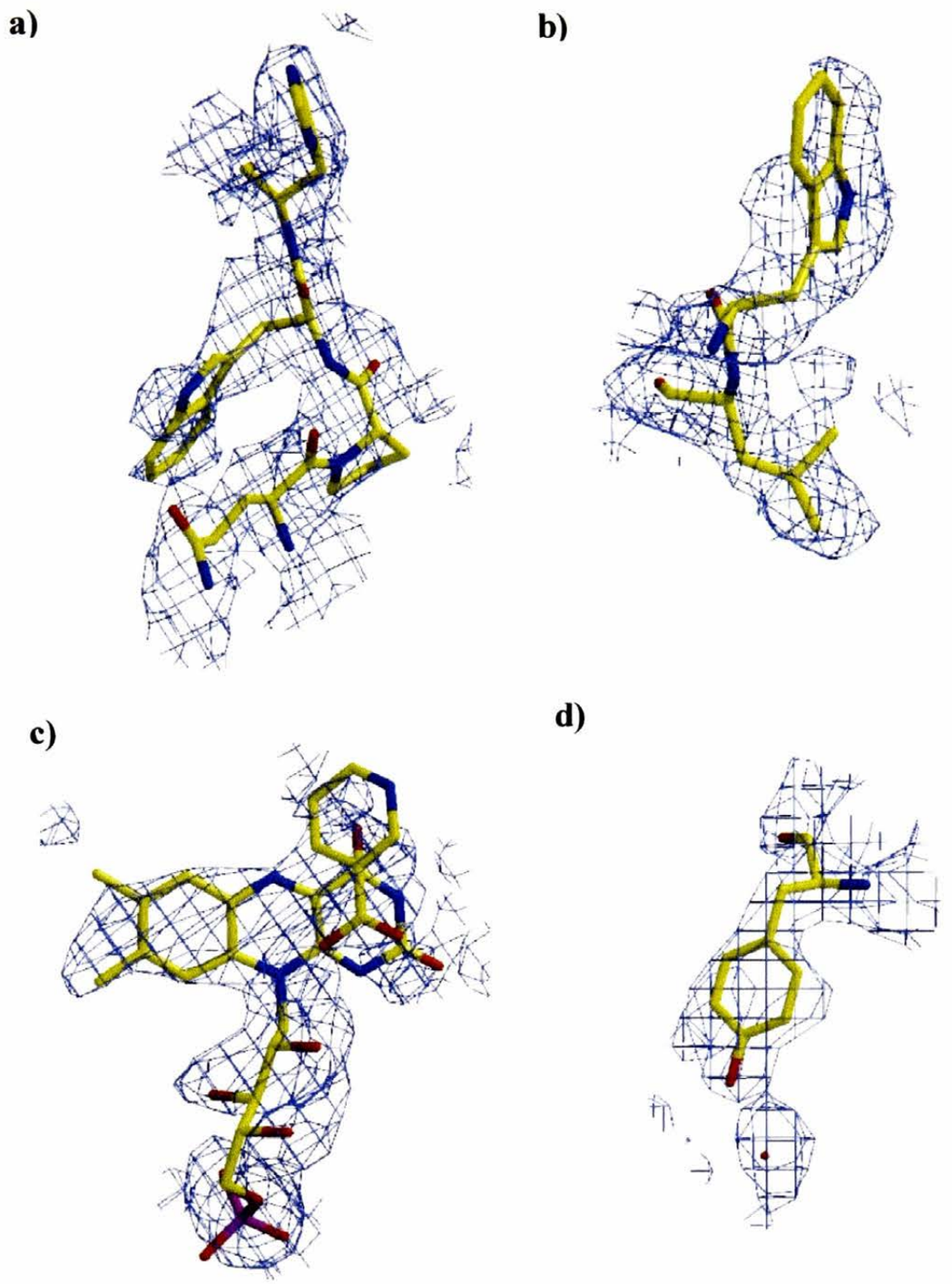


Figure 5.4. Initial electron density map calculated from MAD phases, with an overlay of structural elements from the final refined orthorhombic co-ordinates. Map is 2mFo-DFc, contoured at 1σ, 2.4 Å resolution. Regions of structure are: a) B44-47, b) B94-95, c) FMN & NIC from monomer A and d) B144 and W51. Note the excellent quality of the initial phases and subsequent electron density, where most residues, the cofactor and water molecules can be visualised. Figure prepared using BOBSCRIPT (Esnouf 1997) and RASTER3D (Meritt and Bacon 1997).

maps (Figure 5.4). Electron density from initial maps calculated from the SAD derived phases would have been more difficult to interpret *de-novo*, but significant improvement can be seen upon density modification and NCS averaging (data not shown). However, it was sufficiently easy to use the orthorhombic traced polypeptide chain in both the tetragonal and monoclinic model building steps.

5.3.4 Refinement

The progress of refinement for the orthorhombic structure can be seen in Table 5.5. Final refinement statistics for all three crystal forms are shown in Table 5.6. Refined electron density for the structures can be seen in Figures 5.5 to 5.7.

Refinement Step	Rfactor (%)	Rfree (%)
Initial 2° structure fitting	43.5	43.5
Side chain build	31.6	35.2
Minimisation and B-factor	27.8	31.0
Cofactor Fitting	26.0	28.4
1st round water picking	23.6	26.5
2nd round water picking	22.0	25.1
Optimise weighting	20.9	24.5
Thorough model examination	20.6	24.4
Final model	19.8	24.2

Table 5.5. Refinement progress of the orthorhombic NTR structure determination. Rfree from a random 5 % of reflections, not used in refinement.

5.4 Summary

Having obtained 2 crystal forms for both native and derivative NTR protein crystals, MAD, SAD and MR were used to obtain phase information. Solution of the orthorhombic form by MAD provided the basis for solving the other two crystal forms. SAD phasing of the selenomethionine tetragonal form allowed selenium site comparison with the MAD results,

	Orthorhombic	Tetragonal	Monoclinic
Refinement			
Data Set	RM	Native	Native
Resolution (Å)	2.4	1.7	1.8
No. of Non-hydrogen Atoms	7515	4038	7027
No. Water Molecules	643	586	139
R factor[a]	19.8 %	18.2 %	22.6 %
R_{free}[b]	24.2 %	20.4 %	23.8 %
Average B factor (protein atoms only)	21.0 Å^2	16.9 Å^2	44.9 Å^2
RMSD Bond Angles[c]	1.88°	1.57°	1.12°
RMSD Bond Lengths[d]	0.018 Å	0.015 Å	0.006 Å
Copies in ASU	4	2	4

[a] R factor = $\Sigma_{hkl}| |F_o| - |F_c| |/\Sigma_{hkl}|F_o|$.

[b]R factor based on 5 % of the data withheld from refinement

[c]RMSD bond angles and [d]bond lengths are the root mean square deviations from ideal values

Table 5.6. Final refinement statistics for all three NTR structure determinations.

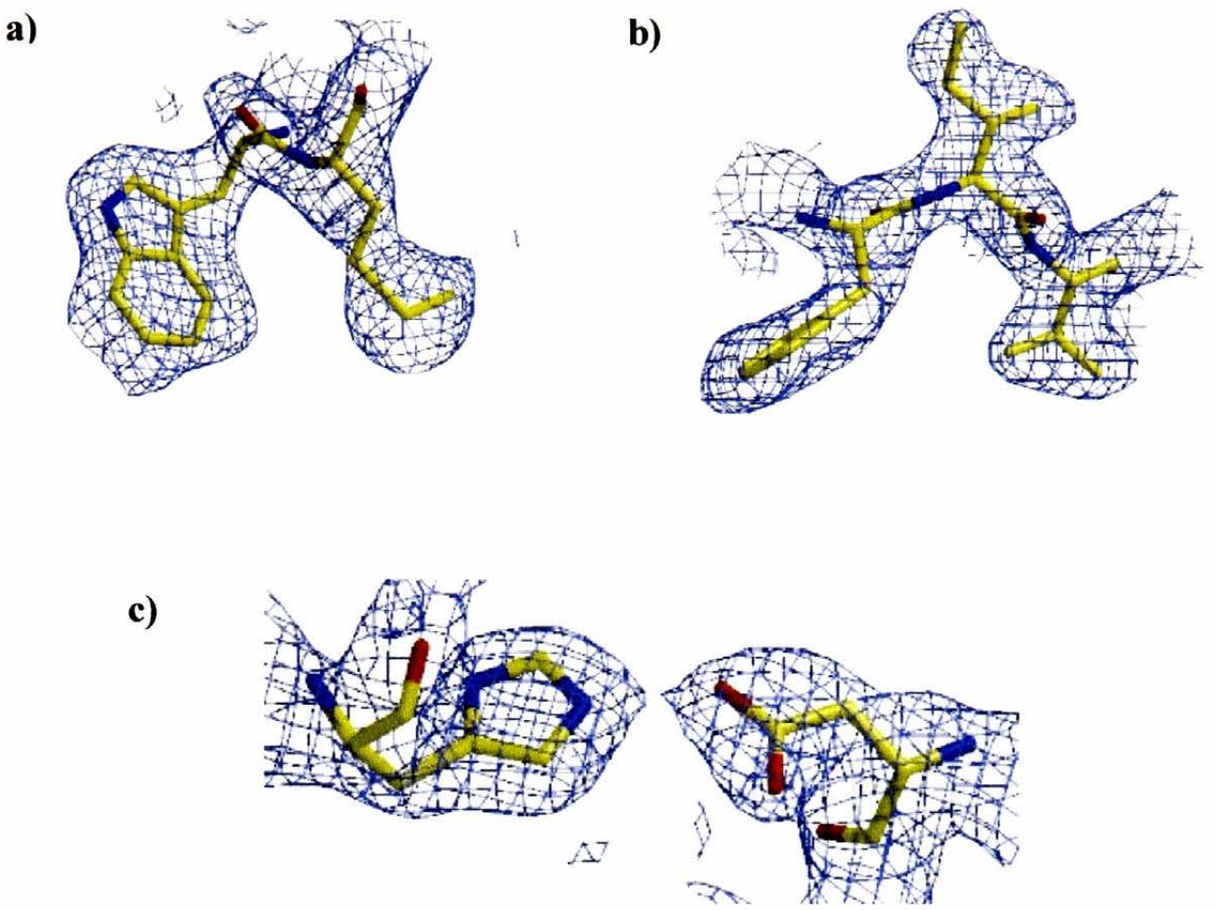


Figure 5.5. Refined electron density and co-ordinates from the orthorhombic NTR structure determination. Map is 2mFo-DFc, SIGMAA-weighted, contoured at 1σ, 2.4 Å resolution . Regions shown are : a) B138-139, b) B48-50 and c) A198 & A11. Figure prepared using BOBSCRIPT (Esnouf 1997) and RASTER3D (Meritt and Bacon 1997).

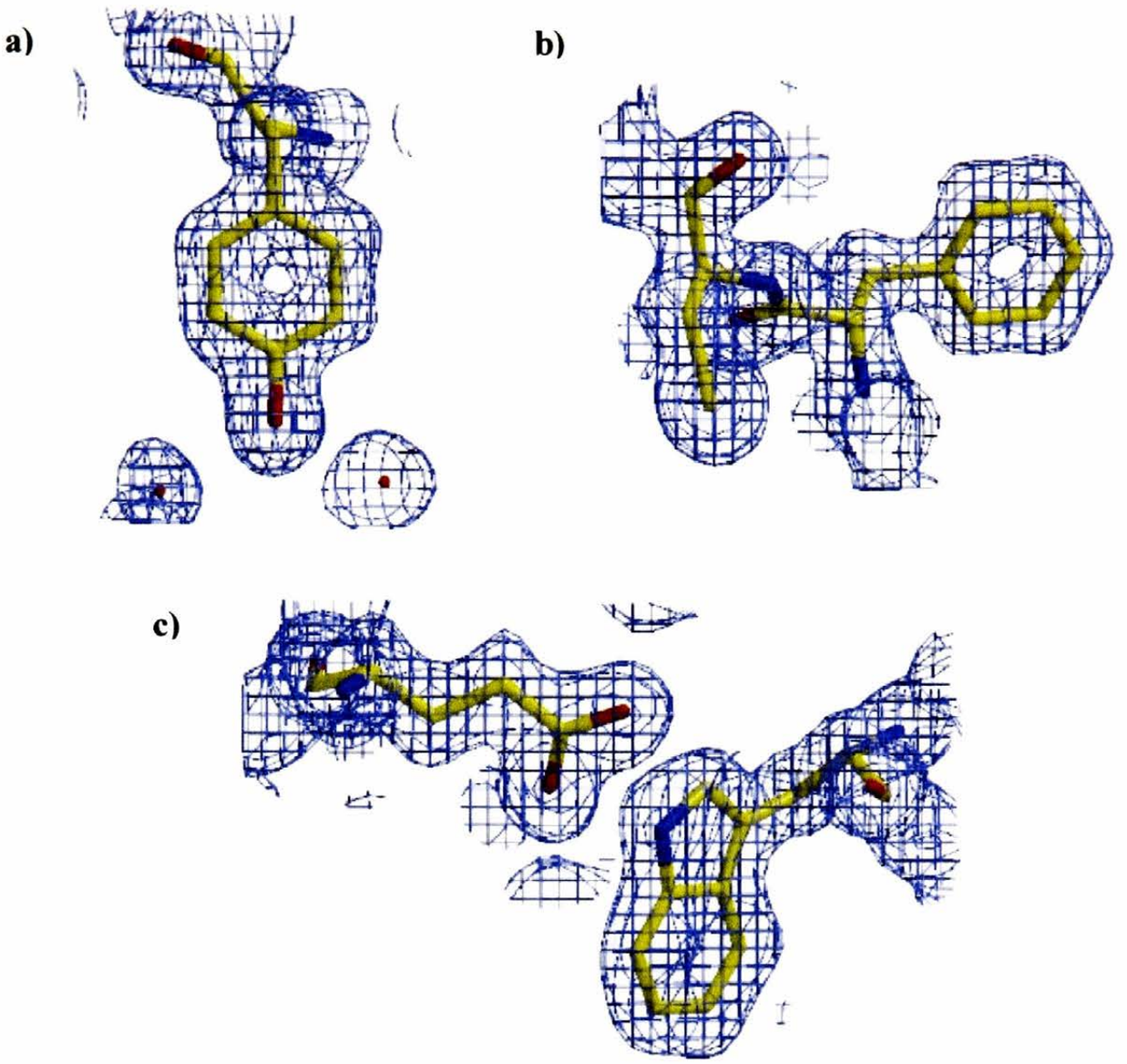


Figure 5.6. Refined electron density and co-ordinates for the tetragonal NTR structure determination. Map is 2mFo-DFc, SIGMAA-weighted, contoured at 1σ, 1.7 Å resolution. Regions shown are : a) B36, W6 & W67, b) A84-85 and c) A165 and B138. Figure prepared using BOBSCRIPT (Esnouf 1997) and RASTER3D (Meritt and Bacon 1997).

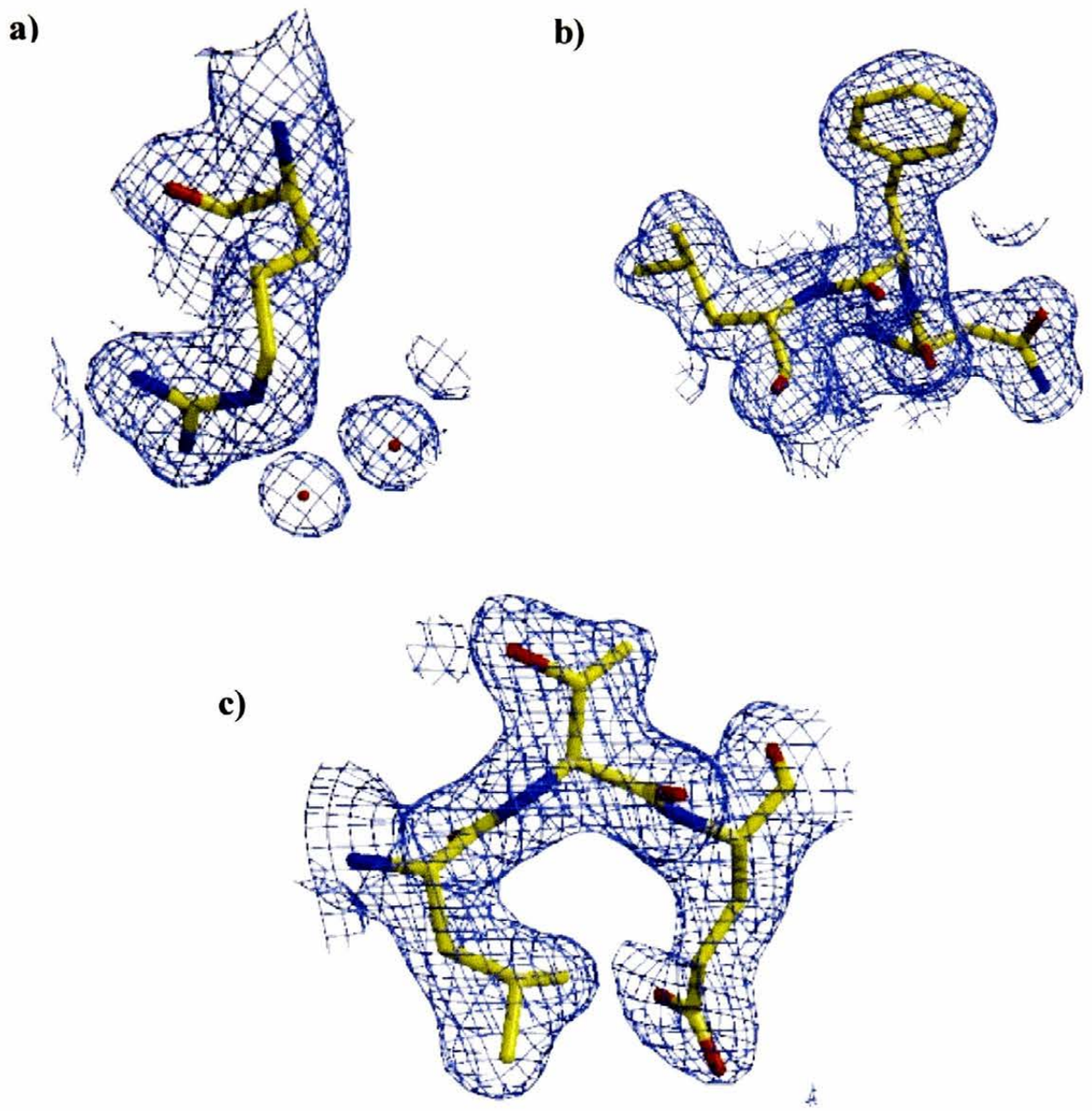


Figure 5.7. Refined electron density and co-ordinates from the monoclinic NTR structure determination. Map is 2mFo-DFc, SIGMAA-weighted, contoured at 1σ, 1.8 Å resolution. Regions shown are : a) A10, W21 and W58, b) A149-151 and c) A214-216. Figure prepared using BOBSCRIPT (Esnouf 1997) and RASTER3D (Meritt and Bacon 1997).

and the tetragonal structure was refined against a higher resolution native dataset. Finally, MR with the orthorhombic form co-ordinates as a model was used to solve the native monoclinic crystal form. A tabulated form of these results is given in Table 5.7.

Crystal Form	Protein Used	Cell	Phasing	Refinement
Orthorhombic $P2_12_12_1$	SeMet	58 x 120 x 144 Å	MAD (Solve)	Maximum likelihood with experimental phases
Tetragonal $P4_12_12$	SeMet (for SAD) Native (to refine)	57 x 57 x 262 Å	SAD (CNS)	Maximum likelihood with experimental phases
Monoclinic $P2_1$	Native	72 x 58 x 117 Å $\beta = 103^o$	MR (CNS)	Maximum likelihood

Table 5.7. Summary of Chapter 5 – Solving the Structure of NTR.

CHAPTER 6

STRUCTURAL ANALYSIS OF NTR

6.1 Introduction

Following the refinement of the three crystal forms of NTR (Chapter 5), we can now start to look at the general fold of the protein and analyse how its various secondary structure elements interact. After the general topology has been investigated, we can look at specific residues in the active site and start to identify key steps in the catalytic process. It will also be of use to study differences between the three crystal forms, and also compare our structures to those of previously determined NTR homologues.

6.2 Results

6.2.1 Secondary Structure Assignments

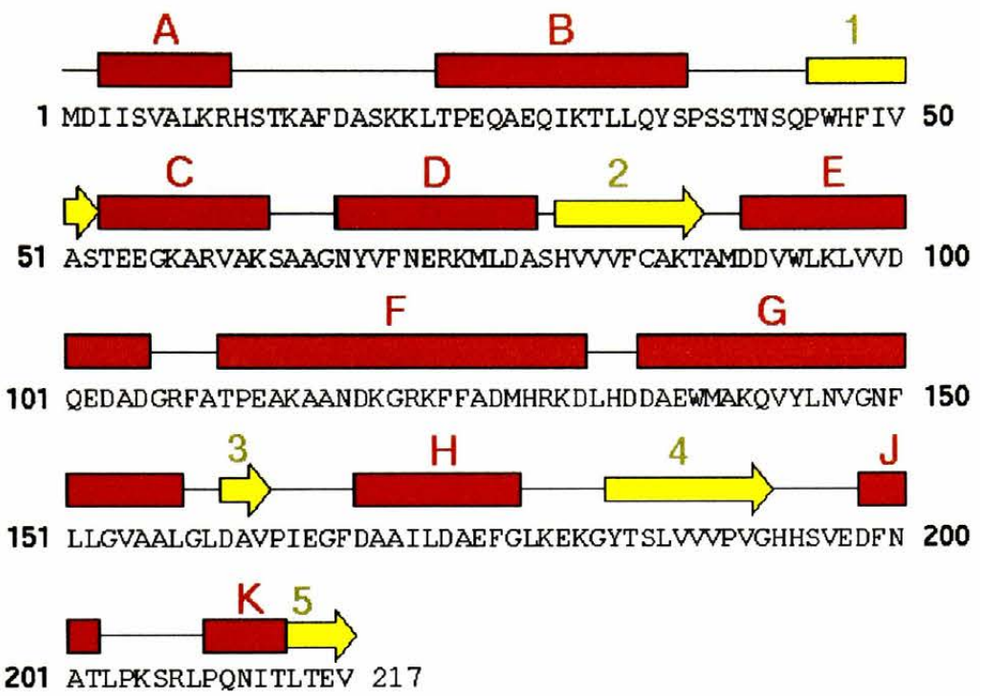
Using the A chain monomer of the refined tetragonal NTR structure as a representative of the different monomers in the three crystal forms, secondary structure assignment was calculated with PROCHECK (Laskowski *et al.* 1993). The 10 α-helices were designated A to K (with I excluded to avoid confusion with 1). The 5 β-strands were designated 1 to 5 inclusive. The occurrence and placement of these secondary structure elements in the NTR primary sequence is shown in Figure 6.1a. Loops are named with respect to the secondary structure elements that proceed and follow them *e.g.* B-1 is the loop between helix B and strand 1.

6.2.2 Topology & Description of the Protein Fold

6.2.2.1 General Topology

The main body of the monomer consists of a large hydrophobic core (residues 2-91 and 131-217), with a small two-helix domain (helices E and F, residues 92-130) protruding from this.

a)



b)

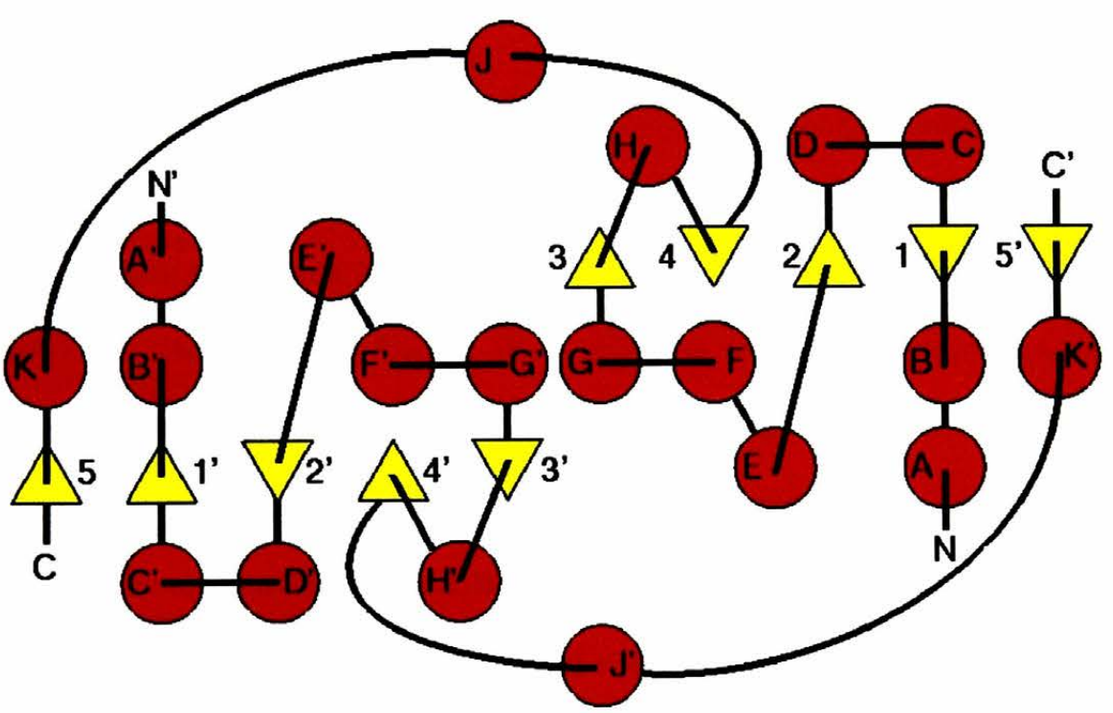


Figure 6.1. (a) Secondary structure assignments for NTR. The α-helices are represented by rectangles and the β-strands by arrows. (b) TOPS (Westhead *et al.* 1999) topology diagram for the NTR dimer. Circles represent α-helices and triangles represent β-strands (downward = into plane of diagram, upward = out of plane of diagram). Connections drawn from edge to centre represent loops that project out of the plane of the diagram, connections drawn from centre of edge represent loops that project into the plane of the diagram. All secondary structure elements were assigned according to PROCHECK (Laskowski *et al.* 1993) analysis.

Two orthogonal views of the NTR dimer and this domain structure can be seen in Figure 6.2. Each monomer is comprised of a central twisted β-sheet, flanked on both sides by α-helices. Each β-sheet contains four anti-parallel strands from one monomer and a fifth parallel strand contributed by the five C-terminal residues of the other monomer. Helices B and G contact one face of the β-sheet, whilst helices C, D and H lie on the other. The general fold of the protein, built up one secondary structure element at a time can be seen in Figure 6.3. The active site contains one tightly bound FMN cofactor, and a molecule of the nicotinic acid ligand used to crystallize the protein. Each active site is comprised of residues from both monomers and lies in a solvent-exposed channel at the dimer interface, with the two separate FMN N5 atoms 26.4 Å apart. Each channel is lined by helix F, and three loops (B-1, C-D and 3-H).

6.2.2.2 The dimer interface

A full description of specific monomer : monomer residue contacts is provided in Table 6.1. The dimer interface is extensive and has a buried surface area of 3,686 $Å^2$ per monomer, as calculated by AREAIMOL (CCP4 1994). The two N-termini are positioned close together, with the D2 side-chains only 3.1 Å apart. The A helices sit between the grooves formed by the G helices, which run parallel together at an angle of approximately 25° to the non-crystallographic axis of symmetry. Helix A also contacts helix B of the opposing monomer. The N-terminus (helices A & B and loop A-B) contacts the C-terminus (helices J & K, loop J-K, strand 5), as the chain feeds into the other monomer and completes the β-sheet. The C-terminus also makes contacts to helix E of the opposing monomer.

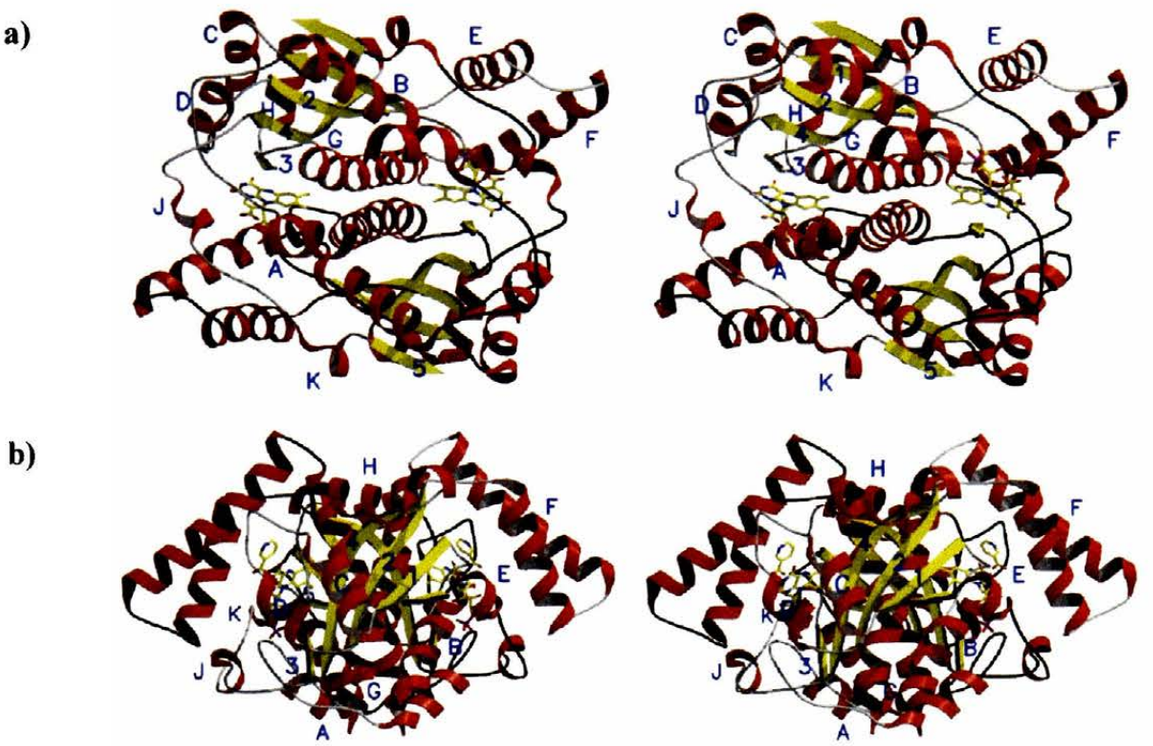


Figure 6.2. Parallel stereo diagram of NTR Fold: (a) view looking down the non-crystallographic 2-fold axis; (b) view of dimer rotated approximately 90° from (a) about the horizontal axis. For one monomer, with the backbone coloured light grey, the α-helices (red) are labeled A to K and β-strands (yellow) are labeled 1 to 5. The FMN cofactor and nicotinic acid ligand are represented in stick form and are coloured by atom type (C, yellow; N, blue; O, red; P, mauve). The diagram was generated using MOLSCRIPT (Kraulis 1991) and RASTER 3D (Meritt and Bacon 1997).

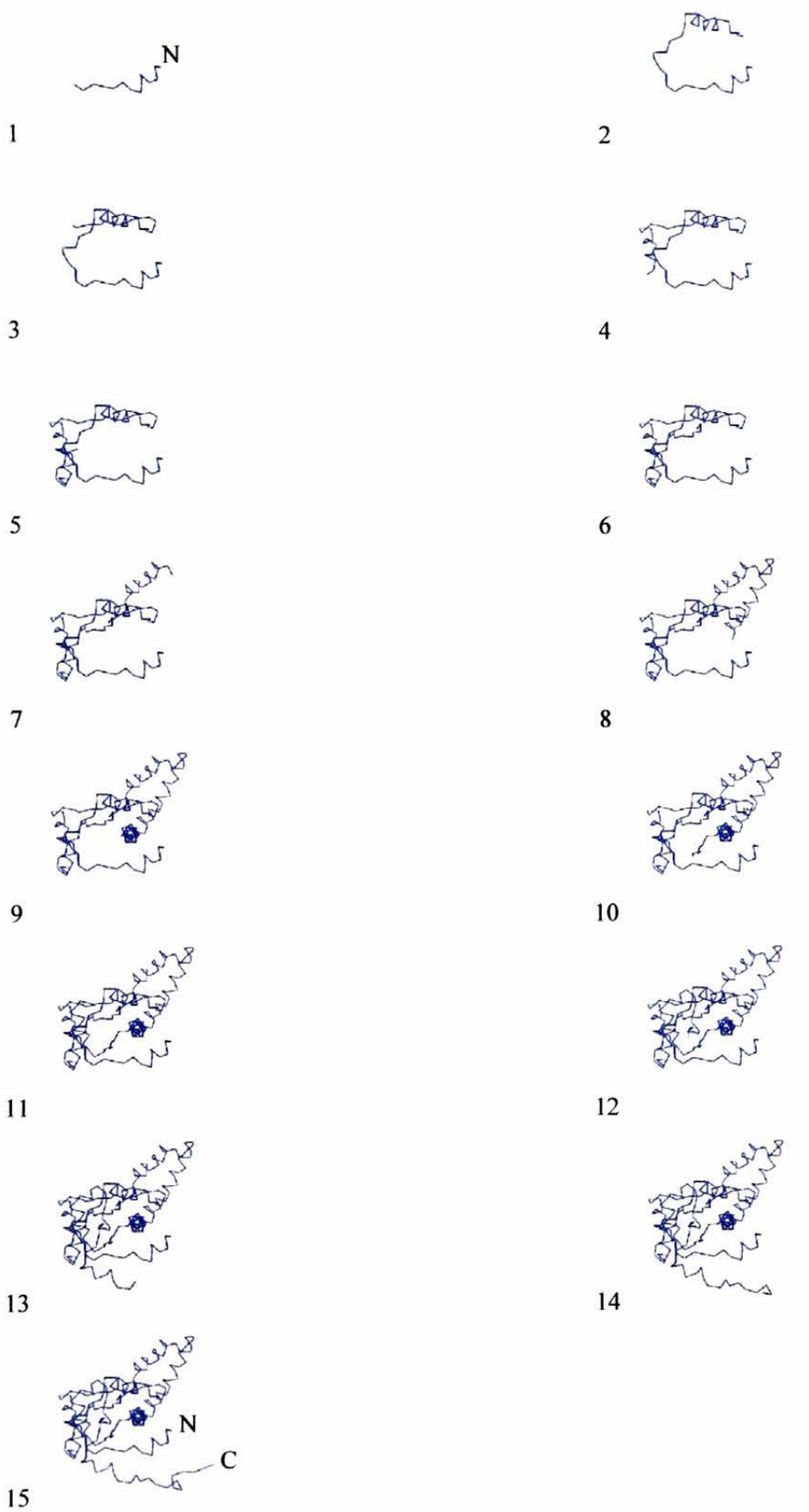


Figure 6.3. Schematic trace of protein, starting at the N-terminus (1) and ending at the C-terminus (15). Secondary structure elements are added to the diagram sequentially. Figure prepared using MOLSCRIPT (Kraulis 1991) and RASTER 3D (Meritt and Bacon 1997).

