The University

of Birmingham

Department of Clinical Neuroscience

"Characterisation of a recombinant human cysteine dioxygenase"

Christopher Harper Barry

A thesis submitted for the degree of Doctor of Philosophy

UNIVERSITY OF BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

This thesis is dedicated to no one in particular. I hope that the work contained within will benefit the next researchers to pursue work on this very interesting protein. The University of Birmingham must be proud to have such a beautiful campus with such outgoing and enthusiastic faculty. I will cherish my memories here forever. God bless England.

Contents

1 Introduction **10**

1.1 Pathologies 14

1.1.1 Prion Disease 14

1.1.2 Huntington's Disease 16

1.1.3 Alzheimer's Disease 17

1.1.4 Multiple Sclerosis 18

1.1.5 Amyotrophic Lateral Sclerosis 20

1.1.6 Hallervorden-Spatz Disease 20

1.1.7 Pick's Disease 21

1.1.8 Progressive Supranuclear Palsy 22

1.1.9 Parkinson's Disease 23

1.2 Overview 24

1.2.1 Mitochondrial Dysfunction 24

1.2.2 Redox State 25

1.2.3 Oxidative Stress 26

1.2.4 Glutathione Metabolism 29

1.2.5 Intracellular Cysteine 30

1.2.6 Inclusion Bodies 31

1.2.7 Taurine and Hypotaurine 32

1.2.8 Dioxygenases 33
1.2.9 Cysteine Dioxygenase 42
1.2.10 Summary 46

2 Aims **47**

3 Expression of Cysteine Dioxygenase **50**
3.1 Aims 52
3.2 Protocols 52
3.2.1 Expression System 52
3.2.2 Confirmation of Expression System 54
3.2.3 Expression of CDO 55
3.2.4 Expression of Isotopically Labelled CDO 56
3.2.5 Purification of CDO 57
3.2.6 Confirmation of Expressed Protein I 58
3.2.7 Confirmation of Expressed Protein II 59
3.3 Results and Discussion 62

4 Physical Characterisation of CDO **70**
4.1 Aims 70
4.2 Methods 71
4.2.1 Mass Spectroscopy 71
4.2.2 Surface Plasmon Resonance 72
4.2.3 Amino Acid Decomposition 73
4.2.4 Dynamic Light Scattering 75
4.2.5 Rheology 77
4.3 Results and Discussion 79

4.3.1 Spectroscopic Determination of CDO Produced by the Bacterial Expression System 79
4.3.2 Concentration Determination of CDO using SPR 81
4.3.3 Determination of Amino Acid Composition of CDO produced in Bacterial System 88
4.3.4 Aggregation State of Bacterially Produced CDO using Dynamic Light Scattering 89

5 Structural Analysis **101**
5.1 Aims 101
5.2 Methods 102
5.2.1 Circular Dichroism 102
5.2.2 Protein Crystallography 104
5.2.3 Nuclear Magnetic Resonance 107
5.2.4 Protocol 108
5.3 Results and Discussion 109
5.3.1 Secondary Structure Determined Using Circular Dichroism 109
5.3.2 Functionality Assay Using NMR 124
5.3.3 Crystallography 129
5.3.4 Nuclear Magnetic Resonance 133

6 Conclusions **138**

7 Acknowledgements **142**

A Sequence Specific Details of CDO **144**

Bibliography **190**

Abbreviations

aa	amino acid
bp	base pairs
BSE	Bovine spongiform encephalopathy
^{13}C	Carbon-13
CD	Circular Dichroism
CDO	Cysteine Dioxygenase
CDO^R	Recombinant CDO (CDO_h^R)
CDO_h	Human CDO
CDO_r	rat CDO
CJD	Creutzfeldt-Jakob disease
CNS	Central Nervous System
CSAD	sulfoalanine decarboxylase
cydx	cysteine dioxygenase (SwissProt)
Cys	L-cysteine

Da	Dalton
DAPI	4'-6-diamindine-2-phenyl indole
DIFP	diisopropyl fluorophosphate
DLS	Dynamic Light Scattering
D_2O	Deuterium oxide
DRPLA	dentatorubral pallidoluysian atrophy
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EtOH	Ethanol
FITC	fluorescein-5-isothiocyanate
FFI	fatal familial insomnia
GSS	Gerstmann-Straussler-Scheinker disease
GSSG	Oxidised glutathione
GSH	Reduce glutathione
GuHCl	guanidinium hydrochloride
HD	Huntington's disease
HDO	Deuterium hydrogen oxide
HSQC	Heteronuclear Single-Quantum Coherence
IgG	Immunoglobulin

IPTG	Isopropyl β-D-thiogalactopyranoside
MALDI	Matrix-Assisted Laser Desorption Ionisation
MeOH	Methanol
MS	Mass Spectroscopy
^{15}N	Nitrogen-15
N_A	Avogadro's Number
n	refractive index
n_0	refractive index of pure solvent
NBIA-1	neurodegeneration with brain iron accumulation type-1
NFT	neurofibrillary tangles
NMR	Nuclear Magnetic Resonance
PAGE	poly-acrylamide gel electrophoresis
PCS	Photon Correlation Spectroscopy (DLS)
PD	Parkinson's disease
PMSF	phenylmethylsulfonyl fluoride
PrP^C	cellular prion protein
PrP^{Sc}	pathogenic prion protein
QELS	Quasi-elastic Light Scattering (DLS)
R_h	Hydrodynamic radius

R_s	Radius of a perfect sphere
$\Re$	Reynolds Number
ROS	Reactive Oxygen Species
SDS	Sodium dodecyl sulfate
SOD	Superoxide Dismutase
SPR	Surface Plasmon Resonance
TEMED	N,N,N',N'-tetramethylethylenediamine
TFA	trifluoroacetic acid
TOF	Time of flight
TSE	transmissible spongiform encephalopathies
β_s	Mandelkern-Flory-Scheraga Constant
η	viscosity
η_o	viscosity of solvent
η_c	characteristic viscosity
η_i	intrinsic viscosity
η_r	relative viscosity
ρ	density

Abstract

Cysteine Dioxygenase is an enzyme that catalyses the reaction of cysteine to cysteinesulfinic acid and is thought to regulate the intracellular concentrations of its substrate. The enzyme may also be involved in the oxidation of exposed sulfidrils of proteins and by implication, involved in cell signalling or regulation. Substrate and products of cysteine dioxygenase are known to be disrupted in a number of neurodegenerative diseases making the protein an important target for medical research. To date, the structure of the protein has not been solved. An understanding of its structure and reaction mechanism will further the understanding of the aetiology of a number of different pathologies. Its structure may also further the development of novel pharmacological drugs.

Chapter 1

Introduction

Cysteine dioxygenase (CDO) is an enzyme that catalyses the reaction of L-cysteine to 3-sulfino-L-alanine (L-Cysteinesulfinic acid). The reaction requires Fe^{2+} and NAD^{+} or $NADP^{+}$as cofactors. The enzyme was originally named cysteine oxidase until it was establised that it uses molecular oxygen as a co-substrate[1]. For this reason, it has been given an enzyme clasification of EC 1.13.11.20 (1-Oxidoreductase, 13-Acting on single donors with incorporation of molecular oxygen, 11-With incorporation of two atoms of oxygen, 20-cysteine dioxygenase).

The precise physiological role of this enzyme has not been establised. It seems to be present in all animals, exists as a single gene copy, and shows an extraordinary sequence homology/identity across organisms. In man, the gene is cytogenetically located on the long arm of chromosome 5, in the region of 22-23 (5q22-q23). The human genome project has since precisely located the gene between 119919360 - 119931270 bp. The mRNA is 1549 bases in length. The gene has 5 exons of 12kb of chromosome 5 and codes for a protein 200 aa residues long[2].

Biochemically, CDO is believed to be the principle regulator of free intra-

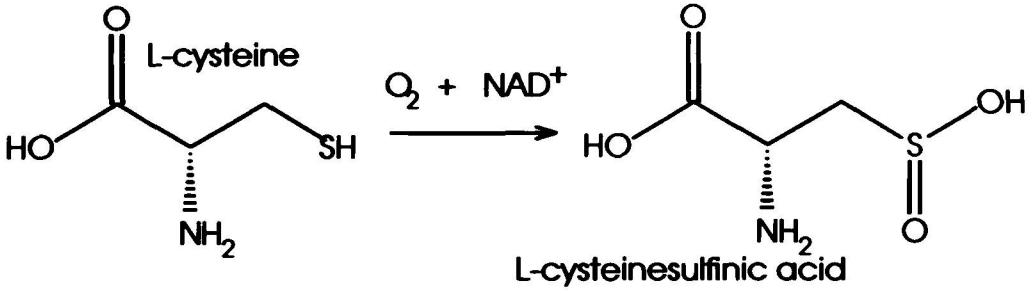


Figure 1.1: Reaction catalyzed by cysteine dioxygenase

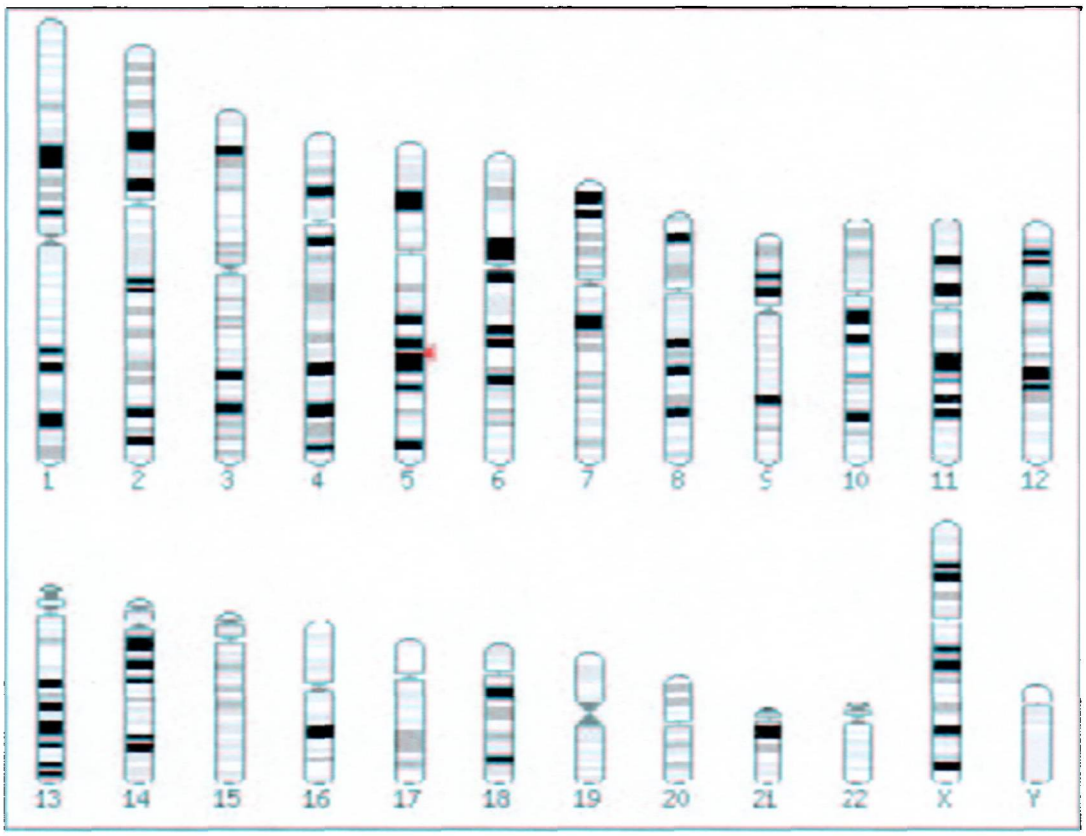


Figure 1.2: Sanger Centre genome map with locus of human CDO gene

cellular L-cysteine by converting it to 3-sulfino-L-alanine, which is ultimately converted to hypotaurine and taurine. As well as being an important constituent of proteins, L-cysteine is also a NMDA agonist and an excitotoxin[3]. Cysteine is also a precursor to glutathione and other anti-oxidants making CDO a tantalising target for pharmacophore development and to study its neuro-physiological role.

A significant amount of research is currently devoted to elucidating the aetiological basis of diseases. Epidemiological and genetic investigations have yielded many exogenous factors and loci as the potential causes of many pathologies. In nearly every neuropathy, differences in the severity, age of onset, gestation period, etc. indicate that there are many non-genetic factors involved in pathogenesis. Developmental genetics and cancer research have shown that truncations of a protein may lead to a loss of function or cause the protein to become constitutively active. Molecular biophysics and prion research are less clear on the functionality of misfolded proteins, but the consequences are equally deleterious. Investigation of enzyme systems that potentially create intracellular conditions permissive for protein denaturation, misfolding, and aggregation is important not only for prion disease, but also for all neurodegenerative diseases.

Many neurodegenerative diseases are characterised by mitochondrial dysfunction, oxidative stress, alterations in the redox state, protein aggregation, and ultimately apoptosis or necrosis. There is no concensus on which aetiological features are causative or consequential. Cysteine dioxygenase is a suspect gene product in many neurodegenerative diseases. Its precise neurophysiological function has never been established primarily because no appropriate model system exists. However, the gene product is likely to be the primary regulator of intracellular L-cysteine. Its substrate, L-cysteine,

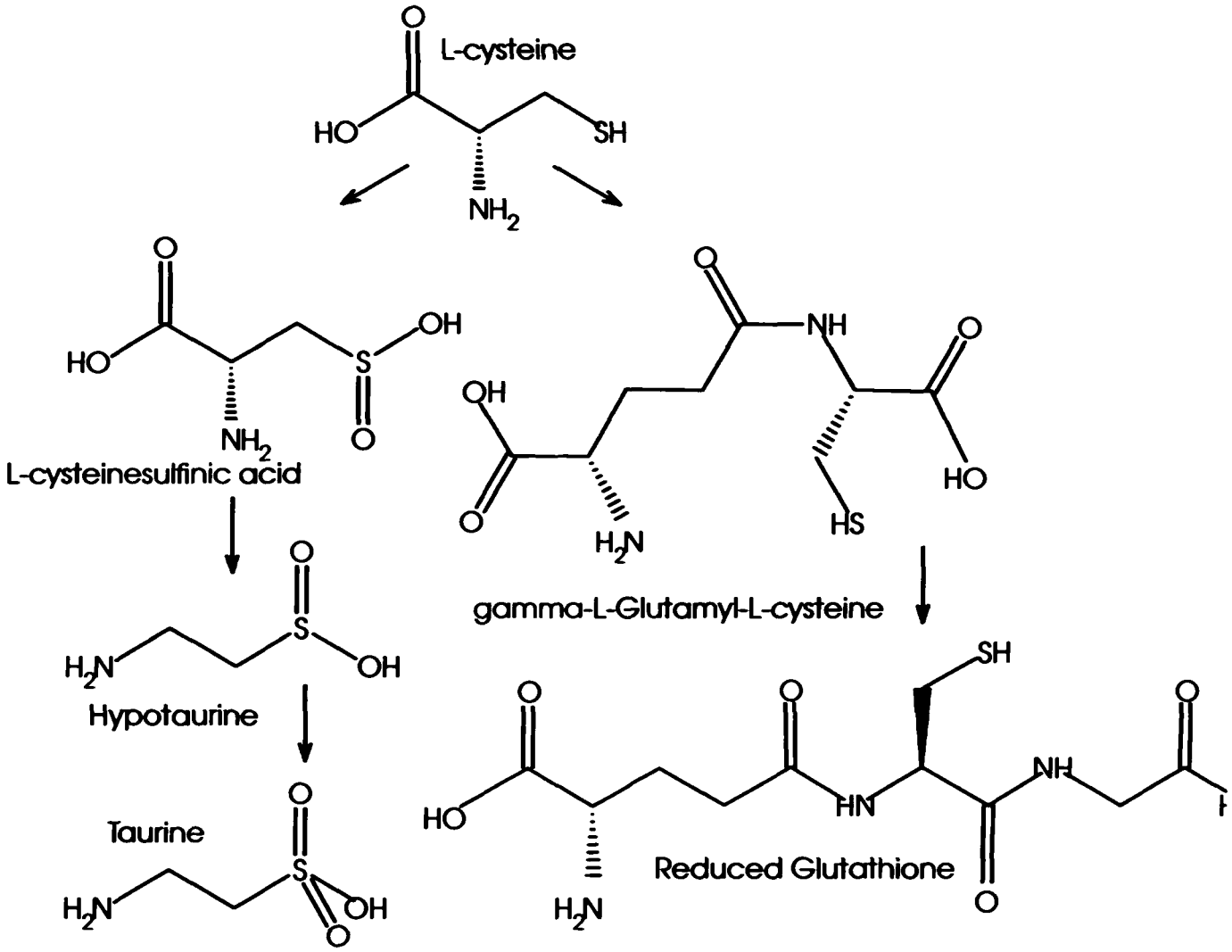


Figure 1.3: End-products of L-cysteine involved in the anti-oxidant defence.

is a neurotoxin as well as being a precursor substrate to the neuronal anti-oxidant defence, glutathione. In addition, two of its end-products, hypotaurine and taurine, are molecules that have a roles in neuromodulation and in anti-oxidant defence. In order to understand what physiological relevance cysteine dioxygenase may have with respect to cell survival, a brief summary of some pathologies and few regulatory systems are discussed.

1.1 Pathologies

Different neuro-pathologies may show very different clinical symptoms. Many have been identified as having a genetic basis or have been shown to be caused by exogenous factors. Most, if not all, of these different diseases have a commonality in the intracellular environment of their respective cell types. The cell groups affected dictate the outward clinical symptoms of the disease. A brief review of a few neurodegenerative diseases will highlight factors involved in pathogenesis and the role CDO may play within the cell.

1.1.1 Prion Disease

There are several human prion diseases: Kuru, fatal familial insomnia (FFI), Gerstmann-Straussler-Scheinker disease (GSS), and Creutzfeldt-Jakob disease (CJD) are the most documented. Prion diseases differ from other pathologies in that the diseases are not caused by virus, bacteria or other conventionally accepted pathogen; they are caused by a otherwise normal protein (PrP^C) that undergoes a conformational change that catalyses the unfolding of the rest of the the PrP^C into a pathogenic form (PrP^{Sc}). CJD, FFI, GSS, and Kuru are all part of the same class of transmissible spongiform encephalopathies (TSEs). Prion diseases are progressively neurodegenerative and are invariably fatal.

Initial clinical symptoms of CJD include fatigue, vertigo, anxiety, impaired memory, weakness, and impaired motor function[4]. As the disease progresses, gross impairments of higher order brain function and motor function ensue[4]. There is tremendous neuronal loss in CJD. The cerebral cortex is the most affected and the deeper lamellae of the frontal and temporal lobes are particularly prone[4]. The striatum (caudate and putamen), thala-

mus, midbrain, and the superficial layers of the cerebellar cortex also show these changes[4]. In the final stages of the disease, victims' mental abilities completely deteriorate and they become vegetative[4]. CJD can occur spontaneously but a new variant epidemic (nvCJD) has attracted international attention in recent years. It is believed that scrapie, a sheep TSE, was passed across species to bovine (becoming bovine spongiform encephalopathy, BSE) by feeding cattle a "meat and bone meal" (MBM) containing scrapie infected sheep remains. Bovine spongiform encephalopathy (BSE) was then transmitted to man (nvCJD). The epidemic was thought to have spread further by the practice of feeding MBM, this time containing BSE infected meat. Kuru is thought to have been CJD and the epidemic a century ago is believed to have originated from cannibalistic practices upon a spontaneous CJD victim[1]. Gestation periods may last several years or decades but once the disease begins manifesting itself clinically, the course of neurodegeneration is rapid[8].

PrP aggregation is a common feature of all prion disease, is apparent histologically, and amyloid plaques present are putative sites harbouring PrP^{Sc}[9]. In CJD, PrP^{C} is a regulator of Cu^{++}/Zn^{++}-superoxide dismutase. It has been shown that the conversion of PrP^{C} to PrP^{Sc} is sensitive to the redox state of the cell and the inability to mediate proper PrP^{C} disulfide bonds[10]. The toxic effects of PrP^{Sc} have been attenuated *in vitro* with drugs that raise intracellular levels of glutathione[11]. A lowered glutathione reductase is reported in a PrP^{C} knockout cell line[12]. A dysfunctional glutathione reductase may indicate a change in the ratio of reduced to oxidised glutathione and to generalised protein folding problems (discussed later). There is widespread labelling for

[1]Kuru and CJD have been known to virulent diseases[5]. The actual discovery of a virus has never been published and it was not until the discovery of the prion[6] that researchers realised a protein was the virulent agent. It has since become accepted that Kuru is caused by the virulent form of the prion protein[7].

markers of oxidative stress in prion diseases and is a proposed cause of neuronal loss[13]. Impaired activities of mitochondrial Mn-superoxide dismutase (SOD), ATPase and cytochrome c in animal prion models suggest mitochondrial dysfunction[14]. Inclusion bodies in prion disease are ubiquitinated[15].

1.1.2 Huntington's Disease

Huntington's disease (HD) is an autosomal dominant disorder. The disease usually progresses from slight impaired motor function to progressive dementia over a course of 15-20 years[4]. There are reductions in the mass of the cerebral cortex, white matter, thalamus and particularly the caudate and putamen[4]. There is progressive degeneration of the corpus striatum and the globus pallidus, hypothalamus, dentate nucleus, brain stem, and Purkinje neurons of the cerebellum may also be affected[4]. Neurofibrillary tangles reminiscent of Alzheimer's disease are sometimes present[4]. Interestingly, a co-worker in this laboratory has shown that CDO is strongly expressed in Purkinje neurons[16].

Huntington's disease is one of a class of $(CAG)_n/Q_n$-expansion disorders that code for extended glutamine repeats[17]. In HD, the glutamine repeat is found on the huntingtin protein (349 kDa). The disease will not manifest itself until a threshold of repeats is reached; with increasing repeats there is a decrease in the age of onset and an increase in the severity of the disease. Other $(CAG)_n/Q_n$-expansion disorders include dentatorubral pallidoluysian atrophy (DRPLA or Smith disease), atrophin, and spinobulbar muscular atrophy (Kennedy disease). Hallmarks of HD and other $(CAG)_n/Q_n$-expansion disorders are aggregates of protein in the nuclei. Inclusion bodies in HD and $(CAG)_n/Q_n$-expansion disorders are ubiquitinated[15]. Although the ge-

netic evidence in $(CAG)_n/Q_n$-expansion disorders is compelling, epidemiological findings amongst twins indicate that environmental differences can affect the influence of genetic predisposition[18]. Oxidative stress and severe mitochondrial impairments of complex II and complex III have been reported in HD with less severe impairments of complex IV[19] (Figure 1.4). There is an accumulation of iron in the basal ganglia in HD[20]. Interestingly, a 50% reduction of oxidised glutathione has been reported in HD[21]. A change in in the ratio of oxidised to reduced glutathione may be of importance for proper protein structure, function, and inclusion body formation.

1.1.3 Alzheimer's Disease

Alzheimer's disease (AD) is a progressive neurodegenerative disease characterised by dramatic cognitive decline, neurodegeneration and characteristic neurofibrillary tangles (NFT). Several genetic mutations have been identified which can predispose an individual to AD, however, many cases are idiopathic.

Histologically, AD has characteristic inclusion bodies in affected areas called senile plaques. As with other diseases, neurofibrillary tangles and plaques are aggregates of insoluble, protease-resistant and misfolded proteins. Several proteins have been identified as constituents of these inclusions: $\beta - amyloid$ (A beta), α1-antichymotrypsin, amyloid precursor protein (APP), presenilin-1(PS1), neurosin, butyrylcholinesterase, synelfin (synuclein, NACP), tissue factor pathway inhibitor-1 (TFPI-1), tissue inhibitor of metalloproteinase (TIMP), tau, heparin sulfate glycosaminoglycan, and many more are likely to be discovered[22, 23, 24, 25, 26, 27, 28, 29, 30]. Mutations in some of these proteins have been shown to cause familial AD; in idio-

pathic cases, the number of proteins identified in proteinaceous inclusions and knowledge of prion proteins as infections agents may indicate that initiation of the disease is due to a more general problem of proper folding.

Research indicates that there is an alteration in the ratio of reduced to oxidised glutathione in the brains of AD patients[31]. Populations of neurons within the brain that are prone to develop NFT's are also part of a subset of neurons more immunoreactive to taurine[32]. Cysteine levels are elevated in AD[33] and an increased cysteine to sulphate ratio in AD links a defect in sulphur metabolism to the two previous observations[34, 35]. Oxidative stress and apoptosis in AD are well documented. There are several *in situ* experiments that have been published linking various transition metals with the generation of reactive oxygen species and solubility problems with amyloid deposits; aluminium is also a suspect metal. Brain magnesium levels have been found to be lowered in AD compared to controls[36, 37]. Experiments have demonstrated that amyloid plaques are sites of free radical generation and oxidative stress; however β-amyloid peptides will not spontaneously generate free radicals without the presence of a transition metal[38].

Epidemiological studies indicate that aluminium exposure is a risk factor for developing AD[39]. Magnesium deficiency may also be a risk factor.

1.1.4 Multiple Sclerosis

Multiple sclerosis (MS) accounting for about one per 1000 deaths around the world is a leading cause of disability amongst young people[4]. Generally it is recognised to affect the spinal neurons and accounts for the disability of movement. Clinically it is characterised by apathy, depression, fatigue weight loss, muscle pains, weakness, numbness, paresthesia, and varying vision

problems[4]. It is a relapsing disease but as the disease takes its course it is invariably chronic and degenerative[4]. The disease also affects the brain and often results in vertigo, incontinence, and other problems associated with brain stem function[4].

Brain sections in MS often reveal plaques[4]. A commonly involved area in MS is the region between the caudate nucleus and the corpus callosum, the striatum, pallidum and the thalamus[4]. There has not been a satisfactory amount of work done investigating MS plaques histologically but inclusion bodies, "pole bodies", have recently been identified[40]. Both reduced and oxidised glutathione levels are elevated in MS[41]. It has been reported that glutathione reductase activity is increased and glutathione peroxidase activity is decreased[42]. Some evidence exists that there is a decrease in the cerebral thiol pool[43]. There are obvious implications of these reports for oxidative stress; they may also have implications for improper protein folding. At least one enzyme, α-2-macroglobulin, that has been reported in MS with an aberrant structure[44]. The presence of inclusion bodies indicates that there are many others. Although the plaques have not been well characterised immunohistochemically; one paper reports that λ-amyloid is a constituent[45]. Iron deposits have been detected in brain tissue in MS and likely contribute to oxidative stress[46].

Initial epidemiological studies have suggested that mercury[47] is a risk factor. There are varying reports linking transition metals as either being protective or risk factors based on soil composition. A preponderance of epidemiological papers implicate a viral risk factor.

1.1.5 Amyotrophic Lateral Sclerosis

Amyotrophic lateral sclerosis (ALS) or motor neuron disease (MND) is a neurodegenerative disorder that affects both the upper and lower motor neuron systems[4]. Cerebral regions are often normal but in protracted cases there is degeneration of the precentral gyrus[4]. Neuronal loss and gliosis are also present in the thalamus, striatum, globus pallidus, subthalamic nucleus, and substantia nigra[4]. Occasionally, inclusion bodies called "Bunina bodies" are found in affected neurons and are immunoreactive to cystatin C, an inhibitor of cysteine proteases[4, 48]. Hyalinated intracellular inclusions are common in affected neurons and are immunoreactive to ubiquitin and neurofilaments[4]. There is a unusually high incidence of ALS in the south Pacific indicating exogenous risk factors[4].

It has been reported that the cerebro spinal fluid in ALS showed an elevated reduced to oxidised glutathione ratio[49]. Alterations in the cysteine to sulphate ratio in ALS have been reported[35]. Prolonged survival in animal models of familial ALS receiving N-acetyl-L-cysteine, a non-toxic pro-cysteine compound, has been documented[50]. Elevated taurine levels have been reported in cervical spinal regions in ALS[51] and in the brain[52].

1.1.6 Hallervorden-Spatz Disease

Hallervorden-Spatz Disease (HSD) or also "neurodegeneration with brain iron accumulation type-1" (NBIA-1) is characterised histologically by the accumulation of iron, protein aggregation, and progressive neurodegeneration[53]. HSD is a very rare disease that starts to manifest itself clinically in late childhood[4]. As the disease progresses, victims progressively lose their ability to articulate themselves. Dementia and dystonia ensue[4]. The accumulation

of iron is visually apparent upon postmortem examination with staining of the globus pallidus and substantia nigra[4]. The globus pallidus and substantia nigra also show neuron loss, gliosis, and inclusion bodies (most also test positive for iron)[4].

Inclusion bodies are immunoreactive to α-synuclein, tau, and ubiquitin[54]. It has been reported that cysteine dioxygenase activity is decreased and there are accumulations of cysteine and cysteine-glutathione mixed disulfide in HSD[55]. An enzyme blockage in cysteine metabolism and the ability of cysteine to chelate metals has been a proposed mechanism for the disorder[4]. While iron accumulation is apparent in HSD, iron metabolism elsewhere in the body appears to be normal[56]. In addition to iron as an aetiological problem, copper and zinc levels are also elevated in HSD[57].

1.1.7 Pick's Disease

Pick's Disease is characterised by middle-age onset dementia with concurrent lobar cerebral atrophy[4]. It distinguishes itself from AD with morphologically distinct inclusion bodies, Pick's bodies, as opposed to the neurofibrillary tangles associated with AD[4]. Pick's bodies are comparable in size to the cell nucleus and appear to be composed of differing linear fibrils[4].

Proteins identified in the inclusion bodies, "Pick bodies", in Pick's disease include ubiquitin[4], tau, neurofilament[58], α-synuclein[59], heme oxygenase-1 (HO-1)[60] and amyloid[61]. Significant accumulations of iron and manganese have been detected in brain tissue[62]. Oxidative stress is believed to be an important factor in the neural demise in Pick's disease with the presence of HO-1, an indicator of oxidative stress[60].

1.1.8 Progressive Supranuclear Palsy

Steele-Richardson-Olsewski disease or Progressive Supranuclear Palsy (PSP) is a late onset disease that is accompanied with parkinsonism and a characteristic loss of the ability to gaze downward[4]. The subcortical region of the brain is mostly affected showing signs of neuronal loss, gliosis, but other regions of the brain and spinal cord may also be involved[4]. The subthalamic nuclei, substantia nigra, and superior colliculi are most affected, together with the corpus striatum, globus pallidus, thalamus, red nucleus, brain stem tegmentum, and dentate nuclei also showing signs of the disease[4]. There is striatal dopamine depletion and depigmentation of the substantia nigra[4]. There is a reduction of dopamine receptors in PSP which accounts for the non-responsiveness of levodopa therapy and distinguishes itself from Parkinson's disease[4].

A 50% reduction of reduced glutathione is observed in the caudate nucleus in PSP[21]. Glutathione is markedly increased in the substantia nigra distinguishing itself again from PD[63]. Taurine levels are markedly increased in the nucleus accumbens, substantia nigra, and globus pallidus[63]. Cellular concentrations of γ-aminobutyric acid, glutamic acid, and glycerophosphoethanolamine levels are also impaired in various regions of the brain compared with controls[63]. Inclusion bodies are present in the brain and spinal cord that are immunoreactive to tau and microtubule associated protein but are not ubiquitinated[4]. Oxidative stress and mitochondrial disfunction have been reported in both clinical and in *in situ* models of PSP.[64, 65].

Epidemiological studies indicate there may be a link between some Caribbean plant toxins and developing PSP[66]. Exhaustive efforts to determine whether or not there are other risk factors have been explored without finding any

links[4].

1.1.9 Parkinson's Disease

Parkinson's disease (PD) is a progressive neurodegenerative disorder that is also present in a number of different neuropathies affecting the central nervous system[67]. Clinical manifestations of the disease include poverty of movement, tremor, a characteristic shuffle step, and loss of balance[4]. As the disease progresses, patients may become susceptible to infection, confined to bed, and dementia may develop[4]. Clinical manifestations of parkinsonism may be primarily degenerative, encompassing features of diseases like HSD and PSP, or secondary to diseases like CJD[4]. The disease usually manifests itself later in life and is often fatal ten years after the onset of clinical symptoms[4].

Depletion of the pigmented dopaminergic neurons is a key feature of PD. Areas of the brain affected include the brain stem, zona compacta of the substantia nigra, locus corelleus, reticularis, dorsal nucleus of the vagus, dorsal raphe nucleus, substantia insomata, hypothalamus, and globus pallidus[4]. In addition to neuronal loss and gliosis, inclusion bodies are present within cells in affected areas. Intracellular inclusions known as Lewy bodies may be present in many areas of the CNS[4].

