INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI®

THE STRUCTURE OF HOMOSERINE DEHYDROGENASE FROM *Saccharomyces cerevisiae*

By

BYRON DELABARRE, B.Sc., M.Sc.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements

for the Degree

Doctor of Philosophy

McMaster University

THE STRUCTURE OF HOMOSERINE DEHYDROGENASE

Doctor of Philosophy (2000) McMaster University

(Biochemistry) Hamilton, Ontario

TITLE: The Structure of Homoserine Dehydrogenase from *Saccharomyces cerevisiae*

AUTHOR: Byron DeLaBarre, B.Sc. (*Applied Chemistry)* (University of Waterloo), M.Sc. (*Chemistry)* (McMaster University)

SUPERVISOR: Dr. A. M. Berghuis

NUMBER OF PAGES: xvi, 241

Abstract

The core of this work is the structural determination of the enzyme homoserine dehydrogenase from *Saccharomyces cerevisiae*. The structure has been determined in three different states of ligation: apo form, in a binary form complexed with nicotinamide cofactor, and in a ternary form complexed with both L-homoserine and a nicotinamide cofactor analogue. The structures were determined by X-ray diffraction studies to resolutions of 2.3, 2.3, and 2.6 Å, respectively. Due to problems with non-isomorphism, the measured amplitudes were phased from a single heavy atom derivative. Data averaging about a local two-fold axis (non-crystallographic symmetry) was used in combination with other density modification techniques to resolve the phase ambiguity problem arising from single isomorphous replacement. Because traditional techniques failed to locate the axis, a novel technique for locating the symmetry axis was devised.

The fold of the catalytic / substrate binding portion of the structure establishes homoserine dehydrogenase as a unique dehydrogenase. The ternary complex reveals that the amino acid substrate binds to the enzyme in an unusual manner. Site directed mutagenesis data based upon the structure validates the selection of active site residues. Furthermore, the combination of residues employed for catalysis of the hydride transfer

suggests that the bound form may be a *gem*-diol form. Reaction mechanisms in both the forward and reverse directions are proposed.

This is the first structural determination of homoserine dehydrogenase. A comparison of the *Saccharomyces cerevisiae* homoserine dehydrogenase primary sequence with sequences from other homoserine dehydrogenases reveals that the fold of this enzyme is conserved across species. The structure explains the effect monovalent cations have upon the enzyme, how it is affected by an inhibitory compound and suggests a possible mechanism for allosteric control.

Acknowledgements

I would like to acknowledge Suzanne Jacques and Catherine Pratt who together developed the overexpression system and purification scheme for HSD. They provided a genesis for this project by demonstrating that HSD could be crystallized. Suzanne Jacques also determined many of the kinetic properties of HSD during her doctoral work. Randy Read at Cambridge gave some great advice which helped in the formulation of a new way to find NCS, and Liang Tong at Columbia was kind enough to lend some unpublished algorithms to do this. Dan Yang provided advice on how to combine heavy atom data for the elucidation of optimal phase information. Paul Thompson made the mutants discussed in chapter 6 and provided useful arguments while the reaction mechanism was being worked out. Christy Thomson showed me how to do the ruthenium oxide staining experiments; many thanks to her. Many thanks to my supervisory committee for allowing me to proceed with a project that was fraught with difficulties; in the end it was worth the struggle. My committee consisted initially of Albert Berghuis, Gerry Wright and Dan Yang. Lynne Howell graciously agreed to join during the final year of my work. All four provided me with high standards that I tried my best to meet. Albert Berghuis is given extra acknowledgement as my immediate supervisor. He suffered my desultory ways, but most importantly he left me alone to do my work. Many thanks to the past and

present members of the Berghuis lab group for their outward tolerance of some of my less agreeable personality traits and for getting me fig newtons during synchrotron trips. Thanks go to Gerhard Gerber for admitting me in to the Biochemistry doctoral program after my stint in the department of Chemistry. He gave me some great advice about the worth of classwork versus benchwork. I may have lost out on MRC, NSERC, and OGS, but I am a better scientist for it. Thanks to the American Crystallographic Association and the International Centre for Diffraction Data for awards given for my work. Their recognition meant a lot to me.

Finally, thanks to my wife Barbara who put up with my absences during the evenings and weekends I spent with the X-ray machine and computer graphics terminal. She helped me to remember that there is more to life than helices, loops, and strands. I dedicate this work to her.

Table of Contents

List of Figures and Boxes

List of Tables

List of Abbreviations Used in This Document

α, β	Alpha helix or beta strand. HSD secondary structure is labeled such that uppercase letters are used for helices and lowercase letters for strands.
AIDS	Acquired immunity deficiency syndrome
AK	Aspartate kinase
ASAD	Aspartate semi-aldehyde dehydrogenase
C Terminal	Carboxyl terminus
CCP4	Collaborative computational project #4
CHESS	Cornell high energy synchrotron source
E. coli	*Escherichia coli*
F, F_o, F_c	Structure Factor (general,observed or calculated), defined as $F_{hkl} = \Sigma f_j \exp[2\pi i(hx_j+ky_j+lz_j)]$
$FADH_2$	Flavin adenine dinucleotide (reduced form)
$FMNH_2$	Flavin mononucleotide (reduced form)
GAPDH	Glyceraldehyde phosphate dehydrogenase
HIV	Human immunodeficiency virus
HSD	Homoserine dehydrogenase
IDH	Isocitrate dehydrogenase
kDa	kiloDalton (1000 g / mol)
K_i	Inhibition constant
K_m	Michealis constant

K_{sp}	Solubility constant
L-ASA	L-Aspartate
L-Hse	L-Homoserine
L-Ser	L-Serine
L-Thr	L-Threonine
MDR	Medium chain dehydrogenases reductases
MIRAS	Multiple isomorphous replacement with anomalous scattering
N Terminal	Amino terminus
NAD	Nicotinamide adenine dinucleotide (oxidized or reduced)
NAD(P)	Ribose 2' phosphorylated form of NAD
NAD^+	NAD in oxidized form
NADA	3-aminopyridine adenine dinucleotide (NAD analogue)
NADH	NAD in reduced form
NCS	Non-crystallographic symmetry
NMR	Nuclear magnetic resonance
OD	Optical density
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Polyethylene glycol
PIP	di-μ-iodobis(ethylenediamine) diplatinum (II) nitrate
r.m.s.d.	Root mean square deviation

R_{cullis}	Heavy atom phasing statistic generally defined as the ratio of the heavy atom vector to the lack of closure
RI-331	*(S)*-2-amino-4-oxo-5-hydroxypentanoic acid
R_{merge}	Crystal data merging statistic
$R_{factor/free}$	Crystallographic statistic defined as Σ ($\|F_{obs} - F_{model}\|$) / $\|F_{obs}\|$) for R_{factor}. R_{free} is defined similiarily, but uses a subset of reflections that are not used during refinement.
SBDD	Structure based drug design
SDR	Short chain Dehydrogenase / Reductases
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SIR	Single isomorphous replacement
UDP	Uracil diphosphate
UGE	UDP galactose epimerase
UV	Ultraviolet
WHO	World health organization

Chapter One: Introduction / Background

1.1 The Problem Of Systemic Fungal Infections

The work in this thesis was broadly motivated by the desire to better understand what it will require to treat fungal infections. There are two ways in which fungi adversely affect human civilization: fungi can attack our crops (Jacques (33) wrote an excellent review on this aspect) and fungi can attack us directly. Infections can be either systemic or topical; because systemic fungal infections are a more serious threat to a person's well being than the discomfort caused by topical fungal infections, the systemic infections will be dealt with in more detail here. Systemic fungal infections differ from topical fungal infections in several ways: systemic infections inflict higher mortality rates, occur less frequently, and are much more difficult to treat.

Topical fungal infections are more familiar to the general population in nuisance diseases such as athlete's foot, caused partially by the dermatophyte fungi *Trichophyton rubrum* (1, 2), and dandruff, caused by the fungal pathogen *Malassezia furur* (3). There is a simple reason for the greater prevalence of topical fungal infections: fungi pose a nominal threat for a healthy immune system. Unfortunately for

immunocompromised individuals who cannot battle invading fungi, the situation is critically different. Mortality rates for some fungal pathogens approach 100% when they are able to colonize a critical organ such as the lungs or brain (4, 5). Immunocompromised individuals include the elderly, organ transplant recipients, chemotherapy recipients, and acquired immunodeficiency syndrome (AIDS) patients. The latter group of people are the fastest growing segment of immunocompromised people in our population, so it will be illustrative to examine their situation in greater detail.

1.1.1 Fungal Infections in AIDS Patients

AIDS is brought on by infection with the human immunodeficiency virus (HIV) and results from a destruction of T-cells, the target of the invading virus (6, 7)

AIDS is epidemic world wide; the United Nations and WHO estimates that 5.8 million people were newly infected with the HIV virus in 1997 alone (8). Sub-saharan Africa is at the centre of the epidemic; of the estimated 30.6 million people in the world living with HIV or AIDS, two thirds live in developing African nations. Conditions are better here in Canada, but they are still problematic. Although the number of reported AIDS cases in Ontario has declined in recent years, the number of reported HIV infections in 1997 has, by some estimates, risen by as much as 100% over those reported in 1996 (9). These seemingly contradictory

facts can be reconciled by taking into consideration the success of the most recent anti-viral therapies, especially the new HIV protease inhibitors (10, 11). Despite recent advances in slowing the progression of HIV infection into AIDS, there are few cases where HIV infection did not ultimately lead to the contraction of AIDS, a decreased quality of life, and eventually death. Furthermore, the current drug regimens are expensive, an important factor when considering that the majority of infected people live in third world nations.

Another pernicious aspect of disease is that the virus is extremely mutable; early successes with reverse transcriptase inhibitors were quickly eradicated by mutations in the reverse transcriptase enzyme. To rest on our laurels with the most recent therapeutic advances would be to foolishly ignore past lessons learned not only from the HIV virus, but from other microbes as well. For example, penicillin was hailed as one of the great medical advances of the 20th century but we are now faced with bacteria that have evolved, because of their overexposure to antibiotics, into pathogens that are untreatable by current medical technologies (12-14). Fortunately, new anti-viral drugs are still being developed to delay the onset of AIDS after contracting an HIV infection. Furthermore, vaccinations against the HIV virus are being actively pursued to prevent new HIV infections (15); success in this area would be of tantamount importance to our current AIDS crisis, but it is still not within our grasp.

Approximately 40% of AIDS fatalities have been attributed to systemic fungal infections (16). The weakened immune system of AIDS patients is obviously the cause of their susceptibility to fungal pathogens, but it is also known that those simply infected with HIV are prone to infection as well. There is a significantly different pattern of fungal infection between HIV-infected and HIV-free individuals who have been infected by fungal pathogens (17, 18). The reasons for this are not entirely clear, but some researchers (19) have found evidence to suggest that it is the protein coating of HIV, specifically glycoprotein 120, which allows fungal infections to elude the immune system of HIV infected individuals.

Current treatment options for systemic fungal infections are surprisingly sparse; furthermore, the occurrence of strains resistant to present day drugs is on the rise (20). There are approximately 60,000 HIV infected individuals in Canada by 1997 estimates (9). Coincident with the growing number of HIV infected people, opportunistic fungal infections will occur more frequently. The limited range of antifungal compounds available for treatment of these infections will undoubtedly accelerate the emergence of drug resistant pathogenic fungi.

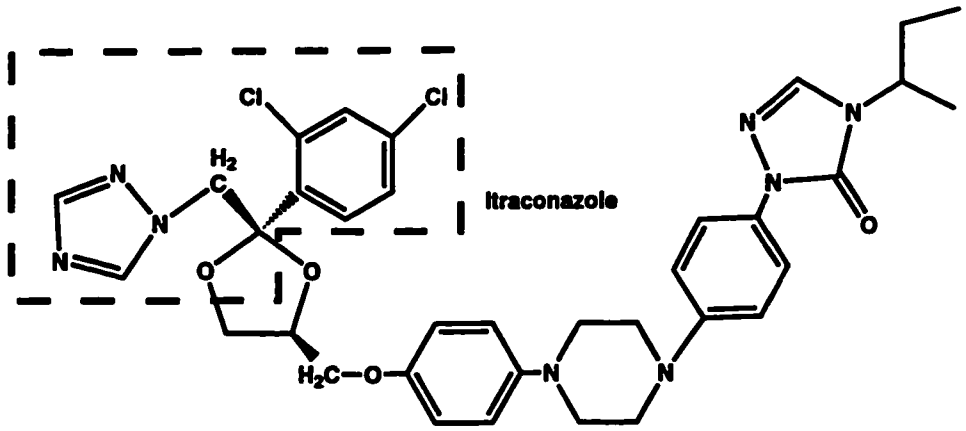
Oral thrush, caused by the fungus *Candida albicans* (21), is a clinically important example of an opportunistic infection. Thrush causes extreme discomfort in those it afflicts; if left unchecked, the disease can spread in the host patient, eventually causing death. *C. albicans* is the most common

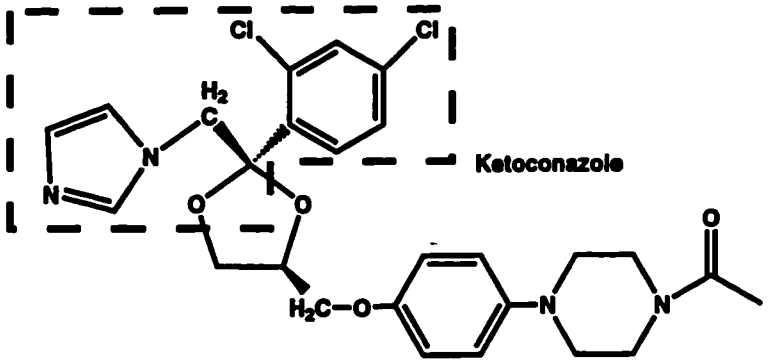
opportunistic fungal pathogen found in AIDS patients, but there are several other examples. These include variant *Candida* species such as *C. glabrata and C. krusei* (22) and different fungi such as *Aspergillus niger* (23) and *Cryptococcus neoformans* (17).

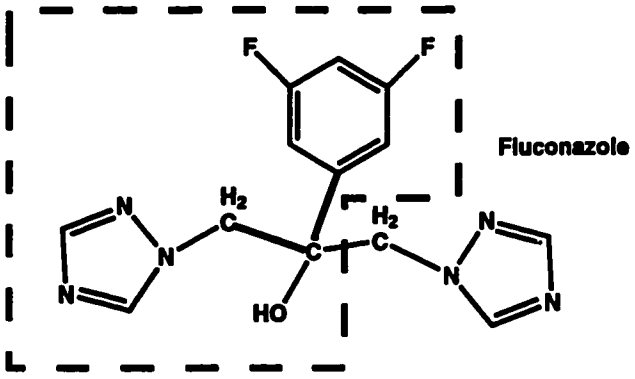
The most common treatment for fungal infections is to prescribe one of the azoles (*e.g.* fluconazole, itraconazole, and ketoconazole; see Figure 1.1). The azoles act by interfering with sterol synthesis in fungi (24), thereby slowing their growth and ultimately resulting in unviable metabolic disturbances.

Unfortunately, strains of clinically relevant fungal species are becoming resistant to azole therapy. One harbinger of drug resistance is the emergence of new strains of pathogens; this was demonstrated (25) on a genetic level with PCR based techniques. In the study of 19 different isolates of *C. albicans*, none of the strains were identical. In a larger study of 700 *C. albicans* isolates (26), the researchers used relative growth methods to determine that nearly 40% of the strains were resistant to some form of azole therapy, and that nearly 19% of the resistant strains were resistant to all common azole therapies. The latter observation denotes the existence of cross-resistance to azole therapy, *i.e.* exposure to one of the azole drugs can effect a resistance to all of the azole drugs. Some researchers (20) think that this is evidence of an active efflux mechanism: fungi have become capable of recognizing the common structural components of azole based compounds and are simply pumping them out of their cells. One group (27) concluded

Figure 1.1 - The Azole Class of Antifungal Drugs







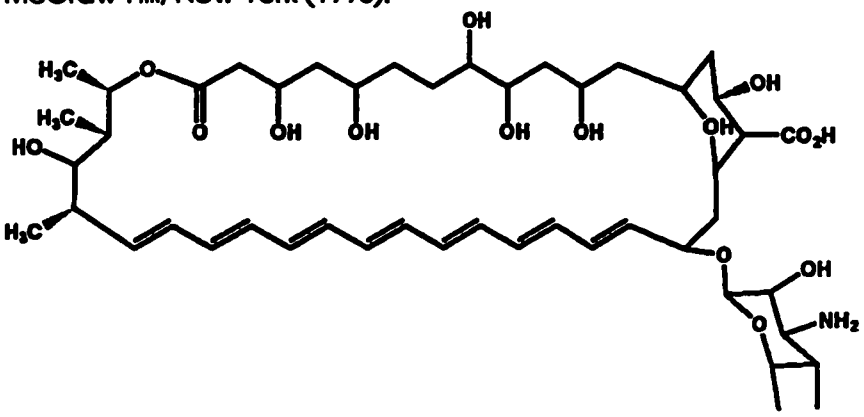
Structures of the azole drugs adapted from Hardman, J. G., and Limbird, L. E., Goodman & Gilman's The Pharmacological Basis of Theraputics, 9th ed, McGraw-Hill, New York (1996). The dashed polygons indicate the similar moiety in the three compounds.

from multivariate statistical analyses of clinical data that the primary risk factor for the development of resistant fungi is the use of fluconazole. Others (28) have noted that the widespread use of azole drugs in the treatment and prophylaxis of fungal infections is becoming an increasing danger: they cite the case of one individual treated with increasing doses of fluconazole who was found to harbour 17 different *C. albicans* isolates over a 2 year period.

The backup therapy when azole drugs fail is amphotericin-B, a drug which is usually prescribed for late stage AIDS patients. Amphotericin-B, shown in Figure 1.2 is structurally different from the azole and functions by destroying the integrity of the fungal cellular membrane (8).

Figure 1.2 Amphotericin B

Structure taken from Hardman, J. G., and Limbird, L. E., Goodman & Gilman's The Pharmacological Basis of Theraputics, 9th ed, McGraw-Hill, New York (1996).



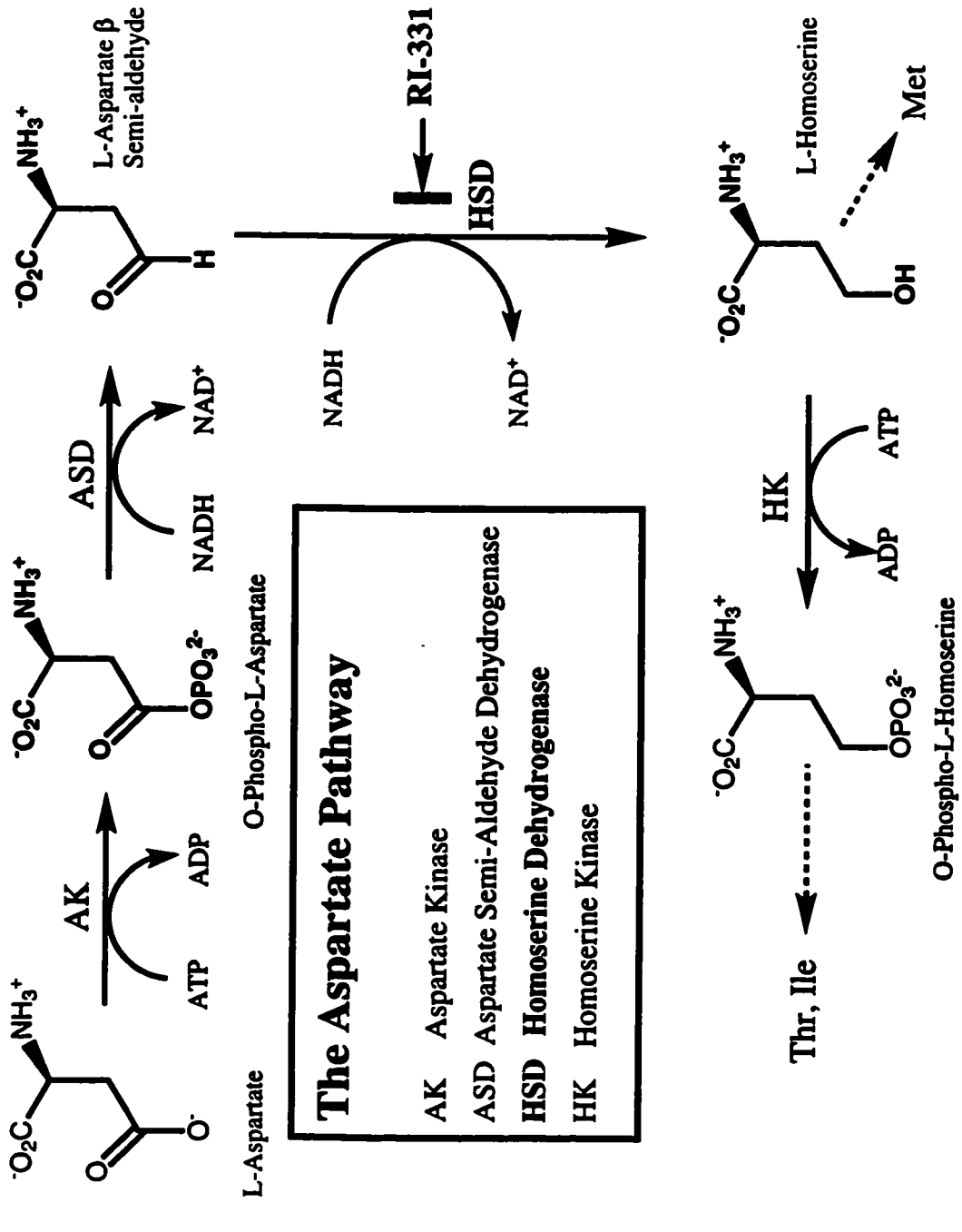
Amphotericin-B therapy requires hospitalization because of the toxic side effects of the drug, and is generally administered *via* an intravenous

catheter. Ironically, the very use of the catheter puts the patient at risk to further nosocomial infections at the site where the catheter is introduced (29). Amphotericin-B is presently the last line of defense against AIDS-related fungal infections.

1.2 The Identification Of A New Antifungal Agent

The work in this thesis was specifically inspired by the discovery of a natural product isolated (30) from a *Streptomyces* culture broth. The chemical name for this compound is *(S)*-2-amino-4-oxo-5-hydroxypentanoic acid but it is referred to in the literature as RI-331. RI-331 was shown to inhibit the growth of a variety of fungal species including *C. albicans, Cryptococcus neoformans* and *Saccharomyces cerevisiae*. Yamaki *et al.*(31) showed that RI-331 was affecting amino acid levels in *S. cerevisiae* and were able to isolate the cause to a disruption in the aspartate pathway. The aspartate pathway is responsible for the conversion of aspartic acid to methionine, isoleucine and threonine (32), as shown in Figure 1.3. Note that the aspartate pathway in fungi is slightly different from other classes of organisms; it does not generate lysine. Lysine synthesis in other organisms starts from L-aspartate β-semi-aldehyde (L-ASA), the substrate for homoserine dehydrogenase (HSD). Thus HSD in other species occupies a branch point in the pathway and could conceivably be a point of regulation. Fungi synthesize their lysine *via* a different pathway, the sacaropine pathway, starting from α-ketoglutarate.

Figure 1.3 - The Aspartate Pathway



by RI-331 is HSD. The role of HSD in the aspartate pathway is to convert L-ASA into L-homoserine (L-Hse).

1.2.1 Structure Based Drug Design

It is important to note that RI-331 is still just an inhibitor of HSD and not yet a successful anti-fungal drug. The development of a drug from RI-331 requires the design, synthesis, and testing of many different analogues of RI-331. The work described herein is concerned only with improving the design stage of this process. Manipulations of the atomic structure of RI-331 to effect a fine tuning of its inhibitory and pharmicokinetic properties are required to elevate the compound from inhibitor to pharmaceutical agent status. Such an undertaking will be greatly accelerated by a direct molecular level understanding of the interaction between HSD and RI-331.

This approach to drug design has been called 'structure based drug design' (SBDD) (34) and has only come to prominence within the past 10 years as a powerful tool for the drug development and discovery process.

One very pertinent example of the role of SBDD in drug development can be found in the elaboration of HIV protease inhibitors. The structure of the protease enzyme from HIV was determined by three different groups (35-37) nearly 10 years ago. Since that time, greater than 300 structures of the protease in complex with experimental inhibitors have been determined, giving researchers unambiguous insight into why certain inhibitors are better than others. There has been a boom in the development of potent protease

inhibitors that can be attributed to the greater understanding we have of the interactions between the inhibitors and target enzyme.

SBDD is of great use in redesigning drugs which have become ineffective because of antimicrobial resistance. There are three main modes of drug resistance: inhibitor modification, target modification, and active efflux. SBDD shows its greatest strength when target modification occurs. Structural characterization of the mutant enzyme enables compensation to be rapidly incorporated into new generations of inhibitors; the rapid structural characterization of mutations of the HIV protease enzyme is an excellent case in point (38).

For inhibitor alteration or inhibitor efflux, knowledge of the key chemical features of inhibitory molecules arising from SBDD studies is still valuable towards the design of new inhibitors. Unfortunately, situations will arise in which the key chemical features of the inhibitor are exactly what make it susceptible to resistance; in these instances, SBDD studies can only serve the purpose of confirming classes of inhibitors that are no longer useful as medicinal agents.

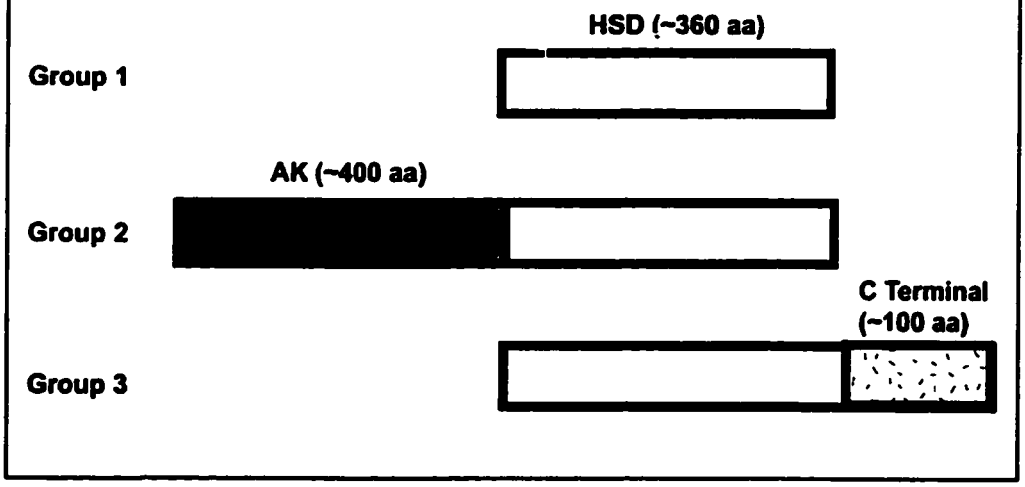
There are no animal homologues of either HSD or the fungal aspartate pathway. This bodes well for the development of a potent drug with minimal side-effects. That a known inhibitor of HSD can kill pathogenic fungi, that the enzyme does not have a homologue in man, and that the enzyme is readily available in pure form makes HSD an ideal candidate for the SBDD process.

1.3 HSD Is A Unique Dehydrogenase

HSD shares some similarities in its manner of dinucleotide binding with the short chain dehydrogenase / reductase (SDR) and 2-oxy acid family (*vide infra*), but it is otherwise unique in its overall fold and key catalytic and substrate binding residues. One of the main goals of this thesis is to demonstrate and understand the unique features of this enzyme.

HSD can be found in organisms from all kingdoms other than animals. The HSD enzyme from all organisms for which the primary sequence has been determined can be categorized into one of three HSD groups: group 1 are short enzymes, group 2 are bifunctional enzymes with both aspartate kinase and homoserine dehydrogenase activity, and group 3 have a C-terminal extension (see Figure 1.4). This review will deal primarily with the enzymes from *S. cerevisiae, Escherichia coli* and *Bacillus subtilis*, as these represent the first discovered member of each of the three groups, respectively. The grouping and a broader coverage of HSD enzymes is presented in chapter 8.

Figure 1.4 – Domain Diagram for The Three HSD Groups



1.3.1 *S. cerevisiae* HSD – Group 1

HSD activity was first isolated from *S. cerevisiae* by Black and Wright (39, 40). Black and Wright purified the enzyme 100-fold from *S. cerevisiae* and demonstrated its requirement for NADH or NADPH in the conversion of L-ASA to L-Hse. Many years later Yumoto *et al.* (41) improved the purification and then used both SDS-PAGE and gel filtration experiments to establish that the enzyme was a homodimer with an approximate molecular weight of 81 kDa. Yumoto *et al.* discovered that L-threonine (L-Thr) inhibited the enzyme in a competitive fashion, contrary to the situation for HSD from other species where L-Thr behaves as an allosteric inhibitor (*vide infra*). Yumoto *et al.* also examined the effect of K^+ ions on HSD, finding that they both afforded heat stabilization and augmented the catalytic ability of the enzyme.

The *hom6* gene encoding *S. cerevisiae* HSD was sequenced by Thomas *et al.* (42). The sequence revealed that the yeast enzyme was most

similar to the *E. coli* enzyme. Furthermore, the gene encoded the shortest of all known HSD enzymes with a predicted molecular mass of 38 kDa for the protomeric species.

Yamaki *et al.* (43) used the *S. cerevisiae* enzyme to study the effects of RI-331. They established that the compound was a competitive inhibitor with respect to L-Hse, implying that it occupies the same binding site as L-Hse. They suggested that the inhibitor was forming a complex with the $NADP^+$ cofactor based upon evidence that preincubation of the enzyme with $NADP^+$ increased the inhibitory effect.

The most recent study of the enzyme by Jacques (33) showed that the enzyme mechanism was an ordered bi-bi process, with the nicotinamide cofactor binding first and releasing last. Cofactor release is the rate determining step of the catalysis. Jacques also observed an augmentation of enzyme activity by monovalent cations.

Jacques further characterized the inhibitory effects of RI-331, showing that the inhibitor was forming a tight binding complex which was dependent upon the presence of $NAD(P)^+$. Unfortunately, the complex could not be isolated for further study.

1.3.2 *E. coli* HSD - Group 2

Despite the fact that HSD was first isolated from *S. cerevisiae*, the enzyme from *E. coli* has received most of the attention. The enzyme is present as two isozymes in *E. coli*: AK-HSDI and AK-HSDII. Of the two,

AK-HSD I is the most abundant enzyme in the cell and was thus more available for study because it can be readily isolated from growing bacteria with classical fractionation methods (44). Work prior to 1973 and for most of 1974 are covered in a review by Truffa-Bachi *et al.* (45).

Both isozymes are bifunctional, exhibiting both aspartate kinase and homoserine dehydrogenase activities, and both isozymes function as a tetramer composed of four homoserine units and four aspartokinase units. However, for the AK-HSDI isozyme, the tetrameric form exhibits a complex allosteric control mechanism, whereas there is a complete absence of allosteric control for the AK-HSDII enzyme.

The allosteric control is effected by L-Thr which has both high and low affinity binding sites on the enzyme (46-48). Occupancy of the high affinity binding site (K_D=0.16 mM for L-Thr) results in 80-90% inhibition of both the kinase and dehydrogenase activities. Furthermore, the high affinity binding site exhibits positive cooperativity with a Hill coefficient of 2.3. The importance of cooperative allosteric control in AK-HSDI was demonstrated by Fontan *et al.* (49) who showed that chemical modification of a cysteine residue in a single subunit of the tetramer was sufficient make the enzyme insensitive to L-Thr. L-Ser has been shown to have inhibitory effects upon the enzyme in a manner similar to L-Thr (50). This effect is probably not physiologically relevant because unlike threonine, serine is not a product of the aspartate pathway (51).

The effect of monovalent cations, specifically Na^+ and K^+, upon the activity of AK-HSDI/II has been known since the earliest studies of these enzymes. Recently, Wedler *et al.* (52) demonstrated with initial velocity kinetics that AK-HSDI has an order of magnitude higher affinity for Na^+ ions over K^+ ions. Furthermore, the presence of the Na^+ ion decreased enzyme activity by an order of magnitude, whereas the presence of the K^+ ion increased enzyme activity 4 fold. Wedler *et al.* proposed a structural role for these ions, but could not rule out a direct role in binding the substrate.

The kinetic mechanism for AK-HSDI was established by Wedler *et al.* (53). Wedler *et al.* used equilibrium isotope exchange kinetics to probe the reaction mechanism and understand how it was effected by the presence of allosteric inhibitors and / or monovalent cations. The results of the experiments supported a complex kinetic model involving 18 different forms of the enzyme. The overall kinetics were predominantly preferred order with the cofactor binding first and releasing last as the rate determining step, similar to the mechanism established for the *S. cerevisiae* enzyme. Random order kinetics could be observed when the pH was lowered from 9.0 to 8.0, supporting the model which had been previously proposed by Angeles *et al.* (54).

1.3.3 *B. subtilis* HSD – Group 3

The HSD from *B. subtilis* was first studied by Skarstedt *et al.* (55) who used the enzyme purified from its natural source. The *Bacillus subtilis* enzyme does not possess bifunctional activity. Skarstedt *et al.* suggested that HSD was inhibited by L-Thr but did not present data to support their claim. The enzyme was subsequently studied by Yeggy *et al.* (56) who also used the enzyme purified from its natural source. They observed only a 25% decrease in HSD activity due to L-Thr.

The *hom* gene encoding HSD was cloned by Parsot *et al.* (57). Parsot *et al.* made three important comparisons between the *E. coli* and *B. subtilis* enzymes.

1. The C terminal half of the *B. subtilis* HSD and both *E. coli* sequences exhibited a similarity that began exactly where the *B. subtilis* AK II and the *E. coli* AK-HSD I/II similarity ended. This was the first direct evidence that the group 2 (bifunctional) HSD enzymes arose from a gene fusion event between AK and HSD.
2. They identified the high degree of similarity which occurred at the then unknown active site region of the enzyme (residues 200-221 of the *B. subtilis* enzyme, *cf.* residues 208-223 of *S. cerevisiae* enzyme).
3. They noted a C terminal extension past the sequence for the *E. coli* HSD. Although the evidence for L-Thr control of the *B. subtilis* HSD

was not clear, they proposed that this C terminal extension was where the allosteric binding site was located.

The allosteric role of the C terminal domain of the group 3 proteins was shown experimentally by Archer *et al.* (58) for HSD from *Corynebacterium glutamicum*. In attempting to understand the basis for allosteric regulation in these enzymes, Archer *et al.* screened mutants for L-Thr insensitivity. The mutants were generated by whole cell UV irradiation. The irradiation was imposed in an iterative manner until a mutant was isolated which was insensitive to L-Thr but otherwise still exhibited wild type catalytic activity. When the gene encoding HSD in the L-Thr insensitive mutant was examined, it revealed that the mutation responsible for the loss of allosteric control was a deletion of G^{1964} in the nucleotide sequence. This base is at codon 429; its deletion resulted in a frameshift which gave rise to an alteration of 10 amino acids past the site of mutation and the introduction of a premature stop codon, thereby truncating seven amino acids from the C terminus of the protein. There is a well conserved arginine at position 441; this residue either constitutes part of the L-Thr binding site or forms a polar interaction with the rest of the protein.

1.3.4 The Control of HSD Within the Aspartate Pathway

The concept of feedback regulation which has made the HSD enzyme the subject of intense research for over thirty years was tested in

a set of experiments performed by Wampler *et al.* (59). An *in vitro* model of the portion of the aspartate pathway which converts L-aspartate into L-threonine was created. Aspartate kinase I, aspartate semi-aldehyde dehydrogenase, AK-HSDI, homoserine kinase, and threonine synthase from *E. coli* and aspartate kinase and HSD from *Rhodopseudomonas spheroides* were purified from natural sources and used to create the coupled assay. The reaction was followed by monitoring the ΔOD_{390} (NADPH oxidation). A quick initial rate for the first 2-3 minutes was observed followed by a slower, steady state rate for the coupled assay. The addition of 1 mM L-Thr to the reaction mixture nearly eliminated the fast initial rate, demonstrating that it was responsible for the feedback inhibition.

The quantities of amino acid intermediates separated from the reaction mixture and measured by both radiolabelling and enzymatic assay methods did not support the L-Thr feedback mechanisms proposed for HSD. None of the intermediates were shown to accumulate in the L-Thr inhibited reaction which suggests the absence of feedback inhibition for any other individual enzymes in the pathway. This result may be misleading. Although Wampler *et al.* claimed to have accounted for the instability of the intermediates, *e.g.* L-ASA, it seems unlikely that this could be completely controlled within the complex reaction mixture. Furthermore, the measurement errors are unknown because the accuracy

of the quantitations was not examined; if the allosteric control exhibited by HSD was subtle it could have been easily missed by the techniques used in the experiments. Finally, the enzymes used in the coupled assay were derived from natural sources; the purification data presented in the paper clearly shows the presence of non-aspartate pathway species within the reaction mixture. If the contaminants interacted with any of the intermediates, it may have skewed the results.

1.4 The Classes Of Dehydrogenases

HSD belongs to a class of enzymes called oxidoreductases, specifically to the dehydrogenase/reductase subgroup of this class. These are enzymes which accomplish an overall oxidation or reduction of a molecule by transfer of a hydride atom from a cofactor such as NAD(P)H, $FMNH_2$ or $FADH_2$. HSD uses the molecule NAD(P)H as its source of hydride atoms and therefore belongs to the group of NAD(P)H dependent dehydrogenases. In order to evoke a more complete understanding of HSD, it would be worthwhile to examine where HSD fits within this group of enzymes.

There are 5 main divisions of dehydrogenase/reductase enzymes. These are the short chain dehydrogenases (60), the medium chain dehydrogenases (61, 62), the aldo-keto reductases (63, 64), the 2-oxyacid dehydrogenases (65, 66), and the aldehyde dehydrogenases (67).

Structural representatives from each of these groups is shown in Figure 1.5.

1.4.1 Short Chain Dehydrogenases

The short chain dehydrogenases (SDR) are a superfamily which are composed of a core group of 57 characterized enzymes and an extended group of 22 enzymes which show a distant relationship to the core group. The group possesses the following characteristics:

1. An approximate length of 250 amino acids in the primary sequence. The extended definition includes proteins which are up to 350 amino acids in length.
2. A single domain architecture for the overall fold
3. An N terminal **GXXXGXG** motif which indicates the fold responsible for dinucleotide binding
4. A strictly conserved tyrosine residue that is frequently accompanied by a lysine residue four positions away to generate a **YXXXK** motif near the middle of the sequence.
5. A dimeric or tetrameric aggregation state.
6. No dependancy upon metal atoms for catalysis.

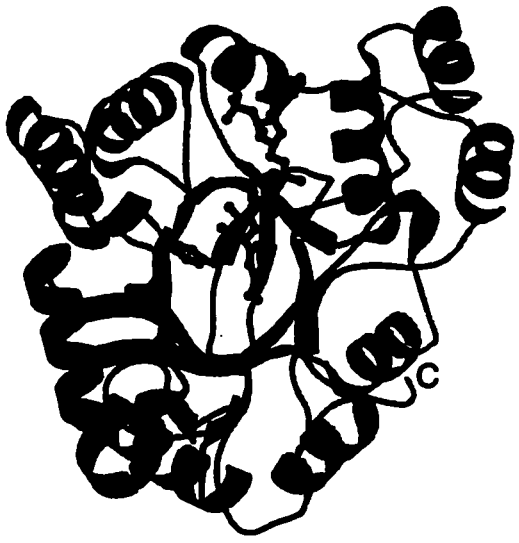
Members of the core group include *Streptomyces* 3α/20-β-Hydroxysteroid dehydrogenase (Figure 1.5A) and *Murine* dihydropteridine reductase. Both of these proteins have been structurally determined by X-ray crystallography (68, 69). The extended group includes UDP-glucose

Figure 1.5(A-C) - Structural Representatives of Oxidoreductase Families Which Employ NAD(P)H as a Cofactor

(A) Short Chain Dehydrogenase Reductase (SDR)

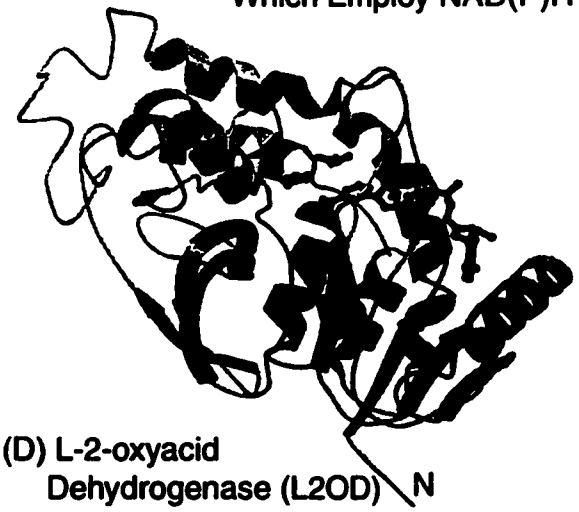
(B) Medium Chain Dehydrogenase Reductase (MDR)

All structures (***A-G***) are shown as monomers. N and C denote the location of amino and carboxyl termini, respectively. Except for (c), structures are oriented with the dinucleotide binding domain at bottom. Secondary structure as defined by original authors. ***(A)*** **SDR** - 3α/20β Hydroxysteroid Dehydrogenase (68) complexed with NAD^+. The arrow indicates the conserved YxxxK residues. ***(B)*** **MDR** - Class 1 Alcohol Dehydrogenase (71) in complex with NAD^+ and DMSO. The arrow indicates the catalytic zinc ion. ***(C)*** **AKR** - Aldose Reductase (87) complexed with NADPH. The view is downwards into the α/β barrel. The proposed catalytic tyrosine is shown. The cofactor is at the top of the barrel.



(C) Aldo-keto Reductase (AKR)

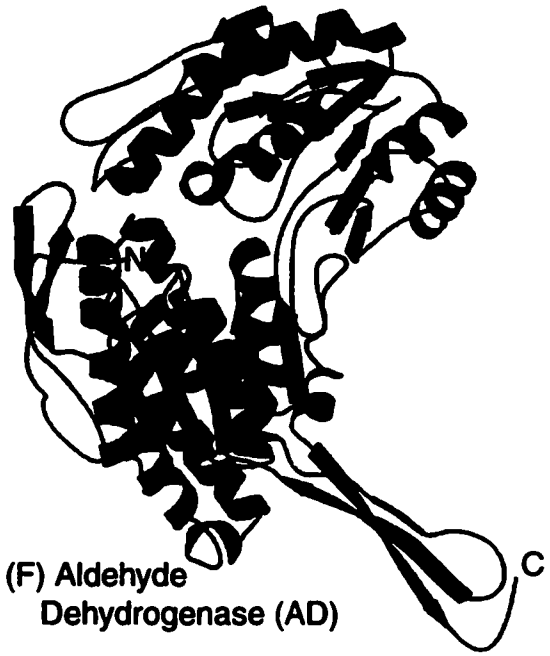
Figure 1.5(D-G) Structural Representatives of Oxidoreductase Families Which Employ NAD(P)H as a Cofactor



(D) L-2-oxyacid Dehydrogenase (L2OD)

(E) D-2-oxyacid Dehydrogenase (D2OD)

***(D)* L20D** -Lactate Dehydrogenase(76) complexed with NAD^+ and oxamate. The arrow inidicates the catalytic histidine residue. ***(E)* D20D** - D-3-Phosphoglycerate Dehydrogenase(88) complexed with NAD. ***(F)* AD** - Aldehyde dehydrogenase(82), apo form. ***(G)* HSD** - Homoserine Dehydrogenase bound with NAD^+ and L-Homoserine. The arrow indicates a key lysine/aspartate pair of residues. HSD is not a family of dehydrogenase structures, but is included here for comparison. All figures prepared with *Molscript/Raster3D.*



(F) Aldehyde Dehydrogenase (AD)

(G) Homoserine Dehydrogenase (HSD)

4-epimerase that has also been structurally characterized (70). All determined structures exhibited similar folds, supporting the deduction (60) made from the alignment of primary sequences that all members of this group share a common fold.

1.4.2 Medium Chain Dehydrogenases

The medium chain dehydrogenase reductase (MDR) family also exists of a core and an extended group, with the extended group differing in the conservation pattern of certain residues. There are 98 proteins in the core group and 8 proteins in the extended group. The MDR members exhibit the following characteristics:

1. They have ~350 residues.
2. They have two domains in their tertiary structure.
3. A **GHE** sequence ~60 residues from the N-terminus in the core group.
4. A $GX_{1-3}GX_{1-3}G$ located near the middle of the primary sequence for the core group. The initial glycine residue in this motif may be replaced with an alanine residue in the extended group. This indicates the dinucleotide binding region.
5. Three strictly conserved glycine residues for the core and extended group which exist at critical turns within the tertiary structure.

Another common feature of the MDR core family is the recruitment of a divalent zinc atom to assist with catalysis. This accounts for the absence of a strict conservation of active site residues within the MDR family of enzymes. The requirement of a divalent zinc can only be established by extensive physical characterization; *i.e.* its presence cannot be established by examining the primary sequence. Nonetheless, the conservation pattern of the glycine residues suggests an overall similarity of fold for the MDR family, an observation which is borne out by an examination of six structurally determined members.

Structurally characterized members from the core group include *Equus caballus* class I alcohol dehydrogenase (71)(Figure 1.5B) and *Homo sapiens* glutathione dependent formaldehyde dehydrogenase (72). A structurally characterized member of the extended class is the *E. coli* quinone reductase (73). The structures are all similar to one another, supporting their classification into a common family. The location of conserved glycine residues at critical regions in the family of sequences can be used to predict that all of the members will have a common fold, but the data is not as compelling for MDR family as it is for the SDR family.

1.4.3 Aldo-keto Reductases

The grouping of dehydrogenases into the aldo-keto reductases (AKR) was instituted when it became clear that there were a large group

of enzymes which were homologous in both their primary and tertiary structures, and yet distinct from both the medium and long chain dehydrogenases. The characteristics of this group of 47 proteins are as follows:

1. They exhibit an overall $(\alpha/\beta)_8$ barrel fold.
2. Members of this group are approximately 320 residues in length and exist as monomers.
3. There is strict conservation of 11 residues*: Gly 22, Gly 45, Asp 50, Lys 84, Asp 112, Pro 119, Gly 164, Asn 167, Pro 186, Gln 190, and Ser 271. Another residue, Tyr 55, is replaced by a threonine residue in one of the family members but is otherwise strictly conserved.
4. The enzymes catalyze A-face hydride transfer from NAD(P)H.

The fold of the AKR family is known from the structure determination of 3α-hydroxysteroid dehydrogenase (74) and of human aldose reductase (75) (Figure 1.5C). The strong primary sequence conservation patterns support the presence of a common fold for this family. The substrate specificity observed between family members arises from two loop regions near the C-terminal region of the protein.

---

* The numbering scheme is based upon human aldose reductase.

1.4.4 2-Oxyacid Dehydrogenases

The fourth and smallest division of dehydrogenases are the 2-oxyacid dehydrogenases. This division is defined by the chemical structure of their substrates. There are two distinct families within this division: the L-isomer D-isomer specific enzymes.

The L-isomer specific enzymes are comprised of lactate (Figure 1.5D) and malate dehydrogenases. These enzymes were the first dehydrogenases to have their structures determined crystallographically (76, 77) and because of this, the first dehydrogenases to be identified as having nearly identical folds. They represent the first experimentally validated family of dehydrogenases.

The D-isomer specific family are comprised of *E. coli* D-3-phosphoglycerate dehydrogenase (79), the cucumber form of hydroxypyruvate dehydrogenase (80), the protein encoded by the *E. coli* *pdxB* gene (a putative erythronate-4-phosphate dehydrogenase) (81), and D-lactate dehydrogenase (66). These four enzymes show greater primary sequence homology among themselves than with the sequences for the L-specific enzymes. Furthermore, the structure of the D-3-phosphoglycerate enzyme (88) (Figure 1.5E) exhibited a unique fold which validates the definition of the D specific enzymes as a separate family. It remains to be seen whether the other family members exhibit structural similarity within the group.

1.4.5 Aldehyde Dehydrogenases

The aldehyde dehydrogenase family is a small group of enzymes with only 16 different members. This group was identified by Hempel *et al.* (67). Family members exhibit the following common features:

1. A primary sequence length of approximately 500 residues.
2. A set of 23 conserved residues, half of which are glycine indicating that the family members probably share a common fold. Key residues include an invariant catalytic cysteine residue at position 366 in the consensus sequence and a well conserved glutamate residue at position 327 in the consensus sequence.
3. Catalysis of aldehyde/*gem*-diol substrates to corresponding carboxylic acids *via* a thioester intermediate.

Family members include methyl malonic semi-aldehyde dehydrogenase and succinic semi-aldehyde dehydrogenase. The original work describing the aldehyde dehydrogenase family was done before structural representatives were available for this family. There are at least two structures in the PDB whose primary sequences fit into the aldehyde dehydrogenase family. These are the structures of bovine mitrochondrial aldehyde dehydrogenase (82) and cod liver betaine aldehyde dehydrogenase (83) (Figure 1.5F). The sequences for these two proteins are 39% identical and 56% similar to one another. The tertiary folds are highly similar: they exhibit a 2.0 Å r.m.s.d. between Cα atoms over 485

residues. The two proteins are nearest neighbours in the PDB as evaluated by VAST† (84) which calculated a 40.8% similarity for the two folds. This structural comparison lends credence to the grouping of the aldehyde dehydrogenases into a common family.

1.4.6 Other Dehydrogenases

Apparently, nature has evolved many ways to catalyze hydride transfer. There are several well characterized dehydrogenases which do not fall within the groups discussed above. Examples from this group of 'other dehydrogenases' include isocitrate dehydrogenase (78), glycerol dehydrogenase (85), L-alanine dehydrogenase (86), and the subject of this thesis, HSD (Figure 1.5G).

1.5 Goals of the Thesis

The primary goal of the thesis is the elucidation of the three dimensional structure of HSD from *S. cerevisiae*. The secondary goals are a molecular level understanding of the reaction mechanism catalyzed by HSD and the inhibitory action of RI-331.

---

† VAST is a computer program that enables rapid comparison at the tertiary structure level between a given model and the entire PDB.