Helix A : Helix B	**B-1 Loop : G-3-H Loops**
I4 : L33 hydrophobic	P38 : L151 hydrophobic
L8 : Y36 hydrophobic	S40 : E165 hydrogen bond
N-terminus : C-terminus	
K31 : L214 hydrophobic	**Helix A : Helix G**
K31 : E216 salt bridge	I3 : L157 hydrophobic
Q35 : L208c hydrogen bond	I3 : A 156 hydrophobic
Y36 : K205 hydrophobic	
S37c : R207 hydrogen bond	**Helix D : Helix F**
Q44 : L208c hydrogen bond	Y68 : F124 hydrophobic
H47 : T215 hydrogen bond	
F48n : T213 hydrogen bond	**Helix E : C-terminus**
F48c : T215n hydrogen bond	W94 : L208 hydrophobic
V50 : L214 hydrophobic	L97 : I212 hydrophobic
V50n : T215c hydrogen bond	Q101 : L208 hydrophobic
A51 : V217 hydrophobic	D105 : S206 hydrogen bond
S52n : V217c hydrogen bond	R107 : N200 hydrogen bond
Helix G : Helix G	R107 : P204c hydrogen bond
Y144 : K141 hydrophobic	
L145 : G148 hydrophobic	**Helix G : 3-H Loop**
L145 : V147 hydrophobic	W138 : E165 hydrogen bond

Table 6.1. Monomer : monomer contacts in the NTR dimer. Backbone nitrogens are denoted by n and carbonyls by c, other contacts involve the residue side-chain. Amino acids labelled using standard single letter code.

6.2.3 The Active Site, FMN & Ligand Binding

6.2.3.1 FMN Binding

The FMN cofactor is located in a pocket comprised of residues from both monomers, as shown in Figure 6.4. The flavin cofactors are tightly bound to the enzyme, and are well-defined in the electron density (see Figure 6.5) with an average B factor of 11.5 $Å^2$ for the tetragonal structure form (the protein has an average B factor of 16.9 $Å^2$). The FMN shows a "butterfly" conformation, observed in many other flavoenzymes (Barber *et al.* 1992), with an angle between the planes of the two outer rings of approximately 15°.

The phosphoribityl tail and dimethylbenzoid parts of the flavin are buried in the protein fold, but the *re* face of the isoallozaxine ring is solvent exposed at its polar end. The *si* face of the

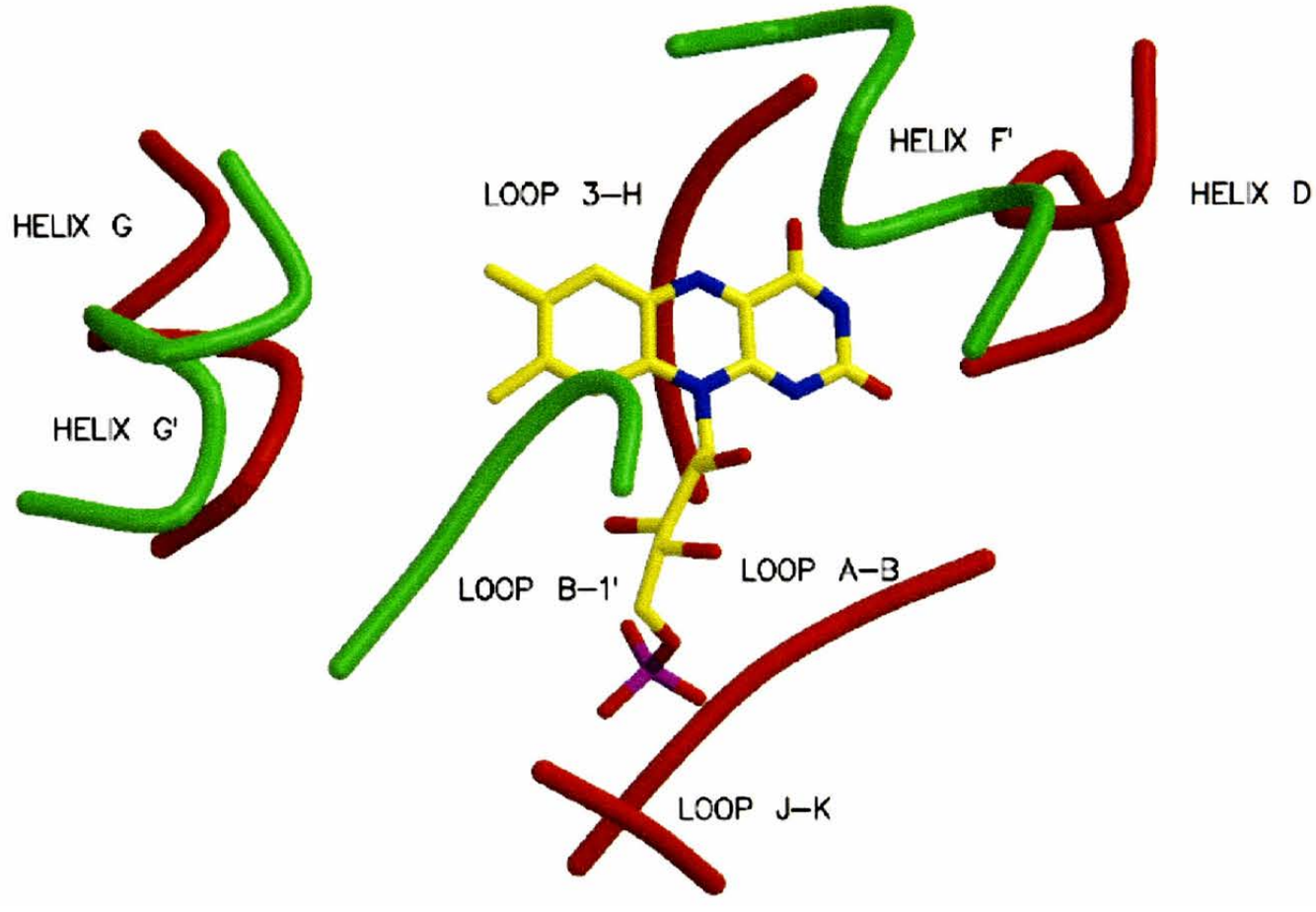


Figure 6.4. Topology of active site and positioning of FMN cofactor with respect to the various secondary structure elements. Monomer A is represented in red, and monomer B in green. The FMN cofactor is coloured according to atom type (C, yellow; O, red; N, blue; P, mauve). A large solvent channel in the protein crystal lies to the right of the diagram, beyond helix D. Most of the cofactor binding is contributed by residues from one monomer (in this case A), with loop B-1 and helix F from the opposing monomer forming a binding pocket over the face of the flavin. The loop 3-H runs down the other side of the active site, forming hydrophobic contacts to the FMN. Both G helices pack hydrophobic residues against the dimethylbenzoid moiety of the cofactor. Figure prepared using MOLSCRIPT (Kraulis 1991) and RASTER 3D (Meritt and Bacon 1997).

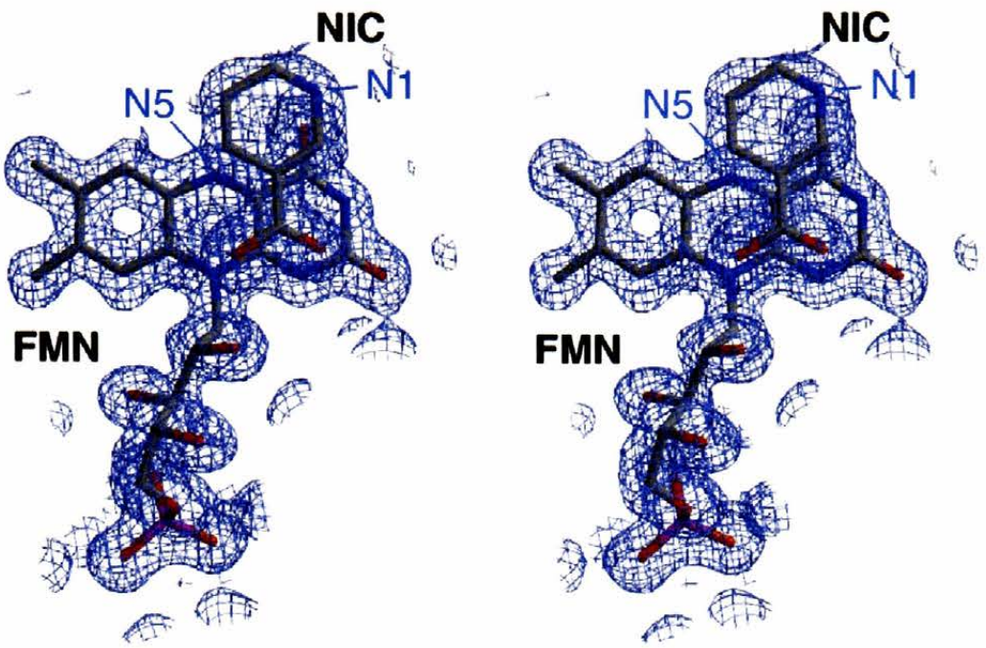


Figure 6.5. Parallel stereo diagram of FMN and NIC refined electron density from the tetragonal crystal form. Map is 2mFo-Dfc, resolution 1.7 Å, SIGMAA-weighted and contoured at 1σ. Diagram coloured by atom type (C, grey; N, blue; O, red; P, magenta). N1 of NIC and N5 of FMN labelled in blue. Note the position of NIC C4 (directly opposite N1 on the ring of NIC) over N5 of the flavin. Diagram prepared using BOBSCRIPT (Esnouf 1997) and RASTER 3D (Meritt and Bacon 1997).

FMN is largely buried and contacts β-strand 3 and some of the loop A-B. Hydrophobic contacts to the non-polar dimethylbenzoid ring of the flavin are provided by the G helices from both monomers of the NTR dimer.

All water molecules mentioned hereafter are numbered with respect to the tetragonal NTR crystal structure. Hydrogen bonding contacts to the FMN are exclusively from one of the monomers only. Asparagine 71 forms two polar contacts with the flavin, N71-ND2 to FMN O4, and N71-OD1 to FMN N3. The FMN O4 also hydrogen bonds to the backbone nitrogen of G166, and the FMN N3 to a water molecule W4. The backbone nitrogen of E165 hydrogen bonds to atom N5 of the cofactor. The FMN O2 makes two hydrogen bonds to the protein, the NZ atoms of lysines 14 and 74, and one to a water molecule W551. The ribityl tail hydrogen bonds to an array of ordered waters, FMN O2* with W74, FMN O4* with W37, and FMN O5* with W86, and only once to the protein – FMN O3* with atom OG of S12. Serine 12 also forms two hydrogen bonds to the phosphate group of FMN, O2P with S12-OG, and O1P with the S12 backbone nitrogen. The phosphate also has several other hydrogen bonds or potential salt bridges with the protein, atom O2P with NH1 of R10, and atom O3P with the NH2 of R207 and atom NZ of K205. A diagram of these contacts is provided in Figure 6.6.

6.2.3.2 Nicotinic Acid Ligand Binding

Although the majority of contacts to the FMN cofactor are provided by just one of the monomers, the opposing monomer contributes several residues that form the binding pocket for the nicotinic acid (NIC) ligand. The NIC ligand has excellent electron density in all three crystal forms (tetragonal example in Figure 6.5) and sits sandwiched between the flavin cofactor and F124, approximately 3.5 Å from both. The ligand also appears to be

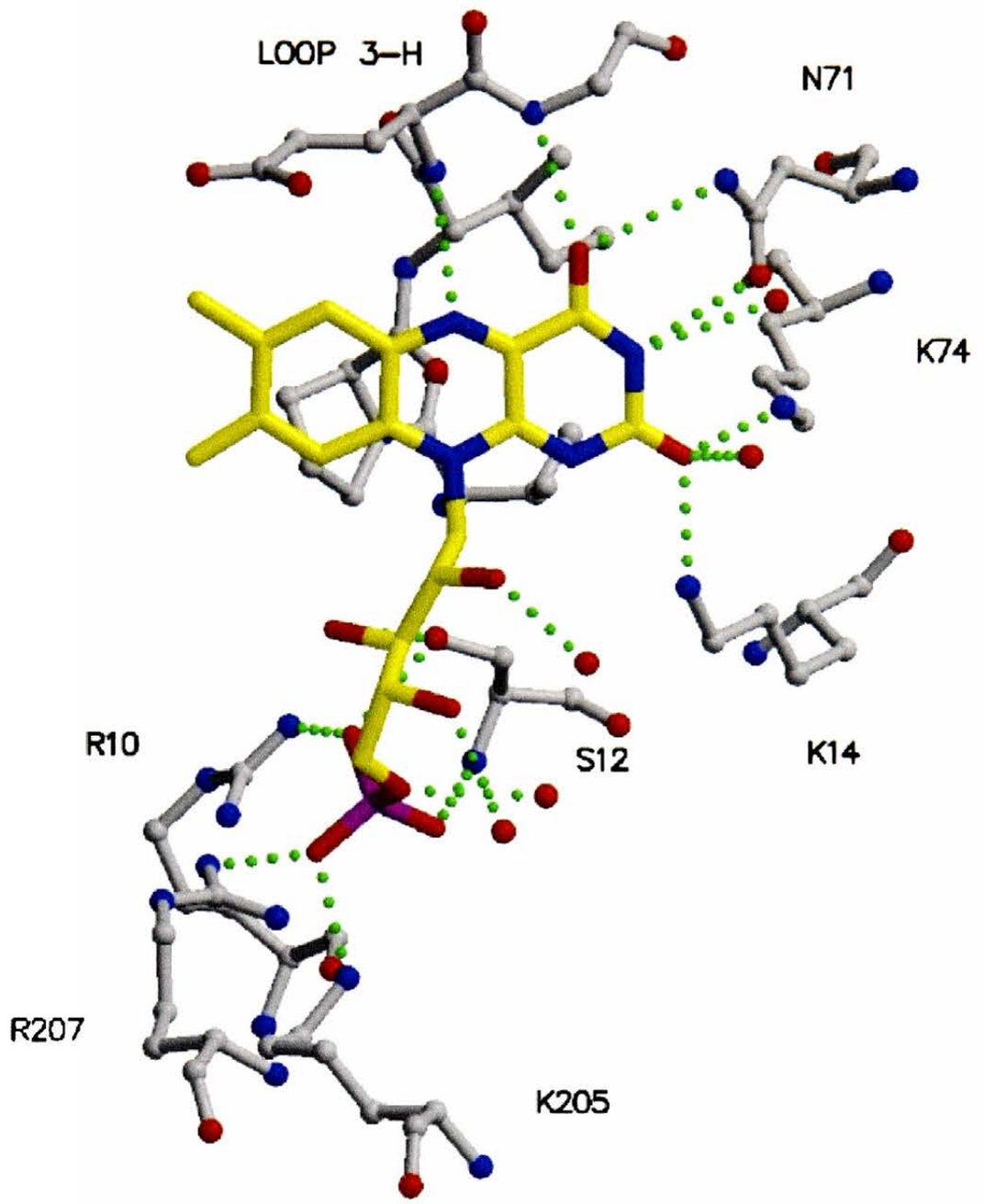


Figure 6.6. Schematic of hydrogen bond and salt bridge contacts between FMN cofactor and protein side chains or bound water molecules. Protein side chain carbons are represented in grey, and FMN carbons in yellow. Oxygen atoms (both of protein, cofactor and water molecules) are coloured red, and nitrogen atoms blue. Hydrogen bonds and/or potential salt bridges are shown by dotted green lines. Loop 3-H consists of residues 162-167 between β-strand 3 and α-helix H. Figure prepared using MOLSCRIPT (Kraulis 1991) and RASTER 3D (Meritt and Bacon 1997).

well-ordered, with an average B-factor of 18.6 $Å^2$. Apart from the Van der Waals contacts between the FMN isoalloxazine and nicotinic acid rings, the cofactor has a further contact to the ligand, with the carbonyl group O7 of the ligand hydrogen bonding to the FMN ribityl O2*. The nicotinic acid only makes one contact to the protein, NIC O7 hydrogen bonding to the backbone nitrogen of T41. The other carbonyl oxygen of the ligand makes two short hydrogen bonds to ordered waters W551 (2.8 Å) and W542 (2.6 Å), and a possible longer interaction to W303 (3.4 Å).

A diagram of all of the active site residues (omitting F124 for clarity, see section 6.3.5 for a detailed description of the role of this residue in ligand binding) is provided in Figure 6.7, showing the residues that interact with both the FMN cofactor and NIC ligand. Other active site residues that line the binding pocket, but make no strong interactions with either cofactor or ligand, include Q142, Y144 and W138 which cap the dimethylbenzoid moiety of the FMN. Tyrosine 68 and F70 from helix D are on the edge of the binding pocket, with F70 bent away from the nicotinic acid ligand and Y68 at the top of the active site channel, 3.8 Å from F124.

6.2.4 Comparison Between Crystal Forms

6.2.4.1 Crystal Packing Arrangements

A view of the packing arrangements down each of the crystallographic axes of all three crystal forms is provided in Figure 6.8. Detailed close-up schematics of the contact points between the symmetry related molecules can be seen in Figures 6.9 to 6.11. Each contact point may incorporate several residues and a full description of these contacts is given in Table 6.2.

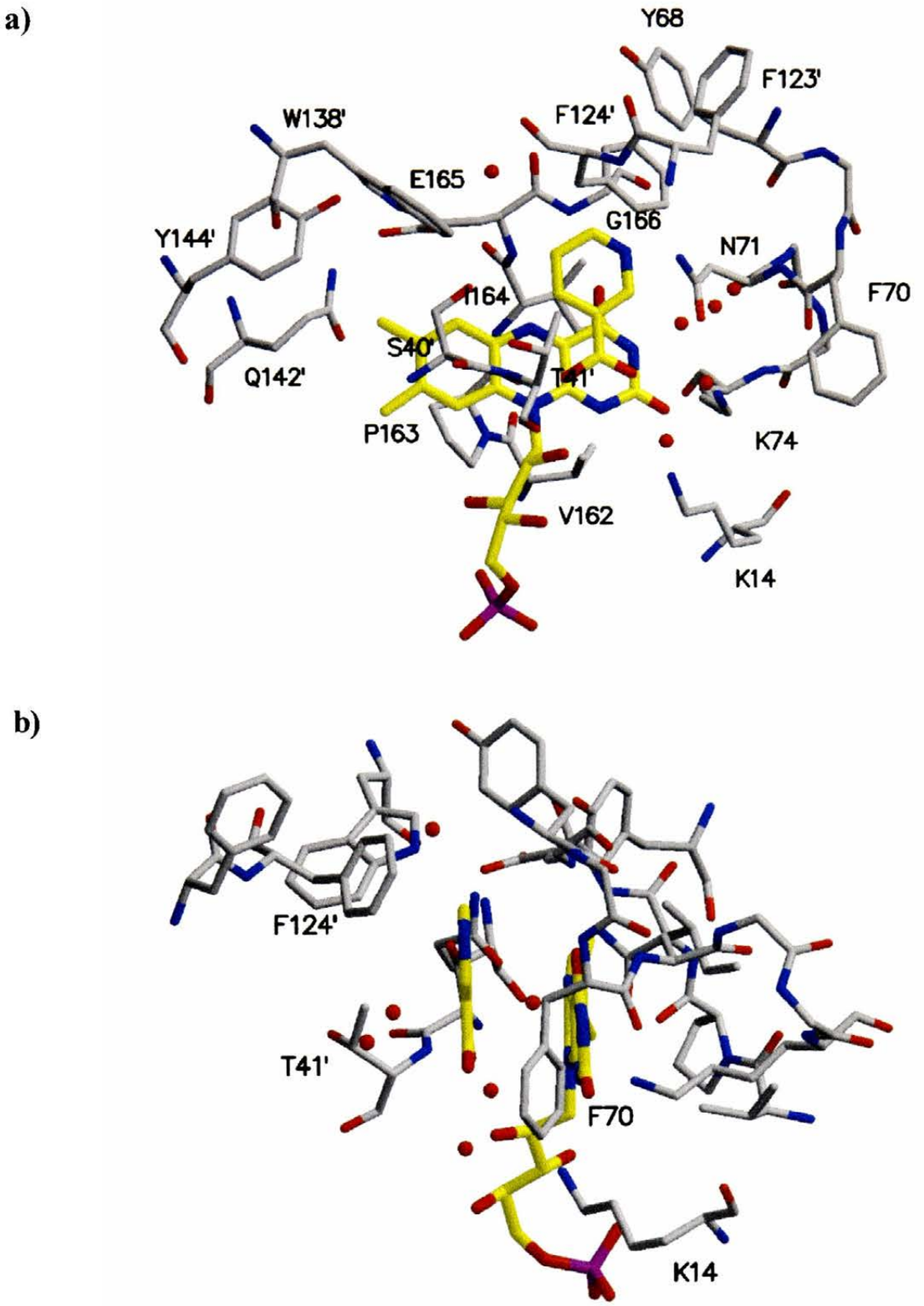


Figure 6.7. Active site residues, bound water molecules, FMN cofactor & nicotinic acid (NIC) ligand. Diagram coloured by atom type (FMN & NIC C, yellow; Protein C, grey; O, red; N, blue; P, mauve). Residues from other monomer of the dimer indicated by a prime. Parts (a) and (b) are related by approximately 90°. All residues are labeled in (a), but only a selection are labeled in (b) to improve clarity. In this diagram the side-chains of residues 69, 72 and 73 have been omitted as they point away from the binding pocket of the enzyme. Figure prepared using MOLSCRIPT (Kraulis 1991) and RASTER 3D (Meritt and Bacon 1997).

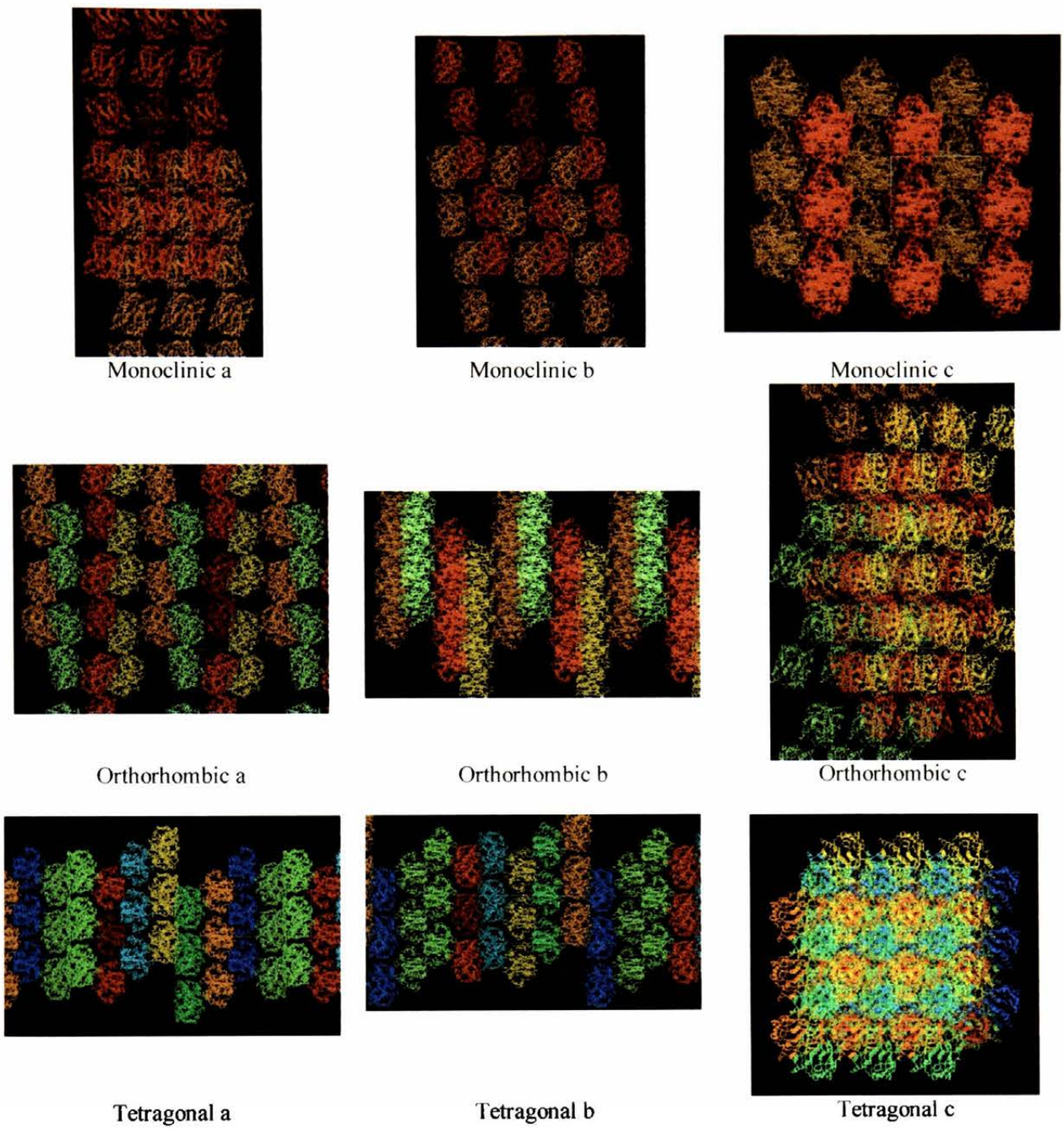


Figure 6.8. Crystal packing between symmetry neighbours. Diagram shows all three NTR spacegroups, and the close crystal packing between symmetry equivalents. Colours represent particular sections of the crystal, and are equivalent within a set of diagrams for a particular spacegroup. Lettering a, b and c refer to crystal unit cell axes. Diagram prepared using Turbo (Roussel *et al.* 1991).

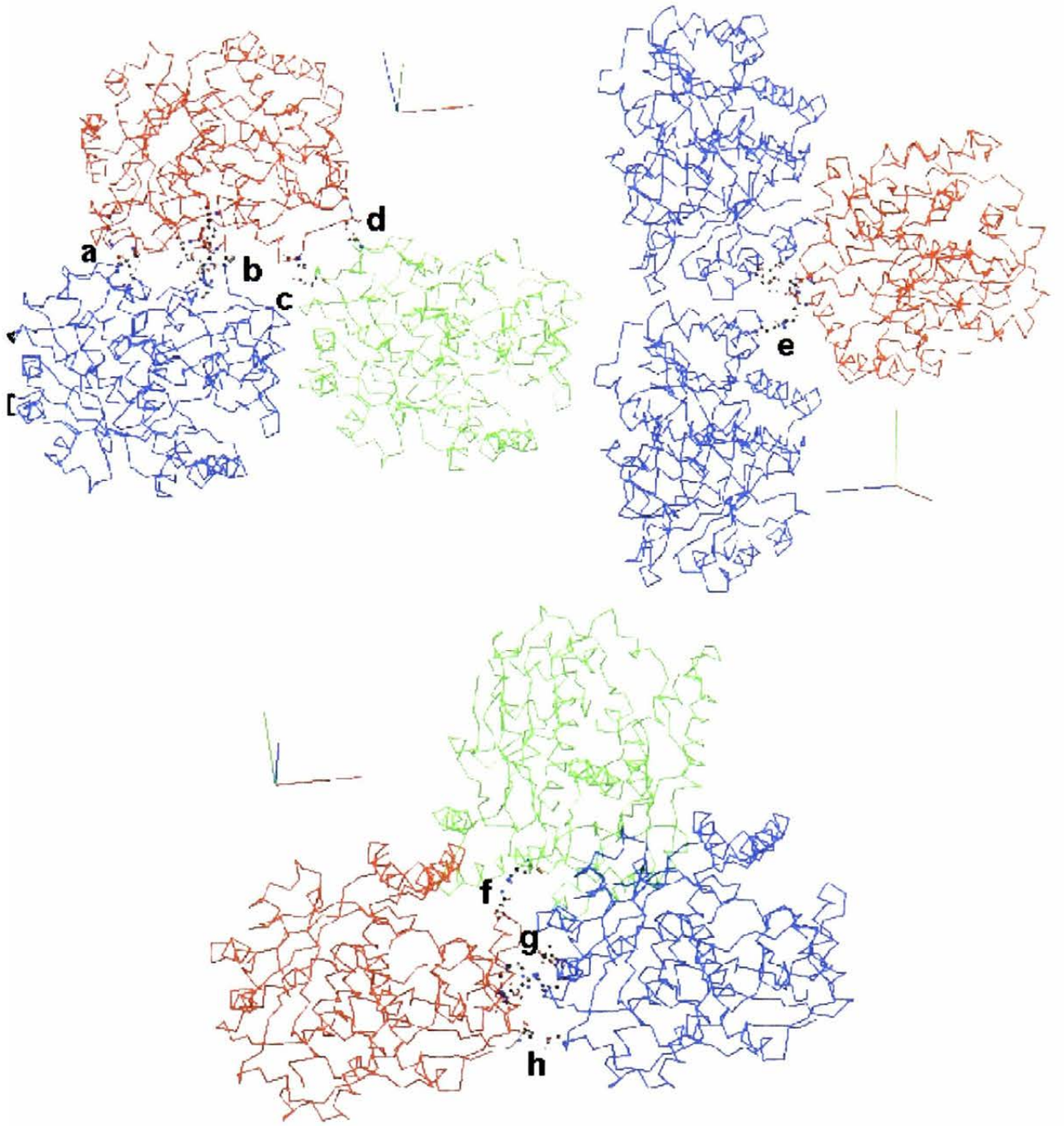


Figure 6.9. Crystal contacts of tetragonal crystal form. NTR is shown as a Cα model, with crystal contact residues represented in ball-and-stick fashion and coloured by atom type (C=grey, N=blue, O=red). The reference molecule (X,Y,Z) is coloured red in all diagrams and symmetry related molecules are coloured green or blue. Crystal axes are given (x=red, y=green, z=blue). The different contact areas are labeled a to h, and are described in more detail in Table 6.2. Contacts labeled e in the diagram come from 2 molecules related by a one unit cell translation along the y axis. Diagram produced using MOLSCRIPT (Kraulis 1991) and RASTER 3D (Meritt and Bacon 1997).

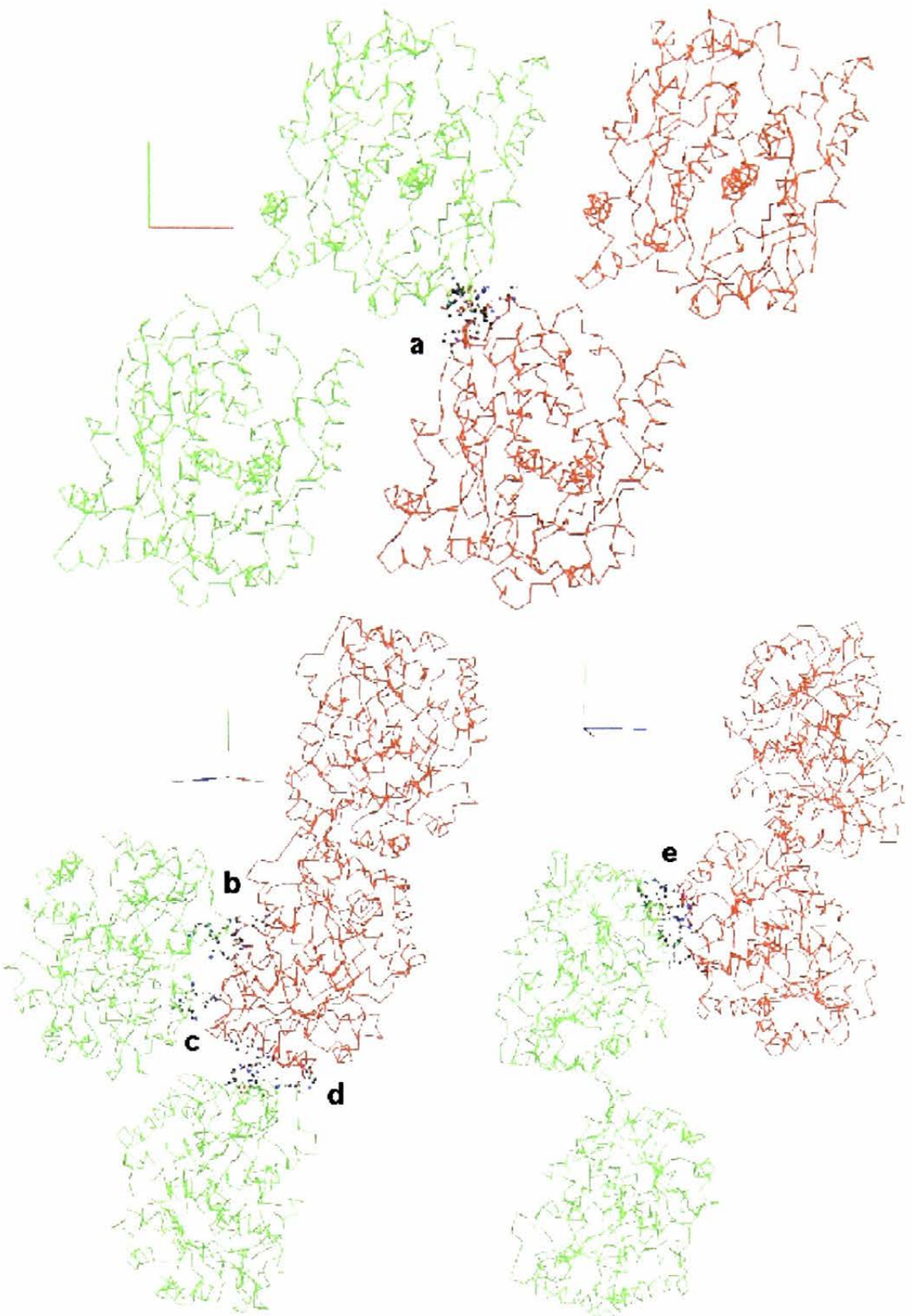


Figure 6.10. Crystal contacts of orthorhombic crystal form. NTR is shown as a Cα model, with crystal contact residues represented in ball-and-stick fashion and coloured by atom type (C=grey, N=blue, O=red). The reference molecule (X,Y,Z) is coloured red in all diagrams and symmetry related molecules are coloured green. Crystal axes are given (x=red, y=green, z=blue). The different contact areas are labeled a to e, and are described in more detail in Table 6.2. Diagram produced using MOLSCRIPT (Kraulis 1991) and RASTER 3D (Meritt and Bacon 1997).