Cysteine levels are elevated in PD[33]. There is a depletion of glutathione in the substantia nigra and the ratio between reduced and oxidised glutathione is depressed[68, 21]. An elevated cysteine to sulphate ratio in PD links a defect in sulfur metabolism to the two previous observations[34, 35]. Cerebral spinal fluid from PD patients show decreased levels of taurine[69]. Inclusion bodies in PD are ubiquitinated, immunoreactive for tau, α-synuclein, amyloid[70],

and have been shown to contain iron[71]. Iron is also a component of the neuromelanin found in the substantia nigra[72].

Epidemiological reports have linked a higher probability of developing PD after prolonged exposure to lead sulfate[73], manganese, copper, lead, iron, mercury, zinc, and aluminium[74]. There is a lower incidence of PD amongst smokers[75]. Exogenous toxins such as MPTP (MPP^+)[76] and cycad plant toxins in Guam[77] and some Caribbean fruits[66] have been demonstrated to cause Parkinsonism. There is abundant literature suggesting that excitotoxcisity (particularly NMDA mediated) is a major contributor to PD neurodegeneration. L-cysteine excitotoxcisity is mediated via the NMDA receptor. It is interesting that magnesium deficiencies may contribute to the pathogenesis of PD[78]; Magnesium is required to block the NMDA receptor, it attenuates the effects of oxidative stress induced by the Fenton reaction[2] *in vivo*[79], and it quenches the Fenton reaction *in situ*[80, 81].

1.2 Overview

1.2.1 Mitochondrial Dysfunction

Mitochondria are responsible for meeting the energy requirements of the cell. This process is established by creating a redox potential across mitochondrial membranes. The process of respiration leads to the generation of ROS and healthy cells have many defences to prevent ROS damage. Mitochondrial dysfunction has been described in neurodegenerative disease. As an organelle with a symbiotic relationship with the rest of the cell, any biochemical change in either system will have an influence on the homoeostasis of the other. In-

[2]The details of the Fenton reaction are described in the section "Oxidative Stress", page 26.

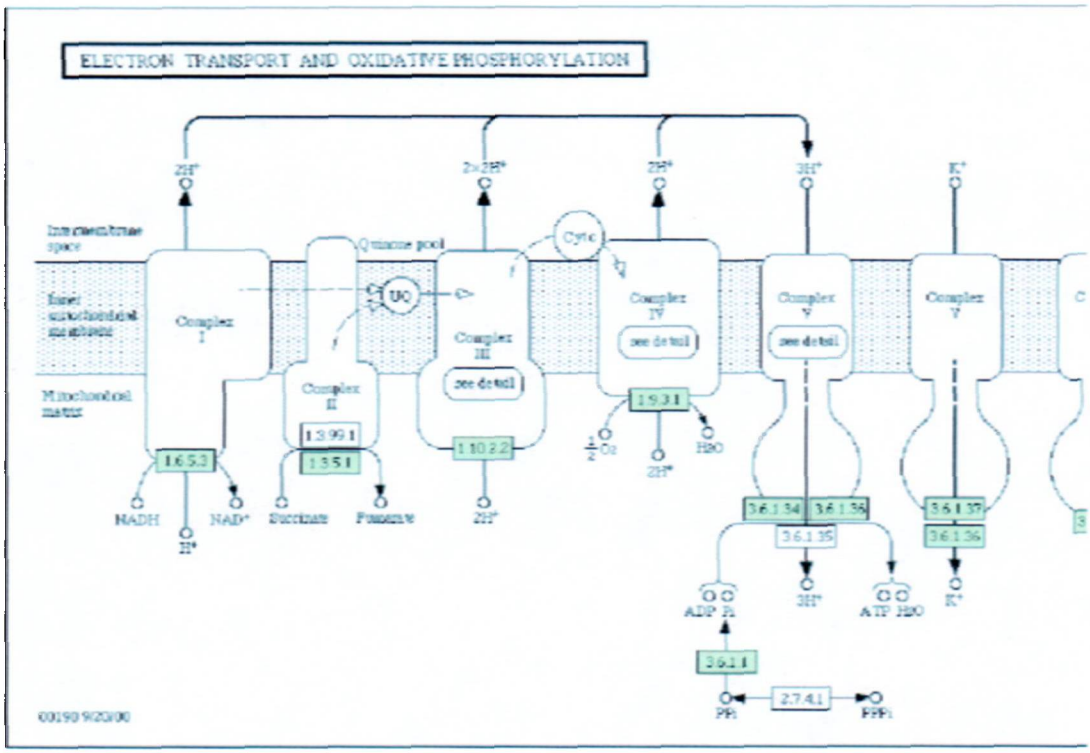


Figure 1.4: Common mitochondrial sites documented a being impaired in different neurodegenerative pathologies. (Image taken from the KEGG database)

hibiting any step in mitochondrial respiration will have a cascade effect that effects cellular energy requirements, change chemical equilibria, biosynthesis, etc.

1.2.2 Redox State

The redox state of a cell is difficult define because of the large number of different process that contribute to the intracellular environment. For simplicity, the redox state could be represented as the ratio of a reduced to oxidised marker, glutathione. The ratio of reduced glutathione is tightly regulated by several different glutathione oxidases, reductases and peroxidases. For the most part, glutathione levels are maintained by a complicated processes that

recycle the substrate. The glutathione redox state then confers its redox status to countless other biochemical systems within the cell.

The redox ratio is not the only important redox parameter; concentration of redox active compounds is equally critical. Biochemically, the redox state is important for maintaining enzymatic reaction kinetics. Physically, the redox state dictates how chemical bonds are formed and this in turn can affect the structural integrity of proteins themselves. The ability to scavenge radicals is also influenced by the redox state.

Although the exact mechanism of paracetamol toxicity is not clear, the toxic effects are related to a depletion of glutathione[82]. Cysteamine and N-acetylcysteine are two prescribed antidotes for acute poisoning. Both these compounds can be converted to glutathione by first being converted to L-cysteine. Implicit with this observation is that cysteine dioxygenase is a redox regulator.

1.2.3 Oxidative Stress

Oxidative stress refers to the generation of reactive oxygen species (ROS) that can react with and damage proteins, lipids, DNA, and RNA. The main reactive species include superoxide radical ($O_2^{\bullet -}$), hydrogen peroxide (H_2O_2), hydroxyl free radical ($HO^{\bullet}$), singlet oxygen, hypochlorous acid, peroxynitrite ($ONOO^-$), and nitric acid. All ROS are generated from oxygen within the cell. Mitochondria are the principal consumers of cellular oxygen and are responsible for generating the majority of ROS's.

A few enzymes are present to transmute ROS into more benign oxygen species. Catalase reduces peroxides (hydrogen and organic) into water and

oxygen.

$$H_2O_2 + H_2O_2 \rightarrow 2H_2O + O_2$$

An extracellular, cytosolic and a mitochondrial superoxide dismutase (SOD) catalyses the conversion of superoxide into hydrogen peroxide and oxygen.

$$O_2^{\bullet -} + O_2^{\bullet -} + 2H^+ \rightarrow 2H_2O_2 + O_2$$

Glutathione peroxidase plays a major role reducing ROS by catalysing peroxide to water; oxidised glutathione is then reduced by NADPH via glutathione reductase[83].

$$2Glu_R + H_2O_2 \rightarrow Glu_{Ox} + 2H_2O$$
$$2Glu_R \leftarrow Glu_{Ox}$$

All of these enzymes require a metal cofactor to coordinate the reactions. SOD requires a zinc, copper, iron or manganese ion depending on the particular isoform. Catalase requires a haem group and manganese. Glutathione peroxidase requires selenium.

Mitochondrial ubisemiquinone and complex I have been recognised as potential sources of superoxide formation via electron leakage during respiration[83]. Several other routes of ROS generation have been proposed to explain oxidative stress observed in pathological conditions such as the Fenton reaction or the analogous Haber-Weiss reaction. The more general Fenton reaction explains how a transition metal can convert superoxide and peroxide (also

hypochlorous acid) into the less reactive but more toxic hydroxyl radical.

$$O_2^{\bullet -} + H_2O_2 \rightarrow HO^{\bullet} + HO^{-} + O_2$$

The hydroxyl radical's toxcisity is due to its longevity and it is one of the most long-lived reactive radicals known. Because there is some disagreement over the mechanism of this reaction, the Haber-Weiss reaction shows the half reactions and illustrate the required Fe^{2+} and Fe^{3+} cycling:

$$O_2^{\bullet -} + Fe^{3+} \rightarrow O_2 + Fe^{2+}$$
$$H_2O_2 + Fe^{2+} \rightarrow HO^{\bullet} + OH^{-} + Fe^{3+}$$

Under anaerobic conditions, the Fenton reaction is also capable of oxidising hydroxyl groups on carbon to ketone groups. Metal chelators may not quench the Fenton reaction. Iron complexed with EDTA has been shown to be 100 fold more catalytic. In biological systems, transition metals are tightly regulated but dysregulation can lead to more rapid catalysis than the transition metal ion alone[84]. *In situ* experiments have demonstrated the effectiveness of many transition metals catalysing the Fenton reaction and damaging DNA[85] and although aluminium is not a transition metal, it has been shown to coordinate with the reaction making iron more catalytic[86].

Chemical factors that influence the Fenton reaction have also been shown to influence pathogenesis. There are papers indicating magnesium as having a protective effect: clinical, epidemiological, and chemical. Two papers have demonstrated magnesium ion as stabilising iron in the Fe^{2+} state[80, 81]. If the chemistry works similarly under physiological conditions, it may prevent

the conversion of $O_2^{\bullet -}$ to the more damaging $HO^{\bullet}$ and prevent the required cycling of Fe^{2+} back to Fe^{3+}. If true, it may indicate $O_2^{\bullet -}$ as the principal ROS generated under pathological conditions.

The superoxide radical is also capable of reacting with nitric oxide to form peroxynitrite, a thiol reactive oxidant that stimulates mitochondrial calcium release[87].

Transition metals are often implicated in the aetiologies of many diseases. It has been shown that thiols such as dithiothreitol (DTT), L-cysteine, glutathione, cysteamine, β-mercaptoethanol, etc. can become toxic by reducing transition metals obviating the need for superoxide in Fenton reactions[88]; this may have ramifications for exposed thiol groups on misfolded proteins that would otherwise be buried within the tertiary structure. Guanosine triphosphate (GTP) also enhances the production of hydroxyl radical formation in the Fenton reaction[89]. Acute toxic exposures to transition metals have also been shown to cause neurodegenerative diseases.

1.2.4 Glutathione Metabolism

Glutathione is a biological radical scavenger and a reductant. Oxidised glutathione (GSSG) is readily converted back to the reduced form (GSH) via a NAD(P)H dependent glutathione reductase (EC 1.6.4.2). GSH is synthesised in a two step reaction by the ATP driven $\gamma - glutamylcysteine$ synthase (EC 6.3.2.2) and ATP driven glutathione synthase (EC 6.3.2.3) (Figure 1.5).

The ratio of GSH/GSSG can affect protein folding. As a nascent peptide is being formed, disulfide bridges are often formed that stabilise adjacent peptide chains. If the environment is too reducing, these disulfide bonds are prevented from forming. If the local environment is too oxidising, disul-

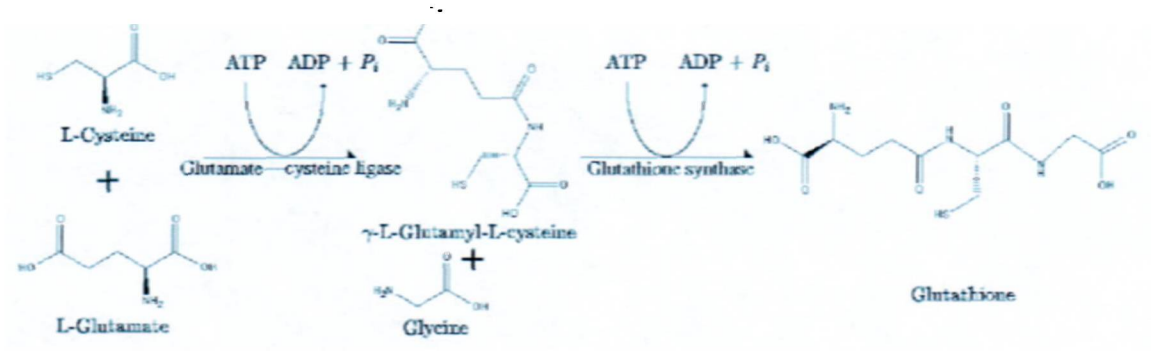


Figure 1.5: Two step process for biosynthesis of glutathione from L-cysteine.

fide bridges will form locking the conformation of the protein. Experiments *in situ* have shown that in order for a protein to fold into its energetically favourable functional tertiary structure, it is necessary for an equilibrium to exist between reduced and oxidised disulfide bonds until the final energetically favourable conformation is reached[90, 91].

Glutathione synthase is the rate limiting enzyme in the biosynthesis of GSH. Overall synthesis is primarily limited by the availability of substrates, namely L-cysteine[92].

1.2.5 Intracellular Cysteine

L-Cysteine is a thiol-containing amino acid whose unique chemistry is vital for protein synthesis, is a required substrate for the biosynthesis of many cellular compounds in many critical pathways, and is also a toxin. L-Cysteine and methionine are two sulphur containing amino acids that play a critical role in coordinating metal ions in redox reactions carried out in enzymes.

Although a critical requirement for the structural integrity of proteins, L-cysteine is also a neurotoxin. For this reason the cell keeps its relative concentration low. It is also a rate limiting substrate in the biosynthesis of

glutathione. L-cysteine is also a regulator of the expression of CDO an γ-glutamylcysteine synthase[93].

1.2.6 Inclusion Bodies

Inclusion bodies are aggregates of misfolded and protease resistant proteins. They may be referred to as Lewy Bodies (LB), senile plaques, amyloid plaques, etc. in histology. They are the result of proteins aggregating together due to misfolding, over-expression, a conformational change that preferentially binds to other proteins, damage, etc. Inclusion bodies are often highly ubiquitinated, a tag for cellular destruction, which indicates that the cell recognises the state of the protein as being aberrant. Controversy exists as to whether the inclusion bodies are causative or a resulting final step in many diseases. They are present in some brains with no apparent pathological symptoms. In others, they are usually detectable long after clinical symptoms become apparent so it has often considered a late event in pathogenesis.

It is not uncommon for recombinantly expressed proteins to form inclusion bodies in host systems. Depending on the host system, a lack of chaperonins or other accessory proteins in the translation process is one proposed explanation for a yield of misfunctional/misfolded aggregated protein. In eukaryotic systems that produce inclusions, an inordinately high level of translation may exhaust the supply of chaperonins, etc. Another likely cause that has been established is the redox state of the cell. Prokaryotic expression systems are more reducing than eukaryotic cells and can prevent disulfide bond formation because of the higher level of reduced glutathione[90]. In order to recover functional protein, procedures have been developed to re-fold the peptide chain. *In situ* "protein folding" work often requires re-suspending the

protein in a solution with a ratio between reduced and oxidised forms of glutathione (or other reductant as β-mercaptoethanol) [90, 94].

Factors such as chaperonins and glycosylation state influence protein folding. There is, however, very little evidence linking chaperonin problems with any neurodegenerative disease. Problems with glycosylation are often attributed to problems with oxidative stress and the redox state of the cell. In nearly every neurodegenerative disease (if not every), there are disruptions of glutathione levels or oxidised to reduced ratios. It is known from *in situ* work with proteins that concentrations of reductant and ratios between reduced/oxidised forms are critical for refolding recombinant protein into functional products. Disruptions in nominal levels of glutathione, precursors, products, and other compounds linked to sulphur metabolism give some insightful aetiological clues because they are all linked by chemical equilibria.

Inclusion bodies and denatured proteins have often been identified as sites of oxidative stress. Various redox active metals have been associated with inclusions and denatured proteins. Research into Fenton or Haber-Weiss chemistry has shown that the reactions are often enhanced above what the metal ions are capable of on their own.

1.2.7 Taurine and Hypotaurine

Taurine is oxidised from hypotaurine by NAD^+ dependant hypotaurine dehydrogenase (EC 1.8.1.3) or from L-Cysteate[3] by sulphoalanine decarboxylase (cysteine-sulphinate decarboxylase)(EC 4.1.1.29). Although there are two pathways, most biosynthesis in the brain is thought to proceed via hypotaurine biosynthesis[96]. Taurine has been assigned several physiologi-

[3]MDCK/EAAT2 have been shown to produce L-Cysteate, a naturally occurring sulfonic acid analog of L-glutamate.[95]

cal properties. Taurine preserves gap junctional intercellular communication in liver hepatocytes subjected to a H_2O_2 stress that normally impairs their function[97]. It can act in an uncompetitive manner on mitochondria for the uptake and sequestering of Ca^{2+}[98]. It has been shown to attenuate MPP^+toxicity[99]. It is believed that taurine is able to act as an osmoregulator by shuttling $K^+$and Cl^-as an electro neutral zwitterion[100] and research has shown that taurine can maintain cell volume[101].

Hypotaurine is synthesised from 3-Sulphino-L-alanine (same as L-Cysteinesulphinic acid) by Sulphoalanine decarboxylase (EC 4.1.1.29) . Hypotaurine, but not taurine, can function as an antioxidant by scavenging free radicals produced *in situ* by the Fenton reaction[102]. In addition, hypotaurine prevents the oxidation, spontaneous or catalyzed, of iron(II) to iron(III)[103]. Preventing the oxidation of Fe^{2+} to Fe^{3+} may be of importance because this step is critical for the recycling of the Fenton reaction and the generation of hydroxyl radical ($HO^{\bullet}$). This may indicate a the biochemical branch between CDO and γ-glutamylcysteine synthase may function in response to the nature of the oxidative insult.

1.2.8 Dioxygenases

Dioxygenases belong to a class of enzyme that use molecular oxygen (O_2) to oxidise a substrate. In the case of CDO, L-cysteine would be the substrate. These enzymes require a transition metal cofactor, typically iron although copper and other atoms have been reported. The reaction also requires an additional dinucleotide, FAD(P)) or NAD(P), cofactor to drive the reaction.

In human, at least fifteen different dioxygenases have been identified. The specificity of their substrates are summarised below:

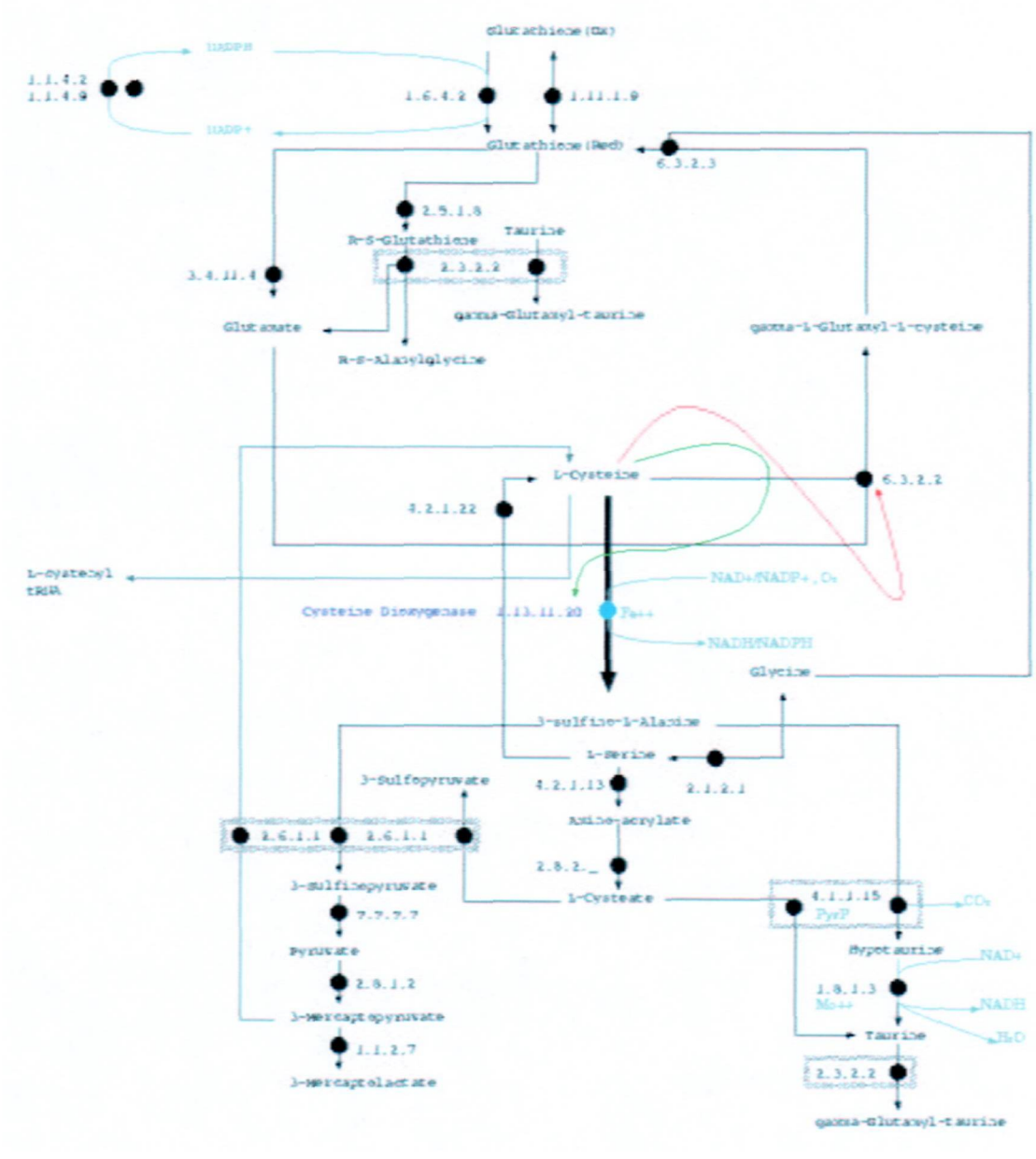


Figure 1.6: Biochemically relevant roles of L-cysteine in intracellular homoeostasis.

CDO	cysteine dioxygenase
HGD	homogentisate 1,2-dioxygenase
PLOD	2-oxaloglutarate 5-dioxygenase
PLOD2	2-oxaloglutarate 5-dioxygenase
PLOD3	2-oxaloglutarate 5-dioxygenase
HPD	4-hydroxyphenyl pyruvate dioxygenase
INDO	indolamine pyruvate 2,3-dioxygenase
P4HA1	2-oxaloglutarate 4-dioxygenase
P4HA2	2-oxaloglutarate 4-dioxygenase
P4HB	2-oxaloglutarate 4-dioxygenase
TDO2	tryptophan 2,3-dioxygenase
BBOX1	γ-oxaloglutarate dioxygenase 1
HAAO	3-hydroxyanthranilate 3,4-dioxygenase
BCDO	β-carotine dioxygenase
BCDO2	β-carotine dioxygenase

The structures and mechanisms of action of several dioxygenases have been solved. The mechanism of quercetin 2,3-dioxygenase and 2,3-dihydroxybiphenyl 1,2-dioxygenase is pictured in the Appendix on pages 145 and 146, respectively [104][105]. For dioxygenases with sequence homology, one could speculate that the reaction mechanisms were also similar. However, in the situation of cysteine dioxygenase, no such sequence homology exists. A phylogenetic