Chapter 1 References

1. **Leyden, J. L.** (1994). *Tinea pedis* pathophysiology and treatment. J Am Acad Dermatol. **31:** S31-33.
2. **Jackson, C. J., Barton, R. C., and Evans, E. G.** (1999). Species identification and strain differentiation of dermatophyte fungi by analysis of ribosomal-DNA intergenic spacer regions. J Clin Microbiol. **37:** 931-936.
3. **Schmidt, A.** (1997). *Malassezia furfur*: a fungus belonging to the physiological skin flora and its relevance in skin disorders. Cutis. **59:** 21-24.
4. **Mori, T., Matsumura, M., Yamada, K., Irie, S., Oshimi, K., Suda, K., Oguri, T., and Ichinoe, M.** (1998). Systemic aspergillosis caused by an aflatoxin-producing strain of *Aspergillus flavus*. Med Mycol. **36:** 107-112.
5. **Nenoff, P., Gutz, U., Tinteinot, K., Bosse-Henck, A., Mierzwa, M., Hofmann, J., Horn, L. C., and Haustein, U. F.** (1996). Disseminated mycosis due to *Scedosporium prolificans* in an AIDS patient with Burkitt lymphoma. Mycoses. **39:** 461-465.
6. **Barre-Sinoussi, F., Chermann, J. C., Rey, F., Nugeyre, M. T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vezinet-Brun, F., Rouzioux, C., Rozenbaum, W., and Montagnier, L.** (1983). Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science. **220:** 868-871.
7. **Klatzmann, D., Champagne, E., Chamaret, S., Gruest, J., Guetard, D., Hercend, T., Gluckman, J. C., and Montagnier, L.** (1984). T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV. Nature. **312:** 767-768.
8. UNAIDS/WHO, Report on the Global AIDS Epidemic, UNAIDS / WHO (1998).
9. HealthCanada, AIDS in Canada: Annual Report on AIDS in Canada, , Division of HIV/AIDS Surveillence, Bureau of HIV/AIDS and STD, Laboratory Centre for Disease Control, Health Canada, Ottawa (1997).
10. **Wlodawer, A., and Erickson, J. W.** (1993). Structure-Based Inhibitors of HIV-1 Protease. Annu. Rev. Biochem. **62:** 543-585.
11. **Flexner, C.** (1998). HIV-protease inhibitors. N Engl J Med. **338:** 1281-1292.

12. **Moellering, R. C., Jr.** (1998). The specter of glycopeptide resistance: current trends and future considerations. Am J Med. **104:** 3S-6S.
13. **Holbrook, K. A., and Lowy, F. D.** (1998). Beta-lactam antibiotics. Cancer Invest. **16:** 405-412.
14. **Dennesen, P. J., Bonten, M. J., and Weinstein, R. A.** (1998). Multiresistant bacteria as a hospital epidemic problem. Ann Med. **30:** 176-185.
15. **Letvin, N. L.** (1998). Progress in the development of an HIV-1 vaccine. Science. **280:** 1875-1880.
16. **Sternberg, S.** (1994). The emerging fungal threat. Science. **266:** 1632-1634.
17. **Thomas, C. J., Lee, J. Y., Conn, L. A., Bradley, M. E., Gillespie, R. W., Dill, S. R., Pinner, R. W., and Pappas, P. G.** (1998). Surveillance of cryptococcosis in Alabama, 1992-1994. Ann Epidemiol. **8:** 212-216.
18. **Teanpaisan, R., and Nittayananta, W.** (1998). Prevalence of *Candida* species in AIDS patients and HIV-free subjects in Thailand. J Oral Pathol Med. **27:** 4-7.
19. **Pietrella, D., Monari, C., Retini, C., Palazzetti, B., Bistoni, F., and Vecchiarelli, A.** (1998). Human immunodeficiency virus type 1 envelope protein gp120 impairs intracellular antifungal mechanisms in human monocytes. J Infect Dis. **177:** 347-354.
20. **Klesper, M. E., Ernst, J. E., and Pfaller, M. A.** (1997). Update on antifungal resistance. Trends in Microbiol. **5:** 372-375.
21. **Sweet, S. P.** (1997). Selection and pathogenicity of *Candida albicans* in HIV infection. Oral Dis. **3 Suppl 1:** S88-95.
22. **Sullivan, D., and Coleman, D.** (1997). *Candida dubliniensis*: an emerging opportunistic pathogen. Curr Top Med Mycol. **8:** 15-25.
23. **Pursell, K. J., Telzak, E. E., and Armstrong, D.** (1992). *Aspergillus* species colonization and invasive disease in patients with AIDS. Clin Infect Dis. **14:** 141-148.
24. **Hardman, J. G., and Limbird, L. E.**, Goodman & Gilman's The Pharmacological Basis of Theraputics, 9th ed, , McGraw-Hill, New York (1996).
25. **Gyanchandani, A., Khan, Z. K., Farooqui, N., Goswami, M., and Ranade, S. A.** (1998). RAPD analysis of *Candida albicans* strains recovered from different immunocompromised patients (ICP) reveals an apparently non-random infectivity of the strains. Biochem Mol Biol Int. **44:** 19-27.
26. **Cartledge, J. D., Midgley, J., and Gazzard, B. G.** (1997). Clinically significant azole cross-resistance in *Candida* isolates from HIV-positive patients with oral candidosis. Aids. **11:** 1839-1844.

27. **Tumbarello, M., Tacconelli, E., Caldarola, G., Morace, G., Cauda, R., and Ortona, L.** (1997). Fluconazole resistant oral candidiasis in HIV-infected patients. Oral Dis. **3 Suppl 1:** S110-112.
28. **White, T. C., Pfaller, M. A., Rinaldi, M. G., Smith, J., and Redding, S. W.** (1997). Stable azole drug resistance associated with a substrain of *Candida albicans* from an HIV-infected patient. Oral Dis. **3 Suppl 1:** S102-109.
29. **Launay, O., Lortholary, O., Bouges-Michel, C., Jarrousse, B., Bentata, M., and Guillevin, L.** (1998). Candidemia: a nosocomial complication in adults with late-stage AIDS. Clin Infect Dis. **26:** 1134-1141.
30. **Yamaki, H., Yamaguchi, M., Imamura, H., Suzuki, H., Nishimura, T., Saito, H., and Yamaguchi, H.** (1990). The mechanism of antifungal action of *(S)*-2-amino-4-oxo-5- hydroxypentanoic acid, RI-331: the inhibition of homoserine dehydrogenase in *Saccharomyces cerevisiae*. Biochem Biophys Res Commun. **168:** 837-843.
31. **Yamaki, H., Yamaguchi, M., Nishimura, T., Shinoda, T., and Yamaguchi, H.** (1988). Drugs Expl. Clin. Res. 14. **14:** 467-472.
32. **Umbarger, H. E.** (1978). Amino acid biosynthesis and its regulation. Annu Rev Biochem. **47:** 532-606.
33. **Jacques, S.**, Characterization of Homoserine Dehydrogenase from *Saccharomyces cerevisae*: A New Fungal Target. Doctoral Thesis, *Dept of Biochemistry*, McMaster University, Hamilton, Ontario, Canada (1999).
34. **Blundell, T. L.** (1996). Structure-based drug design. Nature. **384:** 23-26.
35. **Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B. K., Baldwin, E., Weber, I. T., Selk, L. M., Clawson, L., Schneider, J., and Kent, S. B.** (1989). Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. Nature. **342:** 299-302.
36. **Navia, M. A., Fitzgerald, P. M., McKeever, B. M., Leu, C. T., Heimbach, J. C., Herber, W. K., Sigal, I. S., Darke, P. L., and Springer, J. P.** (1989). Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. Nature. **337:** 615-620.
37. **Miller, M., Jaskolski, M., Rao, J. K., Leis, J., and Wlodawer, A.** (1989). Crystal structure of a retroviral protease proves relationship to aspartic protease family. Science. **245:** 616-621.
38. **Ala, P. J., Huston, E. E., Klabe, R. M., McCabe, D. D., Duke, J. L., Rizzo, C. J., Korant, B. D., DeLoskey, R. J., Lam, P. Y., Hodge, C. N., and Chang, C. H.** (1997). Molecular basis of HIV-1

protease drug resistance: structural analysis of mutant proteases complexed with cyclic urea inhibitors. Biochemistry. **36:** 1573-1580.

39. **Black, S., and Wright, N. G.** (1953). Homoserine Dehydrogease. J. Am. Chem. Soc. **75:** 5766.
40. **Black, S., and Wright, N. G.** (1955). Homoserine Dehydrogenase. J. Biol. Chem. **213:** 51-60.
41. **Yumoto, N., Kawata, Y., Noda, S., and Tokushige, M.** (1991). Rapid Purification and Characterization of Homoserine Dehydrogenase. Arch. Biochem. Biophys. **285:** 270-275.
42. **Thomas, D., Barbey, R., and Surdin-Kerjan, Y.** (1993). Evolutionary relationships between yeast and bacterial homoserine dehydrogenases. FEBS Lett. **323:** 289-293.
43. **Yamaki, H., Yamaguchi, M., Tsuruo, T., and Yamaguchi, H.** (1992). Mechanism of action of an antifungal antibiotic, RI-331, *(S)* 2-amino-4- oxo-5-hydroxypentanoic acid; kinetics of inactivation of homoserine dehydrogenase from *Saccharomyces cerevisiae.* J Antibiot (Tokyo). **45:** 750-755.
44. **Wampler, D. E., and Westhead, E. W.** (1968). Two aspartokinases from *Escherichia coli.* Nature of the inhibition and molecular changes accompanying reversible inactivation. Biochemistry. **7:** 1661-1671.
45. **Truffa-Bachi, P., Veron, M., and Cohen, G. N.** (1974). Structure, function, and possible origin of a bifunctional allosteric enzyme, *Escherichia coli* aspartokinase I-homoserine dehydrogenase I. CRC Crit Rev Biochem. **2:** 379-415.
46. **Bearer, C. F., and Neet, K. E.** (1978). Threonine inhibition of the aspartokinase--homoserine dehydrogenase I of *Escherichia coli.* A slow transient and cooperativity of inhibition of the aspartokinase activity. Biochemistry. **17:** 3523-3530.
47. **Bearer, C. F., and Neet, K. E.** (1978). Threonine inhibition of the aspartokinase--homoserine dehydrogenase I of *Escherichia coli.* Stopped-flow kinetics and the cooperativity of inhibition of the homoserine dehydrogenase activity. Biochemistry. **17:** 3517-3522.
48. **Bearer, C. F., and Neet, K. E.** (1978). Threonine inhibition of the aspartokinase--homoserine dehydrogenase I of *Escherichia coli.* Threonine binding studies. Biochemistry. **17:** 3512-3516.
49. **Fontan, E., and Truffa-Bachi, P.** (1978). The threonine-sensitive homoserine dehydrogenase and aspartokinase activities of *Escherichia coli* K12. Carboxymethylation of a unique cysteine induces a conformational change of the enzyme. J Biol Chem. **253:** 2758-2762.
50. **Costrejean, J. M., and Truffa-Bachi, P.** (1977). Threonine-sensitive homoserine dehydrogenase and aspartokinase activities

62. **Jornvall, H., Hoog, J. O., and Persson, B.** (1999). SDR and MDR: completed genome sequences show these protein families to be large, of old origin, and of complex nature. FEBS Lett. **445:** 261-264.

63. **Jez, J. M., Bennett, M. J., Schlegel, B. P., Lewis, M., and Penning, T. M.** (1997). Comparative anatomy of the aldo-keto reductase superfamily. Biochem J. **326:** 625-636.

64. **Bruce, N. C., Willey, D. L., Coulson, A. F., and Jeffery, J.** (1994). Bacterial morphine dehydrogenase further defines a distinct superfamily of oxidoreductases with diverse functional activities. Biochem J. **299:** 805-811.

65. **Grant, G. A.** (1989). A new family of 2-hydroxyacid dehydrogenases. Biochem Biophys Res Commun. **165:** 1371-1374.

66. **Taguchi, H., and Ohta, T.** (1991). D-lactate dehydrogenase is a member of the D-isomer-specific 2- hydroxyacid dehydrogenase family. Cloning, sequencing, and expression in *Escherichia coli* of the D-lactate dehydrogenase gene of *Lactobacillus plantarum.* J Biol Chem. **266:** 12588-12594.

67. **Hempel, J., Nicholas, H., and Lindahl, R.** (1993). Aldehyde dehydrogenases: widespread structural and functional diversity within a shared framework. Protein Sci. **2:** 1890-1900.

68. **Ghosh, D., Weeks, C. M., Grochulski, P., Duax, W. L., Erman, M., Rimsay, R. L., and Orr, J. C.** (1991). Three-dimensional structure of holo 3-α,20-β-hydroxysteroid dehydrogenase: a member of a short-chain dehydrogenase family. Proc Natl Acad Sci U S A. **88:** 10064-10068.

69. **Varughese, K. I., Xuong, N. H., Kiefer, P. M., Matthews, D. A., and Whiteley, J. M.** (1994). Structural and mechanistic characteristics of dihydropteridine reductase: a member of the Tyr-(Xaa)3-Lys-containing family of reductases and dehydrogenases. Proc Natl Acad Sci U S A. **91:** 5582-5586.

70. **Labesse, G., Vidal-Cros, A., Chomilier, J., Gaudry, M., and Mornon, J. P.** (1994). Structural comparisons lead to the definition of a new superfamily of NAD(P)(H)-accepting oxidoreductases: the single-domain reductases/epimerases/dehydrogenases (the 'RED' family). Biochem J. **304:** 95-99.

71. **Eklund, H., Nordstrom, B., Zeppezauer, E., Soderlund, G., Ohlsson, I., Boiwe, T., Soderberg, B. O., Tapia, O., Branden, C. I., and Akeson, A.** (1976). Three-dimensional structure of horse liver alcohol dehydrogenase at 2.4 Å resolution. J Mol Biol. **102:** 27-59.

72. **Yang, Z. N., Bosron, W. F., and Hurley, T. D.** (1997). Structure of human χ-χ alcohol dehydrogenase: a glutathione- dependent formaldehyde dehydrogenase. J Mol Biol. **265:** 330-343.

73. **Thorn, J. M., Barton, J. D., Dixon, N. E., Ollis, D. L., and Edwards, K. J.** (1995). Crystal structure of *Escherichia coli* QOR quinone oxidoreductase complexed with NADPH. J Mol Biol. **249:** 785-799.

74. **Hoog, S. S., Pawlowski, J. E., Alzari, P. M., Penning, T. M., and Lewis, M.** (1994). Three-dimensional structure of rat liver 3 α-hydroxysteroid/dihydrodiol dehydrogenase: a member of the aldo-keto reductase superfamily. Proc Natl Acad Sci U S A. **91:** 2517-2521.

75. **Wilson, D. K., Tarle, I., Petrash, J. M., and Quiocho, F. A.** (1993). Refined 1.8 Å structure of human aldose reductase complexed with the potent inhibitor zopolrestat. Proc Natl Acad Sci U S A. **90:** 9847-9851.

76. **Adams, M. J., McPherson, A., Jr., Rossmann, M. G., Schevitz, R. W., and Wonacott, A. J.** (1970). The structure of the nicotinamide-adenine dinucleotide coenzyme when bound to lactate dehydrogenase. J Mol Biol. **51:** 31-38.

77. **Tsernoglou, D., Hill, E., and Banaszak, L. J.** (1972). Cytoplasmic malate dehydrogenase--heavy atom derivatives and low resolution structure. J Mol Biol. **69:** 75-87.

78. **Hurley, J. H., Dean, A. M., Koshland, D. E., Jr., and Stroud, R. M.** (1991). Catalytic mechanism of $NADP^+$ dependent isocitrate dehydrogenase: implications from the structures of magnesium-isocitrate and $NADP^+$ complexes. Biochemistry. **30:** 8671-8678.

79. **Tobey, K. L., and Grant, G. A.** (1986). The nucleotide sequence of the *serA* gene of *Escherichia coli* and the amino acid sequence of the encoded protein, D-3-phosphoglycerate dehydrogenase. J Biol Chem. **261:** 12179-12183.

80. **Greenler, J. M., Sloan, J. S., Schwartz, B. W., and Becker, W. M.** (1989). Isolation, characterization and sequence analysis of a full-length cDNA clone encoding NADH-dependent hydroxypyruvate reductase from cucumber. Plant Mol Biol. **13:** 139-150.

81. **Schoenlein, P. V., Roa, B. B., and Winkler, M. E.** (1989). Divergent transcription of pdxB and homology between the *pdxB* and *serA* gene products in *Escherichia coli* K-12. J Bacteriol. **171:** 6084-6092.

82. **Steinmetz, C. G., Xie, P., Weiner, H., and Hurley, T. D.** (1997). Structure of mitochondrial aldehyde dehydrogenase: the genetic component of ethanol aversion. Structure. **5:** 701-711.

83. **Johansson, K., El Ahmad, M., Hjelmqvist, L., Ramaswamy, S., Jornvall, H., and Eklund, H.**, PDB file 1A4S in RCSB PDB (unpublished), (1998).

84. **Madej, T., Gibrat, J. F., and Bryant, S. H.** (1995). Threading a database of protein cores. Proteins. **23:** 356-369.

85. **Leichus, B. N., and Blanchard, J. S.** (1994). Isotopic analysis of the reaction catalyzed by glycerol dehydrogenase. Biochemistry. **33:** 14642-14649.
86. Grimshaw, C. E., Cook, P. F., and Cleland, W. W. (1981). Use of isotope effects and pH studies to determine the chemical mechanism of *Bacillus subtilis* L-alanine dehydrogenase. Biochemistry. 20: 5655-5661.
87. **Wilson, D. K., Bohren, K. M., Gabbay, K. H., and Quiocho, F. A.** (1992). An unlikely sugar substrate site in the 1.65 Å structure of the human aldose reductase holoenzyme implicated in diabetic complications. Science. **257:** 81-84.
88. **Schuller, D. J., Grant, G. A., and Banaszak, L. J.** (1995). The allosteric ligand site in the V_{max}-type cooperative enzyme phosphoglycerate dehydrogenase. Nat Struct Biol. **2:** 69-76.

Chapter Two: The Growth and Data Collection of Native HSD Crystals

The crystallization conditions for HSD were published (1) as a crystallization note in *Acta Crystallographica D* during the course of the thesis work. An abridged version follows.

2.1 Excerpts From Published Conditions

2.1.1 Purification and Crystallization of Homoserine Dehydrogenase.

Homoserine dehydrogenase from *S. cerevisiae* was overexpressed in *E. coli* and purified using procedures described elsewhere (4). Prior to crystallization, the protein was concentrated to an optical density of 9.8 at 280 nm, corresponding to a protein concentration of approximately 10 mg/ml, in either 10 mM HEPES (pH 8.5) or 10 mM Tris-HCl buffer (pH 8.5). Homogeneity was tested by Coomassie and silver stained SDS-PAGE, as well as iso-electric focusing, which indicated that the protein was at least 99% pure.

Crystallization experiments were performed using the hanging drop vapour diffusion method (3). The reservoirs contained 1 ml of precipitant solution and the drops contained 5 µl of a 1:1 mixture of protein solution

and precipitant. Three different groups of experiments were carried out: crystallization of the apo-enzyme, crystallization of the protein in the presence of five-fold molar excess of the co-factor NAD^+, and crystallization of the protein in the presence of five-fold molar excess of the product L-homoserine and the inert NAD^+ analogue 3-aminopyridine adenine dinucleotide (Sigma). For the latter co-crystallization experiment, the product of the reaction catalyzed by homoserine dehydrogenase, L-homoserine, was used instead of the substrate, aspartate semi-aldehyde, since aspartate semi-aldehyde is an unstable compound. The NAD^+ analogue used is a competitive inhibitor of HSD with a K_i of 0.22 ± 0.06 mM (*cf.* NAD^+ K_m of 0.052 ±0.005 mM (4)). Initial crystallization conditions were obtained from a sparse matrix screening of conditions favourable for protein crystal growth (Hampton) (5). Optimal conditions were subsequently determined by varying precipitant and buffer conditions within narrow ranges. Under these conditions, crystals would grow within one to two weeks. Noteworthy in this respect is that the quality of the crystals grown in the presence of NAD^+ was significantly improved by carefully controlling the temperature during crystallization. Without proper temperature control (22 ± 4° C) crystals would only diffract to 3.0 Å resolution. However, by performing the crystallization experiments in a low-temperature incubator (22 ± 0.1° C) the diffraction limit of these crystals could be increased to 2.2 Å resolution.

2.1.2 Data Collection and Processing

Crystals of the apo-enzyme were mounted in thin wall glass capillaries with a small amount of mother liquor. Data collection was performed at room temperature. Crystals of homoserine dehydrogenase grown in the presence of product and/or co-factor (analogue) were soaked briefly in a suitable cryo-protectant and then flash frozen in a stream of cold nitrogen. Data collection for these crystals was performed at 180 K.

The X-ray diffraction data was collected with an Raxis IIc imaging system mounted on a Rigaku RU200B rotating copper anode with power settings of 50 kV and 60 mA. The X-ray beam was passed through a nickel filter and a Supper double mirror focusing system such that the Cu K_α radiation was selected. A 0.3 mm collimator was further used to define the beam. Typical oscillation angles used for data collection were 1.0-1.5°. Data processing was carried out with the Denzo and Scalepack data processing programs (6). Data manipulations were performed with programs from the *CCP4* package (7).

2.1.3 Results and Discussion

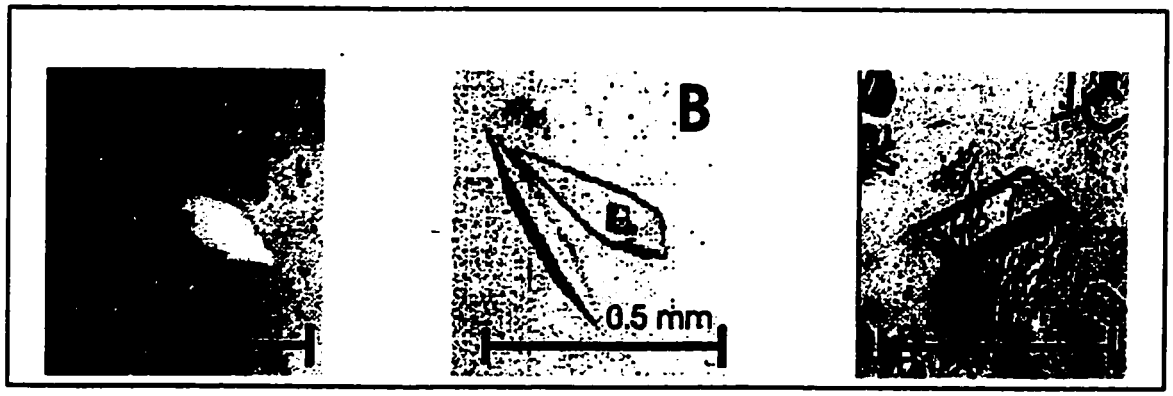
As shown in Table 2.1, crystals of the apo-enzyme diffracted poorly* in comparison to those grown in the presence of L-homoserine and/or

---

* This was further verified after publication by exposure to a more intense X-ray beam at the National Synchrotron Light Source beamline X-8C. Diffraction for this crystal form grown under the conditions listed here does not extend beyond 3.9 Å.

NAD^+ (analogue). Specifically, crystals grown in the presence of NAD^+ diffracted to 2.3 Å[†] resolution. However, due to the long cell-dimensions along the c axis (250.2 Å) and the pixel size of the detector, data collection was complicated and integratable data could only be obtained to a resolution of 2.7 Å.[‡] The crystals are shown in Figure 2.1.

Figure 2.1 - Three different crystal forms of homoserine dehydrogenase.



(A) A crystal of the apo-enzyme. (B) A tetragonal crystal of the enzyme grown in the presence of NAD^+. (C) A monoclinic crystal of homoserine dehydrogenase grown in the presence of L-homoserine and an inert analogue of NAD^+.

[†] Diffraction was observed to 1.9 Å when these crystals were exposed to a more intense X-ray beam at the National Synchrotron Light Source beamline X-8C.

[‡] This was rectified after publication by employing a data collection strategy whereby two data collection passes were made. The first was done with the detector 200 mm from the crystal (low resolution data) and a second with the detetector 120 mm from the crystal (high resolution data). The two datasets were processed separately and then merged together. Rejection criteria were employed such that overlap data in the high resolution data set were not included in the final data set. Such a strategy allowed for a dataset to 2.3 Å to be processed in a satisfactory manner.

Table 2.1 - Crystallization conditions and X-ray diffraction data statistics for the different crystal forms of homoserine dehydrogenase (HSD)

Species	**HSD**	**HSD•NAD$^+$**	**HSD•L-homoserine•NAD$^+$ analogue**
Crystallization conditions[§]	2.0 M ammonium sulfate Tris-HCl pH 8.5	2.2 M (*2.5M*) ammonium sulfate	15-20% (*30%*) PEG 8000
(added/altered in cryo-protectant)		0.1 M sodium acetate pH 4.4	0.1M sodium cacodylate
		(*saturated glucose*)	pH 6.5
Temperature during Data Collection	298 K	180 K	180 K
Space group[**]	P4	$P4_12_12$ / $P4_32_12$	$P2_1$
Unit cell parameters	$a = b = 130$ Å, $c = 240$ Å	$a = b = 80.4$ Å, $c = 250.2$ Å	$a = 58.8$ Å, $b = 104.2$ Å, $c = 120.7$ Å, $\beta = 91.9°$
Molecules per asym. Unit[††]	12	2	4
Diffraction limit	3.9 Å	2.3 Å	2.6 Å
R_{sym} [#](last shell)	-	0.077 (0.492)	0.099 (0.344)
Unique Reflections Measured (Multiplicity)	-	32166 (5.1)	44116 (3.7)
Completeness (Last shell)	-	98.4 (94.8)	99.2 (98.1)

[§] The hanging drop method was used. The conditions of the reservoir solutions are shown.

[**] A preliminary assignment based upon symmetry and systematic absences observed in the reduced data. The apo-enzyme space group assignment was based upon data from a single oscillation frame.

[††] Number of molecules was determined by comparing the Matthews coefficient to typical values for protein crystals.

[#] R_{sym} value is based on intensities. For the apo-enzyme the complete data set was not collected due to the limited resolution

Analyses of the Matthews coefficients (8) indicated that one, two or six homoserine dehydrogenase dimers were present in the asymmetric unit, dependent on the crystal form. This implied that non-crystallographic symmetry should be present in our diffraction data. However, no orientations for a non-crystallographic symmetry axis could be identified from the self-rotation function calculated with data collected from the crystals which likely have one dimer in the asymmetric unit, *i.e.* crystals grown in the presence of NAD^+. The same observations were made for crystals which likely have two dimers in the asymmetric unit, *i.e.* crystals grown in the presence of L-homoserine and the co-factor analogue.

2.1.4 Findings Made After Publication of the Conditions

The most important finding made after the conditions were published was the beneficial effect of Ca^{2+} ions upon the crystallization of HSD in its tetragonal form. It is not necessary to include Ca^{2+} ions to obtain crystals. However, the presence of the ion attenuated the crystal to crystal nonisomorphism. Table 2.2 is a representative summary of the HSD native data sets. Included here are those datasets which diffracted to >3.5 Å resolution and were collected to better than 90% completion. The R_{merge} statistics are for the comparison against native dataset 2 which was ultimately used to generate the initial phasing model. An examination of Table 2.2 reveals that the unweighted crystal to crystal R_{merge} was improved from values in the range of 14% or higher to values on the order

of 7-11% when Ca^{2+} was included in the crystallization conditions at an initial concentration of ~10 mM. Although this is slightly above the ideal figure of 3-4% usually suggested for crystals to be used for isomorphous replacement studies, it is still low enough to detect well substituted derivatives and thereby obtain useful phasing information. R_{sym}[§§] values for these datasets were variable, ranging from 7% up to 18%. The individual R_{sym} values do not correlate with the R_{merge} values. All datasets diffracted to a minimum 3.5 Å resolution and were collected at 180 K using the cryo conditions described in Table 2.1.

Table 2.2 – The Effect of Ca^{2+} Ions Upon Tetragonal Native Crystal Isomorphism

Native Data Set	a (Å)	c (Å)	R_{merge}[•••]	Calcium Added
1[†††]	80.4	250.2	0.112	no?
2[‡‡‡]	81.0	249.7		yes
3	81.1	250.2	0.230	no
4	80.6	249.5	0.096	yes
5	81.3	249.9	0.107	yes
6	81.3	249.4	0.069	yes
7	80.9	249.8	0.177	yes
8	81.5	250.1	0.182	no
9	80.3	249.8	0.167	no
10	80.8	250.1	0.136	no

[§§] R_{sym} is defined as $\Sigma | F_{+} - F_{-} | / \Sigma F_{ave}$ where $F_{+/-}$ are theoretically equivalent reflections.

[•••] R_{merge} is defined as $\Sigma | F_{Native1} - F_{Native2} | / \Sigma F_{Native2}$ where the structure factors are based solely upon the observed intensities. Native 1 in all cases in dataset 2. A 3σ cutoff was used to generate these numbers.

[†††] This was the high resolution data set used for refinement of the tetragonal crystal structure

[‡‡‡] This is the data set which produced the best phasing statistics for the generation of initial electron density maps (see following chapter).

The conditions in which the tetragonal crystals form grow involve ammonium sulphate. Since $Ca(SO_4)$ has a moderately low K_{sp} value of $2.45 * 10^{-5}$ M^2, it was inevitable that some of the calcium would either precipitate or crystallize out in the sulfate form. This was indeed observed: colourless crystals with a hexagonally shaped rod habit could be observed within 1-2 days of setting up the crystallization experiments. The standard rod shaped and occasionally block shaped crystals of NAD^+ bound HSD would appear within three weeks after the appearance of the $Ca(SO_4)$ crystals. Replacement of $CaCl_2$ with $Ca(OAc)_2$ resulted in similar growth patterns for the crystals, suggesting that only the Ca^{2+} ions were responsible for the improvement in isomorphism.

Unfortunately, the result of adding Ca^{2+} to the crystallization condition was not entirely reproducible, as shown by the high R_{merge} statistic of native dataset #7. This type of irreproducibility was characteristic of the HSD crystals in general, and contributed to the difficulties encountered during the search for heavy atom derivatives discussed in the following chapter.

2.2 Summary of Chapter 2

HSD from *S. cerevisae* was crystallized in three different forms: one tetragonal form in the absence of cofactor, a second tetragonal form in the presence of cofactor, and a monoclinic form in the presence of cofactor analogue and product. Crystallographic data was collected for

the latter two forms and processed to generate useful datasets, to 2.3 Å and 2.6 Å respectively.

The Matthews number indicated that the second tetragonal crystal form has two HSD molecules per asymmetric unit and that the monoclinic crystal form has four HSD molecules per asymmetric unit.

Finally, it was demonstrated that the introduction of Ca^{2+} ions into the crystallization medium improved the quality of the tetragonal crystal growth.

Chapter 2 References

1. **DeLaBarre, B., Jacques, S. L., Pratt, C. E., Ruth, D., Wright, G. D., and Berghuis, A. M.** (1998). Crystallization and Preliminary X-ray Diffraction studies of Homoserine Dehydrogenase from *S. Cerevisiae*. Acta Cryst. **D54:** 413-415.
2. **Jacques, S.**, Personal Communication, (1999).
3. **McPherson, A., Jr.** (1976). The growth and preliminary investigation of protein and nucleic acid crystals for X-ray diffraction analysis. Methods Biochem Anal. **23:** 249-345.
4. **Jacques, S.**, Characterization of Homoserine Dehydrogenase from *Saccharomyces cerevisae*: A New Fungal Target. Doctoral Thesis, *Dept of Biochemistry*, McMaster University, Hamilton, Ontario, Canada (1999).
5. **Jancarik, J., and Kim, S.-H.** (1991). J. Appl. Cryst. **24:** 409-411.
6. **Otwinowski, Z., and Minor, W.**, Processing of X-ray Diffraction Data Collected in Oscillation Mode, in *Methods in Enzymology, Volume 276: Macromolecular Crystallography, part A*, C.W. Carter, J. R. M. S., Ed., Academic Press, pp. 307-326 (1997).
7. **CCP4.** (1994.). The CCP4 Suite: Programs for Protein Crystallography. Acta Cryst. D50. 760-763.
8. **Matthews, B. W.** (1968). J. Mol. Biol. **33:** 491-497.

Chapter Three: Phase Determination

3.1 The Initial Phase Model

3.1.1 Choice of Crystal System for Heavy Atom Derivatives

Of the two well diffracting crystal forms, the tetragonal (NAD^+ bound) crystals were easier to reproduce. They were therefore chosen as native crystal form for heavy atom soaking studies. An additional benefit of using the tetragonal form was that data could be collected more quickly on potential derivatives. Their diffraction pattern exhibits P4/mmm symmetry; thus only 45° of data was necessary to generate a complete dataset. The monoclinic system on the other hand, has a diffraction pattern which exhibits P2/m symmetry; this required 180° of data to generate a full data set.

Although it was not known at the outset, it was necessary to collect most of the unique data from derivative crystals in order to identify them as such. In general, derivative crystals can be identified from R_{merge} statistics based on a fraction of the unique data (< 10° of data collection (1)). For the tetragonal HSD crystals, non-isomorphism was problematic. The only way a useful derivative could be identified was by examining

difference Patterson maps. In order to generate reliable difference Patterson maps, a substantial portion of the unique data must be collected.

3.1.2 Heavy Atom Searches

Approximately 100 conditions comprising of a wide variety of heavy atom salts and soaking conditions were tested until a useful heavy atom derivative was found. Cryo-conditions used for the native crystal data collection were used during the heavy atom search. Weighted R_{merge} values ranged from 0.014 to 0.323; unweighted R_{merge} values were higher (see Figure 3.1). Only the unweighted R_{merge} statistics approached theoretical values for actual derivatives. Of all the data sets collected, only one of the derivatives produced an interpretable difference Patterson map. The derivative which produced this map was created by soaking a crystal for 11 days at pH 6.0 in 2 mM $KAuCl_4$. The composition of the solution was otherwise similar to the mother liquor: 3.0 M ammonium sulfate, 80 μM NAD^+, and 1 mM $CaCl_2$. The appropriate Harker sections for this map are shown in Figure 3.2. The strong peak observed in the z=1/2 Harker section was the first indication that this particular derivative would produce useful phasing information. Weak to medium peaks appeared in congruous positions on the other Harker sections.

3.1.3 Patterson Map Interpretation

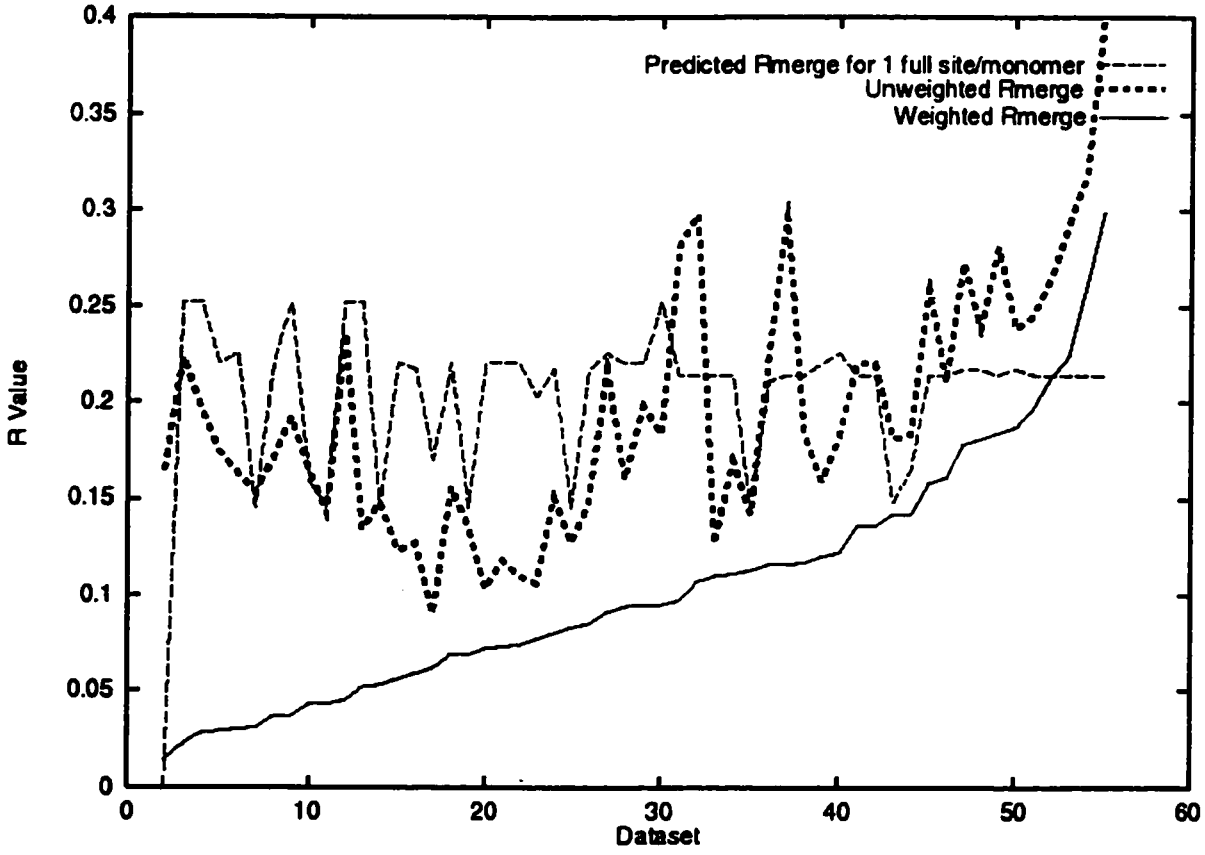
Because of the noise in the difference Patterson map, an automated Patterson interpretation routine was used to ensure that the appropriate cross-vectors were selected. The algorithm implemented in HASSP in the SOLVE suite (2) was able to interpret the difference Patterson maps sufficiently to yield 3 initial heavy atom sites. The SOLVE package also uses local scaling, which probably helped to further reduce the noise in the difference Patterson maps. Local scaling is a resolution dependent reciprocal space weighting scheme. The effect of local scaling is a reduction in the systematic errors made during the data collection.

The heavy atom positions established by the SOLVE package were further refined and ultimately verified by their phasing power and R_{Cullis}* statistics as evaluated by MLPHARE (3), a heavy atom phasing and refinement program which employs approximate maximum likelihood techniques to refine heavy atom positions.

---

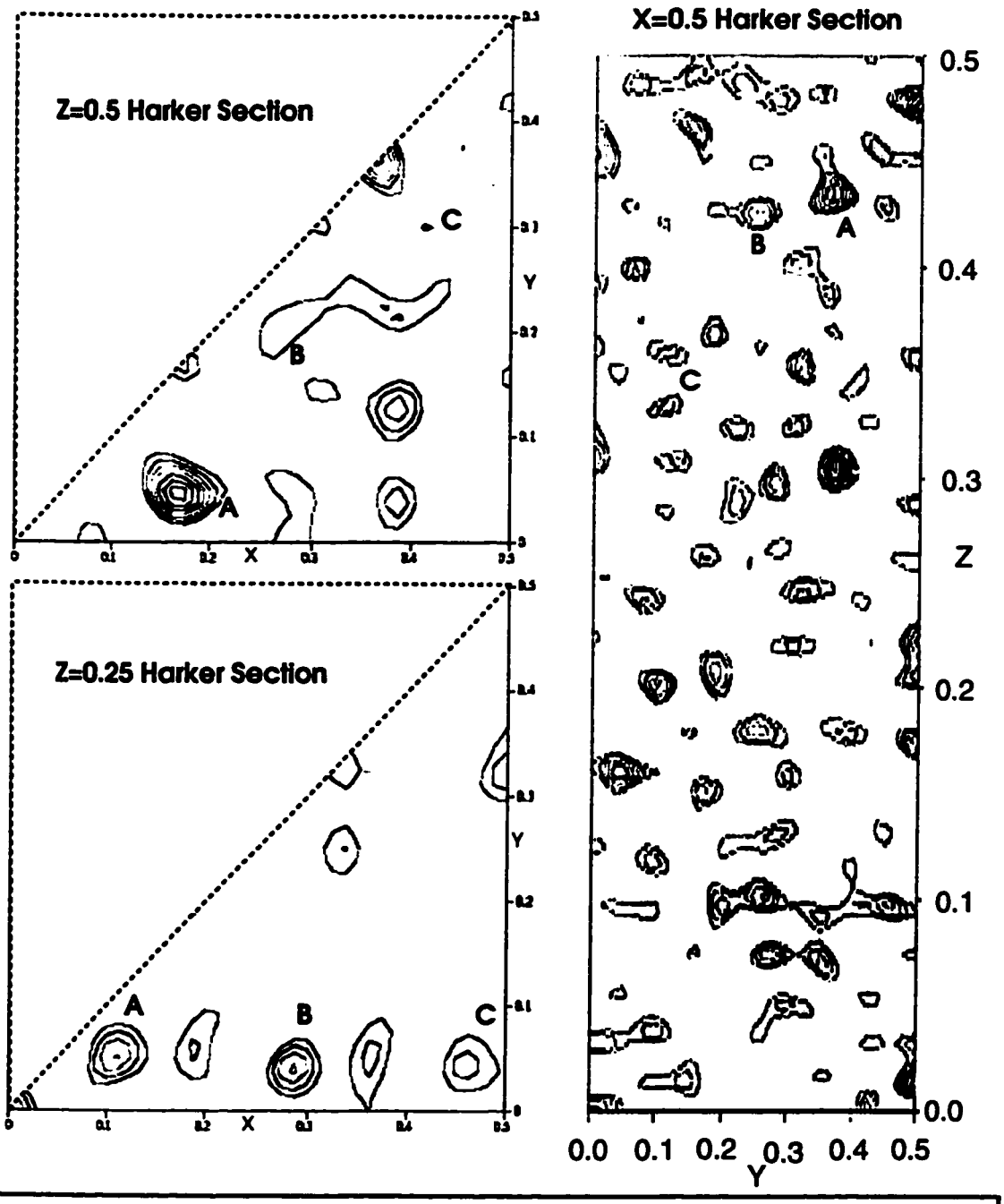
* R_{Cullis} in MLPHARE is the ratio of the lack of closure to the isomorphous difference.

Figure 3.1 – Summary of Derivative R_{merge} Values



The weighted and unweighted merging statistics based on amplitudes from a selection of derivatives are plotted. All merging statistics are based on the low resolution tetragonal form. Weighted R_{merge} statistics take into account the errors measured for the respective data sets. The predicted R_{merge} (Crick, F. H. C., and Magdoff, B. S. (1956). Acta Crystallogr. 9: 901-908) for single heavy atom substitution is also shown, using the atomic number appropriate for each individual dataset.

Figure 3.2 - Harker Sections of Difference Patterson Map from Gold Derivative



Three Harker sections from the difference Patterson map of the gold derivative *vs.* the low resolution tetragonal crystal form. The Patterson function was calculated with amplitudes better than 2σ for each crystal form. The contour levels show the Patterson peaks at 0.5σ levels, starting at 1.5σ. The three heavy atom positions derived from this map are labeled A,B, and C respectively.

Double difference Fourier maps were calculated according to Box[†] 3.3 as implemented in FFT (4) of the CCP4 program suite (5). The maps revealed an additional heavy atom binding site yielding a total of 4 sites as an initial phasing model.

Box 3.3 – The Double Difference Fourier Transformation

The Fourier transform is:

$\rho = \Sigma\Sigma\Sigma$ (terms) (summation over all hkl)

The double difference is carried out by using the following as terms in the Fourier transform equation:

$$w(F_{PH(I)} - (F_P + F_H))$$

where:
ρ=electron density
w=figure of merit
$F_{PH(I)}$ are the measured heavy atom structure factors
F_P are the measured native structure factors
F_H are the calculated structure factors for the known heavy atom positions

3.1.4 Initial Electron Density Maps

Electron density maps generated from the 4 atom SIR gold phasing model were not easily interpreted because of the remaining phase

---

[†] Boxes will be numbered within the same sequence as Figures throughout the document.

ambiguities which arose from using a single derivative. Density modification techniques of solvent flattening, histogram matching, and phase extension were applied to the 4 position Au phasing model. The DM program (6) in the CCP4 suite was used to apply the density modification. The electron density maps generated from a Fourier transform of the phased structure factors exhibited some interpretable features; specifically, backbone atoms in α helical conformations were clearly visible.

The appearance of α helices is a useful indicator since unlike protein solvent boundaries, the presence of helices is not imposed in any of the density modification techniques. Furthermore, helices can be used to identify the proper enantiomorph of the two potential space groups for the tetragonal crystals. Until this point, it was impossible to determine whether the crystals belonged to $P4_12_12$ or the $P4_32_12$ space group. The proper chirality of the helices is observed only if the true enantiomorphic space group is chosen. The space group had been arbitrarily assigned as $P4_12_12$, but it actually was $P4_32_12$. Thus the α-helices appeared as left handed helices in the first maps. The situation was remedied by reindexing the data in $P4_32_12$ and applying an inversion operator to the heavy atom positions.

3.1.5 Additional Heavy Atom Derivatives

Although some secondary structure elements could be observed in the electron density maps generated from the density modified 4 position gold phases, the maps were still of poor quality. Only a few short helices could be clearly identified, and continuous β sheets could not be modeled definitively. Therefore, the datasets collected from previously attempted derivatives were reexamined.

The phasing information obtained from the gold derivative was used to identify other derivatives. Patterson maps which are not readily interpretable do not necessarily imply an absence of heavy atoms. R_{merge} statistics (see Figure 3.1) had indicated many crystals which may have been slight derivatives and could have yielded phase information if their heavy atom positions were located.

Although protein features cannot be readily observed in electron density maps arising from a poor phasing model, the electron rich heavy atom positions in other derivatives can be located with sufficient accuracy to generate further phase information. The presence of heavy atoms in other derivatives can be identified by examining difference Fourier difference maps, calculated according to the equation in Box 3.4.

Box 3.4 - Terms for Difference Fourier Map

The Fourier transform is calculated in a manner analogous to that shown in Box 3.3 with the following terms:

$$w(F_{PH2}-F_{OBS})$$

where:
w=figure of merit
F_{PH2} are the measured structure factor amplitudes for the putative derivative
F_{OBS} are the measured native structure amplitudes with phases calculated from the first derivative

The main difficulty of this approach lies in establishing whether peaks found in the difference maps with coordinates similar to the heavy atom positions from the first derivative represent actual density or are artifacts of phase bias arising from the starting heavy atom model. Such ambiguities can be resolved by treating the questionable peaks as real heavy atoms and subjecting them to refinement in the absence of information from the original derivative. If the occupancy of the heavy atoms drops to zero during the refinement, then the peaks were simply a result of phase bias. The ambiguous sites cannot be ignored because all major heavy atom positions must be identified to get useful phases from a derivative. If all of the positions are identical to the original or any other heavy atom model, the contribution of the derivative in question to the

overall phasing will be slight. This is an important point to consider because if the derivative adds more noise than signal to the overall model, then additional heavy atom derivatives will be detrimental to the quality of the resultant electron density maps.

To minimize bias from the gold heavy atom model, the density modified phases were used to generate the difference Fourier maps. The two strongest derivatives found by this method were from K_2PtCl_4 and PIP (di-μ-iodobis(ethylenediamine) diplatinum (II) nitrate). These derivatives were created by soaking crystals in conditions similar to those which created the Au derivative. The only difference was the nature of the heavy atom used in the soaking conditions.

Many combinations of heavy atom data sets were tried before this optimal combination was discovered. MLPHARE was used to refine the positions determined from the Fourier difference maps and tc calculate phasing statistics. R_{Cullis} and phasing power[‡] statistics were used to determine which combination of derivatives and heavy atom sites would be the most effective.

Each derivative had three sites, some in common with each other and some in common with the Au derivative. There was a total of 5 unique sites among the three derivatives. This phasing model will hereafter be

---

[‡] Phasing power in MLPHARE is defined as the ratio of the heavy atom structure factor to the lack of closure.

referred to as the 3/5 MIRAS model. The statistics are tabulated in Table 3.1.

Table 3.1 – The 3/5 MIRAS Model

Derivative Compound	**Occupancy**	**Anomalous Occupancy**	***Site #* (fractional coordinates)**
$KAuCl_4$	0.6	0	*1* (0.62, 0.17, 0.09)
	0.5	0	*2* (0.91, 0.48, 0.08)
	0.5	0	*3* (0.24, 0.20, 0.03)
	0.3	0	*4* (0.93, 0.22, 0.08)
K_2PtCl4	0.2	0.2	*5* (0.22, 0.02, 0.06)
	0.6	0.5	*1*
	0.5	0.5	*3*
PIP	0.6	0.3	*5*
	0.8	0.4	*1*
	0.7	0.4	*3*

Data for the platinum derivatives was collected using the inverse beam method with a wavelength set to optimize the anomalous signal. The stable refinement of the anomalous occupancy in MLPHARE was taken as evidence for the existence of the anomalous signal. The anomalous signal is a useful source of phase information because it is orthogonal to the phasing information obtained from the isomorphous replacement signal.

Refinements in MLPHARE followed by Fourier difference maps with various combinations of derivative information present were performed until no further sites could be found with the additional heavy atom

derivatives. It was necessary to constrain B factors during these refinements. At the time, noisy data was thought to have necessitated B factor constraint. In retrospect, it is more likely that errors in the heavy atom model forced the B factor constraint.

Statistics for the individual heavy atom data sets are given in Table 3.2. The merging statistics with both the high and low resolution data sets are shown in Tables 3.3 and 3.4. The normal probablility (NP) values were calculated according to the methods given in (7).

Table 3.2 - Datasets used for phasing information						
Dataset	**Res[‡] (Å)**	**R_{sym}[§] (last shell)**	**a (Å)**	**c (Å)**	**Completion (last shell) %**	**λ[••]**
Native 1[††]	2.6	.07 (.25)	80.4	250.2	99 (98)	1.542
Native 2	3.2	.18 (.39)	81.0	249.7	98 (82)	1.542
$AuCl_3$	3.4	.19 (.41)	81.3	249.7	97 (96)	1.542
K_2PtCl_4	3.2	.10 (.16)	81.5	248.4	87 (58)	1.072
PIP	3.4	.15 (.22)	80.7	248.2	95 (85)	1.072

Table 3.3 - Derivative Data *vs.* Native 1						
Crystal	**R_{merge}**	**D_{iso}[‡‡] (Max)**	**D_{ano}**	**NP[§§]**	**R_{Cullis}**	**Phase**
$AuCl_3$	.27 (.06)	152 (1002)	0	6.2	.80	1.6
K_2PtCl_4	.32 (.16)	174 (991)	48 (640)	10.6	0.97	0.8
PIP	.39 (.17)	215 (1490)	52 (674)	9.5	1.0	0.7
Native 2	.15 (.04)	80 (470)	-	3	n/a	n/a

Table 3.4 - Derivative Data *vs.* Native 2						
Crystal	**R_{merge} (wR_{merge})**	**D_{iso} (Max)**	**D_{ano} (Max)**	**NP**	**R_{Cullis}**	**Phase Power[†††]**
$AuCl_3$	.24 (.18)	118 (891)	0	3.8	0.57	2.7 (*1.5*)
K_2PtCl_4	.29 (.19)	141 (1022)	43.4 (446)	5.9	*Not meas.*	*0.81*
PIP	.36 (.29)	181 (1395)	48.7 (607)	6.7	*Not meas.*	*1.1*

[‡]Resolution cutoff determined by shell where I/sigI < 3 for 50% of data

[§] Based on amplitudes.

[••] The shorter wavelength was collected at the CHESS F2 beamline. The higher wavelength datasets were collected on the home source described in chapter 2.

[††] This is the same native dataset discussed in chapter 2 truncated to a lower resolution. Only the single pass data was used to generate this dataset.