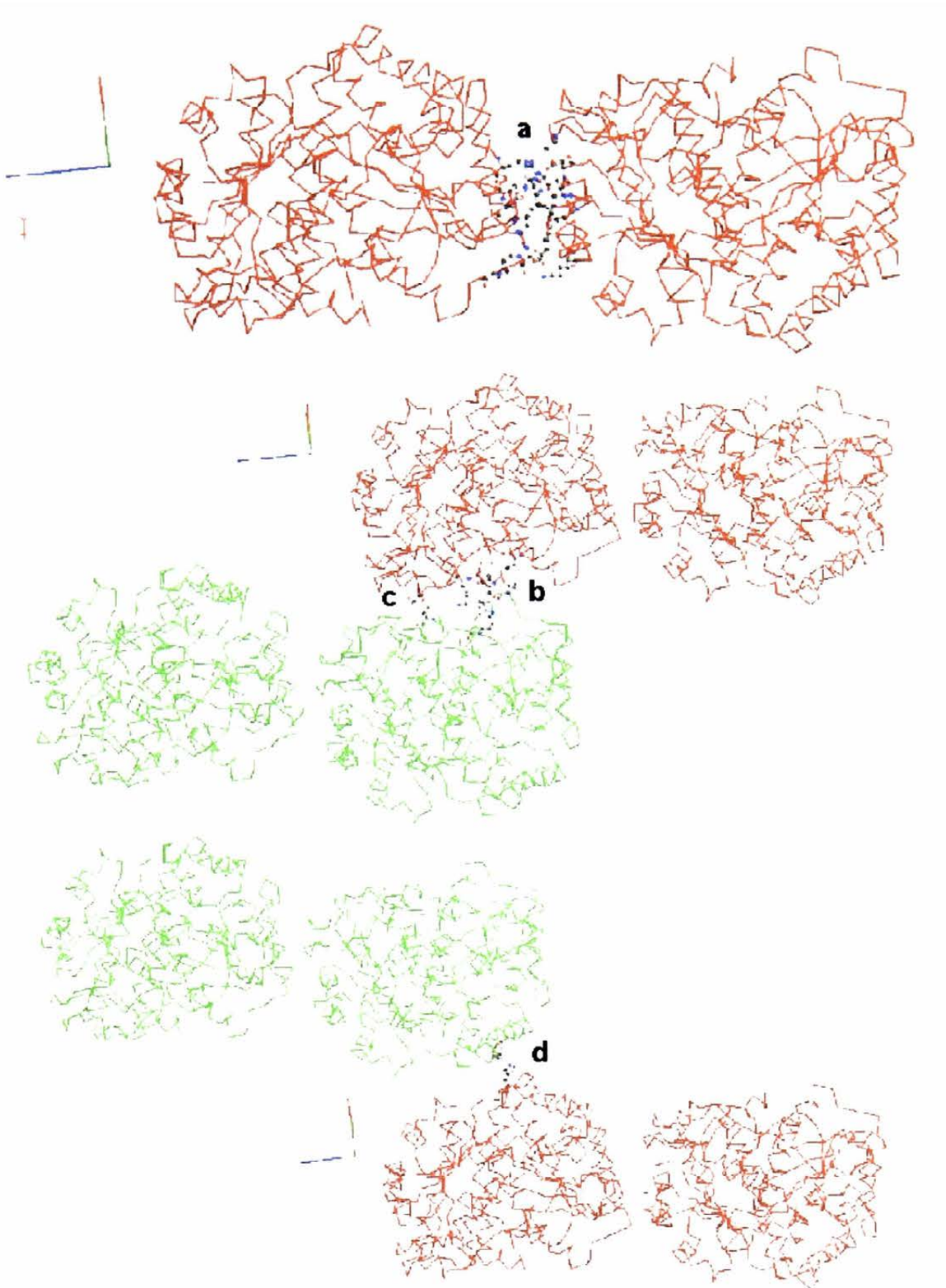


Figure 6.11. Crystal contacts of monoclinic crystal form. NTR is shown as a Cα model, with crystal contact residues represented in ball-and-stick fashion and coloured by atom type (C=grey, N=blue, O=red). The reference molecule (X,Y,Z) is coloured red in all diagrams and symmetry related molecules coloured green. Dimer : dimer contacts within the asymmetric unit are shown at the top of the diagram. Crystal axes are given (x red, y=green, z=blue). The different contact areas are labeled a to d, and are described in more detail in Table 6.2. Diagram produced using MOLSCRIPT (Kraulis 1991) and RASTER 3D (Meritt and Bacon 1997).

Tetragonal Form
Contact (a) A20-A104, A202-B5
Contact (b) B36-A25, A5-A28, A5-A31, A8-A24, A8-A25, A9-A28
Contact (c) B24-B197
Contact (d) B19-A109
Contact (e) B130-A62, B130-B67, B131-A62, B133-A175
Contact (f) B67-B130
Contact (g) B72-A73, B72-A76, B73-A77, B76-A73
Contact (h) B54-A197
Orthorhombic Form
Contact (a) B19-C55, B19-C59, B21-C55, B21-C175, B23-C54, B24-C58, B54-C174
Contact (b) A24-A8, A24-B32, A28-A5, A28-A9, A31-A5
Contact (c) C5-D202, C9-D202, C2-D204
Contact (d) C18-A54, C19-A62, C73-A18, C73-A54, C77-A54, C196-A73, C197-A72 C197-A77
Contact (e) B91-C133, B170-D171, B174-D67, B174-D168, B174-D171, B179-D132 B179-D170, B180-C127, B180-D68, B180-D131, B181-C131, B182-C131
Monoclinic Form
Contact (a) B18-C54, B19-C62, B21-C55, B73-C18, B73-C77, B77-C54, B196-C72 B196-C73, B197-C73, B197-C77
Contact (b) A24-A8, A24-B32, A28-A5, A28-A9, A31-A5
Contact (c) B5-A202
Contact (d) B122-A133, B126-A133

Table 6.2. Residues in crystal contacts (see Figures 6.9 to 6.11). Contacts were defined as residues within 3.5 Å of each other, and calculated using the CCP4 program CONTACT (CCP4 1994). Residues from the reference molecule (X,Y,Z) are given first (identified by chain), followed by those from the symmetry related molecule. Upper-case letters refer to the protein chain of the asymmetric unit. Contact (a) for the monoclinic crystal form lists dimer : dimer contacts in the asymmetric unit, no direct contacts are observed between dimers in the orthorhombic form.

6.2.4.2 Conformation Differences

All three forms give near identical Cα traces and show a small rmsd based on Cα positions (Table 6.3). The main point of backbone conformation variation in the structures occurs loop C-D (residues 64-72), which is solvent exposed (and so more prone to flexibility) and contains residues that form crystal contacts in all 3 of the different spacegroups. The other major deviation in backbone conformation occurs in the tetragonal structure, at loop F-G (residues B126-135), which is again involved in crystal contacts, which are unique to the tetragonal form.

Protein/Chain ID	RMSD[a]	RMSD	Seq ID[b]	PDB Code	Ref
***E. coli* NIC-NTR**					
Tetragonal-B	0.39	0.82		1ICR	Lovering *et al.* 2001
Orthorhombic-A	0.31	0.71		1ICV	
B	0.24	0.70			
C	0.37	0.78			
D	0.38	0.87			
Monoclinic-A	0.27	0.70		1ICU	
B	0.31	0.80			
C	0.34	0.71			
D	0.25	0.72			
***E. coli* NTR**					
A	0.30	0.93	100 %,217,0,0	1DS7	Parkinson *et al.* 2000
B	0.42	0.97			
***E. cloacae* NTR**					
A	0.36		88 %,216,0,0	1NEC	Haynes *et al.* 2002
B	0.28				
C	0.33				
D	0.28				
***V. fischeri* Frase**					
A	0.99		32 %,217,1,2	1VFR	Koike *et al.* 1998
B	1.05				
***T. thermophilus* NOX**					
	1.52		21 %,216,0,19	1NOX	Hecht *et al.* 1995
***E. coli* NfsA**					
	1.59		15 %,218,2,41	1F5V	Kobori *et al.* 2001

[a] RMS deviation in Å between pairs of equivalent atoms with respect to the *E. coli* NIC-NTR tetragonal chain A. First column based on Cα only, second column based on all atoms.[b] Percentage sequence identity, Length of alignment, No. of Insertions, No. of Deletions, calculated from superimposed 3-D structures using Swiss PDB Viewer (Guex *et al.* 1997).

Table 6.3. Comparison of nitroreductase homologues.

The side-chain conformations between the three forms show the greatest variation in areas of solvent exposed residues, as would be expected, and conformation difference around the active site is effectively limited to residue F70, which is discussed in detail in section 6.3.5. Some smaller conformation differences can be seen in R107 and K207 in the orthorhombic structure. These residues contact the FMN phosphate group, and the differences may be attributed to a lower comparative resolution, and lower accuracy in fitting residues to the electron density.

6.3 Discussion

6.3.1 Active Site Residues & Implications for Catalysis

NTR has previously been proposed to adopt a "ping-pong" mechanism (see chapter 1), where the flavin is reduced by NAD(P)H, the nucleotide leaves and then allows the substrate to bind and become reduced, oxidizing the FMN cofactor in the process. A full discussion of flavin oxidation and reduction is provided in Chapter 1. The structures presented here support this hypothesis and provide little allowance of the ternary complex alternative, there being scant extra space in the active site to bind nucleotide and substrate at the same time.

We can assume that the nicotinic acid ligand binding is indicative of how NAD(P)H would bind the protein (due to the structural similarity of nicotinic acid to the nicotinamide headgroup), and that the redox reactions must occur at the *re* face of the FMN, as the *si* face is occluded by contacts to the protein. With the NIC ligand representing the nicotinamide ring of the nucleotides, the remainder of the NAD(P)H would point out of the active site, toward a solvent pocket of the crystal, and give some explanation as to why NTR has no direct preference for NADH or NADPH (with no protein contacts to the nicotinamide ribose). With

the NIC ligand positioned in this orientation, the C4 is 3.4 Å away from the flavin N5. This C4 atom of NIC corresponds to the C4 atom of the nucleotide cofactors that is responsible for hydride transfer in redox reactions. Atoms N5 and N1 of the flavin are known to be the two positions on the isoalloxazine ring system that accept reducing equivalents during reactions of FMN with reductants. The geometry of flavin:nucleotide reactions was studied in depth by Fraaije and Mattevi (2000), who looked at the distances and angles involved in a variety of flavoproteins. Their average value of 3.5 Å between the position of attack (site of electron transfer) and the flavin N5 compares favourably to the 3.4 Å found using our structures (our equivalent being atom C4 of NIC), and the angle between N10, N5 and the site of attack ranged between 96 to 117°, commensurate with our value of 101.6°. These measurements confirm the validity of NIC as a nucleotide substrate analogue in our structures, and the possibility that the reactions of NTR with substrates proceed via hydride transfer.

Fraaije & Mattevi found other aspects of flavoprotein:ligand interactions to be conserved, one of which was the stereochemistry of the hydrogen bond donor to N5 (donor-N5-N10 angle was between 116-170°). The hydrogen bond donor in NTR is the backbone nitrogen atom of Glu165, with a Glu165N-N5-N10 angle of 153.6°. This precise geometry arrangement is thought to increase the oxidative power of the FMN cofactor, the hydrogen bond stabilising a protonated FMN N5 (Fraaije and Mattevi 2000). A diagram of the protein:FMN:NIC geometry is provided in Figure 6.12.

In the reductive half of the ping-pong reaction, a proton must be transferred to the flavin, along with a hydride ion from the nucleotide. The proton may come from any of the residues contacting the isoalloxazine ring as mentioned in section 6.2.3.1 (e.g. both the protonatable

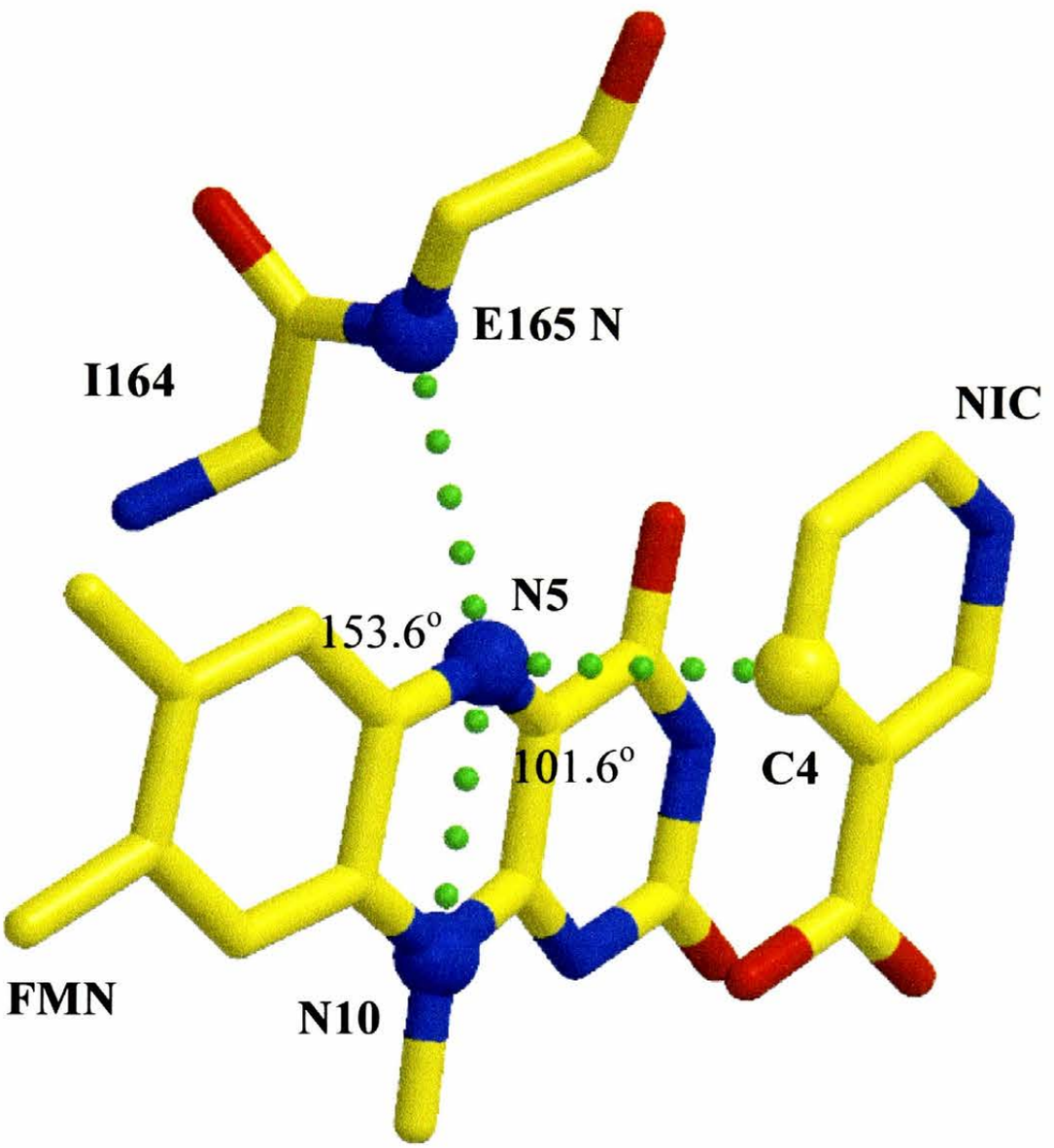


Figure 6.12. Geometry of the protein : flavin : nicotinamide interactions. Atoms coloured according to atom type : C, yellow; N, blue; O, red. Straight lines (in dotted form, green) have been drawn between the centres of atoms E165 N, FMN N5, FMN N10 and NIC C4 (all shown with a larger atom radius above for clarity). The NIC C4 is 3.4 Å from the FMN N5, placed ideally for hydride transfer. Figure was prepared using MOLSCRIPT (Kraulis 1991) and RASTER 3D (Meritt and Bacon 1997

lysine residues near N1 – K14 & K74), or a bound water molecule. Upon reduction of the flavin, tautomerisation can occur and form a negative charge on atom N1, which could be stabilized by K14 which lies 3.4 Å away. Stabilisation of this negative charge, and the environment around the flavin N1 position has long been proposed to affect the flavin redox potential, favouring cofactor reduction (Ghisla and Massey 1989; Massey 1995).

The oxidative half of the ping-pong reaction concerns reducing the substrate and reoxidizing the flavin, presumably obeying a substrate binding orientation similar to that seen in the first half of the reaction. This raises the problem that if substrates bind with the aromatic ring in exactly the same position as NIC, the atoms that would accept the hydride (e.g. NO_2 groups, quinone carbonyls) would be further away from N5, but if substrates bind with the acceptor atom in the equivalent C4 position, the ring stacking interactions would differ. It is possible to model CB1954 with its ring over the NIC ring, but the N of the NO_2 group forms a flatter angle with the N5-N10 axis of ~90°. If the ring of any aromatic substrates bound like that of NIC, then the broad substrate specificity of NTR is partly explained. The protein shows few contacts to NIC (mainly hydrophobic stacking between NIC, FMN and F124), and so a range of substrates could bind, and some like CB1954 could presumably bind in multiple orientations. This would also explain the occurrence of two products upon reduction of CB1954.

6.3.2 Comparison to Unliganded Structures

Shortly after solving the three nicotinic acid liganded forms presented in this work, co-ordinates for 2 unliganded nitroreductases, one from *E .coli* (Parkinson *et al.* 2000) and one from *E .cloacae* (Haynes *et al.* 2002) were deposited in the RCSB Protein Data Bank

(Berman *et al.* 2000). This allows a direct comparison between liganded and unliganded NTR. and the use of the *E. cloacae* structure (and also that of the *V. fischeri* FRase homologue) to reaffirm assumptions made regarding the unliganded state.

In general, changes in the structure attributed to ligand binding are concentrated around the FMN cofactor and helices E and F. In the unliganded structures, helices E and F have high B values and are not well-ordered, the B values for this domain being two to three times that of the average for the protein. Binding of the NIC substrate analogue has increased the order of these domains in our structures, with the B values for this domain being close to those for the protein as a whole. The reason for these changes may lie in helix F contacting the NIC ligand (mainly through F124), and changes in mobility may be important in regulating ligand binding and catalysis.

Some amino acid side-chain conformation differences can be seen between the liganded and unliganded structures. In the active site, F124, which sandwiches NIC against the FMN, rotates away from the FMN in the liganded structures to allow NTR to bind the ligand (Figure 6.13). A smaller effect is seen on the neighbouring residue F123. Another residue F70 shows more variability, but moves away from the active site when NIC binds (Figure 6.13). The F124 acts to clamp ligands to the cofactor (explaining the enzyme preference for aromatic substrates), whereas F70 may be more involved in access to the active site or contacts to larger ligands.

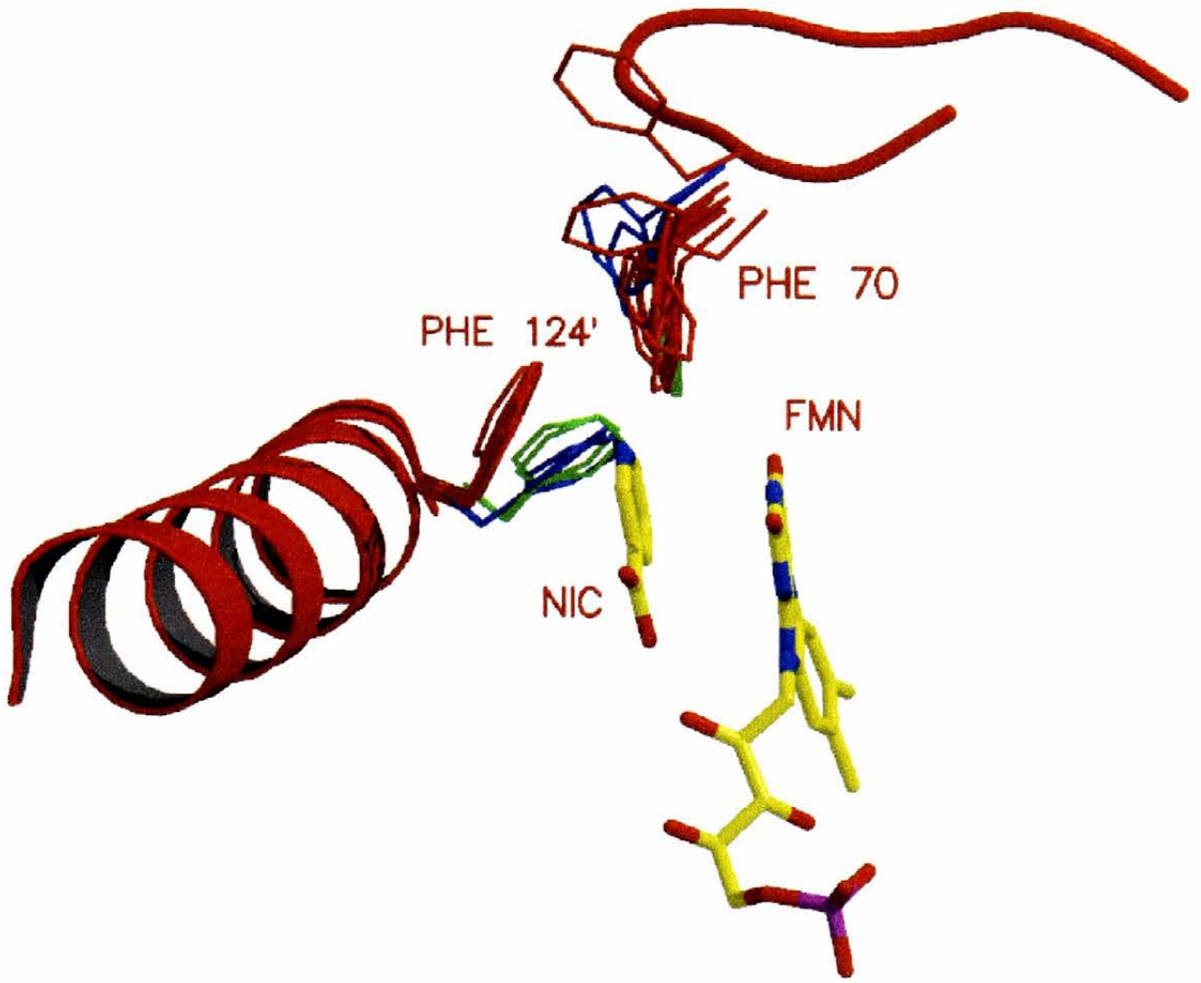


Figure 6.13. Amino acid side-chain shifts upon binding of NIC ligand. The ten NTR NIC monomers are shown in red, the two unliganded *E. coli* monomers from 1ds7.pdb (Parkinson *et al.* 2000) are shown in blue, and the two *V. fischeri* unliganded monomers from 1vfr.pdb (Koike *et a*l. 1998) are shown in green. FMN and NIC coloured by atom type (C, yellow; N, blue; O, red; P, magenta). Notably, F124 can be seen to shift upon the binding of NIC, rotating away from the active site. The bound NIC is sandwiched between the F124 side-chain and the FMN isoalloxazine ring. The F70 side-chain is generally mobile and seen here in many orientations, as opposed to the two distinct populations observed with F124. Diagram prepared using MOLSCRIPT (Kraulis 1991) and RASTER 3D (Meritt and Bacon 1997).

6.3.3 Comparison to Nitroreductase Homologues

The fold of NTR is very similar to that of the nitroreductase from *E. cloacae* and the flavin reductase FRase from *V. fischeri*, the rmsd based on Cα positions being 0.28-0.36 Å and 0.99-1.05 Å, respectively (see Table 6.3). The *E. cloacae* enzyme sequence is 88 % identical to NTR and shows no significant deviations in the fold, and the only residue difference at the active site is the substitution of F123 by Y122, which remains in a similar conformation. Comparison of NTR with FRase shows more variation, and although the secondary structure elements superimpose well, several loop differences can be seen. Loop E-F in FRase (same secondary structure notation as NTR) has one amino acid less than the corresponding loop in NTR and adopts a different conformation. A large difference can be seen in loop F-G, a smaller difference in loop H-4, and the N-terminus has an extension, with a short loop leading into helix A. NTR shows a lower sequence homology to NADH oxidase of *T. thermophillus*, but is structurally similar (21 % identity, rmsd based on Cα of 1.52 Å). Remarkably, NTR and nfsA (see Chapter 1, PDB code 1F5V) share little homology (15 % similarity) but are structurally similar (rmsd based on Cα of 1.59 Å). A major difference exists in the E + F α-helix region, with the nfsA active site being more solvent exposed. The central β-sheet and other secondary structure features are largely super-imposable. The structural similarity between the two proteins raises the prospect of a distant ancestor with a related function.

Shortly after this work on the NIC liganded form of NTR was published (Lovering *et al.* 2001), two liganded and one reduced structures of the *E. cloacae* nitroreductase were published (Haynes *et al.* 2002). The acetate complex is discussed with reference to the NTR acetate structure determination in Chapter 7. The other inhibitor bound structure, a benzoate complex, shows the same amino acid shifts and binding orientation as the NIC-NTR

structures. This is expected as benzoate and nicotinic acid differ only by a single atom, a C to N substitution. The reduced FMN complex of the *E. cloacae* nitroreductase is more interesting, and shows an increased FMN butterfly angle of 25°. No significant amino acid shifts were observed upon FMN reduction (the largest being a 0.15 Å movement of P163). It is therefore tempting to attribute the different behaviour of reduced and oxidised FMN in binding substrates (see Chapter 7) to purely electronic effects, or to small steric effects of the newly introduced hydrogen atoms. The butterfly angle seen in the oxidised and reduced nitroreductase structures may favour reduction of the flavin, and explain why no semiquinone intermediate is formed during catalysis (Zenno *et al.* 1996a). The preference of NTR for this bent, "butterfly" FMN would not favour the formation of a planar semiquinone species.

There are several amino acid differences in the active site when comparing FRase and NTR. Loop 3-H at the *si* face of the flavin has 2 substitutions, V162 for T162 and P163 for T163, possibly playing some role in subtly altering the polarity of the FMN environment. Lysine 74, which hydrogen bonds to atom FMN O2 in NTR, is replaced by an ordered water molecule bound to the shorter residue H76 in FRase. At the *re* face of the FMN, T41 is replaced by I43. F123 and Y68 at the extremity of the substrate binding pocket are replaced by K123 and H70, respectively. Mutation of F123 to K may allow the binding of the larger FMN substrate to the active site. Mutations of NTR that are known to increase FRase activity are discussed in (Zenno *et al.* 1996b). The active sites of NTR and FRase match well and provide a basis for modelling other nitroreductase homologues with potentially lower sequence identity with a high degree of confidence.

6.3.4 Crystal Packing Discussion

When analysing the crystal packing interactions present in the three different spacegroups, it can be seen that the tetragonal form is the most distinct, with the monoclinic and orthorhombic forms sharing a few similar characteristics. The similarity between the monoclinic and orthorhombic forms is also apparent when looking at the unit cell axes (the orthorhombic c axis being approximately a doubling of the monoclinic a axis), and the conversion between the two forms upon a substitution from methionine to seleno-methionine is also an important factor. Looking at the contact diagrams supplied in Figures 6.9 to 6.11, it can be inferred that the (b) + (c) contact regions of the two forms look identical, and that the inter-asymmetric unit monoclinic contacts (a) are approximate to the (d) contact point of the orthorhombic form. Super-imposition of the NTR dimers involved in the (b) + (c) contacts of the two forms shows that the two interactions are not identical - it is possible to overlap one of the dimers with its equivalent from the other spacegroup, but the other dimer involved (across the contact point) will not superimpose as well. This could be attributed to slight non-isomorphism in cell axes between the two forms, but it can be seen that the exact contacts are not completely interchangeable (see Table 6.2), and the differences are more likely to be real. The orthorhombic (d) contact is also subtly different to the monoclinic (a) contact, most clearly seen in contact differences involving residues 72 and 197.

Other contacting regions in the monoclinic and orthorhombic spacegroups appear to be different, with the (a) and (e) contacts unique to the orthorhombic form, and the (d) contact unique to the monoclinic form. The spacegroup "switch" between the two forms has to be attributed to the change between the use of methionine and seleno-methionine. One possible explanation for this switch is the non-equivalence of the monoclinic (d) contact (involving

residue H133) in the orthorhombic form. Here chain C H133 hydrogen bonds to chain B D91 (in contact point (e)), the significance of this difference being that D91 is next to methionine 90 in the primary sequence, and the increase in volume upon substituting sulphur for selenium may change the local environment, favouring one contact type over the other. Also the contact B E180 : D Y68 in the orthorhombic form (again in contact point (e)) is packing near C M127, which may also result in local conformation changes and preferences.

The tetragonal form is different again, and shares few contact similarities with the other two forms, presumably as a result of its markedly different solvent content (43 %, 5 % lower than the monoclinic form, 7 % lower than the orthorhombic form). The tetragonal and the monoclinic or orthorhombic forms growing from the same drop is most likely a result of local differences in protein supersaturation.

CHAPTER 7

LIGAND COMPLEXES & MUTANT ENZYMES

7.1 Introduction

7.1.1 Mutant Nitroreductases

Following the structural determination of the native enzyme in complex with a substrate analogue, it was possible to identify residues that may play a role in the recognition of the CB1954 substrate. As both nicotinate and CB1954 (presumably) bind over the isoalloxazine of the FMN, and both have a six-membered ring, it was assumed that the binding modes of the two compounds would be similar. From observation of those residues closest to nicotinate, several residues were selected as candidates for mutation in an attempt to find a mutant or combination of mutants with increased catalytic activity towards CB1954. Residues deemed essential to FMN cofactor binding (part active site analysis, part sequence alignment-based) were omitted from selection. The final list of candidates was as follows :- S40, T41, Y68, F70, N71, G120, F124, E165 and G166.

7.1.2 Crystal Soaking

Enzymes that utilise a ping-pong mechanism present difficulties with regard to determining a ligand-bound structure – a compound that binds the protein without being turned over or released is required, and this may be complicated by the two different binding steps. Although no major conformational differences occur between the reduced and oxidized enzyme state (Section 7.4.1), the electrostatics and shape of the cofactor change, affecting ligand binding. These factors are less of a problem when dealing with inhibitors as opposed to substrates. Two crystal soaks were attempted, both with oxidized crystals of NTR. The first soak used the inhibitor (and buffer) acetate, which binds in the active site if no nicotinate is present in

crystal growth. The second soak used the antibiotic substrate nitrofurazone, chosen because of the presence of a nitro- group, and also due to the paucity of mechanistic detail on its activation by NTR.

7.2 Materials and Methods

7.2.1 Preparation and Assay of Mutant Nitroreductases

All mutagenesis was the work of Dr. Jane Grove and Dr. Peter Searle (Institute for Cancer Studies, University of Birmingham). Studies on the kinetics of the mutants were the work of Paul Race (School of Biosciences, University of Birmingham).

The full basis of this work is described in detail in Grove *et al.* (2003). A pUC19-based plasmid (pJG12B1) was used as a template for PCR-based site-directed mutagenesis. Mutant libraries for each position were obtained with the use of partially overlapping PCR products that span the 5' and 3' ends of the NTR gene, with NNN (all 4 mixed deoxyneucleotides) at the desired codon. After PCR amplification, the products were digested and cloned into a derivative of the bacteriophage lambda vector λNM1151 (λJG3J1). After packing of the DNA into phage, the phage particles were used to infect the nitroreductase-deficient *E. coli* strain UT5600 (Elish *et al.* 1988). All lysogens were replica plated onto a series of agar plates containing a range of CB1954 concentrations (0-400 μM). Clones more sensitive to CB1954 than the wild-type NTR were amplified by PCR, then sequenced.

7.2.2 Crystallization

For the eventual goal of crystallization, selected mutant phage-infected UT5600 cultures were handled in the same manner as the wild-type enzyme (both in over-expression of NTR and

purification of the enzyme). Crystallization conditions were similar to those reported for the wild-type, with careful screening to compensate for any difference in protein concentration between batches.

7.2.3 Ligand Soaks

All crystal soaks contained 10 % w/v PEG 4000 and 25 % v/v ethylene glycol. The acetate inhibitor soak was buffered by 100 mM sodium acetate, pH 4.6, all other soaks (including "holo" structures, with no ligand bound over the flavin) were buffered using 50 mM sodium citrate, pH 4.6. The nitrofurazone soak contained 15 mM nitrofurazone, solubilised in DMSO (final concentration 15 % v/v). Soaking times for all were approximately 10 minutes.

7.2.4 Data Collection

All crystals were flash cooled to 100 K in a cryostream prior to data collection. Diffraction data were collected at the ESRF using beamline ID14-1 (nitrofurazone substrate soak, acetate inhibitor soak, F124K mutant, N71 mutant – holo and NIC-bound forms) and beamline ID14-4 (T41L mutant, F124N mutant, N71S/F124K double mutant – holo and NIC-bound forms). All data were indexed, integrated and scaled using MOSFLM (Leslie 1992) and SCALA (Evans 1993).

7.2.5 Refinement Protocol

Initial refinement steps used the CNS suite of programs (Brunger *et al.* 1998), as described in Chapter 5. For both forms of the double mutant, N71S NIC-bound form and T41L structure determinations, data were of high resolution and TLS refinement was included in REFMAC runs (Winn *et al.* 2001).

7.3 Results

7.3.1 Mutant Screening Results

(Work by Jane Grove, Institute for Cancer Studies, University of Birmingham). The proposed amino acid positions for mutation were fruitful – mutations that conferred an increased sensitivity to CB1954 were observed at positions S40, T41, Y68, F70, N71, G120 and F124; none were obtained at positions E165 and G166. A histogram showing the various mutations and their relative activities in comparison to the wild-type enzyme is provided in Figure 7.1. From this preliminary characterisation of the activities of the mutant enzymes, structural studies were instigated on the N71S, F124K, F124N and T41L single mutants, along with a N71S F124K double mutant. Kinetic data on these particular enzymes is provided in Table 7.1.

7.3.2 Crystallization

The N71S, F124K and N71S/F124K (double) mutants gave large single crystals characteristic of the tetragonal NTR crystal form. The T41L enzyme yielded very poor crystal clusters which showed no response to screening experiments. Small single crystals were cut from the mass of monoclinic form crystals. The F124N mutant also yielded a crystalline mass rather than single crystals, and an orthorhombic NTR crystal form was finally used. For the ligand soaks, wild-type NTR crystals were grown as documented in Chapter 4.

7.3.3 Data Collection

Data were collected for both NIC-bound and holo forms of the N71S mutant and N71S/F124 double mutant. Data for a NIC-bound form were collected for the F124K, F124N and T41L single mutants. A tetragonal NTR dataset was collected for the acetate soak, and a monoclinic

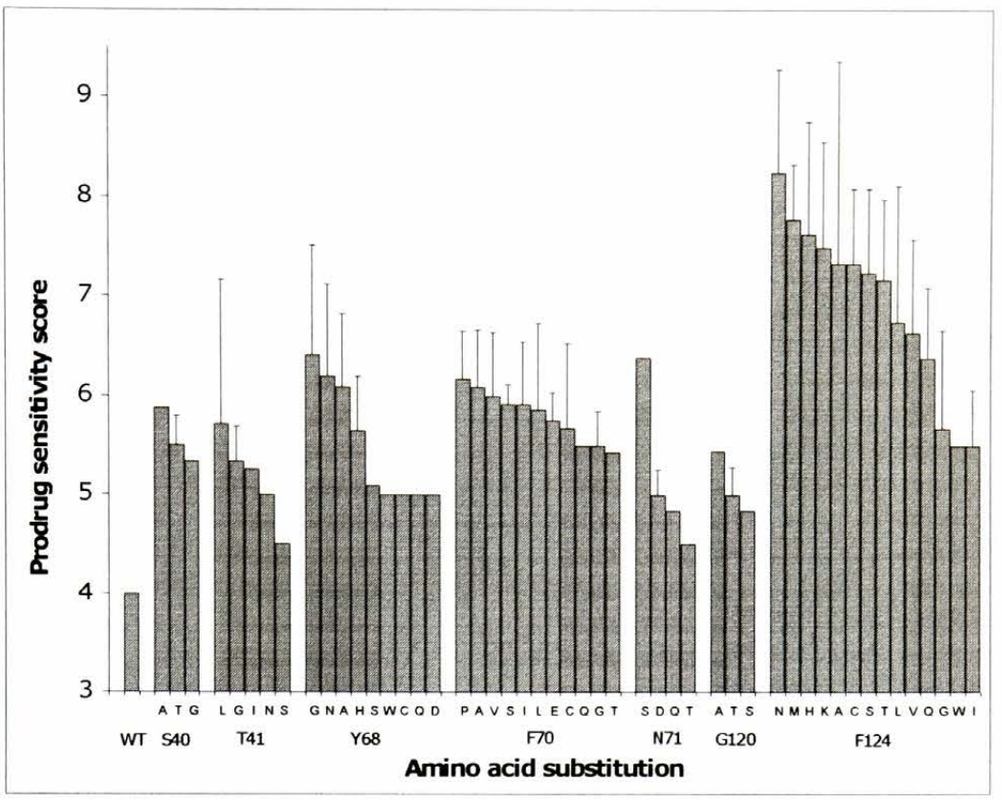


Figure 7.1. Work by Dr. Jane Grove (Institute of Cancer Studies, Birmingham). Relative sensitivity to the prodrug CB1954 by bacterial colonies with mutant nitroreductase. The wild-type enzyme has an arbitrary sensitivity score of 4, mutant enzymes with a score greater than 4 show increased toxicity to bacterial colonies at lower concentrations of prodrug. Assay involves replica plating of bacteria (infected with a phage, vector carries random amino acid at designated position in NTR sequence) onto plates containing 0 – 200 μM CB1954. Colonies that show inhibited growth at lower CB1954 concentrations than required to kill bacteria expressing the wild-type enzyme are picked off and sequenced.