analysis of human dioxygenases (Figure 1.7) and the following multiple alignment, for the same sequences, illustrate the uniqueness of cysteine dioxygenase.

```
 1  MEQTE......VLKPRTLADLIRILHQLFAG.....DEVN  CDO
 1  MAELK........YISGFGNECSSEDPRCPG..SLPEGQN  HGD
 1  MTTYSDKGAKPERGRFLHFHSVTFWVGNAKQ..A..ASFY  HPD
 1  MAHAM.ENSWTISKEYHIDEEVGFALPNPQE..NLPDFYN  INDO
 1  MIWYI.....LIIGILLPQSLAHPGFFTSIG..QMTDLIH  P4HA1
 1  MLRRA..........LLCLAVAALVRADAPE.........  P4HB
 1  MRPLL...........LLALLGWLLLAEAKG.....DAKP  PLOD
 1  MGGCT.VKPQLLLLALVLHPWNPCLGADSEK....PSSIP  PLOD2
 1  MSGCP....................FLGNNFG.......YT  TDO2
 1  MACTI.....QKAEALDGAHLMQILWYDEEE..SLYPAVW  BBOX1
 1  MKLWV...SALLMAWFGVLSCVQAEFFTSIG..HMTDLIY  P4HA2
 1  MTSSG.PGPRFLLLLPLLLPPAASASDRPRG....RDPVN  PLOD3
 1  MERRL................GVRAWVKENRG..SFQPPVC  HAAO
 1  MDIIF.........................G........R  BCDO
 1  MVYRL....PVFKRYMGNTPQKKAVFGQCRGLPCVAPLLT  BCDO2
    !                              *           consensus

30  VEE..VQ.AIMEAYESDPIEWAMYA....KFDQYRY..TR  CDO
31  NPQVCPY.NLYAEQLSGSAFTCPRS...TNKRSWLYRILP  HGD
37  CSKMGFE.PLAYRGLETGSREVVSH..VIKQGKIVFVLSS  HPD
38  DWMFIAK.HLPDLIESGQLRERVEKLNMLSIDHLTDHKSQ  INDO
34  TEKDLVT.SLKDYIKAEEDKLEQIK....KWAEKLDRLTS  P4HA1
22  EEDHVLV.LRKSNFAEALAAHKYLL..VEFYAPWCGHCKA  P4HB
25  EDNLLVL.TVATKETEGFRRFKRSA....QFFNYKIQALG  PLOD
36  TDKLLVI.TVATKESDGFHRFMQSA....KYFNYTVKVLG  PLOD2
15  FKKLPVEGSEEDKSQTGVNRASKGGLIYGNYLHLEKVLNA  TDO2
34  LRDNCPC.SDCYLDSAKARKLLVEA......LDVNIGIKG  BBOX1
36  AEKELVQ.SLKEYILVEEAKLSKIK....SWANKMEALTS  P4HA2
36  PEKLLVI.TVATAETEGYLRFLRSA....EFFNYTVRTLG  PLOD3
24  NKLMHQE.QLKVMFVGGPNTRKDYH..IEEGEEVFYQLEG  HAAO
 8  NRKEQLE.PVRAKVTGKIPAWLQGT..LLRNGPGMHTVGE  BCDO
37  TVEEAPR.GISARVWGHFPKWLNGS..LLRIGPGKFEFGK  BCDO2
                                              consensus

61  N.LVDQGNGKFNLM.ILCWGEGHGSSIH......DHTNSH  CDO
67  S.VSHKPFESI........DEGHVTHNW.DEVDPDPNQLR  HGD
74  A.LNPWNKEMGDHL....VKHGDGV....KDIAFEVEDC.  HPD
77  R.LARLVLGCITMAYVWGKGHGDVRKVLPRNIAVPYCQLS  INDO
69  T.ATKDPEGFVGHP......VNAFKLM.KRLNTEWSELE  P4HA1
59  L.APEYAKAAGKLK.....AEGSEIRLA.KVDATEESDLA  P4HB
60  L.GEDWNVEKGTSA.....GGGQKVRLL.KKALEKHADKE  PLOD
71  Q.GEEWRGGDGINS....IGGGQKVRLM.KEVMEHYADQD  PLOD2
55  QELQSETKGNKIHDEHLFIITHQAYELWFKQILWELDSVR  TDO2
67  L.IFDRKKVYITWPDEH.YSEFQADWLK.KRCFSKQARAK  BBOX1
71  K.SAADAEGYLAHP......VNAYKLV.KRLNTDWPALE  P4HA2
71  L.GEEWRGGDVART....VGGGQKVRWL.KKEMEKYADRE  PLOD3
61  D.MVLRVLEQGKHRDVV.IRQGEIFLLPARVPHSPQRFAN  HAAO
45  S.RYNHWFDGLALLHSFTIRDGEVYYRS.KYLRSDTYNTN  BCDO
74  D.KYNHWFDGMALLHQFRMAKGTVTYRS.KFLQSDTYKAN  BCDO2
                       *  *        *          consensus
```

```
 93  ...................CFLKMLQGNLKE..........T  CDO
 97  WKPFEIPKASQKKVDFVSGLHTLCGAGDI.KSNNGLAIHI  HGD
104  ..DYIVQKARERGAKIMREPWV.....EQ.DKFGKVKFAV  HPD
116  .KKLELPPILVYADCVLANWKK......K.DPNKPLTYEN  INDO
100  ..NLVLKDMSDGFISNLTIQRPVLSNDED.QVGAAKALLR  P4HA1
 92  .........QQYGVRGYPTIKFFRNGDTA.SPKEYTAGRE  P4HB
 93  ..DLVILFADSYDVLFASGPRELLKKFRQ.ARSQVVFSAE  PLOD
105  ..DLVVMFTECFDVIFAGGPEEVLKKFQK.ANHKVVFAAD  PLOD2
 95  .........EIFQNGHVRDERNMLKVVSR..MHRVSVILK  TDO2
104  LQRELFFPECQYWGSELQLPTLDFEDVLRYDEHAYKWLST  BBOX1
102  ..DLVLQDSAAGFIANLSVQRQFFPTDED.EIGAAKALMR  P4HA2
105  ..DMIIMFVDSYDVILAGSPTELLKKFVQ.SGSRLLFSAE  PLOD3
 99  ..TVGLVVERRRLETELDGLRYYVGDTMD......VLFEK  HAAO
 83  ......IEANRIVVSEFGTMAY..PDPCK.NIFSKAFSYL  BCDO
112  ......SAKNRIVISEFGTLA..LPDPCK......NVFER  BCDO2
                                               consensus


106  LFAWPDK..KSNEMVKKSERV...L.RENQCAYI..NDSI  CDO
136  FLCNTSM..ENRCFYNSDGDF.LIV..PQKGNL..LIYTE  HGD
136  LQTYGDT..THTLVEKMNYIG.QFLPGYEAPAFM.DPLLP  HPD
148  MDVLFSF..RDGDCSKGFFLV.SLL..VEIAAA..SAIKV  INDO
137  LQDTYNL..DTDTISKGNLPG...V...KHKSFLTAEDCF  P4HA1
122  ADDIVNW......LKKRTGPA.ATT..LPDGAA..AESLV  P4HB
130  ELIYPDR..RLETKYPVVSDG.KRF..LGSGGFI..GYAP  PLOD
142  GILWPDK..RLADKYPVVHIG.KRY..LNSGGFI..GYAP  PLOD2
124  LLVQQFS..ILETMTALDFNDFREY.LSPASGFQSLQFRL  TDO2
144  LKKVGIV..RLTGASDKPGEV.SKL..GKRMGFLYLTFYG  BBOX1
139  LQDTYRL..DPGTISRGELPG.TKY.....QAMLSVDDCF  P4HA2
142  SFCWPEW..GLAEQYPEVGTG.KRF..LNSGGFI..GFAT  PLOD3
131  WFYCKDLGTQLAPIIQEFFSS.EQY...RTGKPI......  HAAO
114  SHTIPDF..TDNCLINIMKCG.EDF.......YA.TSETN  BCDO
138  FMSRFELPGKAAAMTDDTNVNYVRY....KGDYYLCTETN  BCDO2
                                     *         consensus


138  GLHRVENISHTEPAVSLHLYSPPFDTCHAFDQRTGHKNKV  CDO
169  FGKML..VQPNEI..CVIQRGMRFSIDVFEE....TRGYI  HGD
172  KLPKC.SLEMIDHIVGNQPDQEMVSASEWYL.KNLQFHRF  HPD
181  IPTVF...KAMQM......QERDTLLKALLEIASCLE.KA  INDO
169  ELGKV.AYTEADY..YHTELWMEQALRQLDEGEISTIDKV  P4HA1
151  ESSEV...AVIGF....FKDVESDSAKQFLQ.......AA  P4HB
163  NLSKL..VAEWEG..QDSDSDQLFYTKIFLD.PEKREQIN  PLOD
175  YVNRI..VQQWNL..QDNDDDQLFYTKVYID.PLKREAIN  PLOD2
161  LENKIGVLQNMRV.........PYNRRHYRD.........  TDO2
179  HTWQV..QDKIDA...NNVAYTTGKLSFHTDYPALHHPPG  BBOX1
171  GMGRS.AYNEGDY..YHTVLWMEQVLKQLDAGEEATTTKS  P4HA2
175  TIHQI..VRQWKY..KDDDDDQLFYTRLYLD.PGLREKLS  PLOD3
161  .PDQL..LKEPPFPLSTRSIMEPMSLDAWLD....SHHRE  HAAO
143  YIRKI.NPQTLET.......LEKVDYRKYVAVNLATSHPH  BCDO
174  FMNKV.DIETLEK.......TEKVDWSKFIAVNGATAHPH  BCDO2
                                               consensus
```

```
178  TMTFHSKFGIR....TPNATSGSLENN.............  CDO
201  LEVYGVHFELP.DLGPIGANGLANPRDFL.IPIAWYEDRQ  HGD
210  WSVDD..TQVHTEYSSLRSIVVANYEESIKMPINEPAPGK  HPD
211  LQVFH...QIH.DHVNPKAF.FSVLRIYLSGWKGNPQLSD  INDO
206  SVLDYLSYAVY.QQGDLDKA.LLLTKKLLELDPEHQRANG  P4HA1
177  EAIDDIPFGIT.SNSDVFSK.YQLDKDGVVLFKKFDEGRN  P4HB
198  ITLDH.RCRIF...QNLDGA...LDEVVLKFEMGHVRARN  PLOD
210  ITLDH.KCKIF...QTLNGA...VDEVVLKFENGKARAKN  PLOD2
183  .NFKGEENELL.....LKSE...QEKTLLELVEAWLERTP  TDO2
214  VQLLHCIKQTV.TGGDSEIV..DGFNVCQKLKKNNPQAFQ  BBOX1
208  QVLDYLSYAVF.QLGDLHRA.LELTRRLLSLDPSHERAGG  P4HA2
210  LNLDH.KSRIF...QNLNGA...LDEVVLKFDRNRVRIRN  PLOD3
194  LQAGT.PLSLF...GDTYET......QVIAYGQGSSEGLR  HAAO
175  YDEAGNVLNMGTSIVEKGKTKYVIFKIPATVPEGKKQGKS  BCDO
206  YDPDGTAYNMG....NSFGP.YGFSYKVIRVPPEEVDLGE  BCDO2
                                          *    consensus


201  ........................................  CDO
239  ..VPGGY.TVINKYQGKLFAAKQDVSPFNVVAWHGNYTPY  HGD
248  ....KKS..QIQEYVDYNGGAGVQHIALKTEDIITAIRHL  HPD
246  GLVYEGFWEDPKEFAGGSAGQSSVFQCFDVLLGIQQTAGG  INDO
244  .....NL................KYFEYIMAKEKDVNKSA  P4HA1
215  .....NF................EGEVTKENLLDFIKHNQL  P4HB
231  ..LAYDT.LPVLIHGNGPTKLQLNYLGNYIPRFWTFETGC  PLOD
243  ..TFYET.LPVAINGNGPTKILLNYFGNYVPNSWTQDNGC  PLOD2
214  .....GL....EPHGFNFWGKLEKNITRGLEEEFIRIQAK  TDO2
251  ......I...LSSTFVDFTDIGVDYCDFSVQSK...HKII  BBOX1
246  .....NL................RYFEQLLEEE....REK  P4HA2
243  ..VAYDT.LPIVVHGNGPTKLQLNYLGNYVPNGWTPEGGC  PLOD3
224  ..QNVDV.WLWQLEGSSVVTMGGRRLSLAPDDSLLVLAGT  HAAO
215  PWKHTEV...FCSIPSRSLLSPSYYHSFGVTENYVIFLEQ  BCDO
241  ..TIHGV.QVICSIASTEKGKPSYYHSFGMTRNYIIFIEQ  BCDO2
                                               consensus


201  ........................................  CDO
276  KYN.LKNFM.VINSVAFDHADPSIFTVLTAKSVRPGVAIA  HGD
282  RERGLEFLS..........VPSTYYKQLREKLKTAKIKVK  HPD
286  GHA.AQFLQ.DMRRYMPPAHRNFLCSLESNPSVREFVLSK  INDO
263  SDD..QSDQ.....KTTPKKKGVAVDYLPERQKYEMLCRG  P4HA1
235  PLV.IEFTEQTAPKIFGGEIKTHILLFLPKSVSDYDGKLS  P4HB
268  TVC.DEGLR.SLKGIGDEALPTVLVGVFIEQPTPFVSLFF  PLOD
280  TLC..EFDT..VDLSAVDVHPNVSIGVFIEQPTPFLPRFL  PLOD2
245  EES..EEKE..EQVAEFQKQKEVLLSLFDEKRHEHLLSKG  TDO2
279  ELD..DKGQ.VVRINFNNATRDTIFDVPVERVQPFYAALK  BBOX1
261  TLT..NQTE..AELATPEGIYERPVDYLPERDVYESLCRG  P4HA2
280  GFC..NQDR..RTLPGGQPPPRVFLAVFVEQPTPFLPRFL  PLOD3
261  SYA.WERTQ...........GSVALSVTQDPACKKPLG..  HAAO
252  PFR.LDILK.MATAYIRRMSWASCLAFHREEKTYIHIIDQ  BCDO
278  PLK.MNLWK.IATSKIRGKAFSDGISWEPQCNTRFHVVEK  BCDO2
                                   *           consensus
```

```
201  ........................................  CDO
314  DFV..IFPPRWGVADKTFRPPYYHRNCMSEFMGLIRGHYE  HGD
312  ENIDALEELK....ILVDYDEKGYLLQ...IFTKPVQDRP  HPD
324  GDAGLREAYDACVKALVSLRSYHLQIVTKYILIPASQQPK  INDO
296  EGIKMTPRRQ....................KKLFCRYHDGNR  P4HA1
274  NFKTAAESFKGKILFIFIDSDHTDNQRILEFFGLKKEECP  P4HB
306  QRLLRLHYPQKHMRLFIHNHEQHHKAQVEEFLAQHGSEYQ  PLOD
316  DILLTLDYPKEALKLFIHNKEVYHEKDIKVFFDKAKHEIK  PLOD2
281  ERRLSYRALQGALMIYFYREEPRFQVPFQLLTSLMDIDSL  TDO2
316  EFVDLMNSKESKFTFKMNPGDVITFDNWRLLHGRRSYE.A  BBOX1
297  EGVKLTPRRQ....................KRLFCRYHHGNR  P4HA2
316  QRLLLLDYPPDRVTLFLHNNEVFHEPHIADSWPQLQDHFS  PLOD3
287  ........................................  HAAO
290  RTRQPVQTKFYTGAMVVFHHVNAYEEDGCIVFDVIAYEDN  BCDO
316  RTGQLLPGRYYSKPFVTFHQINAFEDQGCVIIDLCCQDNG  BCDO2
                                              consensus

201  ........................................  CDO
352  AKQGGF..LPGGGSLHSTMTPHGPDADCFEKASKVKLAPE  HGD
345  TLF..LEVIQ..RHNHQGFGAGNFNSLFKAFEEEQNLRGN  HPD
364  ENK.....TS..EDPSKLEAKGTGGTDLMNFLKTVRSTTE  INDO
318  NPK..F.ILA.............PAKQEDEWDKPRIIRFH  P4HA1
314  AVR....LIT....LEEEMTKYKPESEELTAERITEFC.H  P4HB
346  SVK....LVG....PEVRMANADARNMGADLCRQDRSCTY  PLOD
356  TIK....IVG....PEENLSQAEARNMGMDFCRQDEKCDY  PLOD2
321  MTKWRYNHVC.MVHRMLGSKAGTGGSSGYHYLRSTVSDRY  TDO2
355  GTE.....IS..RHLEGAYADWDVVMSRLRILRQRVENGN  BBOX1
319  APQ..L.LIA.............PFKEEDEWDSPHIVRYY  P4HA2
356  AVK....LVG....PEEALSPGEARDMAMDLCRQDPECEF  PLOD3
287  ........................................  HAAO
330  SLYQLF.YLA...NLNQDFKE.NSRLTSVPTLRRFAVPLH  BCDO
356  RTLEVY.QLQNLRKAGEGLDQ.VHNSAAKSFPRRFVLPLN  BCDO2
                                              consensus

201  ........................................  CDO
390  RIADGTMAFMFESSLSLAVTKWGLKASRCLDENYHKCWEP  HGD
381  LTNMETNGVVPGM..........................  HPD
397  KSLLKEG................................  INDO
342  DIISDAE.............IEIVKDLAKPRLSR.ATVHDP  P4HA1
345  RF.LEGK.IKPHLMSQELPEDWDKQPVKVLV.........  P4HB
378  YFSVDADVALTEPNSLRLLIQQNKNVIAPLMTRHGRLWSN  PLOD
388  YFSVDADVVLTNPRTLKILIEQNRKIIAPLVTRHGKLWSN  PLOD2
360  KVFVD....LFNLSTYLIPRHWIPKM.NPTI..HKFLY..  TDO2
388  ........................................  BBOX1
343  DVMSDEE.............IERIKEIAKPKLAR.ATVRDP  P4HA2
388  YFSLDADAVLTNLQTLRILIEENRKVIAPMLSRHGKLWSN  PLOD3
287  ........................................  HAAO
365  .VDKNAEVGTNLIKVASTTATALKEEDGQVYCQPEFLYEG  BCDO
394  .VSLNAPEGDNLSPLSYTSASAVKQADGTICCSHENLHQE  BCDO2
                                              consensus
```

```
201  ........................................  CDO
430  LKSHFTPNSRNPAEPN........................  HGD
394  ........................................  HPD
404  ........................................  INDO
369  ETGKLTTAQY...............................  P4HA1
374  ......GKNFEDVAFDEKKNVFVEFYAPW.CGHCKQLAPI  P4HB
418  FWGALSADGYYARSEDYVDIVQGRRVGVWNVPYISNIYLI  PLOD
428  FWGALSPDGYYARSEDYVDIVQGNRVGVWNVPYMANVYLI  PLOD2
391  .TAEYCDSSYFSSDESD........................  TDO2
388  ........................................  BBOX1
370  KTGVLTVASY...............................  P4HA2
428  FWGALSPDEYYARSEDYVELVQRKRVGVWNVPYISQAYVI  PLOD3
287  ........................................  HAAO
404  L..ELPRVNYAHNGKQY.......RY.VFATG..VQWSPI  BCDO
433  ...DLEKEGGIEFPQIYYDRFSGKKYHFF.YG.CGFRHLV  BCDO2
                                               consensus

201  ........................................  CDO
446  ........................................  HGD
394  ........................................  HPD
404  ........................................  INDO
379  ........RVSKSAWLSGYENPVVSRINMRIQDLTGLDVS  P4HA1
407  WDKLGETYKDHENIVI.AKMDSTANEVEAVKVH.SFPTLK  P4HB
458  KGSALRGELQSSDLFHHSKLDPDMAFCANIRQQDVFMFLT  PLOD
468  KGKTLRSEMNERNYFVRDKLDPDMALCRNAREMGVFMYIS  PLOD2
407  ........................................  TDO2
388  ........................................  BBOX1
380  ........RVSKSSWLEEDDDPVVARVNRRMQHITGLTVK  P4HA2
468  RGDTLRMELPQRDVFSGSDTDPDMAFCKSFRDKGIFLHLS  PLOD3
287  ........................................  HAAO
432  PTKIIKYDILTKSSLKWREDDCWPAEPLFVPAPGAKDEDD  BCDO
468  GDSLIKVDVVNKTLKVWREDGFYPSEPVFVPAPGTNEEDG  BCDO2
                                               consensus

201  ........................................  CDO
446  ........................................  HGD
394  ........................................  HPD
404  ........................................  INDO
411  TAEELQVANYGVGGQYEPHFDFARKDEPDAFKELGTGNRI  P4HA1
445  .FFPASADRTVIDYNGERTLDGFKKFLESG.GQDGAGDDD  P4HB
498  NRHTLGHLLSLDSYRTTHLHNDLWEVFSNPEDWKEKYIHQ  PLOD
508  NRHEFGRLLSTANYNTSHYNNDLWQIFENPVDWKEKYINR  PLOD2
407  ........................................  TDO2
388  ........................................  BBOX1
412  TAELLQVANYGVGGQYEPHFDFSRNDERDTFKHLGTGNRV  P4HA2
508  NQHEFGRLLATSRYDTEHLHPDLWQIFDNPVDWKEQYIHE  PLOD3
287  ........................................  HAAO
472  GVILSAIVSTDPQKLPFLLILDAKSFTELARASVDVDMHM  BCDO
508  GVILSVVITPNQNESNFLLVLDAKNFEELGRAEVPVQMPY  BCDO2
                                               consensus
```

```
201  ........................................  CDO
446  ........................................  HGD
394  ........................................  HPD
404  ........................................  INDO
451  ATWLFYMSDVSAGGATVFPEVGASVWPKKGTAVFWYNLFA  P4HA1
483  DLE.....D.LEEAEEPDME......EDDDQKAVKDEL..  P4HB
538  NYTKALAGK.LVETPCPDVYWFPIFTEVACDELVEEMEHF  PLOD
548  DYSKIFTEN.IVEQPCPDVFWFPIFSEKACDELVEEMEHY  PLOD2
407  ........................................  TDO2
388  ........................................  BBOX1
452  ATFLNYMSDVEAGGATVFPDLGAAIWPKKGTAVFWYNLLR  P4HA2
548  NYSRALEGEGIVEQPCPDVYWFPLLSEQMCDELVAEMEHY  PLOD3
287  ........................................  HAAO
512  DLHGLFITD.........MDW.....DTKKQAASEEQRDR  BCDO
548  GFHGTFIPI..............................  BCDO2
                                              consensus

201  ........................................  CDO
446  ........................................  HGD
394  ........................................  HPD
404  ........................................  INDO
491  SGEGDYSTRHAACPVLVGNKWVSNKWLHERGQEFRRPCTL  P4HA1
509  ........................................  P4HB
577  GQWSLGNNKDNRIQGGYENVPTIDIHMNQIGFEREWHKFL  PLOD
587  GKWSGGKHHDSRISGGYENVPTDDIHMKQVDLENVWLDFI  PLOD2
407  ........................................  TDO2
388  ........................................  BBOX1
492  SGEGDYRTRHAACPVLVGCKWVSNKWFHERGQEFLRPCGS  P4HA2
588  GQWSGGRHEDSRLAGGYENVPTVDIHMKQVGYEDQWLQLL  PLOD3
287  ........................................  HAAO
538  ASDCHGAPLT.............................  BCDO
557  ........................................  BCDO2
                                              consensus

201  ........................................  CDO
446  ........................................  HGD
394  ........................................  HPD
404  ........................................  INDO
531  SELE....................................  P4HA1
509  ........................................  P4HB
617  LEYIAPMTEKLYPGYYTRAQFDLAFVVRYKPDEQPSLMPH  PLOD
627  REFIAPVTLKVFAGYYTKGFALLNFVVKYSPERQRSLRPH  PLOD2
407  ........................................  TDO2
388  ........................................  BBOX1
532  TEVD....................................  P4HA2
628  RTYVGPMTESLFPGYHTKARAVMNFVVRYRPDEQPSLRPH  PLOD3
287  ........................................  HAAO
548  ........................................  BCDO
557  ........................................  BCDO2
                                              consensus
```

```
201  ........................................  CDO
446  ........................................  HGD
394  ........................................  HPD
404  ........................................  INDO
535  ........................................  P4HA1
509  ........................................  P4HB
657  HDASTFTINIALNRVGVDYEGGGCRFLRYNCSIRAPRKGW  PLOD
667  HDASTFTINIALNNVGEDFQGGGCKFLRYNCSIESPRKGW  PLOD2
407  ........................................  TDO2
388  ........................................  BBOX1
536  ........................................  P4HA2
668  HDSSTFTLNVALNHKGLDYEGGGCRFLRYDCVISSPRKGW  PLOD3
287  ........................................  HAAO
548  ........................................  BCDO
557  ........................................  BCDO2
                                               consensus

201  ..............................  CDO
446  ..............................  HGD
394  ..............................  HPD
404  ..............................  INDO
535  ..............................  P4HA1
509  ..............................  P4HB
697  TLMHPGRLTHYHEGLPTTRGTRYIAVSFVDP  PLOD
707  SFMHPGRLTHLHEGLPVKNGTRYIAVSFIDP  PLOD2
407  ..............................  TDO2
388  ..............................  BBOX1
536  ..............................  P4HA2
708  ALLHPGRLTHYHEGLPTTWGTRYIMVSFVDP  PLOD3
287  ..............................  HAAO
548  ..............................  BCDO
557  ..............................  BCDO2
                                     consensus
```

X non conserved
X similar
X conserved
X all match

1.2.9 Cysteine Dioxygenase

Pointing out the Fenton reaction as a mechanism for the generation of ROS in neurodegenerative disease is like beating a dead horse. Understanding what dysregulates the cell environment that allows opportunistic transition metals to bind to misfolded proteins and carry out their unique chemistry is a topic of importance in the field.

Cysteine dioxygenase (EC 1.13.11.20) (CDO) catalyses the conversion of

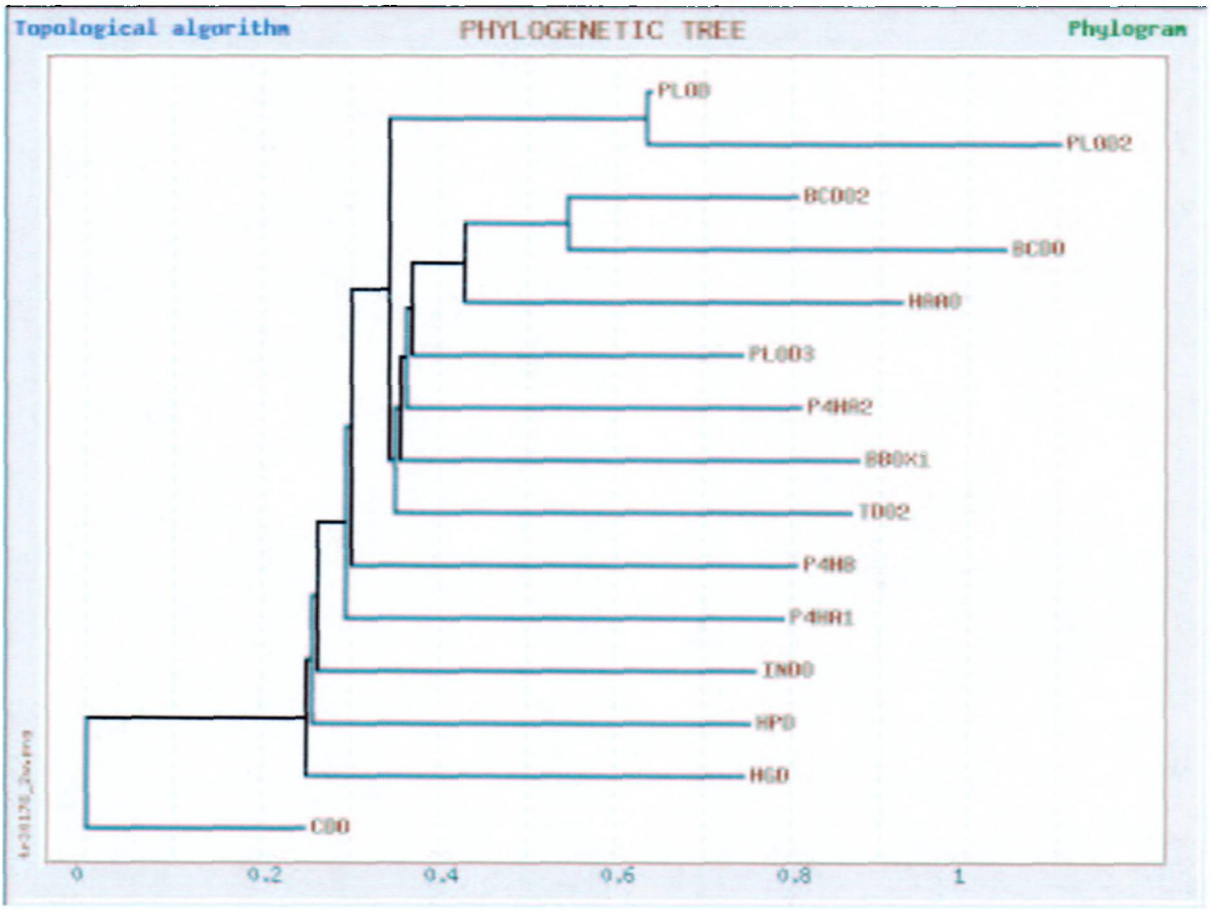


Figure 1.7: Phylogenetic analysis of sequence data of human dioxygenases.

cysteine to 3-Sulphino-L-alanine (also named L-Cysteinesulphinic acid and 3-Sulphinoalanine), a necessary substrate for hypotaurine and taurine biosynthesis. It has 200 amino acid residues in human and rat, has a calculated molecular weight of 22984 Da in man, it uses Fe^{2+} as a cofactor, and requires NAD(P)H. CDO has been identified in *homo sapiens*, *rattus norvegucus*, *drosophila melanogaster* (fly), *caenorhabditis elegans* (nematode)(preliminary) and *schistosoma japonicum* (oriental blood fluke). CDO shares no sequence homology with other proteins (including other dioxygenases) yet there is considerable protein sequence homology amongst CDOs in vastly different organisms (Figure 1.8). In regions where the primary sequences are not as conserved, there may well be a considerable conservation in their secondary/ tertiary structure (Figure 5.4, page 112). Such conservation may indicate a similarly conserved physiological role as well. Substrates and products of CDO are involved in radical scavenging, oxidation and reduction homoeostasis, osmotic regulation, detoxification, and signal transduction.

It is believed that CDO is a principal regulator of intracellular cysteine level[106]. Owing to the fact that glutathione levels are regulated by the availability of cysteine, CDO is automatically suspect as a possible regulator of glutathione levels as well.

Additional interesting physiological parallels of a cysteine dioxygenase from the fungus *histoplasma capsulatum* can be drawn. It exists in nature as a multicellular fungus but is a unicellular pathogenic yeast in an animal host. The phase change of *histoplasma capsulatum* is dependant on a low redox potential, a change in intracellular cysteine levels, as well as cysteine/redox dependant changes in mitochondrial respiration; expression of cysteine dioxygenase is believed to be the enzyme mediating the intracellular redox changes that lead to differentiation into the pathogenic form[107, 108, 109]. Although

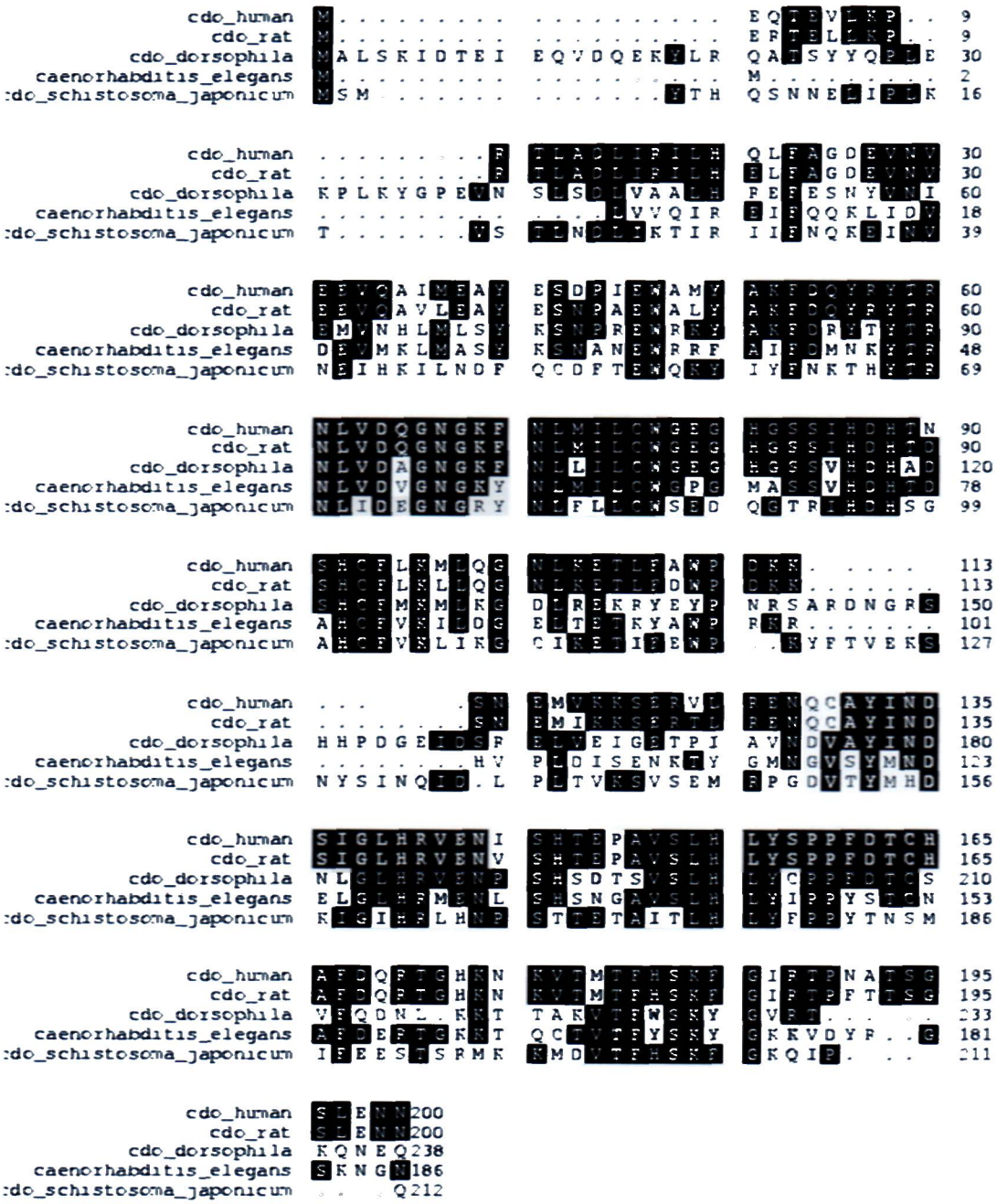


Figure 1.8: Protein sequence homology of CDO

a different organism, CDO has been shown to regulate the redox status as well as influence mitochondrial respiration.

1.2.10 Summary

Cysteine dioxygenase is a rate limiting enzyme that is responsible for regulating intracellular cysteine levels and for the biosynthesis of hypotaurine and taurine. Glutathione and neuromelanin[110] biosynthesis are dependant upon the availability of cysteine. Hypotaurine, taurine, neuromelanin and GSH/GSSG are all important compounds for maintaining homoeostasis and for responding to different stressors present under normal and pathological states. As well as being a key biosynthesis enzyme, CDO also appears to be able to shift the stress response of the cell depending on the nature of the insult.

Chapter 2

Aims

The physical characterisation of CDO is the primary objective. Firstly, to express the protein an confirm the primary structure. Secondly, to find solution conditions in which the protein is stable and appears to be folded. Lastly, to quantify secondary and tertiary structural characteristics of the protein.

A system that provides a continuous supply of protein and chromatography to purify the enzyme is a prerequisite to accomplish these goals. Circular dichroism is a physical technique to measure the secondary structure of the protein. Dynamic light scattering is used to monitor the behaviour of CDO in solution and to design experiments around it. Nuclear magnetic resonance and protein crystallography are the final techniques to quantitatively analyse the actual structure of the protein under near physiological conditions. The ultimate aim is to determine the tertiary structure of the protein or to at least lay down the foundations to solve its structure.

Gaining detailed structural information about CDO is of more than just academic interest. Pharmacophores could be synthesised, based upon structure, specific to the protein that act as antagonists, agonists, or modulate the function in some other manner. Such compounds would be useful to re-

searchers studying cell or animal models of diseases. If the modulation of the enzyme proves that it could play a pivotal role in any pathology, the structure is of benefit pharmacologically for the development of next generation drugs based on structure; "rational drug design". Drugs such as *Relenza*^{TM} [1] . *Amprenavir*^{TM} [2], *BNP7787*[3] and *Entacapone*^{TM} [4] are a few examples of therapeutics rationally designed from 3-D structures of proteins[111, 112, 113].

[1]Relenza or Zanamivir/GG167 is a potent neuraminidase inhibitor used as an anti-influenza drug. Relenza was designed based on the x-ray crystallographic structure of influenza neuraminidase.

[2]Amprenavir is an AIDS protease inhibitor that was designed based on the x-ray crystallographic structure of HIV-protease . It is licensed to GalaxoWellcome/Vertex and is also known as Vertex VX-478, Agenerase or 141W94.

[3]BNP7787 is a drug rationally designed to prevent toxcisity associated with taxanes and cisplatin used in chemotherapy and without side-effects. It was developed by Cray Supercomputer and BioNumerik Pharmaceuticals.

[4]Entacapone is a Catechol O-Methyltransferase inhibitor marketed by Orion Pharma and is being used to extend the effectiveness of levodopa therapy in Parkinson's Disease

1. Expression of CDO in a bacterial system
   (a) Proof of expression
      i. PCR analysis of construct
      ii. SDS-PAGE of product
      iii. Sequencing of the construct insert
   (b) Refine system for frequent bulk expression
2. Analysis of expressed product
   (a) Concentration Determination
      i. Surface Plasmon Resonance
      ii. Spectrophotometry
   (b) Size determination
      i. SDS-PAGE
      ii. Mass spectroscopy
      iii. Dynamic light scattering
         $\longrightarrow$ Hydrodynamic Radius
         $\longrightarrow$ Aggregation State
   (c) Analysis of Secondary Structure
      i. Circular Dichroism
   (d) Analysis of Tertiary/Quaternary Structure
      i. X-ray Crystallography
      ii. Nuclear Magnetic Resonance

Figure 2.1: Representation of the aims of the study.

Chapter 3

Expression of Cysteine Dioxygenase

For the production of cysteine dioxygenase, a bacterial BL21(DE3) expression system was chosen. The vector used to induce human CDO is pET15b, an expression vector that co-expresses a poly-histidine fusion protein (HIS_x). The advantages of this bacterial system are that CDO can be expressed quickly, with minimal risk of cross contamination, and in large yields. The advantage of the HIS_x tag is that CDO can be purified without the need for sophisticated chromatography equipment.

One potential drawback is that prokaryotic cells do not perform post-translational modifications on their expressed proteins. Endogenous CDO has been studied for a number of years and there has been no published reference of any glycosylation site[1]. Furthermore, computational analysis indicates that there is only one probable N-glycosylation site (Figure 3.1 and A.4,

[1]Unpublished reports indicate that CDO is glycosylated. However, the degree to which CDO is glycosylated and the number and sequence of these sites remains unclear. It is also unclear if these sites are glycosylated for cell localisation, protein functionality or modulation. It has also not been established under which conditions and cell types CDO may be glycosylated.

page 147) and one O-glycosylation site (Figure A.5, page 148); Asparagine-144 and Threonine-193 respectively. It is open to speculation whether any potential modification affects the overall tertiary structure. Given the high degree of sequence homology from the nematode to man (Figure 1.8, page 45), in vastly different cells, makes it unlikely they share identical post-translational modifications; the prokaryotically expressed CDO is likely to retain its structure.

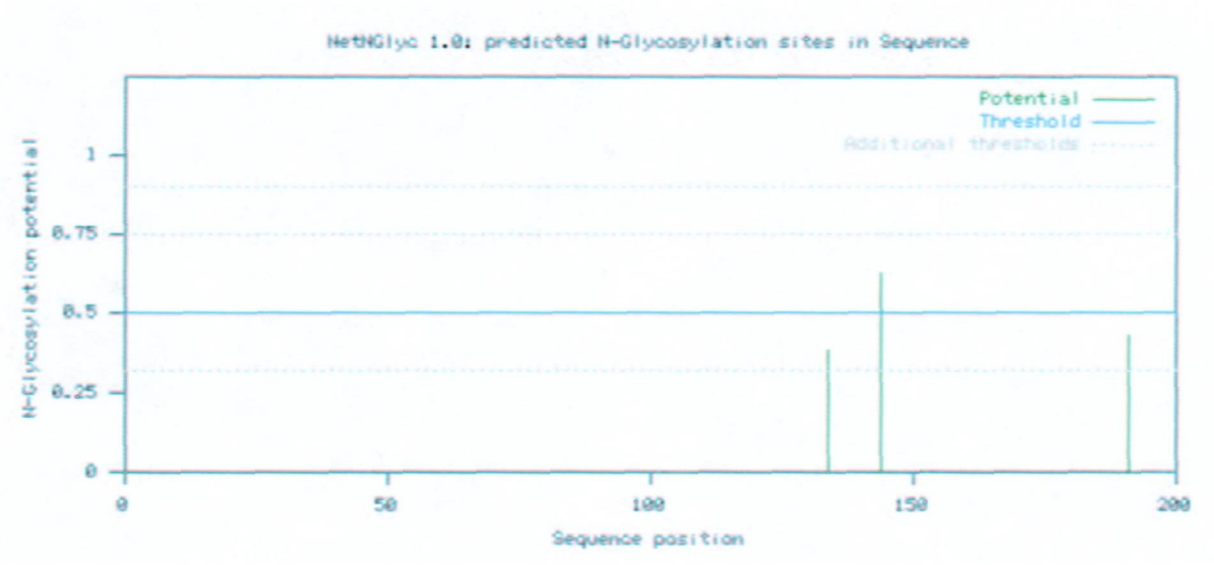


Figure 3.1: Predicted N-glycosylation of endogenous CDO using NetNGlyc 1.0.