[‡‡] $D_{ano/iso}$ mean isomorphous/anomalous differences for all resolution shells.

[§§] Normal probability value (7) obtained from SCALEIT in CCP4. Higher numbers indicate better derivatives in situations where nonisomorphism is not problematic.

[•••] Obtained by individually refining the derivative in SHARP. Statistics falsely indicated that the phasing model was improved when all derivatives were refined together. *N.B.* the phasing power is defined slightly differently from MLPHARE, *i.e.* the values are not directly comparable.

[†††] Values in smaller font italics were derived from MLPHARE refinement.

Combining the information from the 3 derivatives resulted in phasing statistics that were reasonable. Secondary structure elements could be seen in solvent flattened maps. Refinements in MLPHARE gave excellent statistics for lower resolution (up to 5 Å) and reasonable statistics for higher resolution data (up to 3.4 Å). When the heavy atom model was refined in SHARP (8), a heavy atom program which uses a more rigorous implementation of the maximum likelihood approach, better statistics were obtained and more importantly, the resulting electron density maps showed more recognizable features. Phases were transferred to the high resolution dataset, as it was assumed that density modification techniques could be used to improve and extend the phase model. In retrospect, this was an error. A better phasing model was obtained with the lower resolution dataset (*vide infra*).

3.2 Improvement of the Initial Phase Model

3.2.1 Determination of Non-crystallographic symmetry

The Matthews coefficient for the tetragonal form (Table 2.1) was consistent with the presence of a dimer in the asymmetric unit, an observation in agreement with previous characterization of the enzyme (9). This implied that there was a symmetry operation which would describe the relation of the two protomers within the asymmetric unit.

Since the structure factors describe the protein structure, the same symmetry operation could be used to relate certain sets of the structure factors to one another. The crystallographic symmetry operators were eliminated during the data processing step, but the protomer relating symmetry operator was still present and could have describe the structure factors derived from the 3/5 MIRAS phase model.

The protomer relation is referred to as non-crystallographic symmetry (NCS) because it cannot generate all of the crystallographic copies of an entity within the unit cell, *i.e.* it is a local rather than a global operation. The NCS operator is a useful piece of information to have because it can be used at several different stages during the process of a structure determination:

1. NCS can be applied to the initial structure factors derived from heavy atom phasing to generate an improved map.
2. NCS can be used during model building. The NCS operator describes how the monomers are related to one another; *e.g.* for a dimer described by strict NCS, it is only necessary to construct one monomer and then apply the NCS operator to generate the entire dimer.
3. The NCS operator can be used as either a constraint or a restraint during refinement. This can improve the data to parameter ratio which will in turn make it easier to find the best model for the data. The ratio

can be improved by a factor of 2 for the case of strict NCS describing a dimer.

The NCS operator is composed of two parts: it has a rotational and a translational component. The rotational component is obtained first by one of the following methods:

1. The rotational symmetry can be observed in the diffraction data before phases are determined. Patterson functions are used in self-rotation searches to determine the rotational component.
2. The heavy atom model may exhibit the NCS symmetry. For this to provide unambiguous answers, each monomer must be substituted in three sites *i.e.* a total of 6 sites are required for a dimer.
3. If the unaveraged maps are good enough, a crude model can be constructed. The model can then be rotated and translated onto itself to establish the NCS operator.

The translational component can be determined by using the determined rotational component in phased correlation searches once initial phases are available. If the NCS is purely translational, phase information is not required as the operation can be observed in the native Patterson maps.

The 3/5 MIRAS phase model was used to search for the NCS operator. Unfortunately, none of the routes described above were effective for the HSD tetragonal data: the rotational component could not be found in the diffraction data, there were only 5 heavy atom sites, and the unaveraged maps could not be interpreted to produce a model good enough to establish the NCS. This latter point was especially unsettling for obvious reasons; not only was the model insufficient to determine NCS, it was insufficient to give any indications as to how the various secondary structural elements could be connected; thus model building at this stage was unfeasible. Because of the poor quality of the maps and the difficulties of obtaining useful derivatives, considerable effort was put into trying to establish the NCS operator.

3.2.1.1 A Novel Method for Determining the NCS Rotational Component

In the process of trying to elucidate the NCS operator for HSD, a new method was generated. The new method was inspired by the tactics used by Stein *et al.* to locate the NCS for the pertussis toxin structure (10). Briefly, Stein *et al.* extracted spheres of density from their initial maps and placed the spheres into triclinic cells using a transformation described by Main *et al.* (11). Cross-rotation searches between the Patterson functions for the sphere and the entire maps generated from the original crystal form were undertaken until the rotational operator could be found. Phased

correlation searches then revealed the translational component to give the entire NCS operator.

The method described by Stein *et al.* relied on the ability to locate the protein dimer with the unaveraged unique structure factors. This was not directly possible with the initial phasing model for HSD. With the MIRAS phased electron density maps of HSD it was unclear where each of the monomers were located because crystallographic contacts had obfuscated the true molecular boundaries.

The problem was overcome with the Wang-Leslie solvent masks (12) which had been established during solvent flattening in the density modification process. The solvent mask is the barrier between regions of high electron density and regions of low electron density in the electron density maps. The high regions are assumed to be ordered protein atoms and the low regions unordered solvent.

An assumption was made that the crystallographic (interdimer) contacts did not individually represent as much surface area as the noncrystallographic (intradimer) contacts. The solvent mask was gradually shrunk by an algorithm implemented in the MAMA program of the RAVE suite (13). The algorithm is described briefly as follows. Solvent masks are represented computationally as grids on a 0.9 Å grid (~1/3rd of the high resolution limit) with points set to a value of 1 for protein and a value of 0 for solvent. The mask is actually the boundary

between grid points with a value of 0 and grid points with a value of 1. The algorithm entailed examining all neighbouring grid points for each point within the mask. If the grid point had a neighbour which was not in the mask, it was changed to a value of 0 (*i.e.* changed from protein to solvent). This algorithm effectively peels away a layer of the mask while maintaining most of the topographical features. One way of visualizing the effect of the algorithm is to imagine that the unit cell was slowly flooded with excess solvent. The artificial excess solvent covered the narrower regions of connective density, leaving behinds 'islands' which represented the thicker regions of electron density.

After each round of contraction, the masks were examined on a graphics terminal. After five rounds of contraction, separate pieces of mask, or 'islands' could be observed in the solvent masks. The islands corresponded to the location of electron dense regions of the map. It was presumed that among the electron dense regions was the elusive location of the HSD dimer.

A sphere of density centred on one of these locations was extracted and placed into a triclinic cell (space group P1) with the transformation described by Main *et al.* (11) and implemented in GHKL (14). The particular 'island' was chosen because it was completely contained within the asymmetric unit. A radius of 28 Å was used to describe the sphere of data to be transformed because it was the

maximum radius describing the 'island' of density which did not include density from neighbouring 'islands'.

Patterson based self-rotation functions performed on the structure factors from the triclinic cell gave a clear answer, as shown in Table 3.5.

Table 3.5 – Solutions from NCS Rotation Search

Answer #	Alpha	Beta	Gamma	$\sigma^{\ddagger\ddagger\ddagger}$
1	0.0	0.0	0.0	16.6
2	349.3	161	190.7	7.8
3	0.0	180.0	269.6	7.1
4	54.0	139.3	126.0	5.6

The rotational search covered $0 \leq \alpha,\gamma \leq 360$ and $0 \leq \beta \leq 180$ and was done in increments of 5° with a upper limit of 28 Å for the Patterson vectors. Data in the resolution range of 15 to 6 Å were used for the search. The identical search using structure factors generated from the complete native data set resulted in answers which corresponded only to crystallographic symmetry operators. Answer 1 in Table 3.5 is the origin peak, answer 2 is the correct NCS rotational component, answer 3 is a crystallographic symmetry rotational component, and answer 4 is an incorrect NCS component. Thus, the correct answer was adequately separated from the nearest incorrect answer.

---

‡‡‡ The root mean square of the map (3.012) resulting from the Patterson search was taken as σ

The resolution cutoff employed in the transfer of the density 'island' to the triclinic cell had a dramatic effect upon finding the correct answer. If a resolution higher than 5 Å was used, the correct answer was less distinct. The correct answer disappeared completely if a 4 Å cutoff was employed. There are two reasons for this behaviour. First, the phasing statistics became progressively worse as resolution increased. The figures of merit were probably overestimated, so that spurious information (noise) was being transferred along with the structure factors which accurately described the low resolution shape of the dimer. Second, the NCS diad was ultimately determined to be a pseudo-diad only; *i.e.* it was not a strict NCS operator. The inclusion of higher resolution data likely broadened the occurrence of the true answer to the point where it disappeared into the background noise of the search.

3.2.1.2 Translational Component

A real space correlation search, as implemented in the CCP4 program GETAX (15), was used to establish the translational component necessary to complete the NCS operator. The search was performed on a 2 Å grid of a map calculated at 6 Å resolution from the 3/5 MIRAS phasing model. The search was intended to look for a dimer with a radius on the order of 28 Å and protomer units related by the rotational component described above. The results of the search are listed in Table 3.6.

Table 3.6 – Translation Solutions from the Real Space Correlation Search					
Answer	**X (Å)**	**Y (Å)**	**Z (Å)**	**Correlation coefficient**	**σ[§§§]**
1	13.5	78.1	1.5	0.58	11.6
2	4.1	51.9	28.1	0.45	8.7

The first answer in Table 3.6 appeared with slight variations in 17 of the top 20 solutions, whereas the second answer appeared in 3 of the top 20 solutions. The first answer was taken as correct. The complete NCS operator gave a correlation coefficient of ~0.4 and could be refined to better than 0.6. The NCS correlation coefficient helps to assess whether the operation is valid for the data set in question. A perfect agreement of 1.0 indicates perfect correlation, but values > 0.3 can be used to improve experimental phase information.

The presence of two strong answers suggested that multidomain averaging would improve the phase model more than single domain averaging. In order to carry out phase averaging, a mask must be generated which describes the overall shape of the protomer units; for multidomain averaging, a separate mask is required for each domain. The initial mask was generated by examining which parts of the 3.4 Å map from the 3/5 MIRAS phases correlated best after applying the NCS

[§§§] The root mean square of the map (0.052) resulting from the translation search was taken as σ

operator. This operation was implemented by DM. Visual inspection of the correlation masks revealed an obvious bilobal structure for the protein. The automatically generated mask was manually divided into two non-overlapping masks. The NCS operator was separated into two different operators by refining the single original operator with each of the two individual NCS domain masks.

Only manual revision of the NCS masks were done after this point. Manual revisions were necessary to encompass any new portions of the protein model which were not already in the mask.

3.2.2 Density Modification

All density modification was performed with the DM program. Multi-domain averaging, together with solvent flattening and histogram matching provided the best results. Phase extension of the 3.4 Å MIRAS phases to 2.6 Å was attempted; results were ambiguous at the time and in retrospect likely not very helpful. For the experimentally determined phases the density modification produced a remarkable improvement, as demonstrated in Figure 3.5.

3.2.3 Refinement of the Phasing Model

Despite the discovery of the NCS operator, only slightly more than half of the C_α positions could be determined from the resultant maps. Other density modifications, such as cross-crystal averaging with the $P2_1$

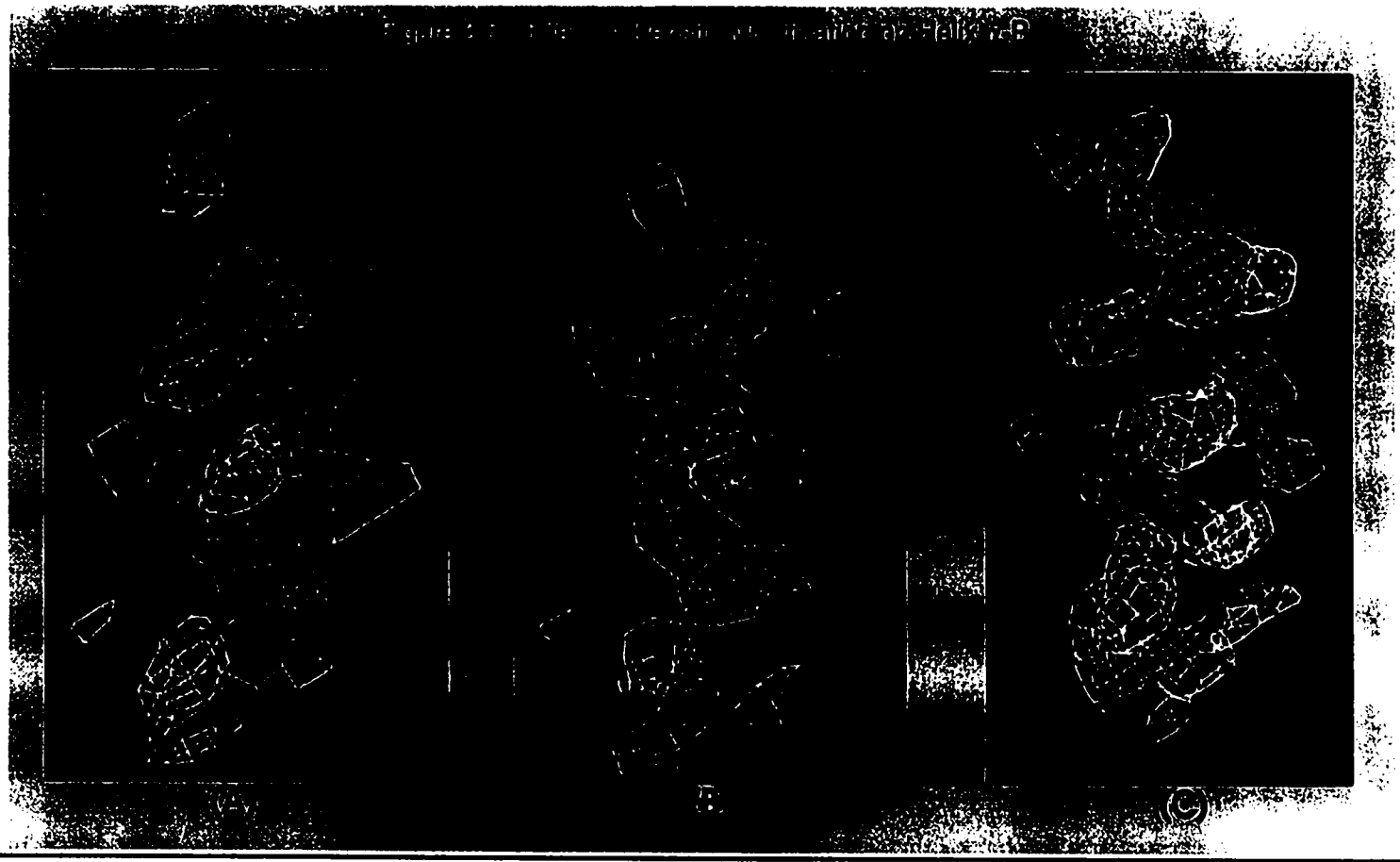
amplitudes, were tried but did not improve the maps. The greatest improvement came from two modifications to the original phasing model:

1. Removal of the two Pt derivatives from the MIRAS model to generate a 4 site Au SIR model.
2. Use of the low resolution data set to generate the initial phases.

The density modified map generated from the low resolution SIR phases enabled better than 90% of the protein backbone model to be constructed.

Model building was performed using the maps generated from the amplitudes measured from the low resolution form of the tetragonal crystal. Phasing statistics were superior with this data, and fewer chain breaks were observed in resultant electron density maps.

Side chain assignment was difficult to accomplish using the electron density maps initially available. The nucleotide binding region was identified by its α/β structure; this enabled approximately 40 residues to be built in from the C-terminus for both the A and B protomers. The overall R factor statistics at this point (R_{factor} and R_{free} were 0.505 and 0.556, respectively) were poor, but better than the values attributed to a random distribution of atoms. The remaining residues were modeled as serine residues, and the initial model was used to generate phases so that a difference map could be calculated in hopes of finding further heavy

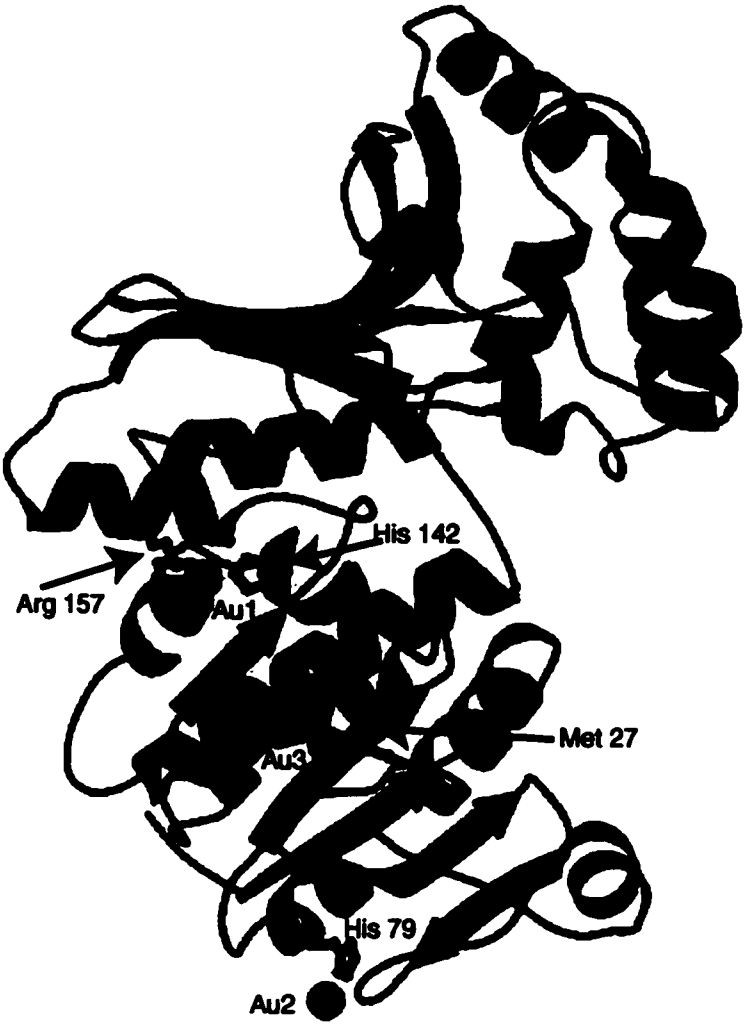


All maps are contoured at 1.2σ around helix α-B. **(A)** Electron density map derived from the unmodified 3/5 MIRAS phases. **(B)** Electron density map derived from the solvent flattened/histogram matched 3/5 MIRAS phases. **(C)** As (B), but with the two fold NCS used to average the phases. Note the improved connectivity and sidechain definition in the map as successive density modifications are applied. Density modification was done with DM. Maps are as displayed in O.

atom positions. A slight improvement in the phasing statistics was obtained by including two additional heavy atoms associated with the Met 27 residue of both the A and B protomer. These sites were located in the difference mapsgenerated with the combined experimental and incomplete model phases. The R_{Cullis} for the acentric reflections decreased from 0.59 to 0.58 and the phasing power for the acentric reflections increased from 2.66 to 2.70 (determined in SHARP). Statistics for the centric reflections improved as well.

The switch from a 4 heavy atom model to a 6 heavy atom model had benefits beyond improving the phasing statistics. It provided confirmation of the NCS operator. The heavy atom positions aligned in an unambiguous manner with each other, suggesting that there were three binding sites on each protomer, as shown in Figure 3.6. This information enabled the fitting of side chains in the polypeptide backbone by limiting the possible side chains in the vicinity of the heavy atom. This work is described in the following chapter.

Figure 3.6 - Gold Binding Sites in HSD



The three gold binding sites in the HSD protomer. The ligating residues are also shown. *NB* - the binding sites are identical in both protomers. Figure prepared with *Molscript/Raster3D*

3.3 Summary of Chapter 3

An excellent derivative for the tetragonal crystal form was obtained by soaking the tetragonal crystals in a solution of $AuCl_3$. The data collected from this derivative crystal produced an interpretable difference Patterson map. The subsequent location of three gold atoms in the derivative structure provided experimental phase information of sufficient quality to locate a fourth gold atom.

Density modification of the phased structure factors deduced from this 4 site model were of sufficient quality to generate a map in which the NCS relation of the two protomers could be determined. Additional phase information from two platinum deriviatives was initially incorporated to generate a MIRAS phasing model, but eventually dropped from the phasing model when it was discovered that maps were much cleaner when the gold derivative data alone was used.

The rotational component of the NCS was determined by a novel method, using the density modified MIRAS map as a starting point. The elucidation of the NCS enabled 2 fold averaging of the electron density maps. The averaged electron density maps could be interpreted well enough for the construction of a nearly complete poly-serine/alanine model.

The calculated phases generated by the poly-serine/alanine model enabled the location of two further gold atoms in the derivative structure.

Chapter 3 References

1. **Gewirth, D.**, The DENZO Manual, Yale, New Haven, CT (1993).
2. **Terwilliger, T. C., and Berendzen, J.** (1999). Automated MAD and MIR structure solution. Acta Crystallogr D Biol Crystallogr. **55**: 849-861.
3. **Otwinowski, Z.**, Maximum Likelihood Refinement of Heavy Atom Parameters, in *CCP4 Study Weekend*, Wolf, W., Evans, P. R., and Leslie, A. G. W., Eds., SERC, Daresbury, U.K., pp. 80-86 (1991).
4. **Ten Eyck, L. F.** (1973). The Fast Fourier Transform Acta Cryst. **A29**: 486.
5. **CCP4.** (1994.). The CCP4 Suite: Programs for Protein Crystallography. Acta Cryst. **D50**. 760-763.
6. **Cowtan, K.**, Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography, pp. 34-38 (1994).
7. **Howell, P. L., and Smith, G. D.** (1992). Identification of heavy-atom derivatives by normal probability methods. J. Appl. Cryst. 81-86.
8. **DeLaFortelle, E., and Bricogne, G.**, Maximum-Likelihood Heavy-Atom Parameter Refinement in the MIR and MAD Methods, in *Methods in Enzymology, Macromolecular Crystallography*, volume 276, R.M. Sweet and C. W. Carter, J., Ed., Academic Press., New York, pp. 472-494 (1997).
9. **Yumoto, N., Kawata, Y., Noda, S., and Tokushige, M.** (1991). Rapid Purification and Characterization of Homoserine Dehydrogenase. Arch. Biochem. Biophys. **285**: 270-275.
10. **Stein, P. E., Boodhoo, A., Armstrong, G. D., Cockle, S. A., Klein, M. H., and Read, R. J.** (1994). The crystal structure of pertussis toxin. Structure. **2**: 45-57.
11. **Main, P., and Rossmann, M. G.** (1967). Relationships among Structure Factors due to Identical Molecules in Different Crystallographic Environments. Acta Cryst. **21**: 67-72.
12. **Wang, B. C.**, Solvent Flattening in *Methods in Enzymology*, Vol. 115, pp. 90-112 (1985).
13. **Kleywegt, G. J., and Jones, T. A.**, Halloween ... Masks and Bones, in *From First Map to Final Model*, Bailey, S., Hubbard, R., and Waller, D., Eds., SERC Daresbury Laboratory, pp. 59-66 (1994).
14. **Tong, L., and Rossmann, M. G.** (1997). Rotation function calculations with GLRF program. *Methods in Enzymology*. 276: 594-611.

15. **Vonrhein, C., and Schulz, G. E.** (1999). Locating proper non-crystallographic symmetry in low-resolution electron-density maps with the program GETAX. Acta Crystallogr D Biol Crystallogr. **55**: 225-229.

Chapter Four: Structure of The Apo/Holo Form from the Tetragonal Crystal Data

4.1 Methodology

4.1.1 Model Building

A benefit of the 6 site Au SIR phasing model was a reduction in the number of possibilities for assigning side chain residues in the unestablished regions of the polypeptide backbone trace. The Au derivative was obtained by soaking gold ions into the crystals at pH 6.0. Previous attempts to generate the derivative at pH 4.4 had been unsuccessful; this suggested that the deprotonation of histidine residues was necessary to bind the gold complexes. Gold is a soft Lewis acid (1) that will bind to the soft Lewis basic nitrogen atoms in the side chains of the histidine residues. The NCS model dictated that there were 3 Au atoms per protomer, a value which matched the number of histidine residues in the primary structure.

Unfortunately, the sidechain assignment was not entirely straightforward. The 5th and 6th Au atoms were bound to a residue near the N terminus already believed to be a methionine, Met 27. The identify

of the residue had been assigned based on its sequential location within the Rossmann fold. The Rossmann fold was identified from the backbone atom positions by its overall α/β fold, analogous to the moderately well conserved Rossmann in other dinucleotide binding proteins. The location of the Rossmann fold had previously been localized to the N-terminal region of the primary structure because of the **GXGXXG** motif starting at residue 12. Because of poor electron density, it was difficult to make definite side chain assignments at this point; however, the chemistry was sensible: the sulpher atom in the methionine residue is also considered a soft Lewis base and will bind gold atoms. Furthermore, the 5^{th} and 6^{th} heavy atom positions refined to a lower occupancy than the other heavy atom binding sites. This was in accord with the weaker binding to be expected of the methionine sulpher atoms relative to the histidine nitrogen atoms.

This meant that a maximum of 2 of the 3 histidine residues could be binding the gold atoms. Fortunately, the remaining gold atoms were ligated to histidine residues in the derivative. The other four heavy atom positions enabled the identification of four of the six histidine residues in the dimer. This information greatly reduced the number of possibilities for assigning side chains and enabled a rapid completion of the model building.

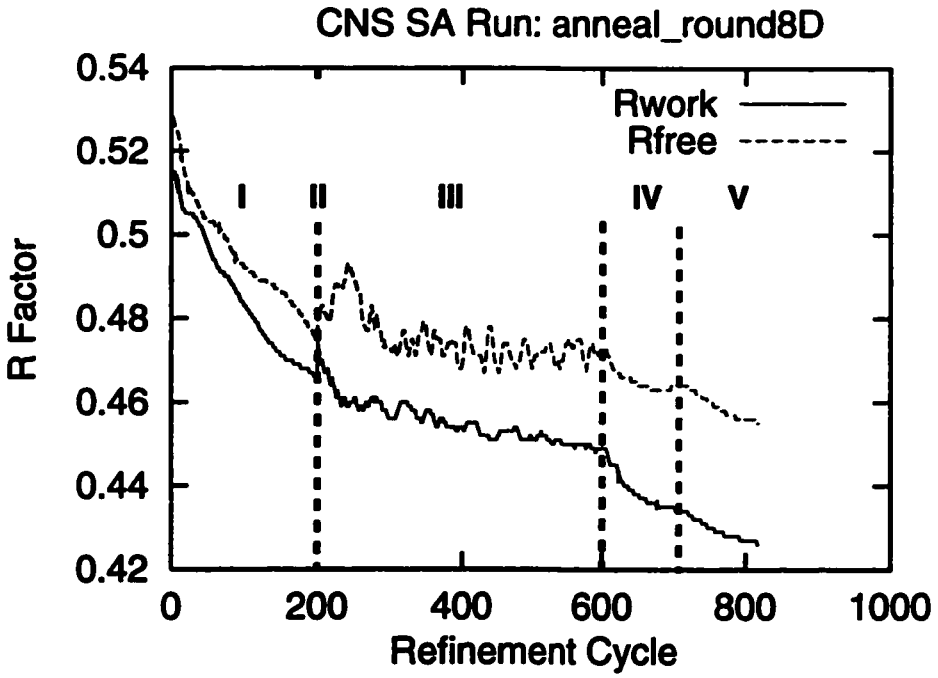
4.1.2 Model Refinement

Model refinement was accomplished by an iterative combination of refinement and manual intervention. Refinement is comprised of simulated annealing in both torsion and Cartesian space; Powell minimization; both grouped and individual B factor refinement; and real space refinement. A further description and results from a typical refinement are shown in Figure 4.1. Manual intervention is effected by model rebuilding at a computer graphics terminal guided by both experimental and difference maps. The most informative difference maps were σ_A weighted (2) 2Fo-Fc and Fo-Fc maps.

The program CNS, versions 0.3 through 0.9 (3) was used for the reciprocal space based refinements. The program O (4) was used for both model building and real space refinement steps. Two programs, OOPS (5) and PROCHECK (6) were used to locate regions of the protein which did not exhibit characteristics typical of a protein structure; this helped in the identification of regions which were not being properly refined by the reciprocal space based methods.

The R_{free} value based on a subset of approximately 830 (5%) of the measured reflections was used to determine which refinement protocol was most effective at each juncture. When no further progress could be made with the amplitudes measured for the low resolution crystal form, amplitudes from the high resolution crystal form were employed for

Figure 4.1 - A Simulated Annealing / Powell Minimization Refinement



I - Powell minimization
II - Powell minimization with no X-ray term
III - Slow cool simulated annealing with torsional dynamics (3000 to 0 K)
IV - Molecular torsion dynamics at 300 K
V - Powell minimization

Refinement is performed such that E_{tot} is minimized

$$E_{tot} = E_{model} + wE_{xray}$$

Where:
E_{model} is derived from the bond lengths and angles of the protein
E_{xray} is a term describing the agreement between observed and calculated structure factors
w is a weighting term to balance the contribution of the two energy terms
NB - if NCS was used in the refinement, it would be included as an additional energy term, i.e. E_{NCS}

refinement. The R factor statistics were 0.390 and 0.449 for R_{factor} and R_{free} respectively, and did not change significantly when the high resolution crystal set of amplitudes were introduced.

After switching to the high resolution set of amplitudes, the map generated by the modified experimental phases from the low resolution amplitudes was no longer useful as guide to rebuilding. Model rebuilding was guided to completion solely by difference maps generated from the model and the high resolution amplitudes. The final refinement statistics are given in Table 4.1, and Ramachandran plots for each protomer are shown in Figure 4.2.

When the data set switch took place, the old R_{free} set was discarded and a new set of approximately 2000 R_{free} (6%) reflections were selected. This is not an entirely correct procedure, since the R_{free} set should always be kept separate from the R_{factor} set to get a truly independent cross validation of the refinement process. However, it was more convenient to select a new R_{free} subset. Subsequent rounds of simulated annealing reestablished the independence of the new R_{free} set, enabling the continued use of this statistic as a measure of refinement progress (7).

Table 4.1 – Refinement Statistics for the Tetragonal Structure*	
Resolution (Å)	40 - 2.3
R_{sym} {last shell} *(%)*	..077 {0.49}
Multiplicity	5.1
# Unique Reflections	32166 (35733)
R_{factor} *(%)*	0.218 (0.234)
R_{free} *(%)*	0.266 (0.281)
Total atoms	5719
HSD protomers	2
Substrate molecules	1 NAD^+
Ions	2 Na^2
Solvent atoms	253
Bulk Solvent Model	flat (k=0.336, B=29.92 $Å^2$)
R.m.s.d bonds (Å)	0.007
R.m.s.d angles (°)	1.4
R.m.s.d torsions (°)	23.1
Coordinate error†(Å)	0.3
Ramachandran plot: % residues in allowed region (core)	98 (86)

* The values are given from statistics generated from F with a 2σ cut off. Where appropriate, the values with no cutoff are given in round brackets.

† This is the cross validated σ_a coordinate error

Figure 4.2 (A and B) Ramachandran Plots of Tetragonal Models

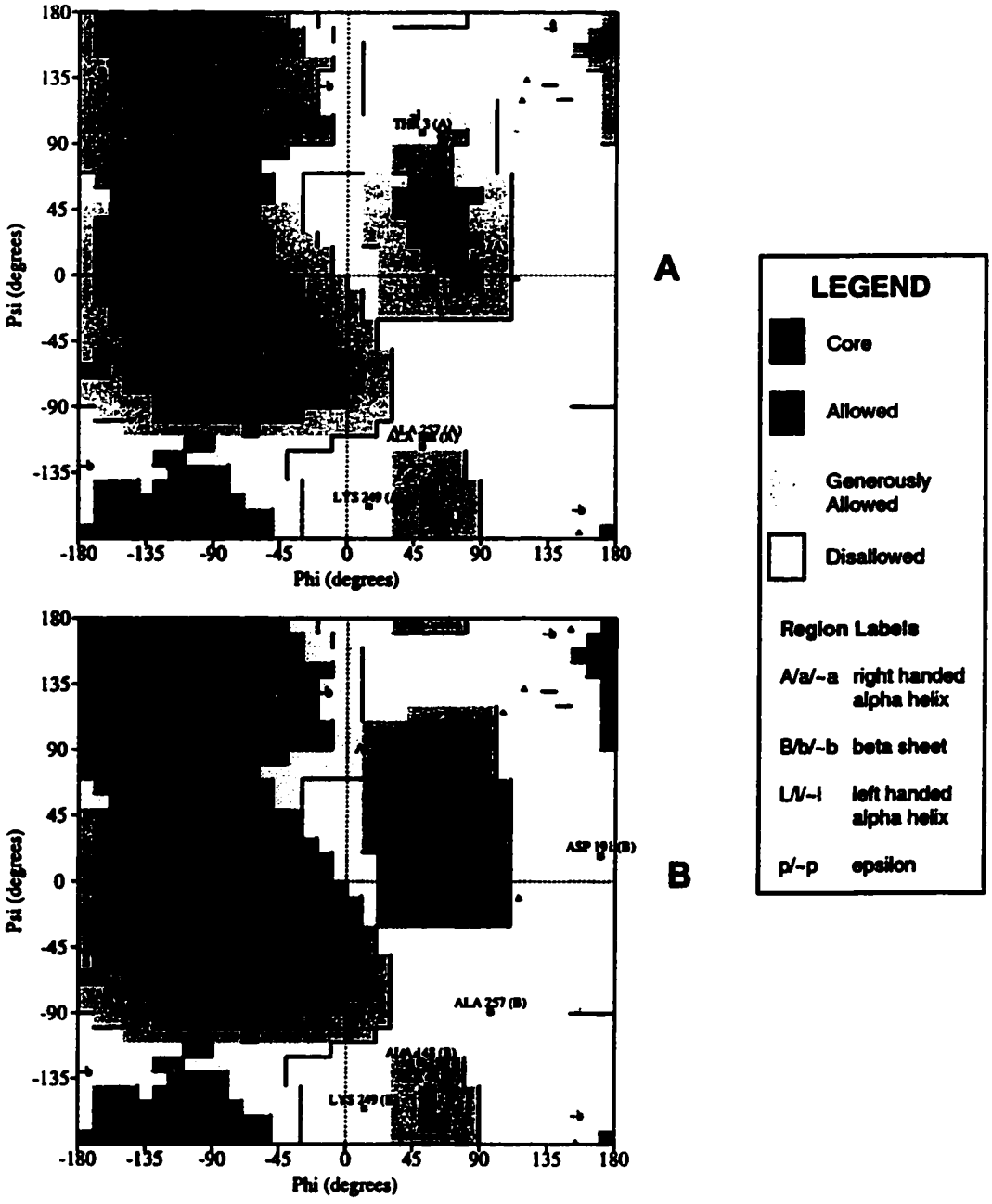


Figure (A) and (B) are the Ramachandran plots for protomers A and B (of the tetragonal crystal form), respectively. Residues in the disallowed regions are labelled with residue number and type. The core, allowed, and generously allowed regions are shaded according to the legend given above and also with appropriate letters in upper case, lower case, or preceeded by a tilda, respectively (*e.g.* A, a, ~a for the alpha helical region). Definitions for the region are as given in Morris, A. L., MacArthur, M. W., Hutchinson, E. G., and Thornton, J. M. (1992). Stereochemical quality of protein structure coordinates. Proteins. 12: 345-364.

4.1.2.1 The Role of the NCS Operator During Refinement

Although it was useful for improving experimental phase information, the NCS did not accurately describe the higher resolution details of the protein structure; this became clear as more stringent requirements for model accuracy were introduced to the model building / refinement process. Initially refinement was carried out with strict NCS constraints, *i.e.* the B molecule was generated by applying the NCS operator to the A molecule. The inaccuracy introduced by the NCS constraint was apparent from the behaviour of the R_{free} statistic. The NCS constraint was reduced to an NCS restraint and then eventually dropped altogether from refinements after the restraint failed to improve the agreement between model and data.

The distinction of constraints *versus* restraints is an important one. Crystallographic refinement is a minimization problem; as such it is affected by the overall data to parameter ratio. The use of either constraints or restraints during the minimization is designed to increase this ratio. Constraints fix certain parameters (*e.g.* atomic coordinates) to a specific value, effectively reducing the number of parameters being refined. Restraints allow certain parameters to only lie within a specific range; the introduction of these ranges effectively increases the amount of data being used for the minimization.

The improvement observed after omission of the NCS operator constraint was examined in detail by examining the r.m.s.d. of the Cα atom positions in structures submitted to simulated annealing refinement in torsion space without the NCS operator constraint. The r.m.s.d. values were determined from the results of 10 separate simulated annealing runs, each started from the same coordinate positions. The protomers exhibited an r.m.s.d. of 0.69 ± 0.03 Å among the C_α positions of either protomer A or protomer B, and an r.m.s.d. of 1.09 ± 0.03 Å between the C_α positions of protomers A and B. This implied that the asymmetry between protomers was real, and not simply an artifact of refinement. Visual examination of the overlap revealed that some regions of the protein obeyed NCS better than others; this variation was modeled so that the NCS operator could be reintroduced as a refinement restraint.

The reason for the NCS breakdown was correctly presumed to be due to preferential binding of NAD^+ in one of the protomers. Density for the NAD^+ molecule in the A protomer and the corresponding absence of density in the B protomer was only definite during the latter stages of refinement. There are several possible reasons for why the density for the cofactor molecule was vague in the initial stages, but the most probable is that density corresponding to the cofactor molecule was degraded during the averaging procedure used to obtain interpretable starting map.

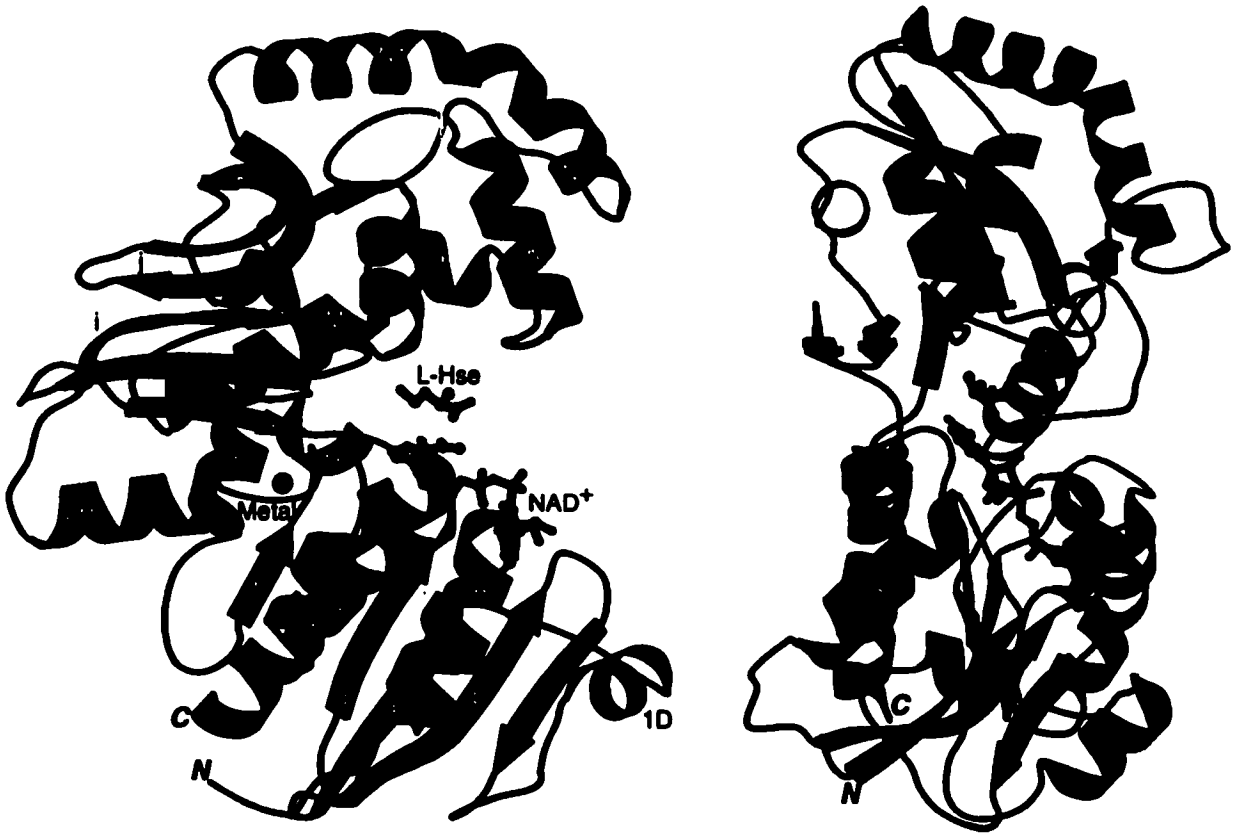
Although it was known that the protomers were not strictly identical, the NCS operator was still used to guide model building. If a region of the electron density maps for one of the protomers were clearer than that of its partner, the model built into the NCS related region was used as a starting point for rebuilding and subsequent refinement. This approach was dropped only during the last few rounds of refinement/model building, when the individual protomers were built and refined completely independent of one another.

4.2 Results and Discussion

4.2.1 Overall Fold

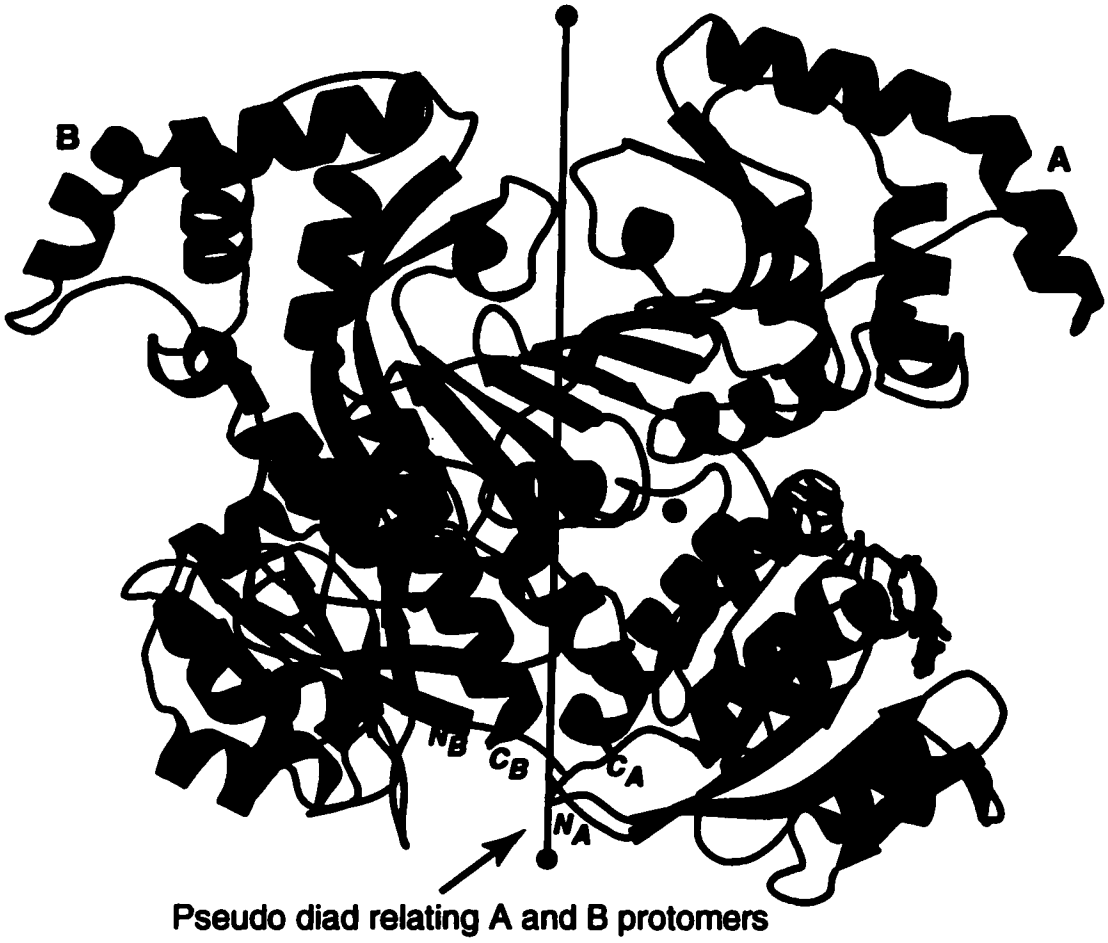
The overall shape of the HSD protomer is trilobal (see Figure 4.3). The three lobes are: the dinucleotide binding region, the dimerization region, and the catalytic domain. A distinction of region *versus* domain is employed here to describe parts of the protein which are in close proximity in the tertiary structure but are either discontinuous (regions) or continuous (domains) in the primary structure. The overall fold is analogous to a bent hairpin. The polypeptide chain begins with the N-terminus located in the dinucleotide binding region and finishes with the C-terminus folded as a helix against this region such that both the N and C termini are located close in space to one another. The location of each region with respect to the primary sequence is shown in Figure 8.1.

Figure 4.3a - The Homoserine Dehydrogenase Protomer in Two Views



Two orthogonal views of the HSD protomer. The leftmost figure is the standard view used in most other figures. The rightmost figure is oriented with the active site facing the viewer. The amino and carboxyl termini are labelled ***N*** and ***C***, respectively. The structure is coloured such that the dinucleotide binding region is red, the dimerization region is green, and the catalytic domain is red. The NAD^+ molecule is shown in yellow. The location of the L-Hse molecule was taken from the ternary complex discussed in chapter 5. The metal atom (Na^+) is shown as a green sphere.

Figure 4.3b - The Homoserine Dehydrogenase Dimer



The HSD dimer. Colouring scheme is the same as shown in Figure 4.3a. The pseudo-diad relating the two protomers is shown as a black baton. The amino and carboxyl termini are labelled *N* and *C*, respectively. The termini are further labelled with subscripts denoting whether they are from the A or B protomers. The NAD^+ molecule bound to the A protomer is shown in yellow. The metal binding site is shown with a green sphere in the A protomer only.

4.2.1.1 Dinucleotide Binding Region

The dinucleotide binding region in HSD (Figure 4.4) is comprised of residues 2 through 147 and residues 340 through 359. It has an α/β fold typical for its function (8) (9). A VAST search of the PDB revealed it to be similar to several NAD dependent enzymes. The HSD dinucleotide binding region is closest to UDP galactose 4-epimerase (UGE) (10) with which it shares 112 similar C_α positions. The two folds exhibit an an r.m.s.d. of 3.1 Å between their aligned C_α positions. The dinucleotide binding region of HSD differs slightly from that of UGE in that HSD has 6 as opposed to 7 β-strands, with the direction of β-c[‡] on the far end of the sheet reversed for HSD.

A unique structural feature of the HSD dinucleotide binding region is a short 3_{10} helix (α-1F) which joins β-e to α-2F. The α-1F helix is positioned at the boundary between the dinucleotide binding region and catalytic domains

---

[‡] The secondary structures are named with lower case letters for β strands and upper case letters for α helices. Loops will be denoted by their surrounding secondary structural elements.

Figure 4.4, 4.5, 4.6 - Schematic Diagrams of the HSD Domain Folding

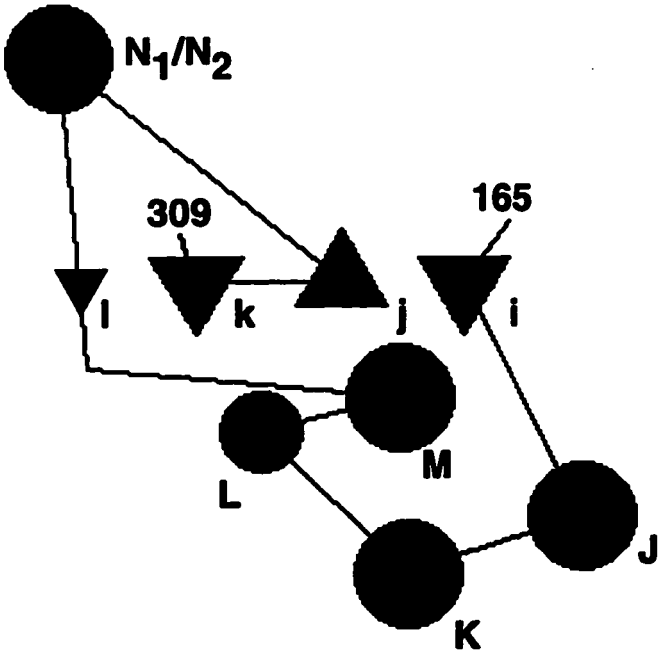


Figure 4.6 - Catalytic Region

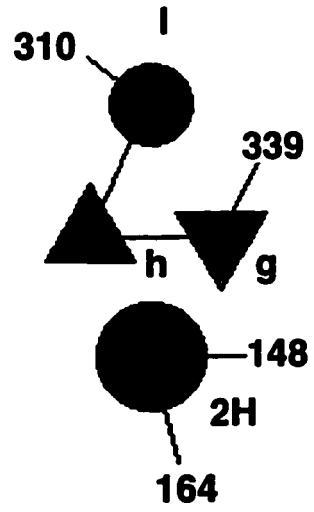


Figure 4.5 - The Dimerization Region

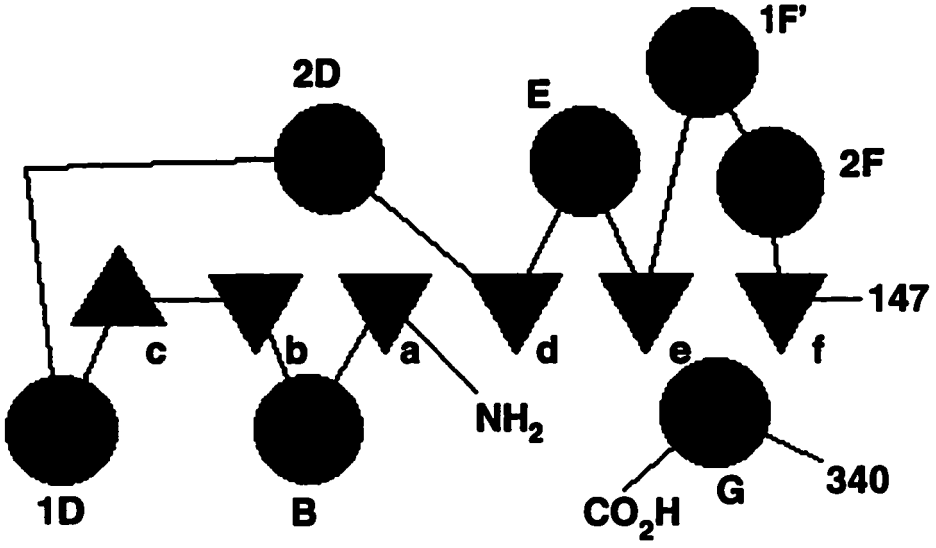


Figure 4.4 - Dinucleotide Binding Domain

and contains a putative catalytic lysine residue (Lys 117) which is strictly conserved among HSD sequences.