Mutation	K_M for CB1954 (μM)	K_{cat} for CB1954 (min^{-1})	k_{cat}/K_M for CB1954 ($min^{-1}\mu M^{-1}$)
Wild-type enzyme	880 ±56	140 ±11	0.2 ±0.02
N71S single mutant	180 ±25	150 ±14	0.8 ±0.2
F124K single mutant	400 ±13	140 ±4	0.4 ±0.03
N71S / F124K double mutant	180 ±16	170 ±12	0.9 ±0.1
F124N single mutant	280 ±21	180 ±21	0.6 ±0.1
T41L single mutant	1000 ±86	400 ±43	0.4 ± 0.03

Table 7.1. Work by Paul Race (School of Biosciences, University of Birmingham, Birmingham). Kinetic parameters (apparent) for NTR enzymes. All calculations used a fixed NADH concentration of 60 μM.

dataset for the nitrofurazone soak. Full data collection statistics are listed in Table 7.2 (native soaks) and Table 7.3 (mutant enzymes).

7.3.4 Phasing and Model Building

Datasets for all mutant enzymes or native soaks yielded interpretable electron density maps when using the corresponding native crystal form to phase the structure. Final refinement statistics for the structures are listed in Table 7.4 (native soaks) and Table 7.5 (mutant enzymes).

7.3.5 Structural Features

7.3.5.1 F124K Mutations

Electron density for K124 in the F124K single mutant and both the NIC and holo forms of the N71S/F124K double mutant is provided in Figure 7.2. This mutation was the most ambiguous of all the mutations to model, with partial disorder in all of the conformations observed. It cannot be discounted in some of the models (single mutant B chain and double mutant holo form A chain especially) that the position of the NZ atom of the introduced lysine is modelled incorrectly, taking the place of an ordered water molecule.

7.3.5.2 N71S Mutations

Electron density for S71 in both the NIC and holo forms of the single and double mutants is provided in Figure 7.3. Density for the mutated residue in all of the forms was excellent. A bound water molecule between S71 and the FMN cofactor could clearly be modelled in all chains apart from chain B of the single mutant holo form.

Crystal Properties	Nitrofurazone	Acetate
Spacegroup	$P2_1$	$P4_12_12$
Cell Dimensions (Å)	70.08 x 56.47 x 116.09 β=103.1°	57.58 x 57.58 x 263.18
Solvent Content	48 %	43 %
Copies in ASU	4	2
Data Collection		
Completeness (%)	99.6 (99.5)	99.8 (98.8)
Redundancy	6.5	8.0
$I/\sigma(I)$	8.9 (5.8)	3.6 (1.6)
R_{sym} (%)[a]	5.6 (11.8)	10.7 (40.3)
Resolution (Å)	1.7	1.7

[a] $R_{sym} = \Sigma_j|<I>-I_j|/\Sigma<I>$, where I_j is the intensity of the j^{th} reflection and <I> is the average intensity
Values in parentheses refer to data in the highest resolution shell

Table 7.2. Data collection statistics of NTR crystal soaks

Crystal Dataset	N71S H	N71S N	N71S/F124K H	N71S/F124K N	F124K N	F124N N	T41L N
Spacegroup	$P4_12_12$	$P4_12_12$	$P4_12_12$	$P4_12_12$	$P4_12_12$	$P2_12_12_1$[c]	$P2_1$
Cell Dimensions (Å)	57.57 x 57.57 x 264.25	57.40 x 57.40 x 263.42	57.64 x 57.64 x 264.40	57.67 x 57.67 x 266.39	57.29 x 57.29 x 261.27	57.47 x 120.75 x 144.50	71.86 x 57.29 x 116.8 β=103.78
Copies in ASU	2	2	2	2	2	4	4
No. Observations	87743	616386	256917	188621	88746	269731	310358
No. Unique Reflections	18297	36110	40055	40264	15087	86057	80580
Completeness (%)[a]	99.3 (97.9)	99.8 (99.2)	94.1 (82.6)	94.3 (76.2)	84.6 (54.9)	91.8 (68.9)	93.9 (69.9)
Redundancy	4.8	17.1	6.4	4.7	5.9	3.1	3.9
I/σ(I)	8.3 (2.1)	9.5 (2.0)	5.7 (2.4)	5.7 (3.6)	8.4 (7.4)	9.2 (1.7)	9.5 (3.2)
R_{sym} (%)[b]	6.8 (30.2)	6.1 (34.5)	8.9 (25.3)	8.3 (19.8)	5.8 (9.0)	6.5 (46.6)	5.7 (22.1)
Resolution (Å)	2.4 (2.5-2.4)	1.9 (2.0-1.9)	1.8 (1.9-1.8)	1.8 (1.9-1.8)	2.4 (2.5-2.4)	1.8 (1.9-1.8)	1.8 (1.9-1.8)

[a]Values In parentheses are for data in the highest resolution shell.

[b]$R_{sym} = \Sigma_j|<I>-I_j|/\Sigma<I>$, where I_j is the intensity of the j^{th} reflection and <I> is the average intensity

[c]Different orthorhombic cell to the SeMet NTR crystal form mentioned in Chapter 5

Table 7.3. Data collection statistics of NTR mutant crystal forms. **H** = holoenzyme structure, **N** = nicotinate-bound structure.

	Nitrofurazone	Acetate
Refinement		
Resolution (Å)	1.7	1.7
No. of Non-hydrogen Atoms	7831	4035
No. Water Molecules	907	653
R factor[a]	14.4 %	17.3 %
R_{free}[b]	16.6 %	19.5 %
Average B factor	12.33 Å^2	12.77 Å^2
RMSD Bond Angles[c]	1.25°	1.17°
RMSD Bond Lengths[d]	0.008 Å	0.010 Å

[a] R factor = $\Sigma_{hkl}| |F_o| - |F_c| |/\Sigma_{hkl}|F_o|$.

[b]R factor based on 5 % of the data withheld from refinement

[c]RMSD bond angles and [d]bond lengths are the root mean square deviations from ideal values

Table 7.4. Final refinement statistics for all NTR ligand:complex determinations.

Refinement							
Data Set	N71S **H**	N71S **N**	N71S F124K **H**	N71S F124K **N**	F124K **N**	F124N **N**	T41L **N**
Resolution (Å)	2.4	1.9	1.8	1.8	2.4	2.0	1.8
No. Water Molecules	105	318	334	351	194	242	499
R factor[a]	22.0 %	17.3 %	16.5 %	15.8 %	19.3 %	24.9 %	17.3 %
R_{free}[b]	28.1 %	21.0 %	21.4 %	20.3 %	23.8 %	28.4 %	19.8 %
Average B factor ($Å^2$)	24.40	21.11	14.22	12.57	17.08	20.68	16.26
RMSD Bond Angles[c] (°)	2.13	1.45	1.45	1.40	1.85	1.83	1.37
RMSD Bond Lengths[d] (Å)	0.006	0.017	0.015	0.016	0.006	0.007	0.012

[a] R factor = $\Sigma_{hkl}| |F_o| - |F_c| |/\Sigma_{hkl}|F_o|$.

[b]R factor based on 5 % of the data withheld from refinement

[c]RMSD bond angles and [d]bond lengths are the root mean square deviations from ideal values

Table 7.5. Final refinement statistics for all NTR mutant enzyme structure determinations. **H** = holoenzyme structure, **N** = nicotinate-bound structure.

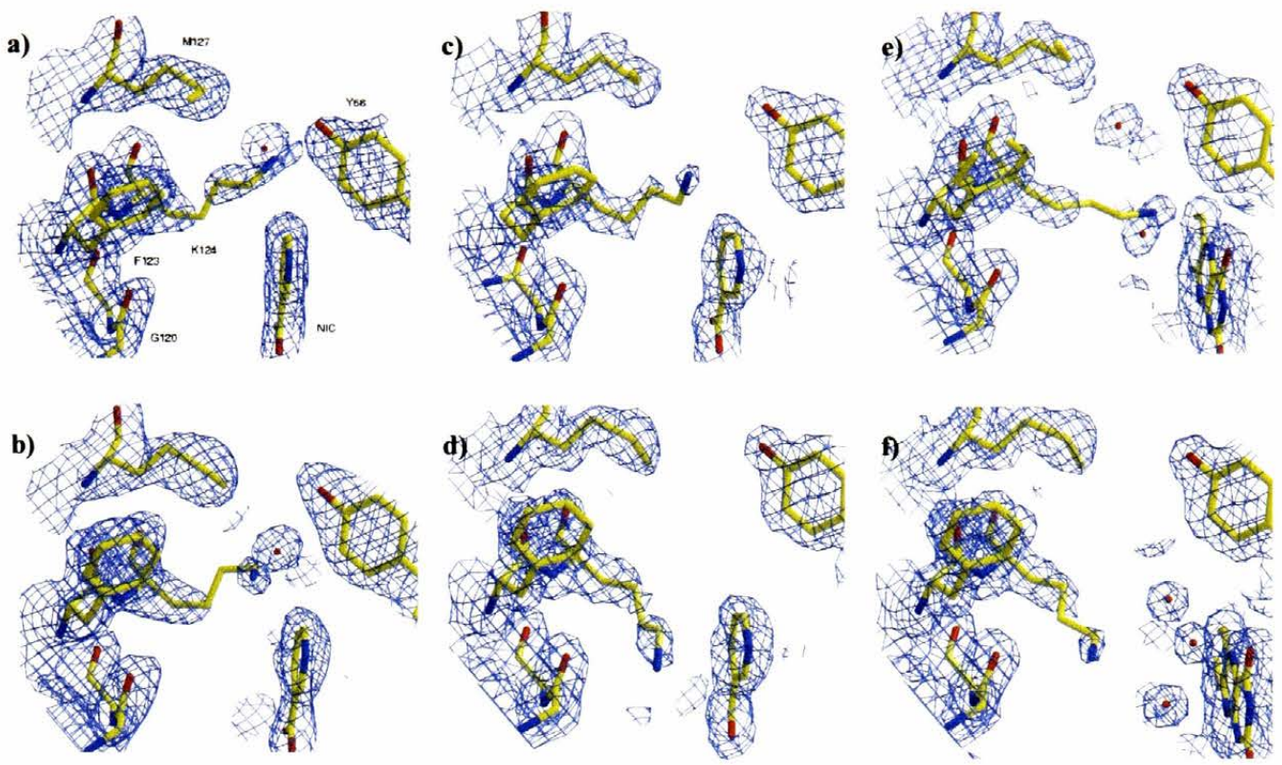


Figure 7.2. Refined electron density for K124 in mutant enzymes. Colouring is by atom type (C, yellow; O, red; N, blue). a) + b) F124K single mutant, c) + d) F124K N71S double mutant NIC form, e) + f) F124K N71S double mutant holo form, all have chain A on the top and chain B on the bottom. Residue labelling in first panel only, omitted from the others. All figures are shown in the same orientation. Maps are 2mFo-Dfc, SIGMAA-weighted, contoured at 1σ. The resolution range of data included in map calculation varies for each structure, see Table 7.3. Figure prepared using BOBSCRIPT (Esnouf 1997) and RASTER 3D (Meritt and Bacon 1997).

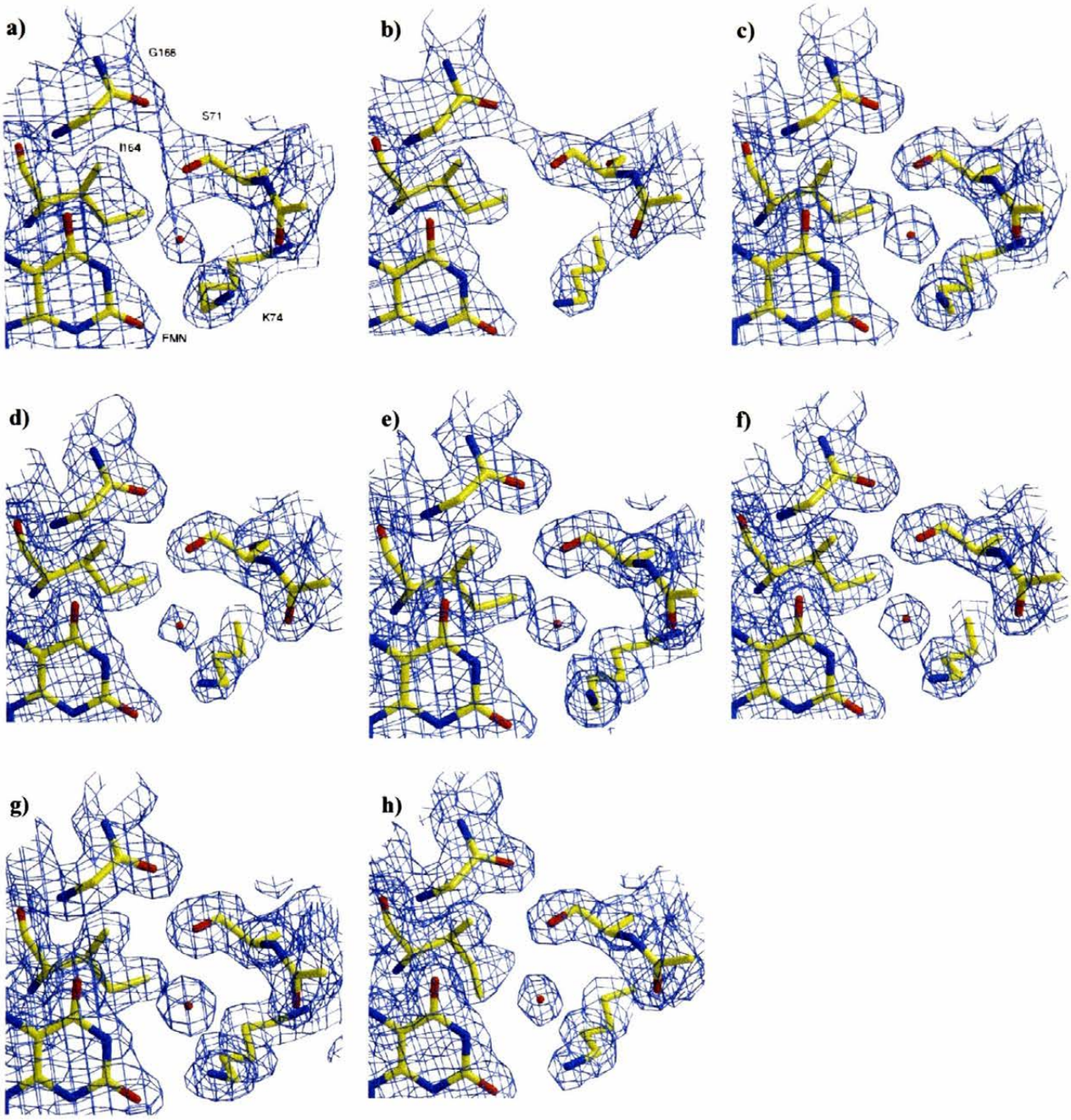


Figure 7.3. Refined electron density for S71 in mutant enzymes. Colouring as Figure 7.2. a) + b) N71S single mutant holo form, c) + d) single mutant NIC form, e) + f) F124K N71S double mutant NIC form, g) + h) double mutant holo form, all show chain A first, then chain B. Residue labelling in first panel only, omitted from the others. All figures are shown in the same orientation. Maps are 2mFo-Dfc, SIGMAA-weighted, contoured at 1σ. The resolution range of data included in map calculation varies for each structure, see Table 7.3. Figure prepared using BOBSCRIPT (Esnouf 1997) and RASTER 3D (Meritt and Bacon 1997).

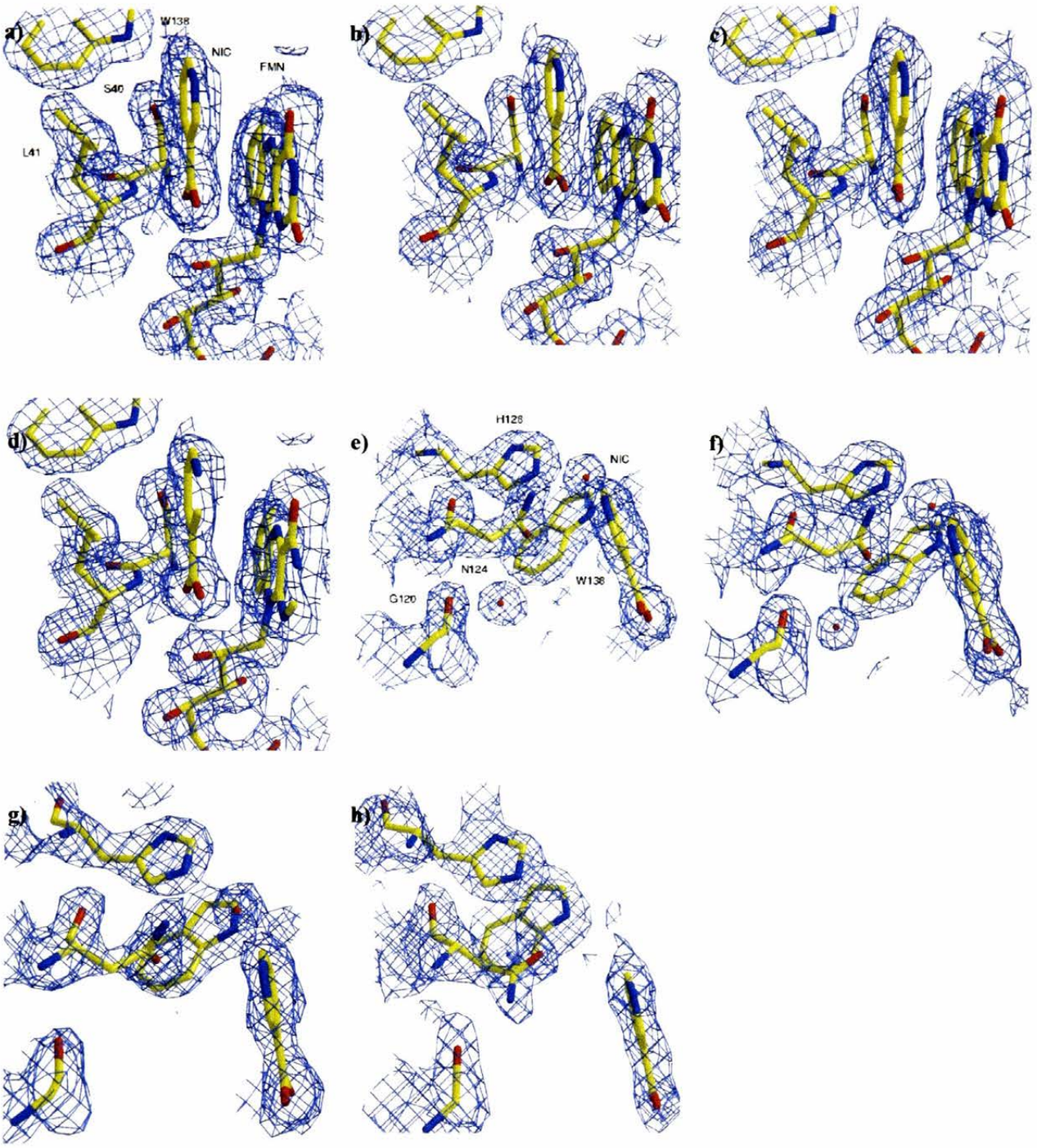


Figure 7.4. Refined electron density for T41L and F124N single mutants. Colouring as Figure 7.2. a) to d) T41L, e) to h) F124N, all chains shown in order A-D. Residue labelling in first panel only, omitted from the others. All figures from same structure are shown in the same orientation. Maps are 2mFo-Dfc, SIGMAA-weighted, contoured at 1σ. The resolution range of data included in map calculation varies for each structure, see Table 7.3. Figure prepared using BOBSCRIPT (Esnouf 1997) and RASTER 3D (Meritt and Bacon 1997).

7.3.5.3 Other Single Mutants

Electron density for the T41L and F124N mutants is provided in Figure 7.4. Density for L41 was excellent, and the residue was modelled in the same conformation in all four chains. Diffraction data for the F124N mutant may be twinned (observation from cumulative intensity analysis, data not shown). However, electron density for N124 was good in all four chains, leading to the inclusion of this data in this investigation.

7.3.5.4 Native Ligand Soaks

Clear electron density for the respective bound molecule was observed in the active site of the nitrofurazone and acetate complexes (Figure 7.5). In both cases only one molecule is bound per active site. Nitrofurazone adopts a planar conformation, with the aliphatic moiety stacking over the two polar rings of the FMN, and the nitrofuran ring projects into a loop formed by residues 68 to 71. This orientation was unexpected as it places the nitrogroup away from the site of hydride transfer. Nitrofurazone is bound with good occupancy, with an average B factor of 19.4 $Å^2$ (*c.f.* the overall average B factor of 12.3 $Å^2$).

Nitrofurazone makes 7 hydrogen bonds to the protein, FMN and ordered waters (Figure 7.5). Atom N4 of nitrofurazone hydrogen bonds to E165 OE1, atom O4 with the T41 backbone nitrogen and the FMN O2*, atom O1 with the F70 backbone nitrogen and a water molecule, and atom O2 with N71 ND2. There are extensive Van der Waals contacts between the nitrofurazone aliphatic region (C5, N2, N3, C6, N4 and O4) and the isoalloxazine ring.

Changes in active site residue conformations exist between the NTR:nitrofurazone complex and the NTR:NIC structure. The F124 sidechain of the nitrofurazone complex more closely

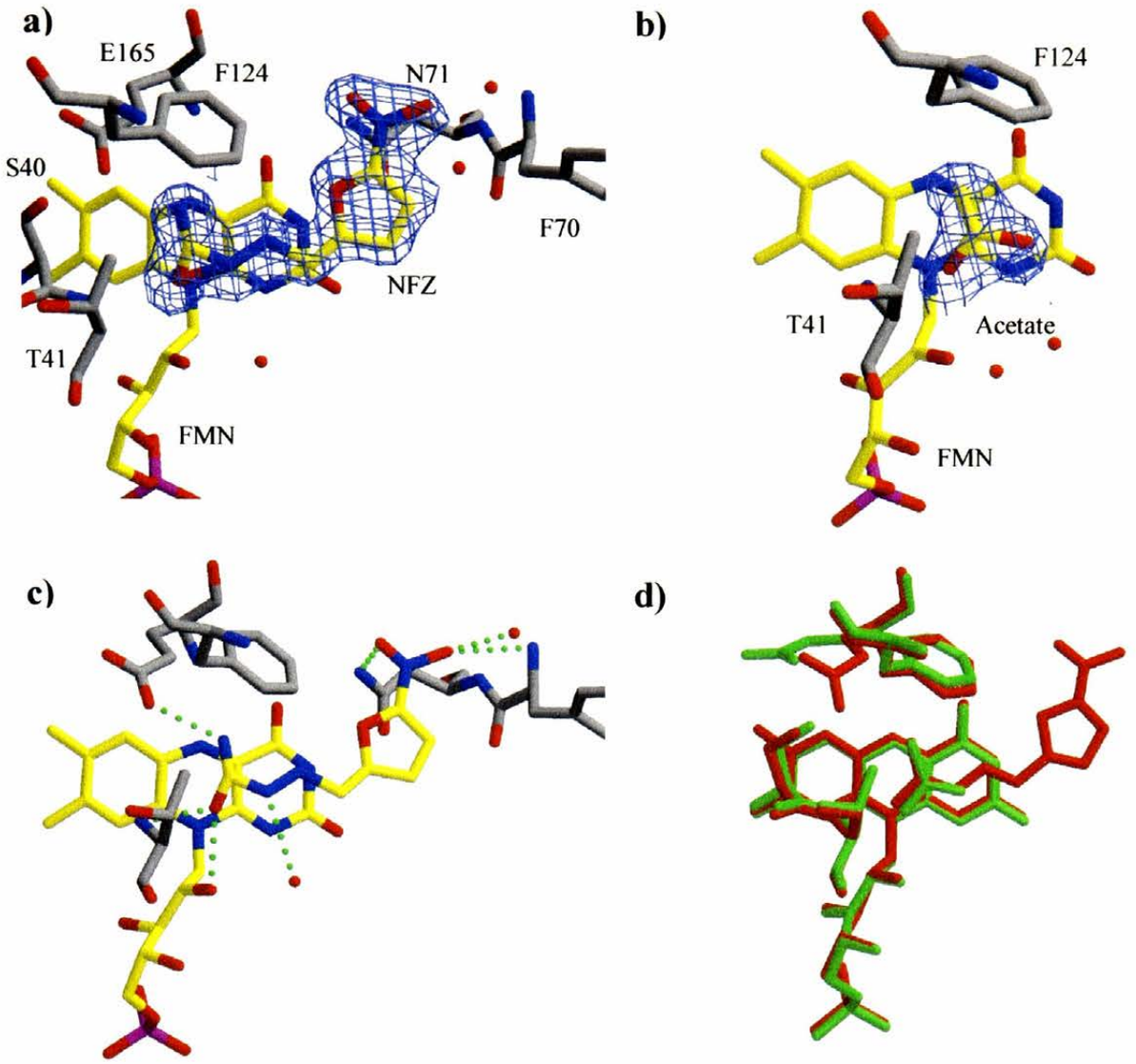


Figure 7.5. Binding of nitrofurazone and acetate to NTR. a) to c) coloured by atom type: protein C, grey; other C, yellow; O, red; N, blue; P, magenta. Refined electron density maps of a) nitrofurazone and b) acetate, 2mFo-DFc, SIGMAA-weighted, contoured at 1σ, resolution 1.7 Å. c) contacts of nitrofurazone to the protein, FMN and water molecules, hydrogen bonds displayed as dotted green lines. d) shift of active site residues when binding the two ligands, nitrofurazone structure in red, acetate structure in green. Figures produced using MOLSCRIPT (Kraulis 1991), BOBSCRIPT (Esnouf 1997) and RASTER 3D (Meritt and Bacon 1997).

resembles the orientation observed in the non-liganded NTR structure (Parkinson *et al.* 2000). F70 has its mobility reduced due to the binding of the nitrofuran ring in the pocket formed by residues 68 to 71. The contact between nitrofurazone and E165 causes the glutamate sidechain to orient itself closer to S40 (2.3 Å), the electron density between the two residues is continuous (Figure 7.6a). A comparison between the nitrofurazone complex and the NIC complex is presented in Figure 7.6b.

Upon elucidation of the NTR:acetate structure, it was apparent that acetate binds NTR in an orientation analogous to the carboxamide moiety of nitrofurazone (Figure 7.5). The bound acetate has good occupancy, with an average B factor of 24.5 $Å^2$ (overall average B factor of 12.8 $Å^2$). Although it would be possible to model the acetate molecule with the methyl carbon in the same position as the nitrofurazone N3 atom (with the acetate oxygens occupying the nitrofurazone N4 and O4 positions), it was decided to model the methyl carbon pointing directly into the ring of F124. This positions one acetate oxygen into a hydrogen bond with the T41 backbone nitrogen, and the other facing bulk solvent.

7.4 Discussion

7.4.1 Ligand Soaks

The role of the antimicrobial agent nitrofurazone in the discovery of NTR is detailed in Chapter 1. Nitrofurazone and the related compound nitrofurantoin remain important in treating burns victims and genitourinary infections (Asnis 1975), proving effective against organisms that have developed resistance to other treatments. Although nitrofurans have been used in such a manner for many years, the mode of their activation into a toxic species by nitroreductases has never been fully investigated.

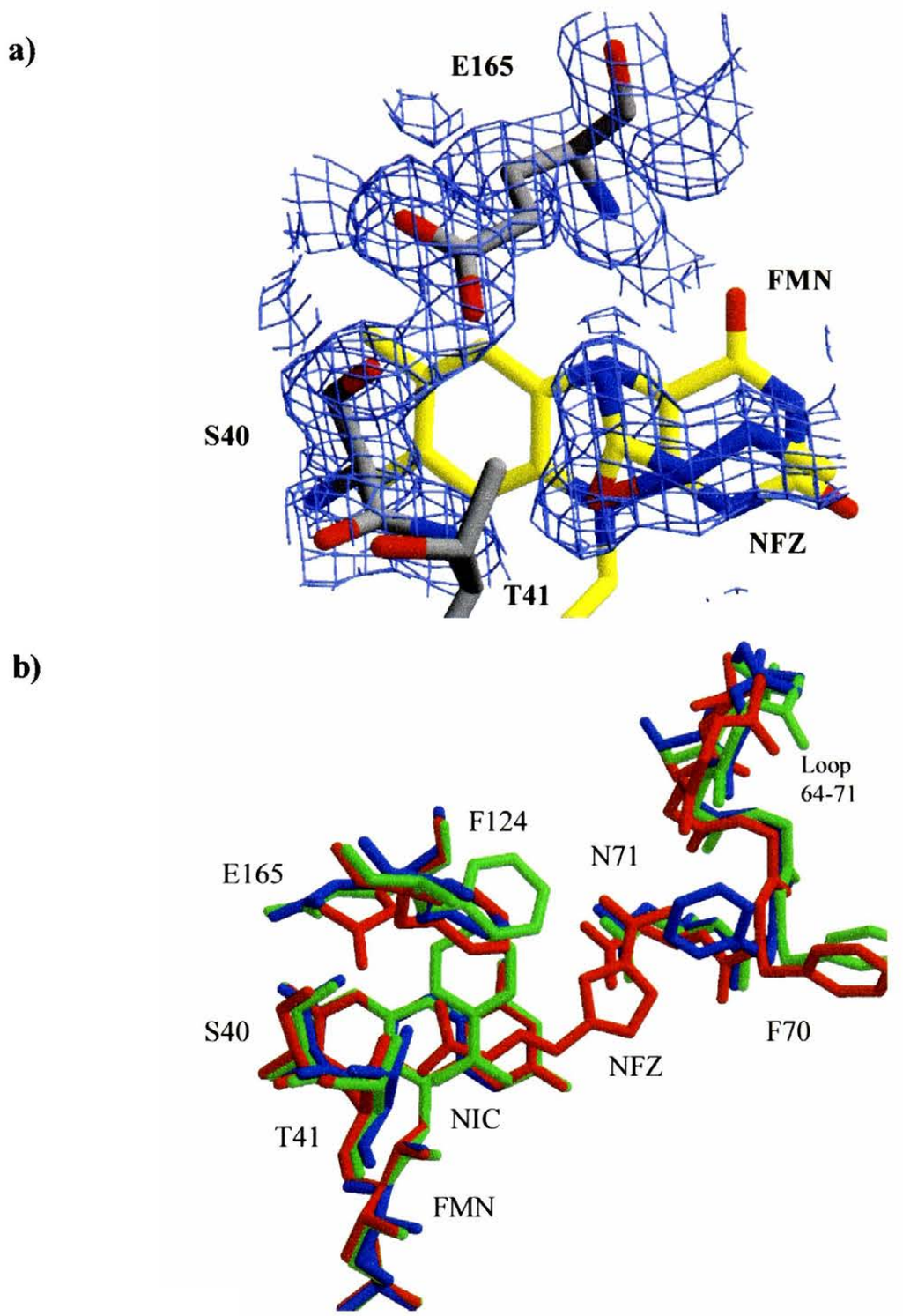


Figure 7.6. Changes observed upon binding nitrofurazone. a) Electron density of contact between E165 and S40, 2mFo-DFc style, SIGMAA-weighted, contoured at 1σ, resolution 1.7 Å. The electron density map shows continuous density between the two sidechains. Diagram coloured identical to Figure 7.2. b) Residue shifts upon binding nitrofurazone. The NTR : NIC complex (Lovering *et al.* 2001) is in green, the holo NTR structure (Parkinson *et al.* 2000) in blue, and the nitrofurazone complex in red. The nitrofurazone complex F124 is similar to the holo form, and F70 more resembles the NIC form. E165 of the nitrofurazone structure moves closer to the flavin, pushing S40 away slightly. The conformation of the loop formed by residues 64 to 71 is different between all 3 structures, but the nitrofurazone complex is different from the other two. Diagram produced using MOLSCRIPT (Kraulis 1991), BOBSCRIPT (Esnouf 1997) and RASTER 3D (Meritt and Bacon 1997).

One hypothesis is that the nitrogroup of nitrofurazone is reduced by nitroreductase to a nitroso-compound (2 electron reduction), and then reduced by a further 2 electrons to a toxic hydroxylamine product (Haynes *et al.* 2002). The final two-electron reduction may be enzymatic or non-enzymatic (reduced directly by NAD(P)H). Another possibility is that nitrofurazone is reduced at the aliphatic end of the molecule, resulting in the liberation of formamide (Guay 2001).

Mutations of NTR that affect the reduction of nitrofurazone (Whiteway *et al.* 1998) are of no use in analyzing amino acid residues important for catalytic activity, simply being interpretable as mutations that affect protein folding or FMN binding. With the crystal structure of the NTR:nitrofurazone complex, and the acetamide portion of the drug bound over the flavin, reduction away from the nitro group remains a possibility. Unfortunately, attempts to obtain a nitrofurantoin complex failed, leaving the chance that the nitrofurazone complex is something of an artefact of crystallization.

The question of nitrofurazone activation by NTR was answered by the use of HPLC and mass spectrometry of the reaction products (personal communication, Paul Race). It confirmed the presence of a nitroso intermediate, and upon anaerobic analysis under nitrogen, the 4 electron reduction hydroxylamine product. This indicated that the orientation of nitrofurazone observed in the crystal structure is that of a non-productive complex. The distance between the nitro group and the flavin N5 is 7.7 Å, with no real possibility of tunneling electrons between the two.

The observation of this non-productive complex can be rationalized in terms of active site changes upon reduction of NTR. The crystal structure with nitrofurazone was obtained with an oxidized enzyme (*i.e.* is a dead-end complex), and presumably nitrofurazone binds differently to the reduced enzyme. No large structural changes are observed upon reduction of the homologous enzyme from *E. cloacae* (Haynes *et al.* 2002), with the greatest change being a shift in the butterfly angle of the flavin from 16° to 25°. It was also noted that this reduced enzyme no longer bound acetate, in contrast to the oxidized form. As nitrofurazone binds to the oxidized NTR with the acetate-like acetamide group over the flavin, the reduction of the enzyme may favour binding of the nitrogroup over this region instead.

Multiple binding modes for ligands between an oxidized and reduced enzyme have been observed in another flavoprotein, PETN reductase (Barna *et al.* 2001). This scenario may be prevalent in relatively non-specific flavoenzymes, although its purpose remains unclear. This substrate "flipping" may be due to the distribution of charge on the flavin, changing with the two electron reduction.

7.4.2 Mutant Enzymes

Conformational diagrams (with descriptions of contacts formed between the mutated residue and other aspects of the active site) and comparisons to the wild-type structure are provided in Figures 7.7 to 7.9. A summary of the kinetic parameters of the mutant enzymes (personal communication, Paul Race) is provided in Table 7.1.