The CDO expression system was also used with the intent of studying its protein product. If a gene product is post-translationaly modified, the modifications are rarely uniform from one particle (protein) to the next. This is especially true if the modifications are in response to cell regulation (targeted destruction or feedback regulation). Using an unmodified gene product eliminates the intractable problem of purifing each particle species of CDO.

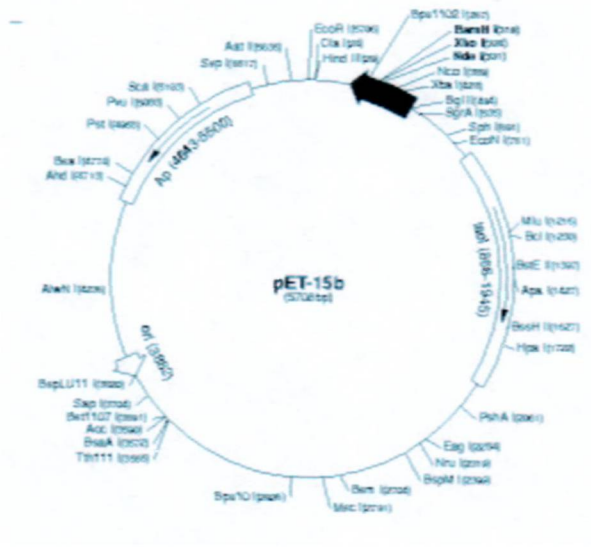


Figure 3.2: pET15b (Novagen, Inc.)

3.1 Aims

The aims of the expression system are firstly to confirm that the insert sequence of the vector matches the sequence of the human gene. Secondly, to express CDO and confirm its presence by electrophoresis and western blotting. Lastly, to develop a system with increased yield and that produces a steady supply of protein.

3.2 Protocols

3.2.1 Expression System

A Novagen pET15b expression vector was used for all experiments (Figure 3.2 and A.6, page 149)[114]. Human CDO cDNA was sub-cloned into the vector and the transfected into a BL21(DE3) cell. The cDNA was created by rtPCR with primers containing BamH1 and Nde1 and force ligated into the vector.

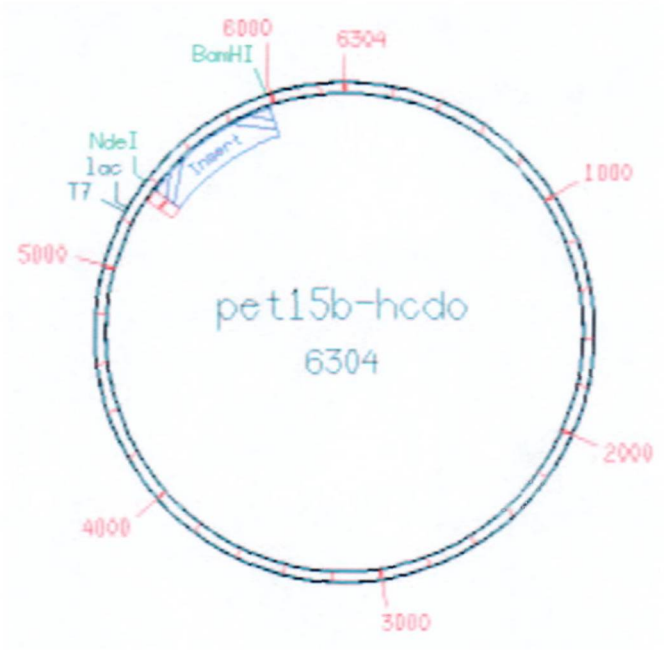


Figure 3.3: Plasmid map of the pET15b-CDO construct

SOB Medium (1ℓ)		**SOB Plating Medium** (1ℓ)	
20g	Peptone	20g	Peptone
5g	Yeast Extract	5g	Yeast Extract
0.5g	Sodium Chloride	0.5g	Sodium Chloride
833$\mu\ell$	3M Potassium Chloride	833$\mu\ell$	3M Potassium Chloride
5mls	2M Magnesium Chloride	5mls	2M Magnesium Chloride
pH$\Rightarrow$7.5$\Longrightarrow$Autoclave		15g	Bactoagar
250$\mu\ell$	200mg/ml Ampicillin	pH$\Rightarrow$7.5$\Longrightarrow$Autoclave	
		250$\mu\ell$	200mg/ml Ampicillin

Figure 3.4: SOB growth medium and plating medium

3.2.2 Confirmation of Expression System

Confirmation of the insert was performed by streaking colonies onto a SOB plate containing ampicillin (Figure 3.4) and detecting the insert by PCR. Plates were sealed with a strip of parafilm and incubated overnight at 37°C in humidity chamber. A pair of primers corresponding to 5' and 3' flanking T7 promoter/terminator regions were synthesised by Alta Bioscience.

Primer	Sequence	Synthesis ID
hCDO pET 5'	TCC-CCA-CGA-GCA-TAT-GGA-ACA-GA	D15103DR
hCDO pET 3'	GTT-TGG-TGC-CGG-ATC-CTT-AGT-TGT	D15104DR
T7 Terminator	GCT-AGT-TAT-TGC-TCA-GCG-G	D26465CB
T7 Promoter	TAA-TAC-GAC-TCA-CTA-TAG-GG	D26463CB
Term.+	GTC-CGG-CGT-AGA-GGA-TC	D27698CB
Pro.+	CGG-ATA-TAG-TTC-CTC-CT	D27697CB

Colonies were transfered directly to PCR reaction mixtures using sterile pipette tips. PCR reaction mixtures consisted of: 28.8$\mu\ell$ ddH_2O, 1$\mu\ell$ dNTP mixture (25 μM), 1$\mu\ell$ 5' primer (100 pM), 1$\mu\ell$ 3' primer (100 pM),5$\mu\ell$ 10x Buffer, 3$\mu\ell$ 25mM $MgCl_2$, 0.25$\mu\ell$ Taq Polymerase. After the mixture was inoculated directly with a bacterial colony, it was mixed with a vortexer and topped with 50$\mu\ell$ sterile mineral oil. Lysis of the cells was done by the initial melting PCR step. PCR reactions were done using a Perkin Elmer thermocycler using a step cycle. Optimal annealing temperatures were provided with the synthesis of the oligos. Thirty cycles were performed using a "1 min 94°" melting step, a "1 min 50°" annealing step and a "2 min 72°" elongation step[2]. No final

[2]Thirty cycles were chosen based on a standard practice for amplification. The melting step temperatures were based upon the theoretical melting temperatures of the synthesised oligoneucleotides. The elongation time was based on the product length.

PCR Mixture	**PCR Cycle**	**50X TAE**	
29.8$\mu\ell$ ddH_2O	(30 cycles)		
1$\mu\ell$ dNTP	1min at 94°C	242g	Tris Base
1$\mu\ell$ 3' Primer	1min at 50°C	57.1	Glacial Acetic Acid
1$\mu\ell$ 5' Primer	2min at 72°C		
4$\mu\ell$ 10X Buffer			
3$\mu\ell$ 25mM $MgCl_2$		100mls	0.5M EDTA
0.25$\mu\ell$ Taq			

Figure 3.5: PCR relevant mixtures.

extention step was used. PCR products were detected by running them on a 1% agarose gel at 80V DC in 1X TAE containing ethidium bromide. Detection of product was made using a UV transilluminator equipped with a digital camera.

For colonies with confirmed inserts of appropriate size, sequencing was performed. Two pairs of primers were synthesised corresponding to a promoter/terminator (p/t) sequence and a region outside the promoter (p+/t+). The sequences were synthesised by Alta Bioscience. The primers were selected by their specificity to the CDO construct and their annealing temperatures.

Primer	Abbreviation	Sequence
promoter sequence	p	TAA-TAC-GAC-TCA-CTA-TAG-GG
terminator sequence	t	GCT-AGT-TAT–TGC-TCA-GCG-G
upstream prom. seq.	p+	CGG-ATA-TAG-TTC-CTC-CT
downstream term. seq	t+	GTC-CGG-CGT-AGA-GGA-TC

3.2.3 Expression of CDO

Induction of CDO was done by adding isopropyl β-D-thiogalactopyranoside (IPTG). This compound in not endogenously expressed in BL21(DE3) and is

effective stimulating an otherwise silent transgenic vector gene. The Lac operator for the IPTG is also not expressed in BL21(DE3) cells making IPTG a very gene specific inducer. Cell cultures grown in 250ml volumes in 1ℓ baffled flasks. They were incubated at 37°C on a rotary shaker. Growth curves were establised by taking aliquots of the sample a different times and measuring the absorbance at λ=600nm. Induction of CDO with IPTG was done by adding $250\mu\ell$ 1M IPTG to each 250ml culture at or near $OD_{\lambda=600nm}$=0.60 for SOB medium[115]. Induction was allowed to proceed for 6 to 8 hours. Cells were harvested by adding $500\mu\ell$ 200mg/ml phenylmethylsulfonyl fluoride (PMSF) and pelleting them by centrifugation. The pellet was resuspended in approximately 20mls 1X binding buffer (10mM sodium phosphate, 500mM sodium chloride, 5mM Imidazole). The pellet was lysed using a french press. Cellular debris was removed by ultracentrifugation. Post-centrifugation supernatant was removed and passed through a $0.22\mu m$ filter to prevent clogging of the affinity column.

3.2.4 Expression of Isotopically Labelled CDO

Another advantage of using the bacterial system is that it is relatively easy to incorporate specific isotopes into the protein. In the absence of amino acids in the growth medium, prokaryotes are able to *de novo* synthesise their own amino acids from a ammonium salts and glucose.

For experiments that required isotopically labelled protein, Ammonium-^{15}N Chloride (98% ^{15}N, GOSS Scientific Instruments, NLM-467) was used as a nitrogen source to label nitrogen. D-Glucose-U-$^{13}C_6$ (99% ^{13}C, Cambridge Isotope Laboratories, CLM-1396) was used as a carbon source to label carbon. The cultures were grown in 250mls as previously described but in minimal

media:

Minimal Media (1ℓ)		**1000X Micronutrients** (2ℓ)	
6g	Na_2HPO_4	4.9g	Boric Acid
3g	KH_2PO_4	1.42g	Cobalt Chloride
5g	NaCl	0.56g	Copper Sulphate
1g	$(NH_4)Cl$	3.17g	Manganese Chloride
10mls	Vitamin Solution	0.57g	Zinc Sulphate
1ml	1000X Micronutrients		
10mls	20% Glucose		
2.6mls	154mg/ml Ampicillin		
1ml	fresh $FeSO_4$		
10μℓ	1M $CaCl_2$		
1ml	1M $MgCl_2$		

filter sterilise

3.2.5 Purification of CDO

An Amersham Pharmacia Biotech C 16/20 column (19-5101-01) with an AC 16 adaptor (19-5109-01) was packed with Chelating Sepharose Fast Flow to a bed volume of 12cm. The column was packed by attaching a peristaltic pump to the base and filling the column with 50mls of Sepharose solution, the pump was run at about 3mls/min while the sepharose packed. The buffer draining the column for this procedure was recycled to the top of the column. The resulting bed volume was 12cm and the dead space above was minimised using the AC 16 adaptor. All solutions used in the affinity purification were filter sterilised using a 0.22 μm syringe filter and degassed.

The column was first charged with 400mM $NiSO_4$. When the the colour of the column changed to blue, the free nickel was washed away with 100mls 1x Binding Buffer (10mM sodium phosphate, 500mM sodium chloride, 5mM Imidazole). The filter sterile CDO in 1X binding buffer was passed through the column. The column was washed with an addition 100mls fresh binding buffer. Non-specific binding of BL21(DE3) protein were removed by washing the column with 200mls 1X Primary Wash Solution(60mM Imidazole, 10mM sodium phosphate, 500mM sodium chloride). The bound protein was washed with an additional 100mls 1X Secondary Wash Solution (400mM sodium fluoride, 20mM Glycine, 10mM sodium phosphate)[3]. The purified CDO was eluted in 1X Elution Buffer (400mM sodium fluoride, 20mM Glycine, 10mM sodium phosphate, 200mM EDTA) and collected with a fraction collector.

3.2.6 Confirmation of Expressed Protein I

This is a modification of method previously described[116]. A Bio Rad Power Pac 300 and Mini Protean II Cell was used for all protein electrophoresis.

The acrylamide running gel (Laemmli, 1970) (Appendix pA) was prepared by mixing 6mls 1.5 M Tris/0.4% SDS, 8mls Ultrapure acrylamide (30% acrylamide : 0.8% bis-acrylamide), 80$\mu\ell$ 10% ammonium persulphate (APS) and 80$\mu\ell$ N,N,N',N'-tetramethyl-ethelenediamine (TEMED). The mixture was poured between the glass plates of a BioRad gel caster leaving 1.5 *cm* clearance from the top of the plate. Water saturated butanol was layered above to prevent dehydration and warping at the acrylamide interface. After the gel had polymerised, the butanol was washed away with water and aspirated. A stacking gel was prepared by mixing 3mls 0.5M Tris/0.4% SDS, 1.5 mls Ultrapure

[3]The purpose of the secondary wash step will be thoroughly discussed in the next chapter under dynamic light scattering

acrylamide, 7.5mls ddH_2O, 60$\mu\ell$ APS and 60$\mu\ell$ TEMED. The mixture was layered carefully above the running gel, an appropriate comb was inserted, and allowed to polymerise.

Electrophoresis was done in a 1x running buffer (192mM Glycine/25mM Tris/0.1% SDS)[115], made from a 10x concentrate. Protein samples were placed in loading buffer (1:1) consisting of 4% SDS, 20% glycerol, 10% β-mercaptoethanol, 125mM tris pH 6.8 and 0.1% bromphenol blue. Just prior to loading, the samples were denatured for 5 minutes in a water bath at 100°C and then centrifuged to collect condensate. Once loaded, gels were run for 1 to $1\frac{1}{2}$ hours at 40mA and with constant current.

Gels were usually run in pairs. After electrophoresis, one gel was stained by incubating the gel in a plate with a Coomassie Blue solution (0.26g Coomassie Brilliant Blue R250 in 90mls methanol:H_2O (1:1 v/v) + 10mls glacial acetic acid). After a one hour incubation on a rotary shaker, gels were de-stained overnight in the same solvent without the dye. The second gel was left unstained and used for western blotting.

3.2.7 Confirmation of Expressed Protein II

It was necessary to have a supply of epitope specific antibodies to detect the presence of CDO in western blots and dot blots. The epitope was chosen from a predicted hydrophilic region of the swiss-prot entry for human CDO[116]. The peptide sequence was synthesised on controlled pore glass (CPG) (AltaBioscience, Birmingham, England) and used as packing material for affinity purification of IgG's from sera. Sheep sera (The Binding Site, Birmingham, England) from animals sensitised to the same epitope was then used as the source of epitope-specific immunoglobulins.

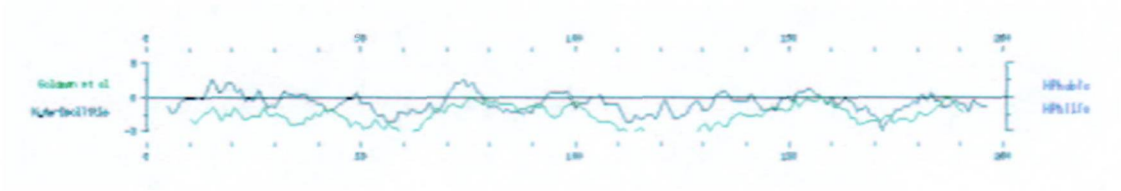


Figure 3.6: GCG representation of predicted hydrophilicity of the SwissProt sequence entry of human cysteine dioxygenase.

```
   human_epitope  ~~~~~~~~~~  ~~~~~~~~~~  ~~~~~~~~~~  ~~~~~  0
cdo_homo_sapiens  MEQTEVLKPR  TLADLIRILH  QLFAGDEVNV  EEVQA  35

   human_epitope  ~~~~~~~~~~  ~~~~~~~~~~  ~~~~~~~~~~  ~~~~~  0
cdo_homo_sapiens  IMEAYESDPI  EWAMYAKFDQ  YRYTRNLVDQ  GNGKF  70

   human_epitope  ~~~~~~~~~~  ~~~~~~~~~~  ~~~~~~~~~~  ~~~~~  0
cdo_homo_sapiens  NLMILCWGEG  HGSSIHDHTN  SHCFLKMLQG  NLKET  105

   human_epitope  ~~~~~~~~~~  ~~~~~~~~~~  ~~~~~~~~~~  ~~~~~  0
cdo_homo_sapiens  LFAWPDKKSN  EMVKKSERVL  RENQCAYIND  SIGLH  140

   human_epitope  ~~~~~~~~~~  ~~~~~~~~~~  ~~~~~~FDQR  TGHKN  9
cdo_homo_sapiens  RVENISHTEP  AVSLHLYSPP  FDTCHAFDQR  TGHKN  175

   human_epitope  KV~~~~~~~~  ~~~~~~~~~~  ~~~~~11
cdo_homo_sapiens  KVTMTFHSKF  GIRTPNATSG  SLENN200
```

Figure 3.7: A local alignment showing the location of the synthesised cysteine dioxygenase epitope with the human cysteine dioxygenase gene product.

The column was equilibrated with PBS pH 7.0 using a peristaltic pump with a flow rate of about 2 $m\ell$/min. Normal sheep sera was filter sterilised through a .22μm Millipore syringe top filter and then recycled through the equilibrated column for 30 min at the same flow rate. Bound immunoglobulins were thoroughly washed with fresh PBS 7.0 for 30 min. To elute the antibodies, the CPG was completely drained, then eluted with two separate 1 ml fractions of 100mM Glycine pH 2.3. The pH of the elute was adjusted to 8.0 using NaOH.

For western blotting analysis, the second gel was placed on a nitrocellulose membrane (0.45μm Bio Rad Transblot Transfer Medium 162-0115) cut to the same dimensions of the gel. Two pieces of damp electrode paper were placed on either side of the gel/membrane to add a little integrity to the gel. Two coarse sponge separators were then placed on either side of the sandwich, the stack was fastened into a transfer clasp and placed in a semi-dry tank. The transfer tank was filled with 1x running buffer containing 20% methanol. Electrophoresis was done a 200 mA for one hour. Upon completion, the membrane was blocked for 1h at T_R in 10% powdered milk in PBS 7.0 on a rotary shaker. Using affinity purified αCDO IgG's in a concentration of 1:10 to 1:100 in PBS 7.0, the membranes were incubated at $T_{4^{\circ}C}$ on a rotary shaker overnight. Membranes were washed three times in a fresh PBS solution. 10mls of 1:500 dilution of donkey α sheep-peroxidase conjugated antibodies in PBS were incubated for 1hr at T_R on a rotary shaker. Membranes were then washed three times in a fresh PBS solution for 10min. A developing solution was made by dissolving 1 tablet of Sigma 4-chloro napthol dissolved in 10mls of MeOH and then adding 50mls of triethanolamine buffered saline (5.5g NaCl + 2.8mls triethanolamine +17mls 1M HCl per litre and pH 7.5) and 30$\mu\ell$30% H_2O_2. The solution was then applied to the membrane until the

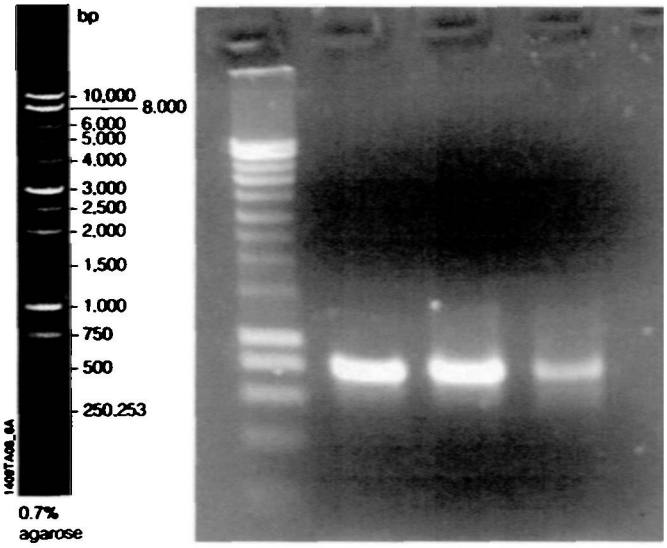


Figure 3.8: Promoter to terminator product from colony PCR of pET15b-hCDO (colonies 1,3,5 respectively) in BL21(DE3)

desired level of development was reached. The reaction was terminated by rinsing the membrane in water.

This same protocol was was used for dot blots where a small aliquot (5$\mu\ell$) of each fraction of purified CDO was placed on a nitrocellulose membrane. The membrane was then blocked and probed for CDO as previously described to determine which factions contained the eluted protein.

3.3 Results and Discussion

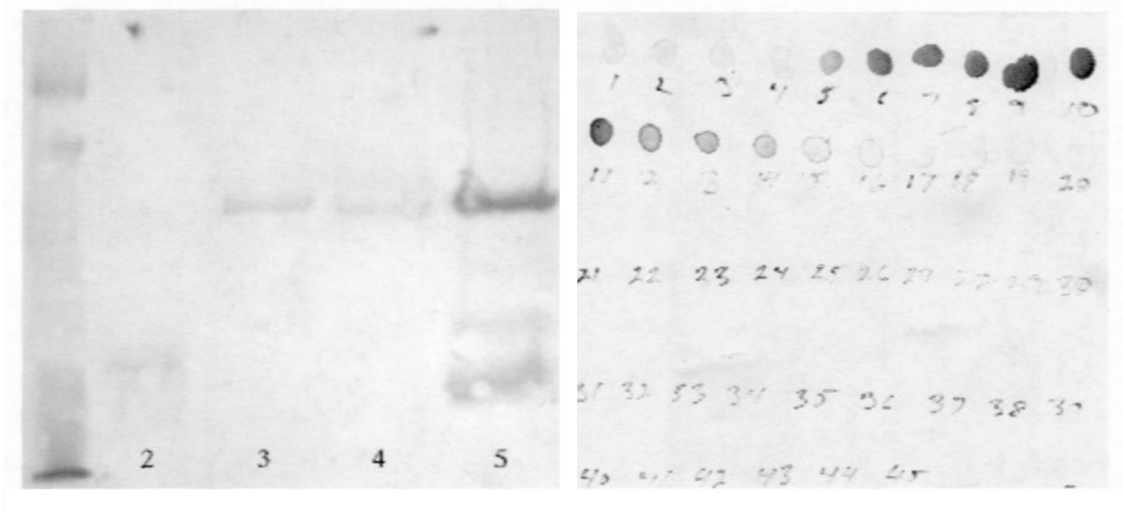
Three separate colonies of sub-cloned CDO were identified. Each separate culture was streaked on a SOB plate and after incubation, an individual colony from each was transfered to a PCR mixture specific for the construct promoter and terminator. After amplification, the PCR product was run on an 1% agarose gel. Figure 3.8 illustrates the results of the amplification. Each of the three colonies contained an insert that approximated the expected size of

the human CDO gene.

Starting with colony one, the PCR product was submitted to Alta Bioscience for sequencing with the same promoter/terminator primers as well as a second outlying set. The initial sequencing runs are listed in the appendix on pages 168 through 173. Without performing any manual base-calling, the sequence data were aligned together and the consensus sequence from the sequencing runs matched the predicted sequence of the CDO insert. The alignments are listed in figures A.14 through A.16 on pages 174 to 176.

To detect the presence of the gene product, it was necessary to first produce an epitope specific antibody for CDO. Neat sheep sera was affinity purified against a column containing the target peptide. To test its affinity, an SDS-PAGE was run on human liver homogenate. After electrophoresis, the proteins from human liver were transfered to a nitrocellulose membrane. The membrane was blocked and probed for cysteine dioxygenase. Figure3.9 illustrates the results. Normal human, horse and sheep sera were also run as positive controls for the secondary anti-sheep antibody. Dot blots of recombinantly expressed CDO where then probed with the same immunoglobulin. Since recombinant CDO was detected, it also indicates that there no fame shifts occurred during the ligation of the insert CDO gene into the pET15b vector.

With an an epitope specific antibody for human CDO, the recombinant CDO was run on SDS-PAGE and western blotted. Figure 3.10 illustrates the results. Human liver homogenate were used as a positive controls and the chromatography flow-through of induced BL21(DE3) lysate were used to gauge protein binding. Since the concentration of purified CDO was far greater than the amount present in liver homogenate, the endogenous CDO band appears very faint by comparison. Also present in the wester blot and

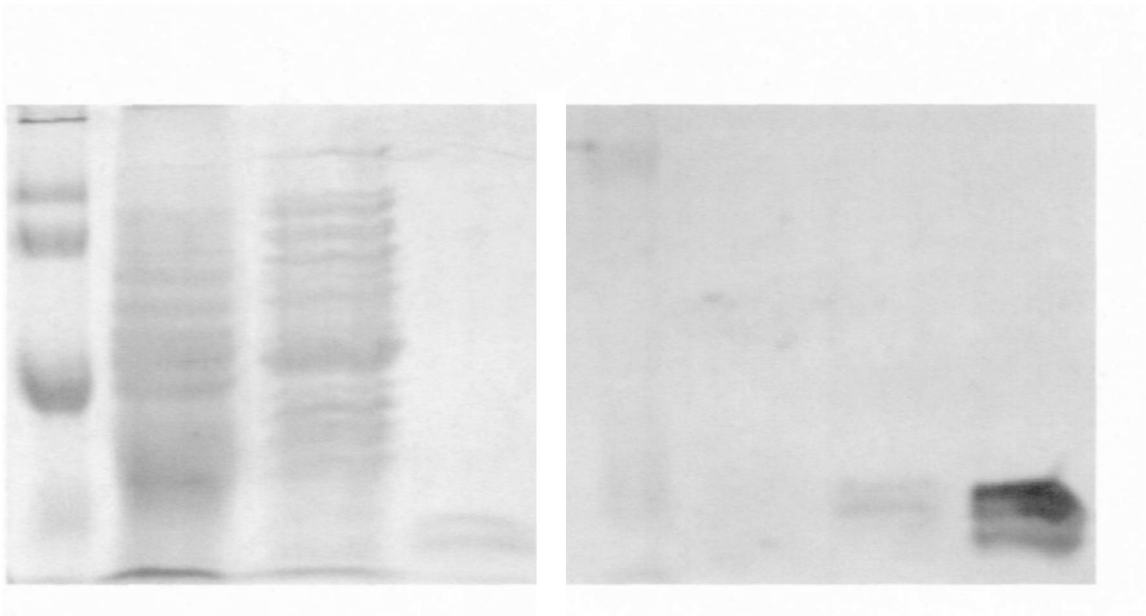


(a) (b)

Figure 3.9: (a) Western blot of CDO (2) with normal human(3), horse(4), and sheep sera(5) as controls.(b) Dot blots to CDO detected immuno-chemically. (Size standards are listed in the Appendix on page 178)

the gel is a smearing of the band from 25kDa downwards indicating proteolysis (Figure 3.10). The protease inhibitor phenylmethylsulfonyl fluoride (PMSF) was later added at all stages of purification and the experiment was repeated. Figure 3.11 Illustrates the dot blot used to detect which fraction contained CDO during purification and a pure protein using protease inhibitors. Trypsinogen was also loaded as a more accurate size marker. The gel was overloaded on one lane to detect the presence of contaminating proteins. The trypsinogen purchased from Sigma shows some contamination at higher molecular weights and the CDO does not indicating the the affinity purified CDO is more pure than a purified commercial product.

CDO is a protein not easily reduced by β-mercaptoethanol during electrophoresis. It generally appears to run as two bands. This property is more evident when run on a higher percentage acrylamide gel. Figure 3.12 illustrates that CDO runs as a single band when β-mercaptoethanol is left out of



(a) SDS-PAGE (b) Western Blot

Figure 3.10: (a)SDS-PAGE and (b)Western blot analysis of human liver homogenate; flow through of homogenate of induced BL21(DE3); and affinity purified CDO **without using protease inhibitors**. (Size standards are listed in the Appendix on page 178). Four lanes are indicated in both gel and blot; starting with the size marker on the left.

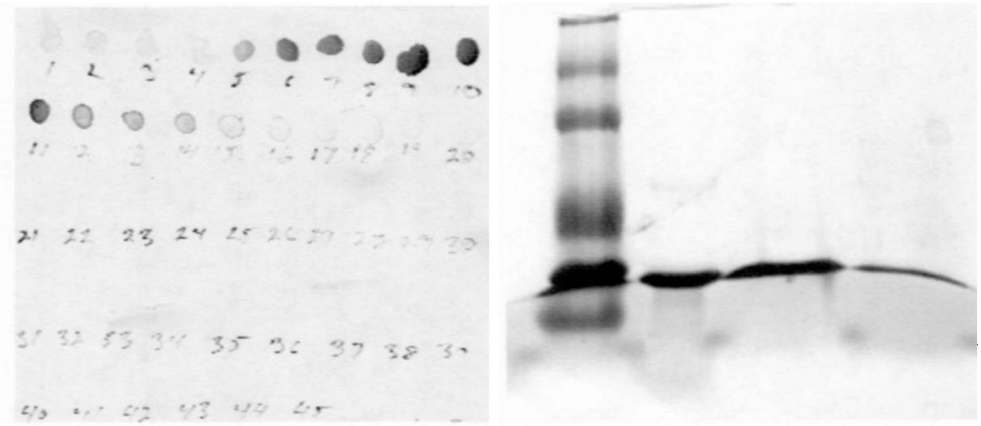
Time (min.)	$OD_{\lambda=280}$
0	0.00
127	0.15
200	0.30
222	0.43
245	0.60 (IPTG added)
290	0.90
330	1.11

Table 3.1: Growth rate for BL21(DE3) transfected with pET15b-hCDO in SOB media at $T_{37°}$.

the gel loading buffer. The second band is not a dimer as it runs as a comparable molecular weight to CDO. The differences are likely due to changes in the protein's hydrodynamic radius and not due to degradation. Since CDO runs as a single band in the absence of β-mercaptoethanol, the change is likely due to disulfide bond stabilisation of the protein. A single band never seems to be present with β-mercaptoethanol regardless incubation time and heating. β-mercaptoethanol has been suggested to be a substrate for CDO. This indicates that at least two of the four cysteine residues in the protein are involved in disulfide bridge formation.

It was also desirable to get an approximate growth curve for CDO in SOB media. In a 250ml inoculated flask. The absorbance of the medium ($OD_{\lambda=280}$) was taken at different time intervals while it was incubating at 37° C on a rotary shaker (Table 3.1). The data were then plotted (Figure 3.13) and fitted by a second order polynomial (Equation 3.1). This equation was used to approximate the time when the BL21(DE3) cell line reached the cell cycle for induction.

$$ABS = 5.997544 \bullet 10^{-3} - 8.176101 \bullet 10^{-4}t + 1.284495 \bullet 10^{-5}t^2 \qquad (3.1)$$



(a) dot blot western of CDO

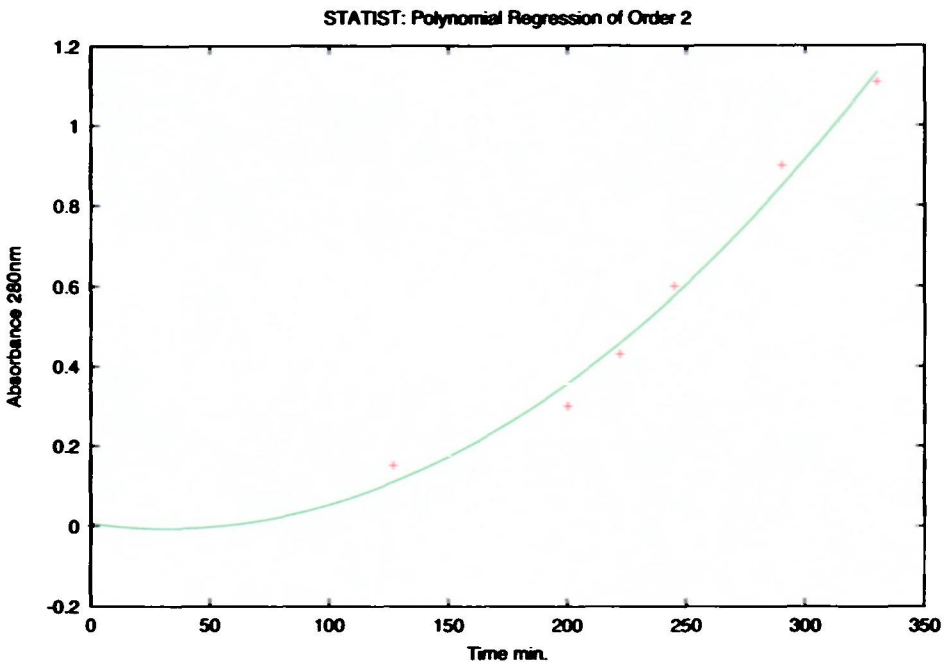
(b) Overload of CDO on SDS-PAGE

Figure 3.11: Dot blot of CDO eluted from a nickel column. Gel overload of trypsinogen, CDO, and a lower concentration of CDO(Size standards are listed in the Appendix on page 178)

(a) Dot blot of elute at pH 6.5 and 7.5

(b) CDO, CDOr6.5, CDOr7.5, Lysate, CDOr++

Figure 3.12: Dot blot of elute at pH 6.5 and pH 7.5. A 15% SDS-PAGE of CDO without β-mercaptoethanol (6.5), CDO with β-mercaptoethanol, flow-through from the column, and concentrated CDO 1 week with β-mercaptoethanol.(Size standards are listed in the Appendix on page 178)



(a) A) time(min.) B) Absorbance (280nm)

Figure 3.13: Growth curve for BL21(DE3) in SOB media

Growth curves were not done for CDO in minimal media. Growth rates for minimal media were greatly depressed. All cultures for expression were started with an overnight inoculum. The growth cycle of this initial inoculum was not easily controlled in minimal media. Owing to this problem, growth curves in minimal media showed too much variation to be useful for expression experiments.