The HSD dinucleotide binding region is completed by the C-terminal helix, α-G. Composed of residues 340-359, helix α-G protrudes from the dimerization domain to pack against the dinucleotide binding region β-sheet between strands β-e and β-f. This has been observed previously among the Rossmann folds: the C-terminal helix of D-glyceraldehyde-(3)-phosphate dehydrogenase (GAPDH) folds in a similar manner (11). The comparison is drawn here because it has some bearing upon the stereospecificity of hydride transfer from the nicotinamide cofactor. This point will be discussed in further detail section 4.2.4.

4.2.1.2 Dimerization Region

The dimerization region is the smallest region of the protein. It is comprised of residues 148-164 and 310-339. The domain consists of four elements, as shown in Figure 4.5: α-2H, which extends out from the dinucleotide binding region; α-I, which is a short 3_{10} helix; and two anti-parallel strands, β-g and β-h. The β strand pair interacts in an intramolecular fashion on the β-h side with the β-i strand of the catalytic domain and in an intermolecular fashion on the β-g side with the dimerization region of the second protomer. Strand β-g hydrogen bonds with its opposing partner from the B subunit in an anti-parallel fashion.

Helix α-G from the dinucleotide binding region also plays a role in stabilizing the subunit interaction.

4.2.1.3 The Catalytic Domain

The catalytic region (Figure 4.6) is comprised of residues 165-309. The fold consists of β-strands i, j and k which extend the anti-parallel sheet of the dimerization region to create a 12 stranded mixed β-sheet in the dimer. A short β strand at the edge of the sheet, β-l, runs parallel to its nearest neighbour β-k. The remainder of the of the strands which comprise the inner 10 strands of the dimeric β sheet associate in an anti-parallel fashion.

The fold is also comprised of helices J through N. Along with the short β-l, this helical region forms a connection from β-i to β-j. Along with its connecting loops, the stretch constitutes 110 residues and thus represents the bulk of the 140 residue catalytic domain.

Two catalytic residues are found in α-M: Asp 219 and Lys 223. Residues responsible for substrate binding are found in α-J (Thr 176), the K-L loop (Glu 208) and α-L (Asp 214). The roles of these residues are known from the ternary structure and site directed mutagenesis experiments discussed in chapters 5 and 6 respectively.

Helix α-1F provides a boundary between the dinucleotide binding region and the catalytic domain, as well as a fourth putative catalytic

residue, Lys 117. The implications of this arrangement are discussed in more detail in later chapters of this thesis.

Helix N is a long (20 residue) broken helix that connects β-I to β-j, traversing the catalytic domain β-sheet in the process. There is a short region of 3_{10} helix in the middle of α-N, dividing the helix into α-N1 and α-N2. The entire α-N helix exhibited higher B factors in both subunits, but was slightly more ordered in the B protomer than in the A protomer. The variability in this region of the protein structure is due to the extensive crystallographic contacts made by the α-N helices (*vide infra*).

4.2.2 The Singular Nature of the HSD Catalytic & Dimerization Regions

Neither the catalytic nor the dimerization region have matches in the current PDB. This was determined by conducting a VAST (12) search of the entire database. The search was undertaken in three different ways: either the catalytic domain, the dimerization region, or the combined catalytic/dimerization domain were used as a search model. The absence of any determined protein structures resembling these regions of HSD establishes that HSD has a novel dehydrogenase fold. The name 'alpha-beta spiral' is proposed for this new motif because when the overall folding pattern for this mixed α/β domain is displayed in a topology diagram (Figure 4.6) it takes on the general shape of a spiral.

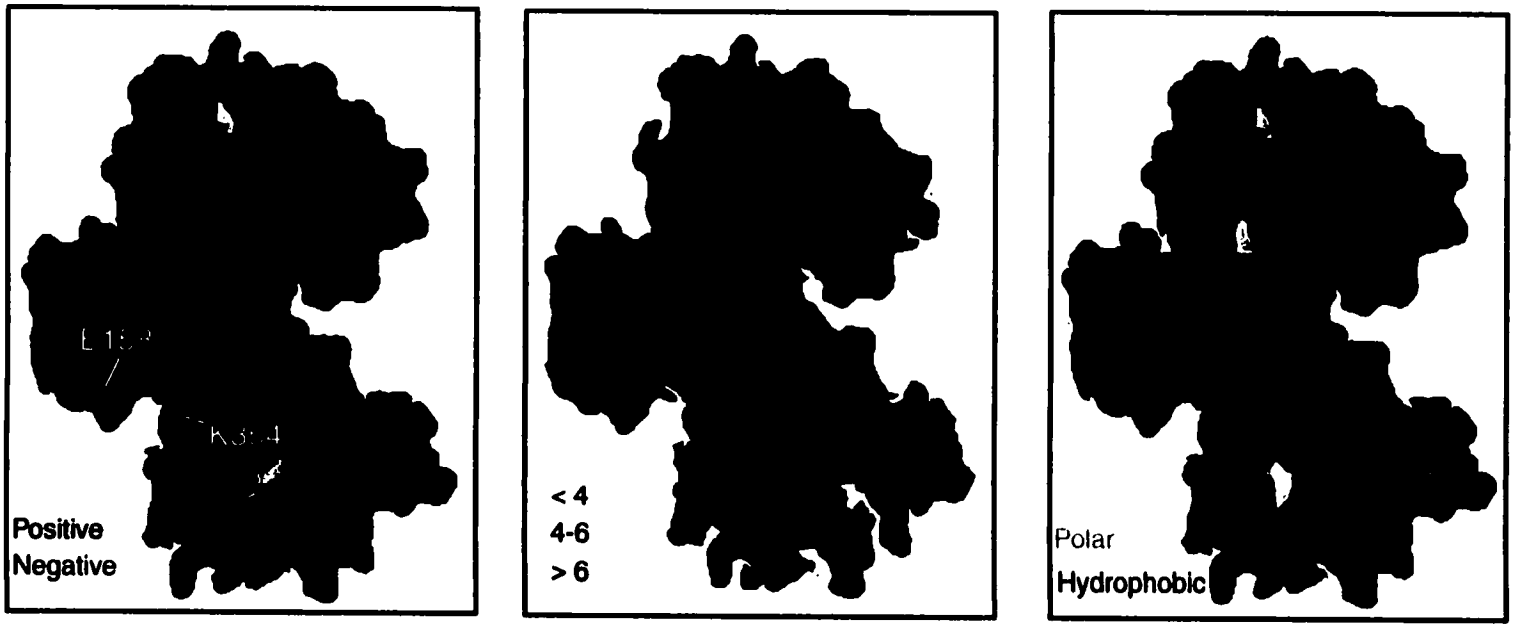
4.2.3 Dimer Interface

HSD forms a homo-dimer with an interface surface area of 1600 $Å^2$; the nature of the interface is shown in Figure 4.7. The main interaction between protomer A and B is appears to be an antiparallel β-strand hydrogen bonding pattern between β-strand g and its diad related partner in the opposing protomer. With one exception, the remainder of the interactions are of a hydrophobic nature, including the residues which constitute the β strand at the interface.

The only ionic interaction is a 2.7 Å salt bridge between Lys 354 of protomer A and Glu 158 of protomer B and *vice versa*. Lys 354 interacts with the strictly conserved Asp 351 (3.2 Å O_γ-N_ε distance) on its on subunit, apparently to orient the lysine residue for an effective interaction with Glu 158 from the other protomer subunit.

Both residues are moderately conserved among the group 1 and 3 HSD enzymes. There is a notable exception within the group 1 enzymes. When the sequence of *Archeoglobus fulgidus* HSD is aligned to the *S. cerevisiae* sequence a reversal of charge polarity can be observed: Glu 158 is replaced with an arginine residue and Lys 354 is replaced with a glutamate residue.

Figure 4.7 The Dimer Interface



Three different properties of the dimer interface are shown, with appropriate values mapped to the van der Waal surface of the protomer generated in GRASP. The standard view of the protomer is shown. The first panel shows the ionic surface potential, and the location of the residues participating in a salt bridge between the two dimers. The second panel is a map of distances between all surface atomic positions of the two protomers, with the colouring indicated for distances of <4, 4-6, and >6 Å, respectively. The third panel shows the hydrophobic nature of the sidechains involved in the interface.

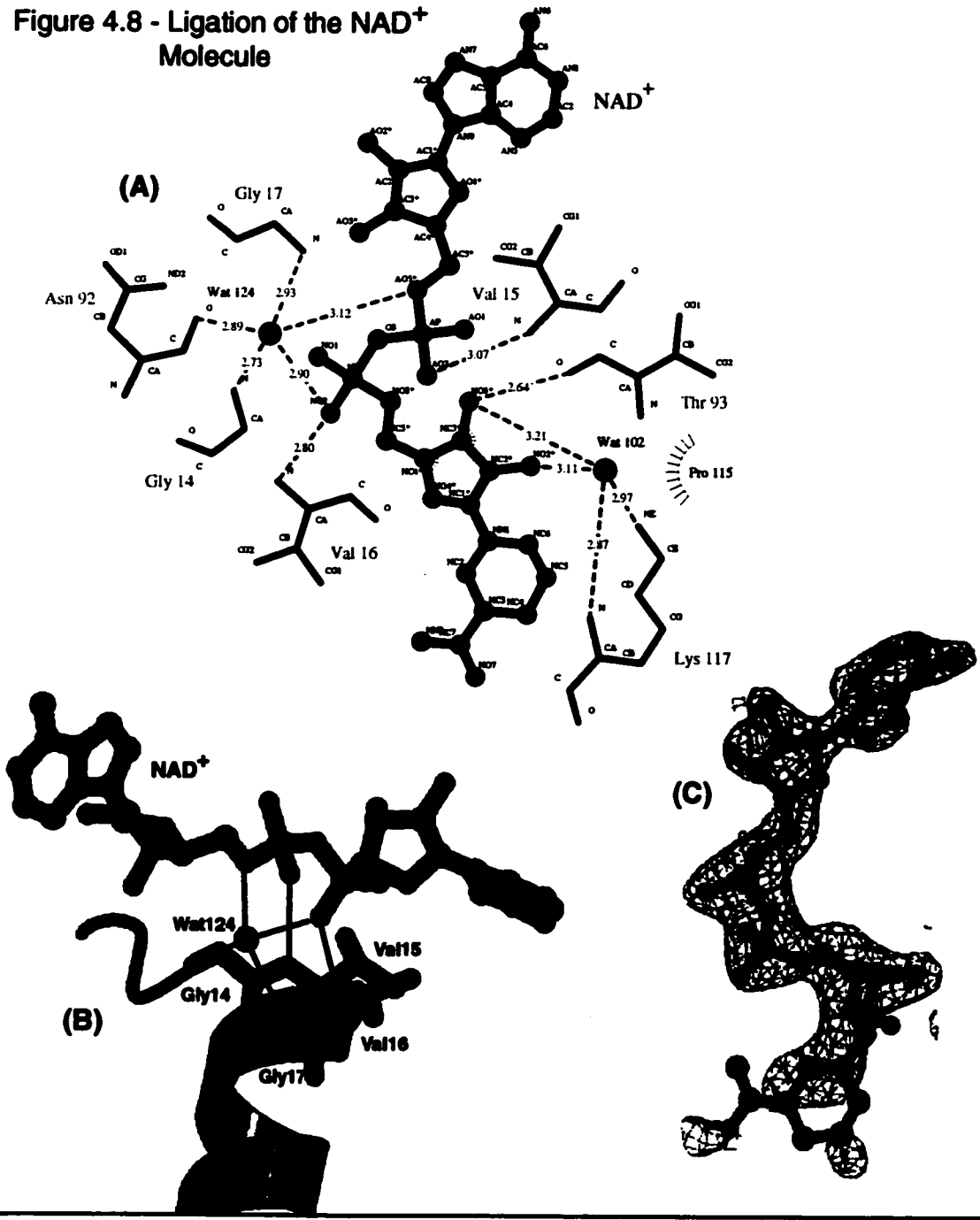
4.2.4 Dinucleotide Binding

Only a single NAD^+ molecule was bound to the tetragonal form crystal structure of HSD. There is no strong evidence for cooperativity in the HSD dimer, so the asymmetric binding is presumably a result of the crystal packing (*vide infra*).

As shown in Figure 4.8, the NAD^+ molecule interacts primarily with HSD at the N-terminus of helix B. The positive dipole of helix B is directed at the negatively charged diphosphate region of the NAD^+ molecule. Such an arrangement is observed for many other dinucleotide binding proteins (13). There are specific interactions between the phosphate oxygen atoms and the residues which comprise the first turn of helix B, namely Gly 14 through Gly 17. These residues represent the latter part of the motif (GxGxxG) which can be used to identify dinucleotide binding proteins (8). Specific interactions occur directly between phosphate oxygen atoms and the amide nitrogen atoms of both Val 15 and Val 16. Water mediated interactions by Wat 124 occur between phosphate oxygen atoms and the amide nitrogen atoms of both Gly 14 and Gly 17. The peptide carbonyl oxygen of Asn 92 also interacts *via* the water molecule Wat 124 with the diphosphate oxygen atoms.

The only other hydrogen bonding between NAD^+ and HSD occurs between the hydroxyl oxygen groups of the pyridyl end ribose ring and residues Thr 93 and Lys 117. There is a direct interaction between the

Figure 4.8 - Ligation of the NAD^+ Molecule



(A) Ligation of the NAD^+ molecule by the A protomer of the tetragonal structure. Hydrogen bonds are shown with dashed lines and Van der Waal's interactions are indicated with dashed semi-circles. Figure generated with *LigPlot*. *(B)* Interaction of the NAD^+ molecule with the B helix. Note that the diphosphate moiety sits directly above the N-terminal dipole. The 4^{th}-6^{th} residues of the GXGXXG motif are shown, as is Wat 124. Figure generated with *Molscript/Raster3D*. *(C)* A σ_a weighted 2Fo-Fc SA-omit map displayed at 0.6σ around the NAD+ molecule - *N.B.* the poor electron density at the pyridyl ring. Figure was generated with *Bobscript/Raster3D*.

carbonyl oxygen atom of Thr 93 and the $C_{3'}$ hydroxyl oxygen of the ribose ring. There are several interactions between Lys 117 and the bound NAD^+ molecule. Lys 117 interacts *via* both its sidechain amino group and its backbone amide group to water molecule Wat 102 which in turn forms hydrogen bonds to either of the 2' or 3' ribose hydroxyl groups of the cofactor molecule. Furthermore, the side-chain nitrogen is close enough (3.8 Å) to interact directly with the $C_{2'}$ hydroxyl oxygen. Lys 117 is adjacent to both Asp 223 and Lys 219 , both of which are key catalytic residues (see Figure 5.5 in the following chapter). Furthermore, Lys 117 is strictly conserved among the HSD sequences and is the only residue which interacts with the cofactor through side chain atoms. Therefore, Lys 117 is likely a key residue not only for binding NAD^+, but also for bringing the cofactor into a favourable position for hydride transfer to occur. This point will be discussed further in chapter 5 when discussing the changes which occur upon binding the product, L-Hse.

There is little evidence of direct interactions between either the adenine ring or the pyridyl group of the NAD^+ molecule and the protein. The paucity of interactions for either end of the NAD^+ molecule explains why the electron density around the pyridyl ring was not very well defined (see Figure 4.8C). The B factors for the pyridyl ring were approximately 10 $Å^2$ higher than the rest of the cofactor molecule.

This situation is not, however, reflected in the adenine moiety of the NAD^+ molecule. The adenine ring has well defined electron density in simulated annealing omit maps and exhibits B factors similar to those of the more tightly bound regions. Despite the lack of any close interactions or evidence for motion, the adenine ring is ideally situated within the shallow pocket lined by the sidechains of residues Ala 41, Thr 70, Leu 73, Ile 98, and the peptide backbone atoms of Glu 40.

The positional semblance of the C-terminal helices of HSD and GAPDH was mentioned previously. The two enzymes share another feature: both catalyze transfer of the *pro-S* (B face) hydrogen atom from the C_4 position of the reduced nicotinamide ring. In GAPDH there are two conserved residues, Asn 334 and Glu 335, on the loop connecting the C-terminal helix to its preceding secondary structural element. The function of these residues in GAPDH is to orient the pyridyl ring such that *pro-S* hydride transfer is favoured. HSD also has this loop but the sequence of residues near to the NAD molecule are Gly 338, Ala 339, Gly 340, and Ala 341. These residues do not have side chains capable of forming hydrogen bonds with the NAD molecule and are too distant to form hydrogen bonds through backbone atoms or to provide Van der Waal's type interactions. However, after the first turn of the C terminal helix there is a strictly conserved threonine residue, Thr 344, which is fully capable of forming a hydrogen bond interaction with its sidechain hydroxyl group to

either the carbonyl oxygen or amino group on the pyridyl ring. However, the interatomic distances are all beyond 4 Å in the crystal stucture of the holo enzyme.

4.2.5 Crystal Packing Analysis of the Tetragonal Crystal Form

Since the principal experimental method employed in this thesis is X-ray crystallography, some discussion of crystal packing is warranted here. One of the criticisms leveled against X-ray crystallography as a means for gathering information on biological molecules is that when they are crystallized, the molecules are not necessarily in an aqueous phase conformation. While there is some truth to this statement, the similarity between structures determined in different space groups, as is the case for HSD, as well as the similarity between structures determined by both X-ray and NMR methods shows that the structures determined in crystalline form are indeed relevant to the conformation of the structure in solution phase. The space between protein molecules in the crystalline form is filled with solvent molecules, mainly consisting of disordered water molecules. Any regions of the protein not participating in crystal contacts can be assumed to be facing solvent, and thus in an environment similar to that found *in vivo*. However, in order to attach biological relevance with confidence to the structures presented in this work it is important to examine the crystal packing forces that affect the structure of HSD.

Figure 4.9 - Crystallographic Contacts in the Tetragonal Data

Segments of the A and B protomers involved in crystal contacts (closer than 3.6 Å interaction) resulting from crystal packing are indicated in red. Molecules are shown in the standard view. Figures generated with Molscript/Raster3D.

The location of the crystal contacts are shown graphically in Figure 4.9. Overall, 9.5 % of the residues are involved in crystallographic contacts, defined here as interactions that are within 3.6 Å of a symmetry related molecule. The most obvious interaction is between the dinucleotide binding region of molecule B and Helix 1D from a symmetry related copy of molecule A. The 4_3 screw operation (y+0.5, -x+0.5, z+0.25) generates a copy of the A molecule in which a helix, α-1D occupies the same area that the adenine ring of an NAD molecule would occupy in the B protomer. This explains why only one of the molecules in the crystallized dimer was bound with NAD^+. The preclusion of NAD^+ by the α-1D crystal packing interaction does not detract from the observation that the unbound protomer should be similar, if not identical, to the apo form of the enzyme. This is so because the NAD^+ adenine ring makes no discernable contacts with the protein. It is unlikely that the protein conformation could be affected in a global manner by the crystallographic contacts described above.

Another interesting feature of the crystal packing is the clear presence of a solvent channel located approximately 25° off of the z-axis (shown in Figure 4.10). The width of these channels is slightly larger than 19 Å, a distance which could easily accommodate the NAD molecule which measures slightly less than 9 Å at its widest point. The channel is notable because HSD crystals were observed to crack when quantities of

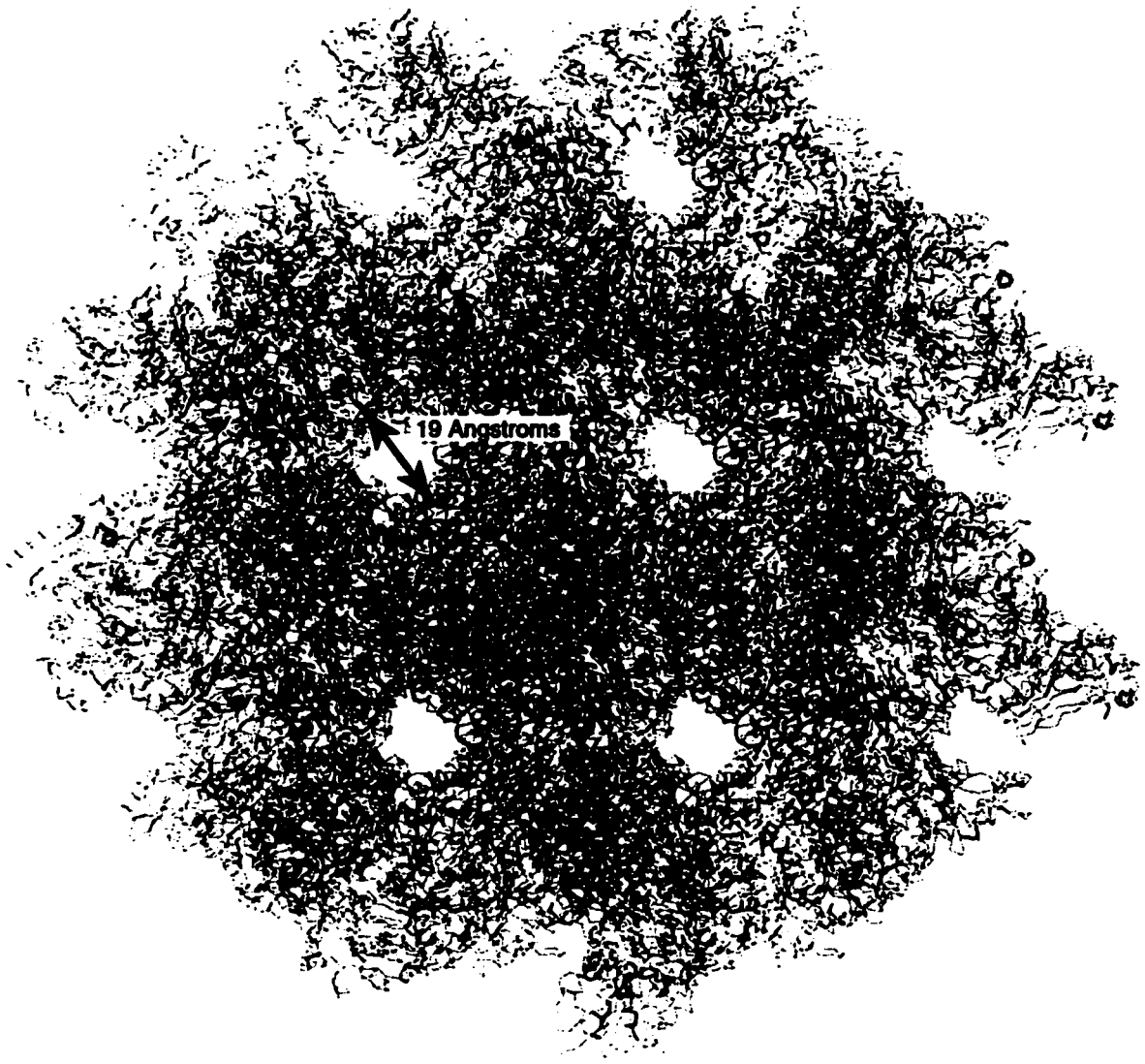
NAD^+ greater than 5:1 cofactor:protein molar ratio were included in the mother liquor. It is quite possible that the crystals were destroyed when NAD^+ molecules traveled down the solvent channels and occupied the region in the B protomer blocked by the symmetry generated helix 1D from the A protomer.

4.2.6 Differences Between Bound and Unbound Forms

A fortuitous result of the crystal packing is that the dimeric HSD bound with a single NAD^+ allows a view of the enzyme in both its NAD^+ bound (holo or A) form and unbound (apo or B) form. HSD undergoes an ordered reaction process as shown in Figure 6.2. Strictly speaking, the apo and holo form of the enzyme from the tetragonal structure represents the initial and ultimate catalytic forms of the enzyme, respectively. Because it is unlikely that significant differences would occur between the NADH bound structure and the NAD^+ structure, the holo form of HSD presented herein is a reasonable model for the second complex in the reduction process as well.

The apo and holo form of the enzyme were aligned by a brute force search for maximal overlap and then globally optimized for the best overall alignment, as implemented in the program lsqman (14). The r.m.s.d. of the C_α positions is given in Table 5.5 in the following chapter. The lack of gross domain motion between the different forms is striking. Using the structure of only a single protomer as a guide, one would suspect that a

Figure 4.10 - Solvent Channels in the Tetragonal Crystal Structure



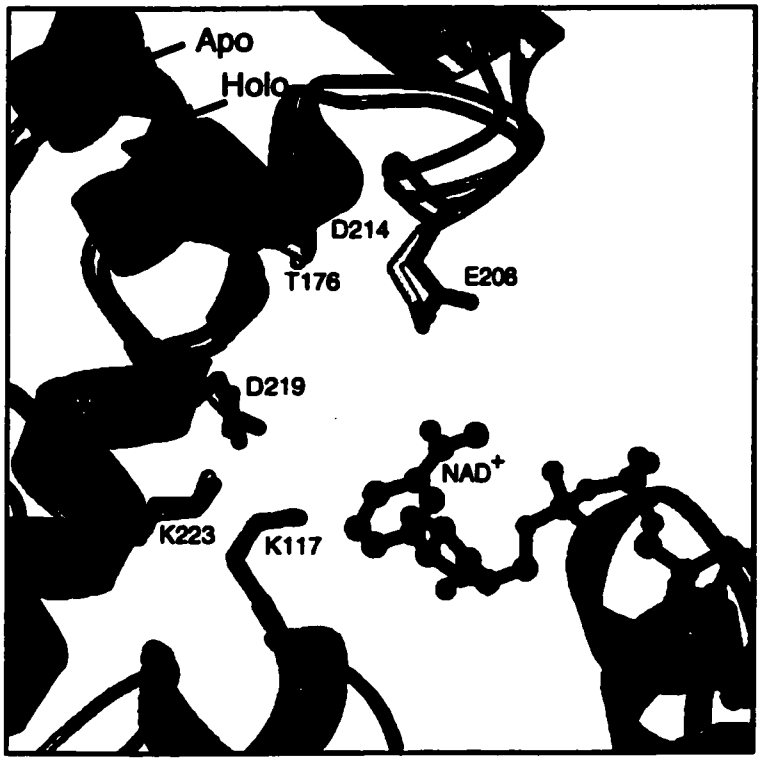
Crystal packing viewed 25^{o} off the z axis to reveal a series of solvent channels in the tetragonal crystal form. Figure generated with O.

clamping type motion would be feasible during the reaction mechanism, with the catalytic lobe moving down to facilitate the hydride transfer reaction. A more critical examination reveals that the dinucleotide binding region is connected to the catalytic region with a double hinge attachment. A clamping type motion of the catalytic region would require extensive rearrangement in several regions of the protein; this type of conformational change is not easily accomplished.

Despite the lack of gross motions in the protein, there are some smaller movements which are significant for the catalysis (shown in Figure 4.11). As expected for the binding of the nucleotide, the most significant differences between the A and B protomers occur in the dinucleotide binding region. Helix B apparently undergoes a slight tilting motion of 1.1° with an overall r.m.s.d. of 0.4 Å for residues 14-27. There is a crystallographic contact for Gly 14 in the B protomer, so the motion may be due partially to crystal packing forces.

The differences that occur for α-J, α-K, α-L and their connecting loops, all elements which contain key catalytic or substrate binding residues, are due mainly to crystallographic contacts. The crystal contacts occur in the B protomer and involve 14 out of 40 residues. There are some large differences: the C_α atoms of α-J is approximately 0.9 Å closer to the NAD binding domain; some of the residues in the J-K loop move as much as 3 Å as measured by the difference of their C_α positions; and

Figure 4.11 - Overlap of the Apo and Holo Protomers at the Active Site



A superposition of the two protomers discussed in this chapter. The apo form is semitransparent, the holo form is solid. Key residues are shown and labeled. The secondary structure elements are coloured as before, with red for the dinucleotide binding region and blue for the catalytic domain. The NAD^+ molecule is coloured by atom, with the C_4 carbon of the pyridyl ring coloured in black. The overlap is based on a global alignment of the two protomers. Positional differences for the secondary structural elements are small *e.g.* the α-J helix moves 0.9 Å from the holo to the apo form.

certain key residues such as Glu 208 change their Cα position by 1.6 Å. The extensive nature of the crystallographic contacts makes it difficult to disentangle which differences occur as a result of NAD^+ binding and which simply accommodate crystal packing. The differences in regions remote from the bound dinucleotide are more easily attributed to crystal packing forces. A good example are the α-N1 and α-N2 regions that show apparent motions in excess of 3 Å in some areas. Helices α-N1 and α-N2 have a high number of crystallographic contacts.

Helix α-2F on the other hand, is free from crystallographic influence. It exhibits a positional change which is likely a breathing mode for the protein structure. The overall r.m.s.d. for the helix is 1.1 Å between the two forms. Furthermore, the C_α atoms of α-2F helix in the B protomer have thermal (B) factors which are on average 25 $Å^2$ higher than those in the A protomer. Helix α-2F immediately follows the 3_{10} helix α-1F which bears Lys 117, a residue which has an important role in binding the NAD^+ molecule. Only slight differences were observed for α-1F, but the greater mobility of α-2F suggests significant differences in the total motions of the dinucleotide binding region between the A and B protomers. This can be construed as further evidence that the two molecules are indeed in different states of ligation. Furthermore, the motion of the α-2F region belies its role as a switch region controlling enzyme activity, an issue that will be discussed further in chapter 8.

4.3 Summary of Chapter 4

The six site gold derivative enabled the construction of a complete model for HSD. The model was refined by simulated annealing dynamics, Powell minimization, B factor refinement, and manual intervention to an R_{factor} and R_{free} of 0.218 and 0.266, respectively. Data extending to 2.3 Å was used for the refinement.

The structure of the enzyme is trilobal, consisting of a dinucleotide binding region, a dimerization region, and a catalytic domain. Both the dimerization region and catalytic domain represent novel protein folds.

The nicotinamide cofactor binds in a typical fashion. The main interactions occur between the diphosphate moiety of the molecule and the α-B helix. The enzyme binds the cofactor asymmetrically in the crystal structure, in that one protomer has the cofactor bound to it while the other protomer is in apo form. The lack of cofactor binding in one of the enzyme protomers is due to crystal packing, and cannot be taken as evidence of a cooperative enzymatic mechanism.

A metal ion binding site was located near the dimer interface. The metal ion was modeled as a sodium ion based upon arguments presented in Chapter 7.

Chapter 4 References

1. **Cowan, J. A.**, *Inorganic Biochemistry*, VCH Publisher, New York (1993).
2. **Read, R. J.** (1986). Improved Fourier Coefficients for Maps Using Phases from Partial Structures with Errors. Acta Cryst. **A42**: 140-149.
3. **Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L.** (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr. **54**: 905-921.
4. **Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M.** (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Cryst. **A47**: 110-119.
5. **Kleywegt, G. J., and Jones, T. A.**, Model-building and refinement practice, in *Methods in Enzymology, Volume 276: Macromolecular Crystallography, part A*, Vol. 277, C.W. Carter, J. R. M. S., Ed., Academic Press, pp. 208-230 (1997).
6. **Laskowski, R. A., MacArthur, M. W., Moss, J. S., and Thornton, J. M.** (1993). Procheck. J. Appl. Cryst. **26**: 283-291.
7. **Kleywegt, G. J., and Brunger, A. T.** (1996). Checking your imagination: applications of the free *R* value. Structure. 897-904.
8. **Carugo, O., and Argos, P.** (1997). NADP-dependent enzymes. II: Evolution of the mono- and dinucleotide binding domains. Proteins. **28**: 29-40.
9. **Rossmann, M. G., Moras, D., and Olsen, K. W.** (1974). Chemical and biological evolution of a nucleotide-binding protein. Nature. **250**: 194-199.
10. **Thoden, J. B., Frey, P. A., and Holden, H. M.** (1996). Crystal structures of the oxidized and reduced forms of UDP-galactose 4-epimerase isolated from *Escherichia coli*. Biochemistry. **35**: 2557-2566.
11. **Vellieux, F. M., Hajdu, J., Verlinde, C. L., Groendijk, H., Read, R. J., Greenhough, T. J., Campbell, J. W., Kalk, K. H., Littlechild, J. A., Watson, H. C.** *et al.* (1993). Structure of glycosomal glyceraldehyde-3-phosphate dehydrogenase from *Trypanosoma brucei* determined from Laue data. Proc Natl Acad Sci U S A. **90**: 2355-2359.

12. **Madej, T., Gibrat, J. F., and Bryant, S. H.** (1995). Threading a database of protein cores. Proteins. **23**: 356-369.
13. **Hol, W. G., van Duijnen, P. T., and Berendsen, H. J.** (1978). The alpha-helix dipole and the properties of proteins. Nature. **273**: 443-446.
14. **Kleywegt, G. J.** (1999). Experimental assessment of differences between related protein crystal structures. Acta Crystallogr D Biol Crystallogr. **55**: 1878-1884.

Chapter Five: Structure of Ternary Complex from the Monoclinic Crystal Data

5.1 Methodology

5.1.1 Phasing of Monoclinic Structural Data

The Matthews coefficient for the monoclinic crystal form (Table 2.1) was consistent with the presence of 4 protomers within the asymmetric unit. Under the assumption that these were present as dimers similar to those found in the tetragonal data, the tetragonal form dimer was used as a search model in a molecular replacement search.

The molecular replacement search is divided into two components: a rotational search and a translational search. The entire procedure was accomplished with algorithms implemented in CNS 0.5-0.9 (1).

5.1.1.1 Rotational Search

The 'fast direct' rotational algorithm as implemented in CNS 0.5 (2) was used to establish the rotational components of the two dimers. The 'fast direct' method uses a real space algorithm to find significant rotation solutions and then improves these solutions with a direct algorithm. The nomenclature of the two subtly different algorithms is misleading; both use

Patterson based methods to find the rotational operator which describes how the search model is oriented in the experimental data. The difference between the two methods is the order in which the necessary calculations are performed.

The real space approach calculates a Patterson function from the search model which has been placed into a triclinic cell of suitable dimensions and then applies a rotational operator to the resultant Patterson function, effectively rotating the model in Patterson space. It is necessary to apply an upper limit to the vectors used for the search because of the noise generated by intermolecular vectors.

The direct rotation algorithm changes the order of events so that the intermolecular vectors are no longer a problem. The model is placed in a cell identical to that observed for the experimental data and rotated in real space before the Patterson function is calculated. The direct rotation approach is more accurate and can succeed in cases where the real space method fails. The use of Patterson correlation refinement (3) amplifies the difference between the two methods because of the additional vectors used in the direct rotation algorithm. Unfortunately, the direct space method is computationally intensive because a new Patterson function must be generated for each orientation. The combination of the two approaches in the 'fast direct' algorithm is an efficient and accurate way to determine the rotational component.

Reflections in the resolution range from 15 to 4 Å were used as the experimental data for the search. The coarse rotation search was performed on an angular grid of 12°; the top twenty answers were refined on a 2.4° grid. Results which differed by less than 10° in the three dimensional search space were treated as a single answer. The search was limited to the rotational asymmetric unit for the $P2_1$ space group (4). Patterson vectors between 46 and 5 Å inclusive were used for the real space search. The upper boundary was automatically generated by calculating the furthest distance from the geometric centre of the dimer. Only the lower boundary, necessary to remove the self vector peak, was used for the direct rotation algorithm.

The top five answers from the rotational search are shown in Table 5.1. Answers 1 and 3 are clearly related by an ~180° rotation about θ_1 and describe the answer for one of the dimers and its diad rotation. Answers 2 and 4 describe closely related answers for a second molecule which does not appear to posses diad symmetry. Answer 5 is an incorrect answer. Thus, the results of the rotational search agree with the presence of two dimers in the asymmetric unit. The rotational answers were subjected to PC refinement against the E2E2 target and used in the translational search.

Table 5.1 – Solutions from the Molecular Replacement Rotational Search				
Answer #	**θ_1 (°)**	**θ_2 (°)**	**θ_3 (°)**	**RF-function***
1	95.6	33.0	275.6	0.17
2	197.4	66.7	271.8	0.15
3	250.3	51.4	256.4	0.13
4	149.1	59.4	251.1	0.13
5	238.4	41.7	264.3	0.05

5.1.1.2 Translational Search

Because $P2_1$ is a polar space group, only the x and z coordinates were varied during the search to place the first dimer molecule. The search was performed against the FastF2F2 target†. The grid spacing was 0.33 Å. The results were refined with rigid body refinement against the E2E2 target, treating each protomer as a separate rigid body. The refined answers are tabulated in Table 5.2.

---

* The direct RF-function is defined as:

$$R_{direct}(\Omega) = \langle|E_{obs}|^2|E_m(\Omega)|^2 - \langle|E_{obs}|^2\rangle\langle|E_m(\Omega)|^2\rangle \rangle / (\langle|E_{obs}|^4 - \langle|E_{obs}|^2\rangle^2\rangle * (\langle|E_m(\Omega)|^4 - \langle|E_m(\Omega)|^2\rangle^2\rangle)^{1/2}$$

where:

Ω is a rotational operator matrix

$E_{obs/m}$ are the normalized structure factors for the observed (obs) or modeled (m) data

$\langle\rangle$ indicates an average over all hkl reflections

† The FastF2F2 and E2E2 targets discribe the energy minimization functions in CNS. See reference (1) for further detail.

Table 5.2 – Placement of the First Dimer in the Monoclinic Cell				
Rotation Answer from Table 5.1	**x (Å)**	**z (Å)**	**Correlation Coefficient‡**	**Packing Function§**
1	2.92	3.22	0.36	0.31
2	2.27	34.15	0.34	0.31
3	0.37	42.09	0.34	0.31
4	-1.23	19.82	0.33	0.31
5	2.04	31.75	0.11	0.31

The first answer shown in Table 5.2 was taken as the correct orientation of one of the dimers in the asymmetric unit. It was fixed in place while translational searches with the remaining orientations were carried out to locate the second dimer in the monoclinic asymmetric unit. A third dimension had to be included in the search because of the fixed orientation of the first dimer, an operation which defines the origin of the unit cell along the unique (y) axis in the $P2_1$ spacegroup. Search parameter and energy targets were otherwise the same as those used for the first translational search. Rigid body refinement was applied to all 4 protomers individually, using the highest scoring translational answers from the second search as starting point.

---

‡ The correlation coefficient of the translation function is defined similarly to the rotational function, with regular (unnormalized) structure factors and the rotational operator matrix replaced by a translational vector.

§ The packing function is defined as the ratio of occupied grid points to unoccupied grid points. The occupied grid points are those regions on a finite grid covering the entire unit cell which are within van der Waal's radius of a model atom or its symmetry related mate. Thus a packing function of 1 indicates a unit cell completely filled with protein atoms.

Table 5.3 – Placement of the Second Dimer in the Monoclinic Cell					
Rotational Answer from Table 5.1	**x (Å)**	**y (Å)**	**z (Å)**	**Correlation Coefficient**	**Packing Function**
2	33.9	91.1	-26.0	0.41	0.63
4	0.6	1.3	-40.3	0.40	0.63
3	31.6	24.3	-18.3	0.30	0.37

The refined translational components are tabulated in Table 5.3. The highest scoring answer, obtained with rotation answer 2, was taken as correct. The difference in the correlation coefficient was very slight over the translational answer obtained with rotational answer 4. The very similar nature of the two answers is likely a result of the refinement techniques (*e.g.* rigid body, PC refinement) used to improve individual answers. The packing function result obtained with rotational answer 3 suggests that it was placed on top of the dimer located in the first translational search. This agrees with the previous observation that rotational answers 1 and 3 are diad related.

5.1.2 Model Refinement

The software used for both reciprocal space refinement and graphic rebuilding was identical to that described previously for the refinement of the tetragonal crystal structure in chapter 4.

The model was refined initially without NCS restraints using simulated annealing in torsion space. Rounds of simulated annealing

refinement were alternated with grouped[**] B-factor refinement, reducing the R_{factor} and R_{free} from 0.453 and 0.480, respectively to 0.342 and 0.427, respectively. At this point, no further improvement in the R factors could be accomplished with the reciprocal space refinements, *i.e.* manual intervention was necessary to complete the refinement. The options at this juncture were to undertake model rebuilding efforts, or to establish a way of increasing the radius of convergence for the reciprocal space refinement. The latter option was a more efficient route to the correct answer because model rebuilding is generally the rate limiting step during structure refinement. Model rebuilding against the monoclinic data required twice as much work as what was required for the tetragonal data due to the presence of an extra dimer in the asymmetric unit.

One way of improving the radius of convergence is to introduce restraints on the variables being affected by the refinement. This can be accomplished through the use of NCS operators. By examining combinatorial overlaps of the Cα positions for the individual protomers, it was established that protomer A[††] was most similar to protomer C and that protomer B was most similar to protomer D. The resulting NCS symmetry

---

[**] The B factors were grouped into 2 values / residue such that one set represented the sidechain atoms while the others represented the backbone peptide atoms.
[††] The dimers in the asymmetric unit are defined such that dimer 1 consists of protomers A and B and dimer 2 consists of protomers C and D.

model reflected an *inter*dimer symmetry rather than an *intra*dimer symmetry for the protomers in the monoclinic cell.

The alternating simulated annealing and grouped B factor refinement was restarted from the rigid body refined molecular replacement solution with two NCS restraints describing the relation between protomers A and C and between protomers B and D, respectively. The resulting R_{factor} and R_{free} fell from 0.453 and 0.480, respectively to 0.295 and 0.347, respectively before manual intervention was necessary. Thus, the introduction of NCS restraints allowed the model to be refined in a more satisfactory manner. In addition to limiting the inefficient graphical rebuilding steps, the improved refinement yielded a more accurate model which could be used to create more interpretable difference maps, thereby accelerating the entire refinement process.

Alternating rounds of reciprocal space refinement with model rebuilding using σ_A weighted 2Fo-Fc and Fo-Fc maps as a guide brought the structure to a state in which the placement of ligands was possible. The ligands in question are the L-Hse and the NADA molecule included in the crystallization conditions. The R factors at this point were 0.236 and 0.298 for R_{factor} and R_{free}, respectively. Difference maps generated before this stage of refinement were ambiguous with regards to the presence of ligand molecules.

The only active site which showed clear electron density for both the L-Hse molecule and the NADA molecule was molecule D. There was no interpretable density in the other active sites. After placing the ligands in the active site of monomer D, the NCS restraints between protomers B and D were removed, under the assumption that ligation would result in conformational changes of molecule D relative to B. The NCS restraints between A and C were reduced to cover only those portions of the molecules which remained identical, as identified from visual inspection of the overlap of C_α positions of the two protomers. In the final rounds of positional refinement, all NCS restraints were removed. The B factor refinement was expanded such that individual B factors were modelled towards the latter stage of refinement so that the best possible difference maps could be calculated for the placement of the ligands. The individual B factor refinement was initiated when the R_{factor} and R_{free} were at 0.236 and 0.298, respectively. These values improved slightly to 0.234 and 0.297 after 20 cycles of individual B factor refinement. NCS restraints were maintained during B factor refinement and were justified by the behaviour of the R_{free} statistic which increased without the NCS restraints in place. The data to parameter ratio was approaching unity with the introduction of individual B factors. This poor ratio likely necessitated the NCS restraints during the B factor refinement.

Two rounds of water picking on the Fo-Fc maps were undertaken to place 248 water molecules. Further rounds of water picking after this point did not improve the R factors. The metal atom positions were located in the first round of water picking. These were modeled as sodium atoms based upon the arguments presented in chapter 7.

After water picking was finished, a simulated annealing omit map of the L-Hse molecule and cofactor analogue was used to guide manual rebuilding of these ligands. The B factors for the rebuilt ligands were set to the average value of the surrounding residues (35 $Å^2$). The ligands were then subjected to simulated annealing in Cartesian space. All other components in the model (protein, water, and metal atoms) were held fixed during the simulated annealing runs. The refinement was repeated multiple times because there is some random element inherent in the simulated annealing process. Each run refined the atomic coordinates of the ligands to precisely the same positions, so the final refined positions of the ligands are an accurate result. The refined ligand coordinates agreed well with the omit maps. The combined model rebuilding, simulated annealing, and NCS restrained individual B factor refinement resulted in a reduction of R_{factor} and R_{free} from 0.220 and 0.286, respectively to 0.217 and 0.285, respectively. Another round of water picking was performed added 103 water atoms to produce the near final model. A final water molecule, Wat721, was positioned based upon σ_a weighted Fo-Fc

simulated annealing omit maps centred around the L-Hse molecule in the D protomer. The final refinement statistics are shown in Table 5.4 and the Ramachandran plots are shown in Figure 5.1.

Table 5.4 – Refinement Statistics for the Monoclinic Structure[‡‡]	
Resolution (Å)	40 - 2.6
Multiplicity	3.7
# Unique Reflections	44116 (44409)
R_{factor}	0.215 (0.217)
R_{free}	0.283 (0.285)
Total atoms	11636
HSD monomers	4
Substrate molecules	1 NADA, 1 L-Hse
Ions	4 Na^+
Solvent atoms	371
Bulk Solvent Model	flat (k=0.322, B=34.08 $Å^2$)
R.m.s. bonds (Å)	0.008
R.m.s. angles (°)	1.5
R.m.s. torsions (°)	23.3
Coordinate error[§§] (Å)	0.52
Ramachandran plot: % residues in allowed region (core)	99 (86)

[‡‡] Values were calculated with a 2σ cut off on amplitudes, the same cut off used throughout the refinement. Where appropriate, the final values obtained with no data cut off are given in round brackets.

[§§] This is the cross validated σ_a coordinate error.

Figure 5.1 (A and B) Ramachandran Plots of Monoclinc Models

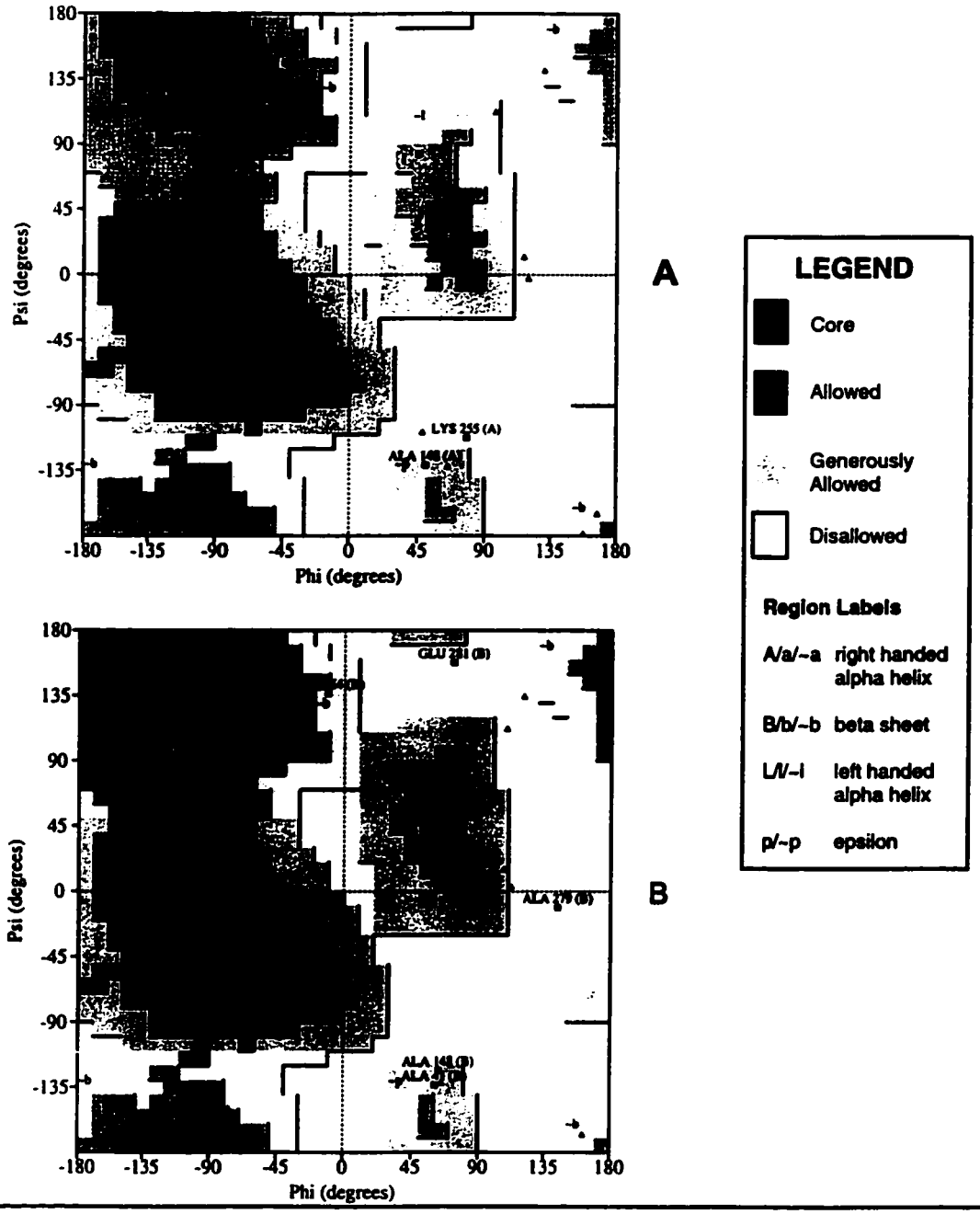


Figure (A) and (B) are the Ramachandran plots for protomers A and B (of the monoclinic crystal form), respectively. Residues in the disallowed regions are labelled with residue number and type. The core, allowed, and generously allowed regions are shaded according to the legend given above and also with appropriate letters in upper case, lower case, or preceeded by a tilda, respectively (*e.g.* A, a, ~a for the alpha helical region). Definitions for the region are as given in Morris, A. L., MacArthur, M. W., Hutchinson, E. G., and Thornton, J. M. (1992). Stereochemical quality of protein structure coordinates. Proteins. 12: 345-364.