The mutation of N71 to serine yielded two distinct groups of structures, one with a bound water molecule between the serine and the FMN, and one without. The absence of a water

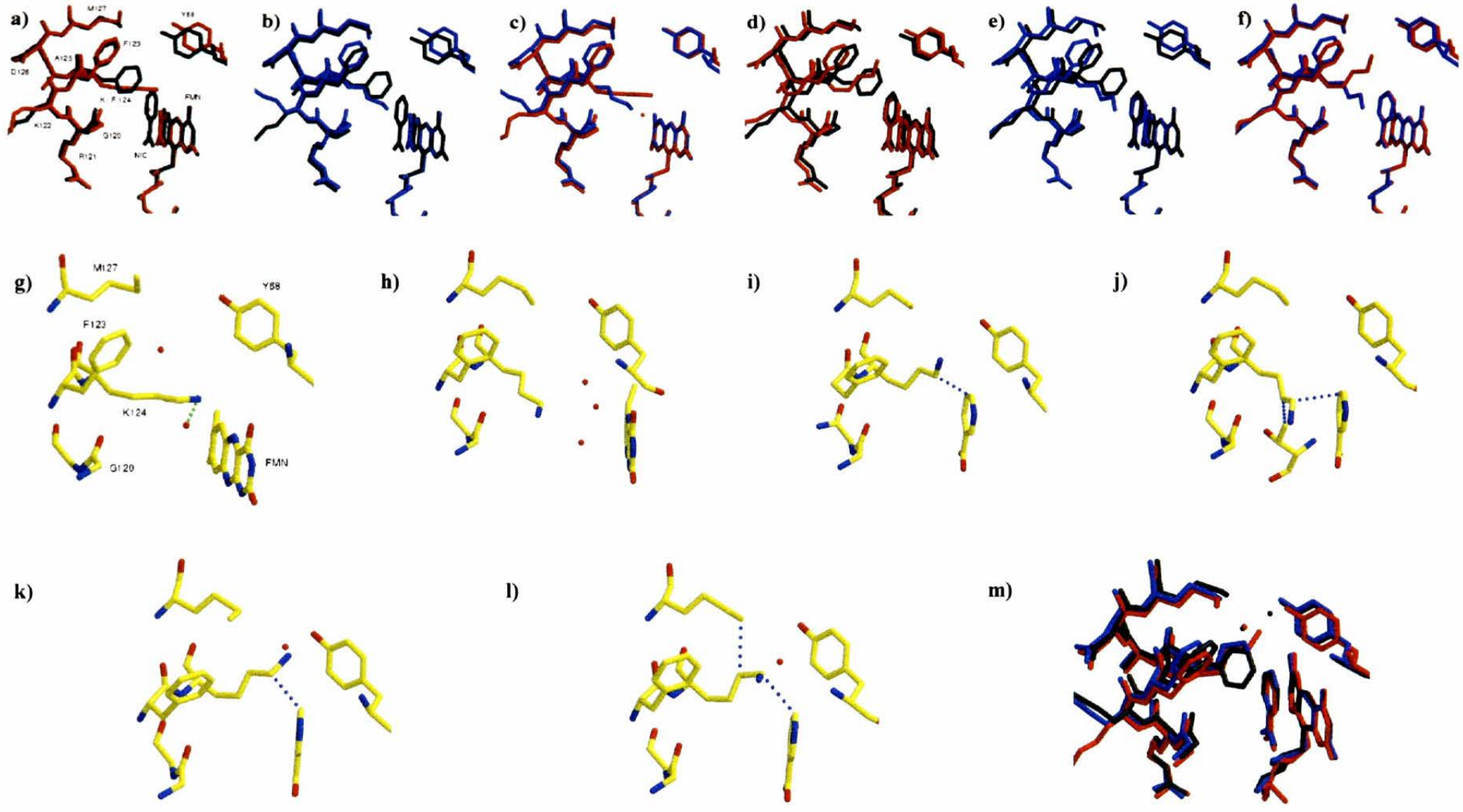


Figure 7.7. Contacts of K124 residue. a) to f) comparison of F124K structure to native [a) to c) - double mutant holo form chain A red, chain B blue, native enzyme grey ; d) to f) - double mutant NIC form chain A red, chain B blue, native enzyme grey]. g) to l) contacts of K124, coloured as Figure 7.2. g) + h) double mutant holo form, i) + j) double mutant NIC form, k) + l) single mutant, all show chain A followed by chain B. m) comparison of single mutant (chain A red, B blue) to native enzyme (grey). All figures are shown in the same orientation. Prepared using MOLSCRIPT (Kraulis 1991) and RASTER 3D (Meritt and Bacon 1997).

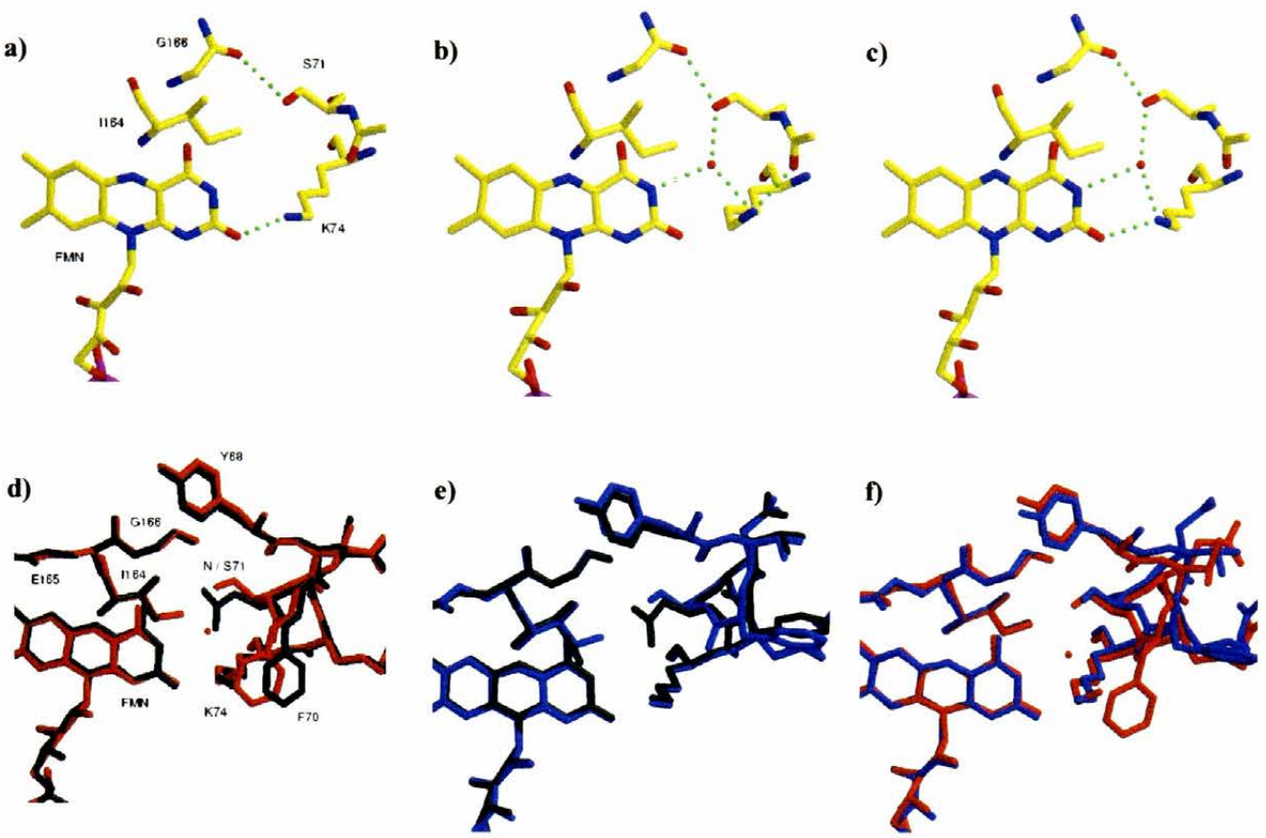


Figure 7.8. Contacts of S71 residue. a) to c) colouring as Figure 7.2. a) N71S mutants with no water bound near flavin, b) N71S mutants with water adjacent and K74 facing away from flavin, c) N71S mutants with water adjacent and K74 facing towards flavin. Hydrogen bonds shown as dotted green lines. d) to f) coloured according to structure : native enzyme, grey; N71S mutant with water adjacent, red; N71S mutant with no water adjacent, blue. Residue labelling in first panel only, omitted from the others for clarity. All figures are shown in the same orientation. Figure prepared using MOLSCRIPT (Kraulis 1991) and RASTER 3D (Meritt and Bacon 1997).

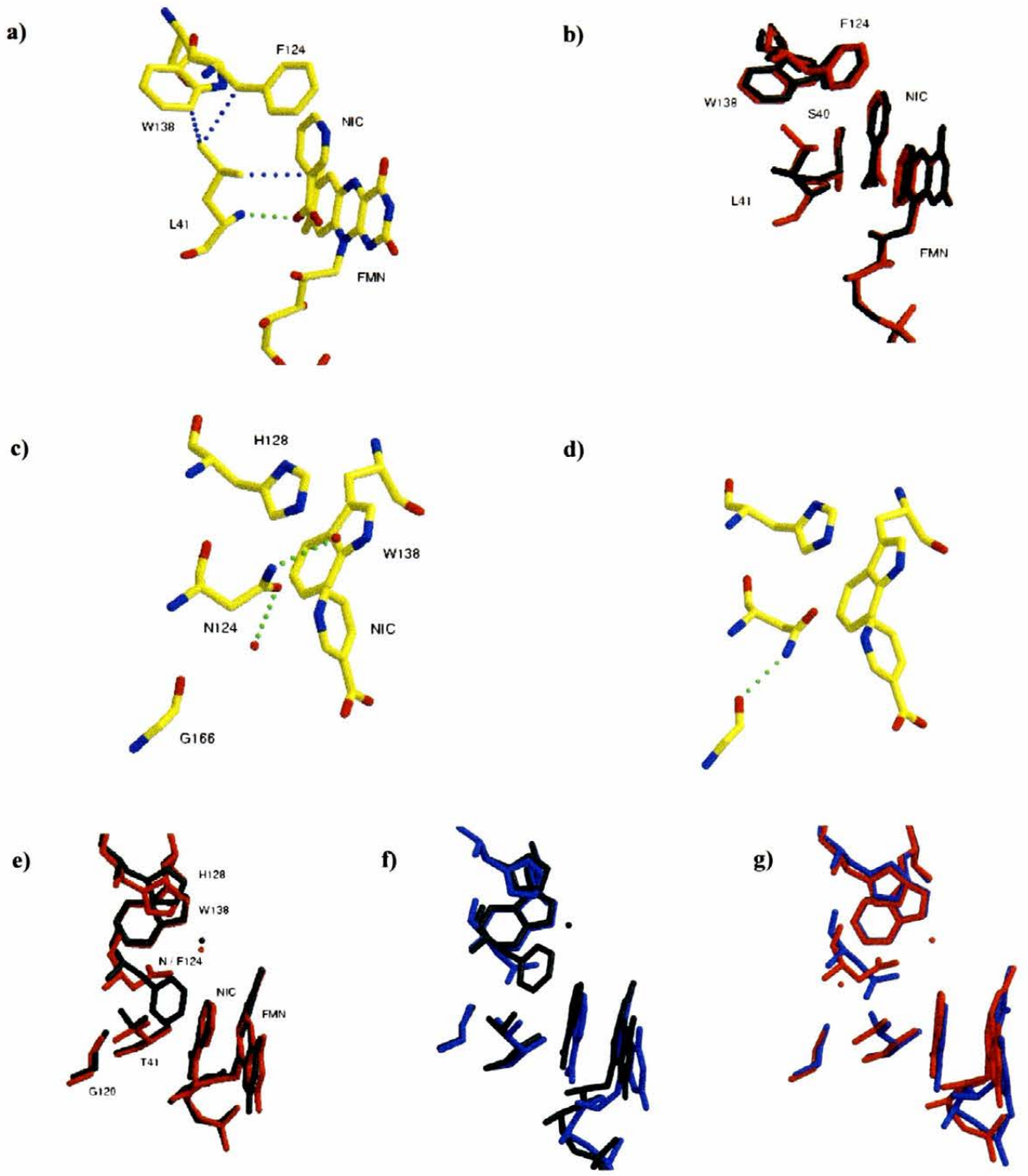


Figure 7.9. Contacts of other single mutants. a) T41L, coloured as Figure 7.2, hydrogen bonds as green dotted lines, VDW contacts as blue dotted lines. b) comparison of T41L structure (red) with native enzyme (grey). c) + d) contacts formed by N124 in the F124N mutant, c) chain A and d) chain D. Colouring as above. e) to g) comparison of F124N structure (chain A red, chain D blue) to native enzyme (grey). Residue labelling in first panel only, omitted from the others for clarity. All figures from same structure are shown in the same orientation. Figure prepared using MOLSCRIPT (Kraulis 1991) and RASTER 3D (Meritt and Bacon 1997).

molecule in chain B of the single mutant holo form cannot be attributed to low resolution data (2.4 Å) as the water molecule is observed in chain A. The water molecule sits 2.8 Å from the S71 OG, 2.9 Å from the FMN N3, and 2.6 Å from K74 NZ (measurements from the A chain, double mutant holo form), forming a hydrogen bonding network (Figure 7.8). In comparing the 7 mutant chains that have the water molecule bound, chain A of the single mutant holo form, and to a lesser extent, chain A of the double mutant NIC form, have the end of K74 rotating away from the FMN. In the former case this is quite extreme, and instead of hydrogen bonding to the FMN O2, K74 NZ now contacts G166 O (2.8 Å). Why differences are observed between these different forms remains unclear. Small movements can also be seen in the I164 sidechain, which although can be in multiple orientations (see chain B of the double mutant holo form, Figure 7.8), prefers to orient itself away from the cofactor, possibly as a result of the introduced water molecule (although the converse is not observed in the chain with the water molecule absent). The N71S mutation also causes general movement of the main chain loop in this region (Figure 7.8). The lack of recovery of a N71G mutant may indicate that increased flexibility in this area is detrimental to enzyme activity.

A reason for the improved catalytic activity of the N71S mutant (at least with regards to CB1954 turnover) must exist. An attractive proposition is that this mutation increases the polarity of the isoalloxazine environment (chiefly via hydrogen bonding from the introduced water molecule), favouring reduction of the FMN cofactor. The wild-type asparagine residue is too far away to form hydrogen bonds to the FMN (ND2 : FMN O4 = 3.8 Å, OD1 : FMN N3 = 3.8 Å), and there is no space for an intervening water molecule. Other mutations at this position (D, Q, T) support this theory somewhat, although the preference of Q over E suggests that charge too close to the cofactor is not tolerated. The lack of an improved enzyme with

histidine at position 71 may be due to detrimental charge repulsion with the neighbouring K74 residue.

Any modulation of the enzyme redox potential will have two effects – one on the relative rate of reduction by nucleotide, and an opposite effect on the relative rate of oxidation by substrate. The redox potential of NTR is –210 ± 5 mV (personal communication, Dr Simon Daff, Department of Chemistry, University of Edinburgh), that of the NAD^+/NADH couple -320 mV and CB1954 -385 mV. There exists the potential to lower the redox potential of NTR toward that of NAD^+/NADH, sacrificing the rate of the reductive half-reaction of the ping-pong mechanism to facilitate a more favourable oxidative half-reaction. The actuality may be more complex, however, with any residue mutation also likely to have unrelated effects, e.g. substrate fit to active site. It would be advantageous to have accurately measured redox potentials for all of the mutant enzymes.

The mutation of F124 to lysine gave a greater conformational variability than the N71S mutant (Figure 7.7), presumably due to the increased number of torsion angles possible in a lysine sidechain. No distinct pattern emerges in the analysis of the six forms, other than that when NIC is not present K124 is free to extend further toward the flavin. When NIC is present the sidechain moves away to accommodate the ligand. With NIC being a fairly small molecule, K124 may have to move even further when binding bulkier substrates such as CB1954. The mutation of F124 to lysine has a minimal effect on other residues in the active site (Figure 7.7), merely causing slight movement of the neighbouring F123 sidechain. In the variety of conformations observed K124 makes mainly VDW contacts to the NIC ligand (where present) and active site residues (T41 and M127). In the double mutant holo structure

chain A, the lysine NZ makes a hydrogen bond (2.6 Å) to a water molecule occupying the space where NIC would bind.

Although the conformation of K124 is variable, the Cα – Cε atoms of the sidechain are usually in the same position as the C atoms of the F124 sidechain (Figure 7.7), indicating a potential preference of the active site in positioning aliphatic sidechains at this residue. The lack of any large scale movement of residues in the active site, or definite position of the K124 NZ makes it difficult to analyse how this mutant possesses an increased activity for CB1954 reduction compared to the wild-type enzyme. This may become apparent in an enzyme : substrate complex structure.

The introduction of a positive charge at position 124 may increase the activity of NTR by favourable interactions with the negatively charged substrate groups. Indeed F124H also shows increased activity, F124R presumably not obtained from the mutant screening procedure due to unfavourable steric clashes upon introduction of such a large sidechain, leading to a lower enzyme activity. This hypothesis may have some promise as no F124D or F124E mutants were recovered. The only other F124 mutation not recovered was F124P, understandable in that a proline residue could act as a "helix-breaker" in helix F. With such a wide variety of mutations at F124, it would be prudent to say that not all mutations may lead to an improved NTR for precisely the same reasons. The best candidates at this position all possess some flexibility (M, H, K ; N less so), and are usually polar. The error bars involved in such a scoring protocol (Figure 7.1) can mean that the wild-type enzyme, W124, Y124 and I124 can be grouped together. It would not be unreasonable to suggest that (for CB1954 reduction) bulky hydrophobic residues at this position hinder turnover. This may be due to

substrate access, or unfavourable interactions with the polar substituents on the aromatic ring of the substrate. The mutation of F124 to lysine is flexible enough to accommodate a range of binding orientations of CB1954. The prevalence of K124 over G124 suggests that interaction of some sort is preferred.

No reason can be given for chain D of the F124N mutant occupying a different conformation to the other three chains (Figure 7.9), other than that position 124 is near to the protein surface and may be influenced by subtleties in crystal packing. Using chain A as an example, N124 makes two hydrogen bonds to separate water molecules, both 3 Å away (Figure 7.9). Chain D of F124N is different from chains A-C, atom ND hydrogen bonding to G120 O (2.9 Å). Both conformations show the polar interactions possible upon mutation from phenylalanine. As N124 does not alter any of the surrounding residues (although an alternate conformation of T41 is observed, this can also be seen in some high resolution wild-type studies, unpublished data), it must increase NTR activity by interaction with substrate.

The mutation of T41 to leucine is intriguing as the K_M and k_{cat} for CB1954 are increased compared with the wild-type enzyme. The T41L mutant is also the only mutant investigated that alters the ratio of the 2 / 4-hydroxylamine products of CB1954 reduction (personal communication, Paul Race), from 1 : 1 to 1 : 3, yielding more of the desired toxic product. The leucine residue makes Van der Waals contacts with NIC, F124 and W138 (Figure 7.9). The backbone conformation of this mutant L41 residue remains identical to T41 of the wild-type enzyme, forming a hydrogen bond between the backbone nitrogen and the NIC oxygen. The mutation does not affect any other active site residues, and presumably the increased activity is due to a direct interaction with substrate. Analysis of other mutations

recovered at position 41 does not yield any clear result (G, I, N and S). Presumably S41 behaves similarly to T41, and G41 allows more possible orientations of substrate binding but has less substrate interaction. With L/I/N 41 it appears that medium-sized flexible, uncharged residues are preferred, making the inclusion of G and S, but the omission of A puzzling. This may be an artefact of the mutant screening procedure.

To understand why L41 may change the 2 / 4-hydroxylamine ratio, CB1954 was modelled into the active site, its central ring superimposed with that of NIC (Figure 7.10). The 2 possible "right" orientations have the 4-nitro group at position A (3.5 Å from the nitro N atom to the flavin N5 for hydride transfer), and the 2 "wrong" orientations switch the 4-nitro group for the 2-nitro group (which would lead to formation of the non-toxic product). The modelling serves only to act as a guide, as the ring systems of CB1954 and NIC may not bind to NTR identically. Table 7.6 describes the environment of the protein near the substituents of the ring.

Position (pocket)	Description
A	Over FMN N5, clash with S40, near T41
B	Clash with E165, near S40
C	Very close to F124, near Y68 and G166
D	Near F124 and N71
E	Projecting into solvent
F	Near T41

Table 7.6. Environment of protein near modelled CB1954 (also see Figure 7.10)

Positions A and F are closest to residue 41. Position A is occupied by a nitrogroup in all four of the modelled orientations, position F is occupied by a H atom or aziridine group in the "right" models and a H atom or $CONH_2$ group in the "wrong" models. Although both sets allow a hydrogen atom substituent at position F, in justification of the T41L mutant producing

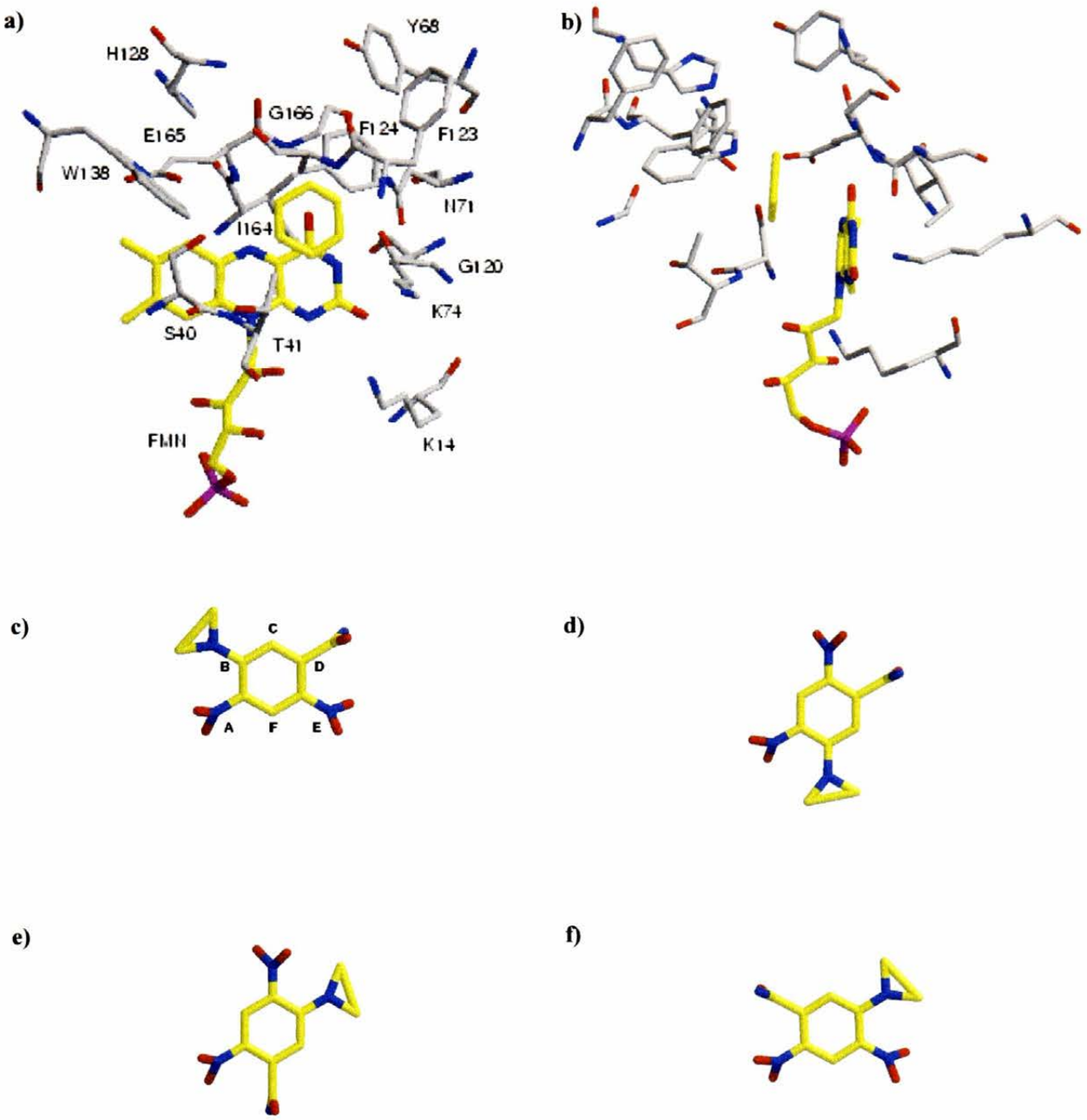


Figure 7.10. Modelling of CB1954 into active site of NTR. Diagram coloured according to atom type : protein C, grey; FMN and ligand C, yellow; O, red; N, blue; P, magenta. a) + b) two orthogonal views of active site, with the ring of NIC (acid group removed) used to model CB1954 upon. c) to f) possible orientations when placing a nitrogroup of the ligand over N5 of the FMN. c) + d) illustrate the two "right" orientations, where the desired 4-hydroxylamine would be produced. e) + f) illustrate the two "wrong" orientations, where the non-toxic 2-hydroxylamine would be produced. Protein environment of the ring substituents labelled as A to F. Figure prepared using MOLSCRIPT (Kraulis 1991) and RASTER 3D (Meritt and Bacon 1997).

more of the 4- than the 2-hydroxylamine, the hydrophobic leucine would interact more favourably with an aziridine group than a $CONH_2$ group.

The improvement of NTR by mutation becomes a complex issue when considering multiple mutants. The combination of single mutants may be additive or multiplactive, or indeed they may negate one another. There is also no indication that mutant enzymes will behave identically in *E. coli* and tumour cells.

7.4.3 Other Mutants – No Structural Data

Mutation at serine 40 recovered A, T and G mutations, indicating the need for a small residue at this position, large residues presumably clashing with bound substrate. Mutation at tyrosine 68 recovered a range of mutants (G, N, A, H, S, W, C, Q, D), the omission of a phenylalanine mutant a possible artefact of the screening protocol. Small residues at position 68 score well, indicating a possible role in accessibility to the active site or removing steric hindrance. Mutations recovered at position 70 (P, A, V, S, I, L, E, C, Q, G, T, D) show a similar lack of cohesion, other than the preference for a smaller residue (no W, Y or R). Mutations at glycine 120 (A, T, S) indicate the importance of a small side chain near the entrance to the active site.

CHAPTER 8

THE STRUCTURE OF AKR1C3

8.1 Introduction

From the starting point of a pure and homogenous source of AKR1C3, it was desired to obtain well-diffracting protein crystals to lead to eventual use in structure-based drug design and binding-mode studies. It would therefore be advantageous to have high resolution data and crystals amenable to soaking in substrates and inhibitors (good mechanical stability, solvent channels in close proximity to the enzyme active site). Derivitisation for phasing purposes would not be necessary due to the high sequence similarity of previously solved homologues (Hoog *et al.* 1994; Bennett *et al.* 1996; Nahoum *et al.* 2001), but the structure / function differences between family members would have to be addressed with respect to the small number of amino acid residue changes between enzyme subtypes.

8.2 Materials & Methods

8.2.1 Sample Preparation

Purified HSD was a generous gift from Dr Jon Ride (School of Biosciences, University of Birmingham). The protein was supplied at a concentration of 20 mg/ml in a buffer containing 10 mM KH_2PO_4, pH 7.0, 1 mM DTT, 1 mM EDTA and 2 mM $NADP^+$.

8.2.2 Crystallization

Using a starting point of sparse matrix screening (solutions 1-96 from screening kits I and II from Molecular Dimensions), a series of hanging-drop vapour diffusion trials were set up, using 3 µl of protein stock mixed with 3 µl of reservoir solution, equilibrated against 1 ml of reservoir solution in VDX plates. From the analysis of crystallization conditions for

homologues of HSD (Zhu *et al.* 2001), salt concentration appeared to be critical in obtaining different crystal forms, and was treated as the most important factor in subsequent fine screening optimization of the growth conditions. Final crystallization conditions were 100 mM sodium citrate, pH 6.0, 0.4 or 0.8 M ammonium acetate, 25 % w/v PEG 4000 and 2.5 % v/v MPD, and are discussed in more detail in section 8.3.1

8.2.3 Inhibitor Soaking

Crystals were soaked in a buffer containing 100 mM sodium citrate, pH 6.0, 0.8 M NaCl, 25 % w/v PEG 4000, 10 % v/v ethylene glycol, 5 % v/v DMSO and 5 mM indomethacin or flufenamic acid for a period of 2 hours. Crystals could be flash-cooled direct from this solution.

8.2.4 Cryoprotection

For the binary complex structures ($NADP^+$ and acetate bound), crystals were soaked for 2 minutes in a buffer identical to the reservoir solution but supplemented with an extra 12 % v/v MPD. This was performed in a single step, with the crystals then being transferred directly to the cryo-stream at 100 K.

8.2.5 Data Collection

Data for 1) the low-salt $P3_221$ crystal form, 2) the indomethacin:NADP complex $P2_12_12_1$ form and 3) the flufenamic acid:NADP complex $P2_12_12_1$ form were collected at the ESRF on beamline ID14-2. Data for the high salt $P2_12_12_1$ crystal form were collected at the ESRF on beamline ID14-1.

8.2.6 MR Phasing

The protein chain (all atoms) of the human type III 3α-hydroxysteroid dehydrogenase (AKR1C2) complexed with $NADP^+$ and ursodeoxycholate (PDB entry 1IHI, Jin *et al.* 2001) was used as the search model for molecular replacement. Data for the $P3_221$ crystal form were entered into the CNS suite molecular replacement package (Brunger 1990). The fastdirect option and a resolution cutoff of 15-4 Å were used. The $P2_12_12_1$ crystal form was solved in a similar manner, with the refined $P3_221$ form of AKR1C3 used as the search model.

8.2.7 Refinement Protocol (all 4 structures)

All models were built using the molecular graphics program TURBO-FRODO (Roussel *et al.* 1991). Initial refinement steps were analogous to those described for the NTR protocol (see Chapter 5), followed by several rounds of TLS refinement using REFMAC5 (Winn *et al.* 2001). Due to the high resolution of the $P2_12_12_1$ form data, individual anisotropic B-factors were refined, as opposed to isotropic B-factors for the other three structures. Geometric restraint files for the NSAID inhibitors were generated interactively using CCP4i sketcher (CCP4 1994).

8.3 Results

8.3.1 Crystallization

Poor crystals were identified growing in the well from MDL Crystal Screen I solution 2 (0.1 M sodium acetate, pH 4.6, 0.2M ammonium acetate, 30 % w/v PEG 4000). From optimisation of this initial "hit", two conditions giving different crystal morphologies were identified, differing only in the concentration of ammonium acetate present (0.4 M gave diamond-shaped crystals, 0.8 M cuboids). Typical refined conditions for crystal growth were

100 mM sodium citrate, pH 6.0, 0.4 or 0.8 M ammonium acetate, 25 % w/v PEG 4000 and 2.5 % v/v MPD. The two optimised crystal forms and initial crystalline results can be seen in Figure 8.1. The crystals grew over a period of 2-3 days and reached maximum dimensions within a week (~0.4 x 0.3 x 0.3 mm for the diamond-like crystals, ~0.6 x 0.2 x 0.2 mm for the cuboids).

8.3.2 Data Collection & Phasing

Upon diffraction analysis of the crystals, it was apparent that the two crystal morphologies corresponded to two distinct spacegroups, both possessing one monomer per ASU (V_m calculations, results not shown). The diamond-shaped, low salt crystals belonging to the trigonal spacegroup $P3_221$ and diffracting to ~2.1 Å resolution, and the cuboids belonging to the orthorhombic spacegroup $P2_12_12_1$ and diffracting to ~1.1 Å resolution. Full data collection statistics for both forms and the inhibitor soaks are shown in Table 8.1. The $P3_221$ form possesses a large c axis of ~250 Å, subject to disorder and responsible for the choice of the $P2_12_12_1$ form for the inhibitor soaking. Typical diffraction images for both spacegroups are shown in Figure 8.2.

The list of peaks obtained from the cross-rotation function in the first step to solve the $P3_221$ form are shown in Table 8.2, with a clear solution as the top peak. The translation function was used to determine the $P3_1$ or $P3_2$ screw-axis ambiguity, with top solutions of monitor = 0.3 and packing = 0.5 for $P3_1$, and monitor = 0.6 and packing = 0.5 for $P3_2$. The molecular replacement step solving the $P2_12_12_1$ form with the refined $P3_221$ structure gave a clear solution in the cross-rotation function, and a top translation function answer with monitor = 0.7 and packing = 0.6.

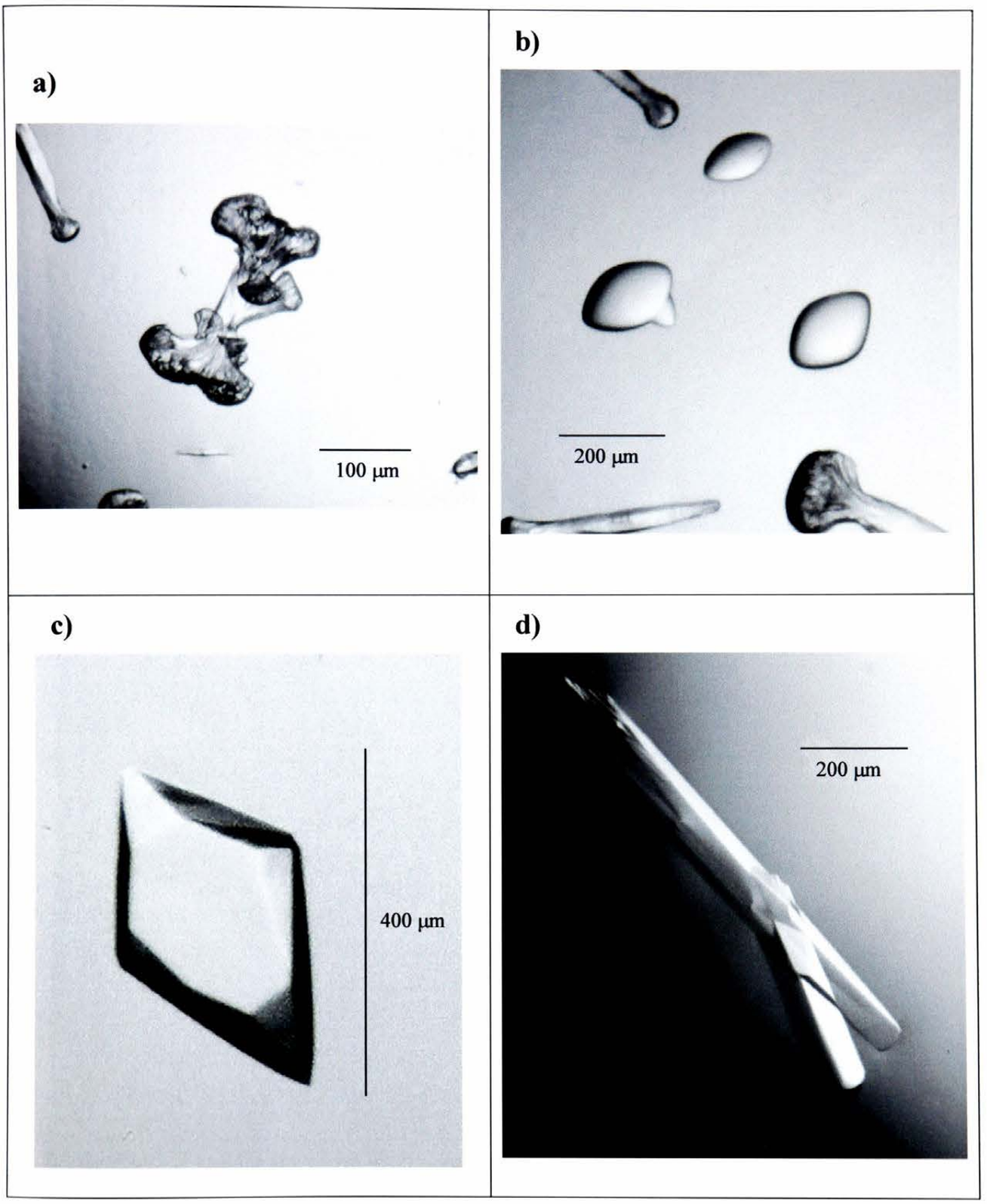


Figure 8.1. AKR1C3 crystallization photographs: a) Initial crystalline results using MDL screen I solution 2; b) Round crystals, possibly obtained by too fast an equilibration between protein drop and reservoir solution; c) Diamond-like $P3_221$ crystal form, low salt (0.4M ammonium acetate); d) Cuboid-like $P2_12_12_1$ crystal form, high salt (0.8M ammonium acetate).