Chapter 4

Physical Characterisation of CDO

Once a system was developed that produced a steady supply of CDO , attempts were made to characterise the protein structurally. Computational analysis of CDO (see Appendix A, page 149 to 165) had given a theoretical model. Experimental techniques were necessary to test whether CDO fit the predicted model.

4.1 Aims

To measure the mass of the recombinant protein using mass spectroscopy. Explore the use of surface plasmon resonance to accurately determine the concentration of CDO in a non-destructive manner. Perform an amino acid decomposition of CDO as a crude method of confirming its sequence and as a second method of protein concentration determination. To analyse the secondary structure of CDO using circular dichroism and measure the overall stability of the protein using dynamic light scattering (DLS).

4.2 Methods

4.2.1 Mass Spectroscopy

Matrix assisted laser desorption time of flight mass spectroscopy (MALDI-TOF MS, MALDI) is a technique that allows the precise mass of a protein to be measured. Other mass spectroscopic techniques are either unable to fully ionise a protein or are too destructive to measure the entire protein. It works by suspending a protein in some matrix, ionising the protein with a laser, and measuring the time it takes for the protein to reach a detector. Inertial differences between proteins are resolved by the time-of-flight to the detector; this allows the masses to be measured extremely accurately.

Protocol

Affinity purified CDO was de-salted using Millipore $ZipTip_{C4}$, a small scale C4 reverse phase packing material in a pipette tip. The protocol for desalting was taken directly from the Millipore $ZipTip_{C4}$ guidelines.

Affinity purified CDO was flash frozen with liquid nitrogen in a Pyrex tube and lyophilised overnight. The freeze dried protein was partially resuspended in 1.6M GuHCl/0.5%TFA made from a 5x stock solution. The $ZipTip_{C4}$ was equilibrated with $10\mu\ell$ of 70% acetonitrile (chromatography grade) and then re-equilibrated with two $10\mu\ell$ volumes of 0.1% TFA (mass spectroscopy grade). CDO was bound to the C4 packing material with 10 repeated aspirations from a protein suspension in 1.6M GuHCl/0.5%TFA. The bound CDO sample was de-salted with 10 x $10\mu\ell$ repeated aspirations of 5%MeOH/0.1%TFA. The protein was eluted from the C4 matrix by recycling the same $4\mu\ell$ solution of 75% acetonitrile/0.1% TFA for 10 cycles. The $4\mu\ell$ sample was submitted

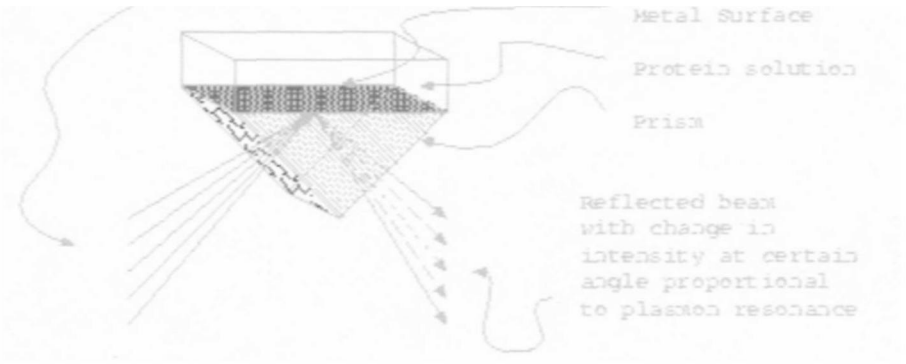


Figure 4.1: Practical theory of SPR

to the University of Birmingham chemistry department for molecular mass determination by MALDI-TOF.

4.2.2 Surface Plasmon Resonance

Surface plasmon resonance (SPR) is a technique to measure the resonance of surface plasmons and this was used to determine protein concentration quickly, accurately, and without wasting/changing the sample. Plasmons are waves that describe charge densities fluctuations of a metal at some frequency. Detecting these resonant frequencies can be done by directing a beam of light at some angle toward a metal surface, a thin gold film in these experiments, and then detecting a change in intensity of the reflected light. Plasmons are created on the opposite surface of the gold film; the surface in contact with a protein solution. Resonance of this surface occurs when vector components of the incident wave match that of the plasmon. When this occurs, there is a decrease in the reflected intensity on the opposite surface. These resonances vary with changes in the solution at the interface and are very sensitive to changes in the refractive index[117]. Taking advantage of this phenomenon, SPR was used to determine protein concentration.

Protocol

A Biacore 3000 instrument was used for all SPR measurements. A Pioneer Sensor Chip J1 was used for all measurements of refractive index (Biacore AB, Uppsala, Sweden). This system is normally set up to measure binding of small molecules. In these experiments, an unmodified gold surface (Pioneer J1 chip) was used with the assumption that it would be chemically inert. The output data for the Biacore unit is in terms of resonance units (RU). However, there is a linear (one-to-one) relationship between the refractive index (n) and the surface plasmon resonance. Salt and protein standard gradients were first tested both to establish the linearity between SPR and n; to test how prone a gold surface is to protein adsorbtion; and to develop a precise value for the change in concentration of CDO with a change in RU ($\frac{\partial RU}{\partial c}$). With a known value of the change in refractive index with concentration ($\frac{\partial n}{\partial c}$.) it is possible to calculate a protein concentration.

$$n = k \cdot c \cdot \left(\frac{\partial n}{\partial c}\right) \tag{4.1}$$

$$c = \frac{n}{k \cdot \left(\frac{\partial n}{\partial c}\right)} \tag{4.2}$$

A more detailed description of the method with results can be found in a paper submitted for publication (Appendix page 182)

4.2.3 Amino Acid Decomposition

An amino acid decomposition is a qualitative way to access the relative ratios of individual amino acid residues in a protein. With the advent of protein sequencing, this method is not commonly used any more to characterise pri-

mary sequence. The method is however the most effective way to determine a protein concentration. The results of experiment are in terms of concentration. For a protein of known primary sequence, it is a trivial task to convert from the residue concentration to the overall protein concentration.

In the case of cysteine dioxygenase, it became apparent that it was not easy to reproducibly measure the CDO concentration due to irreversible aggregation. The predicted molar extinction coefficient was also different than the actual molar extinction coefficient. The amino acid decomposition is necessary to get a starting concentration with which to base an actual extinction coefficient. One drawback is that Asn and Gln are converted to Asp and Glu; and Thr and Ser have 5% and 10% losses in the hydrolysis process. In these experiments, this is not a concern since the other amino acids can be referenced and an accurate $\frac{\partial n}{\partial c} \cdot \frac{\partial RU}{\partial c}$, and molar extinction coefficient could be developed.

Protocol

Amino acid decomposition involves the complete hydrolysis of the protein into its individual amino acids by boiling it in HCl. The resultant mixture is then passed through a reverse phase HPLC system. The resultant peaks from the separation of hydrolysed product are then integrated to give a percentage breakdown of each amino acid. It is also the most accurate way available to biologists to determine a proteins concentration. Other methods require a standard, such as lysozyme or BSA, which may not accurately reflect the protein of interest or it may consume too much precious sample.

4.2.4 Dynamic Light Scattering

It is absolutely critical for structural studies, in X-ray crystallography and NMR, to have a mono-disperse protein sample. Dynamic light scattering is by far the most versatile technique to access the dispersity of a protein. It is a way of measuring the brownian motion of a particle (CDO) in solution. There is an empirical relationship between the diffusion coefficient of a particle and its hydrodynamic radius. This technique allows one to deduce the volume of a particle. If a protein is aggregating, one would expect to see an increase in its volume with time.

DLS measures the diffusion coefficient (D) of a particle. The diffusion coefficient is related to the temperature (T) and the frictional coefficient (f) where k is the Boltzmann constant. The frictional coefficient is related to viscosity (η) and the radius (R). The entire relation is explained by the Stokes-Einstein equation (Equation 4.3)

$$\begin{aligned} D &= \frac{kT}{f} \\ &= \frac{kT}{6\pi\eta R} \end{aligned} \tag{4.3}$$

No special care need to be taken to present a sample to the instrument save filtering it. It works over a wide range of concentrations and is sensitive to particles as small as 1nm to several microns. A single measurement can be done in less then a half hour which makes this an unparalled technique to screen conditions that may be amenable to a protein. The only drawback is that it is more sensitive to larger particles in solution at the exclusion of detecting the smaller monomeric protein polymers. Thus, this technique should

be used more qualitatively than to infer real physical data about the sample.

Protocol

Dynamic light scattering (also known as quasi-elastic light scattering of photon correlation spectroscopy) is a spectroscopic means of measuring the diffusion coefficient of a particle (CDO). The radius of the protein can then be approximated via the Stokes-Einstein equation (Equation 4.3). In a typical experiment, a laser beam passes through a sample and some incident light is scattered by the proteins in suspension (elastic scattering). The scattered light is detected by a photomultiplier tube (PMT) at some angle relative to the incident beam (90° in these experiments). The photon counting PMT measures changes in the intensity of the reflected photons to establish a correlation curve. This is an automated process and in simplified terms, the the correlation curve determines the brownian motion of the proteins/particles and references a diffusion coefficient to each species of particle in solution.

$$R_m = \frac{kT}{6\pi\eta D}$$

With a measured value for the diffusion coefficient (D) and solving equation 4.3 for R yields a value for the radius of a particle. This equation is only concerned with the radius of a perfect sphere though. For asymmetrical particles like proteins, the apparent radius can be much larger (Figure 4.2) and require a perin factor take the asymmetry into consideration.

Cysteine dioxygenase measurements were largely qualitative to determine stability of the protein under different buffer conditions. It is also very effective in quantitating a degree of aggregation in solution for establishing limits on salinity and protein concentration.

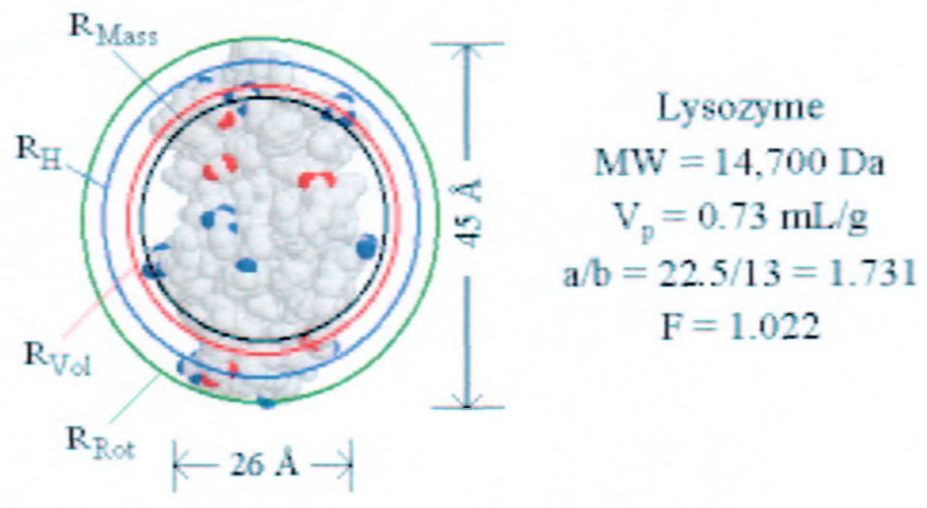


Figure 4.2: Lysozyme. (Image taken from Protein Solutions, Inc.)

A Zetasizer 3000HS (Malvern Instruments, Ltd.) was used for all DLS experiments. This was equipped with a 532 nm laser that brought the sensitivity of the instrument to at least 2 nm particle size.

4.2.5 Rheology

Rheology is the study of viscosity and was used to make some of the DLS measurements more quantitative. A digital plate viscometer was used to quantitate the viscosity of samples. There are several empirical relationships between viscosity and the physical state of a compound. In the instance of a protein, viscosity in itself yields little useful information. However, it is a necessary variable to refine to describe more accurately other physical quantities.

Because of fluctuations in the measured viscosity with temperature and instrument, relative viscosities (η_r) were measured (where η is the measured quantity and η_o is the viscosity of water under the same conditions used to measure a buffer):

$$\eta_r = \frac{\eta}{\eta_o}$$

The relative viscosity η_r provides a more accurate measurement of the actual quantity when other variables, as temperature, are not easily controlled. Often it is the viscosity of the protein that is of interest. Because a plate viscometer can induce shear and cold deform a biopolymer in solution. Specific and characteristic viscosities had to be quantitated indirectly via the Stokes-Einstein Equation (Equation: 4.6) where the diffusion coefficient, D, was measured from light scattering experiments.

$$\eta_s = \frac{\eta - \eta_{buffer}}{\eta_{buffer}} \tag{4.4}$$

$$\eta_c = \frac{\eta - \eta_{buffer}}{\rho \cdot \eta_{buffer}} \tag{4.5}$$

$$D = \frac{kT}{f} = \frac{kT}{6\pi\eta R} \tag{4.6}$$

The viscosity measurements provide some preliminary background parameter for dynamic light scattering. For η_s and η_r, and accurate protein viscosity, η, is acquired by measuring a diffusion coefficient of a latex microsphere of known radius, R, in a protein solution. Solving the Stokes-Einstein equation for η will yield an accurate measurement of the viscosity of CDO without inducing shear. It is usually not possible to directly measure the viscosity of polymers (and biopolymers) because the measuring process induces shear, alters their structure, and hence alters their η_c under ambient conditions.

4.3 Results and Discussion

4.3.1 Spectroscopic Determination of CDO Produced by the Bacterial Expression System

An initial sample of CDO in standard buffer failed to ionise when it was submitted for electrospray mass spectroscopy. Instead, a CDO sample was purified as previously explained and prepared in acetonitrile using a ZipTip^{TM}. The sample was submitted to the mass spectroscopy facility at the Chemistry Department at the University of Birmingham for MALDI-TOF. The spectrograph of the sample (Figure 4.3) indicated two masses, a singly and doubly charged species of CDO. Actual and measured masses for both the monoisotopic and distributed isotopic averages are listed in table 4.1.

There was a difference in mass of 147 Da between the measured mass and the predicted mass of CDO for the singly charged species. Although this difference could be explained by the addition of an amino acid residue into the protein, DNA sequence analysis indicated otherwise. Furthermore, a change of 147 Da for a protein of this size is within the margin of error for MALDI-TOF. The likely explanation is that there was a slight calibration error of the instrument.

	Expected	Measured	Δm
Monoisotopic ($m^+ + H$)	25,132.2298 Da	n/a	n/a
Average($m^+ + H$)	25,148.38 Da	25,295 Da	$\sim$ 147 Da
Monoisotopic($m^{2+} + 2H$)	12,566.6188 Da	n/a	n/a
Average($m^{2+} + 2H$)	12,574.69 Da	12,576 Da	$\sim$ 1 Da

Table 4.1: Calculated and measured MALDI masses for recombinant CDO with both singly and doubly charged species.

When a phenylalanine is inserted into the presumed CDO sequence, the

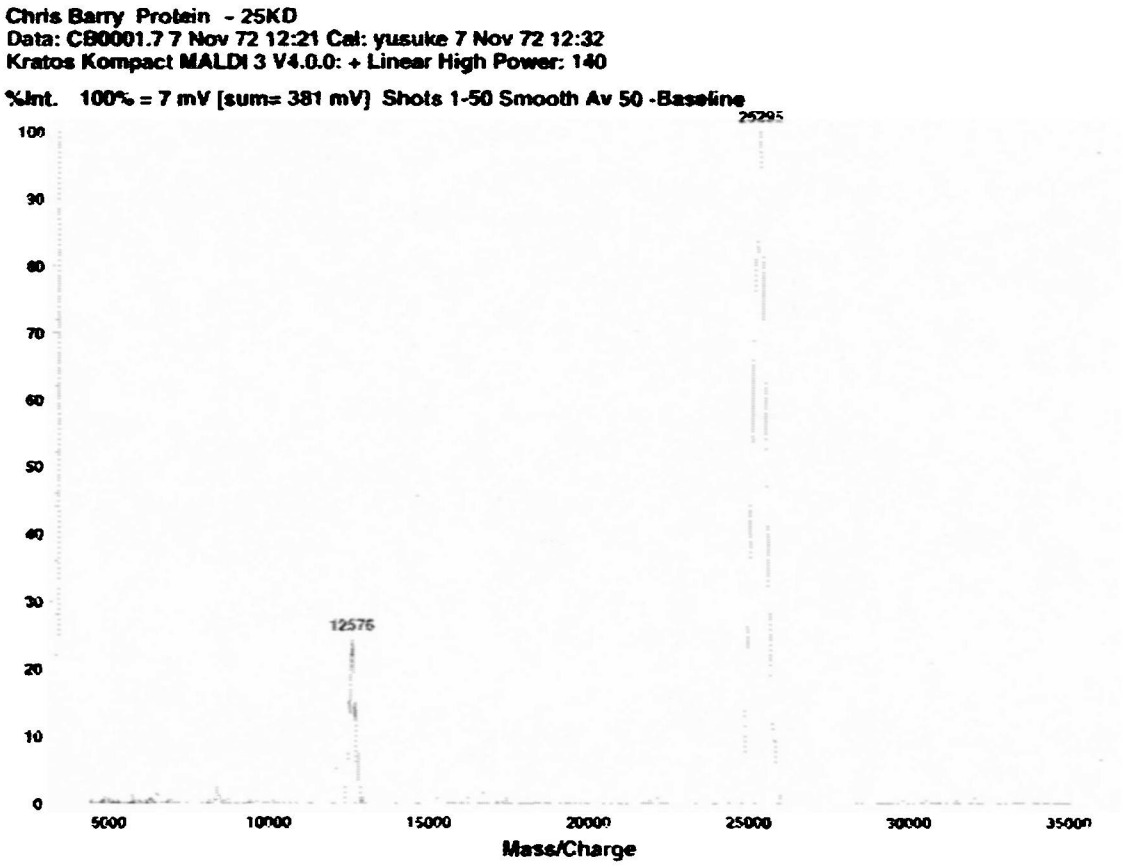


Figure 4.3: MALDI-TOF of CDO

mass is recalculated to be 25,295.56 Da. This is nearly the precise mass measured for the recombinant CDO. The likely location of the additional amino acid is in the fusion protein.

4.3.2 Concentration Determination of CDO using SPR

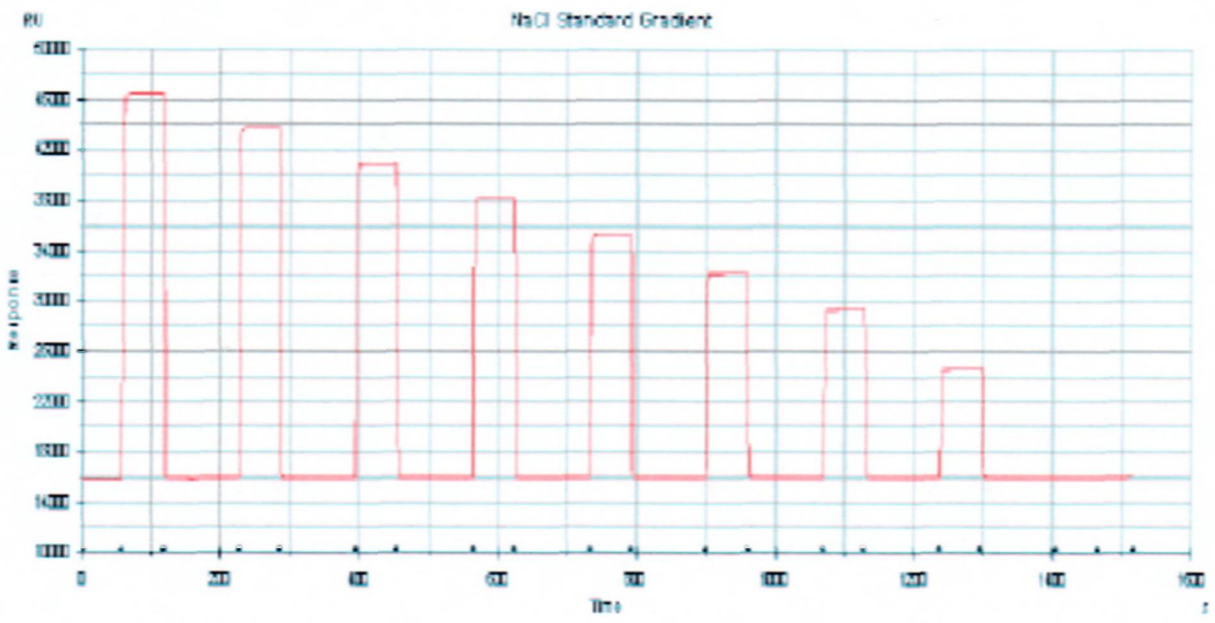
The interpretation of the data for some experiments was dependant upon having an accurate value for the concentration of the protein. It was also desirable to fulfil this requirement in a non-destructive manner. It was known that this could be done by measuring the refractive index of the sample. However, refractometry is not very responsive to changes in protein concentration and is thus very prone to error. Surface plasmon resonance (SPR) was tested to evaluate the concentration of CDO.

The theory first had to be tested on a salt gradient to confirm the relationship of SPR with refractive index. In order to calibrate the Biacore instrument, a sodium chloride standard was used (Figure: 4.6). Sodium chloride has an established index of refraction (n) at different concentrations (c) and has a $\frac{\partial n}{\partial c}$ of 1.854E-3$\frac{ml}{mg}$. Aqueous sodium chloride was measured with SPR (in Resonance units, RU) and the value was compared with known refractive indices. The calculated $\frac{\partial RU}{\partial c}$ was measured to be 1,617,586 $\frac{ml}{mg}$.

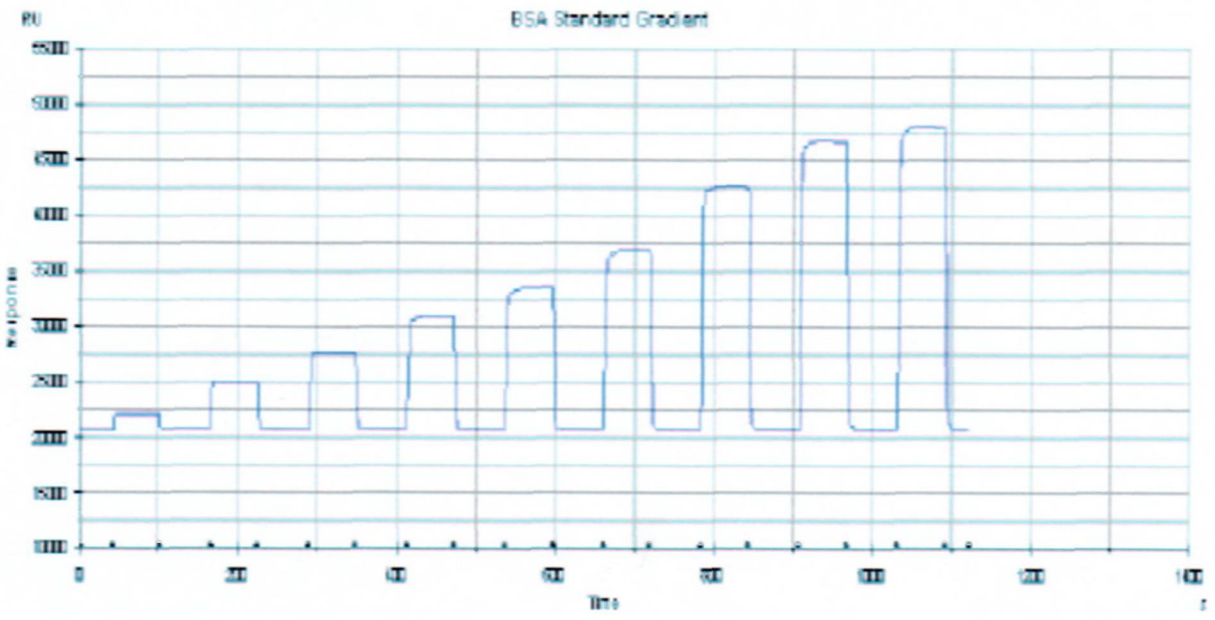
$$n = k_n \cdot c \cdot \frac{\partial n}{\partial c} \longrightarrow RU = k_{RU} \cdot c \cdot \frac{\partial RU}{\partial c}$$

$$c = \frac{RU}{k\left(\frac{\partial n}{\partial c}\right)}$$

This procedure was then repeated with two protein gradients, bovine serum albumin (BSA) and lysozyme (Figures 4.7 & 4.8). This procedure was done in



(a) SPR data of NaCl gradient.



(b) SPR data of BSA gradient.

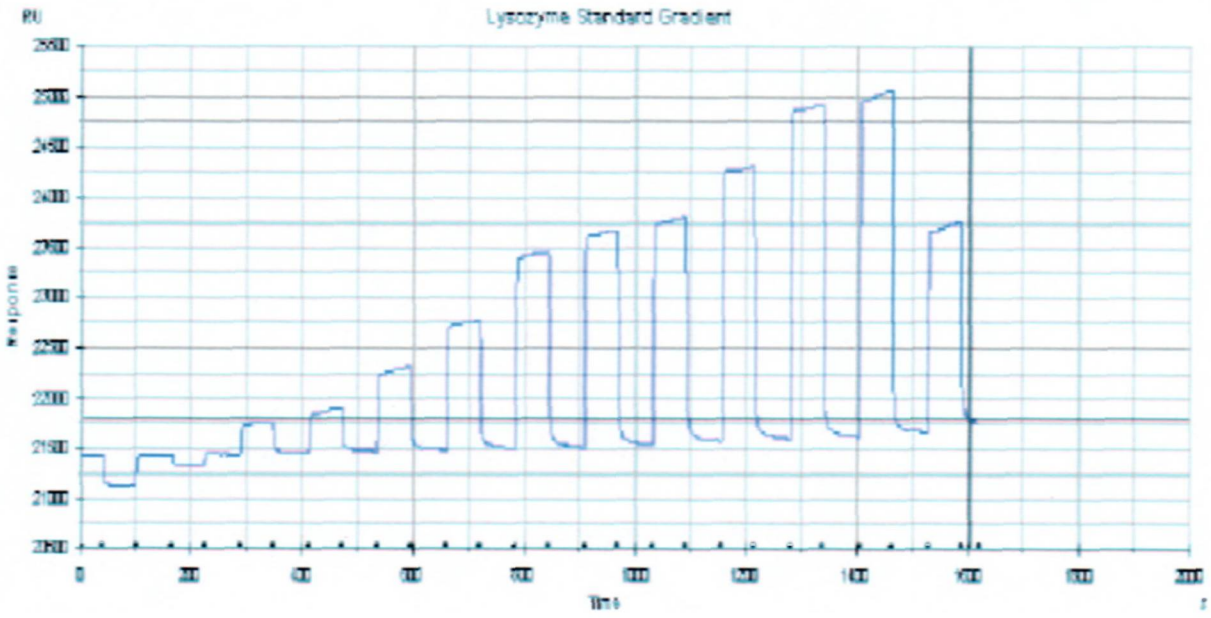
Figure 4.4: Acquired SPR data. (Tabulated data listed on following pages)

order to test the validity of SPR theory with respect to protein concentration determination. The calculated $\frac{\partial RU}{\partial c}$ for BSA was found to be 2676.298 $\frac{ml}{mg}$ and the $\frac{\partial RU}{\partial c}$ for lysozyme was found to be 421.632 $\frac{ml}{mg}$.

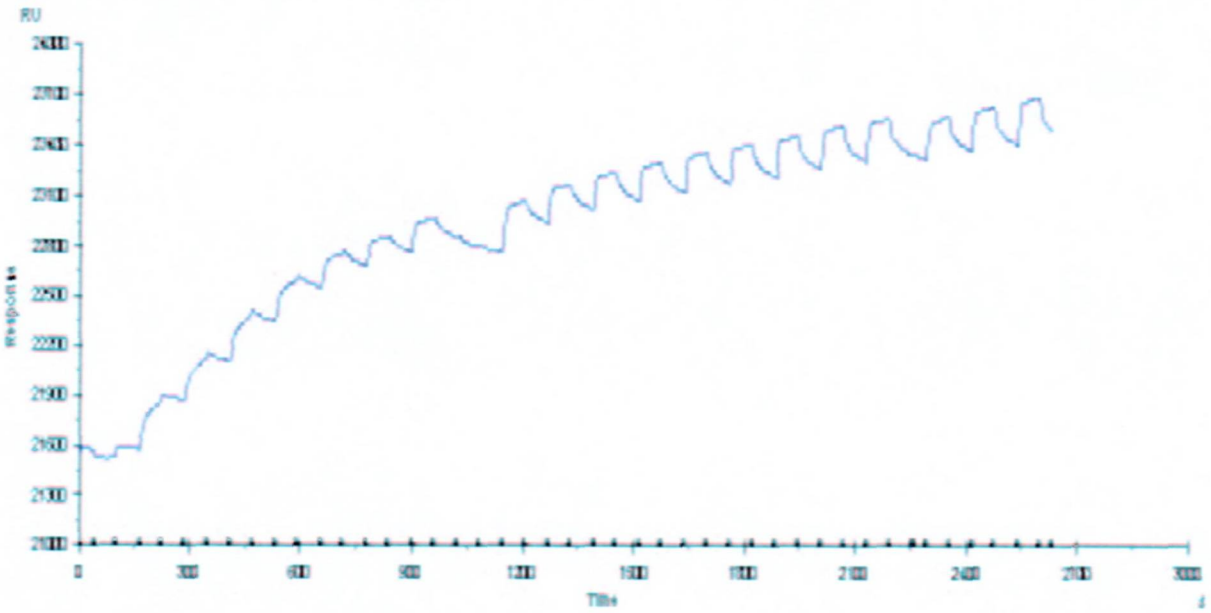
	$\frac{\partial RU}{\partial c}$(ml/mg)
NaCl	1,617,586
BSA	2,676.298
Lysozyme	421.632
CDO	∞

Table 4.2: Calculated values of $\frac{\partial RU}{\partial c}$ for standard gradients and CDO.

A gradient of affinity purified CDO was then run on the SPR apparatus. Unlike the NaCl, BSA and lysozyme gradients, the CDO gradient did not give a step-wise response in resonance (Figure 4.5). This was due to adsorbtion of the protein to the gold surface of the sensor chip. Addition of L-cysteine to the intermediate washing steps was not effective in returning resonance changes to baseline indicating the nature of binding is not via a sulfhydryl group. It is not clear whether this affinity is from the fusion protein or CDO itself. Since CDO binds iron and other metals, it may indicate the recombinant protein retains it ability to chelate metal. It may also indicate that CDO has an affinity for the adsorbed L-Cysteine on the gold surface. This was an unexpected result and made the method inappropriate for concentration determination of CDO.