Figure 5.1 (C and D) Ramachandran Plots of Monoclinc Models

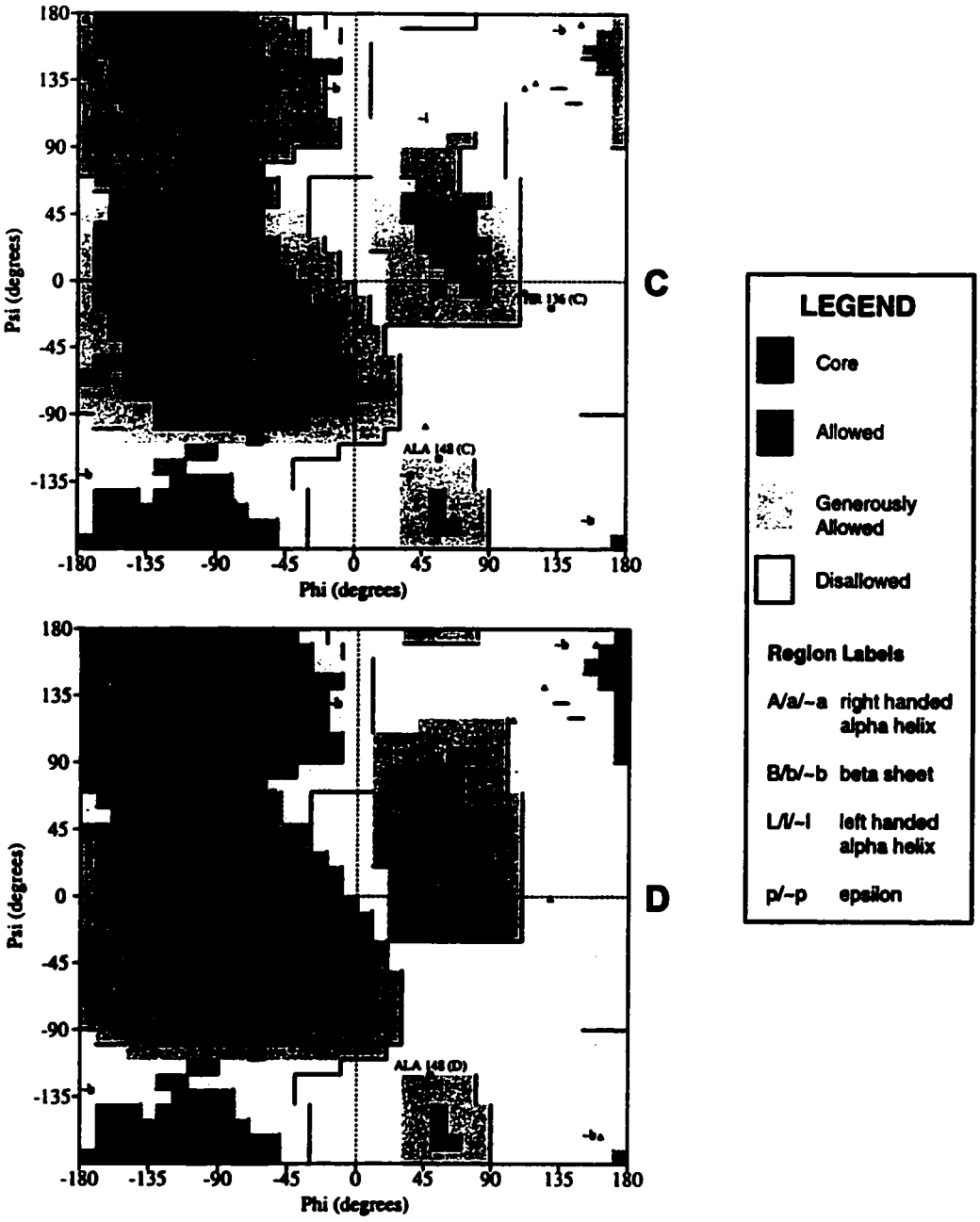


Figure (C) and (D) are the Ramachandran plots for protomers C and D (of the monoclinic crystal form), respectively. Residues in the disallowed regions are labelled with residue number and type. The core, allowed, and generously allowed regions are shaded according to the legend given above and also with appropriate letters in upper case, lower case, or preceeded by a tilda, respectively (*e.g.* A, a, ~a for the alpha helical region). Definitions for the region are as given in Morris, A. L., MacArthur, M. W., Hutchinson, E. G., and Thornton, J. M. (1992). Stereochemical quality of protein structure coordinates. Proteins. 12: 345-364.

5.2 Results and Discussion

5.2.1 Crystal Packing

The crystal contacts are shown in Figure 5.2. Crystallographic contacts are defined as atoms in the models that are within 3.6 Å of a symmetry related molecule. They affect 10.1, 8.9, 6.7, and 7.8% of the residues for protomers A, B, C, and D, respectively.

The L-Hse molecule and NADA molecule are bound to only one of the four protein molecules, the D protomer. The crystallization conditions included the ligands at a minimum 3:1 ligand:protomer molar ratio, so one would expect a random distribution among the active sites. Since a random distribution is not observed, there must be something influencing the ligand binding. The most likely place to find an answer for the asymmetric ligation is in the crystal packing. However, none of the active sites are blocked by symmetry related molecules, as was the case for the asymmetric binding of NAD^+ in the tetragonal crystal form. Furthermore, there are no clear solvent channels which may explain the presence of ligands in only one of the active sites.

Figure 5.2 - Crystallographic Contacts in the Monoclinic Structure

Segments of the A, B, C and D protomers of the monoclinic crystal form involved in crystal contacts (< 3.6 Å interaction) resulting from crystal packing are indicated in red. Molecules are shown in standard views. Figures generated with *Molscript/Raster3D*.

The true reason for the asymmetric ligation is subtle, but points toward an important component of catalysis. For the D molecule, there are crystallographic contacts which occur at the C terminus of α-J at Glu 183 and in the J-K loop, specifically at Thr 186, Gln 188, and Asn 190. The J-K loop is an omega loop (5); these structures have limited flexibility so contact at any point would affect the other secondary structures joined by this loop, *i.e.* α-J and α-K. Of the other 3 protomers, only molecule B shows a contact in this area at Asp 191 in the J-K omega loop. The overall effect of the crystallographic contacts for both the B and D protomers is a translation of α-J in the direction of its N terminus; this difference is 1.1 Å for the D molecule and approximately 0.5 Å for the B molecule, both with respect to the A/C molecules. The result of the positional change of α-J is to bring it closer towards the L-Hse binding site, thereby moving the helical dipole into a position where it can interact with the bound L-Hse more favourably. In this conformation, the protein apparently exhibits a greater affinity for the L-Hse molecule because of the increased binding energy provided by the α-J dipole.

To understand why the NADA molecule is also found only in the D protomer, one must consider the established order of substrate binding for HSD. The NADA molecule is a substitute for the NAD^+ molecule which dissociates from HSD after the L-Hse molecule has departed (6) (Figure

6.2). This kinetic model is supported by the crystallographic data regarding how the two ligands bind to the active site. The L-Hse molecule appears to block the dissociation of NADA from the enzyme.

5.2.2 Differences Between Holo Structure and Ternary Complex

In comparing the holo and ternary complex, the region of interest lies principally within the catalytic domain, specifically those structural elements which bear the catalytic residues. These elements are α-J, α-L, α-M, and the K-L loop. Unfortunately, there are numerous crystal packing contacts for helices α-J and α-K and their intervening 10 residue omega loop. Although none of the residues implicated in catalysis are directly involved in crystal packing contacts, any consideration of differences which occur in these regions as a result of binding L-Hse must be done under the caveat of crystal packing forces.

Overall, the motions are subtle – there are no gross domain motions. This was also true of the difference between the dinucleotide bound and unbound structures. A comparison of the r.m.s.d. differences in the C_α positions for the 6 different crystallographic copies of HSD is presented in Table 5.5. The comparison uses the alignment of C_α positions from residues 5 through 355.

Table 5.5 –Cα r.m.s.d. of All 6 Determined HSD Protomers (values expressed in Å)					
Protomer*	**B1 (apo)**	**A2 (apo)**	**B2 (apo)**	**C2 (apo)**	**D2 (ternary)**
A1(holo)	1.1	0.8	0.9	0.8	0.9
B1(apo)		1.4	1.3	1.3	1.2
A2(apo)			0.9	0.4	0.8
B2(apo)				0.8	0.7
C2(apo)					0.8

From these comparisons it is apparent that the enzyme has a fairly rigid structure. This can also be seen when the three different catalytic forms (apo, holo, ternary) are overlaid upon one another, as shown in Figure 5.3A. There is independent evidence for the rigidity of the enzyme during the catalytic cycle. The tryptophan fluorescence spectrum does not differ when ligands are added to the protein, which means that the environment of the tryptophan residues remains consistent throughout the catalytic cycle (6). The location of the two tryptophan residues in the protein structure is shown in Figure 5.3B.

The overall rigidity can be explained by both the tertiary and quaternary structures of the protein. The double hinge construction was described in the previous chapter; the multimeric nature of the enzyme will

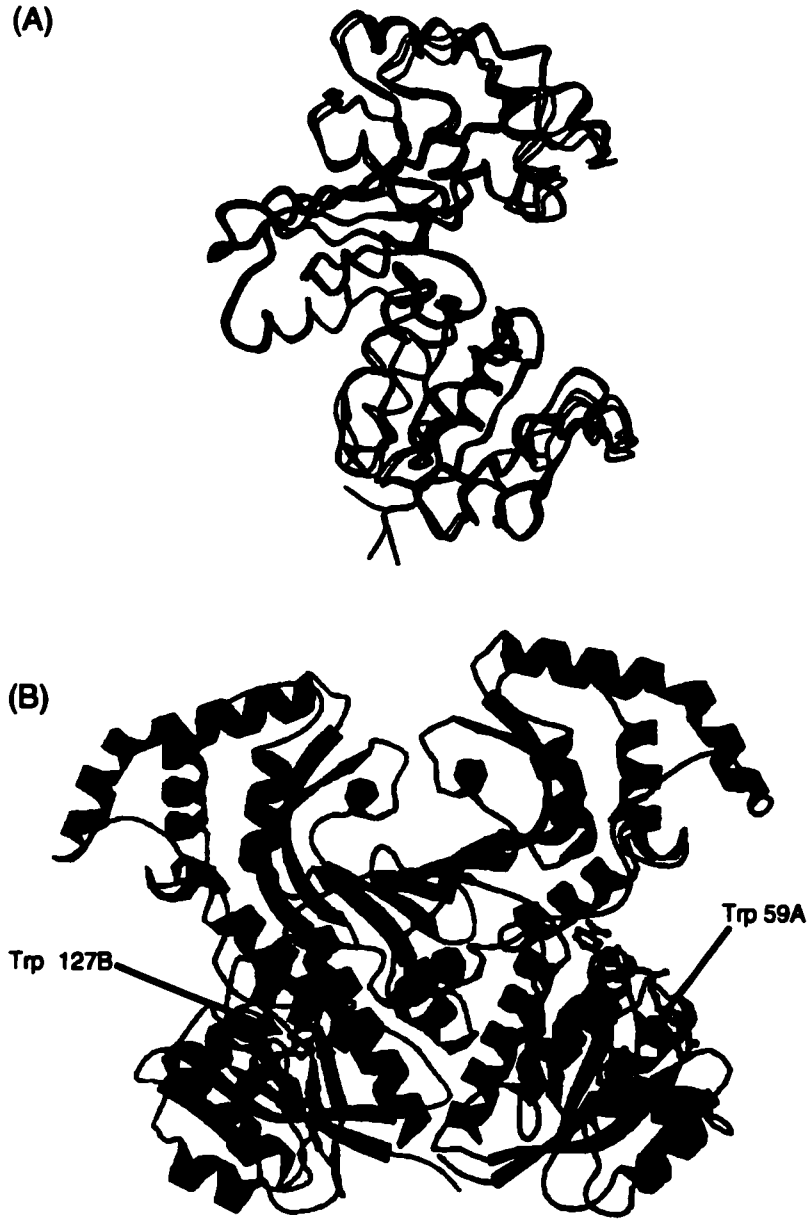
---

*** Protomers are labelled A, B, C, or D within crystal forms (C/D for monoclinic only), and 1 or 2 representing tetragonal or monoclinic forms, respectively.

be discussed here. Apart from the L-Thr inhibition, there is scant evidence of cooperativity in the enzymatic mechanism for HSD from any species. The conformational motion of a single protomer is, therefore, restricted by the contacts it makes with its partner through the dimer interface. This point is further supported by a consideration of the group 2 HSD enzymes. The group 2 enzymes function as tetramers; there would be numerous contacts disturbed if the catalytic mechanism required gross conformational changes in the protein.

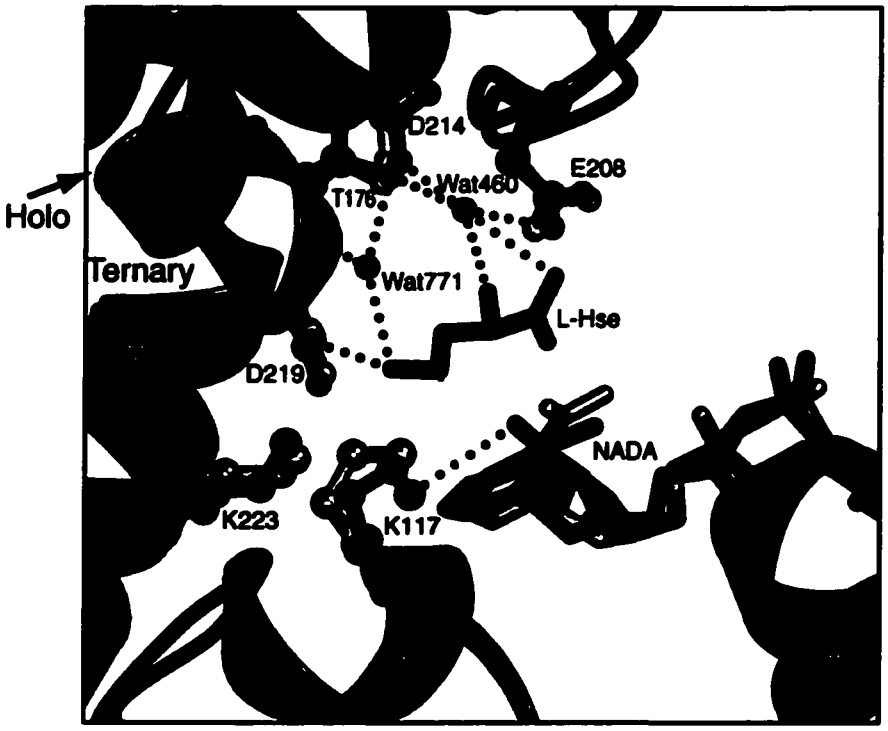
The subtle protein motions which do occur arise from the requirement that during the catalytic mechanism, the cofactor molecule and substrate molecule must be brought into a position which facilitates hydride transfer. These motions can be divided into two types: those which affect the bound L-Hse molecule and those which affect the NADA molecule. The observed differences between the holo and ternary form active site structural components are shown in Figure 5.3C.

Figure 5.3 - Crystallographic Evidence for Minimal Domain Motion Between the Apo, Holo, and Ternary Forms of HSD



(A) Overlap of the Cα positions of the apo (B1), holo, and ternary complex. The structures are coloured green, red, and blue respectively and are shown in standard view. *(B)* The location of the two tryptophan residues, W59 and W127, are shown on the HSD dimer. Note that W127 inserts itself between the dimerization region and dinucleotide binding region. The environment of this residue would be altered considerably if large domain motions were to occur. Figures generated with *Molscript/Raster3D*.

Figure 5.3C - Overlap of the Holo and Ternary Protomers at the Active Site



A superposition of the holo protomer (chapter 4) and the ternary protomer discussed in this chapter. The view is approximately the same as that shown in Figure 4.11. The holo form is semitransparent, the ternary form is solid. Key residues are shown and labeled, and hydrogen bonds are indicated with dashed red lines. The two water molecules present in the ternary complex are shown. The secondary structure elements are coloured as before, with red for the dinucleotide binding region and blue for the catalytic domain. The NADA molecule is coloured by atom, with the C_4 carbon of the pyridyl ring ccloured in black. The NAD^+ molecule is semitransparent. The L-Hse molecule is coloured green on its carbon atoms with its C_4 carbon atom coloured in black. The overlap is based on a global alignment of the two protomers. Positional differences for the secondary structural elements are small *e.g.* the α-J helix moves 1.1 Å from the holo to the ternary form.

5.2.2.1 Conformational Differences Affecting L-Hse

With respect to the holo enzyme, α-J exhibits a shift towards the active site on the order of 0.8 Å. This brings the N terminal helix dipole closer to the substrate hydroxyl oxygen atom. Helix α-K also exhibits a large difference with an overall r.m.s.d. of 1.0 Å for the C_α atoms of residues 194-204. The implied motion of the helix is towards the bound L-Hse, and is also seen in the K-M loop. The K-M loop residues 205-211 show an overall r.m.s.d. of 0.9 Å, again in the direction of the bound L-Hse molecule. The differences are less for the subsequent helices α-L and α-M, both of which bear important residues. Helix α-L, which contains Asp 214 has an overall r.m.s.d. of 0.4 Å of its C_α atoms and α-M, which contains both Asp 219 and Lys 223, has an overall r.m.s.d. of 0.4 Å of its C_α atoms. Neither of the helices have residues which are directly affected by crystal packing interactions, so these smaller motions may be more indicative of the type of motions which occur for the *in vivo* state of the protein, underscoring the overall rigidity of the molecule. This degree of motion is within the error values associated with the holo and ternary complexes (Tables 4.1 and 5.4 respectively), so the α-L and α-M helices likely exhibit little to no motion during the catalytic cycle.

5.2.2.2 Conformational Differences Affecting NADA

When considering the dinucleotide binding region, the most obvious region for scrutiny is α-B. The positional change of the helix is a tilting motion of 2.6° towards the active site for the ternary complex with respect to the holo complex. The C_α atoms of the helix show an overall r.m.s.d. of 0.7 Å. The effect of this difference upon the cofactor molecule can be examined by measuring the change in the phosphorus atom positions. These atoms are the most electron rich in the entire molecule and should therefore be the most accurately determined atomic positions in the ligand (7). The phosphorus atom closest to the pyridyl ring moves 0.62 Å while the phosphorus atom closest to the adenine ring moves by 0.54 Å. The overall difference can be described by a combined twist and translation that is communicated down the entire length of the NADA molecule.

The other region to consider is the 3_{10} helix α-1F which bears Lys 117. This helix changes only slightly: it exhibits a difference of 0.3 Å r.m.s.d. away from where the cofactor is bound in the holo form enzyme. The Lys 117 residue changes the conformation of its side chain such that the ε-amino group is 1.7 Å closer to the original (holo) position of the NAD molecule. Note that this change opposes the overall positional difference of α-1F. This motion is required to maintain optimal contacts with the ribose $C_{2'}$ hydroxyl group. Helix α-2F exhibits a 0.5 Å r.m.s.d. in a similar direction to α-1F for the ternary complex with respect to the holo complex.

These observations suggest that Lys 117 plays an important role in controlling the orientation of the cofactor.

Other significant motions in the dinucleotide binding region occur, but their functional significance is not entirely obvious. The elements β-b and α-E, portions of which comprise the nominal adenine region of the dinucleotide binding pocket, each exhibit positional changes which mirror that of the adenine ring. These differences are on the order of 0.7 Å but are affected by at least 4 crystal contacts (1 for α-E and 3 for β-b), so their significance is unclear.

5.2.3 Geometry of Ligands and of the Active Site

The dual presence of the L-Hse and NADA molecules in the active site of the enzyme is an invaluable aid to determining how HSD is catalyzing the hydride transfer between the cofactor and the substrate molecules (see Figures 5.4 and 5.5). The ternary complex gives a precise picture of what is occurring for the reverse reaction whereby L-Hse is converted to ASA with concomitant reduction of NAD^+ to NADH. The catalytic and/or substrate binding residues may have been identified from the binary complex alone by site-directed mutagenesis of conserved residues in the vicinity of the pyridyl ring of the bound cofactor. However, a number of key features important to the reaction mechanism could only have been established by or discovered from the crystallographic data.

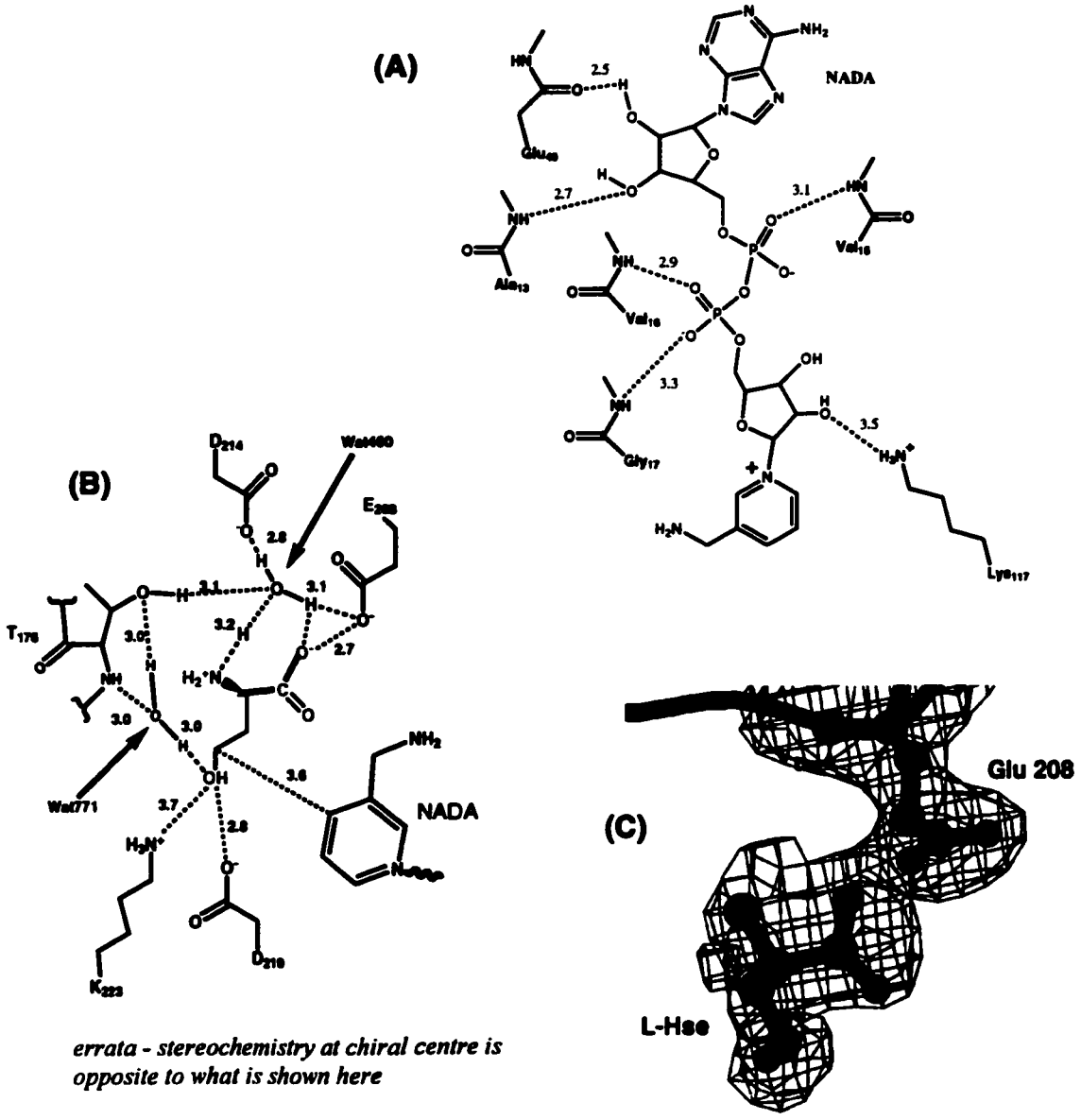
Specifically, the presence of an N-terminal helix dipole directed at the active site, two water molecules in the active site, and an unusual arrangement for binding the carboxylate group of the substrate molecule.

With a few minor exceptions, the active site residues are strictly conserved. The exceptions are the replacement of Glu 208 with an arginine residue in *Pseudomonas aeruginosa*, and the functional conservation of Thr 176 as either a threonine residue or a serine residue. The carboxylate group of the L-Hse molecule is located only 2.7 Å away from the side chain carboxylate group of Glu 208. Within experimental error, the interaction could be considered a low barrier hydrogen bond.

The definition of a low barrier hydrogen bond is that the participants are 2.6 Å or less from each other and have matched pKa values (8, 9). Since both oxygen atoms are part of carboxylate groups in amino acid molecules, it is likely that the second criterion of the definition is met as well. Low barrier hydrogen bonds are very strong interactions and can be considered essentially covalent. Their energy of formation approaches 30 kcal/mol in the gas phase. The energy derived from their formation can be used to drive catalysis, but it is unclear how this would be so for the L-Hse molecule bound to HSD.

The apposition of two negatively charged oxygen atoms must be balanced by a positively charged atom for a stable interaction to exist. The most likely candidate is a hydrogen atom from the nearby water

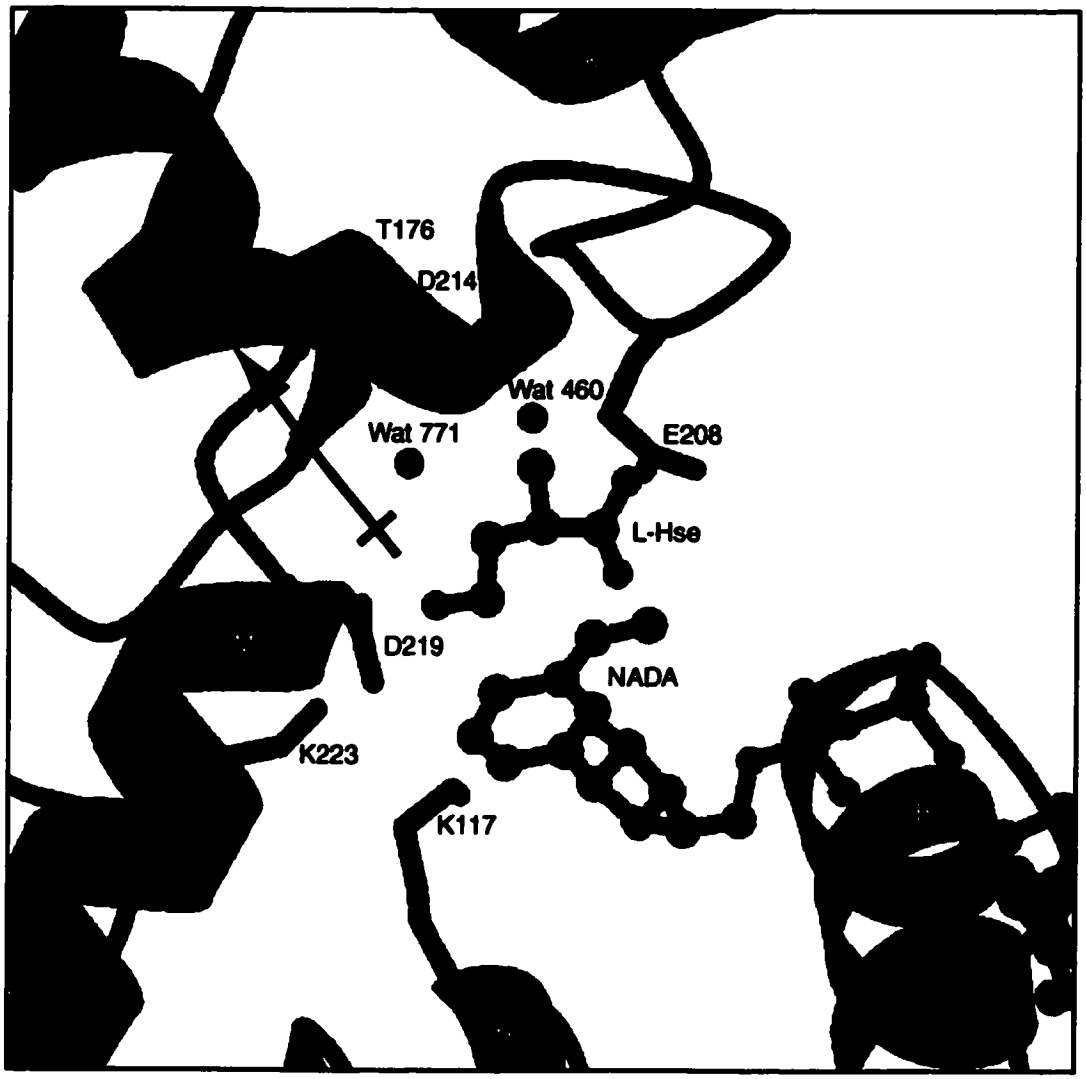
Figure 5.4a - Ligation of L-Hse and NADA by Active Site Residues



errata - stereochemistry at chiral centre is opposite to what is shown here

(***A***) Protein-cofactor interactions of the NADA molecule in the ternary complex. Unlike the binary complex, there are no apparent intervening water molecules. Also note the increased number of contacts over those observed in the binary complex (*cf.* Fig. 4.8a). (***B***) Hydrogen bond contacts for the L-Hse molecule and active site water molecules. The values in panels (A) and (B) are in Å and indicate distances between non-hydrogen atoms. Hydrogen atoms are included to illustrate the hydrogen bonding network. (***C***) A σ_a weighted Fo-Fc SA-omit map generated around the bound L-Hse molecule and Glu 208. The map is contoured at 3σ. The density clearly shows the presence of the ligand and the close interaction of the two carboxylate groups as discussed in the text.

Figure 5.4b - Active Site with Bound Ligands



A view of the active site in the ternary complex. Helix α-G has been removed for clarity. The cofactor analogue (grey carbon atoms), reaction product (green carbon atoms), and key residues are depicted. Two key water molecules are also shown. The carbon atoms involved directly in hydride transfer are coloured in black. The N-terminal (*i.e.* positive) dipole of helix α-J points directly at where the oxyanion will form during the hydride transfer. The dipole is indicated with standard notation (arrowhead indicates the negative end).

molecule (Wat 460). The oxygen of the water molecule is 3.1 Å from the Glu 208 carboxyl oxygen and 3.4 Å from the carboxyl oxygen of bound L-Hse molecule. The water molecule appears to be a key component for the substrate binding. Wat 460 also forms hydrogen bond interactions with the amino group of the L-Hse molecule (3.2 Å), the hydroxyl group of the functionally conserved Thr 176 (3.1 Å), the side chain carboxylate of Asp 214 (2.8 Å) and the backbone amide nitrogen of Glu 208 (3.3 Å).

The second water molecule in the active site, Wat 771, is found 3.2 Å away from the first Wat 460. This water molecule forms hydrogen bond interactions with both the backbone amide nitrogen of Thr 176 and the sidechain hydroxyl of Thr 176. It also hydrogen bonds to the hydroxyl group of the bound L-Hse molecule. Furthermore, it is close (3.7 Å) to the C_4 carbon of the bound L-Hse. This has an important implication for the nature of the bound substrate in the forward direction reaction, a point which will be discussed in the following chapter.

At the opposite end of the substrate molecule, the hydroxyl group of L-Hse hydrogen bonds with the side chain carboxylate of Asp 219 (2.8 Å) and is close (3.7 Å) to the side chain amino group of Lys 223. The disparity between the two bond lengths is significant and of some importance to the reaction mechanism proposed in Chapter 6.

Lys 117 is also relatively near to the hydroxyl group (4.7 Å) but is considerably closer (3.5 Å) to the $C_{3'}$ hydroxyl group of the pyridyl end

ribose ring in the bound NADA molecule. Lys 117 interacts indirectly with the L-Hse molecule in that Asp 219 is flanked by both Lys 117 (3.9 Å) and Lys 223 (2.9 Å). Unlike the interaction between Lys 117 and the NAD^+ molecule in the holo form, there are no crystallographic water molecules intervening between the residue and the cofactor, suggesting that a water molecule in this location is not crucial to the reaction mechanism.

5.2.3.1 Hydride transfer Geometry

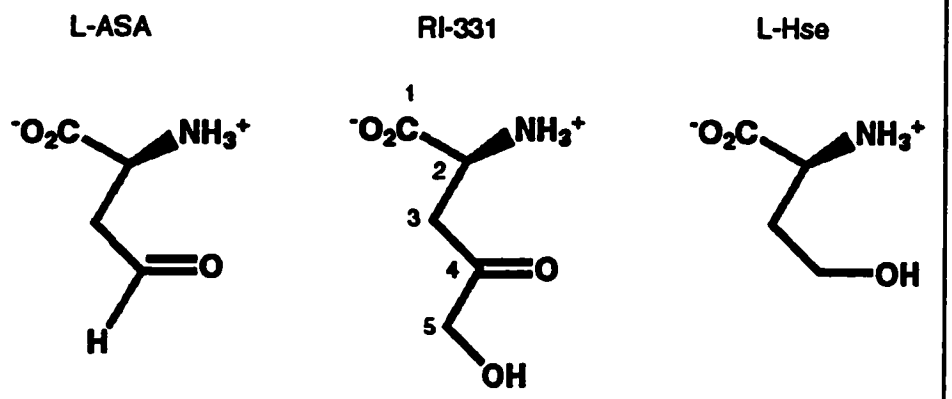
The C_4 carbon atom of the L-Hse group is located 3.6 Å away from the hydride bearing (C_4 on the pyridyl ring) carbon of the NADA molecule (Figure 5.4 and Figure 5.5). Furthermore, the observed geometry is in agreement with what has been observed in other ternary complexes of hydride transfer reactions. Other crystal structures of ternary complexes of NAD^+/NADH binding enzymes with substrates and cofactors exhibit distances between the atoms involved in hydride transfer of 3 to 5 Å and ($N\text{-}C_4\text{-}H$) angles < 180° (references cited in (10)).

5.2.4 Implications for RI-331 Inhibitor Mechanism

The ternary complex can be used to make a supposition on the nature of the inhibitory complex that RI-331 forms with the cofactor. RI-331 is a competitive inhibitor of HSD (11). This implies that both RI-331 and L-Hse bind in a similar manner to the enzyme active site. The

chemical structures of these two amino acids are quite similar (see Figure 5.6). Thus, one would could predict that RI-331 will bind in a manner highly similar to the binding observed for L-Hse.

Figure 5.6 – Chemical Structure of L-ASA, RI-331, and L-Hse



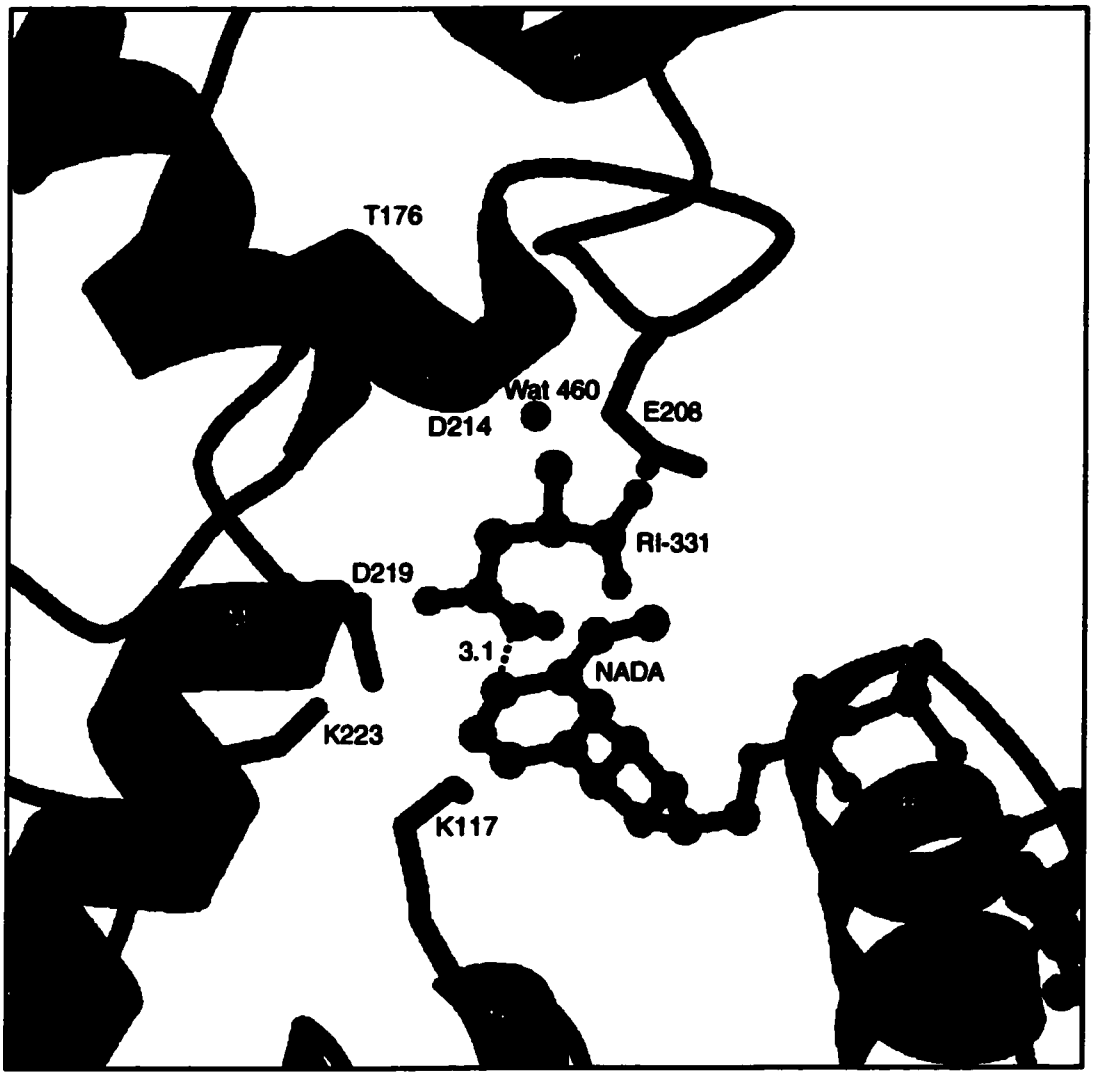
Furthermore, others (6) have suggested that RI-331 forms a covalent complex with the oxidized cofactor because preincubation of the enzyme with NAD^+ augments the inhibitory effects. The NADA molecule in the ternary structure is an excellent model for how the NAD^+ molecule binds. The difficulty Jacques had in obtaining an isolated RI-331•NAD complex suggests that the two ligands must bind to the enzyme for the complex to exist.

Therefore, simply replacing the bound L-Hse in the stuctural model with an energy minimized form of RI-331 should give an accurate picture of the situation in the active site before the covalent complex is formed. This is shown in Figure 5.7. There are two very interesting features

apparent in this model. First, the carbonyl oxygen of RI-331 aligns very well with the hydroxyl oxygen of L-Hse. This puts the carbonyl under the influence of the N-terminal dipole from α-J. Second, the C_5 carbon of the RI-331 molecule is 3.1 Å away from the C_4 carbon of the pyridyl ring of the NADA molecule. The helix dipole will stabilize the enolate form of RI-331, and the C_5 carbon of RI-331 is within striking range of the C_4 position of the pyridyl ring of the NAD molecule. A reasonable reaction mechanism for the formation of the RI-331·NAD^+ complex can be proposed on the basis of this model (shown in Figure 5.8).

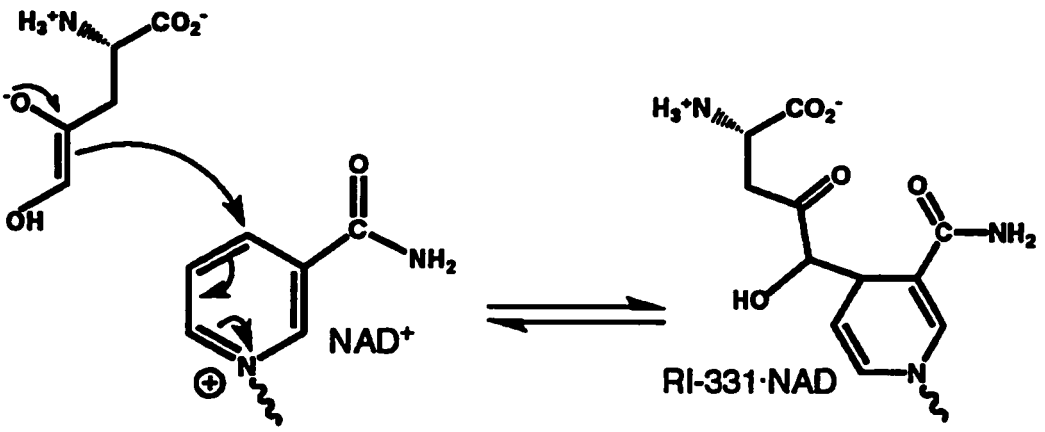
There is evidence for this type of chemistry from another enzyme. Benach *et al.* (12) obtained a series of high resolution data sets from crystals of complexes between a variety of small ketones and the NAD^+ molecule bound to *Drosophila* alcohol dehydrogenase. The complexes are clearly visible in the electron density generated from the crystallographic data. The proposed chemistry entailed an enolate species attacking the C_4 postion of the pyridyl ring, similar to what is being proposed for the RI-331 complex formation here.

Figure 5.7 - RI-331 Modeled Into the Active Site



RI-331 modeled into the active site based on the L-Hse coordinates. Atoms involved in putative bond formation are shown in black, with the distance indicated in Angstroms. The RI-331 molecule was subjected to energy minimization prior to being modeled into the active site.

Figure 5.8 – Proposed Reaction Mechanism for RI-331 NAD^+ Complex Formation



5.3 Summary of Chapter 5

The ternary complex of HSD•NADA•L-Hse was solved *via* molecular replacement with coordinates from the binary complex. The structure was refined to an R_{factor} and R_{free} of 0.215 and 0.283, respectively. Data to 2.6 Å was used for the refinement.

The key residues for the amino-acid substrate binding are Thr 176, Glu 208, and Asp 214. The Glu 208 residue interacts in an unusual fashion with the carboxylate group of the bound amino acid substrate. Two water molecules are bound with the amino acid substrate.

The catalysis of the hydride transfer is putatively accomplished by two residues: Asp 219 and Lys 223. The nicotinamide cofactor analogue

makes more direct contacts with the dinucleotide binding region of the enzyme.in the ternary complex than the NAD^+ molecule does in the binary complex.

The cofactor analogue and substrate could only be located in only one of the protomer active sites. The probable reason for this was the slight translation of α-J and its resultant dipole towards the location of the bound L-Hse molecule. The positional difference is apparently due to the effects of crystal packing forces from symmetry related molecules upon the J-K omega loop. No gross domain motions among the three different forms (apo, holo, and ternary) were observed.

The determination of the ternary complex permitted the modeling of a RI-331 into the active site of the enzyme. The modeled HSD•NAD^+•RI-331 complex suggests that the formation of a covalent complex between RI-331 and NAD^+ through an RI-331 enolate intermediate is possible.

Chapter 5 References

1. **Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L.** (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr. **54**: 905-921.
2. **DeLano, W. L., and Brunger, A. T.** (1995). The direct rotation function: rotational patterson correlation search applied to molecular replacement. Acta Cryst D. 740-748.
3. **Brunger, A.** (1990). Extension of Molecular Replacement: a New Search Strategy based on Patterson Correlation Refinement. Acta Cryst. **A46**: 46-57.
4. **Rao, S. N., Jih, J.-H., and Hartsuck, J. A.** (1980). Rotation-Function space Groups. Acta Cryst. **A36**: 878-884.
5. **Leszczynski, J. F., and Rose, G. D.** (1982). Loops in globular proteins: a novel category of secondary structure. Biochimie. **64**: 203-209.
6. **Jacques, S.**, Characterization of Homoserine Dehydrogenase from *Saccharomyces cerevisae*: A New Fungal Target. Doctoral Thesis, *Dept of Biochemistry*, McMaster University, Hamilton, Ontario, Canada (1999).
7. **Cruikshank, D. W. J.** (1999). Remarks about Protein Structure Precesion. Acta Cryst. **D55**: 583-601.
8. **Gerit, J. A., Kreevoy, M. M., Cleland, W., and Frey, P. A.** (1997). Understanding enzymic catalysis: the importance of short, strong hydrogen bonds. Chem Biol. **4**: 259-267.
9. **Ash, E. L., Sudmeier, J. L., De Fabo, E. C., and Bachovchin, W. W.** (1997). A low-barrier hydrogen bond in the catalytic triad of serine proteases? Theory versus experiment. Science. **278**: 1128-1132.
10. **Mesecar, A. D., Stoddard, B. L., and Koshland, D. E., Jr.** (1997). Orbital steering in the catalytic power of enzymes: small structural changes with large catalytic consequences. Science. **277**: 202-206.
11. **Yamaki, H., Yamaguchi, M., Tsuruo, T., and Yamaguchi, H.** (1992). Mechanism of action of an antifungal antibiotic, RI-331, *(S)* 2-amino-4- oxo-5-hydroxypentanoic acid; kinetics of inactivation of homoserine dehydrogenase from *Saccharomyces cerevisiae*. J Antibiot (Tokyo). **45**: 750-755.
12. **Benach, J., Atrian, S., Gonzalez-Duarte, R., and Ladenstein, R.** (1999). The catalytic reaction and inhibition mechanism of *Drosophila* alcohol dehydrogenase: observation of an enzyme-

bound NAD-ketone adduct at 1.4 Å resolution by X-ray crystallography. J Mol Biol. **289**: 335-355.

Chapter Six: Proposed Reaction Mechanisms

6.1 Review of Dehydrogenase Reaction Mechanisms

The active site of HSD is significantly different from other structurally characterized NAD dependent dehydrogenases. This observation suggests a novel catalysis of the hydride transfer between the cofactor and the substrate. Thus, before proposing a reaction mechanism for the hydride atom transfer from NADH to L-ASA catalyzed by HSD, it would be worthwhile to examine how this same reaction is catalyzed by other NAD dependent dehydrogenase enzymes. The most recent review covering the mechanisms of a variety of dehydrogenases was presented 15 years ago by Fersht (Chap 15 in (1)). It was apparent even from this earlier review that hydride transfer chemistry could be accomplished by a diverse range of catalytic mechanisms.

Enzyme catalyzed hydride transfer reactions are generally reversible processes. The exceptions to this generalization are those enzymes that combine addition or elimination reactions along with the hydride transfer, *e.g.* nucleophilic water attack to generate acids from oxidized primary alcohols. In such instances, the overall reaction may not

be reversible; the irreversibility is not, however, due to the hydride transfer.

Physiological hydride transfers are observed in both reductive and oxidative processes. The reductive process is the physiologically relevant direction for HSD. However, the principle of microscopic reversibility states that the transition state structure will be identical for both the forward and reverse directions (Chapter 2 in (1) and references cited therein). The transition state of the reaction catalyzed by HSD occurs during hydride transfer. For this reason, examples will be taken from both oxidative and reductive processes, with no special importance attached to the distinction between the two. Attention will be focused on the actual hydride transfer step, as this is the common feature relevant to the mechanism for HSD.

With only a few exceptions, the reaction mechanisms discussed below were established by data from crystal structures and enzyme kinetics. Enzyme kinetics is a term used herein to describe experiments involving, in various combinations, the following: activity *vs.* pH profiles of substrates and substrate analogues, kinetic isotope effects, dead-end inhibitor studies, and the effects upon enzymatic activity resulting from the change of one or more residues in the protein.

The two approaches are complementary: crystal structures provide a clear but static view of one or more steps in the reaction process

whereas enzyme kinetics provide insight into the dynamic nature of reaction mechanisms. Reaction mechanism models arising out of the integration of both approaches represent the most accurate reaction models and should therefore be given greater weight in any discussion of enzyme catalysed reaction mechanisms. The reaction mechanisms discussed below are tabulated in Table 6-1 and shown schematically in Figure 6.1.

Table 6–1 A Survey of Catalytic Residues in Dehydrogenases

Enzyme (Species*)	Physiological Reaction†	Catalysis‡	Dehydrogenase Family	Panel in Figure 6.1
alcohol dehydrogenase (*Drosophila melanogaster*)	oxidation	Tyr, Lys	SDR	A
alcohol dehydrogenase (*Homo sapiens*)	oxidation	Zn^{2+}, H_2O	MDR	B
3α-hydroxysteroid dehydrogenase (*Rat*)	reduction	Tyr, Asp, Lys, His	AKR	C
malate dehydrogenase (*E. coli*)	oxidation / reduction	His, Asp, Arg	L specific 2-oxyacid dehydrogenases	D
lactate dehydrogenase (*Squalus acanthias*)	oxidation / reduction	His, Asp	L specific 2-oxyacid dehydrogenases	See D
D-2-hydroxyisocaproate dehydrogenase (*Lactobacillus casei*)	unknown	His, Asp, Arg	D specific 2-oxyacid dehydrogenases	E
aspartate semi-aldehyde dehydrogenase (*E. coli*)	oxidation	His, Cys	Aldehyde dehydrogenases	F
glyceraldehyde 3-phosphate dehydrogenase (*Homo sapiens*)	oxidation	His, Cys	Aldehyde dehydrogenases	See F
aldehyde dehydrogenase (*Vibrio harveyi*)	oxidation	Cys, Glu, Tyr?	Aldehyde dehydrogenases	-
glycerol dehydrogenase	oxidation	K^+, Asp/Glu?	None at present	G
isocitrate dehydrogenase	oxidation	Asp, Mg^{2+}	None at present	H
glutamate dehydrogenase (*Clostridium symbosium*)	reduction	Lys, Asp/Glu?	None at present	I
6-phosphogluconate dehydrogenase (*Sheep*)	oxidation	Lys, Glu	None at present	J
homoserine dehydrogenase (*Saccharomyces cerevisiaee*)	reduction	Lys, Asp	None at present	See Figures 6.5-7

* The species for the structurally determined enzyme
† The oxidation/reduction is defined with respect to the non-nicotinamide substrate.
‡ Question mark indicates residues identified without structural validation.

Figure 6.1 A-E - Hydride Transfer Mechanisms

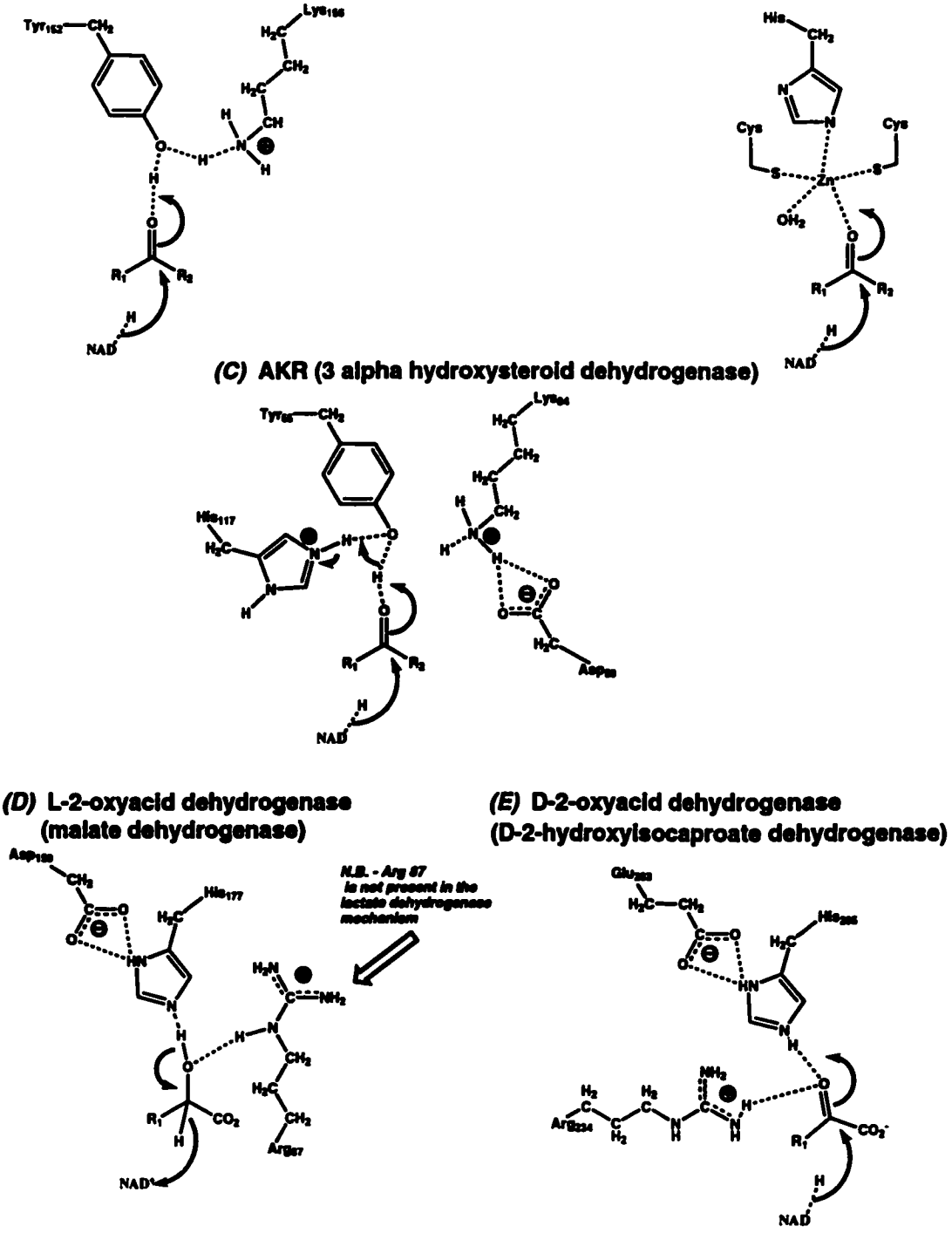
(A) **SDR (Drosophila alcohol dehydrogenase)**

(B) **MDR (generalized)**

(C) **AKR (3 alpha hydroxysteroid dehydrogenase)**

(D) **L-2-oxyacid dehydrogenase (malate dehydrogenase)**

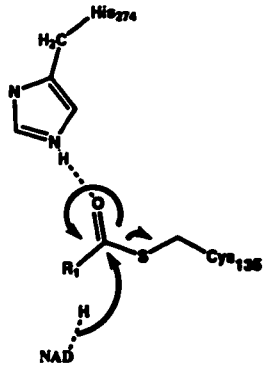
(E) **D-2-oxyacid dehydrogenase (D-2-hydroxyisocaproate dehydrogenase)**



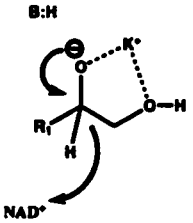
Reaction mechanisms discussed in the text. The residue numbering arises from the crystal structures of the enzymes (indicated in parenthesis).