Space Group	$P2_12_12_1$			$P3_221$
Ligands	Acetate, $NADP^+$	Flufenamic Acid, $NADP^+$	Indomethacin, $NADP^+$	Acetate, $NADP^+$
Cell Parameters (Å,°)	55.81 x 63.07 x 96.20	55.80 x 63.00 x 96.29	55.68 x 63.12 x 96.11	58.06 x 58.06 x 206.32
	$\alpha, \beta, \gamma = 90$	$\alpha, \beta, \gamma = 90$	$\alpha, \beta, \gamma = 90$	$\alpha, \beta, = 90, \gamma = 120$
Data Collection Statistics [a]				
Resolution (Å)	38 – 1.2 (1.26 – 1.20)	48 – 1.8 (1.90 – 1.80)	20 – 1.7 (1.79 – 1.70)	51 – 2.1 (2.21 – 2.10)
Completeness (%)	98.5 (90.1)	99.7 (98.1)	98.0 (88.3)	94.7 (74.6)
Multiplicity	5.0 (2.6)	4.4 (2.8)	3.4 (2.7)	10.6 (1.9)
$I/\sigma I$	16.4 (4.1)	13.6 (3.8)	18.7 (4.7)	22.0 (2.6)
No of Observations	522,433 (36,225)	141,396 (12,414)	127,566 (12,797)	243,584 (4,813)
No of unique reflections	105,032 (13,800)	32,084 (4,459)	37,296 (4,785)	23,069 (2,550)
Rsym (%)	5.9 (24.2)	7.8 (25.6)	4.9 (20.0)	8.7 (22.0)

[a] values in parentheses are for the outer shell

Table 8.1. Data collection statistics for all four HSD structures.

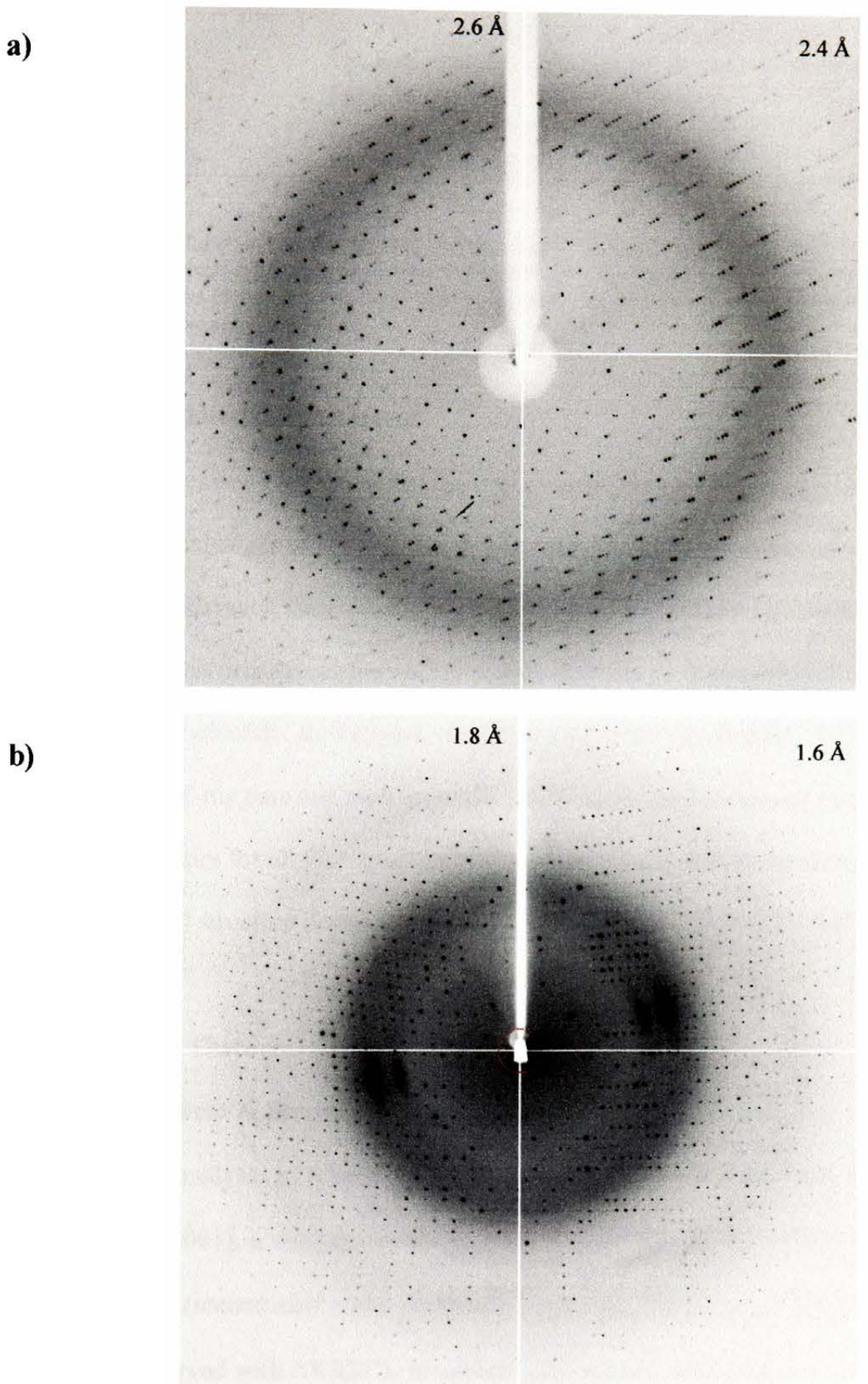


Figure 8.2. Diffraction patterns of AKR1C3 crystals. a) $P3_221$ crystal form, 0.5° oscillation, ESRF Beamline ID14-2. b) $P2_12_12_1$ crystal form, 1° oscillation, ESRF beamline ID14-1. Note the larger spacing between spots on the smaller cell $P2_12_12_1$ form, and the long reciprocal c-axis spot arrangement of the $P3_221$ form (smaller spacing between diffraction spots). Resolution given at corner and edge of diagram.

Trial Number	Theta1	Theta2	Theta3	RF value
1	267.711	56.250	257.035	0.1056
6	329.828	47.135	243.387	0.0364
7	252.674	47.135	263.351	0.0349
9	275.275	67.500	241.681	0.0336
10	269.966	45.573	256.643	0.0331

Table 8.2. Rotation function peak list for cross rotation step in molecular replacement with $P3_221$ crystal form. Theta angles are defined in CNS (Brunger *et* al. 1990), RF value represents scoring of rotation function step, comparing the vectors of the diffraction data with those generated from the placement of the trial model. Peaks shown are the top five scoring solutions.

8.3.3 Model Building & Refinement

Initial electron density from the molecular replacement-determined phases was excellent, allowing residue changes to be made from the search model to our sequence ($P3_221$ form, Figure 8.3), and showing clear density for $NADP^+$ in all 4 structures. In the inhibitor-soaked structures, electron density corresponding to indomethacin or flufenamic acid was visible adjacent to the nucleotide. Refinement of all four structures was straightforward owing to the high resolution X-ray data and high sequence similarity to the MR search model. Final refinement statistics for all four structures are shown in Table 8.3. Refined electron density for the individual structure determinations can be seen in Figures 8.4, 8.7 and 8.8.

8.3.4 Structural Features

8.3.4.1 Protein Fold & Secondary Structure Assignments

From structural analyses of related AKR family members (Hoog *et al.* 1994; Bennett *et al.* 1996; Jin *et al.* 2001), it was presumed that AKR1C3 would fold in a similar manner, that of a TIM $\alpha_8\beta_8$ barrel (named after triose phosphate isomerase, Banner *et al.* 1976). The $\alpha_8\beta_8$ fold was indeed observed with AKR1C3, in tandem with several other features common to the AKR family of proteins. A schematic of the secondary structure assignments (as designated

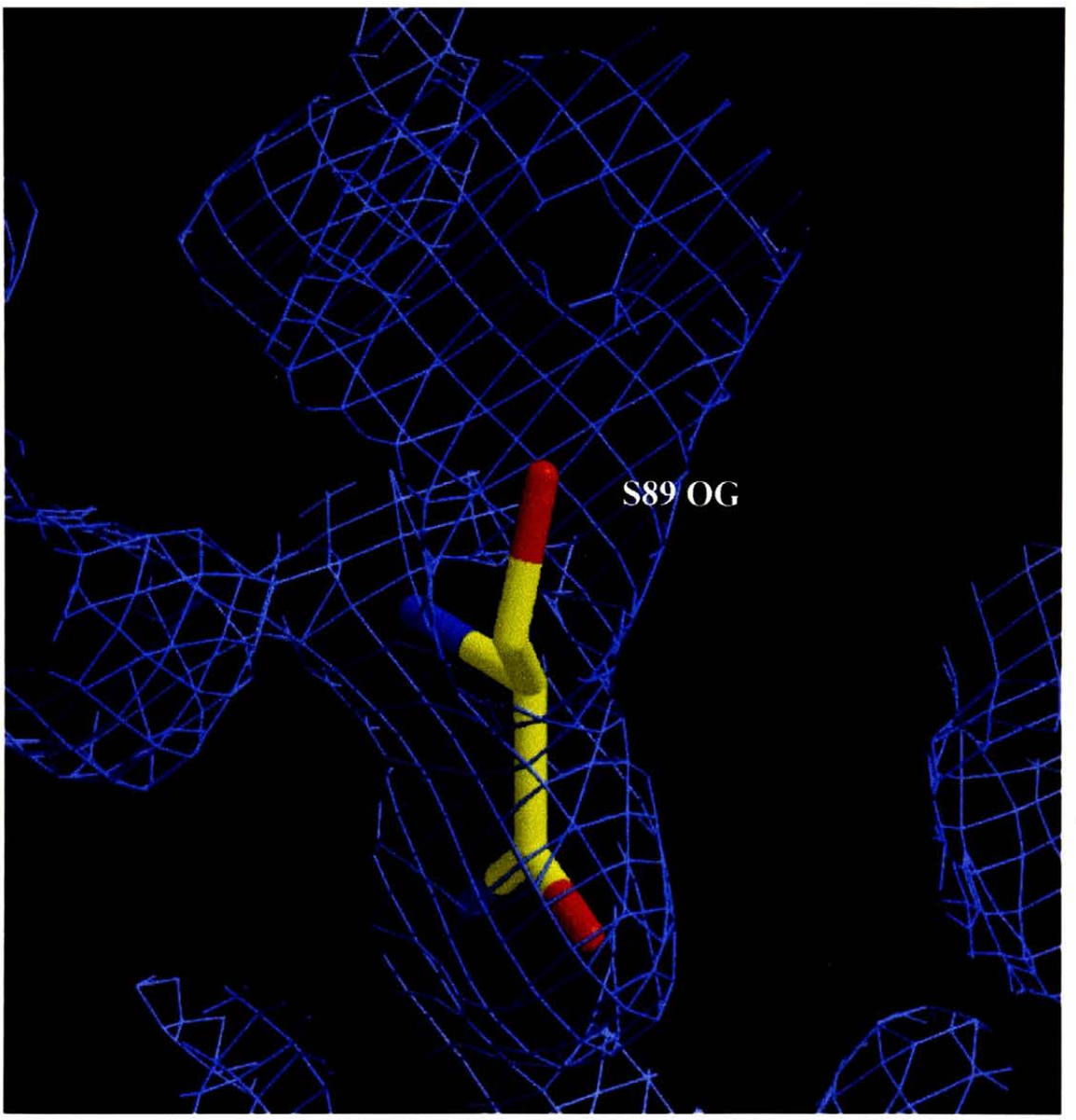


Figure 8.3. Lack of pre-existing model bias in AKR1C3 molecular replacement results. Electron density map is 2mFo-DFc, SIGMAA-weighted, 1.5 Å resolution. Map calculated from phases of correctly placed 1IHI.pdb (Jin *et al.* 2001) molecular replacement solution, and amplitudes from $P3_221$ crystal form dataset. The residue shown is that of S89 of 1IHI.pdb, with clear density for the corresponding residue F89 of AKR1C3. Diagram produced using BOBSCRIPT (Esnouf 1997) and RASTER 3D (Meritt and Bacon 1997).

Crystal Form	Acetate Low Salt	Acetate High Salt	Indomethacin Soak	Flufenamic Acid Soak
Ligands	$NADP^+$ acetate	$NADP^+$ Acetate	$NADP^+$ indomethacin	$NADP^+$ Flufenamic acid
Spacegroup	$P3_221$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Refinement				
Resolution Range (Å)	30-2.1	30-1.2	30-1.7	30-1.8
No. of Non-hydrogen Atoms	2889	3075	2964	2858
No. Water Molecules	296	434	363	237
R factor[a]	15.2 %	12.7 %	17.2 %	15.7 %
R_{free}[b]	19.5 %	14.4 %	19.8 %	20.3 %
Average B factor ($Å^2$)	13.87	8.91	16.03	11.66
RMSD Bond Angles[c] (°)	1.35	2.02	1.60	1.82
RMSD Bond Lengths[d] (Å)	0.012	0.020	0.009	0.016

[a] R factor = $\Sigma_{hkl}| |F_o| - |F_c| |/\Sigma_{hkl}|F_o|$.

[b]R factor based on 5 % of the data withheld from refinement

[c]RMSD bond angles and [d]bond lengths are the root mean square deviations from ideal values

Table 8.3. Refinement statistics of HSD structures.

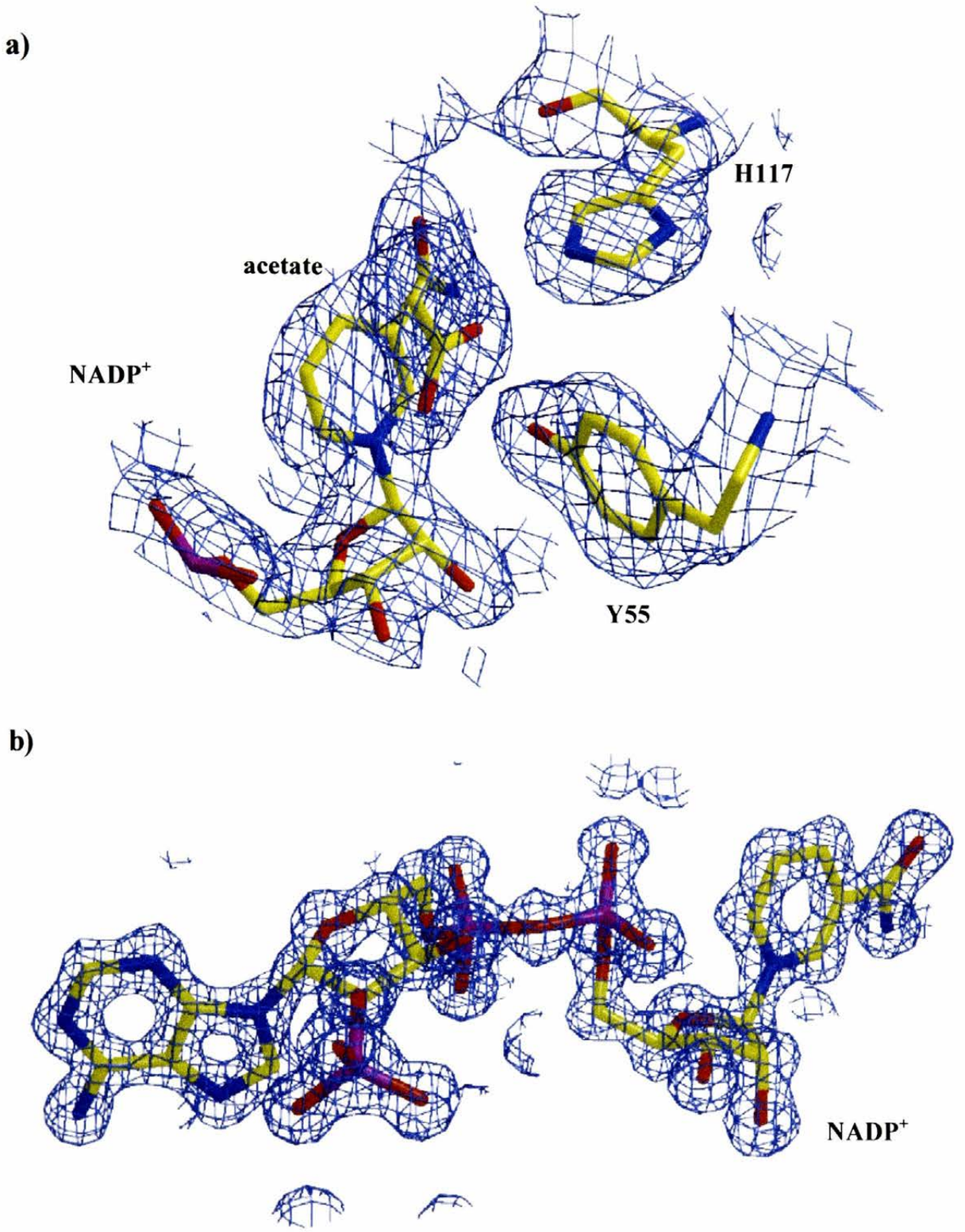


Figure 8.4. Refined electron density from both crystal form structure determinations. Both maps are 2mFo-DFc, SIGMAA-weighted, contoured at 1σ. a) $P3_221$ crystal form (2.1 Å), showing the nicotinamide headgroup of the NADP' cofactor, bound acetate molecule, Y55 and H117. b) entirety of NADP' cofactor, $P2_12_12_1$ crystal form (1.2 Å). Diagram produced using BOBSCRIPT (Esnouf 1997) and RASTER 3D (Meritt and Bacon 1997).

by PROCHECK, Laskowski *et al.* 1993) is given in Figure 8.5. Secondary structure nomenclature is adapted from aldose reductase (Wilson *et al.* 1992).

The barrel is closed at the N-terminal end by an antiparallel β-hairpin formed from residues 7-17 of the protein. The hairpin then feeds into the first stave of the barrel, β-strand 1. The $\alpha_8\beta_8$ compact fold is augmented by an extra two α-helices, H1 preceding α7, and H2 following α8. These two extra helices contact one another on the periphery of the barrel. Smaller helical regions are present in some of the loops. Several loops extend from the C-terminal end of the barrel, forming the active site of the enzyme. Loop A is composed of residues 118-144, loop B from residues 217-238, and loop C from the C-terminal of the protein sequence (from H2 onward). Residues 1-5, residue 321 onwards and the C-terminal His tag are presumably disordered as no electron density for these was observed in experimental maps. The overall protein fold and loop structure can be seen in Figure 8.6.

8.3.4.2 The Active Site

8.3.4.2.1 The Cofactor

All four structures have a bound $NADP^+$ molecule, indicative of the tight binding constant of the nucleotide (K_d = 40 nM, personal communication, J.P. Ride). The electron density for all parts of the $NADP^+$ cofactor is excellent in all of the maps, especially so in the $P2_12_12_1$ acetate complex at 1.2 Å (Figure 8.4). The $NADP^+$ is present in an extended conformation, bound by the loops present at the C-terminal end of the barrel. Looking along the axis of the barrel, down at the cofactor (Figure 8.6b), loops A and B are roughly opposite one another, with loop C on one side and an open pocket on the other. Another opening at 90° to the barrel axis allows entry of the nucelotide.

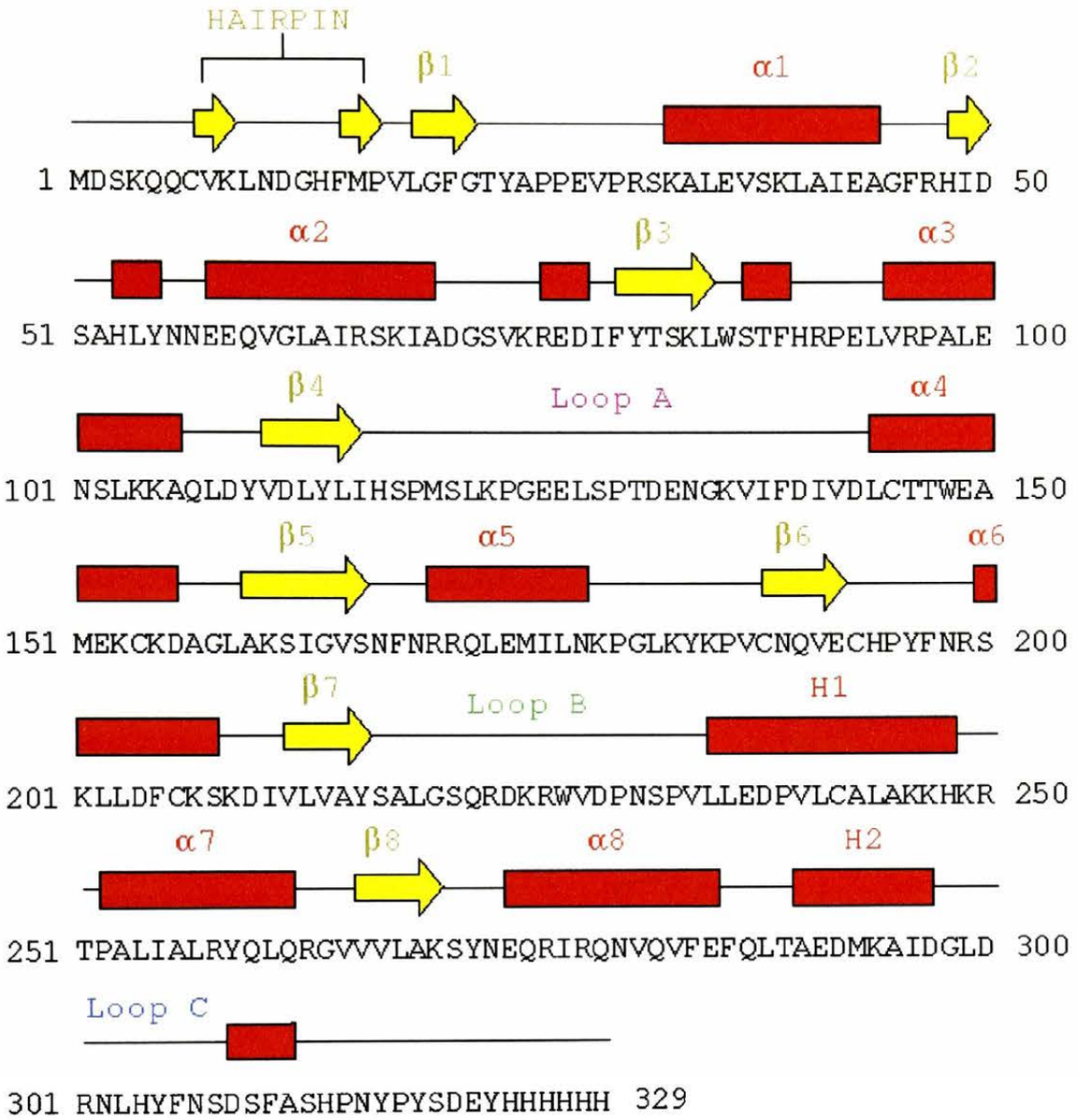


Figure 8.5. Secondary structure assignments for AKR1C3. N-terminal hairpin, beta strands 1-8, alpha helices 1-8, H1 and H2 labelled. Smaller helices present in some of the loops remain unnumbered. Colouring corresponds to Figure 8.6. Secondary structure was assigned to the final model by use of PROCHECK (Laskowski *et al.* 1993). Nomenclature adapted from aldose reductase (Wilson *et al.* 1992)

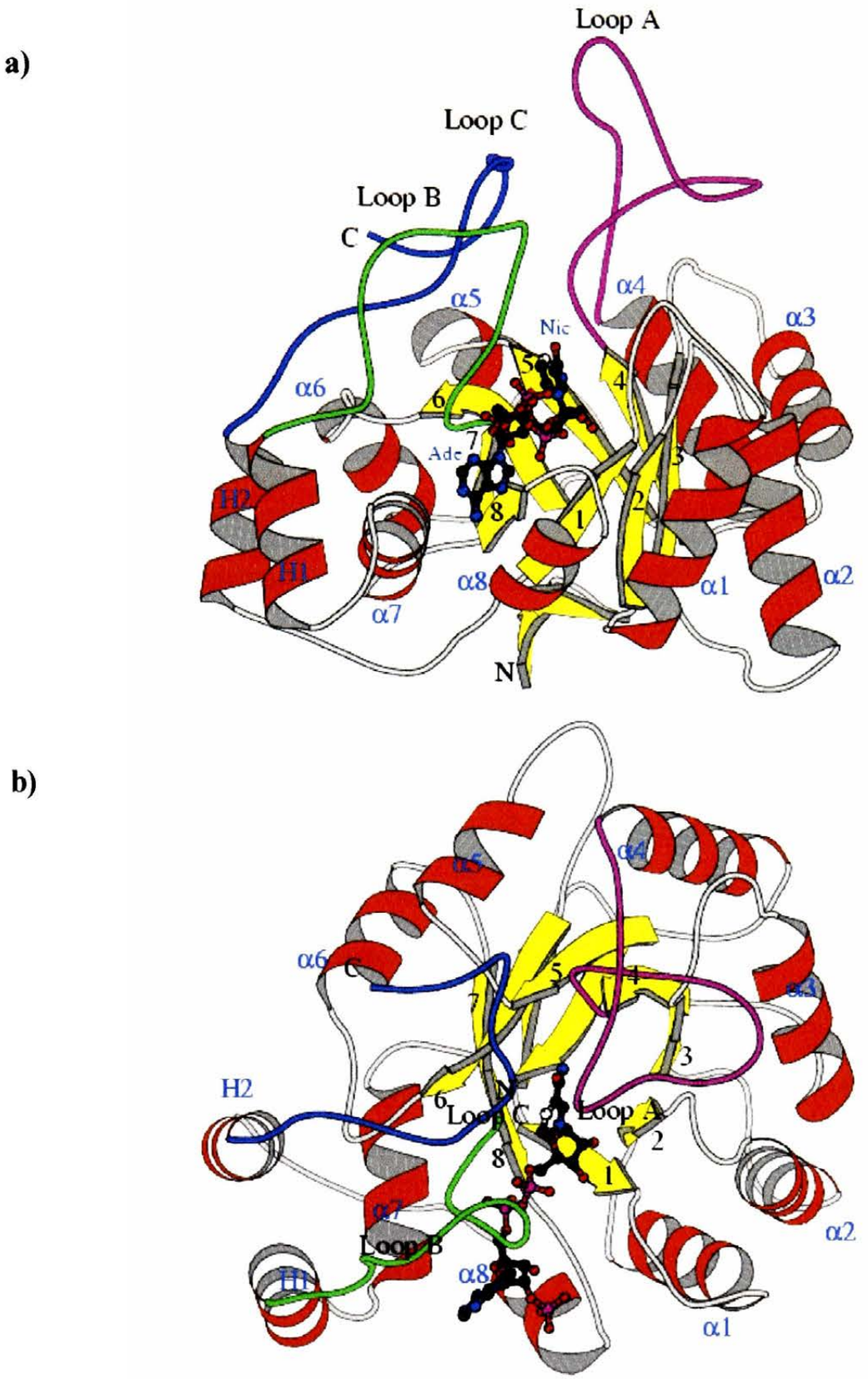


Figure 8.6. Two orthogonal views of secondary structure of AKR1C3. Nucleotide present as ball-and-stick model. Labelling is as follows : N, N-terminus; C, C-terminus; Ade, adenine ring of NADP'; Nic, nicotinamide ring of NADP; secondary structure elements labelling identical to that of Figure 8.5. Figure produced using MOLSCRIPT (Kraulis 1991).

Residues K270 and R276 form salt bridges with the 2' phosphate, and S271 and Y272 contribute hydrogen bonds. AKR1C3 shows a preference for NADP(H) over NAD(H), which can be attributed to these charged contacts to the 2` phosphate group of the nucleotide. The diphosphate of $NADP^+$ sits between a gap at the ends of β-strands 7 and 8, hydrogen bonding to S217N, S217OG, L219N, S221N, Q222NE2 and K270N. Several members of the AKR family possess a salt bridge covering the diphosphate moiety of the bound nucleotide, known as the "safety belt" (e.g. AKR1A1, Wilson *et al.* 1992). This is not observed in AKR1C3, which has a hydrogen-bonding arrangement instead. The bulk of the adenine ring of $NADP^+$ forms Van der Waals contacts between the R276, L219, L236 and A253 sidechains, whilst the edges of the ring hydrogen bond to the protein – N6 of $NADP^+$ with N280OD1 and Q279NE2, and N7 of $NADP^+$ with N280ND2. Ordered water molecules hydrogen bond with N1 and N3 of the adenine ring.

Due to inconsistencies between the traditional nucleotide labelling system of dehydrogenases and AKR enzymes, the labelling used in the first AKR:nucleotide complex will be used here (Wilson *et al.* 1992). This defines the A side of the nicotinamide ring as the side with atom numbers increasing in a clockwise fashion, and the B side with atom numbers increasing in a counter-clockwise fashion. The nicotinamide ring of $NADP^+$ is positioned above the centre of the barrel, with the A side stacked against the side chain of Y216, and the B side exposed to solvent. This orientation of the cofactor allows substrates to bind over the B face, with the C4-pro-R position available to take part in hydride transfer. The nicotinamide O7 hydrogen bonds to N167ND2, and the N7 hydrogen bonds to both S166OG and Q190OE1.

8.3.4.2.2 Acetate Complexes & Active Site Residues

Despite the difference in crystal packing, loops A, B and C retain the same conformation in both spacegroups. Both crystal structures show the presence of a molecule of acetate in the active site, bound over the $NADP^+$ cofactor, no crystals could be grown in the absence of acetate. Both the acetate complexes have a bound molecule of MPD, which sits at the edge of the active site pocket near the base of loop A. The MPD makes one hydrogen bond to the protein, atom O2 of MPD to S129OG.

Besides the stacking arrangement with the nucleotide, each acetate makes contacts to several of the conserved active site amino acids. Oxygen O of the acetate hydrogen bonds to an ordered water molecule. The other oxygen forms one standard hydrogen bond to H117 NE2 (2.9 Å), and another shorter hydrogen bond to Y55OH (2.5 Å). The methyl group of the acetate molecule points into a hydrophobic pocket comprised of Y24, Y55, L54, W227 and F306.

The residues that form the active site steroid binding pocket (as observed from studies on AKR1C2, see section 8.4.4) are largely hydrophobic. A list of these residues showing the variance between different AKR family members is presented in Table 2.3.

8.3.4.2.3 NSAID Complexes

Both the indomethacin (IMN) and the flufenamic acid (FLA) AKR1C3 complexes show strong, clear electron density for the NSAID compound at the active site (Figures 8.7 and 8.8). The carboxylate of FLA is in approximately the same orientation as the acetate in the acetate complex structures, FLA forming a short hydrogen bond with Y55OH (2.55 Å). The two rings

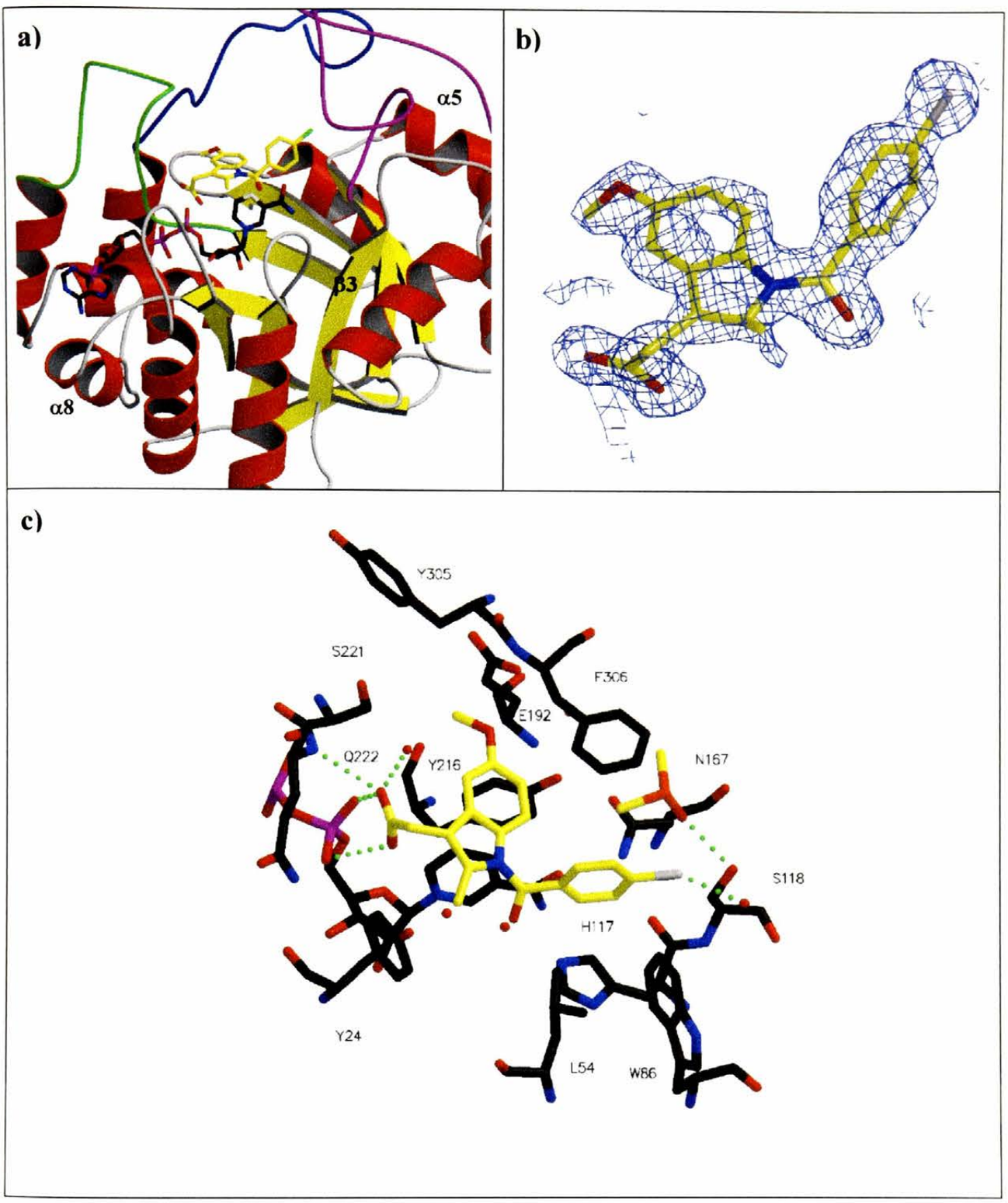


Figure 8.7. Binding of IMN to AKR1C3. (a) Fit of IMN into active site. Ligands in stick form, $NADP^+$ C grey, IMN C yellow. Loop A coloured magenta, loop B green, loop C blue. (b) Electron density map of IMN, 2mFo-DFc, SIGMAA-weighted, contoured at 1σ, 1.7 Å resolution. (c) Contacts of IMN to AKR1C3. Coloured according to atom type : protein C, grey; $NADP^+$ C, black; IMN & DMSO C, yellow; O, red; N, blue; P, magenta; S, orange; Cl, white. Hydrogen bonding shown as dotted green lines. Amino acid numbering using standard single letter code. Water molecules represented as red spheres. Diagrams prepared using MOLSCRIPT (Kraulis 1991), BOBSCRIPT (Esnouf 1997) and RASTER 3D (Meritt and Bacon 1997).