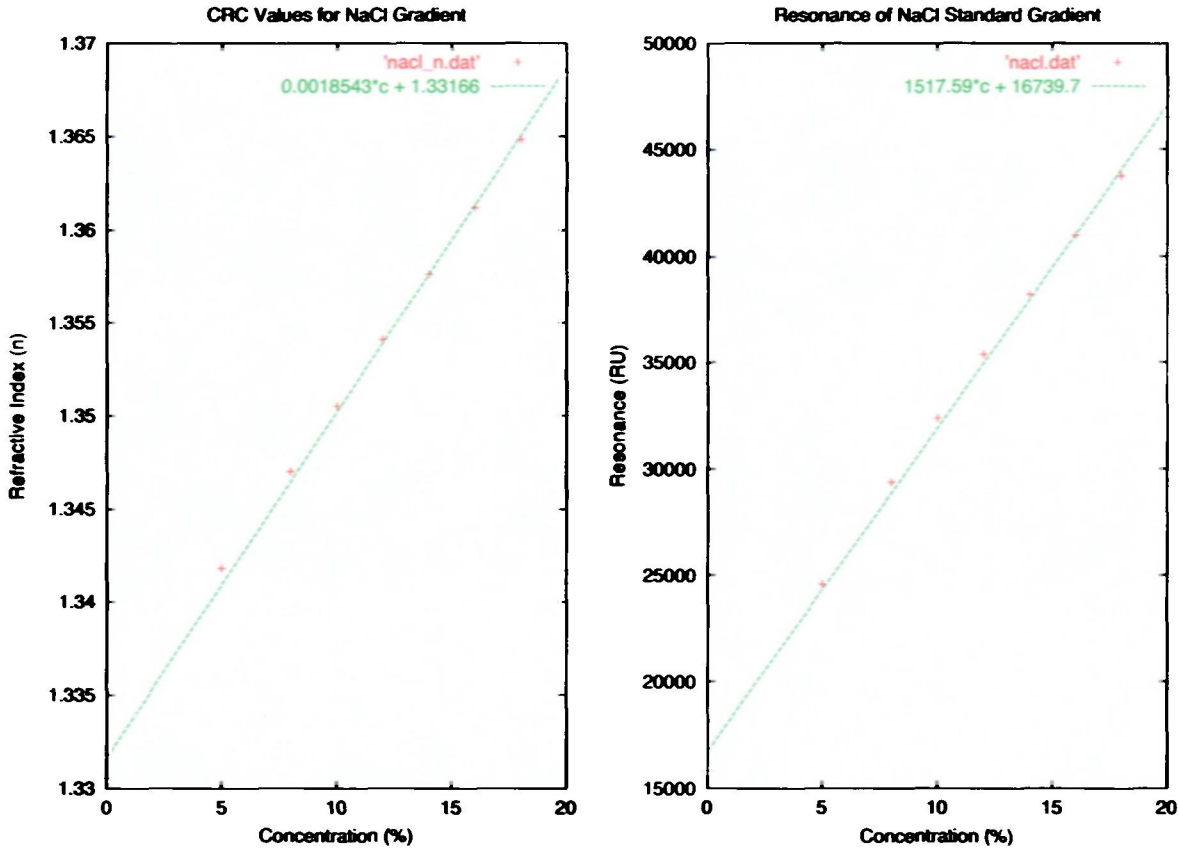


(a) SPR data of lysozyme gradient.



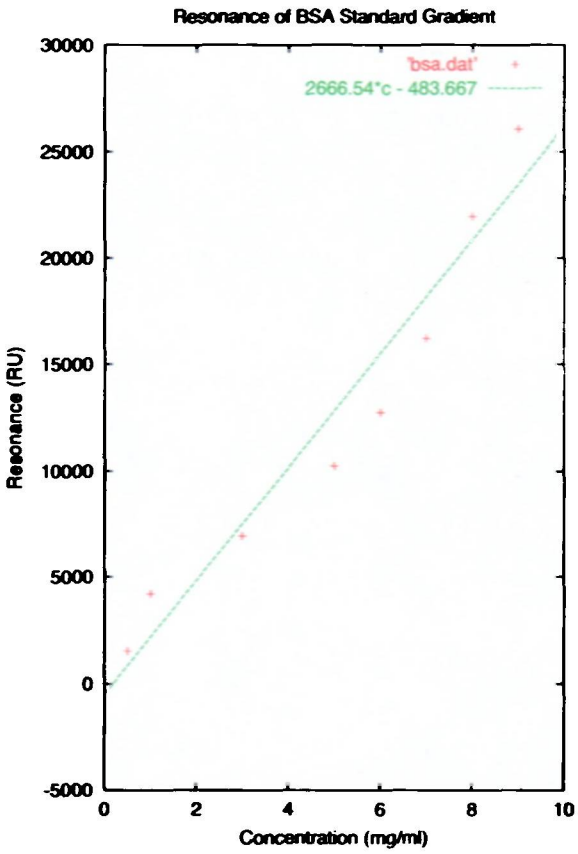
(b) SPR of a CDO standard gradient

Figure 4.5: Acquired SPR data. (Tabulated data listed on following pages)



Concentration (%)	Refractive Index (n)	Resonance Units (RU)
0	1.33054	15826.7
5	1.3418	24576.7
8	1.347	29378.6
10	1.3505	32364.2
12	1.3541	35369.4
14	1.3576	38229.1
16	1.3612	40983.6
18	1.3648	43757.5
20	1.3684	46483.1

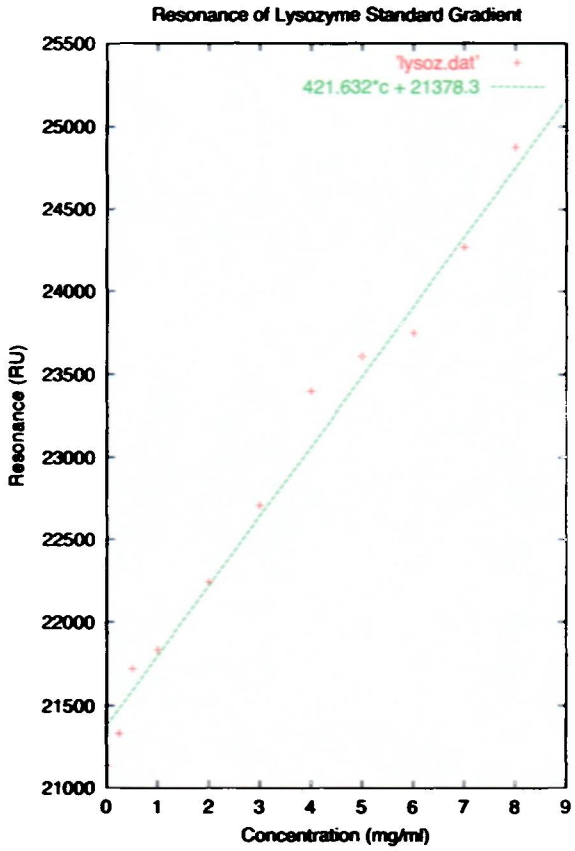
Figure 4.6: SPR of NaCl Standard Gradient



(a) SPR plot of BSA standard

Concentration	RU	Baseline RU	ΔRU
0	20669	20669	0
0.5	22192.4	20669.8	1522.6
1	25092.2	20895.9	4196.3
3	27786	20850.2	6935.8
5	31071.1	20829.4	10241.7
6	33575.8	20822	12753.8
7	37009.3	20811.5	16197.8
8	42765.6	20816.6	21949
9	46882.2	20820.3	26061.9
10	48146	20847.6	27298.4

Figure 4.7: BSA standard gradient



Concentration (mg/ml)	Resonance (RU)
0	21134.1
0.250	21329.6
0.5	21720.0
1	21833.2
2	22241.2
3	22704.5
4	23399.4
5	23607.3
6	23746.9
7	24263.8
8	24874.1
9	24975.4

Figure 4.8: Lysozyme standard gradient

4.3.3 Determination of Amino Acid Composition of CDO produced in Bacterial System

The alternative technique to measure the concentration of CDO was spectroscopically. To develop a molar extinction coefficient, an amino acid decomposition was performed. The results are tabulated in Figure 4.3 along with the actual and expected composition of CDO. The composition of CDO is consistent with what was expected based on its sequence. Differences between the actual composition and expected concentration are due to the conversion of Asn and Gln to Asp and Glu respectively during the hydrolysis process; Thr and Ser also have 5% and 10% losses respectively in the hydrolysis process.

Amino Acid	Conc. (nmol/ml)	Composition	Act. Comp.	Exp.Comp.
Aspartic Acid	2000	10.9%	4.545%	10.909%
Threonine	1000	5.5%	5.455%	4.910%
Serine	1400	7.6%	8.182%	7.369%
Glutamic Acid	2000	10.9%	7.273%	10.873%
Proline	780	4.3%	3.636%	4.3%
Glycine	1400	7.6%	6.364%	7.6%
Alanine	1500	8.2%	5%	8.2%
Cysteine	n/a	n/a	1.818%	0%
Valine	1100	6.0%	5%	6%
Methionine	380	2.1%	3.636%	2.1%
Isoleucine	750	4.1%	4.545%	4.1%
Leucine	1700	9.3%	8.182%	9.3%
Tyrosine	350	1.9%	2.727%	1.9%
Phenylalanine	710	3.9%	4.091%	3.9%
Histidine	1300	7.1%	8.182%	7.1%
Lysine	980	5.3%	5.455%	5.3%
Arginine	980	5.3%	4.545%	5.2%
Glutamine	0	0%	3.636%	0%
Asparagine	0	0%	6.364%	0%
Tryptophan	0	0%	1.364%	0%

Table 4.3: Initial amino acid decomposition of CDO by Alta Bioscience. (Alta code A4768, CDO B3) with measured composition, actual composition, and expected measured composition.

An extinction coefficient was developed based upon the data from the decomposition. It was later discovered that spectroscopic concentration measurements using the calculated extinction coefficient did not appear to be consistent. Soon afterward it was discovered that CDO aggregated under most conditions and may explain the lack of consistency. It was not logistically possible to perform an amino acid decomposition for each sample to accurately determine protein concentration. Since the decomposition is a long and expensive procedure, concentrations were approximated using the theoretical molar extinction coefficient based solely upon its sequence under mono-disperse conditions (Explained under Dynamic Light Scattering).

4.3.4 Aggregation State of Bacterially Produced CDO using Dynamic Light Scattering

Dynamic light scattering (DLS) was used to measure the aggregation state of CDO. A freshly purified sample eluted with high imidazole and salt was measured. A measurement was taken every two hours. Figure 4.9 illustrates the results of each measurement. Each coloured plot represents a measurement at a different time. The figure clearly shows a progression in the size of the protein from about 4-5nm to 10μm over a 24 hour time period at room temperature.

The mass of CDO from the MALDI data indicated it was 25,295 Da. The predicted size of a perfect sphere of mass 25,295 Da is 2.4381 nm (based on prediction from Protein Solutions molecular weight calculator). The masses measured in figure 4.9 and in all DLS experiments are based upon the diffusion coefficient (D). Calculation of the hydrodynamic radius is based upon temperature (T) and viscosity (η) where k and π are constants.

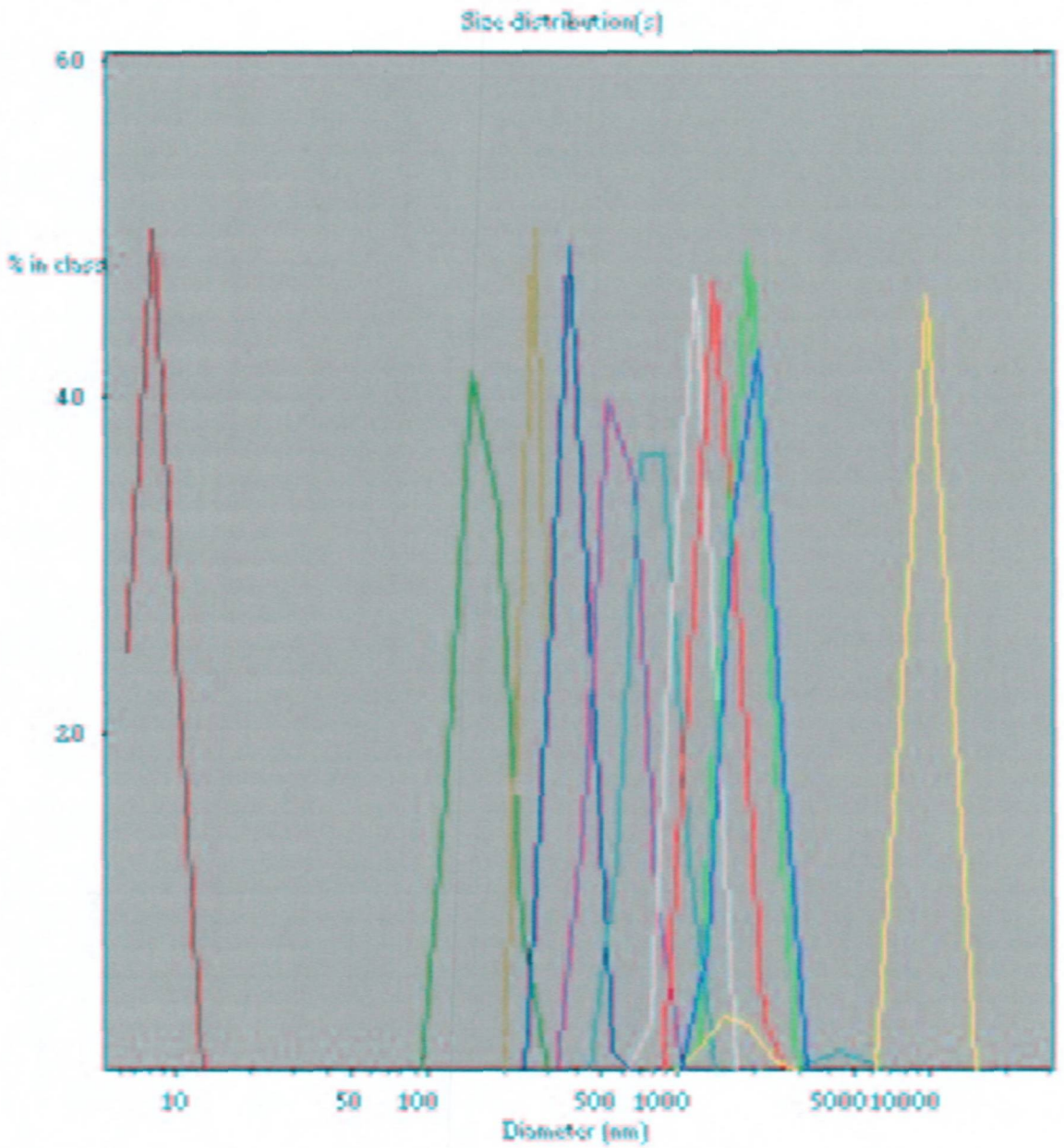


Figure 4.9: Measurement of time dependant aggregation of CDO using dynamic light scattering.

$$R_m = \frac{kT}{6\pi\eta D}$$

Most proteins are asymmetrical so the perfect sphere model is not accurate. They also contain a layer of hydration. Both these characteristics will increase its hydrodynamic radius above that of a perfect sphere. A second constant could have been added to the above equation take symmetry into consideration but, since it was not possible to regulate T and to measure precisely η, it was not practical. The results are more qualitative but clearly illustrate aggregation and is strongly suggestive that CDO is mono-disperse at the time of purification. Since many recombinantly expressed proteins form inclusion bodies, showing that the sample is monodisperse at the time of purification is important; studying its structure is not an intractable problem.

The fidelity of all experimental data was contingent upon finding a buffer that was appropriate for CDO that prevented aggregation. The protein was not stable in any number of standard buffers so DLS experiments were set up, albeit qualitatively, to assay various additives on CDO. It soon became apparent that CDO has a tendency to aggregate (at almost any concentration) over a period of hours to days.

Another important consideration in DLS is that larger particles scatter much more strongly than smaller particles. To compensate for this, Malvern software is able to plot DLS data according to intensity as well as by number. Figure 4.11 illustrates the same measurement of CDO according to number and intensity. This is a monodisperse sample but the presence of a very few larger species skewed the result when viewed by intensity. At some point, if too many larger species are present, detection of the monomer becomes impossible. For all experiments, it was necessary to re-suspend samples in a

starting buffer before aggregation has progressed to this point. The samples were then aliquotted and frozen until they were ready to be assayed.

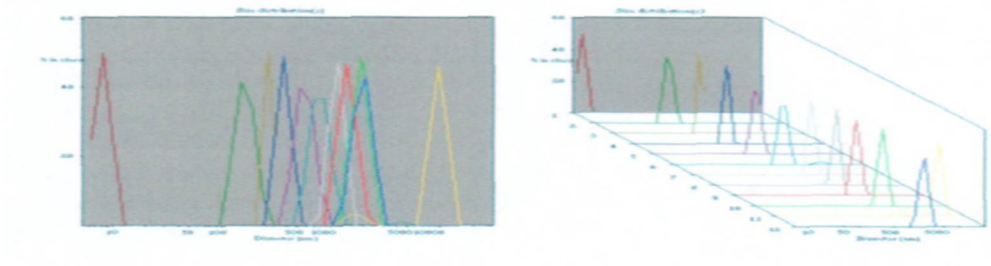
An experiment was set up where freshly purified CDO was dialysed and concentrated for half a day. The solution was then aliquotted and flash frozen. A matrix of additives described in the literature were added in turn to each aliquot and the aggregation state of CDO with time was monitored. Figure 4.10 illustrates the most striking result. It is important to note that the protein had already started to aggregate the moment it was eluted off the column and that at least 12 hours had transpired when the measurements were taken. It is also important to note that the scale for this data on the x-axis is logarithmic, not linear. Everything taken into consideration, there is no question that glycine has a profound beneficial effect on the dispersity of the protein.

It was also found that sodium fluoride (NaF) was more beneficial than sodium chloride in preventing the aggregation of CDO. Imidazole seemed to accelerate aggregation under some of the conditions tested and EDTA was chosen as an alternative in the elution step during affinity purification. A suitable buffer with a pH lower than 7.8 was not found. It was not logistically possible to measure the viscosity of every buffer or combination of buffers. As a result, interpretation of DLS experiments were based upon changes in apparent hydrodynamic radii of samples rather than absolute values. Addition of glycerol to a monodisperse sample illustrates the impact viscosity can have on the apparent size of CDO (Figure 4.12). Subtle changes in a buffers pH, ionic strength, ion type, etc. can have an impact on the buffers viscosity. Table 4.4 are a few measured viscosities. Since particle sizes are calculated from $R_m = \frac{kT}{6\pi\eta D}$. Any miscalculation of viscosity will be exaggerated 6πD times in the size calculation.

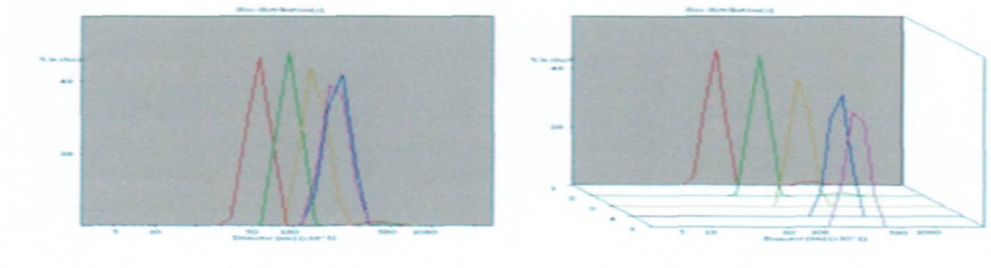
Buffer	$\frac{\eta}{\eta_0}$	η
100mM NaF, 20mM Gly, 10mM PO_4 pH 8	1	.8904
100mM NaF, 20mM Gly, 10mM PO_4 pH 7.4	1	.8904
100mM NaF, 20mM Gly, 10mM PO_4 pH 7.0	.9950	.8859
100mM NaF, 20mM Gly, 10mM PO_4 pH 6.5	.9847	.8768
200mM NaF, 25mM Gly	1.0827	0.9900
200mM NaF, 50mM Gly, 5% Glycerol	1.2052	1.0731
200mM NaCl, 1mM Gly,100mM PO_4, pH 7.5	1.0205	.9087
above ↑ + 1μM $FeSO_4$	1.0205	.9087

Table 4.4: Measured viscosities for some buffers used for DLS experiments

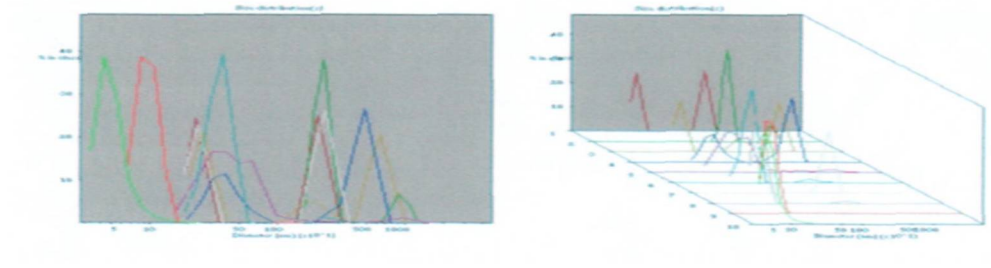
To test the results more quantitatively, glycine was added to every solution (1mM) during and following the purification step. Figure 4.11 shows the result. The discovery that glycine prevents aggregation was very fortuitous. Finding conditions that keep a protein mono-disperse is usually the rate limiting step to determining the tertiary structure by NMR and in growing protein crystals. The final preparatory step necessary for structural characterisation is to reproducibly concentrate the protein without aggregation and to determine the point of maximal concentration while preserving the dispersity. The most appropriate method to concentrate the sample was to elute CDO at high concentration EDTA, in a well packed column, and to use the most concentrated fraction. 400mM Sodium fluoride with 10mM Glycine and 10mM Na_2HPO_4 at a pH of 7.8 to 8.0 were the most stable buffers found. Dialysis against polyetheleneglycol to concentrate CDO did not seem to preserve the dispersity of the sample. Dialysis against desicants like sephadex were equally un-beneficial. Centrifugal concentrators were not used for fear of aggregation at the membrane interface.



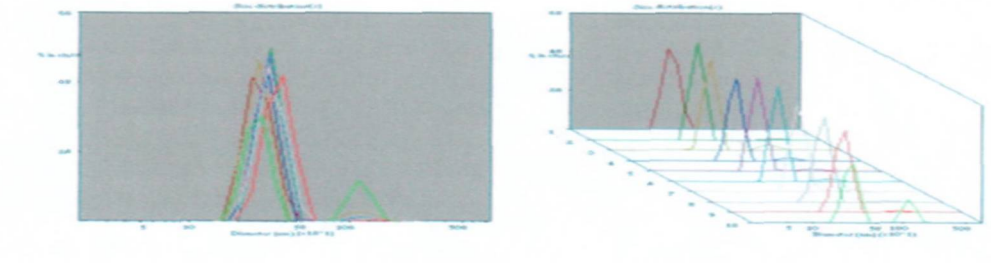
(a) PBS



(b) BME

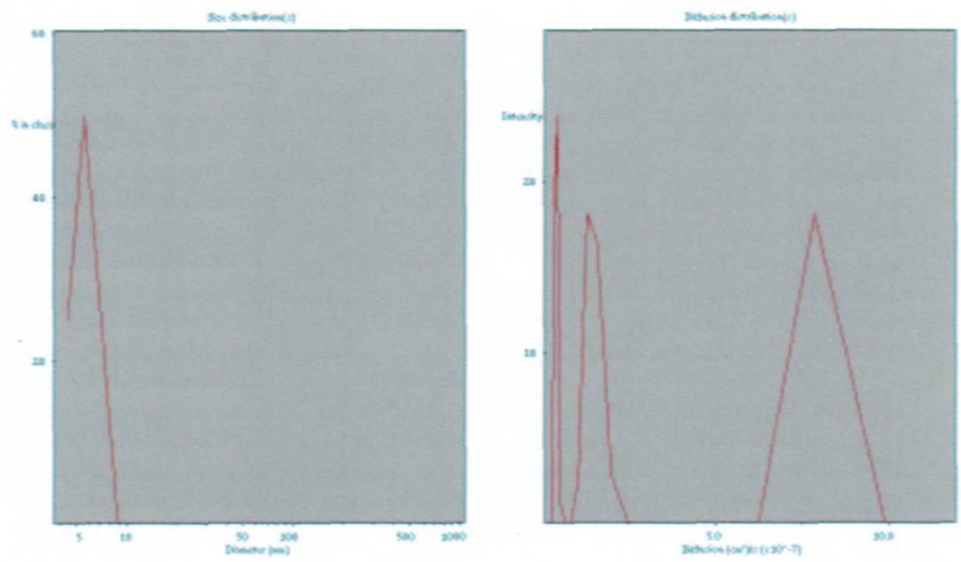


(c) SDS



(d) Glycine

Figure 4.10: Time dependant measurements: (a) CDO in PBS (b) CDO in PBS with 0.01% β-mercaptoethanol (c) CDO in PBS with 0.01% SDS (d) CDO in PBS with 1mM Glycine



(a) Difusion Coeficient by number

(b) Difusion Coeficient by intensity

Figure 4.11: Cysteine Dioxygenase suspended in 25mM Glycine/200mM NaF

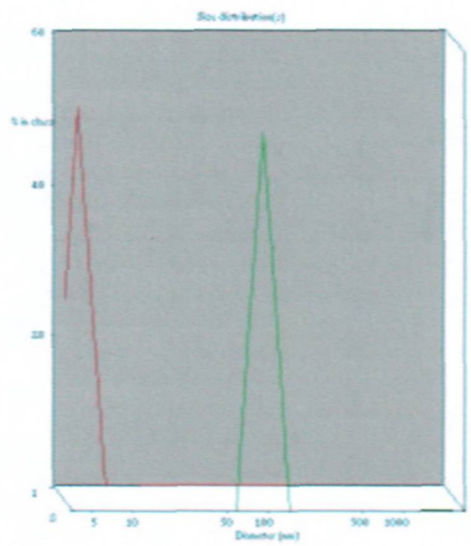


Figure 4.12: Influence of viscosity on particle size. Monodisperse CDO in NaF without (Red) and with 10% glycerol (Green)

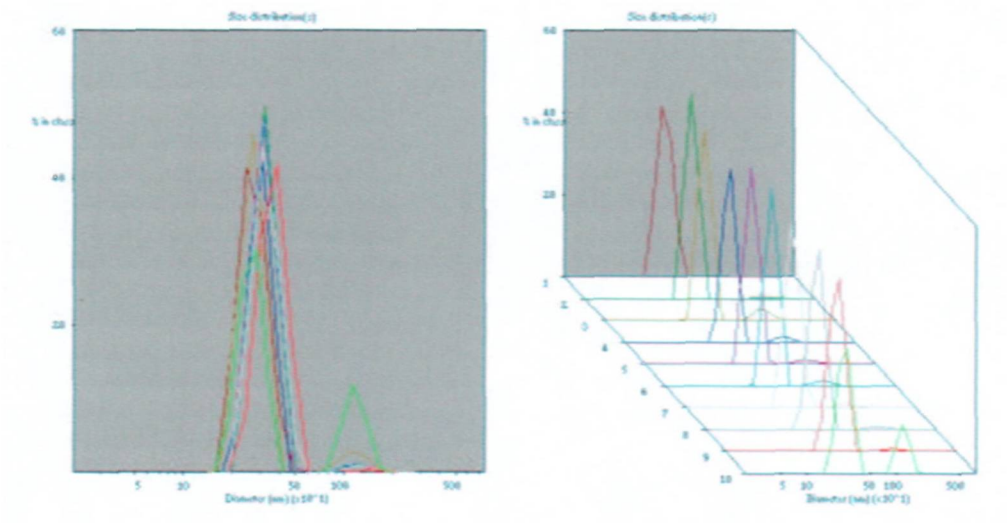


Figure 4.13: Cysteine Dioxygenase in 1mM Glycine/1M NaCl/10mM Phosphate pH 6

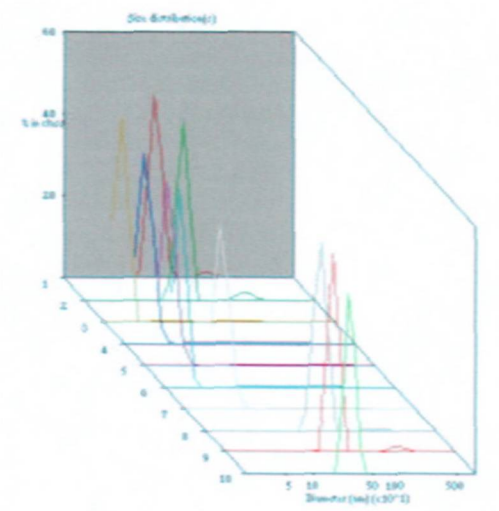


Figure 4.14: Cysteine Dioxygenase in 1mM Glycine/1M NaCl/10mM Phosphate/excess iodoacetic acid pH 6

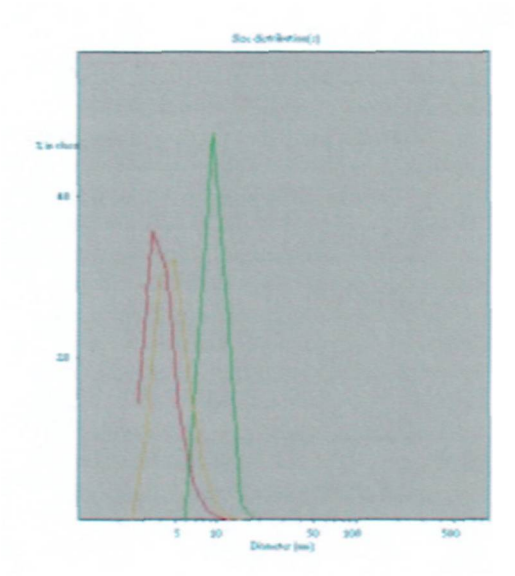
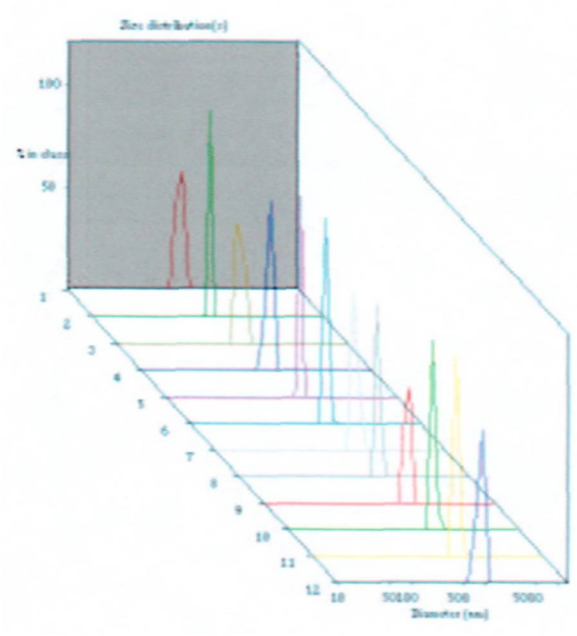
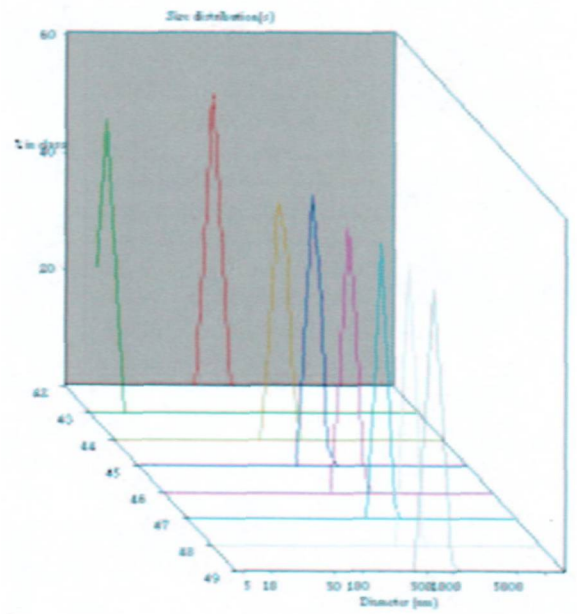


Figure 4.15: Cysteine Dioxygenase in 100mM PO_4/1mM Gly/200mM NaCl pH 7.5, 10mM PO_4/20mM Gly/100mM NaF pH 8.0, 10mM PO_4/20mM Gly/100mM NaF pH 7.4