Figure 6.1 F-J - Hydride Transfer Mechanisms

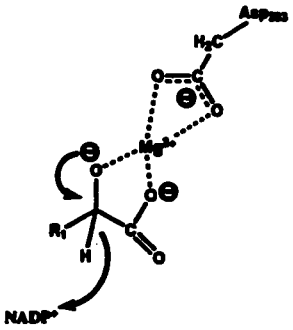
(F) **Aldehyde Dehydrogenase (Aspartate Semi-aldehyde dehydrogenase)**



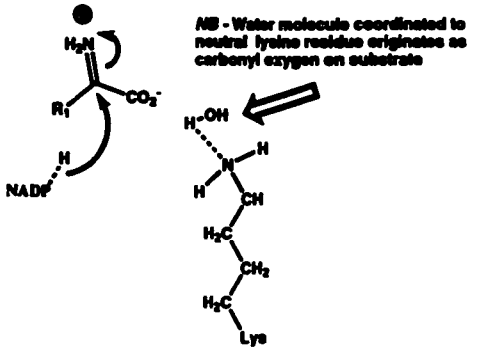
(G) **Glycerol dehydrogenase**



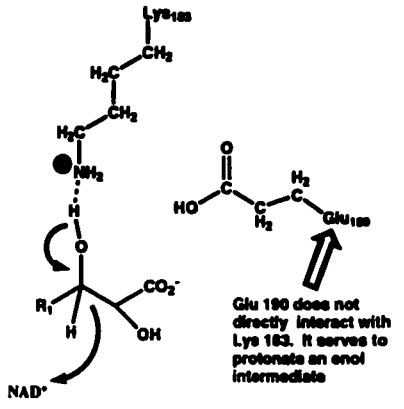
(H) **Isocitrate Dehydrogenase**



(I) **Glutamate dehydrogenase**



(J) **6-phosphosgluconate dehydrogenase**



6.1.1 Short Chain Dehydrogenases

The short chain dehydrogenase/reductase (SDR) family of alcohol dehydrogenases utilizes a strictly conserved active site tyrosine residue. A tyrosine residue, Tyr 152[§], abstracts the proton from the hydroxyl group while the hydride atom is transferred from the substrate alcohol to the NAD^+ cofactor. In the reductive direction, the same tyrosine would polarize the carbonyl oxygen through protonation. The tyrosine is assisted by a lysine residue, Lys 156. The reaction mechanism (Figure 6.1A) was proposed by McKinley-Mckee *et al.* (2) based upon kinetic studies of the alcohol dehydrogenase enzyme from *Drosophila melanogaster* and was supported by the crystal structure of 3α/20β-hydroxysteroid from *Streptomyces hydrogenans* (3) and murine dihydropteridine reductase (4). The lysine residue is not present in some SDR family members (5).

6.1.2 Medium Chain Dehydrogenases

The medium chain dehydrogenase/reductase (MDR) group of dehydrogenases employ a water coordinated zinc atom in the active site (Figure 6.1b). Information on this family of enzymes comes from structural studies of the various zinc dependent alcohol dehydrogenases as well as numerous enzyme kinetic studies (6).

The four coordinate zinc atom is typically ligated by two cysteine residues, a histidine residue, and a water molecule. The zinc atom itself plays only part of the role carried out by the tyrosine residue in the SDR alcohol dehydrogenases. The zinc atom stabilizes the oxyanion in the oxidative direction or polarizes the carbonyl oxygen in the reductive direction (7). The rest of the catalysis is assisted by the bound water molecule. There is some controversy over whether the water molecule remains directly attached to the catalytic zinc atom during the reaction (8), but the consensus view is that the water molecule is the immediate source/acceptor of protons during the reductive/oxidative reactions.

6.1.3 Aldo-keto Reductases

The aldo-keto reductases (AKR) are a large group of NAD dependent dehydrogenases which bind their co-factor without using the otherwise ubiquitous Rossmann fold motif. The hydride transfer is, however, similar in its catalytic requirements to the other enzymes discussed here (Figure 6.1c).

A tyrosine residue, Tyr 55[§], along with an aspartate residue, Asp 50, and a lysine residue, Lys 84, comprise the key catalytic residues conserved among the 42 AKR family members. The aspartate and lysine do not interact with the substrate but rather serve the role of modulating

---

[§] The residue numbers refer to the *Streptomyces* 3α/20β hydroxysteroid dehydrogenase

the pKa of the catalytic tyrosine. A histidine residue, His 117, has also been implicated in the catalytic mechanism (9). Like the lysine/aspartate pair, it modulates the pKa of the tyrosine residue. The overall mechanism has been described as a 'push-pull' mechanism, with the histidine and lysine/aspartate pair alternately pushing or pulling electron density from the tyrosyl oxygen atom, dependent on the direction of the catalyzed reaction.

6.1.4 2-Oxyacid Dehydrogenases

There are two subfamilies within the 2-oxyacid dehydrogenase group: the L substrate specific and the D substrate specific enzymes.

6.1.5 L-2-Oxyacids

6.1.5.1 Lactate Dehydrogenase

Lactate dehydrogenase catalyses the conversion of L-lactate to L-pyruvate. The hydride transfer requires deprotonation of an alcohol group and hydride abstraction by NAD^+. The deprotonation is assisted by a histidine residue in the active site (10) (see Figure 6.1d). This was established from steady state enzyme kinetic studies of the porcine enzyme from and a crystal structure of the homologous lactate

---

** The numbering scheme is for the rat liver 3α-hydroxysteroid dehydrogenase

dehydrogenase/NAD^+/pyruvate ternary complex (11) from *Squalus acanthias*.

6.1.5.2 Malate Dehydrogenase

A similar mechanism has been proposed for *E. coli* malate dehydrogenase which catalyzes the conversion of L-malate to oxaloacetate (Figure 6.1d). The mechanism has been modeled computationally (12) based on previous experimental studies (13-15). The deprotonation, similar to the step observed for lactate dehydrogenase is also assisted by an active site histidine / aspartate pair (His 177/Asp 150 for the *E. coli* enzyme). There is an additional arginine residue, Arg 87, which helps to stabilize the negative charge on the oxygen atom as it is converted from a hydroxyl to a carbonyl group. Arg 87 accomplishes this by hydrogen bonding to the oxygen atom through the proton of the N_ε sidechain atom. A reaction for the reverse (substrate reducing) direction based upon kinetic isotope effects was proposed and is essentially a mirror of the forward (substrate oxidizing) reaction mechanism (16).

6.1.6 D-2-hydroxy Acids

6.1.6.1 D-2-Hydroxyisocaproate Dehydrogenase

An example from the D-specific subfamily is given here for the catalysis carried out by D-2-hydroxyisocaproate dehydrogenase enzyme

from *Lactobacillus casei.* The physiological role of this enzyme is not yet fully understood. A 1.9 Å crystal structure of a ternary complex has been determined for the enzyme (17). A mechanism (Figure 6.1e) was proposed for the reduction of the carbonyl at the 2 position to a hydroxyl group with D-stereochemistry. The hydride transfer is facilitated by three active site residues. The carbonyl group is polarized by a protonated His 295/Glu 263 pair and through the $N_{\eta 1}$ nitrogen proton of Arg 234. Similar to the mechanisms given for L-malate and L-lactate, the acidic residue serves to modify to the pKa of the histidine residue and does not bind directly to the substrate.

The His/Asp or His/Glu is a well represented catalytic diad in dehydrogenase mechanisms. In addition to the examples given above, it is also present in the well studied reaction model of glucose-6-phosphate dehydrogenase from *Leuconostoc mesenteroides* (18, 19).

6.1.7 Aldehyde Dehydrogenases

The aldehyde dehydrogenases that catalyse the oxidation of aldehyde substrates to either carboxylic acids or phosphate esters share a common feature. All of the reaction mechanisms involve the formation of a thio ester with an active site cysteine residue. This requirement was proposed in the review by Oppenheimer *et al.* (6) as necessary for driving the reaction in the physiological direction for the case of dehydrogenases which catalyse the oxidation of aldehyde substrates to carboxylic acids. If the nucleophilic

cysteine residue is absent, the reaction equilibrium shifts against the physiological direction. This generalization does not appear to be true of enzymes which combine hydride transfer with substrate phosphorylation/dephosphorylation, as shown by the following two examples of aspartate semi-aldehyde dehydrogenase and glyceraldehyde 3-phosphate dehydrogenase.

6.1.7.1 Aspartate Semi-aldehyde Dehydrogenase

There are several examples of reaction mechanisms for aldehyde dehydrogenases. The mechanism for L-ASA dehydrogenase (ASAD) from *E. coli* has been worked out by steady state kinetics methods (20) and is essentially supported by the recently determined crystal structure of the enzyme from the same species (21) (Figure 6.1f). ASAD catalyzes a reductive dephosphorylation of aspartate O-phosphate to L-ASA and is the step immediately preceding the reaction catalyzed by HSD in the aspartate pathway. Based upon steady state kinetic studies, Viola *et. al* proposed the involvement of a lysine residue, but the crystal structure showed that a histidine residue, His 274, was a more likely choice. The histidine residue polarizes the carbonyl group twice during the reaction cycle: initially during the sulhydryl attack by Cys 135 to generate the thio ester and again during the hydride transfer reaction to generate the product L-ASA.

6.1.7.2 Glyceraldehyde 3-Phosphate Dehydrogenase

The reaction mechanism for ASAD is similar to the mechanism proposed for glyceraldehyde 3-phosphate dehydrogenase (GAPDH). GAPDH catalyses the oxidative phosphorylation of glycerol 3-phosphate to 1,3-diphosphoglycerate as part of glycolysis. The reaction mechanism is based upon steady state kinetics studies (22, 23) of the homologous pig and human muscle enzymes and crystal structure (24) of the human muscle enzyme. As one would expect for enzymes employing similar reaction mechanisms, GAPDH and ASAD have nearly identical active sites. The histidine residue implicated in deprotonating the active site cysteine nucleophile are one example of the similarity. The N_ε atom of the active site histidine (His 274) of ASAD is less than 1 Å away from N_ε atom of the histidine residue which plays the identical role in GAPDH (His 176), even though these residues are located on strands in the individual proteins which do not superimpose.

This is an important point to consider because ASAD is also being considered as a target for the design of antimicrobial compounds. The similarity of the active sites between GAPDH and ASAD has important repercussions for any molecules designed to inhibit ASAD because of the ubiquitous nature of GAPDH and its role in glycolysis metabolism. It is likely that any compounds that inhibit ASAD will also show activity against GAPDH and could therefore cause serious side effects.

6.1.7.3 Aldehyde Dehydrogenase

Another example of an aldehyde dehydrogenase group member can be observed for the reaction carried out by aldehyde dehydrogenase from *Vibrio harveyi*. This enzyme can oxidize a wide variety of aldehyde substrates to their corresponding carboxylic acids. A study (25) of the steady state kinetics revealed that two different glutamate residues are important for catalysis. Glu 253 plays a key catalytic role in both the sulfhydryl attack and the hydride transfer whereas Glu 377 plays a key catalytic role only in the hydride transfer. Except for the cysteine residue required for forming a thio ester, it was unclear whether other residues were important for catalysis. Similar studies on the homologous human mitochondrial aldehyde dehydrogenase also revealed the importance of a glutamate residue for the catalysis. One interpretation of the observations made by Vedadi *et al.* is that Glu 253 increases the nucleophilicity of the attacking sulfhydryl and Glu 377 assists Glu 253 in polarizing the carbonyl oxygen, either directly or through an intermediate histidine or lysine.

An examination of the unpublished structure of the homologous cod liver betaine dehydrogenase (26) did not clarify the role of the glutamate residues. While there are two glutamate residues in the vicinity of the catalytic cysteine, the nearest potential catalytic residue is a tyrosine, Tyr 167, 7.6 Å from the cysteine residue. The tyrosine residue is moderately conserved among the aldehyde reductase family. The reaction

mechanism model for this class of enzymes is either variable or not yet fully established.

6.1.8 Other Dehydrogenases

There are several dehydrogenases which do not fall under the classifications outlined in the introduction. These examples reinforce the point that the arrangement of residues which catalyse hydride transfer is widely varied.

6.1.8.1 Glycerol Dehydrogenase

Glycerol dehydrogenase is another dehydrogenase which uses a metal atom for catalysis (*cf.* MDR family enzymes). Glycerol dehydrogenase oxidizes glycerol to 1,3 dihydroxyketone and is dependent upon the presence of a monovalent cation. Leichus *et al.* (27) have proposed a mechanism (Figure 6.1g) for the enzyme from *Cellulomonas* species NT3060 in which a potassium ion stabilizes the oxyanion created by deprotonation of the C2 hydroxyl group in glycerol. An examination of the steady state enzymology pH profile suggested that the residue was either an aspartate or a glutamate. Hydride transfer from glycerol to NAD^+ occurs after formation of the potassium stabilized oxyanion form. There is currently no structural data to either identify the key residues or to support the proposed mechanism.

6.1.8.2 Isocitrate Dehydrogenase

Another example of a dehydrogenase reaction where a metal atom plays an important role is observed in the mechanism proposed for isocitrate dehydrogenase (IDH). IDH catalyses the oxidative decarboxylation of isocitrate to α-ketoglutarate as part of the citric acid cycle. The 2.5 Å crystal structure of the ternary complex of magnesium bound isocitrate, $NADP^+$ and IDH from *E. coli* (28) afforded an excellent opportunity to examine the catalytic mechanism (Figure 6.1h). Both Asp 283 and the Mg^{2+} atom function to stabilize the oxyanion of the enolate which forms during the hydride transfer step. Asp 283 also binds to the catalytic magnesium atom. The reaction is completed by protonation at the C_3 carbon of the magnesium stabilized enolate form of α-ketoglutarate.

6.1.8.3 Glutamate Dehydrogenase

Another class of hydride transfers which have been studied are those which also involve the transfer of an amino functionality to or from the substrates. An example of reductive amination can be found by examining the mechanism proposed for the catalysis accomplished by glutamate dehydrogenase from bovine liver (29), *i.e.* the conversion of α-ketoglutarate to glutamic acid. The reaction mechanism (Figure 6.1i), initially proposed based upon steady state enzymatic studies, entailed

hydride transfer occurring after attack by an ammonia molecule on the bound substrate. No key residues were identified as necessary to catalyze the hydride transfer from NADPH to the imino functionalized carbon atom, but a lysine and a carboxylate group were proposed to assist the steps leading to the formation of the imino group. A crystal structure of the homologous enzyme from *Clostridium symbosium* (30) revealed the presence of two active site lysine residues, but no carboxylate groups could be identified in the vicinity of the active site.

6.1.8.4 L-Alanine Dehydrogenase

The reverse reaction, oxidative deamination, is accomplished by L-alanine dehydrogenase which catalyses the conversion of L-alanine to pyruvate. The enzyme from *Bacillus subtilis* was studied by steady state kinetics by Grimshaw *et al.* (31). They proposed a reaction mechanism similar to the reverse of the L-glutamate dehydrogenase mechanism. The hydride transfer step occurs first to generate an imino group. The imino group is displaced in a stepwise fashion by a water molecule to generate the carbonyl product. The studies did not identify which residues were responsible for protonation/deprotonation steps, and there are no crystal structures of L-alanine dehydrogenase to confirm the proposed mechanism.

6.1.8.5 6-Phosphogluconate Dehydrogenase

6-phosphosgluconate dehydrogenase catalyses the oxidative decarboxylation of 6-phosphogluconate to ribulose 5-phosphate as part of the pentose phosphate pathway. A mechanism for the reaction has been proposed (Figure 6.1j) based upon the crystal structure (32) and steady state kinetics (33) of the enzyme from sheep liver.

The mechanism entails a deprotonation of the C_4 hydroxyl group by an unionized active site lysine, Lys 183. The deprotonation is concerted with hydride transfer from the C_4 carbon atom to the NAD^+ cofactor. The resulting carbonyl oxygen is protonated by the same lysine residue so that the enol form is stabilized, thereby facilitating decarboxylation. The enol form is subsequently protonated by an active site glutamine, Glu 190, to generate the product molecule.

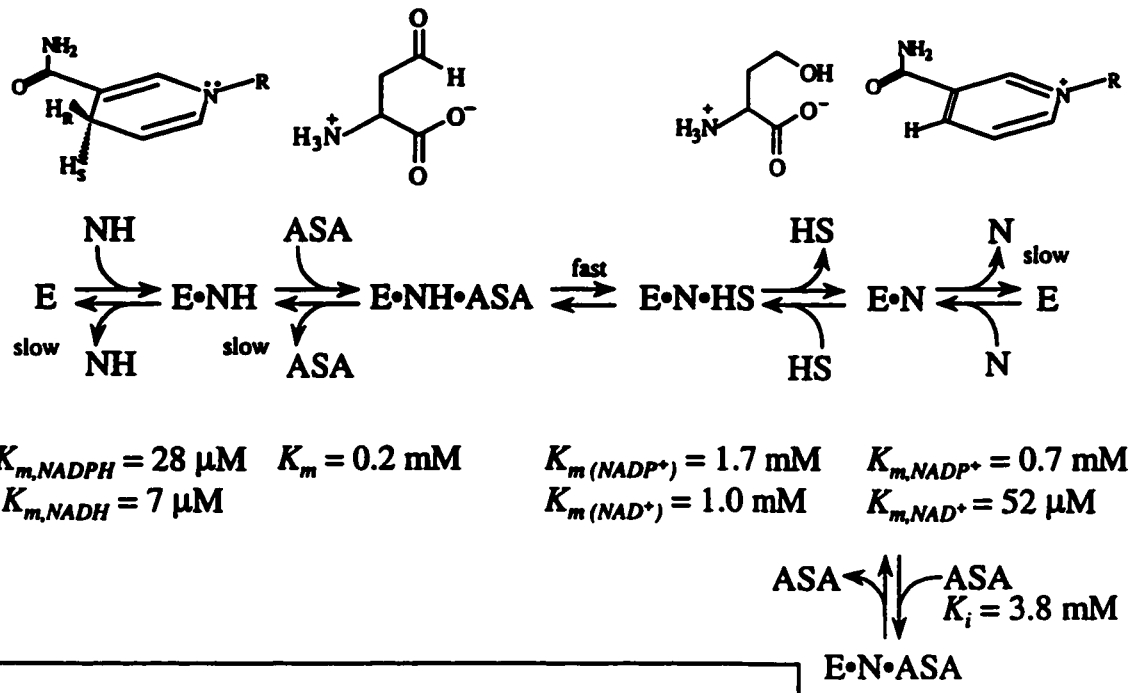
The presence of both a lysine and an acidic residue in the reaction mechanism is similar to that found in HSD. However, the geometry (they are 5.6 Å apart from each other) and roles are different.

6.2 Proposed Homoserine Dehydrogenase Reaction Mechanisms

The order of reaction participant binding or release was determined from steady state enzyme kinetic studies. The experiments revealed that HSD follows an ordered bi-bi kinetic reaction mechanism (34) with NAD(P)H binding first, followed by L-ASA. After the co-factor and

substrate are bound to the enzyme, the hydride transfer is catalyzed. There is some crystallographic evidence to suggest that a slight conformational change involving α-helix J occurs at this point. After hydride transfer, the protonated product is released from the active site. The final and rate determining step is the release of the oxidized co-factor, NAD^+. The reaction order is shown schematically in Figure 6.2.

Figure 6.2 The Reaction Order for HSD.



NH is reduced cofactor, N is oxidized cofactor, HS is L-Homoserine. *NB*-Michaelis constants are given and the stereochemistry of the nicotinamide ring is indicated. Adapted From (34)

Given the three different structures discussed in this thesis, a complete reaction mechanism for both the reductive and oxidative processes can be proposed if the following assumptions are made:

1. The NADH bound enzyme has a conformation very similar to the NAD^+ enzyme. This assumption is based upon the minimal differences observed for the three different states of the enzyme (apo, holo, and ternary complex) which have been structurally characterized.
2. The NAD molecule binds in a manner similar to the NAD analogue. Others (35) have observed that the NAD analogue behaves in a manner similar to that of NAD. Furthermore, millimolar concentrations of the NAD analogue inhibits HSD activity (36).
3. The orientation and position of the L-ASA molecule are similar to the L-Hse orientation and position. Large substrate conformational shifts are not important for the reaction mechanism. This assumption is justified by the observation that hydride transfer is not the rate determining step of the kinetic mechanism.

6.2.1 Active Site Geometry

The active site geometry was discussed in Chapters 4 and 5. One of the key observations from the crystal structure data was the presence of an N-terminal helix dipole in the active site. The N-terminus of α-helix J is very close (under 5 Å) to the oxygen atom of the hydroxyl in the bound

L-Hse molecule of the ternary complex. A brief overview of the literature covering the role of helix N-termini in enzyme catalysis is given below.

The role of helix N-termini in stabilizing negative charges was first studied in a comprehensive manner by Hol *et al.* (37) in a paper which focused on the preponderance of phosphate groups at helix N-termini. The study was expanded later by Hol (38) to include interactions with other anionic species, *e.g.* ionized serine or cysteine residues, flavin cofactors, and sulfate molecules.

Lockhart and Kim produced a series of papers (39, 40) which directly proved that α helix N-termini could stabilize a negative charge. This was accomplished by studying short synthetic peptides with a probe attached to the N-terminus. The peptides were designed to form isolated α-helices, as identified by circular dichroism, and the probe was located such that it was affected by the resultant N-terminal dipole. Both the pKa and the spectral characteristics of the probe were affected by the helix dipole.

The artificial helices were studied on a computational level by finite difference Poisson-Boltmann calculations (41). The N-terminal dipole of α-helices have also been dealt with on a theoretical or computational level by others (42), including Hol in the studies mentioned above.

Direct evidence for the role of a helix dipole in the reaction mechanism of a naturally occurring protein was obtained from ^{15}N-NMR

studies of the pKa of an active site histidine residue in triosephosphate isomerases (43). It was shown that an active site histidine residue located at the N terminus of a short α-helix had its pKa depressed by 2 units. The pKa depression resulted in a neutral histidine residue at physiological pH. The presence of a neutral histidine residue was shown to be of key importance for the catalytic reaction mechanism of triosephosphate isomerase.

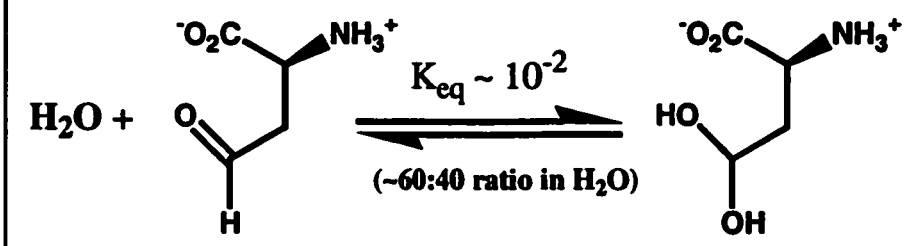
6.2.2 Forward reaction mechanism for HSD

Physiologically, the forward reaction mechanism is the transfer of a hydride atom from NAD(P)H to L-ASA to generate $NAD(P)^+$ and the product alcohol, L-Hse.

6.2.2.1 Aldehyde/diol equilibrium

The proposition of a forward reaction mechanism is complicated by the nature of the substrate, L-ASA. Aldehydes can undergo nucleophilic attack by water to generate *gem*-diols, as shown in Figure 6.3.

Figure 6.3 – The *gem*-diol / aldehyde equilibrium



The equilibrium constant for the aldehyde to diol conversion has not been established for L-ASA, but it can be approximated with values obtained from similar aldehydes. The K_H (defined as the ratio of *gem*-diol to aldehyde) for alkyl aldehydes of 4 and 5 carbon lengths is 0.4 and 0.7, respectively (44, 45). One can assume that the K_H value for L-ASA is similar. This means that the enzyme substrate will exist approximately 60-70% in aldehyde form and 30-40% in *gem*-diol form. Thus, the two forms of the substrate likely exist in quantities ample enough that neither can be ignored. It would therefore be prudent to consider reaction mechanisms entailing binding of each of the substrates until the true composition of the bound substrate can be determined.

The mechanistic requirement that an aldehyde substrate bind in its *gem*-diol has not been observed in naturally occuring enzymes. It was, however, proposed as a modified reaction for the GAPDH mutant enzyme in which the active site cystine was replaced with an alanine (46). The *gem*-diol form of aldehydes has been studied extensively with respect to the chemistry of *Drosophila melanogaster* alcohol dehydrogenase (47). For the alcohol dehydrogenase, the *gem*-diol was suggested as a substrate for the dismutation reaction, in which a carboxylic acid is generated from an aldehyde molecule bound in its *gem*-diol form with the oxidized NAD cofactor.

Figure 6.4 – The Balanced Reaction Equation Catalysed by HSD

$$NADH + L\text{-}ASA + H^{+} \leftrightarrow NAD^{+} + L\text{-}Hse$$

The balanced reaction catalyzed by HSD, shown in Figure 6.4, involves the addition of both a proton and a hydride to the aldehyde substrate; any catalytic mechanism proposed must account for all of these components. The mechanism is based on the three different crystal structures discussed in this thesis as well as a series of HSD mutants generated subsequent to the determination of the crystal structure of the NAD^{+} bound enzyme (see Table 6.2 and Figure 5.5).

Table 6.2 – Kinetic Data from Mutational Studies[††]				
	Forward direction			
	L-ASA[1]		**NADH[2]**	
	K_m (mM)	**k_{cat} (s^{-1})**	**K_m (mM)**	**k_{cat} (s^{-1})**
Wild-type	0.23 ±0.03[6]	560 ±40	0.007 ±0.001	370 ±10
K117A[5]	-	<0.005 ±0.001	-	<0.005 ±0.001
E208L[5]	-	<0.008 ±0.002	-	-
E208Q[5]	-	<0.4 ±0.1	-	-
E208D	9.6 ±0.9	330 ±20	0.027 ±0.003	123 ±3
D219L	0.36 ±0.09[6]	3.1 ±0.5	0.026 ±0.008[6]	2.5 ±0.3
K223V[5]	-	<0.005 ±0.001	-	-
	Reverse direction			
	L-Hse[3]		**NAD^+[4]**	
	K_m (mM)	**k_{cat} (s^{-1})**	**K_m (mM)**	**k_{cat} (s^{-1})**
Wild-type	1.0 ±0.1	21.9 ±0.6	0.052 ±0.005	17.3 ±0.5
K117A[5]	-	<0.005 ±0.001	-	<0.005 ±0.001
E208L[5]	-	<0.001 ±0.000	-	-
E208Q[5]	40 ±20	0.11 ±0.03	4±3	0.2 ±0.1
E208D	15 ±2	5.8 ±0.4	0.9 ±0.1	4.2 ±0.3
D219L	0.36 ±0.09[6]	3.1 ±0.5	0.026 ±0.008[6]	2.5 ±0.3
K223V[5]	-	<0.005 ±0.001	-	-

[1]NADH was maintained at concentration of 0.2 mM for Wild-type, and D219L, and a concentration of 0.3 mM for all other mutants. [2]L-ASA was maintained at concentration of 0.8 mM for Wild-type, and D219L, and a concentration of 10 mM for E208D. [3]NAD^+ was maintained at concentration of 0.3 mM for Wild-type, E208L, D219L and K223V, and a concentration of 2 mM for all other mutants [4]L-Hse was maintained at concentration of 20 mM for Wild-type, E208Q, E208D, and D219L. [5]Owing to the extremely low values of k_{cat}, K_m values for substrates and kinetic parameters for co-factors could not be estimated. [6]Substrate inhibition was observed for Wild-type L-ASA (K_i of 3.8 mM ±0.9), D219L L-ASA (K_i of 1.9 mM ±0.7), and D219L NADH (K_i of 1.5 mM ±1.0).

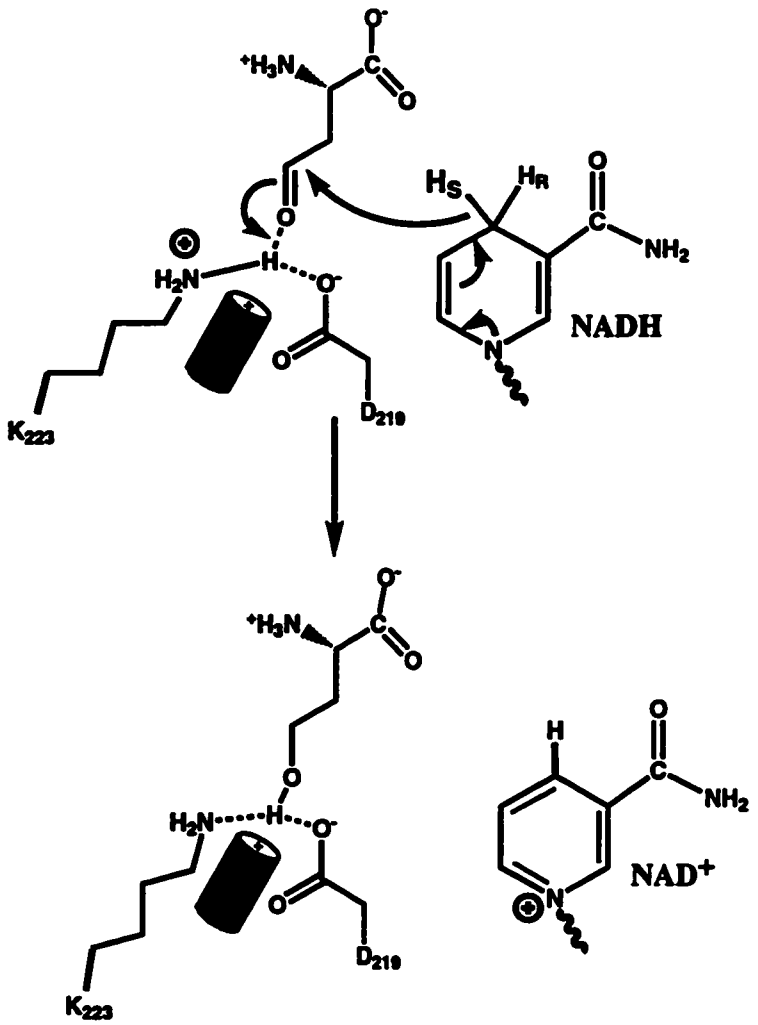
[††] The mutations and kinetic characterization were performed by Paul Thompson.

6.2.2.2 Reduction of the L-ASA Aldehyde

The substrate is bound within a hydrogen bonding network involving Thr 176, Asp 214, Glu 208, and two water molecules, Wat 460 and Wat 771 (Figure 5.4B). The catalytic residues are Asp 219 and Lys 223. The reaction mechanism requires both Asp 219 and Lys 223 to act as a pair, as HSD mutants of D219L and K223V were both hampered in their catalytic activity (Table 6.2). The proposed reaction mechanism, along with the catalytic residues, is in Figure 6.5.

During the hydride transfer, a negative charge will develop on the carbonyl oxygen. Catalysis of the hydride transfer is accomplished by stabilizing the developing negative charge. There are three possible candidates in the active site which may fulfill this role: Asp 219, Lys 223, and the N-terminal dipole of α-J. Asp 219 would have to be protonated in order to stabilize the negative charge. The pKa of an isolated aspartate side chain is 3.90 (48). Because of the adjacent N-terminal dipole of the α-J helix, the pKa of Asp 219 may be even lower. It is likely, therefore, that Asp 219 is in a deprotonated state in the active site. This will affect the next candidate, Lys 223, in its ability to stabilize the negative charge on the carbonyl oxygen atom. The interaction of a deprotonated Asp 219 carboxylate group with a protonated Lys 223 amino group would likely reduce the lysine residue's ability to polarize the substrate carbonyl group.

Figure 6.5 - Proposed Forward Reaction Mechanism for Substrate Bound as Aldehyde



The forward reaction mechanism for the L-ASA substrate bound as an aldehyde. Residues involved in binding the amino acid substrate are not shown (see Figure 5.4B). The role of the α-J N dipole is also shown. See text for a detailed description of the reaction.

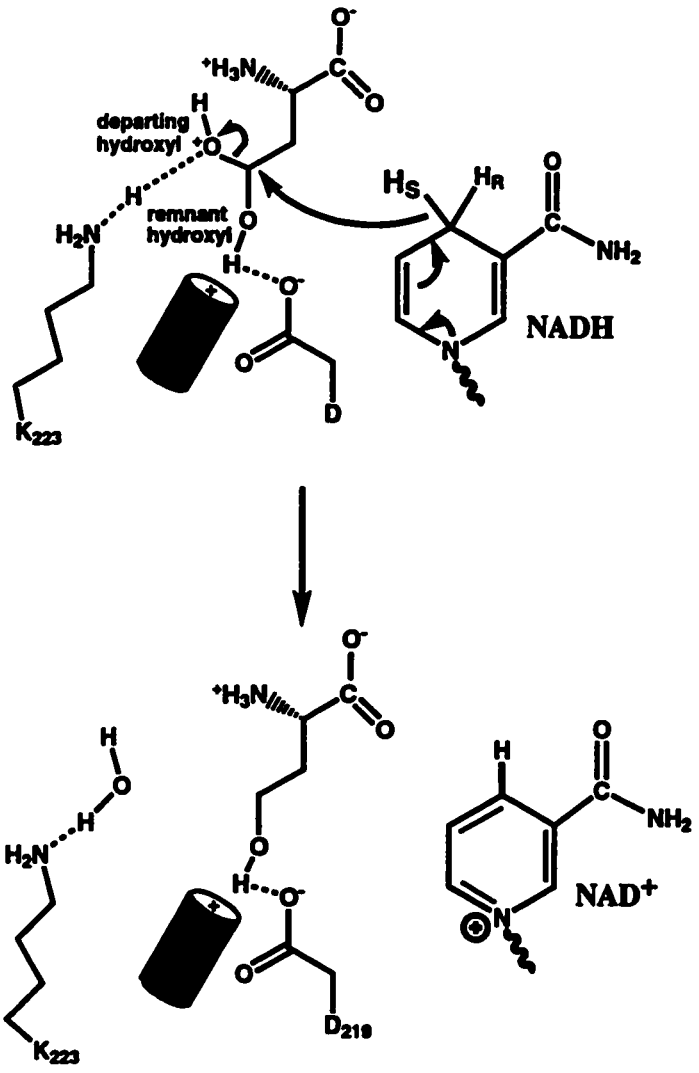
The final candidate, the α-J dipole, is the most probable active site component for stabilizing the negative charge which will develop on the carbonyl oxygen atom during the hydride transfer. The absence of bound L-Hse in crystallographic copies of HSD wherein the α-J helix is pulled away from the active site is a telling observation. Taken in combination with the proven role of α-helix N-terminal dipoles, the evidence for the role of this helix dipole as an oxyanion hole is compelling.

The charge polarization on the carbonyl group would generate a slight positive charge on the C_4 carbon of L-ASA. This would favour the hydride transfer from the bound co-factor to the substrate at the C_4 carbon. The resulting oxyanion could be protonated by Lys 223 to generate the product alcohol. The departure of this alcohol from the active site is followed by the departure of the oxidized co-factor. Protonation of the amino group of Lys 223 by the aqueous solvent would regenerate the enzyme for another round of catalysis.

6.2.2.3 Reduction of the L-ASA *gem*-Diol

The proposed reaction for the *gem*-diol form is shown in Figure 6.6. The amino and carboxylate groups of the substrate are bound in a manner identical to the aldehyde form described above. The catalytic residues are again Lys 223 and Asp 219; however, they have a slightly expanded role when the substrate is bound as a *gem*-diol.

Figure 6.6 - Proposed Forward Reaction Mechanism for Substrate Bound as *gem*-Diol



The forward reaction mechanism for the L-ASA substrate bound in *gem*-diol form. Residues involved in binding the amino acid substrate are not shown (see Figure 5.4B). The role of the α-J N dipole is shown. See text for a detailed description of the reaction. Note that with the exception of the water molecule on the lysine residue, the orientation of the final products for this reaction closely resemble the experimentally determined coordinates for both the amino acid sidechains and the product alcohol from the ternary complex. There is a water molecule located 6.4 Angstroms away from K223.

The two hydroxyl groups of the *gem*-diol form of L-ASA are hereafter referred to as the departing and remnant groups, in accord with their proposed behaviour during the reaction process. The departing hydroxyl group accepts a hydrogen atom from Lys 223, thereby generating a positive charge on the oxygen atom. The remnant hydroxyl group forms a hydrogen bond to the side chain carboxylate group of Asp 219.

The reaction may then proceed in a concerted manner with the hydride attack on the L-ASA C_4 carbon atom accompanied by a departure of the departing hydroxyl group as a water molecule coordinated to a neutral Lys 223.

The effect of the N-terminal dipole from α-J, which points directly at the remnant hydroxyl group would be to stabilize the slight negative charge on the oxygen atom in the remnant hydroxyl group, thereby polarizing the carbon-oxygen bond and maintaining the non-nicotinamide substrate in a state which more closely resembles the product (alcohol) form.

Departure of the protonated alcohol followed by the oxidized NAD^+ molecule and regeneration of the water coordinated Lys 223 residue (*via* protonation from the solvent) would reset the enzyme for another round of catalysis.

6.2.2.4 Which Mechanism is Correct?

It is difficult to select one mechanism over the other based upon the available data. However, the bound diol mechanism has several features which suggest that it is the predominant mechanism.

1. The crystal structure of the ternary complex reveals that the C_4 hydroxyl group oxygen atom is located 0.9 Å closer to Asp 219 than to Lys 223. The *gem*-diol mechanism would require the remnant hydroxyl group to be bound in this manner.
2. For the *gem*-diol based mechanism, the protonation/ionization states of both Asp 219 and Lys 223 are as they would be expected to be at physiological pH.
3. The generation of 3 product molecules in the diol mechanism (water, alcohol, oxidized co-factor) would be favoured by entropy over the generation of just 2 product molecules for the aldehyde mechanism. The increase in entropy from an extra molecule produced by the reaction would be offset by the endothermic requirement that the carbon oxygen bond be broken to release the water molecule. The exact amount of the offset is unknown. Although a thermodynamic argument cannot be used to describe the rate of a reaction, it can be used to describe the final ratio of products to reactants. In this situation, the physiologically relevant direction is entropically favourable if the diol mechanism is invoked. The entropy cost of

forming the *gem*-diol is prepaid by the *gem*-diol/aldehyde equilibrium before substrate binding occurs.

4. The role of the strictly conserved Asp 219 was shown by the kinetics of the D219L HSD mutant to be less important for catalysis. The D219L mutant, although hampered, was still able to catalyze the hydride transfer. This is described by the *gem*-diol based model, where Asp 219 plays mainly a substrate binding role. The D219L mutant K_M values are virtually unaffected in the forward direction. This could be interpreted in two different ways: the *gem*-diol based mechanism may be the wild-type mechanism since an increase in K_M should have been observed in the D219L mutant, or the D219L mutant has altered the enzyme such that the mutant kinetics are not directly comparable to the wild-type kinetics.
5. There is a water molecule in the ternary complex (Wat 771) only 3.7 Å away from the bound L-Hse molecule.

The predominant form of the bound substrate could be determined by studying the effect of shifting the *gem*-diol/aldehyde equilibrium upon K_M for the forward reaction mechanism. Such a shift in equilibrium could be accomplished by synthesizing modified substrates bearing an electron withdrawing substituent at the C_3 carbon (*e.g.* fluorine) or alternatively by replacing the hydrogen atom on the C_4 carbon with a deuterium atom.

6.2.3 Reverse reaction mechanism

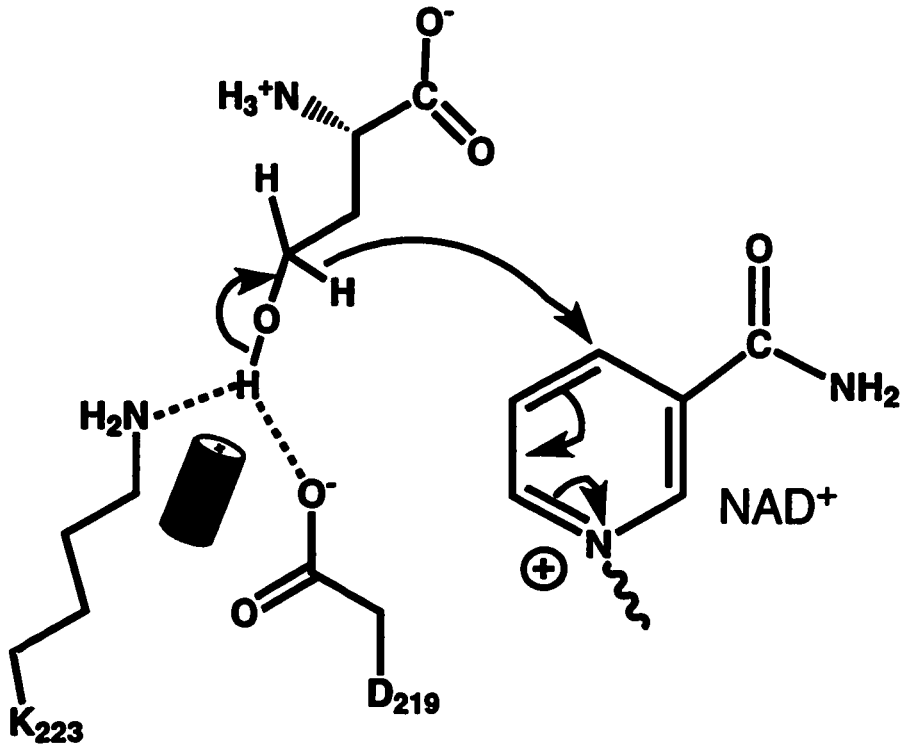
The reverse reaction mechanism requires the reduced cofactor to act as an acceptor for the hydride atom attached to the C_4 carbon atom of the L-Hse molecule. The hydride transfer will generate a carbocation which would quickly be converted to a carbonyl carbon. The process would be catalyzed by the presence of a general base to abstract the C_4 hydroxyl proton. From the crystal structure alone, the most logical choice for this base is Asp 219. However, kinetic studies carried out on an Asp 219 mutant indicate that the aspartate is important, but not crucial for the reverse direction catalysis. The D219L mutant shows an ~7 fold drop in k_{cat} for the reaction in which L-Hse is oxidized.

The lack of observed activity for the K223L mutant indicates that Lys 223 may also be important, but it is unclear why this should be so. Three possible roles can be envisioned for the Lys 223 residue: substrate orientation, coordination of a water molecule, or as a modifier of the Asp 219 pKa.

With regards to substrate orientation, Lys 223 under physiological conditions should have a positive charge that would repel the amino group on L-Hse, thereby ensuring that L-Hse binds in productive manner. A second possibile role for Lys 223 may be to coordinate a water molecule, so that the reverse reaction completely mirrors the forward *gem*-diol reaction mechanism. There is a water molecule in the active site (Wat

771), but it does not interact directly with Lys 223. This water molecule is 6.4 Å away from Lys 223, but is only 3.7 Å from the C_4 of the bound L-Hse. The involvement of this water molecule in the formation of a diol for the reverse reaction is tantalizing, and would satisfy the requirements of microscopic reversability. However, the assignment of the Wat 771 water molecule as a reaction participant in the reverse reaction mechanism is speculative at this point. The third possible role is that Lys 223 is a pKa modifier of the aspartate. This implies a reaction mechanism whereby the lysine / aspartate pair stabilize a polarized carbon-oxygen bond to assist hydride transfer to NAD^+. The resultant transition state of such a mechanism is identical to that proposed for the aldehyde bound forward direction mechanism, thereby satisfying the requirement of microscopic reversibility. The proposed reverse direction reaction mechanism is shown in Figure 6.7.

Figure 6.7 - Proposed Reverse Reaction Mechanism



Residues involved in binding the amino acid substrate are not shown (see Figure 5.4B). The reaction is drawn in a concerted manner. The role of the α-J N dipole is also shown. See text for a detailed description of the reaction.

6.3 Summary of Chapter 6

A review of dehydrogenase reaction mechanisms revealed that the a catalytic lysine residue during hydride transfer catalysis has been observed previously (6-phosphogluconate dehydrogenase). However, the combination of lysine and aspartate as a catalytic pair had not been observed previous to the HSD structure. Based upon the crystal structure, several point mutants were designed, created, and evaluated by steady state kinetics. Lys 223 was shown to be a key catalytic residue. Asp 219 was shown to affect the rate of catalysis but was not critical for the reaction mechanism. Glu 208 and Lys 117 were shown to be amino acid substrate and nicotinamide cofactor binding residues, respectively, based upon their location in the active site and the enzymatic activity after their individual mutations.

Two different forward reactions were proposed because of the aldehyde to *gem*-diol equilibrium of the amino acid substrate. There is no clear evidence for the nature of the bound form of the L-ASA molecule. Crystallographic evidence supports the reaction mechanism in which the *gem*-diol form is bound to the enzyme, but such a mechanism must be considered unusual. Kinetic evidence, specifically the virtually unaffected K_M value observed for the D219L mutant, supports the reaction mechanism in which the aldehyde form is bound to the enzyme.

Furthermore, the aldehyde bound reaction mechanism is a better mirror of the proposed reverse direction mechanism, for which only a single reaction mechanism is likely.

Chapter 6 References

1. **Fersht, A.**, *Enzyme Structure and Mechanism*, WH Freeman, New York (1985).
2. **McKinley-McKee, J. S., Winberg, J. O., and Pettersson, G.** (1991). Mechanism of action of *Drosophila melanogaster* alcohol dehydrogenase. Biochem Int. **25**: 879-885.
3. **Ghosh, D., Weeks, C. M., Grochulski, P., Duax, W. L., Erman, M., Rimsay, R. L., and Orr, J. C.** (1991). Three-dimensional structure of holo 3α/20β-hydroxysteroid dehydrogenase: a member of a short-chain dehydrogenase family. Proc Natl Acad Sci U S A. **88**: 10064-10068.
4. **Varughese, K. I., Xuong, N. H., Kiefer, P. M., Matthews, D. A., and Whiteley, J. M.** (1994). Structural and mechanistic characteristics of dihydropteridine reductase: a member of the Tyr-(Xaa)3-Lys-containing family of reductases and dehydrogenases. Proc Natl Acad Sci U S A. **91**: 5582-5586.
5. **Jornvall, H., Persson, B., Krook, M., Atrian, S., Gonzalez-Duarte, R., Jeffery, J., and Ghosh, D.** (1995). Short-chain dehydrogenases/reductases (SDR). Biochemistry. **34**: 6003-6013.
6. **Oppenheimer, N. J., and Hanlon, A. L.**, Mechanism of NAD Dependent Enzymes, in *The Enzymes*, Vol. 20, Sigman, D., Ed., Academic Press, N York (1992).
7. **Klinman, J. P.** (1981). Probes of mechanism and transition-state structure in the alcohol dehydrogenase reaction. CRC Crit Rev Biochem. **10**: 39-78.
8. **Ryde, U.** (1995). Molecular dynamics simulations of alcohol dehydrogenase with a four- or five-coordinate catalytic zinc ion. Proteins. **21**: 40-56.
9. **Schlegel, B. P., Jez, J. M., and Penning, T. M.** (1998). Mutagenesis of 3 alpha-hydroxysteroid dehydrogenase reveals a "push- pull" mechanism for proton transfer in aldo-keto reductases. Biochemistry. **37**: 3538-3548.
10. **Holbrook**, Lactate Dehydrogenase, in *The Enzymes*, Vol. 11, Boyer, P., Ed., Academic Press, New York (1975).
11. **White, J. L., Hackert, M. L., Buehner, M., Adams, M. J., Ford, G. C., Lentz, P. J., Jr., Smiley, I. E., Steindel, S. J., and Rossmann, M. G.** (1976). A comparison of the structures of apo dogfish M4 lactate dehydrogenase and its ternary complexes. J Mol Biol. **102**: 759-779.
12. **Cunningham, M. A., Ho, L. L., Nguyen, D. T., Gillilan, R. E., and Bash, P. A.** (1997). Simulation of the enzyme reaction mechanism of malate dehydrogenase. Biochemistry. **36**: 4800-4816.