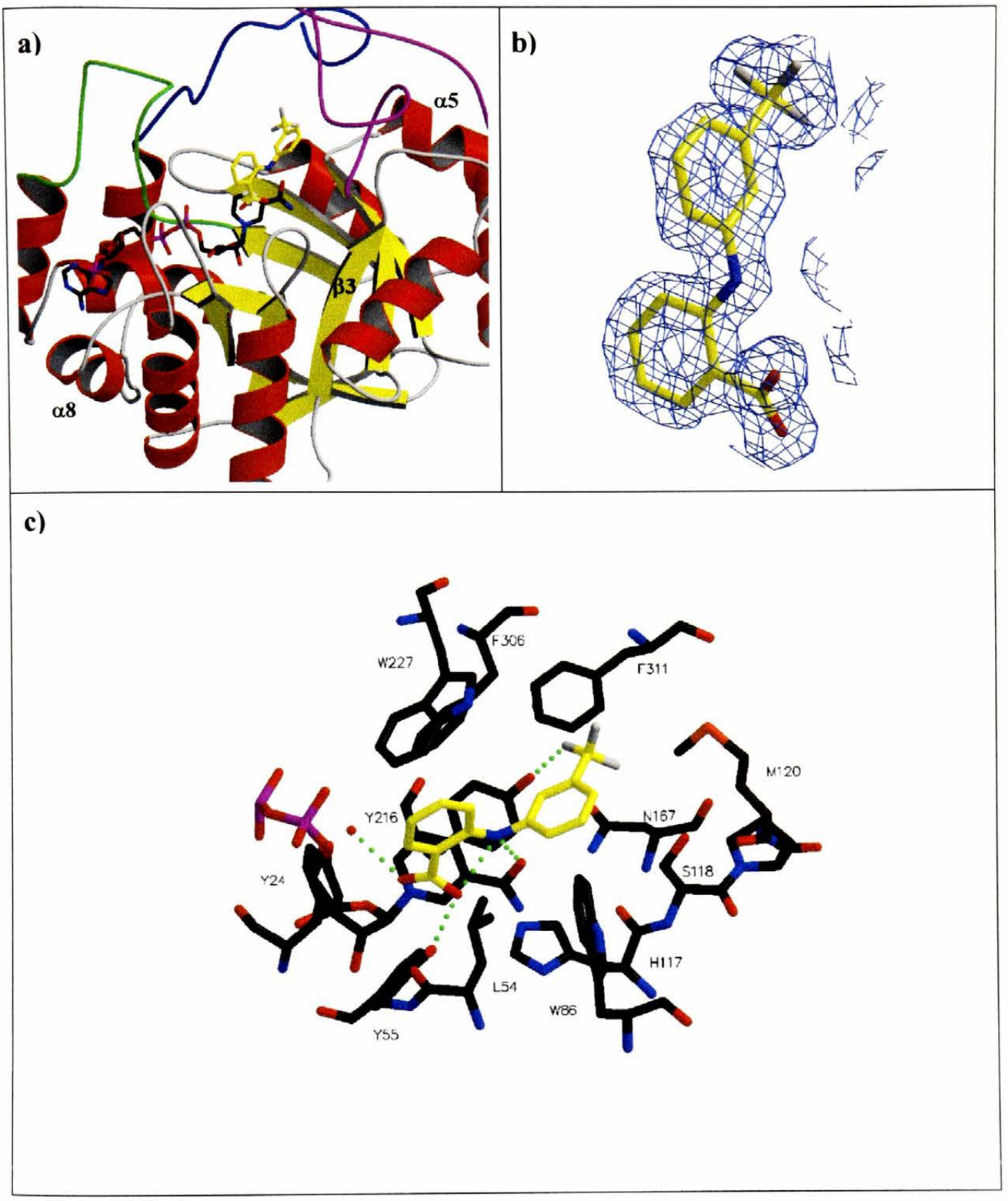


Figure 8.8. Binding of FLA to AKR1C3. (a) Fit of FLA into active site. Ligands in stick form, $NADP^+$ C grey, FLA C yellow. Loop A coloured magenta, loop B green, loop C blue. (b) Electron density map of FLA, 2mFo-DFc, SIGMAA-weighted, contoured at 1σ, 1.8 Å resolution. (c) Contacts of FLA to AKR1C3. Coloured according to atom type : protein C, grey; $NADP^+$ C, black; FLA C, yellow; O, red; N, blue; P, magenta; S, orange; F, white. Hydrogen bonding shown as dashed green lines. Amino acid numbering using standard single letter code. Water molecules represented as red spheres. Diagrams prepared using MOLSCRIPT (Kraulis 1991), BOBSCRIPT (Esnouf 1997) and RASTER 3D (Meritt and Bacon 1997).

of FLA lie at ~70° to each other. The proximal aromatic ring of FLA (the ring with the carboxylate group) sits in a hydrophobic pocket comprised of L54, F306, W227, Y24 and Y55. The secondary amine linkage participates in a bifurcated hydrogen bond between the FLA nitrogen , FLA O1 atom, and the nicotinamide amide oxygen O7. The distal ring of FLA (the ring with the trifluoromethyl group) sits in a more hydrophilic pocket comprised of residues H117, W86, S118, M120, N167 and Y216. The F1 atom of FLA also hydrogen bonds to Y216OH (2.9 Å). With residuesY55 and Y216, FLA forms a "molecular clamp" around the nicotinamide, surrounding it on all sides.

A second, lower occupancy molecule of FLA can be observed near the N-terminal end of the barrel. The FLA sits in a groove formed between residues 6-8 and 258-264. Electron density for this second molecule is better for the proximal ring than for the trifluoromethyl group of the distal ring.

The carboxylate of IMN is in a different orientation than the carboxylate of FLA. Instead of binding near the nicotinamide of the nucleotide, the carboxylate binds next to the diphosphate moiety (in the oxyanion hole), hydrogen bonding to oxygens NO1 and NO2. This mode of binding indicates that the IMN carboxylate group must be protonated – avoiding charge repulsion. Atom O2 of the IMN carboxylate also makes a contact to the protein, hydrogen bonding to Q222 backbone nitrogen. The carbonyl linkage of IMN points in toward the oxyanion hole, with the classical carboxylate oxygen position occupied by a solvent peak (see discussion).

The chlorobenzoyl part of IMN sits between where the two rings of FLA were situated, just out of the plane of the proximal ring. The chlorine atom forms a hydrogen bond with an ordered water molecule. The IMN indole ring system sits orthogonal to the nicotinamide, displacing F306 from its position in the acetate-bound form. The movement of F306 away from the active site causes the corresponding movement of F311, which rotates away to allow the NSAID to bind. The two planes of the IMN rings are roughly perpendicular to one another. An ordered DMSO molecule binds where the distal ring of FLA sat, the oxygen atom of DMSO hydrogen bonding to S118OG. No bound DMSO can be observed in the FLA soak structure.

8.4 Discussion

8.4.1 Acetate Binding & Implications for Catalysis

The interaction of acetate with H117 and Y55 is particularly relevant, with these two residues forming the proposed catalytic tetrad (see Chapter 2) along with K84 and D50. The residues are positioned such that the D50O2 hydrogen bonds to K84NZ, which in turn hydrogen bonds to Y55OH. In the context of comparing the tetrad to the arrangement observed in serine proteases (see Figure 8.10), one of the acetate oxygens occupies the oxyanion hole. This acetate oxygen forms an extremely short hydrogen bond with Y55OH (2.5 Å).

The residue arrangement in serine proteases functions to stabilise any negative charge present at the position of the oxyanion hole. AKR1C3 was crystallized at pH 4.6, equivalent to the pKa of acetate, so 50 % of the free acetate in solution will be charged and the other 50 % will be neutral. Due to the symmetry of the acetate molecule and electron delocalisation effects, it would normally be assumed that each oxygen of a charged acetate molecule would carry half

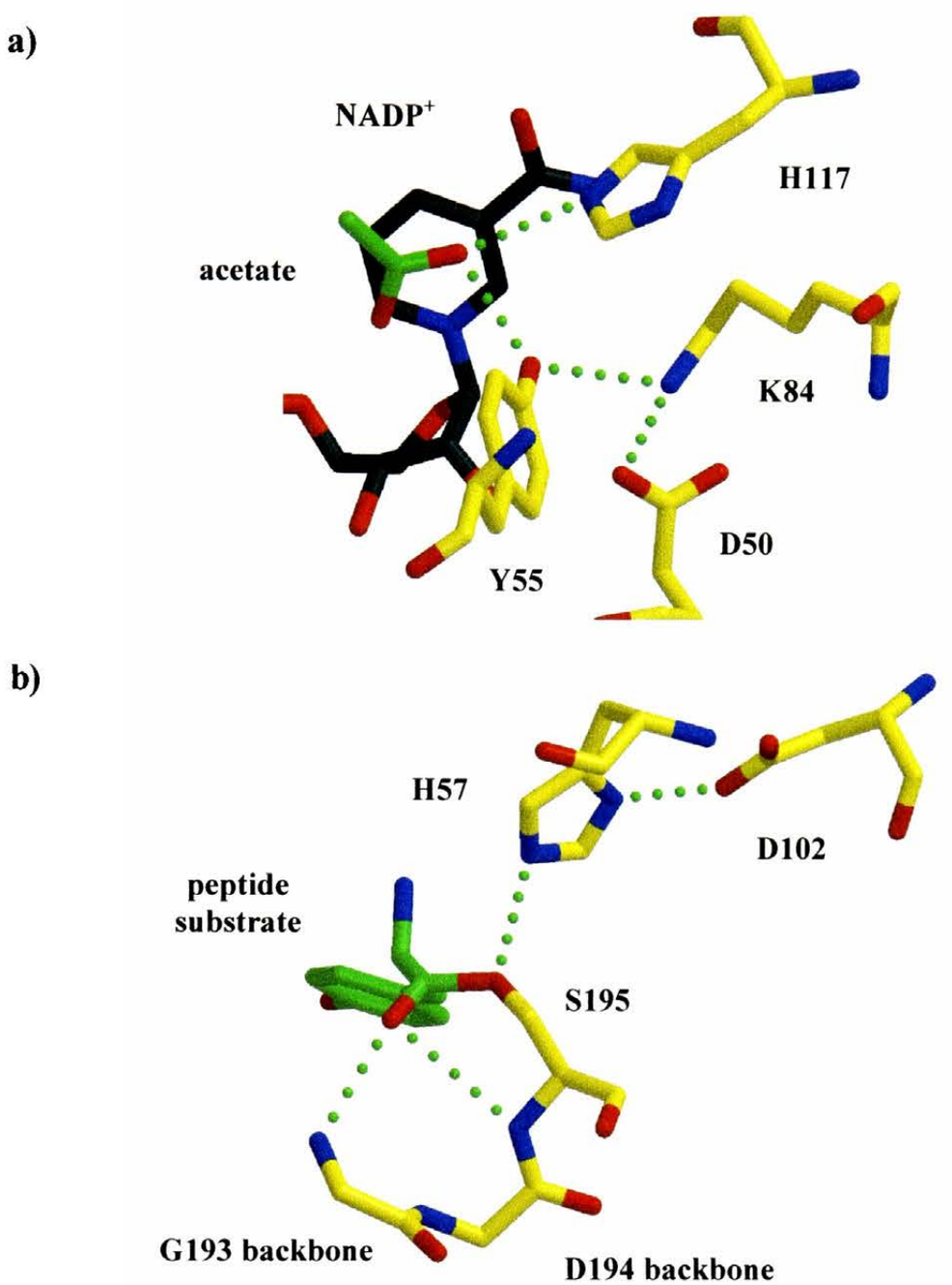


Figure 8.10. Analogy of AKR1C3 active site residues with those of the serine proteases. a) AKR1C3 acetate complex. b) chymotrypsin with bound peptide (PDB code 1AB9, Yennawar *et al.* 1994). Atoms coloured by atom type : protein C, yellow; nucleotide C, grey; substrate C, green; O, red; N, blue. Hydrogen bonding represented by dotted green lines. In the chymotrypsin structure, the peptide ligand (only tyrosine section shown) is covalently bound to the active site serine (Tyr Cα to Ser Oγ). The substrate carbonyl is sat in the oxyanion hole, with two hydrogen bonds to backbone amides. The catalytic serine is positioned by a histidine, itself held in place by an aspartate. With AKR1C3, the catalytic Tyr residue forms part of the oxyanion hole, in association with an active site histidine. The positioning of the catalytic residue is via a lysine : aspartate relay. Although differences are apparent, both function to stabilise the build up of charge on the substrate, and hold a catalytic acid or base in place. Diagram prepared using MOLSCRIPT (Kraulis 1991) and RASTER 3D (Meritt and Bacon 1997).

a formal negative charge. The hydrogen bond distance between acetate and Y55OH is 2.5 Å, far shorter than normal hydrogen bonds and classified as a LBHB (low barrier hydrogen bond, review in Cleland *et al.* 1998). This type of bond is favoured by a charged atom and so it was decided that the acetate molecule in the active site carries a negative charge, entirely localised on the OXT atom.

The observation of acetate in this orientation matches several other structural studies on the AKR family. Nine polypeptide chains (from 8 structures) of AKR members in the Protein Data Bank (Berman *et al.* 2000) possess a compound with a carboxylic acid group bound at the active site. In all of these structures, a carboxylate oxygen is hydrogen bonded to an active site tyrosine and histidine. The group can take one of two orientations, seven of the nine agreeing with the orientation found in the presented AKR1C3 acetate/FLA complexes. A further three family members from the PDB have acetate itself bound. One of these, AKR1C2, binds acetate concurrently with testosterone, displacing the substrate (Nahoum *et al.* 2001). The ubiquity of observing a carboxylate oxygen in this position may indicate that it is acting as an analogue of the oxyanion intermediate when reducing ketones/aldehydes or oxidizing alcohols (see Chapter 2).

The LBHB between acetate and Y55 is only present in the $P2_12_12_1$ complex. This may be attributed to the higher resolution of this crystal form, where the model is more accurate. The overall estimated co-ordinate error is ± 0.037 / 0.034 / 0.028 Å based on R value / FreeR value / maximum likelihood, respectively. Of the AKR family members with a bound compound at the active site, 5 show a LBHB (≤ 2.6 Å) to the tyrosine, 3 show a LBHB to the histidine.

The importance of LBHBs remains highly controversial. At short hydrogen bonding distances (ideally at 2.5 Å), the energy barrier for proton transfer between the two hydrogen bonding atoms is overcome, and the hydrogen is proposed to be able to move freely between them. Such a scenario has been proposed to provide a large part of the catalytic power of oxidoreductase enzymes. A review by Warshel (Warshel 1998) documents the argument for other possible explanations of catalytic power, but it can be agreed that LBHBs provide a significant contribution to substrate or transition state binding.

8.4.2 NSAID Complexes

The electron density maps for both NSAID compounds were excellent and it was easy to model the drug complexes. However, one complication arose when analysing the IMN inhibitor soak. A prospective solvent peak was observed in the oxyanion hole, just 2.0 Å away from the IMN carbonyl oxygen, and 2.6 Å from Y55OH. The latter interaction is indicative of a LBHB but the former is far too short to be any kind of hydrogen bond. Due to the replacement of ammonium acetate with sodium chloride in the NSAID soaks, and the preference of LBHBs for a charged species, it is highly likely that a chloride ion occupies this position. It has also been noted that IMN inhibition (but not FLA inhibition) is sensitive to chloride ions when assaying steroid turnover (personal communication, J.P. Ride).

Although the data for the IMN dataset are good, chlorine does not have enough of an anomalous signal (0.27 e at 0.933 Å) to be able to test the above hypothesis by the construction of an anomalous difference map. Whilst refining the IMN complex structure, a

test refinement using a hydroperoxide derivative of IMN ruled out a covalent linkage of this solvent peak to the NSAID.

Having analysed the amino acid residue shifts of AKR1C3 upon binding the different NSAIDs, it is immediately apparent that more of a change is seen upon IMN binding than that of FLA. In both cases, residues on loops near the solvent exposed end of the active site show more of a shift than those at the bottom of the pocket, near the barrel edge. The largest changes take place on the C-terminal loop, around residues 310-311 (Figure 8.9a). The binding of the IMN indole moiety in the region of F306 causes the sidechain to rotate away from the nicotinamide. The FLA complex imposes no such steric clash and F306 retains the same conformation observed in the acetate complex. The binding of the MPD cryoprotectant molecule in the acetate complexes leads to F311 being bent away from that part of the pocket, towards W227. Although no MPD is present in the FLA soak, F311 in the FLA complex is in much the same position but has a completely different Cα position and ring orientation. This may be to place the ring of F311 orthogonal to F3 of the trifluoromethyl group, much like the interactions observed between methyl groups and aromatic sidechains. Strikingly, F311 of the IMN complex shows no such main chain movement, simply rotating 90° around the Cα-Cβ bond. This rotation allows F306 of the IMN complex to occupy part of the space occupied by F311 in the other structures – this is due to the displacement of F306 from its normal position by the indole of IMN. Indeed, F306 and F311 show the largest positional changes upon NSAID binding, but smaller shifts can be seen in residues such as W227, S310 and W86, all of which line the active site pocket. No large changes are observed in the residues of the catalytic tetrad, small variations in H117 orientation can be seen in both NSAID structures,

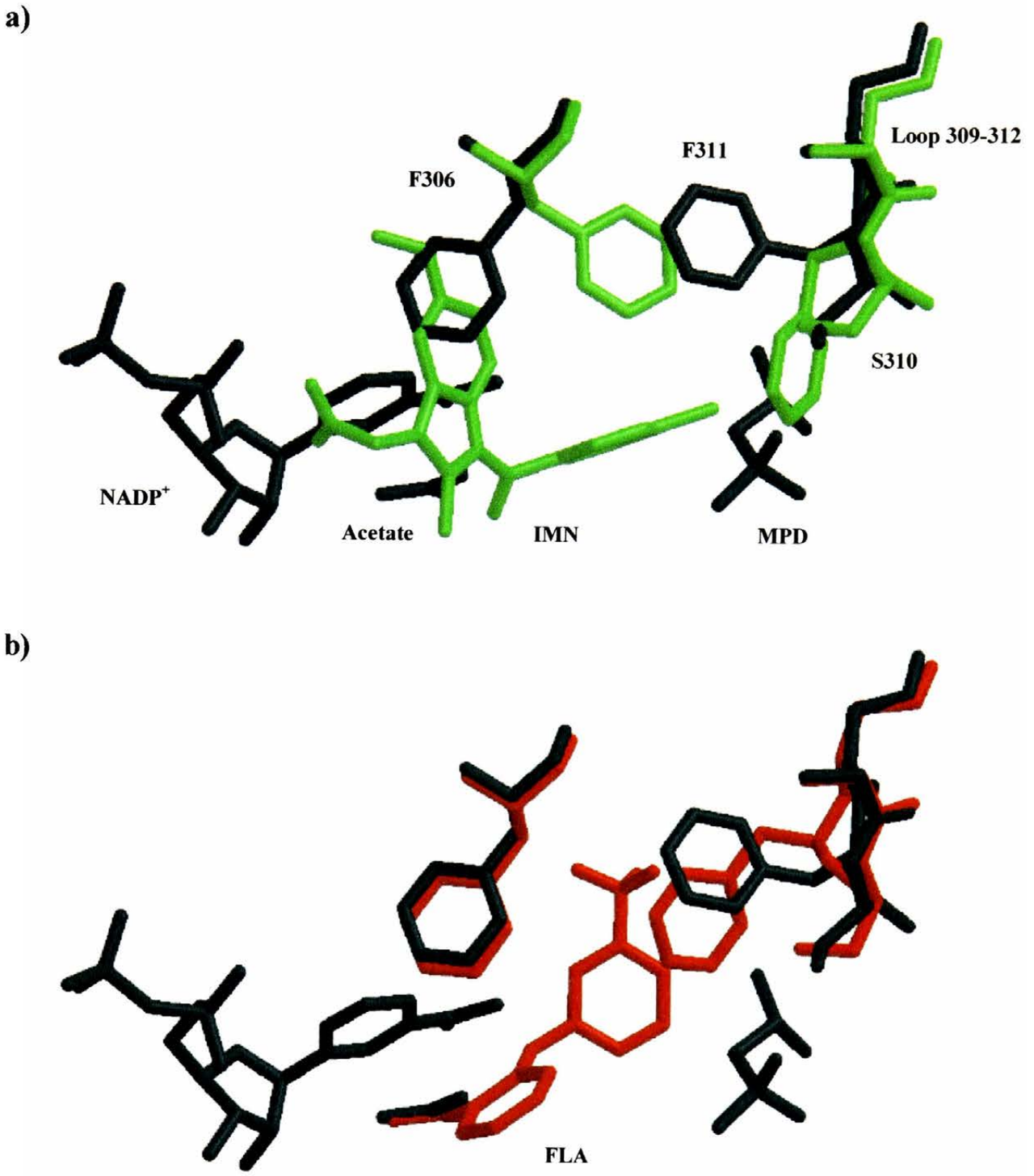


Figure 8.9. Amino acid shifts upon NSAID binding. Acetate complex coloured in grey, indomethacin soak in green (a), and flufenamic acid soak in red (b). Diagrams only show largest amino acid shifts for clarity, full list of changes given in body of text. The indole ring of indomethacin binds where F306 is normally placed, shifting both F306 and F311. The flufenamic acid soak shows a change in orientation of F311, causing a re-arrangement of the C-terminal loop, most apparent between residues 309-312 shown above. Diagram prepared using MOLSCRIPT (Kraulis 1991) and RASTER 3D (Meritt and Bacon 1997).

presumably due to the proximity of this sidechain to the chlorobenzoyl ring of IMN or distal ring of FLA.

The binding mode of both NSAIDs may possibly be influenced by several factors. The FLA carboxylate group has a lower pK_a than that of IMN (3.9 versus 4.5), and so is more likely to be charged at the pH of crystallization (4.6), making it more likely to bind in the oxyanion hole. More experiments testing the pH sensitivity of the AKR1C3 crystals and feasibility of a pH range of NSAID soaks would be required to test this hypothesis in a structural manner. However, no change in the K_i of either NSAID (when varying the pH around this region) is observed in kinetic studies (personal communication, J.P. Ride). Similar kinetic studies do, however, show a 10-20 fold increase in K_i for the NSAID inhibitors when inhibiting reductive compared to oxidative reactions. This may indicate a higher affinity for the enzyme : $NADP^+$ complex than the enzyme : NADPH complex, and even possibly different binding modes.

An equilibrium dialysis study on AKR1C9 inhibition by IMN (Penning *et al.* 1983), revealed a low affinity enzyme:IMN binary complex (K_d = 22 μM) and a different, high affinity enzyme:NAD^+:IMN ternary complex (K_d = 1-2 μM). It could be argued that this may represent two different binding modes of the NSAID, or this may show consistency with the observations presented here that IMN has interactions with the nucleotide as well as the protein.

Regarding the use of these two NSAID complexes to model other NSAID binding to AKR1C3 – the plasticity of the active site residues and the vastly different binding modes of IMN and FLA may show the difficulties inherent in attempting this. It may, however, be a

more simple affair, with subsequent NSAID complex determinations resembling either one or the other. The prospect of there being just these two binding modes is clear cut when comparing sulindac with IMN, or meclofenamic acid with FLA, but becomes complicated with NSAIDs such as peroxicam. Both NSAID complex structures are attractive in terms of drug design – there is plenty of space in the active site to introduce new functional groups, and several bound water molecules exist that could be replaced by new atoms on the NSAID. The large positional variability observed in several of the active site residues may present more of a problem when optimising the fit of the drug to the protein.

8.4.3 Comparison to Cyclo-oxygenase Indomethacin Complexes

Although COX is mechanistically and structurally distinct from the AKR family, it is interesting to compare the reported COX : IMN complexes with the structure obtained in this study. The structure of COX-1 bound to IMN (Loll *et al.* 1996) has a high resolution limit of 4.5 Å, and so a trans model (PDB code 1PGG) and a cis model (PDB code 1PGF) were fitted to the ambiguous electron density. Due to the lack of confidence in the model, it was decided to use the COX-2 : IMN complex (PDB code 4COX, data to 3 Å, Kurumbail *et al.* 1996) in comparisons with AKR1C3. It must however be stressed that at 3 Å some ambiguity in fitting a bound ligand (with limited information in stereochemical restraints) remains.

As no structural homology exists between COX-2 and AKR1C3, all IMN complex comparisons were achieved by overlapping the indole rings of the NSAID molecule. In COX-2, the acid group retains a similar conformation to that seen in AKR1C3, but twisted around the C17-C18 bond by approximately 60°. The major difference between the two IMN complexes involves the cis / trans placement of the carbonyl and chlorobenzoyl groups. With

the COX-2 structure, the carbonyl group is on the other side of the indole plane to the acid, with the chlorobenzoyl group on the same side as the acid group. In AKR1C3 this situation is reversed, with the carbonyl on the same side as the acid and the chlorobenzoyl group on the other side. A smaller difference is observed with the O-methyl part of IMN, pointing in opposite directions in the two structures.

Interactions between IMN and the protein differ markedly between COX-2 and AKR1C3. In general, AKR1C3 has more contacts to the NSAID than COX-2 does. In the COX-2 structure, IMN binds between two helices near the surface of the protein, 8 Å from the heme. Unlike AKR1C3 which had a solvent peak or chloride ion very close to the carbonyl (which itself was in close proximity to active site residues), the carbonyl of IMN in the COX-2 structure has no contacts with the protein or solvent. This may be due to the low resolution of the COX-2 : IMN structural data and poor solvent clarity, or due to different binding environments employed by the two proteins. The acid group of IMN salt bridges to COX-2 R120, different to the uncharged indomethacin carboxyl seen in the AKR1C3 complex. The chlorine of the chlorobenzoyl ring is positioned in a hydrophobic pocket comprised of G526, F381, L384 and M522. There are no other major contacts between IMN and COX-2, just several Van der Waals interactions between the indole ring and the protein.

These differences are somewhat expected due to the vastly different mechanistic schemes of the two proteins. With AKR1C3, IMN binds in the active site, next to the nucleotide and it is easy to see how this would prevent substrate binding. In the case of COX-2 and IMN, the mode of binding is very different. Although IMN binds 8 Å away from the active site heme, it has to be remembered that the arachidonate substrate of the COX enzymes is very long and

flexible. Superimposition of the COX-2 : arachidonate structure (PDB code 1DIY, Malkowski *et al.* 2000) with the COX-2 : IMN complex shows that the two ligands do indeed bind in the same part of the structure. With the large difference in IMN : protein contacts between COX-2 and AKR1C3, the prospect of designing an NSAID selective for one and not the other becomes an attractive prospect.

8.4.4 Comparison to HSD Steroid / Bile Acid Complexes

The "traditional" view of the role of AKR1C3 is in steroid metabolism, catalysing the reduction or oxidation of a wide range of steroids (Penning *et al.* 2001). Several members of the hydroxysteroid dehydrogenase subgroup of the AKR family have been co-crystallized with steroids or steroid-like bile acids. The first, rat liver 3α-HSD (AKR1C9) was co-crystallized with testosterone (PDB code 1AFS, Bennett *et al.* 1997). AKR1C2 was also co-crystallized with testosterone (PDB code 1J96, Nahoum *et al.* 2001), but a molecule of acetate used as a buffer blocks the active site, shifting the steroid out of the oxyanion hole. AKR1C2 has also been complexed with the bile acid ursodeoxycholate in a different study (PDB code 1IHI, Jin *et al.* 2001).

No binary complex (enzyme and nucleotide) has been reported for AKR1C2, but one does exist for AKR1C9 (PDB code 1LWI, Bennett *et al.* 1996). All three steroidal complexes have similar loop orientations. The loop orientation observed in the four structures in this study is different to this, resembling the AKR1C9 binary complex instead. In comparing the HSD enzymes with and without steroid bound, loop A shows the greatest difference, followed by loop C and then loop B. Smaller shifts can be seen in the other shorter loops near the active site. When steroid binds, loop A moves away from the active site pocket to accommodate it,

loop C comes closer in to contact the steroid, and loop B rolls away from loop C slightly. Although acetate is bound in the AKR1C2 testosterone complex, the position of the steroid is similar to that of the four fused rings in the ursodeoxycholate complex, with the bound acetate molecule sat in the same position as the carboxylate of the bile acid. The AKR1C9 testosterone complex is different, bound much further into the active site pocket, the C3 ketone group located in the oxyanion hole. As the two testosterone complexes are different in nature but show a similar loop structure, this indicates perfectly the plasticity of these enzymes in binding substrates in a variety of orientations.

Despite the large loop movements and lack of a full understanding of the different steroid binding modes (for 3α, 17β and 20α metabolism), it is possible to use the steroid complexes to identify residues in AKR1C3 likely to interact with any bound steroids. Comparing AKR1C3 with AKR1C9, residues W227 (invariant), R226 (rat T226), Y24 (T24) and F306 (N306) would all have to move considerably to accommodate any steroid in this pocket. In comparison to the bile acid of AKR1C2, residues W227 (invariant), R226 (P226), Y24 (invariant), L54 (V54), L128 (V128) and P27 (A27) would need to move. Obviously, the differences in amino acid residues may also play a part in steroid accessibility, particularly the bulky residue R226, only present in AKR1C3 of the HSD enzymes. Some consideration must also be made that loops will shift when steroids bind, but the residues involved would still remain in close proximity to the steroid.

Mutational studies have also implicated residues in steroid binding. A Y55F mutant in AKR1C9 (Pawlowski *et al.* 1994) led to a 30-fold increase in K_d for testosterone, and a H117 mutant in aldose reductase led to an 80,000-fold increase in K_m for its substrate (Tarle *et al.*

1993). These two observations indicate that the oxyanion hole and its surrounding residues play a substantial role in substrate binding as well as catalysis. Interestingly, W86 mutants have a moderate effect on steroid binding, but with W227 mutants the effect is more severe (Jez *et al.* 1996b). This situation is reversed with NSAID binding, W86 having a large effect and W227 virtually no effect at all (Penning 1999). This reflects the position of the two residues in the active site pocket, W86 next to the chlorobenzoyl ring of IMN or distal ring of FLA, W227 on the opposite side of the active site, clamped against bound steroid.

For a fuller understanding of the determinants of steroid stereospecificity, it would be necessary to have a hydroxysteroid dehydrogenase series of co-complexes, with 3α, 17β and 20α steroids or steroid mimics. The plasticity of the active site residues permits any meaningful modelling of such complexes, and it may be wise to utilise one or more of the AKR proteins for this purpose. It is intriguing that as well as the binding of NSAIDs and steroids, AKR1C3 also has the versatility to bind prostaglandins. Due to their hydrophobic nature these may indeed bind in much the same way as a steroid, but their increased flexibility may mean much of the molecule is disordered in an electron density map if the contacts to protein are not strong (given that crystals could be obtained). If the co-crystallization of such short-lived labile species as prostaglandins is a problem, and they are too large to soak into the active site, small analogues of the prostaglandin headgroups may have to be used. No complex of any protein with a prostaglandin is available in the PDB at present. Any complex of AKR1C3 with a prostaglandin would also hopefully reveal why this enzyme is unique amongst the AKR1C subfamily in its role of prostaglandin metabolism. Simple sequence alignments provide little clue, apart perhaps from residue S118 (F118 in all the other members of the subfamily, also a serine in bovine prostaglandin synthase).

Publications

Lovering *et al.* (2001) The Structure of Escherichia coli Nitroreductase Complexed with Nicotinic Acid : Three Crystal Forms at 1.7 Å, 1.8 Å and 2.4 Å resolution *J.Mol.Biol* **309**,203-213

Lovering *et al.* (2003) Inhibition of human aldo-keto reductase AKR1C3 by nonsteroidal anti-inflammatory drugs: crystal structures of binary and ternary complexes – manuscript to be submitted

Race, P. R., Lovering, A. L., Ossor, A., White, S. A., Searle, P. F., Wrighton, C. J. and Hyde, E. I. (2003) Structural and mechanistic studies of *E. coli* Nitroreductase with the antibiotic nitrofurazone: Reversed binding orientations in different redox states of the enzyme. – manuscript to be submitted

Grove, J.I., Lovering, A.L., Guise, C., Race, P.R., Wrighton, C.J., White, S.A., Hyde, E.I., Searle, P.F. (2003) Generation of Escherichia coli Nitroreuctase Mutants Conferring Improved Cell Sensitisation to the Prodrug CB1954 – submitted

Patents

International Patent No PCT/GB2002/003833 – Improved nitroreductase enzymes

References

Alekshun, M. N. and S. B. Levy (1997). "Regulation of chromosomally mediated multiple antibiotic resistance: the mar regulon." Antimicrobial Agents and Chemotherapy **41**(10): 2067-2075.

Anlezark, G. M., R. G. Melton, et al. (1992). "The Bioactivation of 5-(Aziridin-1-Yl)-2,4-Dinitrobenzamide (Cb1954) .1. Purification and Properties of a Nitroreductase Enzyme from *Escherichia coli* - a Potential Enzyme for Antibody- Directed Enzyme Prodrug Therapy (Adept)." Biochemical Pharmacology **44**(12): 2289-2295.

Asnis, R. E. (1975). "The reduction of furacin by cell-free extracts of furacin-resistant and parent-susceptible stains of *Escherichia coli*." Archives of Biochemistry and Biophysics(**66**): 208-216.

Banner, D. W., A. C. Bloomer, et al. (1976). "Atomic co-ordinates for triose phosphate isomerase from chicken muscle." Biochemical and Biophysical Research Communications(**72**): 146.

Barber, M. J., P. J. Neame, et al. (1992). "Correlation of X-Ray Deduced and Experimental Amino-Acid- Sequences of Trimethylamine Dehydrogenase." Journal of Biological Chemistry **267**(10): 6611-6619.

Barbosa, T. M. and S. B. Levy (2002). "Activation of the *Escherichia coli* nfnB gene by MarA through a highly divergent marbox in a class II promoter." Molecular Microbiology **45**(1): 191-202.

Barna, T. M., H. Khan, et al. (2001). "Crystal structure of pentaerythritol tetranitrate reductase: "Flipped" binding geometries for steroid substrates in different redox states of the enzyme." Journal of Molecular Biology **310**(2): 433-447.

Benach, J., S. Atrian, et al. (1999). "The catalytic reaction and inhibition mechanism of Drosophila alcohol dehydrogenase: Observation of an enzyme-bound NAD- ketone adduct at 1.4 angstrom resolution by X-ray crystallography." Journal of Molecular Biology **289**(2): 335-355.