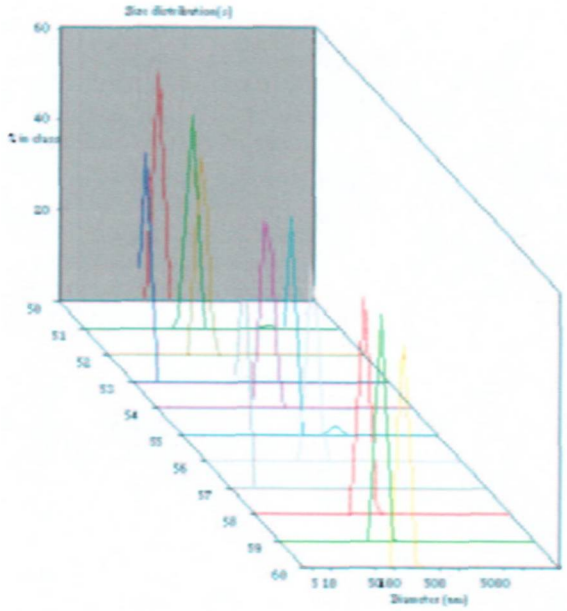


(a) NaCl pH 6.0

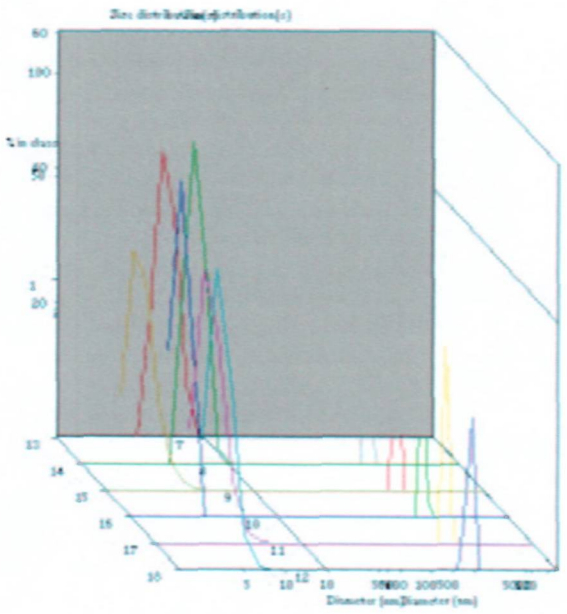


(b) NaF pH 6.0

Figure 4.16: Effect of pH on CDO aggregation

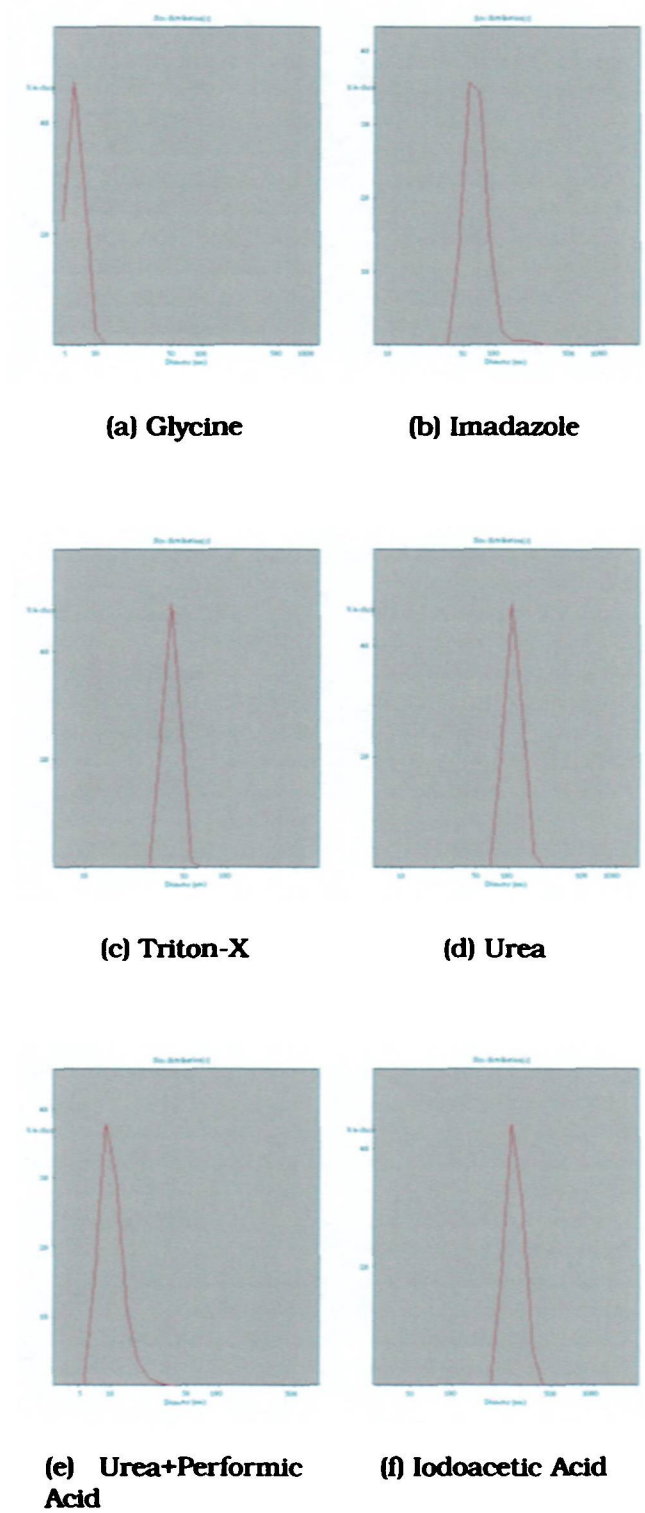


(a) NaF pH 7.4



(b) NaF pH 7.8

Figure 4.17: Effect of pH on CDO aggregation



(a) Glycine (b) Imadazole

(c) Triton-X (d) Urea

(e) Urea+Performic Acid (f) Iodoacetic Acid

Figure 4.18: Photon correlation spectroscopy results of affinity purified CDO.

Chapter 5

Structural Analysis

The structure of CDO has not been characterised before. An understanding of its three dimensional shape will eventually lead to an understanding of how the protein functions. Knowledge of its active sites and any regulatory domains may eventually lead to the development of pharmacophores of therapeutic value.

Recombinant CDO is not a perfect model of the endogenous protein and may not represent the actual structure. Regardless of possible discrepancies between the two proteins, it has a unique protein sequence with no known structures for any domain. For this reason, structural information is valuable for computational modellers. If in the future the structure of endogenous CDO becomes available, the structural information of the unmodified protein may prove to be important model for comparison.

5.1 Aims

The objective of the following sets of experiments gain as much information about the structure of CDO. Circular dichroism will be used to measure the

secondary structure of the protein and compare it to sequence predictions. Crystallography assays are set up to grow protein crystals that can be used in x-ray defraction studies. Nuclear magnetic resonance (NMR) were designed for functionality tests of the recombinant protein by analysis of the structure of its substrate and product. Two and three dimensional NMR experiments were also set up as an alternative method for solving the overall tertiary structure of CDO.

5.2 Methods

5.2.1 Circular Dichroism

Circular dichroism (CD) is a spectroscopic technique that can measure the secondary structure of a protein. Generally, this secondary structure is broken down into three different categories: α-helix, β-sheet, and random coil. α-helical elements are generally stretches of the protein backbone that form straight helical elements. Regions forming β-sheets are regions that are able to pair up with adjacent stretches of the amide backbone and may be parallel or anti-parallel. The random coil are stretches that have no apparent defined structure. These secondary structures ultimately contribute to the overall tertiary structure of the protein.

Circular dichroism refers to technique that characterises the optical characteristics of an asymmetric molecule, CDO in this instance. The results are independent of the known sequence. This information can be used to qualitatively determine secondary structure information about a peptide. Theoretically, it could also be used to determine quaternary structure information (such as monomer and dimer). It was initially discovered that peptides

of known secondary structures exhibited very different CD spectra and that a linear combination of these spectra can give a representative distribution of these secondary structures in a protein[118]. Today, with the growing databases of known protein structures, this technique has been refined in complex computer algorithms to semi-empirically determine a proteins secondary structure.

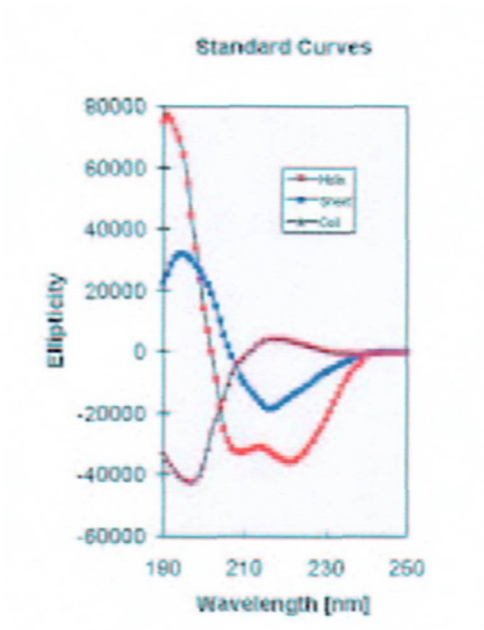


Figure 5.1: Standard curves for pure peptide solutions containing 100% α-helix, β-sheet, and random coil. Image taken from Protein Solutions, Inc.

In a CD measurement, plane polarised light (*with an initial angle*, θ_o) passes through a sample (*of path length* ℓ) and the transmitted light is analysed to determine to what degree the sample rotated the incident beam (ellipticity, θ) and to what extent the sample absorbed at a particular wavelength (λ). Typical measurements are between λ_{190nm} and $\lambda_{>300nm}$. The ellipticity at a given wavelength and the optical rotation (ϕ) are measured in terms of degrees[119].

$$\phi = \frac{180 \cdot \ell \cdot (n_L - n_R)}{\lambda}$$

$$\theta = \frac{2.303 \cdot (A_L - A_R) \cdot 180}{4\pi}$$

Deconvolution of CD spectra to attribute percentages of secondary structure were done predominantly with the K2D programme. Variables required by this programme are wavelength, ellipticity, path length, and protein concentration.

5.2.2 Protein Crystallography

Protein crystallography is a technique that seeks to solve a proteins structure by crystallising a protein. Proteins, just as salts or pure compounds, can join together in a highly ordered lattice. The resultant crystals are not dissimilar in appearance to a crystal of salts, sugars, gems, etc. Protein crystals contain a high percentage of water and are generally very fragile.

The resultant protein crystal is analysed by an intense x-ray beam. An intense and highly columnated beam is passed through the crystal and analysed by a detector. Because crystals are periodic lattices by definition, the x-rays reflect between and defract between and within each unit protein. The defraction pattern eventually leads to an electron density map of an individual protein. It is possible to then "fit" the peptide sequence into this map and yield the overall structure of the protein.

A protein sample must be monodisperse to crystallise. During crystallisation, as opposed to aggregation, the protein comes together and interlocks in a periodic manner. This is achieved during a process of concentration near its solubility limit. A protein must be concentrated very slowly in order to crystallise.

Several procedures have been developed to crystallise a protein but the

"hanging drop" method is the most basic. A schematic of this method is detailed in figure 5.2. A protein sample is mixed with a 50% mixture of a

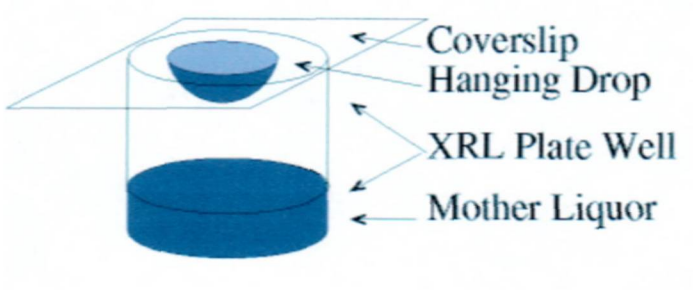


Figure 5.2: Schematic of the hanging drop method

crystallisation solution, the mother liquor. A volume small enough to be suspended as a drop is placed on a glass coverslip and inverted. The coverslip is sealed above a well containing mother liquor. The mother liquor can contain any number of buffers that may be conducive to crystallisation; it also contains hygroscopic compounds such as polyetheleneglycol (PEG). Over time, water vapour from the suspended sample diffuses into the mother liquor below. The hanging drop method relies on this vapour diffusion to slowly concentrate CDO.

There is no prescribed mother liquor solution that will crystallise all proteins so a matrix of solutions need to be tested. Proteins often crystallise near their isoelectric points[1] but this variable needs to be assayed as well. The rate of crystallisation is dependant upon the temperature and PEG concentrations and need to be tested in turn.

[1]The predicted isoelectric point for CDO is listed in Fig. 5.8 on page 118.

300 mM Sodium Acetate / 25% PEG 2000 MME	Sodium Acetate pH 4.5
200 mM Li_2SO_4/ 25% PEG 2000 MME	Sodium Acetate pH 5.5
200 mM $MgCl_2$ / 25% PEG 2000 MME	Sodium Cacodylate pH 6.5
200 mM KBr / 25% PEG 2000 MME	TRIS pH 7.5
200 mM KSCN / 25% PEG 2000 MME	TRIS pH 8.5
800 mM Sodium Formate / 25% PEG 2000 MME	
300 mM Sodium Acetate / 15% PEG 4000	
200 mM Li_2SO_4 / 15% PEG 4000	
200 mM $MgCl_2$/ 15% PEG 4000	
200 mM KBr / 15% PEG 4000	
200 mM KSCN / 15% PEG 4000	
800 mM Sodium Formate / 15% PEG 4000	
300 mM Sodium Acetate / 10% PEG 8k+ / 10% PEG 1k	
200 mM Li_2SO_4 / 10% PEG 8k+ / 10% PEG 1k	
200 mM $MgCl_2$ / 10% PEG 8k+ / 10% PEG 1k	
200 mM KBr / 10% PEG 8k+ / 10% PEG 1k	
200 mM KSCN / 10% PEG 8k+ / 10% PEG 1k	
800 mM Na Formate / 10% PEG 8k+ / 10% PEG 1k	
300 mM Na Acetate / 8% PEG 20k+ / 8% PEG 550 MME	
200 mM Li_2SO_4 / 8% PEG 20k+ / 8% PEG 550 MME	
200 mM $MgCl_2$ / 8% PEG 20k+ / 8% PEG 550 MME	
200 mM KBr / 8% PEG 20k+ / 8% PEG 550 MME	
200 mM KSCN / 8% PEG 20k+ / 8% PEG 550 MME	
800 mM Na Formate / 8% PEG 20k+ / 8% PEG 550 MME	

Table 5.1: Clear Strategy 1 crystallisation matrix solutions.

Protocol

Coverslips and XRL plates were purchased from Molecular Dimensions, Ltd. A matrix of buffers were supplied in with a Clear Strategy 1 protein crystallisation kit (Molecular Dimensions) (Figure 5.1). Coverslips were siliconised making them more hydrophobic. Coverslips were sealed to the XRL plates with vacuum grease (DuPont).

Purified CDO was mixed with a 50% mixture of each solution. The pH of each solution was adjusted beforehand with each of the 5 supplied buffers.

Approximately 50$\mu\ell$ to 100$\mu\ell$ were placed at the centre of the coverslip. A 1ml volume of mother liquor was used at the bottom of the 24 well XRL plate. The samples were left to incubate at room temperature for a period of days to weeks. Samples were the evaluated to see whether the drops remained clear, showed precipitation, or contained crystalline material. Evaluation was done with a light microscope equipped with a digital camera.

If a sample remained clear at a given condition, the procedure was then repeated at a higher starting concentration of CDO. For samples that showed precipitation, the procedure was repeated at a lower starting concentration of CDO.

5.2.3 Nuclear Magnetic Resonance

Nuclear magnetic resonance (NMR) is a technique that involves subjecting a sample to intense modulated magnetic fields. This method can gather information about the primary, secondary, tertiary, and quaternary structure and even the dynamics of a protein. The same technique may be referred to a magnetic resonance imaging (MRI) in medicine when it is applied to macroscopic structures, as organs. A detailed understanding of the physics behind the technique is not necessary; as long as a qualitative understanding of the theory is understood together with the methodology required to interpret spectra. NMR is concerned with resonances induced in chemical bonds by a magnetic field.

Different atoms respond differently to intense magnetic fields by virtue of whether they are magnetically polar or not and the symmetry of their polarity. The polarity and symmetry thereof is dictated by the laws of quantum physics and is of no concern to the molecular biologist. Typically, a biologist is only

concerned with hydrogen, nitrogen, and carbon as these elements are the most abundant in a protein and are the centre of interest in determining protein structure.

Hydrogen, as opposed to deuterium (^2H) and tritium (^3H), is magnetically polar and is the most commonly practised NMR, ^1H NMR. Deuterium is apolar so it is a common practice to use deuterated solvents (ie: D_2O vs. H_2O) with a buffer to render them invisible, if one is interested in only the protein or substrate/product. Most naturally occurring carbon (^{12}C) and nitrogen (^{14}N) are either apolar or not sufficiently symmetrical enough making them effectively apolar for all intents and purposes. Completely incorporating ^{13}C and ^{15}N, which are polar non-radioactive isotopes, into a protein makes them "visible" to an NMR spectra and yields structural information.

As a general rule, NMR requires a concentrated, mono-disperse, homogeneous protein sample. Presently, it is limited to proteins roughly 50kDa or smaller in size. The single greatest limitation of NMR is the expense of ^{13}C products. It has been used in these experiments to probe the functionality of CDO by testing substrate/product conversion (functionality) as well as two and three dimensional NMR of the CDO protein itself.

5.2.4 Protocol

For the functionality tests, NMR spectra were first obtained of CDO substrates, products and co-factors both with and without enzyme. Samples were dissolved in an appropriate aqueous buffer which was quickly frozen. Frozen samples were then lyophilised to remove H_2O and were resuspended in D_2O (99.9% purity, Sigma). Spectra for all functionality tests were performed on either Brucker AP300S 300MHz or on an AV300 300MHz.

Substrates and product standards for NMR functionality tests of CDO were dissolved in PBS, frozen, lyophilised overnight, and resuspended in 99% D_2O: β-nicotinamide adenine dinucleotide, reduced form (Sigma, N-6005), β-nicotinamide adenine dinucleotide (Sigma, N-0632), L-cysteinesulfinic acid (Sigma, C-4418), L-cysteine (Sigma, C-8152). A cystine standard was not performed because NMR spectra because its spectra is nearly indistinguishable from L-Cysteine. Furthermore, deuterated additives that prevent L-Cysteine/cystine interconversion are cost prohibitive.

For experiments involving direct structural studies of CDO, spin labels were incorporated into the protein by growth and induction in minimal media supplemented with the label of choice: Ammonium-^{15}N Chloride (98% ^{15}N, GOSS Scientific Instruments, NLM-467) and D-Glucose-U-$^{13}C_6$ (99% ^{13}C, Cambridge Isotope Laboratories, CLM-1396). Protein samples were supplemented with 10% Deuterium Oxide (99.9% D, Aldrich, 26,979-4) and analysed on either a Brucker 500MHz or a Varian 600MHz magnet.

5.3 Results and Discussion

5.3.1 Secondary Structure Determined Using Circular Dichroism

Since a bacterial system was used to express CDO, concerns were raised that the protein may not fold properly because the bacterial environment is much more reducing that its eukaryotic partner. CD was then used to determine if the recombinant CDO had comparable secondary structure as would be expected from its primary sequence.

Secondary structures are formed by the chemical nature of the amino acid

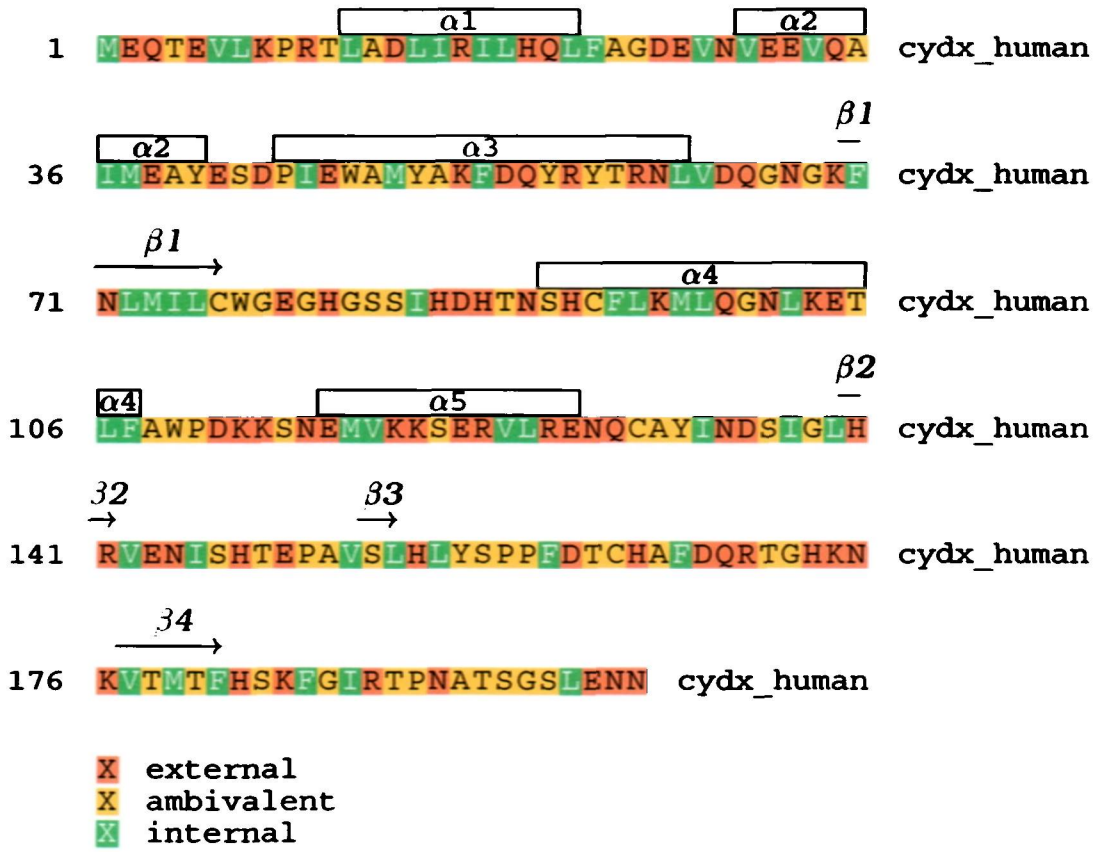


Figure 5.3: Predicted secondary structure of cysteine dioxygenase aligned with its primary sequence.

ligands (hydrophilic/hydrophobic, aromatic/aliphatic, etc.). There are several computer programs that take a primary amino acid sequence a predict what percentage of the protein should be helical, sheet, and coil. The sequence of cysteine dioxygenase was submitted to one of these programmes (PHDsec). The results of this analysis are listed in in the appendix on page 149 to 165. To present this data in a more useful format, the results were aligned with the CDO sequence using TEXshade (Figure 5.3). It is useful to compare these predictions with the experimentally measured distributions from CD. The data for secondary structure prediction are listed in the Appendix on pages A to A.

An initial CDO sample was dialysed into 200mM Na_2HPO_4 pH 7.0 to avoid

absorbance problems associated with chloride ion. Because the sample was dilute, A 0.2mm path length quartz cuvette (Hellma, Uk) was used. The sample was scanned at 100nm/min, with a 1 second response time, a 1nm bandwidth, and the approximate concentration was 3mg/ml (Figure 5.9). The initial spectrum was then submitted to the K2D deconvolution program for deconvolution. The results then indicated that CDO was 40% alpha-helix, 8% beta-sheet and 52% random coil. The CD spectroscopic results do not vary much from what was predicted by the GCG programme based on the CDO primary sequence alone.

Method	$\alpha - helix$	$\beta - sheet$	*random coil*
CD Results	40%	8%	52%
GCG Predicted	≈35%	≈4.5%	≈60.5%

Table 5.2: Tabulation of predicted secondary structure of CDO based on CD spectroscopic results and theoretical predictions based on computational analysis of primary sequence.

It is important to note that the the critical deconvolution was performed within the range of 200nm and 240nm and the absorbance problems below 200nm were minimised or eliminated. This was an encouraging result because it indicated that the protein maintained secondary structure. Because the result was so similar to what would have been expected based on CDO's primary sequence, it also indicates the actual secondary structures of CDO in other organisms (because of the strong sequence homology) may also be homologous in their secondary and presumably tertiary structures as well (Figure 5.4).

Pursuant to the initial folding of recombinant CDO, it was discovered that the protein has a tendency to aggregate, particularly at higher concentrations (See section: Dynamic light scattering). Further CD studies were carried out

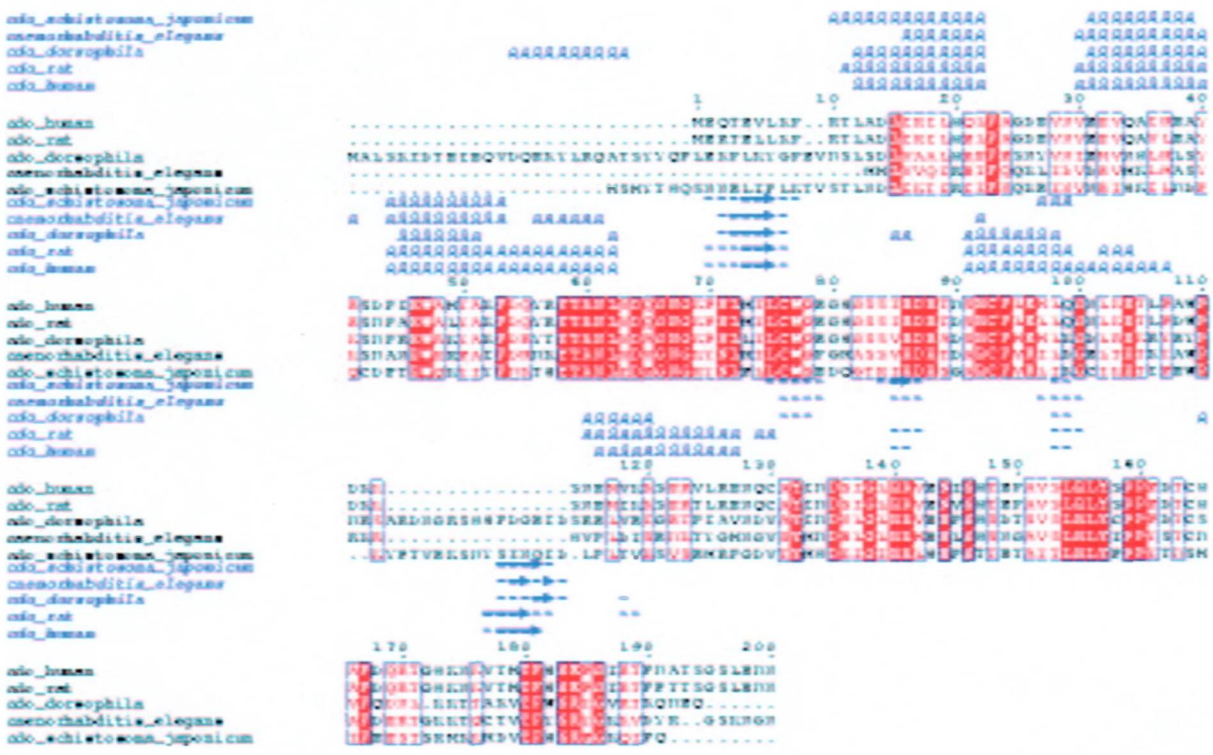


Figure 5.4: Predicted secondary structural homology of CDO

on CDO in buffers that prevent aggregation and are at concentrations necessary to be detected by NMR to ensure the protein was not being denatured in the process of preventing its aggregation.

Subsequent results had to be performed at higher concentrations. They also had to be performed in buffers that are either chiral (Tris)[2] or highly absorbing in the far-UV range (Glycine). To surmount this problem, a special cuvette, 0.10 mm in path length, had to be used. The symmetry of the spectra do not vary appreciably from the initial results indicating that CDO is a stable protein structurally. Moving to a 0.01 mm path length cuvette was not possible because the protein CD spectra was lost. Thoroughly testing the

[2]Chirality is a property where the mirror image of a chemical results in a chemically distinct different compound. There is a requirement of a certain number of mirror planes to meet achieve this property. Chiral compounds, such as L-Cysteine and D-Cysteine (Levorotary and Dextrorotary), rotate light in opposite directions. Chiral buffers, such as Tris, will influence spectra taken from protein samples.

CD spectra below the isoelectric point (Figure 5.8) was not feasible because the conditions found to keep the protein mono-disperse were only appropriate above the isoelectric point.

Samples of affinity purified CDO were dialysed against 100mM NaF, 20mM Glycine, 10mM PO_4. The concentrations of the samples were estimated by taking the absorbance at λ=280 nm and using the extinction coefficient for the recombinant CDO sequence. All samples were of the same concentration of about 1.24mg/ml but varied in pH.

$$\begin{aligned} C_{rCDO} &= \frac{ABS_{\lambda=280}}{\varepsilon_{\lambda=280}} \\ &= \frac{1.243}{25230}\,(1\,cm\,path\,length) \\ &= 4.93E-5\,M \\ &= 4.93E-5\,M \cdot 25.148\frac{g}{mol}) \\ &= 1.24\frac{mg}{ml} \end{aligned}$$

These results (Figure 5.5) indicate that there is a shift in the secondary structure of the recombinant protein as a function of pH. The change occurred in buffers slightly more acidic than pH 7.4. There was no detectable structural change between the range of 7.4-8.0. All these measurements were performed at nearly identical ionic strengths. The near identical absorbance of these samples indicate that there was virtually no change in protein concentration during the dialysis of the samples. The conformational changes appear to be real structural changes and not an artifact; As a control, a CD spectrum of CDO, pH 7.4, was run under identical conditions as the previous save the addition of 250 nM $FeSO_4$ (Figure 5.6). Iron sulphate has bee

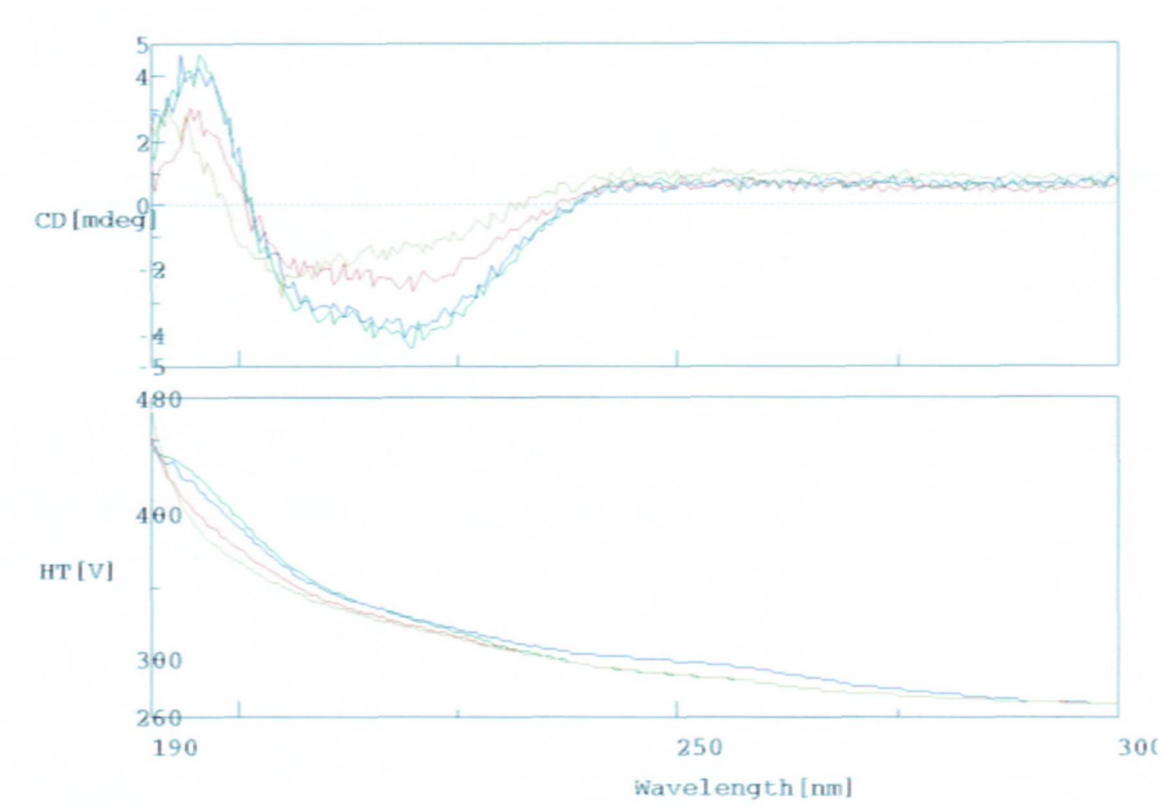


Figure 5.5: Recombinant cysteine dioxygenase pH dependant changes in the optical rotation of light in 100mM NaF,20mM Gly, 10mM PO_4. (pH 8-blue, pH 7.4-green, red-pH7.0, and pH 5-yellow

shown in dynamic light scattering experiments to cause flocculation under these conditions. The resultant spectra yielded no secondary structural information.