13. **Hall, M. D., and Banaszak, L. J.** (1993). Crystal structure of a ternary complex of *Escherichia coli* malate dehydrogenase citrate and NAD at 1.9 Å resolution. J Mol Biol. **232**: 213-222.
14. **Parker, D. M., Lodola, A., and Holbrook, J. J.** (1978). Use of the sulphite adduct of nicotinamide-adenine dinucleotide to study ionizations and the kinetics of lactate dehydrogenase and malate dehydrogenase. Biochem J. **173**: 959-967.
15. **Lodola, A., Shore, J. D., Parker, D. M., and Holbrook, J.** (1978). Malate dehydrogenase of the cytosol. A kinetic investigation of the reaction mechanism and a comparison with lactate dehydrogenase. Biochem J. **175**: 987-998.
16. **Wright, S. K., Zhao, F. J., Rardin, J., Milbrandt, J., Helton, M., and Furumo, N. C.** (1995). Mechanistic studies on malate dehydrogenase from *Escherichia coli.* Arch Biochem Biophys. 321: 289-296.
17. **Dengler, U., Niefind, K., Kless, M., and Schomburg, D.** (1997). Crystal structure of a ternary complex of D-2-hydroxyisocaproate dehydrogenase from *Lactobacillus casei*, NAD^+ and 2-oxoisocaproate at 1.9 Å resolution. J Mol Biol. **267**: 640-660.
18. **Cosgrove, M. S., Naylor, C., Paludan, S., Adams, M. J., and Levy, H. R.** (1998). On the mechanism of the reaction catalyzed by glucose 6-phosphate dehydrogenase. Biochemistry. **37**: 2759-2767.
19. **Viola, R. E.** (1984). Kinetic studies of the reactions catalyzed by glucose-6-phosphate dehydrogenase from *Leuconostoc mesenteroides*: pH variation of kinetic parameters. Arch Biochem Biophys. **228**: 415-424.
20. **Karsten, W. E., and Viola, R. E.** (1991). Chemical and kinetic mechanisms of aspartate-beta-semialdehyde dehydrogenase from *Escherichia coli.* Biochim Biophys Acta. **1077**: 209-219.
21. **Hadfield, A., Kryger, G., Ouyang, J., Petsko, G. A., Ringe, D., and Viola, R.** (1999). Structure of aspartate-beta-semialdehyde dehydrogenase from *Escherichia coli*, a key enzyme in the aspartate family of amino acid biosynthesis. J Mol Biol. **289**: 991-1002.
22. **Harris, J. I.**, Glyceraldehyde 3-Phosphate Dehydrogenase, in *The Enzymes*, Vol. 13, Boyer, P., Ed., Academic Press, New York (1976).
23. **Harrigan, P. J., and Trentham, D. R.** (1973). Kinetic studies of the acylation of pig muscle D-glyceraldehyde 3- phosphate dehydrogenase by 1,3-diphosphoglycerate and of proton uptake and release in the overall enzyme mechanism. Biochem J. **135**: 695-703.

24. **Mercer, W. D., Winn, S. I., and Watson, H. C.** (1976). Twinning in crystals of human skeletal muscle D-glyceraldehyde-3- phosphate dehydrogenase. J Mol Biol. **104**: 277-283.
25. **Vedadi, M., and Meighen, E.** (1997). Critical glutamic acid residues affecting the mechanism and nucleotide specificity of *Vibrio harveyi* aldehyde dehydrogenase. Eur J Biochem. **246**: 698-704.
26. **Johansson, K., El Ahmad, M., Hjelmqvist, L.**, Ramaswamy, S., Jornvall, H., and Eklund, H., PDB file 1A4S in RCSB PDB (unpublished), (1998).
27. **Leichus, B. N., and Blanchard, J. S.** (1994). Isotopic analysis of the reaction catalyzed by glycerol dehydrogenase. Biochemistry. **33**: 14642-14649.
28. **Hurley, J. H., Dean, A. M., Koshland, D. E., Jr., and Stroud, R. M.** (1991). Catalytic mechanism of NADP(+)-dependent isocitrate dehydrogenase: implications from the structures of magnesium-isocitrate and $NADP^+$ complexes. Biochemistry. **30**: 8671-8678.
29. **Rife, J. E., and Cleland, W. W.** (1980). Determination of the chemical mechanism of glutamate dehydrogenase from pH studies. Biochemistry. **19**: 2328-2333.
30. **Baker, P. J., Britton, K. L., Engel, P. C., Farrants, G. W., Lilley, K. S., Rice, D. W., and Stillman, T. J.** (1992). Subunit assembly and active site location in the structure of glutamate dehydrogenase. Proteins. **12**: 75-86.
31. **Grimshaw, C. E., Cook, P. F., and Cleland, W. W.** (1981). Use of isotope effects and pH studies to determine the chemical mechanism of *Bacillus subtilis* L-alanine dehydrogenase. Biochemistry. **20**: 5655-5661.
32. **Adams, M. J., Ellis, G. H., Gover, S., Naylor, C. E., and Phillips, C.** (1994). Crystallographic study of coenzyme, coenzyme analogue and substrate binding in 6-phosphogluconate dehydrogenase: implications for NADP specificity and the enzyme mechanism. Structure. **2**: 651-668.
33. **Price, N. E., and Cook, P. F.** (1996). Kinetic and chemical mechanisms of the sheep liver 6-phosphogluconate dehydrogenase. Arch Biochem Biophys. **336**: 215-223.
34. **Jacques, S.**, Characterization of Homoserine Dehydrogenase from *Saccharomyces cerevisae*: A New Fungal Target. Doctoral Thesis, in *Dept of Biochemistry*, McMaster University, Hamilton, Ontario, Canada (1999).
35. **Fisher, T. L., Vercellotti, S. V., and Anderson, B. M.** (1973). Interactions of 3-aminopyridine adenine dinucleotide with dehydrogenases. J Biol Chem. **248**: 4293-4299.
36. **Jacques, S.**, Personal Communication, (1999).

37. **Hol, W. G., van Duijnen, P. T., and Berendsen, H. J.** (1978). The alpha-helix dipole and the properties of proteins. Nature. **273**: 443-446.
38. **Hol, W. G.** (1985). Effects of the alpha-helix dipole upon the functioning and structure of proteins and peptides. Adv Biophys. **19**: 133-165.
39. **Lockhart, D. J., and Kim, P. S.** (1992). Internal stark effect measurement of the electric field at the amino terminus of an alpha helix. Science. **257**: 947-951.
40. **Lockhart, D. J., and Kim, P. S.** (1993). Electrostatic screening of charge and dipole interactions with the helix backbone. Science. **260**: 198-202.
41. **Sitkoff, D., Lockhart, D. J., Sharp, K. A., and Honig, B.** (1994). Calculation of electrostatic effects at the amino terminus of an alpha helix. Biophys J. **67**: 2251-2260.
42. **Aqvist, J., Luecke, H., Quiocho, F. A., and Warshel, A.** (1991). Dipoles localized at helix termini of proteins stabilize charges. Proc Natl Acad Sci U S A. **88**: 2026-2030.
43. **Lodi, P. J., and Knowles, J. R.** (1993). Direct evidence for the exploitation of an alpha-helix in the catalytic mechanism of triosephosphate isomerase. Biochemistry. **32**: 4338-4343.
44. **Gruen, L. C., and McTigue, P. T.** (1963). Hydration Equilibria of Aliphatic Aldehydes in H_2O and D_2O. J. Chem. Soc. **5**: 5217-5229.
45. **Bell, R. P.**, The Reversible Hydration of Carbonyl Compounds, in *Advances in Physical Organic Chemistry*, Vol. 4, Gold, V., Ed., Academic Press, New York, pp. 1-29 (1966).
46. **Corbier, C., Della Seta, F., and Branlant, G.** (1992). A new chemical mechanism catalyzed by a mutated aldehyde dehydrogenase. Biochemistry. **31**: 12532-12535.
47. **Winberg, J. O., and McKinley-McKee, J. S.** (1998). *Drosophila melanogaster* alcohol dehydrogenase: mechanism of aldehyde oxidation and dismutation. Biochem J. **329**: 561-570.
48. **Dawson, R. M. C., Elliott, D. C., Elliott, W. H., and Jones, K. M.**, *Data for Biochemical Research*, Oxford University Press (1969).

Chapter Seven: The Metal Ion Binding Site and the Nature of its Occupant in the Crystal Structure

7.1 Background

The presence of a metal ion binding site in HSD was discovered while assigning density observed in the σ_a weighted F_o-F_c maps. Typically the density in these maps is ascribed to the presence of ordered water molecules in the crystal structure. However, the coordination sphere around a peak near the 3_{10} helix α-1H was unusual for a water atom. The peak was surrounded by 6 oxygen atoms, 5 of which were backbone carbonyl oxygen atoms while the 6th was an oxygen atom from the carboxylate group of a glutamate residue. The hydrogen bonding requirements of this site could not be satisfied by a water molecule. The the peak was identified as a metal ion binding site.

The metal ion in the binding site was initially modeled as a calcium ion, but was subsequently changed to a sodium ion. There was some evidence to support the first choice, but it was ultimately outweighed by evidence for the latter.

7.2 Materials and Methods

7.1.1 Crystallization

Crystallization of the HSD tetragonal form was accomplished by the hanging drop method. Conditions were identical to those described in chapter 2. The addition of $CaCl_2$ or $Ca(OAc)_2$ was accomplished by adding 1 µl of a 0.1 M solution to the 7.5 µl drop containing HSD, NAD^+, and crystallization solution.

Lanthanide atoms have a large L_{III} absorption edge that can give rise to anomalous diffraction. The anomalous signal can be used to obtain experimental phasing information for the crystallographic structure factors. Furthermore, lanthanide atoms such as Ho^{3+} have been substituted into calcium binding sites of other proteins (1). Before the structure of HSD was available, the presence of a calcium in the protein was hypothesized based upon the beneficial effect calcium had on crystallization. Attempts were made to grow lanthanide substituted crystals to generate experimental phase information. The lanthanide co-crystallization experiments were performed under conditions similar to those described in Chapter 2. Lanthanides were included at a concentration of 2 µM to 20 mM. The lanthanide species used were: $HoCl_3$, $Ho(NO_3)_3$, $Ce(OAc)_3$, and $Nd(OAc)_3$. No calcium was added to these crystallization experiments.

7.1.2 Fluorescence Based Detection of Lanthanide Atoms in Crystals

The presence of atoms capable of generating anomalous diffraction can be detected by examining the fluorescent radiation they emit when the energy of the irradiating X-rays is above the absorption edge of the element in question. The experiments were carried out at the Cornell High Energy Synchrotron Source (CHESS) F2 beamline using wavelength energies from 6 to 8 keV. The fluorescence detector was placed in a position orthogonal to the irradiating X-ray beam and was calibrated with samples of the appropriate lanthanide. The incorporation of lanthanides into the HSD crystals was tested by flash cooling a few small crystals in a cryo-loop and then scanning across the appropriate energy range with the irradiating X-ray beam.

7.1.3 Simulated Annealing Omit Maps / B factor Refinement Tests

Simulated annealing omit maps were calculated with CNS 0.9 (2). Observed structure factors were obtained from the high resolution tetragonal data set. Calculated structure factors from the tetragonal model were used with the exception of an 8 Å sphere centred on the metal ion position. The sphere omitted model was subjected to a 1000 K slow cool anneal in torsion space followed by 100 steps of Powell minimization. A σ_a weighted F_o-F_C difference map was generated to establish the presence of the map feature identified as a metal ion.

The temperature (B) factor refinement tests were done in CNS 0.9. The B factors of the metal ions were set to 30 $Å^2$ and then the entire tetragonal and monoclinic models were respectively subjected to 30 rounds of individual B factor refinement. The B factors for water atoms in the model were held fixed during the refinement. The entire resolution range was used for each structure. The density modified SIR phase data was used as an additional source of information in the refinement of the tetragonal structure. The 'MLHL' target was used to incorporate the experimental phase information. The monoclinic data was refined against unphased structure factors derived from the observed data.

7.1.4 Ruthenium Oxide Assay

The Ca^{2+} binding assay described by Ewart *et al.* (3) was used to determine if HSD could bind Ca^{2+} ions. 10 µg of protein was run on a 12% SDS-PAGE gel. The gel was washed twice for 20 minutes with transfer buffer to remove SDS so the proteins could refold during transfer. The transfer solution was buffered to pH 8.3 with 25 mM Tris, 192 mM glycine and 20% v/v methanol. The mini trans-Blot transfer cell (BioRad) was used to electroporetically transfer the proteins from the SDS-PAGE gel to Hybond-ECL (Amersham) nitrocellulose. The transfer was accomplished in 1 hour with power settings of 100 V and 250 mA. The system was cooled with ice to avoid thermal runaway during the transfer. Hen egg

white lysosyme (Sigma) and herring gull antifreeze (gift of Dr. D-S.C. Yang) were used as negative and positive controls, respectively.

The nitrocellulose was bathed for ~15 minutes at room temperature in a solution of 25 mg/L ruthenium oxide, 60 mM KCl, 5 mM $MgCl_2$ buffered with 10 mM Tris-HCl at pH 7.5. Proteins capable of binding Ca^{2+} will stain pinkish-red during this step.

7.1.5 HSD Activity Assay

HSD was assayed in the reverse direction by continuously monitoring the increase in absorbance at 340 nm at room temperature using a Cary 3E UV-VIS spectrophotometer. The assay mixture contained 25 mM D,L-homoserine, 0.2 mM NAD^+ and 100 mM Tris-HCl at pH 9.0 in a total volume of 995 µL. 5 µL of enzyme at 1 mg/ml was added to start the assay. The reaction was allowed to proceed for one minute. The slope between two fixed points, typically 25 and 50 seconds was taken as a measure of activity. Measurements of individual conditions were performed at least twice.

To examine the effects of calcium, $CaCl_2$ at either 100 or 500 µM was used. To test whether calmodulin was affecting the protein, 25 µg of calmodulin from bovine brain (Sigma) was added to the assay mixture.

7.3 Results

7.3.1 Beneficial effects upon crystallization

Small amounts of $CaCl_2$ improved the reproducibility of tetragonal crystal forms, as discussed in Chapter 2. This observation by itself does not constitute evidence that HSD binds Ca^{2+} ions, since the formation of $Ca(SO_4)$ crystals could simply be affecting the crystal growth kinetics. Another distinct possibility was that the chloride ions were affecting the protein. However, both $CaCl_2$ and $Ca(OAc)_2$ had the same effect, demonstrating that the calcium ions were responsible for the observed improvement. The effect of $Ca(SO_4)$ formation upon crystallization kinetics remains a distinct possibility in explaining the observed improvement.

Calcium was not added to the crystallization conditions used to produce the monoclinic crystal form. The counter ion in the buffer was Na^+, in sodium cacodylate and was present at 0.1 M concentration in the initial conditions of the hanging drop experiment.

7.3.2 Lanthanide Experiments

Small crystals of HSD could be grown in the presence of a variety of lanthanides: Ho^{3+}, Nd^{3+}, and Ce^{3+}. These crystals diffracted to approximately 3 Å when exposed to intense (*i.e.* synchrotron) radiation.

Unfortunately no fluorescence was observed for the crystals grown in the presence of lanthanides when the wavelength scans were performed, indicating little to no incorporation of lanthanide ions in the crystal structure. Fluorescence could be detected from the calibration samples of lanthanides.

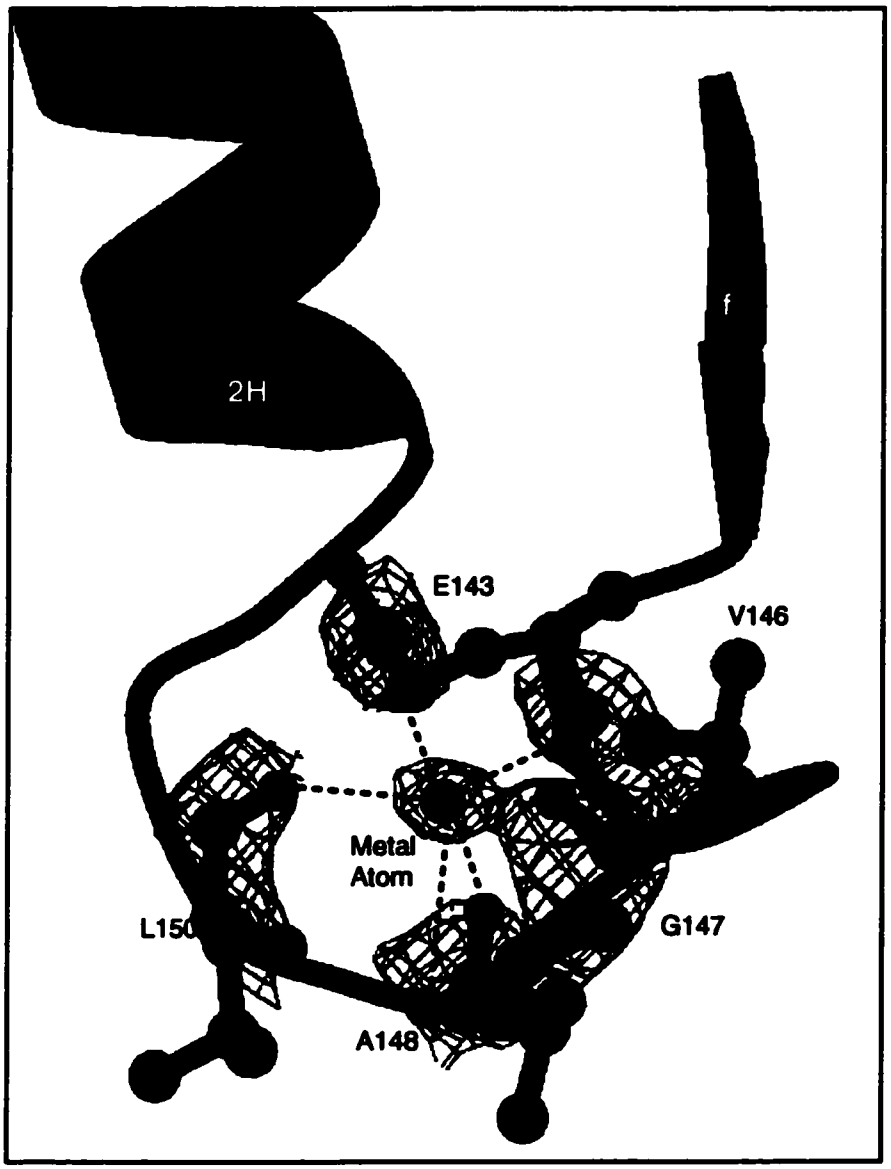
7.3.3 Difference Maps and B Factor Refinement

The simulated annealing σ_a weighted F_o-F_c difference omit map is superimposed on the protein model in Figure 7.1. The map feature interpreted as a metal ion is clearly visible at contour levels greater than 3σ. Difference maps generated in a similar manner from the monoclinic data and model showed similarly strong features in the corresponding regions.

7.3.3.1 Coordination geometry

The density peaks were located in nearly identical positions for each monomer: in the f-2H loop. The 6 coordinate geometry surrounding the metal ion exhibits a distorted trigonal biprismatic geometry, as shown in Figure 7.1. The interactions arise mainly from carbonyl backbone atoms of residues Glu 143, Val 146, Gly 147, Ala 148, and Leu 150. Glu 143 also interacts through its side chain

Figure 7.1 - Simulated Annealing Omit Map Superimposed on the Metal Atom Binding Site



The residues which constitute the metal (Na+) binding site are shown. The secondary structure elements around the site are provided for reference. The 3_{10} helix, α-1H containing most of the residues is shown as a coil. A σ_a weighted SA-omit Fo-Fc map generated as described in the text is shown, contoured at 2.0σ. The density describing the metal atom persists up to 3.2σ. Dashed red lines indicate the bonds described in Table 7.1 and demonstrate the distorted octahedral geometry of the metal atom. The metal atom is depicted as a green sphere.

carboxylate oxygen. The various oxygen-metal distances are given in Table 7.1 and do not vary significantly among the various monomers. Note that the interaction between the carbonyl oxygen atom of Gly 147 and the putative metal ion is far from ideal. Thus the metal ion may be only coordinated by five ligand atoms. Five coordinate sodium ions are unusual, but not unheard of. There are several examples of five coordinate sodium ions in the Cambridge Crystallographic Database (11).

Table 7.1 – Coordination Geometry

Residue	Metal-Oxygen Distance (Å)	Metal-Oxygen-Carbon Angle (°)
Glu 143 (carboxylate)	2.9	118.9
Glu 143	2.7	128.1
Val 146	2.9	127.2
Gly 147	3.5	78.6
Ala 148	2.8	119.2
Leu 150	2.8	140.4

7.3.3.2 B-Factor Refinement

Tables 7.2 and 7.3 show the results of the B factor refinement for the ion modeled as sodium and calcium respectively. The B factors for all atoms in residues 143 to 150 were taken as a measure of protein thermal motion / disorder near the metal binding site. R factors did not change significantly when the metal ion was changed from sodium to calcium. In all cases, the calcium ions exhibited B factors higher then their coordinating residues, whereas the sodium atoms show B factors similar

to their coordinating residues. The results therefore favour the presence of a sodium atom in the metal binding site in both crystal forms.

Table 7.2 Results for Site Modeled with a Sodium Ion			
Chain (143 – 150)	**Ave. B (Å²)**	**Range B (Å²)**	**Metal B (Å²)**
Monoclinic A	17 ± 6	2 – 41	24
Monoclinic B	13 ± 5	2 – 28	21
Monoclinic C	14 ± 4	4 – 25	22
Monoclinic D	13 ± 6	2 – 30	16
Tetragonal A	14 ± 2	9 – 18	14
Tetragonal B	15 ± 2	10 – 23	15

Table 7.3 – Results for Site Modeled with a Calcium Ion			
Chain (143 - 150)	**Ave. B (Å²)**	**Range B (Å²)**	**Metal B (Å²)**
Monoclinic A	17 ±6	2 – 40	55
Monoclinic B	13 ±5	2 – 26	51
Monoclinic C	14 ±4	3 – 27	51
Monoclinic D	13 ±6	2 – 30	49
Tetragonal A	15 ±2	10 – 21	50
Tetragonal B	16 ±2	12 – 25	61

7.3.4 Ruthenium Oxide Assay

The blot stained by ruthenium oxide is shown in Figure 7.2. The results indicate that HSD is capable of binding calcium ions. The intensity of the band in comparison with the herring gull antifreeze protein indicates that HSD either binds Ca^{2+} atoms with weak to moderate affinity, or that it

was not able to refold as efficiently as the positive control during the electroporetic transfer.

7.3.5 Enzyme Activity Assays

The effect of $CaCl_2$ and calmodulin on enzyme activity are given below in Table 7.4.

Table 7.4 – Effect of Calcium and Calmodulin on HSD* Activity

[$CaCl_2$] (μM)	[Calmodulin] (μg/ml)	Activity (nmol NAD^+ converted / minute†). Values are ± 1
0	0	42
100	0	34
500	0	33
100	25	33
500	25	30

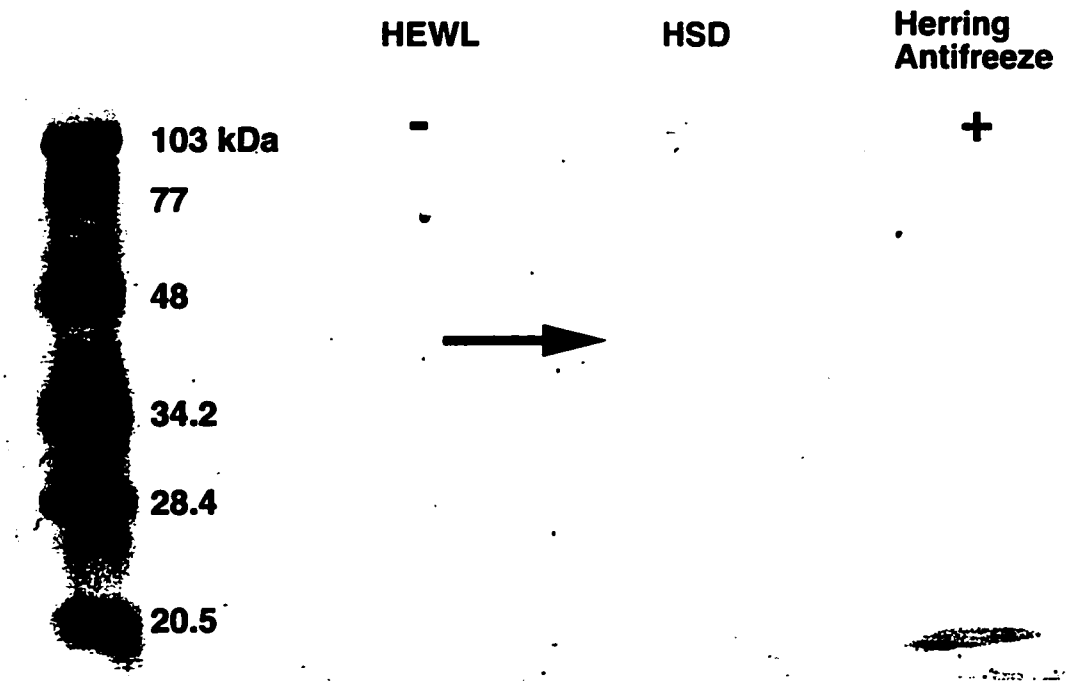
Within error limits, there is no difference for measured values of $CaCl_2$, nor is there a difference when calmodulin is added. There is an approximately 25% decrease in activity resulting from the addition of calcium to the reaction mixture. A similar experiment (data not shown) where the $CaCl_2$ was increased from 0 to 10 mM showed a slight *increase* in HSD activity. This indicates that the actual error values are likely larger

---

* HSD was at a concentration of 5 μg/ml

† This value was calculated by dividing the slope by the extinction coefficient of NADH (6300) and multiplying by the reaction volume (1 ml). The error is based on two measured values.

Figure 7.2 Ruthenium Oxide Stain of Proteins Transferred to Nitrocellulose



The ruthenium oxide assay described in the text is shown. Hen egg white lysozyme (HEWL) (14 kDa, negative control) and herring antifreeze (17 kDa, positive control) surround the lane in which HSD (36 kDa) was run. Low range molecular markers are in the leftmost lane. The position of monomeric HSD is indicated with an arrow.

than those quoted from the duplicate experiments. The reverse (oxidizing) direction activity is most likely unaffected by either divalent calcium or calmodulin within the timespan of these experiments.

7.4 Discussion

The peaks identified during the water picking stage were verified by the simulated annealing omit maps. They were initially modeled as calcium ions for two reasons: the beneficial effect of calcium ions upon crystallization and the coordination geometry calcium ions can exhibit (4). Any metal occupying the site defined by the residues shown in Figure 7.1 would have to be capable of flexible coordination geometry; this is true more so for calcium ions than it is for sodium ions. Furthermore, the ruthenium oxide assay demonstrated that HSD could bind calcium ions.

There was evidence that calcium has a role in HSD from other species. The activity of the spinach HSD enzyme, purified from natural sources, exhibits modestly enhanced activity in the presence of both calcium and calmodulin (5). However, these results could not be duplicated for the *S. cerevisiae* enzyme.

Other biochemical data suggest that the metal ion in the crystal structure is a sodium ion. HSD from *S. cerevisiae* and other species are affected by the presence of monovalent ions, *e.g.* sodium and potassium, as discussed in Chapter 1. Increasing concentrations of potassium ions were shown to decrease the lag time for the *S. cerevisiae* enzyme to

reach its maximal velocity (6). Potassium ions were also shown to afford some protection against heat denaturation for the *S. cerevisiae* enzyme (7). HSD denatures at 62° C in the presence of 0.5 M KCl and at 45° C in the absence of KCl.

These observations are supported by an examination of the B factors of the metal ions and the atoms of the ligating residues. The sodium ion refined to B factor values (Table 7.2) that were more similar to the surrounding atoms whereas the refined B factors for the calcium ions (Table 7.3) were quite different from the surrounding atoms. The higher B factors obtained with the calcium atoms were initially thought to arise from poor occupancy in the metal ion binding site, but it appears more likely that the elevated B factors were due to a misidentification of the metal ion.

The formal charge balance of the binding site favours a monovalent ion over a divalent ion. The only sidechain that coordinates the metal ion is a carboxylate group from Glu 143. The coordination sphere of the metal ion is completed by backbone carbonyl groups from Glu 143 and the other 4 residues which make up the binding site. Thus when sodium or potassium bind, the overall charge will be neutral, as opposed to the positive charge generated when calcium ions bind. The presence of this excess charge may destabilize the local conformation of the protein.

The failure of the lanthanide substitution experiments can be construed as support for this observation. When trivalent lanthanide ions

(La^{+3}) are substituted for divalent calcium ions (Ca^{+2}), there is a charge mismatch that must be accommodated by the protein. It appears that a similar charge imbalance is accommodated when HSD binds divalent calcium in place of the monovalent sodium or potassium (Na^{+} or K^{+}), as demonstrated by the ruthenium oxide based assay. However, the charge discrepancy is likely too great when the trivalent lanthanides are in the metal binding site.

The final and best evidence for the identification of the atom as sodium is that calcium ions were not deliberately added to the crystallization conditions which yielded the crystals ultimately used for the structure refinement. Sodium was not deliberately added either, but it was present at a minimal concentration of 0.1 M as a counter ion to the buffers used in the crystallization conditions for both the tetragonal and monoclinic crystals. The initial identification of the atom as calcium was made because the deliberate addition of calcium ions appeared to have had a beneficial effect upon crystal to crystal R_{merge}. Because the high resolution tetragonal crystal which generated the data used for structure refinement merged well with crystals grown in the presence of calcium ions, it was assumed that the growth conditions for that crystal had been contaminated with calcium ions. The error was carried over to the monoclinic data set when the tetragonal model was used during molecular replacement. Although the calcium ions were not included in the search

model, when density was observed in the map at the metal binding sites, it was assumed to be a calcium ion.

7.4.1 The Role of the Metal Ion

It is possible to rationalize the effects of the metal ions with the protein crystal structure. The metal ion binding site is flanked by β-f and α-2H, putting it near to the dimer interface. This may explain how the metal ions are able to exert their effect upon the enzyme: the ion serves a structural role in keeping the dimer interface intact so that the enzyme active site may exist in an optimal conformation. An intact enzyme interface appears to be necessary for maximal enzyme function, since mutation of a histidine residue, His 309, near the dimer interface was shown to decrease enzyme activity by approximately 40% (6). Others (8-10) have speculated on a regulatory role for the metal ions. The evidence given above, while obtained from a variety of experimental techniques, is certainly not compelling in demonstrating the presence of a sodium ion binding site in the *S. cerevisiae* HSD. Further work will be required to definitively show the presence of a metal ion binding site in the protein structure. Possible experiments include Tb^{3+} substitution followed by fluorescent experiments (12) or collecting X-ray data on HSD crystals which have been soaked in potassium ions.

7.5 Summary of Chapter 7

The presence of a metal ion in the HSD crystal structure was proposed based upon the conformation of residues 143, 147, 148, 149 , and 150. The five residues form a distorted six coordinate antiprismatic geometry around an electron density peak which appears at better than 3σ levels in simulated annealing omit maps.

Several different investigations were used in an attempt to identify the nature of the occupant of the metal ion binding site. Although it was demonstrated with a ruthenium oxide staining test that HSD could bind Ca^{2+} ions, the crystallographic B factors indicated that the species is more likely to be either a Na^{+} ion or a water molecule.

Neither Ca^{2+} nor calmodulin had any measurable effect upon enzyme kinetics, contrary to results reported for the spinach form of HSD.

Because of the minimal difference between Na^{+} and H_2O X-ray scattering factors, the identity of the peak as a metal ion could not be made with crystallographic data alone. Further work will be required to definitively identify the nature of this atom currently being modeled as a sodium ion.

References Chapter 7

1. **Read, R. J.** (1996). As MAD as can be. Structure. **4**: 11-14.
2. **Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L.** (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr. **54**: 905-921.
3. **Ewart, K. V., Yang, D. S., Ananthanarayanan, V. S., Fletcher, G. L., and Hew, C. L.** (1996). Ca^{2+}-dependent antifreeze proteins. Modulation of conformation and activity by divalent metal ions. J Biol Chem. **271**: 16627-16632.
4. **Cowan, J. A.**, *Inorganic Biochemistry*, VCH Publisher, New York (1993).
5. **Pavagi, S., Kochhar, S., Kochhar, V. K., and Sane, P. V.** (1995). Purification and characterization of homoserine dehydrogenase from spinach leaves. Biochem Mol Biol Int. **36**: 649-658.
6. **Jacques, S.**, Characterization of Homoserine Dehydrogenase from *Saccharomyces cerevisae*: A New Fungal Target. Doctoral Thesis, in *Dept of Biochemistry*, McMaster University, Hamilton, Ontario, Canada (1999).
7. **Yumoto, N., Kawata, Y., Noda, S., and Tokushige, M.** (1991). Rapid Purification and Characterization of Homoserine Dehydrogenase. Arch. Biochem. Biophys. **285**: 270-275.
8. **Wedler, F. C., and Ley, B. W.** (1993). Homoserine dehydrogenase-I (*Escherichia coli*): action of monovalent ions on catalysis and substrate association-dissociation. Arch Biochem Biophys. **301**: 416-423.
9. **Wedler, F. C., and Ley, B. W.** (1994). Homoserine dehydrogenase-I (*Escherichia coli*): action of monovalent ions on catalysis and substrate association-dissociation. Biophys J. **67**: 2251-2260.
10. **Wedler, F. C., and Ley, B. W.** (1993). Kinetic and regulatory mechanisms for (*Escherichia coli*) homoserine dehydrogenase-I. Equilibrium isotope exchange kinetics. J Biol Chem. **268**: 4880-4888.
12. **Glusker, J. P.** (1991). Structural Aspects of Metal Liganding to Functional Groups in Proteins. Advances in Protein Chemistry. **42**: 1-76.
13. **Brennan, J.** (2000). Personal communication.

Chapter Eight: Cross Species Comparison of Homoserine Dehydrogenase

The purpose of this chapter is to extend the structural work done for the HSD of *S. cerevisiae* to other species. The results described in this thesis are the first and currently only structure available for HSD. Therefore, the extension will be accomplished by comparing HSD primary amino acid sequences from the various species. Thirty different HSD sequences will be considered. These sequences arise from 27 different species with representatives from the archeabacteria, bacterial, fungal, and plant kingdoms. The notable exception is the animal kingdom, for which there are no known HSD proteins.

The conservation pattern of HSD among species has also been uncovered during massive cross analyses of the Genbank bacterial genomes (1). The gene encoding HSD is found in a cluster of orthologous group of proteins (COG) which describes threonine metabolism. Furthermore, Aravind *et al.* (2) have noted that this group of enzymes is related to enzymes which are allosterically controlled by L-Ser *via* their C terminus.

8.1 Methodology

8.1.1Primary Sequence Alignment

The unique homoserine dehydrogenase primary amino acid sequences were taken from the current GenBank database (3). Alignment was performed with ClustalW 1.7 (4). Default settings were used to generate the initial alignment: blosum matrix, gap opening penalty of 10.0, gap extension penalty of 0.05, and gap openings in hydrophilic or loop forming residues (GPSNDQERK) were favoured. The initial alignment was hand adjusted to take into account secondary structural elements of the *S. cerevisiae* HSD structure in the region surrounding β-strand d. The sequence alignment was reproduced with a variety of gap opening and extension penalty values. Furthermore, randomized input order of the sequences did not significantly alter the resulting alignments.

8.1.2 Primary Structure Alignment Mapped to the Tertiary Structure

The primary sequence conservation is expressed as the percentage of conservation a residue from the *S. cerevisiae* sequence appears in all of the other sequences on a residue by residue basis.

The amount of buried sidechain in the *S. cerevisiae* HSD structure was calculated from values determined by the methods of Richards and Lee (5, 6), as implemented in CNS 0.9. The model tripeptide values for each of the 20 amino acids were used to calculate the percentage of

buried sidechain. Tabulated values for residues which could exist in multiple conformations were averaged.

The surfaces were calculated with GRASP (7) using the HSD tetragonal A molecule.

8.2 Results

8.2.1 Alignment of 30 HSD primary sequences

The multiple sequence alignment is shown in Figure 8.1. The numbering scheme and secondary structural elements displayed at the top of the alignment are for the *S. cerevisiae* enzyme. The numbering scheme which appears after the C terminus of the *S. cerevisiae* enzyme is for the *C. glutamicum* enzyme (8).

The sequence alignments displays three distinct groups of HSD proteins:

1. HSD proteins which are monofunctional and lack a C terminal extension (sequences 1-9). The proteins in this group associate as dimers and do not exhibit allosteric regulation.
2. Bifunctional proteins with both HSD and aspartate kinase domains (sequences 10-19). The aspartate kinase domain is located in the N terminal region of the protein. The proteins in this group associate as

tetramers; the tetrameric forms are sensitive to allosteric regulation by L-Thr.

3. HSD proteins which possess a C terminal extension (sequences 20-30). The C terminal extension is an approximately 100 residue domain which occurs after the C terminus in the *S. cerevisiae* enzyme. These enzymes associate as dimers and are allosterically regulated by L-Thr.

8.2.2 Phylogenetic Tree

The distances between sequences are based on the multiple sequence alignment. The tree was generated from primary amino acid sequences based on a parsimony method of Felsenstein (23) employing the PAM matrix (24) and takes into account the genetic code underlying amino acid mutations. The tree was generated from a randomized entry order of the sequences. Subsequent phylogeny trees generated from other randomly generated sequence entry orders were essentially identical to the tree presented herein. The programs in the Phylip suite (25) were used to perform the calculations. The phylogenetic tree is shown in Figure 8.2.

Figure 8.1a - Multiple Alignment of HSD Sequences (Caption Follows)

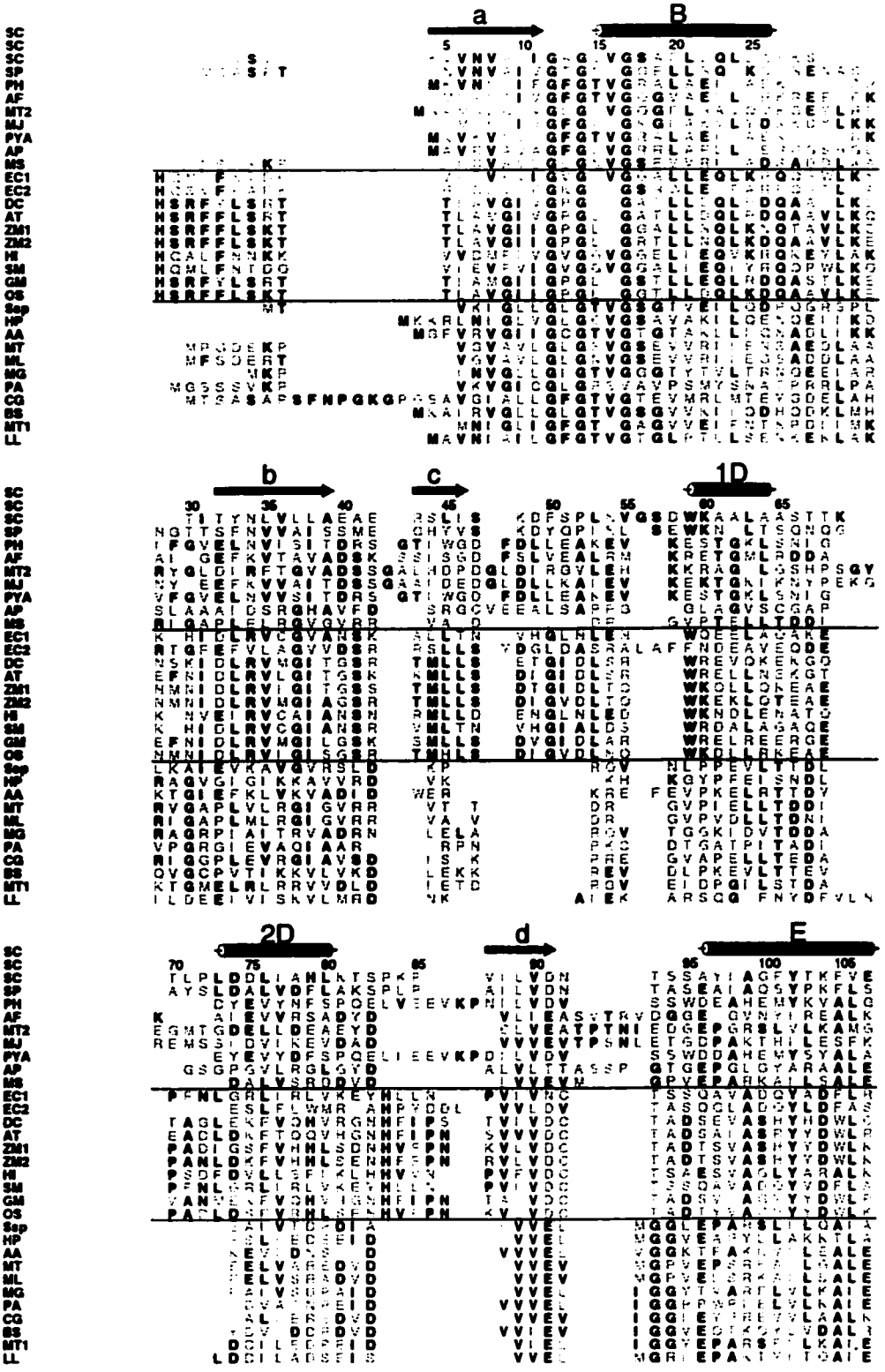


Figure 8.1b - Multiple Alignment of HSD Sequences

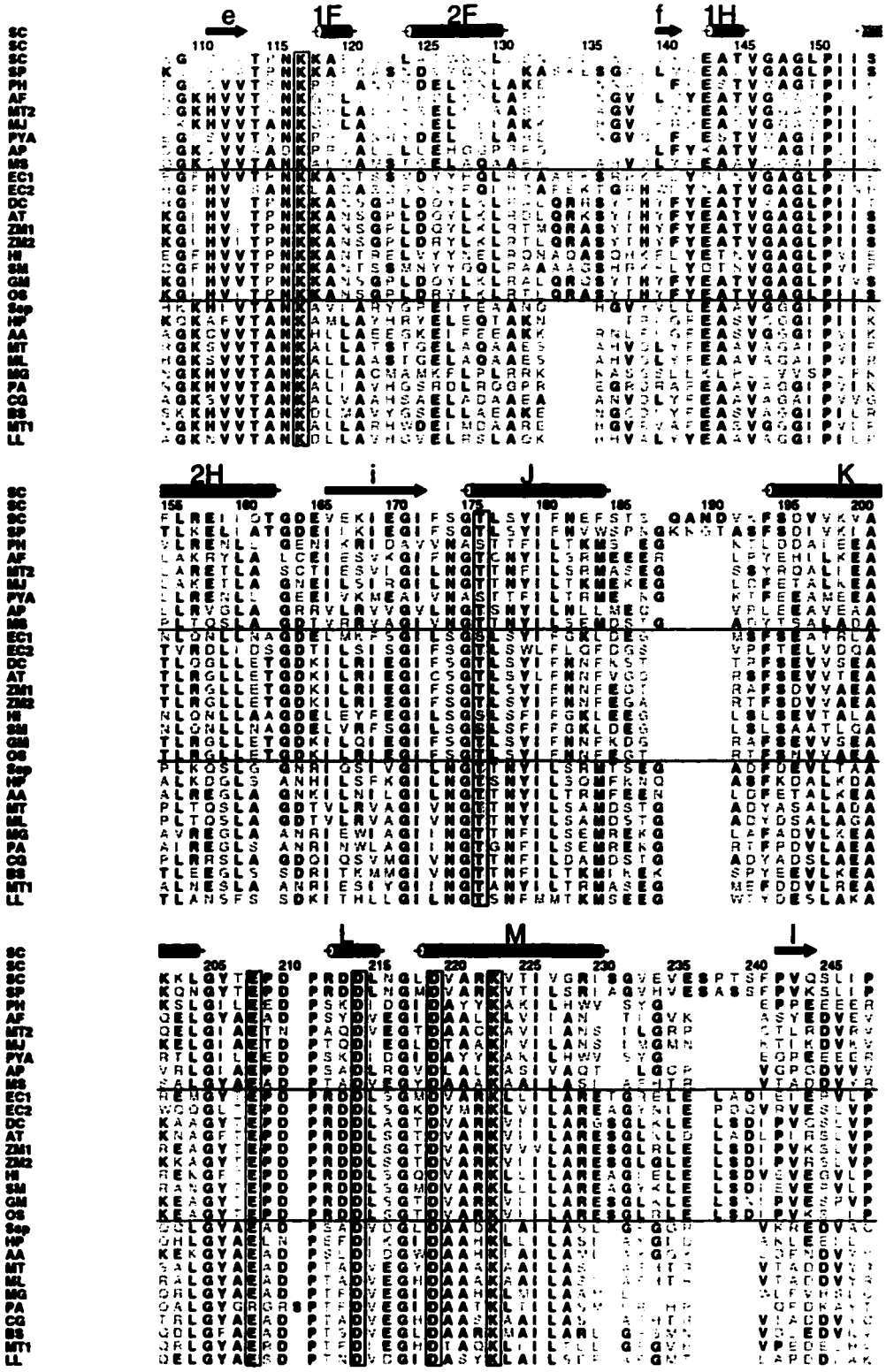


Figure 8.1c - Multiple Alignment of HSD Sequences

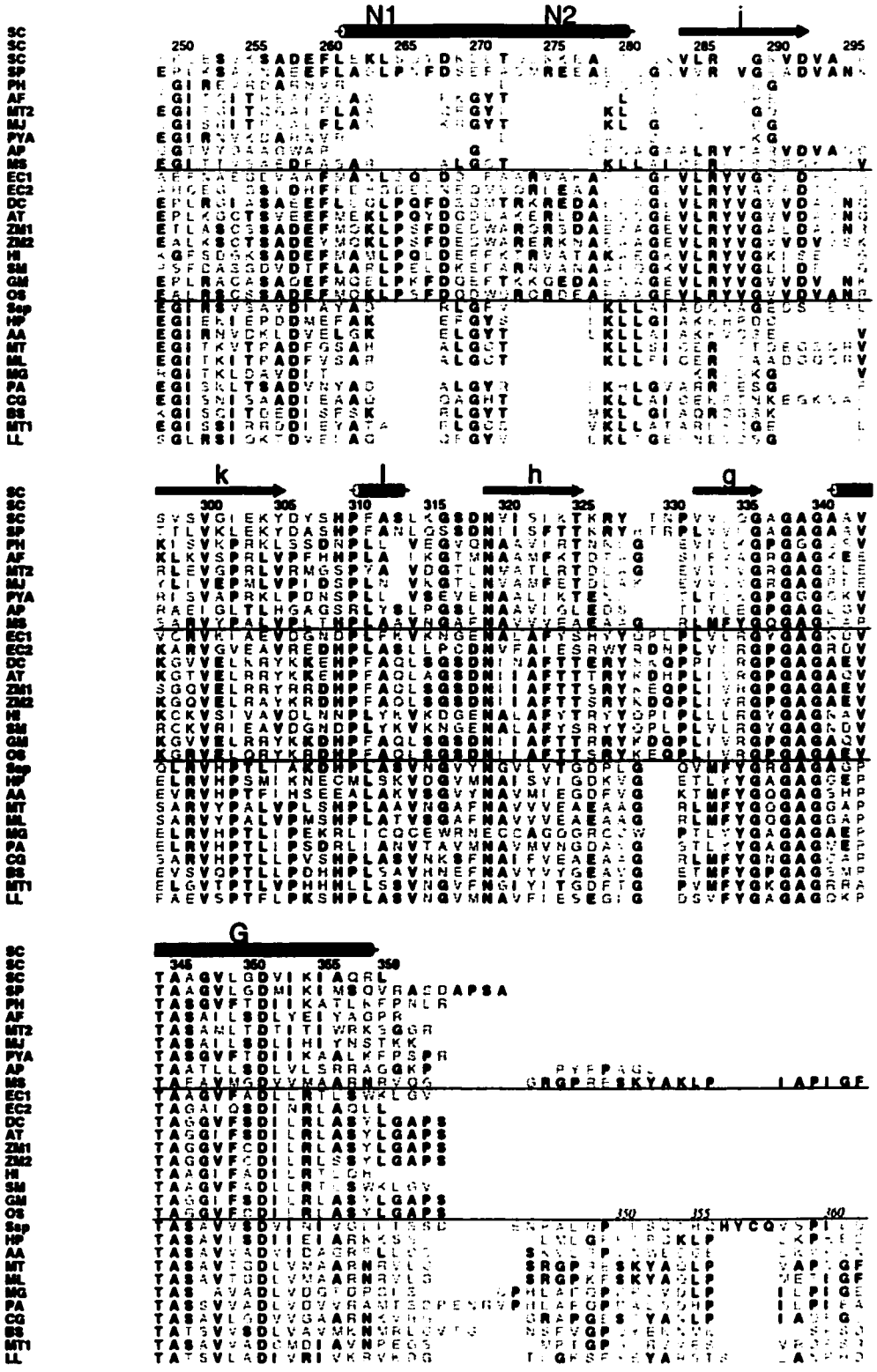
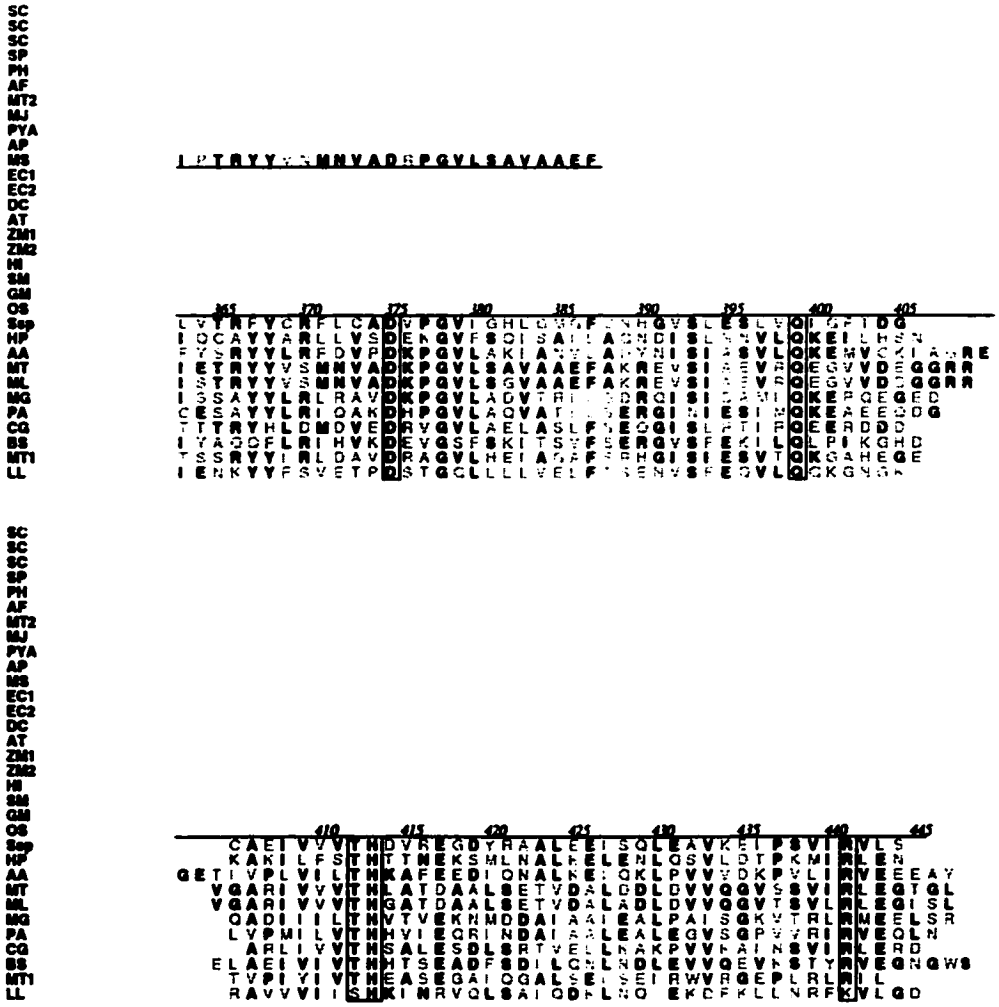


Figure 8.1d - Multiple Alignment of HSD Sequences



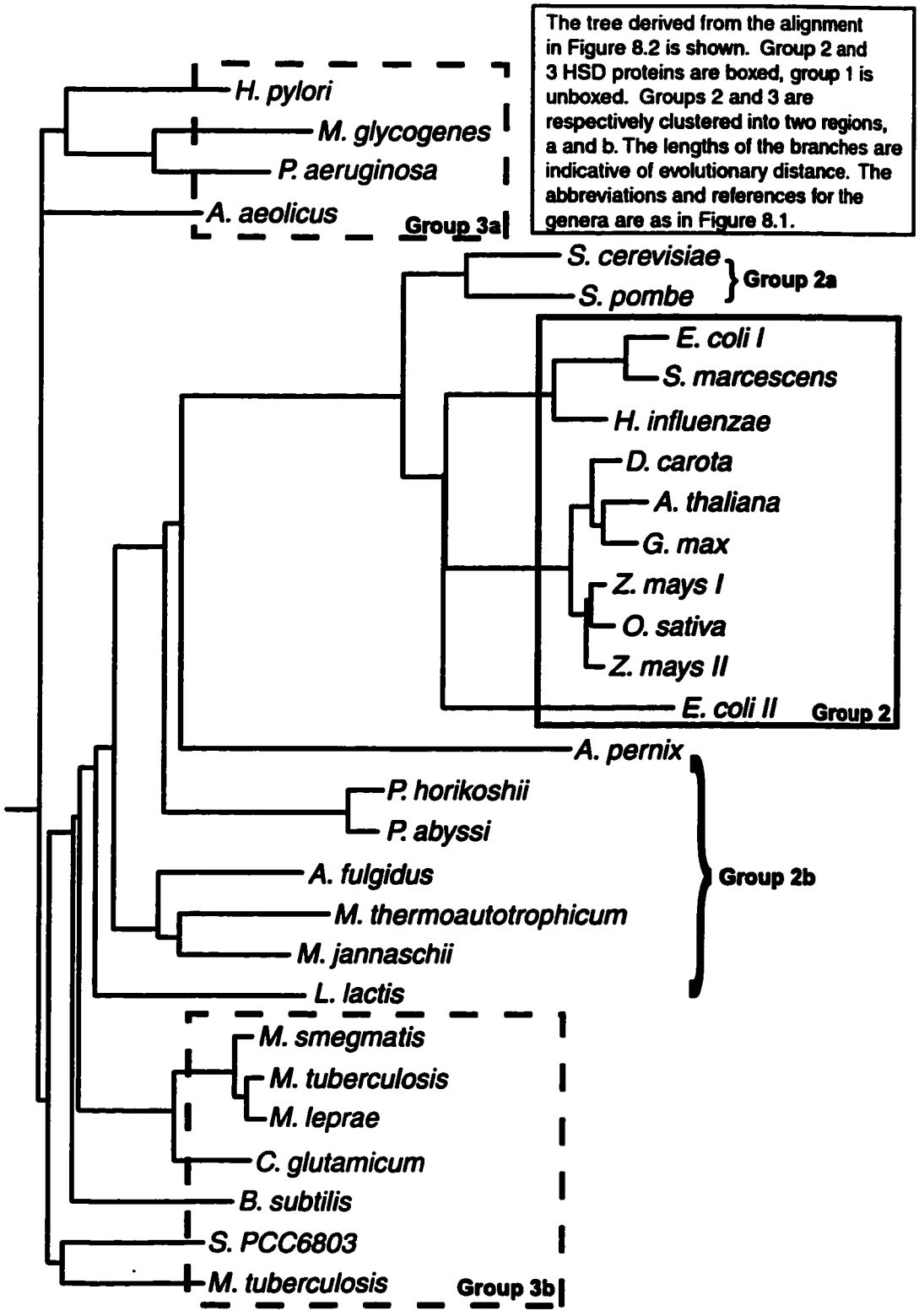
The multiple sequence alignment of 30 different HSD sequences is shown in stretches of ~40 residues. The most frequently conserved residue in each position is shown in bold. Residues implicated in substrate binding and / or catalysis are contained within shaded boxes, and residues in the C terminal extension for putative binding of the allosteric inhibitors are contained within open boxes. The species names are given in shortform down the left hand side of the page. The shortforms are given in the legend immediately following this page. The horizontal rule divides the group 1, 2, and 3 HSD enzymes which are shown in each stretch at top, middle, and bottom, respectively. The numbering scheme is according to the *S. cerevisiae* enzyme. The N termini (aspartate kinase bearing) region of the group 2 enzymes are truncated. For the group 3 HSD enzymes, the numbering scheme for the *C. glutamicum* enzyme is used past the C terminus of S. cerevisiae. The secondary structure, as defined by the *S. cerevisiae* enzyme, is given across the top of each stretch with helices denoted by cylinders and strands denoted by arrows. The figure was generated with *ALSCRIPT*.