Bennett, M. J., B. P. Schlegel, et al. (1996). "Structure of 3 alpha-hydroxysteroid/dihydrodiol dehydrogenase complexed with NADP." Biochemistry **35**(33): 10702-10711.

Bennett, M. J., R. H. Albert, et al. (1997). "Steroid recognition and regulation of hormone action: Crystal structure of testosterone and NADP(+) bound to 3 alpha-hydroxysteroid dihydrodiol dehydrogenase." Structure **5**(6): 799-812.

Berman, H. M., J. Westbrook, et al. (2000). "The Protein Data Bank." Nucleic Acids Research **28**(1): 235-242.

Bernal, J. D. and D. Crowfoot (1934). "X-ray photographs of crystalline pepsin." Nature(**794**): 133-134.

Blattner, F. R., G. Plunkett, et al. (1997). "The complete genome sequence of *Escherichia coli* K-12." Science **277**(5331): 1453.

Blundell, T. and L. N. Johnson (1976). London, Academic Press.

Boland, M. P., R. J. Knox, et al. (1991). "The Differences in Kinetics of Rat and Human Dt Diaphorase Result in a Differential Sensitivity of Derived Cell-Lines to CB1954 (5-(Aziridin-1-Yl)-2,4-Dinitrobenzamide)." Biochemical Pharmacology **41**(6-7): 867-875.

Bradford, M. M. (1976). "A rapid and sensitive method for the quantitation of microgram quantities of protein utilising the principle of protein-dye binding." Analytical Biochemistry(**72**): 248-254.

Bridgewater, J. A., R. J. Knox, et al. (1997). "The bystander effect of the nitroreductase CB 1954 enzyme prodrug system is due to a cell-permeable metabolite." Human Gene Therapy **8**(6): 709-717.

Brunger, A. T. (1990). "Extension of Molecular Replacement - a New Search Strategy Based on Patterson Correlation Refinement." Acta Crystallographica Section A **46**: 46-57.

Brunger, A. T., P. D. Adams, et al. (1998). "Crystallography & NMR system: A new software suite for macromolecular structure determination." Acta Crystallographica Section D-Biological Crystallography **54**: 905-921.

Bryant, D. W., D. R. McCalla, et al. (1981). "Type-I Nitroreductases of *Escherichia coli*." Canadian Journal of Microbiology **27**(1): 81-86.

Bryant, C., L. Hubbard, et al. (1991a). "Cloning, Nucleotide-Sequence, and Expression of the Nitroreductase Gene from *Enterobacter cloacae*." Journal of Biological Chemistry **266**(7): 4126-4130.

Bryant, C. and M. Deluca (1991b). "Purification and Characterization of an Oxygen-Insensitive NAD(P)H Nitroreductase from *Enterobacter cloacae*." Journal of Biological Chemistry **266**(7): 4119-4125.

Bunce, C. M., P. J. French, et al. (1994). "Indomethacin Potentiates the Induction of HL-60 Differentiation to Neutrophils, by Retinoic Acid and Granulocyte-Colony- Stimulating Factor, and to Monocytes, by Vitamin-D-3." Leukemia **8**(4): 595-604.

Bunce, C. M., J. C. Mountford, et al. (1996). "Potentiation of myeloid differentiation by anti-inflammatory agents, by steroids and by retinoic acid involves a single intracellular target, probably an enzyme of the aldoketoreductase family." Biochimica et Biophysica Acta-Molecular Cell Research **1311**(3): 189-198.

Carter, C. W. J. and C. W. Carter (1979). "Protein crystallisation using incomplete factorial experiments." Journal of Biological Chemistry(**254**): 12219-12223.

CCP4 (1994). "Collaborative Computational Project No.4 - The CCP4 Suite : programs for protein Crystallography." Acta Crystallographica Section D-Biological Crystallography(**50**): 760-763.

Clay, C. E., A. M. Namen, et al. (2000). "15-deoxy-Delta(12,14)PGJ(2) induces diverse biological responses via PPAR gamma activation in cancer cells." Prostaglandins & Other Lipid Mediators **62**(1): 23-32.

Cleland, W. W., P. A. Frey, et al. (1998). "The low barrier hydrogen bond in enzymatic catalysis." Journal of Biological Chemistry **273**(40): 25529-25532.

Cobb, C. M., T. Connors, et al. (1969). "2,4-dinitro-5-ethyleneiminobenzamide (CB1954): A potent and selective inhibitor of the growth of the Walker carcinoma 256." Biochemical Pharmacology(**18**): 1519-1527.

Connors, T., J. A. Hickman, et al. (1975). Biochemical Pharmacology(**24**): 1665.

Cowtan, K. (1994). Joint CCP4 and ESF-EACBM newsletter on protein crystallography. **31:** 34-38.

Dauter, Z., M. Dauter, et al. (2002). "Jolly SAD." Acta Crystallographica Section D-Biological Crystallography **58**: 494-506.

Denny, W. A. and W. R. Wilson (1993). "Bioreducible Mustards - a Paradigm for Hypoxia-Selective Prodrugs of Diffusible Cytotoxins (HPDCs)." Cancer and Metastasis Reviews **12**(2): 135-151.

Denny, W. A. (2001). "Prodrug strategies in cancer therapy." European Journal of Medicinal Chemistry **36**(7-8): 577-595.

Dewitt, D. and W. L. Smith (1995). "Yes, But Do They Still Get Headaches?" Cell **83**(3): 345-348.

Dodd, M. C. and W. B. Stillman (1944). "The *in vitro* bacteriostatic action of some simple furan derivatives." J. Pharmacol. Expertl. Therap.(**82**): 11-18.

Dodson, G. and A. Wlodawer (1998). "Catalytic triads and their relatives." Trends in Biochemical Sciences **23**(9): 347-352.

DuBois, R. N., J. Y. Shao, et al. (1996). "G(1) delay in cells overexpressing prostaglandin endoperoxide synthase-2." Cancer Research **56**(4): 733-737.

Elish, M. E., J. R. Pierce, et al. (1988). "Biochemical analysis of spontaneous frpA mutants of *Escherichia coli*." Journal of General Microbiology(**134**): 1355-1364.

Engh, R. A. and R. Huber (1991). "Accurate Bond and Angle Parameters for X-Ray Protein-Structure Refinement." Acta Crystallographica Section A **47**: 392-400.

Enroth, C., B. T. Eger, et al. (2000). "Crystal structures of bovine milk xanthine dehydrogenase and xanthine oxidase: Structure-based mechanism of conversion." Proceedings of the National Academy of Sciences of the United States of America **97**(20): 10723-10728.

Esnouf, R. M. (1997). "An extensively modified version of MolScript that includes greatly enhanced coloring capabilities." Journal of Molecular Graphics & Modelling **15**(2): 132.

Evans, G. and G. Bricogne (2002). "Triiodide derivatization and combinatorial counter-ion replacement: two methods for enhancing phasing signal using laboratory Cu K alpha X-ray equipment." Acta Crystallographica Section D-Biological Crystallography **58**: 976-991.

Evans, P. R. (1993). Data reduction proceedings of CCP4 study weekend, on data collection and processing**:** 114-122.

Ferredamare, A. R. and S. K. Burley (1994). "Use of Dynamic Light-Scattering to Assess Crystallizability of Macromolecules and Macromolecular Assemblies (Vol 2, Pg 357, 1994)." Structure **2**(6): 567-567.

Flower, R. J., S. Moncada, et al. (1980). Analgesic-antipyretics and anti-inflammatory agents: Drugs employed in the treatment of gout. The Pharmacological Basis of Therapeutics. A. G. Gilman, L. S. Goodman and A. Gilman. New York, Macmillan**:** 682-698.

Fosslien, E. (2000). "Biochemistry of cyclooxygenase (COX)-2 inhibitors and molecular pathology of COX-2 in neoplasia." Critical Reviews in Clinical Laboratory Sciences **37**(5): 431-502.

Fraaije, M. W. and A. Mattevi (2000). "Flavoenzymes: diverse catalysts with recurrent features." Trends in Biochemical Sciences **25**(3): 126-132.

GCG (1998). Wisconsin Package Version 10.2. Madison, Wisc, Genetics Computer Group (GCG).

Ghisla, S. and V. Massey (1989). "Mechanisms of flavoprotein-catalyzed reactions." The European J ournal of Biochemistry **181**(1): 1-17.

Gormley, G. J. (1991). "Role of 5-Alpha-Reductase Inhibitors in the Treatment of Advanced Prostatic-Carcinoma." Urologic Clinics of North America **18**(1): 93-98.

Grosse-Kunstleve, R. W. and A. T. Brunger (1999). "A highly automated heavy-atom search procedure for macromolecular structures." Acta Crystallographica Section D-Biological Crystallography **55**: 1568-1577.

Grove, J. I., A. L. Lovering, et al. (2003). "Generation of Escherichia coli nitroreductase mutants conferring improved cell sensitisation to the prodrug CB1954." Submitted.

Guay, D. R. (2001). "An update on the role of nitrofurans in the management of urinary tract infections." Drugs **61**(3): 353-364.

Guex, N. and M. C. Peitsch (1997). "SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling." Electrophoresis **18**(15): 2714-2723.

Hanif, R., A. Pittas, et al. (1996). "Effects of nonsteroidal anti-inflammatory drugs on proliferation and on induction of apoptosis in colon cancer cells by a prostaglandin-independent pathway." Biochemical Pharmacology **52**(2): 237-245.

Hannink, N., S. J. Rosser, et al. (2001). "Phytodetoxification of TNT by transgenic plants expressing a bacterial nitroreductase." Nature Biotechnology **19**(12): 1168-1172.

Haynes, C. A., R. L. Koder, et al. (2002). "Structures of nitroreductase in three states - Effects of inhibitor binding and reduction." Journal of Biological Chemistry **277**(13): 11513-11520.

Hecht, H. J., H. Erdmann, et al. (1995). "Crystal Structure of Nadh Oxidase from *Thermus thermophilus*." Nature Structural Biology **2**(12): 1109-1114.

Hoog, S. S., J. E. Pawlowski, et al. (1994). "3-Dimensional Structure of Rat-Liver 3-Alpha-Hydroxysteroid Dihydrodiol Dehydrogenase - a Member of the Aldo-Keto Reductase Superfamily." Proceedings of the National Academy of Sciences of the United States of America **91**(7): 2517-2521.

Jancarik, J. and S. H. Kim (1991). "Sparse-Matrix Sampling - a Screening Method for Crystallization of Proteins." Journal of Applied Crystallography **24**: 409-411.

Jez, J. M., T. G. Flynn, et al. (1996a). Enzymology and Molecular Biology of Carbonyl Metabolism. H. Weiner, R. Lindahl, D. W. Crabb and T. G. Flynn. New York, Plenum Press. **6:** 579-589.

Jez, J. M., B. P. Schlegel, et al. (1996b). "Characterization of the substrate binding site in rat liver 3 alpha-hydroxysteroid dihydrodiol dehydrogenase." Journal of Biological Chemistry **271**(47): 30190-30198.

Jez, J. M., M. J. Bennett, et al. (1997). "Comparative anatomy of the aldo-keto reductase superfamily." Biochemical Journal **326**: 625-636.

Jin, Y., S. E. Stayrook, et al. (2001). "Crystal structure of human type III 3 alpha-hydroxysteroid dehydrogenase/bile acid binding protein complexed with NADP(+) and ursodeoxycholate." Biochemistry **40**(34): 10161-10168.

Kavanagh, K. L., M. Klimacek, et al. (2002). "The structure of apo and holo forms of xylose reductase, a dimeric aldo-keto reductase from *Candida tenuis*." Biochemistry **41**(28): 8785-8795.

Kissinger, C. R., D. K. Gehlhaar, et al. (1999). "Rapid automated molecular replacement by evolutionary search." Acta Crystallographica Section D-Biological Crystallography **55**: 484-491.

Klampfer, L., J. Cammenga, et al. (1999). "Sodium salicylate activates caspases and induces apoptosis of myeloid leukemia cell lines." Blood **93**(7): 2386-2394.

Kleywegt, G. J. and T. A. Jones (1994). Halloween.....Masks and Bones. From First Map to Final Model. S. Bailey, R. Hubbard and D. Waller. Daresbury, Warrington, SERC Laboratory: 59-66.

Knox, R. J., M. P. Boland, et al. (1988). "The Nitroreductase Enzyme in Walker Cells That Activates 5- (Aziridin-1-Yl)-2,4-Dinitrobenzamide (Cb-1954) to 5-(Aziridin- 1-Yl)-4-Hydroxylamino-2-Nitrobenzamide Is a Form of NAD(P)H Dehydrogenase (Quinone) (E.C.1.6.99.2)." Biochemical Pharmacology **37**(24): 4671-4677.

Knox, R. J., F. Friedlos, et al. (1991). "Bioactivation of CB1954 - Reaction of the Active 4-Hydroxylamino Derivative with Thioesters to Form the Ultimate DNA DNA Interstrand Cross-Linking Species." Biochemical Pharmacology **42**(9): 1691-1697.

Knox, R. J., F. Friedlos, et al. (1992). "The Bioactivation of 5-(Aziridin-1-Yl)-2,4-Dinitrobenzamide (CB1954) .2. A Comparison of an *Escherichia coli* Nitroreductase and Walker DT Diaphorase." Biochemical Pharmacology **44**(12): 2297-2301.

Knox, R. J., F. Friedlos, et al. (1993). "The Bioactivation of CB1954 and Its Use as a Prodrug in Antibody-Directed Enzyme Prodrug Therapy (Adept)." Cancer and Metastasis Reviews **12**(2): 195-212.

Knox, R. J., T. C. Jenkins, et al. (2000). "Bioactivation of 5-(aziridin-1-yl)-2,4-dinitrobenzamide (CB 1954) by human NAD(P)H quinone oxidoreductase 2: A novel co-substrate-mediated antitumor prodrug therapy." Cancer Research **60**(15): 4179-4186.

Kobori, T., H. Sasaki, et al. (2001). "Structure and site-directed mutagenesis of a flavoprotein from *Escherichia coli* that reduces nitrocompounds - Alteration of pyridine nucleotide binding by a single amino acid substitution." Journal of Biological Chemistry **276**(4): 2816-2823.

Koike, H., H. Sasaki, et al. (1998). "1.8 Å crystal structure of the major NAD(P)H : FMN oxidoreductase of a bioluminescent bacterium, *Vibrio fischeri*: Overall structure, cofactor and substrate-analog binding, and comparison with related flavoproteins." Journal of Molecular Biology **280**(2): 259-273.

Kozma, E., E. Brown, et al. (2002). "The crystal structure of rat liver AKR7A1 - A dimeric member of the aldo-keto reductase superfamily." Journal of Biological Chemistry **277**(18): 16285-16293.

Kraulis, P. J. (1991). "Molscript - a Program to Produce Both Detailed and Schematic Plots of Protein Structures." Journal of Applied Crystallography **24**: 946-950.

Kurumbail, R. G., A. M. Stevens, et al. (1996). "Structural basis for selective inhibition of cyclooxygenase-2 by anti-inflammatory agents." Nature **384**(6610): 644-648.

Land, K. M. and P. J. Johnson (1999). "Molecular basis of metronidazole resistance in pathogenic bacteria and protozoa." Drug Resistance Updates **2**(5): 289-294.

Laskowski, R. A., M. W. Macarthur, et al. (1993). "Procheck - a Program to Check the Stereochemical Quality of Protein Structures." Journal of Applied Crystallography **26**: 283-291.

Lei, B. F., M. Y. Liu, et al. (1994). "*Vibrio harveyi* NADPH-Flavin Oxidoreductase - Cloning, Sequencing and Overexpression of the Gene and Purification and Characterization of the Cloned Enzyme." Journal of Bacteriology **176**(12): 3552-3558.

Leslie, A. G. W. (1992). Joint CCP4 and ESF-EACBM newsletter on protein crystallography. **26**.

Liochev, S. I., A. Hausladen, et al. (1999). "Nitroreductase A is regulated as a member of the soxRS regulon of *Escherichia coli*." Proceedings of the National Academy of Sciences of the United States of America **96**(7): 3537-3539.

Loll, P. J., D. Picot, et al. (1996). "Synthesis and use of iodinated nonsteroidal antiinflammatory drug analogs as crystallographic probes of the prostaglandin H- 2 synthase cyclooxygenase active site." Biochemistry **35**(23): 7330-7340.

Lovering, A. L., E. I. Hyde, et al. (2001). "The structure of *Escherichia coli* nitroreductase complexed with nicotinic acid: Three crystal forms at 1.7 Å, 1.8 Å and 2.4 Å resolution." Journal of Molecular Biology **309**(1): 203-213.

Luong, C., A. Miller, et al. (1996). "Flexibility of the NSAID binding site in the structure of human cyclooxygenase-2." Nature Structural Biology **3**(11): 927-933.

Malkowski, M. G., S. L. Ginell, et al. (2000). "The productive conformation of arachidonic acid bound to prostaglandin synthase." Science **289**(5486): 1933-1937.

Massey, V. and S. Ghisla (1983). Biological oxidations. H. Sund and V. Ullrich. Berlin, Springer: 114-139.

Massey, V. (1995). "Flavoprotein Structure and Mechanism - Introduction." Faseb Journal **9**(7): 473-475.

McCalla, D. R., A. Reuvers, et al. (1970). "Mode of action of nitrofurazone." Journal of Bacteriology(**104**): 1126-1134.

McCalla, D. R., A. Reuvers, et al. (1971). "Breakage of bacterial DNA by nitrofuran derivatives." Cancer Research(**31**): 2182-2188.

McCalla, D. R., P. Olive, et al. (1975). "Nitrofurazone-reducing enzymes in *E. coli* and their role in drug activation in vivo." Canadian Journal of Microbiology(**21**): 1485-1491.

McCalla, D. R., C. Kaiser, et al. (1978). "Genetics of nitrofurazone resistance in *Escherichia coli*." Journal of Bacteriology(**133**): 10-16.

McCartney, M. A., B. F. Chatterjee, et al. (1986). "Airplane Emissions - a Source of Mutagenic Nitrated Polycyclic Aromatic-Hydrocarbons." Mutation Research **171**(2-3): 99-104.

McCormack, T. and K. McCormack (1994). "Shaker K+ Channel Beta-Subunits Belong to an NAD(P)H-Dependent Oxidoreductase Superfamily." Cell **79**(7): 1133-1135.

McPherson, A. (1999). Crystallization of Biological Macromolecules. New York, Cold Spring Harbor Laboratory Press.

Meritt, E. A. and D. J. Bacon (1997). "Raster 3D photorealistic molecular graphics." Methods In Enzymology(**277**): 505-524.

Mewies, M., W. S. McIntire, et al. (1998). "Covalent attachment of flavin adenine dinucleotide (FAD) and flavin mononucleotide (FMN) to enzymes: The current state of affairs." Protein Science 7(1): 7-20.

Moolten, F. L. (1986). "Tumor Chemosensitivity Conferred by Inserted Herpes Thymidine Kinase Genes - Paradigm for a Prospective Cancer Control Strategy." Cancer Research **46**(10): 5276-5281.

Morita, I., M. Schindler, et al. (1995). "Different Intracellular Locations for Prostaglandin Endoperoxide-H Synthase-1 and Synthase-2." Journal of Biological Chemistry **270**(18): 10902-10908.

Morris, R. J., A. Perrakis, et al. (2002). "ARP/wARP's model-building algorithms. I. The main chain." Acta Crystallographica Section D-Biological Crystallography **58**: 968-975.

Murshudov, G. N., A. A. Vagin, et al. (1997). "Refinement of macromolecular structures by the maximum- likelihood method." Acta Crystallographica Section D-Biological Crystallography **53**: 240-255.

Nahoum, V., A. Gangloff, et al. (2001). "Structure of the human 3 alpha-hydroxysteroid dehydrogenase type 3 in complex with testosterone and NADP at 1.25 Å resolution." Journal of Biological Chemistry **276**(45): 42091-42098.

Narhi, L. O. and A. J. Fulco (1986). "Characterization of a Catalytically Self-Sufficient 119,000- Dalton Cytochrome-P-450 Monooxygenase Induced by Barbiturates in *Bacillus megaterium*." Journal of Biological Chemistry **261**(16): 7160-7169.

Nettelbeck, D. M., V. Jerome, et al. (2000). "Gene therapy - designer promoters for tumour targeting." Trends in Genetics **16**(4): 174-181.

Nishino, S. F., G. C. Paoli, et al. (2000). "Aerobic degradation of dinitrotoluenes and pathway for bacterial degradation of 2,6-dinitrotoluene." Applied and Environmental Microbiology **66**(5): 2139-2147.

Ohta, T., A. Elnemr, et al. (2002). "Thiazolidinedione, a peroxisome proliferator-activated receptor-gamma ligand, modulates the E-cadherin/beta-catenin system in a human pancreatic cancer cell line, BxPC-3." International Journal of Oncology **21**(1): 37-42.

Oneill, G. P. and A. W. Fordhutchinson (1993). "Expression of Messenger-RNA for Cyclooxygenase-1 and Cyclooxygenase-2 in Human Tissues." Febs Letters **330**(2): 156-160.

Oshima, M., J. E. Dinchuk, et al. (1996). "Suppression of intestinal polyposis in Apc(Delta 716) knockout mice by inhibition of cyclooxygenase 2 (COX-2)." Cell **87**(5): 803-809.

Otwinowski, Z. and W. Minor (1997). "Processing of X-ray diffrcation data collected in oscillation mode." Methods In Enzymology **276**: 307-326.

Pai, E. F. and G. E. Schulz (1983). "The Catalytic Mechanism of Glutathione-Reductase as Derived from X-Ray-Diffraction Analyses of Reaction Intermediates." Journal of Biological Chemistry **258**(3): 1752-1757.

Palackal, N. T., S. H. Lee, et al. (2002). "Activation of polycyclic aromatic hydrocarbon trans-dihydrodiol proximate carcinogens by human aldo-keto reductase (AKR1C) enzymes and their functional overexpression in human lung carcinoma (A549) cells." Journal of Biological Chemistry **277**(27): 24799-24808.

Palmer, C., G. Chung-Faye, et al. (2000). "A phase I dose escalation trial and pharmacokinetics of the GDEPT prodrug CB1954." British Journal of Cancer **83**: 169.

Parkinson, G. N., J. V. Skelly, et al. (2000). "Crystal structure of FMN-dependent nitroreductase from *Escherichia coli* B: A prodrug-activating enzyme." Journal of Medicinal Chemistry **43**(20): 3624-3631.

Pawlowski, J. E. and T. M. Penning (1994). "Overexpression and Mutagenesis of the Cdna for Rat-Liver 3- Alpha-Hydroxysteroid Dihydrodiol Dehydrogenase - Role of Cysteines and Tyrosines in Catalysis." Journal of Biological Chemistry **269**(18): 13502-13510.

Penning, T. M. and P. Talalay (1983). "Inhibition of a Major NAD(P)-Linked Oxidoreductase from Rat- Liver Cytosol by Steroidal and Non-Steroidal Anti-Inflammatory Agents and by Prostaglandins." Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences **80**(14): 4504-4508.

Penning, T. M. (1999). "Molecular determinants of steroid recognition and catalysis in aldo-keto reductases. Lessons from 3 alpha-hydroxysteroid dehydrogenase." Journal of Steroid Biochemistry and Molecular Biology **69**(1-6): 211-225.

Penning, T. M., M. E. Burczynski, et al. (2000). "Human 3 alpha-hydroxysteroid dehydrogenase isoforms (AKR1C1- AKR1C4) of the aldo-keto reductase superfamily: functional plasticity and tissue distribution reveals roles in the inactivation and formation of male and female sex hormones." Biochemical Journal **351**: 67-77.

Penning, T. M., M. E. Burczynski, et al. (2001). "Structure-function aspects and inhibitor design of type 5 17 beta-hydroxysteroid dehydrogenase (AKR1C3)." Molecular and Cellular Endocrinology **171**(1-2): 137-149.

Petrash, J. M., I. Tarle, et al. (1994). "Aldose Reductase Catalysis and Crystallography - Insights from Recent Advances in Enzyme Structure and Function." Diabetes **43**(8): 955-959.

Phillips, T. A., R. A. VanBogelen, et al. (1984). Journal of Bacteriology(**159**): 283-287.

Picot, D., P. J. Loll, et al. (1994). "The X-Ray Crystal-Structure of the Membrane-Protein Prostaglandin-H(2) Synthase-1." Nature **367**(6460): 243-249.

Prescott, S. M. and F. A. Fitzpatrick (2000). "Cyclooxygellase-2 and carcinogenesis." Biochimica Et Biophysica Acta-Reviews on Cancer **1470**(2): M69-M78.

Rao, K. S. and F. Lederer (1998). "About the pKa of the active-site histidine in flavocytochrome b(2) (yeast L-lactate dehydrogenase)." Protein Science **7**(7): 1531-1537.

Read, R. J. (1986). "Improved Fourier Coefficients for Maps Using Phases from Partial Structures with Errors." Acta Crystallographica Section A **42**: 140-149.

Ries-Kautt, M. (2001). An alternative to sparse-matrix screens. Protein Crystallization. T. M. Bergfors. La Jolla, CA, IUL: 91-109.

Robert, A. (1983). "Cytoprotection of the Gastrointestinal Mucosa." Advances in Internal Medicine **28**: 325-337.

Roberts, J. J., F. Friedlos, et al. (1986). "CB1954 (2,4-Dinitro-5-Aziridinyl Benzamide) Becomes a DNA Interstrand Cross-Linking Agent in Walker Tumor-Cells." Biochemical and Biophysical Research Communications **140**(3): 1073-1078.

Rondeau, J. M., F. Tetefavier, et al. (1992). "Novel NADPH-Binding Domain Revealed by the Crystal Structure of Aldose Reductase." Nature **355**(6359): 469-472.

Rosenkranz, H. S., E. C. McCoy, et al. (1981). "A Cautionary Note on the Use of Nitroreductase-Deficient Strains of *Salmonella typhimurium* for the Detection of Nitroarenes as Mutagens in Complex-Mixtures Including Diesel Exhausts." Mutation Research **91**(2): 103-105.

Roussel, A. and C. Cambillau (1991). Silicon Graphics Geometry Partners Directory. Mountain View, CA, Silicon Graphics. **86**.

Sarraf, P., E. Mueller, et al. (1999). "Loss-of-function mutations in PPAR gamma associated with human colon cancer." Molecular Cell **3**(6): 799-804.

Schoonjans, K., B. Staels, et al. (1996). "The peroxisome proliferator activated receptors (PPARs) and their effects on lipid metabolism and adipocyte differentiation." Biochimica et Biophysica Acta-Lipids and Lipid Metabolism **1302**(2): 93-109.

Scrutton, N. S., J. Basran, et al. (1999). "Electron transfer in trimethylamine dehydrogenase and electron- transferring flavoprotein." Biochemical Society Transactions **27**(2): 196-201.

Shiff, S. J. and B. Rigas (1997). "Nonsteroidal anti-inflammatory drugs and colorectal cancer: Evolving concepts of their chemopreventive actions." Gastroenterology **113**(6): 1992-1998.

Shiff, S. J. and B. Rigas (1999). "The role of cyclooxygenase inhibition in the antineoplastic effects of nonsteroidal antiinflammatory drugs (NSAIDs)." Journal of Experimental Medicine **190**(4): 445-450.

Shimizu, T. and L. S. Wolfe (1990). "Arachidonic-Acid Cascade and Signal Transduction." Journal of Neurochemistry **55**(1): 1-15.

Stout, G. H. and L. H. Jensen (1968). X-Ray Structure Determination. New York, Macmillan.

Surette, M. E., J. D. Winkler, et al. (1996). "Relationship between arachidonate-phospholipid remodeling and apoptosis." Biochemistry **35**(28): 9187-9196.

Suzuki-Yamamoto, T., M. Nishizawa, et al. (1999). "cDNA cloning, expression and characterization of human prostaglandin F synthase." Febs Letters **462**(3): 335-340.

Takayama, T., S. Katsuki, et al. (1998). "Aberrant crypt foci of the colon as precursors of adenoma and cancer." New England Journal of Medicine **339**(18): 1277-1284.

Tarle, I., D. W. Borhani, et al. (1993). "Probing the Active-Site of Human Aldose Reductase - Site- Directed Mutagenesis of Asp-43, Tyr-48, Lys-77, and His-110." Journal of Biological Chemistry **268**(34): 25687-25693.

Tegeder, I., J. Pfeilschifter, et al. (2001). "Cyclooxygenase-independent actions of cyclooxygenase inhibitors." Faseb Journal **15**(12): 2057-2072.

Terwilliger, T. C. and J. Berendzen (1999). "Automated MAD and MIR structure solution." Acta Crystallographica Section D-Biological Crystallography **55**: 849-861.

Thompson, H. J., C. Jiang, et al. (1997). "Sulfone metabolite of sulindac inhibits mammary carcinogenesis." Cancer Research **57**(2): 267-271.

Tietze, L. F., M. Neumann, et al. (1989). "Proton-Mediated Liberation of Aldophosphamide from a Nontoxic Prodrug - a Strategy for Tumor-Selective Activation of Cytocidal Drugs." Cancer Research **49**(15): 4179-4184.

Tomarev, S. I., R. D. Zinovieva, et al. (1984). "A Novel Type of Crystallin in the Frog Eye Lens – 35 kDa Polypeptide Is Not Homologous to Any of the Major Classes of Lens Crystallins." Febs Letters **171**(2): 297-302.

Traver, R. D., T. Horikoshi, et al. (1992). "NAD(P)H-Quinone Oxidoreductase Gene-Expression in Human Colon- Carcinoma Cells - Characterization of a Mutation Which Modulates Dt-Diaphorase Activity and Mitomycin Sensitivity." Cancer Research **52**(4): 797-802.

Tsujii, M., S. Kawano, et al. (1998). "Cyclooxygenase regulates angiogenesis induced by colon cancer cells." Cell **93**(5): 705-716.

Urzhumtsev, A., F. TeteFavier, et al. (1997). "A 'specificity' pocket inferred from the crystal structures of the complexes of aldose reductase with the pharmaceutically important inhibitors tolrestat and sorbinil." Structure **5**(5): 601-612.

Usami, N., T. Yamamoto, et al. (2002). "Substrate specificity of human 3(20)alpha-hydroxysteroid dehydrogenase for neurosteroids and its inhibition by benzodiazepines." Biological & Pharmaceutical Bulletin **25**(4): 441-445.

Vane, J. R. (1971). "Inhibition of prostaglandin synthesis as a mechanism of action of the aspirin-like drugs." Nature(**231**): 232-235.

Versteeg, H. H., P. Henegouwen, et al. (1999). "Cyclooxygenase-dependent signalling: molecular events and consequences." Febs Letters **445**(1): 1-5.

Walker, M. C. and D. E. Edmonson (1987). Flavins and flavoproteins. D. E. Edmonson and D. B. McCormick. Berlin. New York, W. deGruyter: 149-157.

Walsh, M. A., G. Evans, et al. (1999). "MAD data collection - current trends." Acta Crystallographica Section D-Biological Crystallography **55**: 1726-1732.

Warshel, A. (1998). "Electrostatic origin of the catalytic power of enzymes and the role of preorganized active sites." Journal of Biological Chemistry **273**(42): 27035-27038.

Watanabe, K., Y. Fujii, et al. (1988). "Structural Similarity of Bovine Lung Prostaglandin-F Synthase to Lens Epsilon-Crystallin of the European Common Frog." Proceedings of the National Academy of Sciences of the United States of America **85**(1): 11-15.

Watanabe, M., M. Ishidate, et al. (1990). "Nucleotide-Sequence of *Salmonella typhimurium* Nitroreductase Gene." Nucleic Acids Research **18**(4): 1059-1059.

Weeks, C. M. and R. Miller (1999). "The design and implementation of SnB version 2.0." Journal of Applied Crystallography **32**: 120-124.

Westhead, D. R., T. W. F. Slidel, et al. (1999). "Protein structural topology: Automated analysis and diagrammatic representation." Protein Science **8**(4): 897-904.

Whiteway, J., P. Koziarz, et al. (1998). "Oxygen-insensitive nitroreductases: Analysis of the roles of nfsA and nfsB in development of resistance to 5-nitrofuran derivatives in *Escherichia coli*." Journal of Bacteriology **180**(21): 5529-5539.

WHO (1997). The World Health Report. Geneva, Switzerland, World Health Organisation.

Whong, W. Z. and G. S. Edwards (1984). "Genotoxic Activity of Nitroaromatic Explosives and Related- Compounds in *Salmonella typhimurium*." Mutation Research **136**(3): 209-215.

Wick, M., G. Hurteau, et al. (2002). "Peroxisome proliferator-activated receptor-gamma is a target of nonsteroidal anti-inflammatory drugs mediating cyclooxygenase-independent inhibition of lung cancer cell growth." Molecular Pharmacology **62**(5): 1207-1214.

Wilson, D. K., K. M. Bohren, et al. (1992). "An Unlikely Sugar Substrate Site in the 1.65 Angstrom Structure of the Human Aldose Reductase Holoenzyme Implicated in Diabetic Complications." Science **257**(5066): 81-84.

Wilson, W. R., W. A. Denny, et al. (1996). Proceedings of the 10th International Congress of Radiation Results. U. Hagen. D. Harder, H. Jung and C. Streffer. **2:** 791-794.

Winn, M. D., M. N. Isupov, et al. (2001). "Use of TLS parameters to model anisotropic displacements in macromolecular refinement." Acta Crystallographica Section D-Biological Crystallography **57**: 122-133.

Yennawar, N. H., H. P. Yennawar, et al. (1994). "X-Ray Crystal-Structure of Gamma-Chymotrypsin in Hexane." Biochemistry **33**(23): 7326-7336.

Zenno, S. and K. Saigo (1994). "Identification of the Genes Encoding NAD(P)H-Flavin Oxidoreductases That Are Similar in Sequence to *Escherichia coli* Fre in 4 Species of Luminous Bacteria – *Photorhabdus luminescens*, *Vibrio fischeri*, *Vibrio harveyi*, and *Vibrio orientalis*." Journal of Bacteriology **176**(12): 3544-3551.

Zenno, S., H. Koike, et al. (1996a). "Gene cloning, purification, and characterization of NfsB, a minor oxygen-insensitive nitroreductase from *Escherichia coli*, similar in biochemical properties to FRase I, the major flavin reductase in *Vibrio fischeri*." Journal of Biochemistry **120**(4): 736-744.

Zenno, S. H., H. Koike, et al. (1996b). "Conversion of NfsB, a minor *Escherichia coli* nitroreductase, to a flavin reductase similar in biochemical properties to FRase I, the major flavin reductase in *Vibrio fischeri*, by a single amino acid substitution." Journal of Bacteriology **178**(15): 4731-4733.

Zhang, X. P., S. G. Morham, et al. (1999). "Malignant transformation and antineoplastic actions of nonsteroidal antiinflammatory drugs (NSAIDs) on cyclooxygenase-null embryo fibroblasts." Journal of Experimental Medicine **190**(4): 451-459.

Zhu, D. W., L. Cantin, et al. (2001). "Crystallization and preliminary X-ray crystallographic analysis of the human type 3 3 alpha-hydroxysteroid dehydrogenase at 1.8 Å resolution." Acta Crystallographica Section D-Biological Crystallography **57**: 589-591.

Appendix A

Lovering, A. L., E. I. Hyde, et al. (2001). "The structure of Escherichia coli nitroreductase complexed with nicotinic acid: Three crystal forms at 1.7 Å, 1.8 Å and 2.4 Å resolution." Journal of Molecular Biology **309**(1): 203-213.