Table 8.1 and Legend for Figure 8.1 and 8.2 – Species Used in Multiple HSD Sequence Alignment

Code	Organism	Reference / Accession*
SC	*Saccharomyces cerevisiae*	(9)
SP	*Schizosaccharomyces pombe*	AL035263
PH	*Pyrococcus horikoshii*	3257491
AF	*Archaeoglobus fulgidus*	2649661
MT2	*Methanobacterium thermoautotrophicum II*	2621481
MJ	*Methanococcus jannaschii*	(10)
PYA	*Pyrococcus abyssi*	CAB49823
AP	*Aeropyrum pernix*	BAA80129
MS	*Mycobacterium smegmatis*	AAD32592
EC1	*Escherichia coli I*	(11)
EC2	*Escherichia coli II*	(12)
DC	*Daucus carota*	(13)
AT	*Arabidopsis thaliana*	3250680
ZM1	*Zea mays I*	1351904
ZM2	*Zea mays II*	1351905
HI	*Haemophilus influenzae*	(14)
SM	*Serratia marcescens*	(15)
GM	*Glycine max*	2970447
OS	*Oryaza sativa*	1777375
Ssp	*Synechocystis PCC6803*	(16)
HP	*Helicobacter pylori J99*	(17)
AA	*Aquifex aeolicus*	2984093
MT	*Mycobacterium tuberculosis*	1706418
ML	*Mycobacterium leprae*	1169334
MG	*Methylobacillus glycogenes*	(18)
PA	*Pseudomonas aeruginosa*	(19)
CG	*Corynebacterium glutamicum*	(20)
BS	*Bacillus subtilis*	(21)
MT1	*Methanobacterium thermoautotrophicum I*	2622343
LL	*Lactococcus lactis*	(22)

* The accession number for the NCBI database is given in cases where sequences were submitted directly to the database without accompanying publication in a traditional journal.

Figure 8.2 Phylogenetic Tree Derived from Sequence Alignment

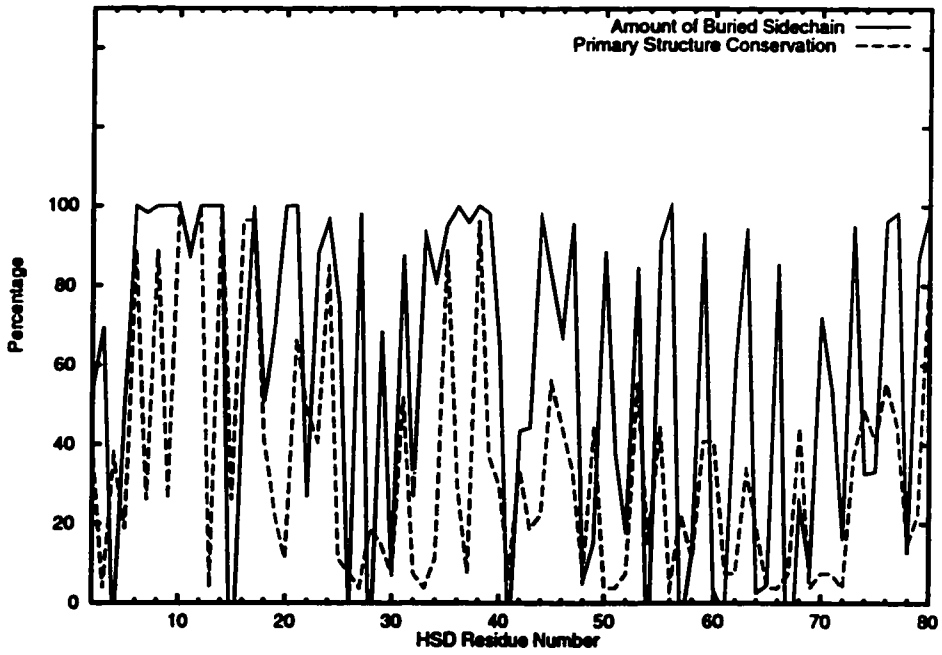


8.2.3 Primary Sequence Conservation Mapped to the HSD Structure

A plot of the percentage of buried side chain and primary sequence conservation on a residue by residue basis is shown in Figure 8.3a-e. The comparison suggests that the *S. cerevisiae* structure is representative of HSD structures from all species. The assumption is made that the buried residues cannot be extensively mutated without changing the overall fold of the protein. The mutation of a single buried residue is an unlikely occurrence because it would necessitate one or more compensating mutations to maintain similar packing of the protein core.

The surface of the protein, however, will not have the same packing constraints as the interior of the protein. Thus, greater variability in amino acid residues would be expected and is shown in the conservation patterns mapped onto the surface of HSD (Figure 8.4). An exception to this is the conservation which occurs at the exposed active site of the protein, shown in Figure 8.4.

Figure 8.3A - Comparison of Buried and Conserved Residues in HSD
(caption follows last plot)



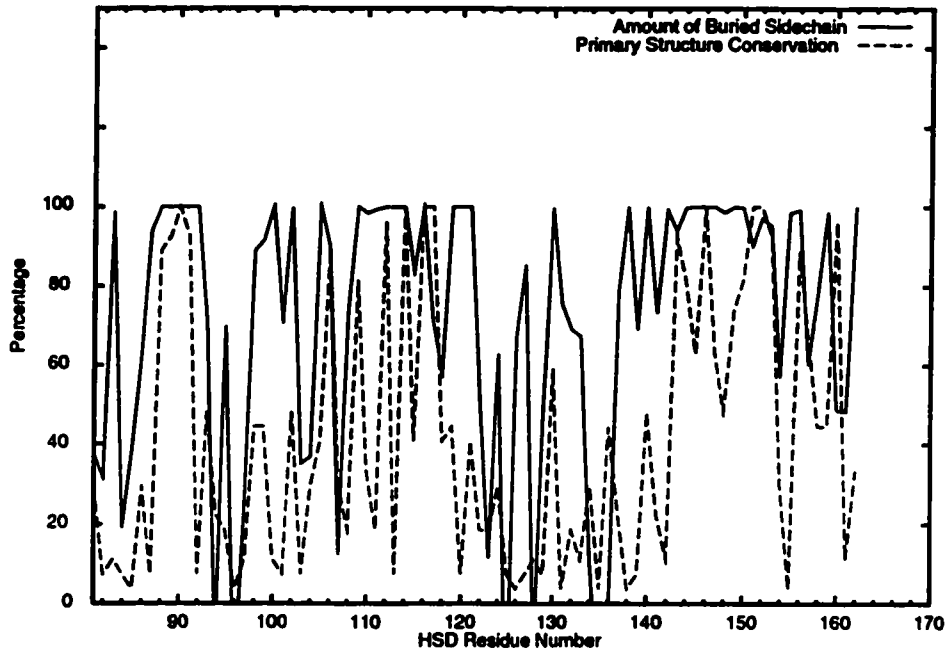
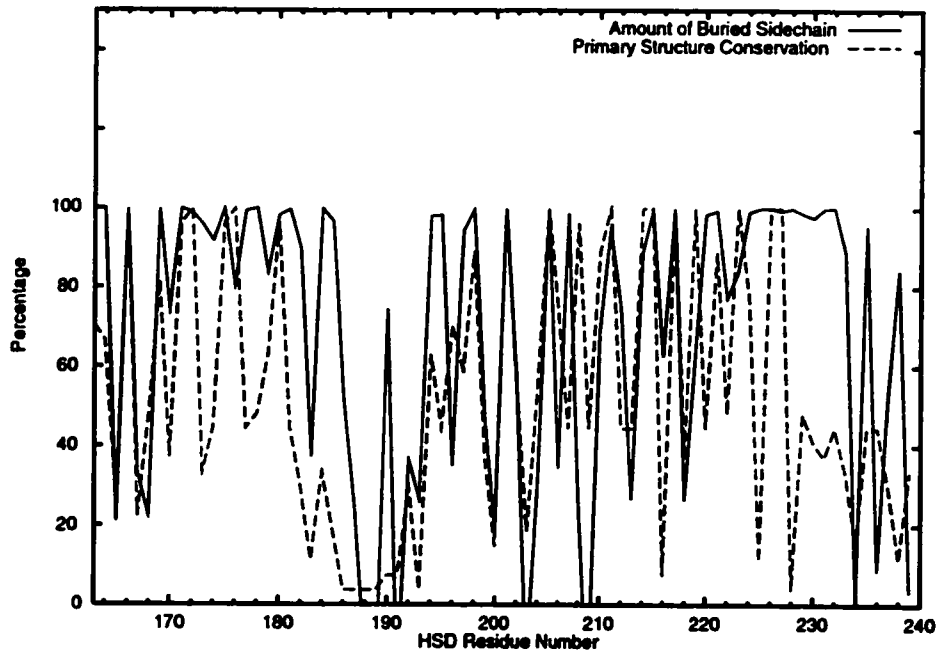


Figure 8.3B - Comparison of Buried and Conserved Residues in HSD



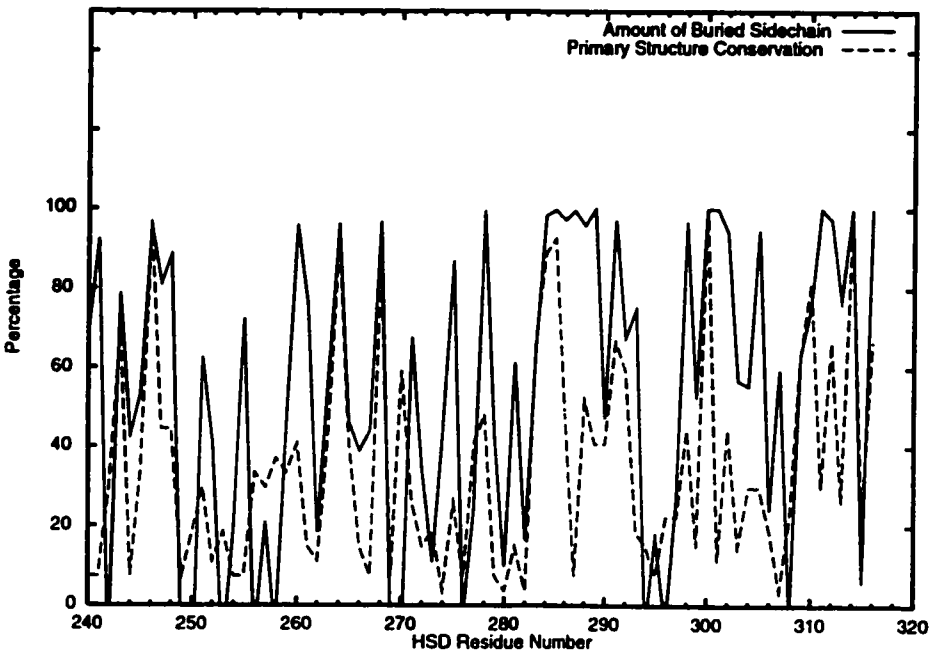
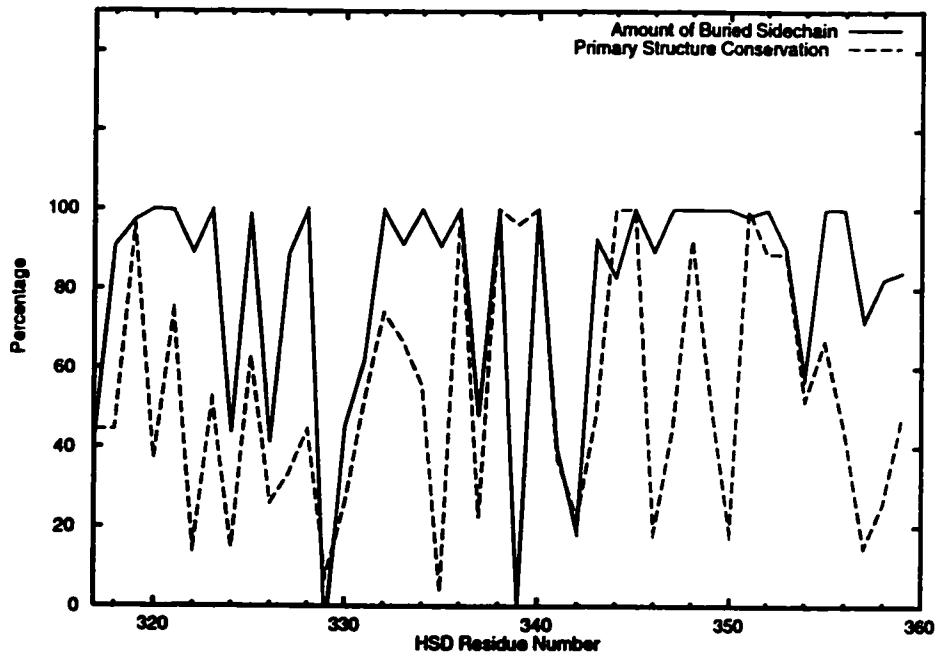
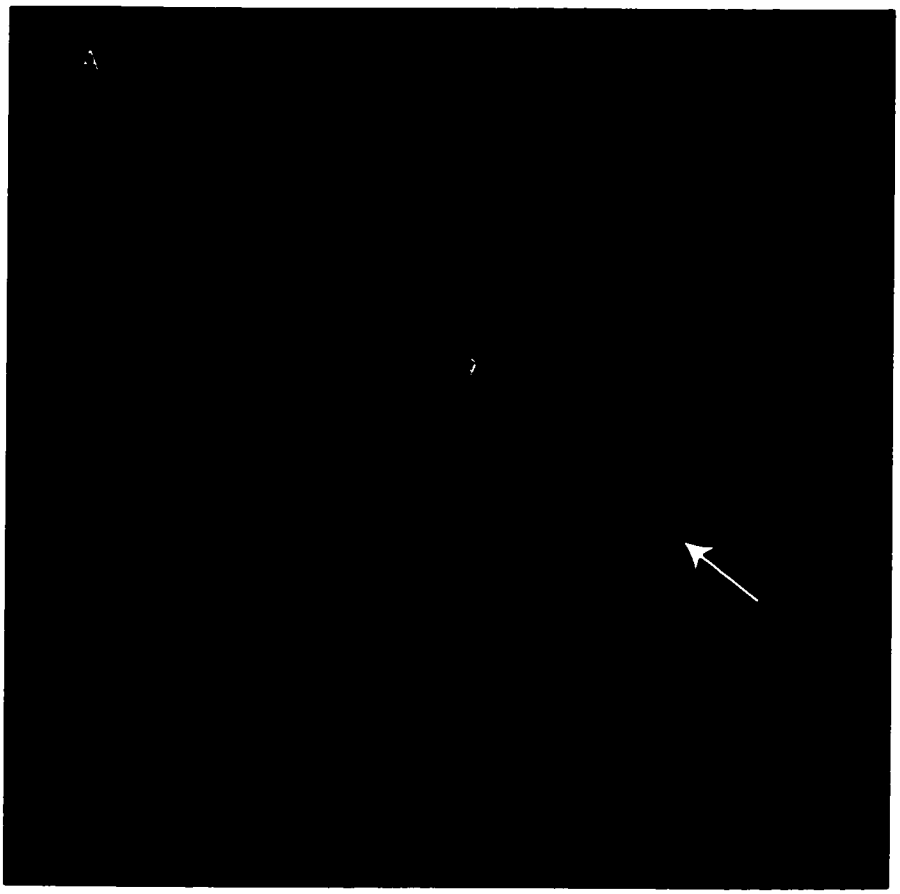


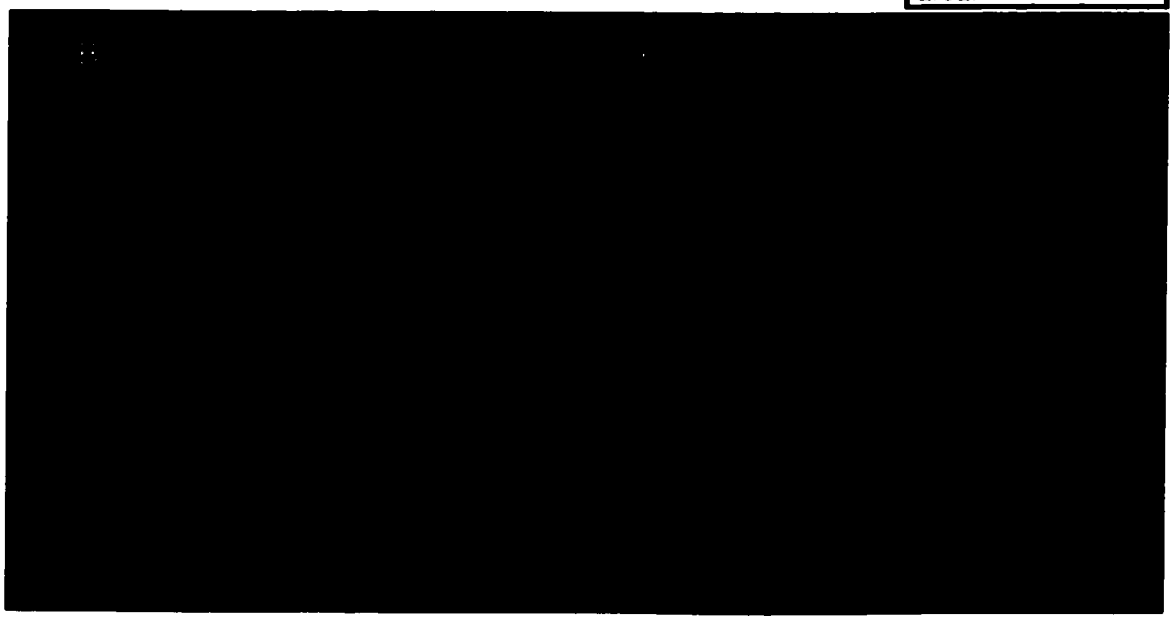
Figure 8.3C - Comparison of Buried and Conserved Residues in HSD



The preceeding 5 panels show co-plots of the primary structure conservation and the amount of buried sidechain, on a residue by residue basis. Primary structure conservation was calculated by how often the residue appearing in the *S. cerevisae* HSD primary structure appeared in other species' HSD primary structures (see Figure 8.1). The amount of buried sidechain is the reciprocal of the amount of solvent exposed sidechain, calculated according to standard methods described in the text. The values were calculated from the *S. cerevisae* quaternary structure. The amount of buried sidechain can be taken as a measure of how likely a positional mutatation will affect the overall fold, with higher values indicating a greater likelihood of fold variation. Both values have been normalized to 100 and expressed as a percentage. 100% thus corresponds to either a fully conserved residue in the primary structure or a fully buried residue in the quaternary structure.

Figure 8.4 - Variability of Homoserine Dehydrogenase Residues on the Protein Surface





(A) A van der Waal surface of the *S. cerevisiae* HSD dimer in (orientation as shown in Figure 4.3B). The primary structure conservation has been mapped to the surface, with red indicating high variability and blue indicating high conservation. The arrow indicates the location of the active site. ***(B)*** is the protomer in the standard view to show the conservation at the dimer interface (indicated with a green box) ***(C)*** A close-up of the active site, showing the conservation of residues in this area.

8.3 Discussion

8.3.1 Active Site Conservation

The overall fold of HSD appears to be conserved. The strongest evidence for this is the strict conservation of the active site residues in the multiple sequence alignment and the area they map to on the surface of the protein. It is difficult to conceive of alternate folds which would bring these residues into a similarly close proximity.

The L-Hse binding/catalytic residues (Lys 117, Thr 176, Glu 208, Asp 214, Asp 219, Lys 223) represent the most striking conservation among the different species. In addition to these key residues, there are other strictly conserved residues in the active site region (Figure 8.1). These residues were first assumed to be important for catalysis by Parsot *et al.* over 10 years ago (26) in a comparison of the sequences from *B. subtilis* and *E. coli* HSD. The single exception to the strict conservation is the substitution of an arginine for Glu 208 in the *P. aeruginosa* sequence. The substitution is explicable in light of how the substrate is bound in the enzyme, as discussed in chapters 5 and 6.

8.3.2 Other Conservation Patterns of HSD Across the Species

Other important regions in the enzyme are well conserved. These include the dinucleotide binding fold initiating turn at position 12, the metal

ion binding site at residues 142 through 149, and the g-G loop. The g-G loop precedes the helix at the C terminus to the dinucleotide binding domain of the N terminus.

The residue by residue comparison of side chain burial and sequence conservation also supports the view of a well conserved HSD fold. Further evidence for the common fold arises from the gaps which were necessary to generate the best alignment. The gaps occurred primarily in loop regions of the *S. cerevisiae* structure. Conservation among the secondary structural elements and variability in the loop regions would be expected for proteins with highly similar folds.

The main exception to the conserved secondary structure / variable gap observation is the α-N1/N2 region. The primary sequence comprising the 2 N helices is highly variable. Of particular note are the large gaps which occur for most of the group 1 and all of the group 3 HSD sequences. The primary sequence of the N helices are highly variable for these two groups of enzymes. The helices may be altogether absent from the non-fungal group1 and all group 3 HSD structures. These helices had high B factors and poor electron density in both crystal structures discussed in this thesis. The two helices do not appear to have a well defined functional or structural role, apart from connecting the elements which surround them. It is likely that they could be replaced with a

random coil structure without adverse effects on the enzymatic function of the protein.

The two fungal forms of HSD in the alignment, *S. cerevisiae* and *S. pombe* exhibit some interesting common features in their primary sequence. These sequences are more similar to the bifunctional group than they are to the short monofunctional group. This is obvious in regions around the 1F 3_{10} helix, the substrate binding regions of loop K-L, and helices L and M. There are several comparisons in the dinucleotide binding fold also, as observed in β-strand d and α-helix E. A third fungal sequence, *C. albicans,* whose sequence has not yet been deposited in the GenBank shows a high degree of similarity to both enzymes.

8.3.3 Evolutionary Aspects of HSD

The overall fold for all of the HSD enzymes is likely to be similar, with three different variations in the common fold. Parsot *et al.* (26) have suggested that HSD evolved from a nonregulated enzyme to a regulated enzyme in two different genetic fusion events, splitting the enzyme at that point into two groups evolved from one common ancestor. One fusion event was with the gene encoding aspartate kinase generating the bifunctional aspartate kinase / HSD proteins (group 2). The second fusion event generated the longer (group 2) proteins by the acquisition of an approximately 100 residue C terminal domain. The tripartite grouping of HSD enzymes is reflected in the conservation patterns observed for the

HSD enzymes. Sequence gaps (*e.g.* the c-1D loop, the 2D-d loop, the M-I loop) are conserved within groups, with some similarities between groups 1 and 3. Some residues and motifs are strictly conserved within groups and only moderately conserved in other groups (*e.g.* GXHV at positions 109-112 in group 2, lysine at position 110 in group 3, VTANKXXXA at positions 113-121 of group 3).

However, an alternate model in which the ancestral form was more like the group 3 HSD enzymes is also possible. The order of events can be described such that the first change was the loss of the C terminal extension. This was followed by fusion of the HSD encoding gene with the aspartate kinase encoding gene to create the bifunctional enzyme. In this model, the evolution of the HSD enzyme is the result two sequential gross genetic events rather than the branching genetic events described by Parsot *et al.*

The latter model is supported by the phylogenetic tree created from the sequence alignment, the location of the N and C termini in the structure, and the nature of allosteric inhibition in the group 2 and 3 enzymes.

The horizontal distances of phylogenetic tree represent the evolutionary distance calculated from the sequence alignment. The group 3 enzymes are closer, albeit only slightly, than the group 1 enzymes are to the root of the tree. Furthermore the group 1 enzymes are clearly closer

in their evolutionary pattern to the group 2 enzymes than the group 3 enzymes.

The second point to consider is the mutually exclusive nature of group 2 and group 3 enzymes. When HSD activity is found in the same protein as AK activity (*i.e.* group 2), the HSD functional domain is encoded by a sequence comparable to the length of the group 1 proteins. The tertiary structure of HSD can easily explain this phenomenon. In the structure, the N and C termini are located less than 5 Å away from one another. Since the aspartate kinase domain is always attached to the N terminus of HSD, a C terminal extension would likely interrupt the interface between the two domains. When the AK/HSD interface is disrupted in the group 2 enzymes, decreased activity is observed for both enzymes and allosteric regulation is lost (27).

Further support for the sequential evolution model comes from the nature of the allosteric regulator of HSD. Inhibition of these enzymes occurs at both the genetic and protein level. The genetic level inhibition will not be discussed here. The protein level inhibition arises from an end product of the aspartate pathway, L-threonine. One hypothesis for the occurrence of allosteric regulation is that enzymes in metabolic pathways evolve from the enzyme which catalyses the final step of the reaction (28). The evolutionary pressure arises from a scarcity of the starting products, beginning at the substrate for the ultimate reaction in the pathway and

continuing backwards to the present day starting point, *i.e.* L-aspartate for the aspartate pathway. Thus allosteric inhibition is recognition by enzymes in the middle or beginning of the pathway for their original ligands which are now at the end of the pathway. The sequential evolution model of HSD fits very well with the allosteric evolutionary model. The C terminal extension is likely a remnant of the original L-Thr producing enzymes. In the group 1 HSD enzymes, the binding site was eliminated when the C terminal tail was lost, ostensibly by the creation of an early stop codon. Aspartate kinase, as part of the aspartate pathway, would also posses the ancestral recognition site, so when the AK/HSD gene fusion occurred, the allosteric inhibition was regained in the bifunctional form of HSD.

8.3.3.1 A Switch Region in HSD May be Responsible for Allosteric Control

There are two different manifestations of allosteric control in the HSD enzymes, one arising from the bifunctional interface (group 2) and a second arising from the C terminal extension domain (group 3). The allosteric behaviour of group 2 enzymes is better understood because of work done on the *E. coli* enzyme. There is no structural data for enzymes from either of the regulated HSD groups, so a model describing the fine details of regulation would be speculative at this point. However, given the structure of the *S. cerevisiae* enzyme, the assumption of a common

fold among the species, and differences observed in the primary sequence alignment, a crude model may be proposed.

The two different forms of HSD in *E. coli*, *E. coli* HSD I and *E. coli* HSD II differ in one important property: the former is allosterically affected by L-threonine whereas the latter is not. The two enzymes are 36% identical and 55% similar in the primary structure of their HSD components, so a simple pairwise comparison of the two gives little insight into the reason for the difference. However, when these sequences are compared in the context of the other group 2 sequences a key difference for residues 113 to 139 becomes apparent.

Within this span of residues, the group 2 motif of TPNKKAN(T/S) is broken only for the *E. coli* HSDII enzyme. The entire region is comprised of α-1F, α-2F and their surrounding loops in the crystal structure. The region contains an important residue, Lys 117. There is a nearby substitution at position 115. In *E. coli* HSDI residue 115 is a proline and in *E. coli* HSDII it is an alanine. Prolines have highly restricted torsion angles in comparison with alanine. Such a substitution will likely change the local conformation of the protein.

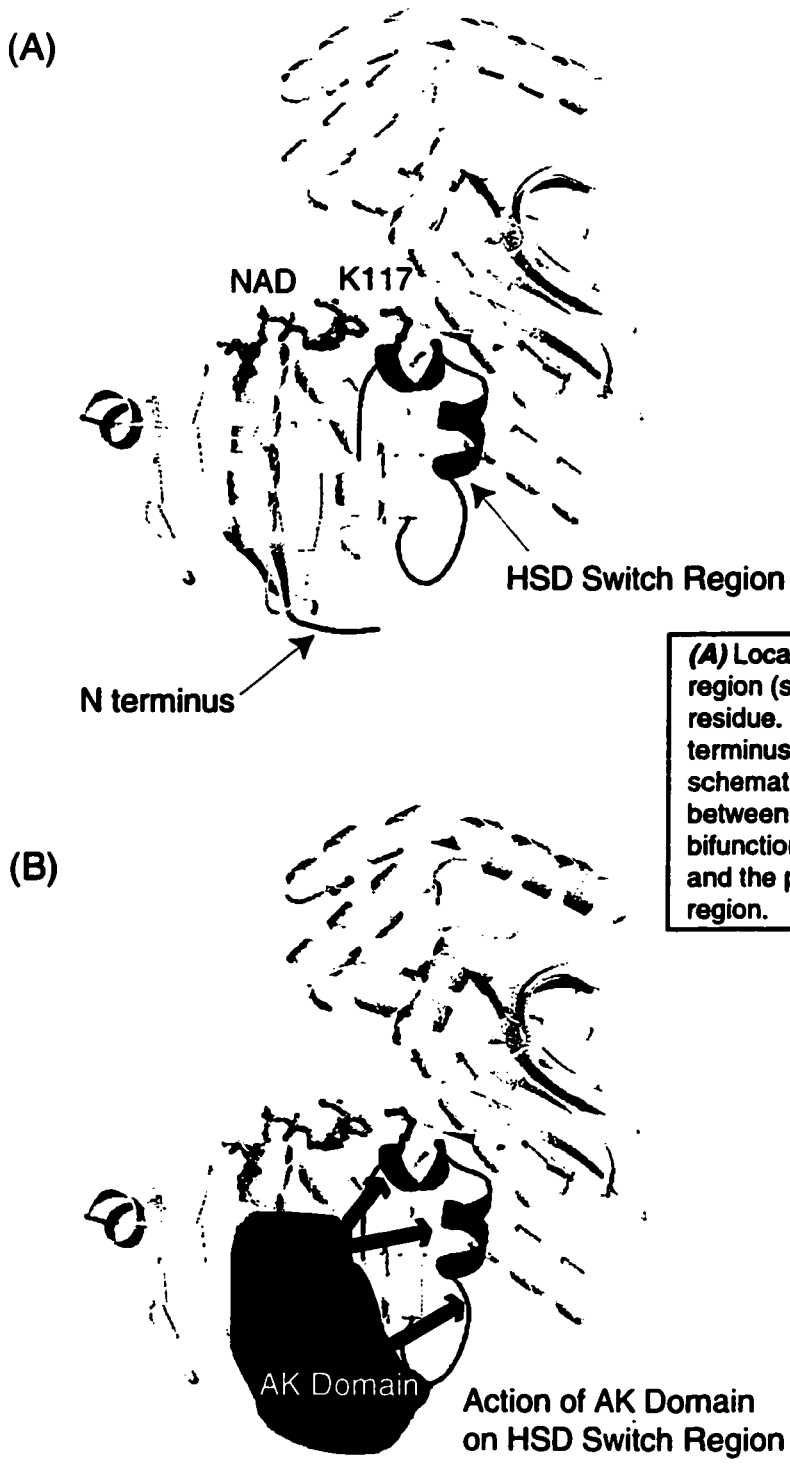
Lys 117 is important because it is the only strictly conserved residue which has side chain interactions with the NAD^+ molecule. Specifically, it forms a hydrogen bond with the 2' hydroxyl group of the ribose ring connected to the pyridyl group of the cofactor. It is the closest

contact to the pyridyl end of the molecule; as such the residue is in a position to affect active site catalysis. This is supported by the lack of enzymatic activity in the HSD K117A mutant (Table 6.2). The difference between the *E. coli* HSDI and the *E. coli* HSDII sequences that occur in this region is strong evidence for a model in which the allosteric inhibition arises from a disruption of the cofactor orientation. This disruption is effected by a conformational change of the Lys 117 bearing α-1F.

This model is analogous to the switch regions observed in the G proteins (29). The HSD switch region is shown in Figure 8.5, along with a view of the potential interaction which gives rise to the allosteric effect. The proposed interaction results from the adjacent aspartate kinase domain in the tetrameric form of the protein. The proximity of the two N termini in the dimer structure suggests an intermolecular interaction. This is in agreement with the observation that allosteric inhibition is observed only for the tetramer.

In addition to the intermolecular proximity of the N termini, the structure revealed an intramolecular proximity of the N and C termini. Since the allosteric inhibition exhibited by the group 3 HSD proteins is caused by a C terminal extension, the α-1F switch region may be applicable to these HSD proteins as well. However, because relatively little is known about the allosteric regulation of the group 3 enzymes, further speculation on this matter would be unwise.

Figure 8.5 - A Model For Allosteric Inhibition Showing the Putative Homoserine Dehydrogenase Switch Region



(A) Location of the proposed switch region (shaded) and the key Lys 117 residue. The proximity of the N terminus is indicated. ***(B)*** A schematic view of the interaction between the AK domain of the bifunctional Group 2 HSD enzymes and the proposed HSDswitch region.

8.4 Summary of Chapter 8

A multiple sequence alignment of HSD primary structures from thirty different sources was presented. The secondary structure assignments for *S. cerevisae* HSD were used to guide ambiguous regions of the alignment. Several residues are strictly conserved, including those which indicate the dinucleotide binding region, the putative metal binding region, and the active site residues. It is highly likely that the fold of HSD is conserved across all species.

The phylogenetic tree derived from the sequence alignment suggests a sequential, rather than the previously proposed parallel evolutionary model for the three different groups of HSD enzymes. Within the parallel evolutionary model, group 3 HSD enzymes evolved from the group 2 enzymes, which in turn evolved from the group 1 enzymes.

A comparison of primary structures for the group 2 HSD enzymes suggests that α-1F, which bears Lys 117, may be part of a switch region responsible for the L-Thr based regulation observed in group 2 and 3 enzymes.

Both the parallel evolutionary model and the identification of the HSD switch region remain highly speculative at this point.

Chapter 8 References

1. **Tatusov, R. L., Koonin, E. V., and Lipman, D. J.** (1997). A genomic perspective on protein families. Science. **278**: 631-637.
2. **Aravind, L., and Koonin, E. V.** (1999). Gleaning non-trivial structural, functional and evolutionary information about proteins by iterative database searches. J Mol Biol. **287**: 1023-1040.
3. **Benson, D. A., Boguski, M. S., Lipman, D. J., Ostell, J., Ouellette, B. F., Rapp, B. A., and Wheeler, D. L.** (1999). GenBank. Nucleic Acids Res. **27**: 12-17.
4. **Thompson, J. D., Higgins, D. G., and Gibson, T. J.** (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**: 4673-4680.
5. **Richards, F. M.** (1974). The interpretation of protein structures: total volume, group volume distributions and packing density. J Mol Biol. **82**: 1-14.
6. **Lee, B., and Richards, F. M.** (1971). The interpretation of protein structures: estimation of static accessibility. J Mol Biol. **55**: 379-400.
7. **Nayal, M., Hitz, B. C., and Honig, B.** (1999). GRASS: a server for the graphical representation and analysis of structures. Protein Sci. **8**: 676-679.
8. **Archer, J. A., Solow-Cordero, D. E., and Sinskey, A. J.** (1991). A C-terminal deletion in *Corynebacterium glutamicum* homoserine dehydrogenase abolishes allosteric inhibition by L-threonine. Gene. **107**: 53-59.
9. **Thomas, D., Barbey, R., and Surdin-Kerjan, Y.** (1993). Evolutionary relationships between yeast and bacterial homoserine dehydrogenases. FEBS Lett. **323**: 289-293.
10. **Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., Blake, J. A., FitzGerald, L. M., Clayton, R. A., Gocayne, J. D., Kerlavage, A. R., Dougherty, B. A., Tomb, J. F., Adams, M. D., Reich, C. I., Overbeek, R., Kirkness, E. F., Weinstock, K. G., Merrick, J. M., Glodek, A., Scott, J. L., Geoghagen, N. S. M., and Venter, J. C.** (1996). Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii.* Science. **273**: 1058-1073.
11. **Katinka, M., Cossart, P., Sibilli, L., Saint-Girons, I., Chalvignac, M. A., Le Bras, G., Cohen, G. N., and Yaniv, M.** (1980). Nucleotide sequence of the *thrA* gene of *Escherichia coli.* Proc Natl Acad Sci U S A. **77**: 5730-5733.
12. **Zakin, M. M., Duchange, N., Ferrara, P., and Cohen, G. N.** (1983). Nucleotide sequence of the metL gene of *Escherichia coli.*

Its product, the bifunctional aspartokinase II-homoserine dehydrogenase II, and the bifunctional product of the *thrA* gene, aspartokinase I-homoserine dehydrogenase I, derive from a common ancestor. J Biol Chem. **258**: 3028-3031.

13. **Weisemann, J. M., and Matthews, B. F.** (1993). Identification and expression of a cDNA from *Daucus carota* encoding a bifunctional aspartokinase-homoserine dehydrogenase. Plant Mol Biol. **22**: 301-312.
14. **Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A., Merrick, J. M. *et al.*** (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science. **269**: 496-512.
15. **Omori, K., Imai, Y., Suzuki, S., and Komatsubara, S.** (1993). Nucleotide sequence of the *Serratia marcescens* threonine operon and analysis of the threonine operon mutations which alter feedback inhibition of both aspartokinase I and homoserine dehydrogenase I. J Bacteriol. **175**: 785-794.
16. **Kaneko, T., Tanaka, A., Sato, S., Kotani, H., Sazuka, T., Miyajima, N., Sugiura, M., and Tabata, S.** (1995). Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis sp. strain PCC6803.* I. Sequence features in the 1 Mb region from map positions 64% to 92% of the genome. DNA Res. **2**: 153-166, 191-158.
17. **Tomb, J. F., White, O., Kerlavage, A. R., Clayton, R. A., Sutton, G. G., Fleischmann, R. D., Ketchum, K. A., Klenk, H. P., Gill, S., Dougherty, B. A., Nelson, K., Quackenbush, J., Zhou, L., Kirkness, E. F., Peterson, S., Loftus, B., Richardson, D., Dodson, R., Khalak, H. G., Glodek, A., McKenney, K., Fitzegerald, L. M., Lee, N., Adams, M. D., Venter, J. C. *et al.*** (1997). The complete genome sequence of the gastric pathogen *Helicobacter pylori.* Nature. **388**: 539-547.
18. **Motoyama, H., Maki, K., Anazawa, H., Ishino, S., and Teshiba, S.** (1994). Cloning and nucleotide sequences of the homoserine dehydrogenase genes (hom) and the threonine synthase genes (thrC) of the gram-negative obligate methylotroph *Methylobacillus glycogenes.* Appl Environ Microbiol. 60: 111-119.
19. **Clepet, C., Borne, F., Krishnapillai, V., Baird, C., Patte, J. C., and Cami, B.** (1992). Isolation, organization and expression of the *Pseudomonas aeruginosa* threonine genes. Mol Microbiol. **6**: 3109-3119.
20. **Reinscheid, D. J., Eikmanns, B. J., and Sahm, H.** (1991). Analysis of a *Corynebacterium glutamicum hom* gene coding for a

feedback- resistant homoserine dehydrogenase. J Bacteriol. **173**: 3228-3230.

21. **Kunst, F., Ogasawara, N., Moszer, I., Albertini, A. M., Alloni, G., Azevedo, V., Bertero, M. G., Bessieres, P., Bolotin, A., Borchert, S., Borriss, R., Boursier, L., Brans, A., Braun, M., Brignell, S. C., Bron, S., Brouillet, S., Bruschi, C. V., Caldwell, B., Capuano, V., Carter, N. M., Choi, S. K., Codani, J. J., Connerton, I. F., Danchin, A. *et al.*** (1997). The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. Nature. **390**: 249-256.
22. **Madsen, S. M., Albrechtsen, B., Hansen, E. B., and Israelsen, H.** (1996). Cloning and transcriptional analysis of two threonine biosynthetic genes from *Lactococcus lactis MG1614*. J Bacteriol. **178**: 3689-3694.
23. **Felsenstein, J.** (1996). Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. Methods Enzymol. **266**: 418-427.
24. **Dayhoff, M. O., Barker, W. C., and Hunt, L. T.** (1983). Establishing homologies in protein sequences. Methods Enzymol. **91**: 524-545.
25. **Felsenstein, J.** (1988). Phylogenies from molecular sequences: inference and reliability. Annu Rev Genet. **22**: 521-565.
26. **Parsot, C., and Cohen, G. N.** (1988). Cloning and nucleotide sequence of the *Bacillus subtilis hom* gene coding for homoserine dehydrogenase. Structural and evolutionary relationships with *Escherichia coli* aspartokinases-homoserine dehydrogenases I and II. J Biol Chem. **263**: 14654-14660.
27. **Truffa-Bachi, P., Veron, M., and Cohen, G. N.** (1974). Structure, function, and possible origin of a bifunctional allosteric enzyme, *Escherichia coli* aspartokinase I-homoserine dehydrogenase I. CRC Crit Rev Biochem. **2**: 379-415.
28. **[illegible]owitz, N. H.** (1945). On the Evolution of Biochemical Syntheses. PNAS. **31**.
29. **Sprang, S. R.** (1997). G protein mechanisms: insights from structural analysis. Annu Rev Biochem. **66**: 639-678.

9.1 Overall Summary and Achievements of the Thesis Work

The core work of this thesis was the *de novo* structure determination of HSD from *S. cerevisae*. Three different intermediates along the reaction pathway were determined to near atomic resolution. The collective structures provide a snapshot view of the reaction mechanism. The crystal structure permitted the design of several mutant HSD enzymes and the interpretation of their enzymatic behavior. The crystallographic structures together with the kinetic data from the mutants has enabled the proposal of accurate reaction mechanisms. These mechanisms will be useful in design of new HSD inhibitors and thereby active against species reliant upon HSD activity. The comparison of the HSD primary structures has revealed that the tertiary structure of the enzyme is likely well conserved; thus the target species of HSD inhibitors need not be limited to the fungal pathogens discussed in the introduction.

Although technical difficulties prevented the collection of crystallographic data for the HSD•NAD^+•RI-331 complex, the HSD•NADA•L-Hse complex is an excellent approximation. The model of the inhibitor ternary complex based on the substrate ternary complex provides strong support for the formation of a covalent complex between RI-331 and the oxidized nicotinamide cofactor. The modeled complex will enable the atomic structure of RI-331 to be altered rationally so that its

inhibitory activities may be enhanced. Alternatively, RI-331/NAD^+ hybrids may be designed with the confidence that they will be effective inhibitors.

While overcoming the technical difficulties of obtaining interpretable electron density maps, a new method for determining NCS was developed. The method is applicable to any protein crystal data, with only two requirements. The first requirement is that the protein possesses at least one globular region which obeys NCS. The second requirement is that the structure factors have at least low resolution (< 5 Å) phase information, *e.g.* isomorphous replacement data from a minimum of one heavy atom derivative.

While trying to understand the unique nature of HSD, a review of NAD(P) dehydrogenase structural families and the mechanisms for their catalysis was undertaken. A current review of HSD enzymes has also been provided, including a formal basis for the classification of HSD enzymes into three different groups. It is the desire of this author that the information be of use to other researchers looking at similar problems.

9.2 Future Directions

There are several unresolved issues left from the work presented herein. Foremost is experimental evidence for the manner in which RI-331 is bound to the enzyme. The most efficient route to this information will be to soak the RI-331 compound into the tetragonal form crystals. Some care will have to be taken to avoid occupying the apo site with

NAD^+•RI-331, as this is likely to cause crystal cracking. The proper conditions will be found within a very narrow concentration range of RI-331 and NAD^+.

The exact composition of L-ASA as it binds to the enzyme remains to be determined. It would be useful to obtain crystals of the enzyme in complex with the L-ASA compound or analogues thereof. One commercially available analogue is the 4-thio derivative of L-ASA. Obtaining a crystal structure of the enzyme in complex with the L-ASA substrate or substrate analogue would complete the determination of all HSD kinetic intermediates, allowing the reaction to be fully structurally characterized. Despite the large number of enzyme structures in the literature, relatively few have multiple, let alone all of their kinetic intermediates structurally determined. Such an achievement could lead to new insights for enzyme catalyzed reaction mechanisms.

The issue of how the amino acid substrate binds to the enzyme may be pursued by other experimental means. Enzyme kinetic studies of the substrate analogues suggested in chapter 6 would be an excellent way to determine which of the two proposed reaction mechanisms is correct. The elimination of one of the reaction mechanisms will further accelerate the design of new HSD inhibitors.

The issue of metal ion is binding may be solved by simply soaking the tetragonal form crystals in a solution of K^+ ions. X-ray data collected

from K^+ soaked crystals will be extremely informative. The scattering factors between K^+ and H_2O differ enough to provide unambiguous information on the assignment of the metal ion binding site. With luck, it may also be apparent why the K^+ ions augment HSD catalytic activity in the putatively isomorphous *E. coli* HSD enzymes.

The evolutionary speculations presented in the final chapter will be difficult to test. The continued efforts of the various genome sequencing projects will hopefully provide additional support for the models proposed herein.

The switch region proposed for HSD will be similarly difficult to test. Domain swapping experiments with the AK-HSD enzymes in *E. coli* currently being performed by Viola *et al.* at the University of Akron represent the current best efforts on this front. The structure determination of either an intact group 2 HSD enzyme (*e.g.* the entire *E. coli* AK-HSDI) or a group 3 HSD enzyme will probably be the final word on whether the assigned switch region is correct or not.

References for Figure Generation

Molecular graphics, schematics, pictures and diagrams are invaluable to the structural biologist attempting to communicate results. Here is some of the well deserved credit for the software that I used to generate my figures.

O:

Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Cryst. **A47**: 110-119.

Molscript:

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

Bobscript:

Esnouf, R. M. (1999). Further additions to MolScript version 1.4, including reading and contouring of electron-density maps. Acta Crystallogr D Biol Crystallogr. **55**: 938-940.

Raster3D:

a) **Merritt, and Murphy.** (1994). Raster3D Version 2.0 - A Program for Photorealistic Molecular Graphics. Acta Cryst. **D50**: 869-873.

b) **Bacon, and Anderson.** (1988). A Fast Algorithm for Rendering Space-Filling Molecule Pictures. **J. Molec. Graphics**. 6: 219-220.

GRASP:

Nayal, M., Hitz, B. C., and Honig, B. (1999). GRASS: a server for the graphical representation and analysis of structures. Protein Sci. **8**: 676-679.

TOPS:

Westhead, D. R., Slidel, T. W., Flores, T. P., and Thornton, J. M. (1999). Protein structural topology: Automated analysis and diagrammatic representation. Protein Sci. **8**: 897-904.

LigPlot:

Wallace, A. C., Laskowski, R. A., and Thornton, J. M. (1995). LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. Protein Eng. **8**: 127-134.

ALSCRIPT

Barton, G. J. (1993). ALSCRIPT: a tool to format multiple sequence alignments. Protein Eng. **6**: 37-40.

Procheck

Laskowski, R. A., MacArthur, M. W., Moss, J. S., and Thornton, J. M. (1993). Procheck. J. Appl. Cryst. **26**: 283-291.