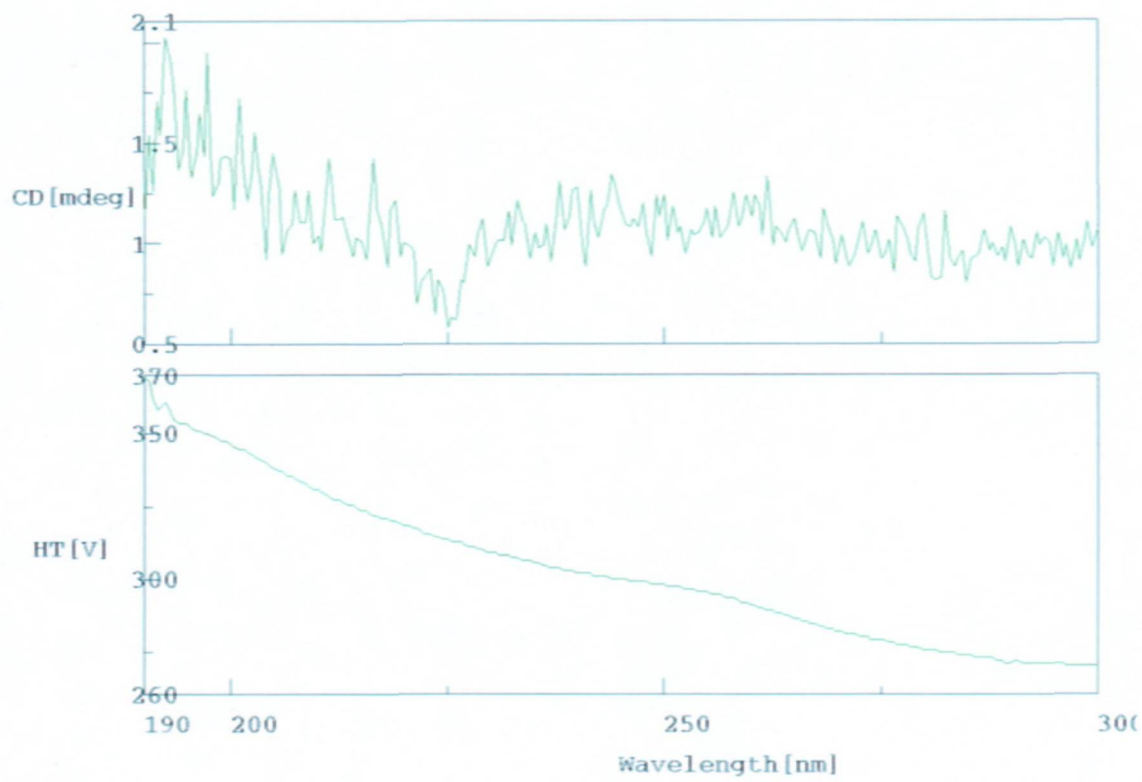


Figure 5.6: Total loss of secondary structural information under conditions of protein aggregation of the recombinant CDO. (1.24mg/ml, 100mM NaF, 20mM Gly, 250nm $FeSO_4$ pH 7.4)

Data acquired from the CD spectra were saved as text files and were imported into MS Excel. Data points corresponding to wavelengths 200nm to 240nm were saved as a separate text file for analysis by the K2D programme, a CD deconvolution programme developed at EMBL Heibelberg. The source code, written in C for Solaris, was compiled under Linux using g++. Interpretation of the spectra led to the tabulated percentages of random coil listed in table 5.3.

A more detailed analysis of the spectra using the CDNN 2.1, a programme that tries to draw more secondary structural information from a spectra. The

pH	% Random Coil
8.0	49
7.4	48
7.0	48
5.0	45

Table 5.3: Deconvolution of CD spectra for CDO in 100mM NaF, 20mM Gly, 10mM Phosphate at varying pH

concentration of the sample was defined as 1.25mg/ml, 25kD, 220aa residues long, and with a path length of 0.01cm. Figure 5.7 tabulates the results. There is a linear relation between a change in secondary structure with the pH and the most dramatic change occurs between pH 7.4 and 7.0. This is also a change around the predicted isoelectric point, pH 7.02, for CDO (Figure 5.8). Sums of secondary characteristics exceeding 100% may be viewed as a window of error in interpreting the accuracy of the defined secondary structures. Contributions to this error include changes in concentration, variability of the cuvette pathlength or position, as well as possible structural characteristics that share similarity to one or more of the defined categories. However, visual inspection of the plotted data indicate that there is some decrease in the apparent secondary structure of the protein that is very pH dependant (Figure 5.5).

Overview of Results

CD spectra of CDO in 200mM Phosphate buffer alone indicated a defined secondary structure of 40% α-helix, 8% β-sheet, and 52% random coil. In the absence of chloride ions (from NaCl), there was a dramatic decrease in the associated fluorescence which allowed higher concentrations of protein to be used with less noise. Addition of NaF as a less fluorescent alternative

to NaCl and addition of glycine introduced noise. DLS experiments indicated these were favourable conditions for CDO stability. In these buffers, samples had to be integrated ten or more times and at slower scan rates to obtain continuous curves. However, the sample stability with time, acquisition times and instrument availability were considerations that had to be balanced. The results were not far different from the standard CDO spectra in phosphate alone.

Because of the variability involved in deconvolution from one software package to the next, many groups will argue that it is not a reliable technique to assign secondary structures. The main purpose of these experiments was to at least show that the protein did not conform to an molten globule with no secondary structure. In establishing that that the protein is at most 50% random coil, a statement can be made that the remaining 50% of the rCDO conforms to some ordered structure. Some of the variability could be attributed to the poly-histidine fusion protein. While it would have been desirable to remove it with thrombin, chromatographic equipment needed to fractionate thrombin, CDO and the fusion protein apart was not available. Attempts to remove the fusion protein with thrombin-agarose were not successful. Contaminants such as agarose (even after centrifugation), glycerol, and possibly albumin prevented ionisation of the sample for MALDI and prevented measurement with DLS.

pH	αH.	Anti-Parallel	Parallel	βTurn	R.C.	Sum
5.0	13.3%	18.5%	17.1%	22.2%	52.3%	123.4%
7.0	14.3%	17.5%	16.3%	21.8%	50.9%	120.7%
7.4	16.4%	15.7%	14.9%	21.0%	48.1%	116.1%
8.0	16.9%	15.3%	14.6%	20.8%	47.5%	115.1%

Figure 5.7: CDNN 2.1 interpretation of CD spectra using the 210nm-260nm method.

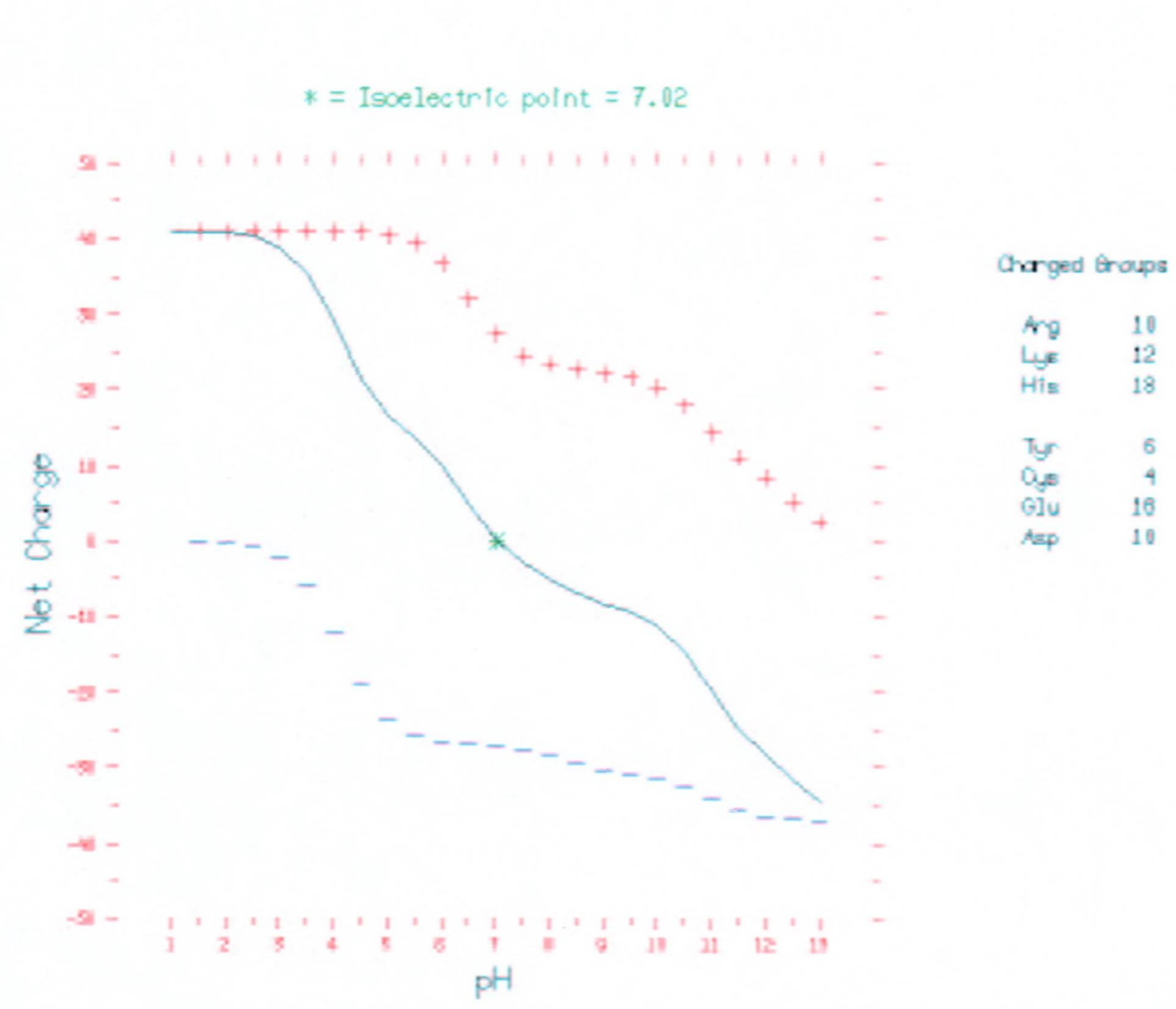


Figure 5.8: Predicted isoelectric point of CDO calculated using GCG Predictions based on primary sequence of the 220aa recombinant CDO.

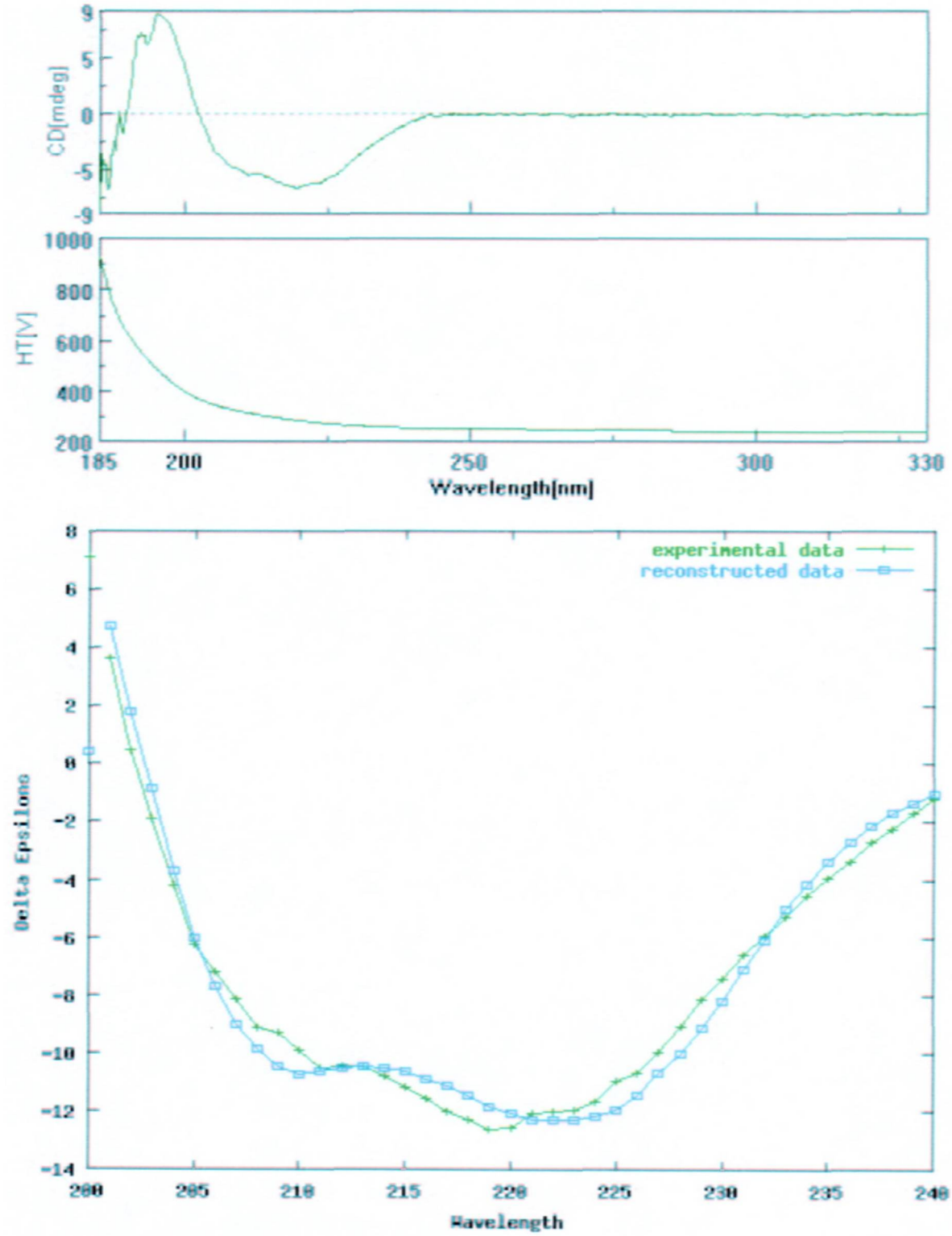
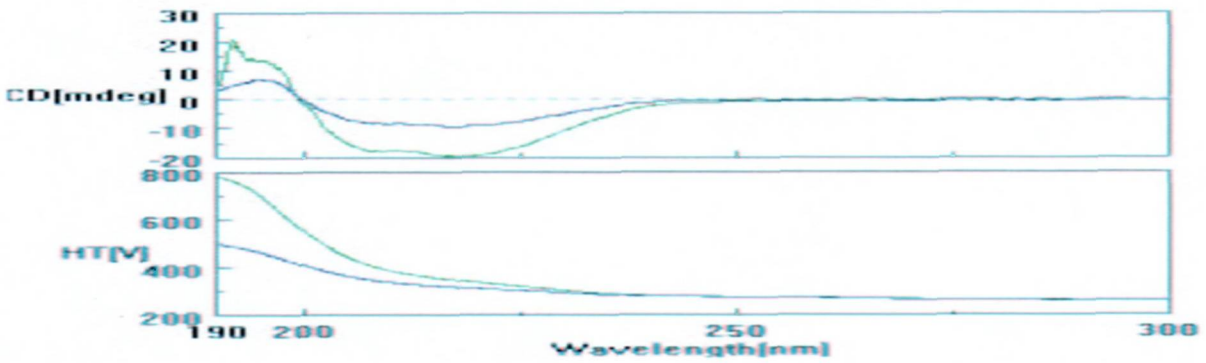
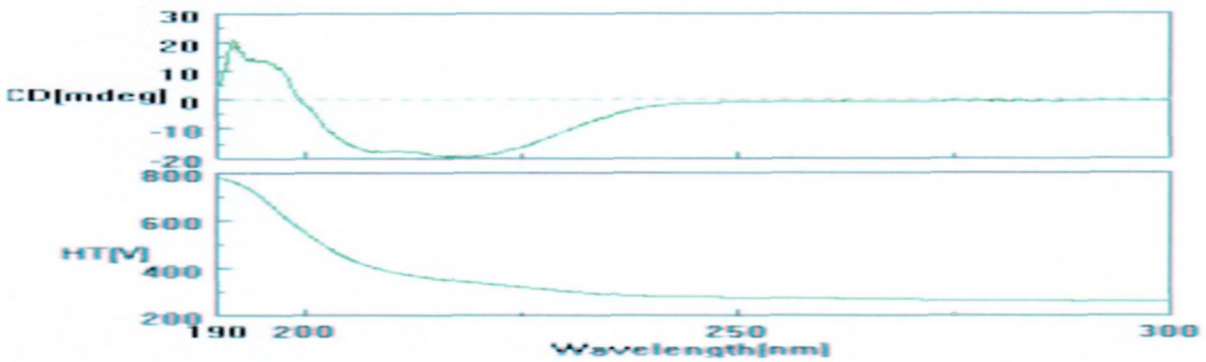


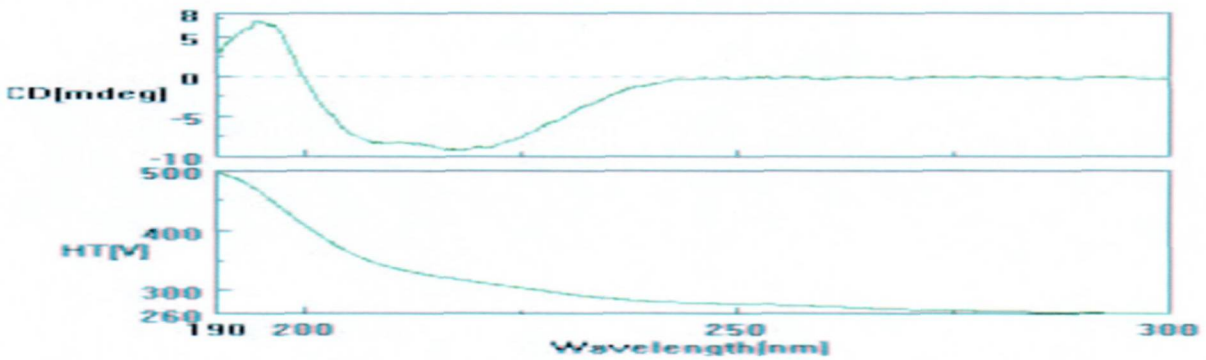
Figure 5.9: CD spectrum and deconvolution of CDO at 200mM Na_2HPO_4 pH 7.0



(a) Overlay of plots (b) and (c)

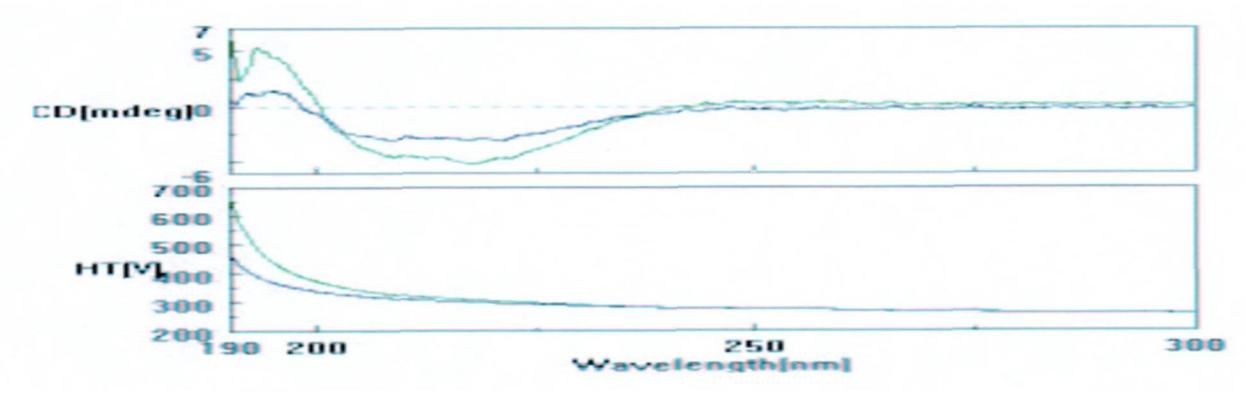


(b) f1 100mM NaF/200mM Glycine/10mM Phosphate pH 8

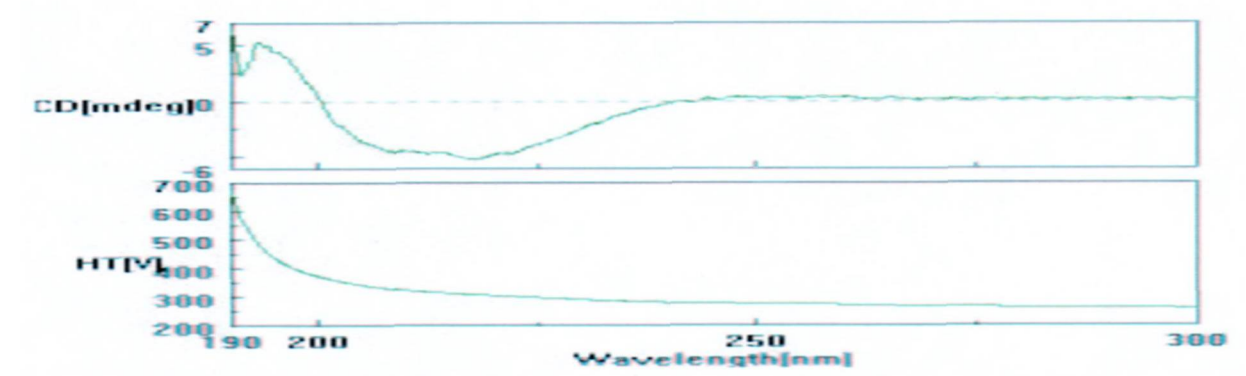


(c) f2 50mM NaF/100mM Glycine/5mM Phosphate pH 8

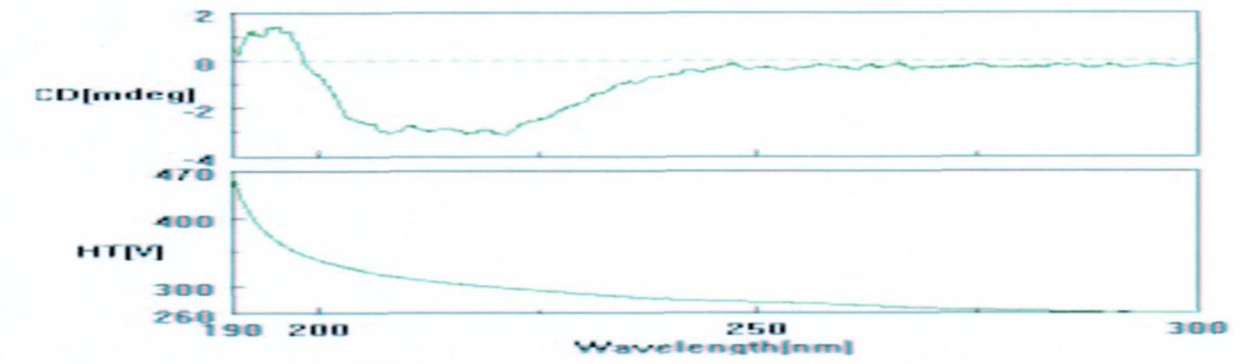
Figure 5.10: f1f2



(a) Overlay of plots (b) and (c)

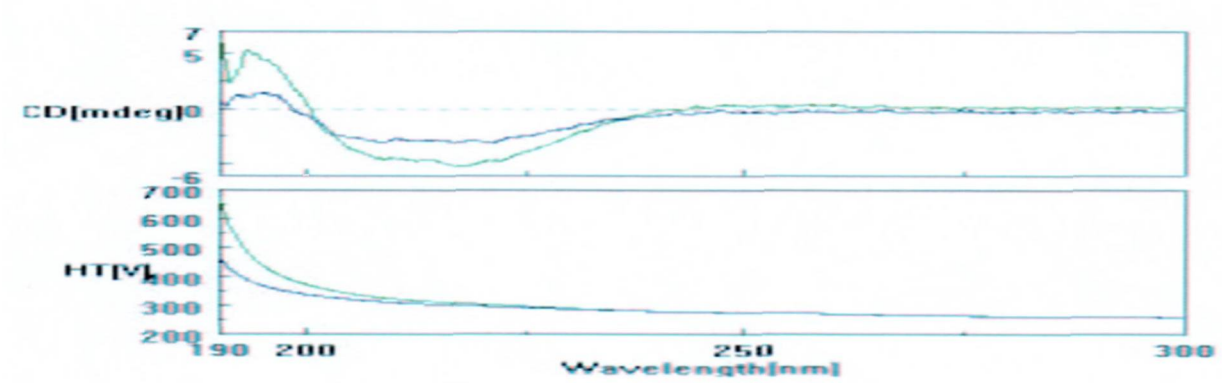


(b) f3 200mM NaCl/1mM Glycine/100mM Phosphate pH 7.4

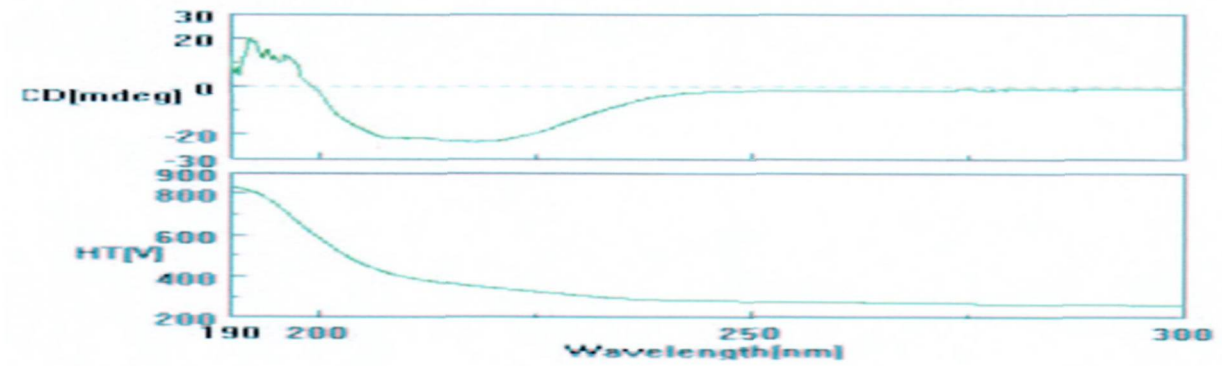


(c) f4 100mM NaCl/100uM Glycine/50mM Phosphate pH 7.4

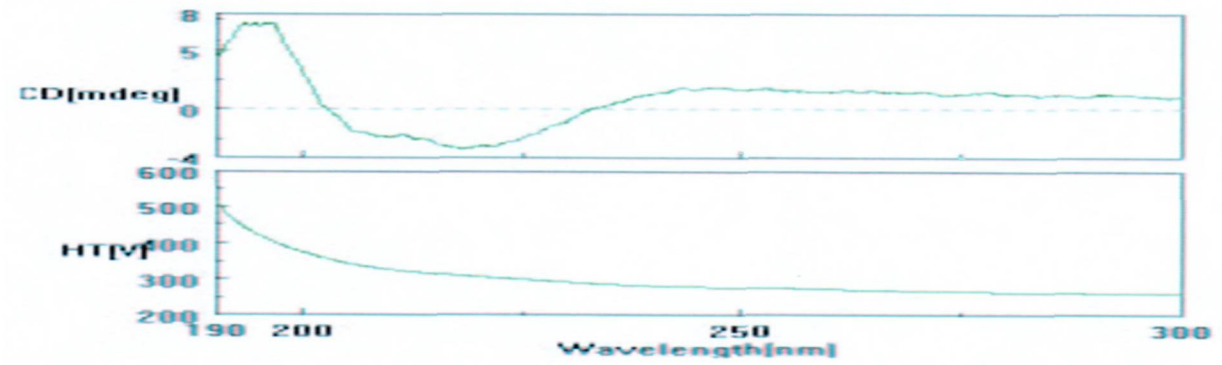
Figure 5.11: f3f4



(a) Overlay of plots (b) and (c)

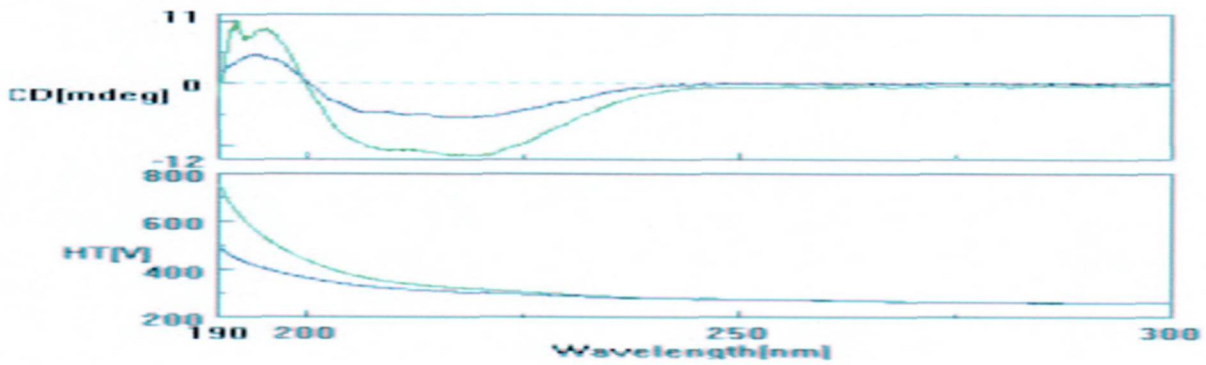


(b) f5 100mM NaF/20mM Glycine/10mM Phosphate pH 7.4

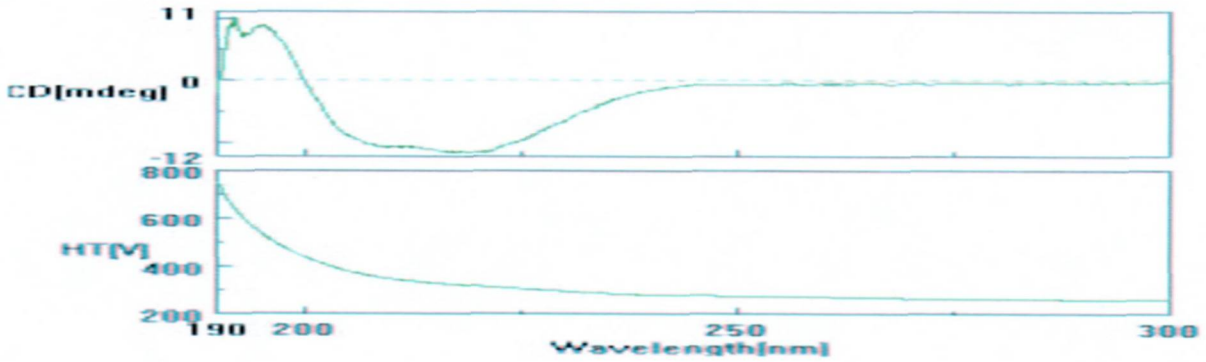


(c) f6 50mM NaF/10mM Glycine/5mM Phosphate pH 7.4

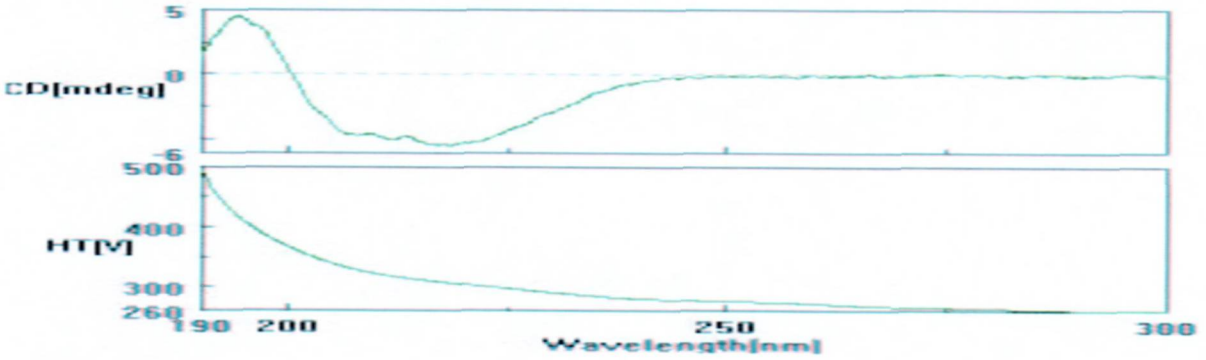
Figure 5.12: f5f6



(a) Overlay of plots (b) and (c)



(b) f7 200mM NaCl/1mM Glycine/100 mM Phosphate pH 7.5



(c) f8 100mM NaCl/100uM Glycine/50mM Phosphate pH 7.5

Figure 5.13: f7f8

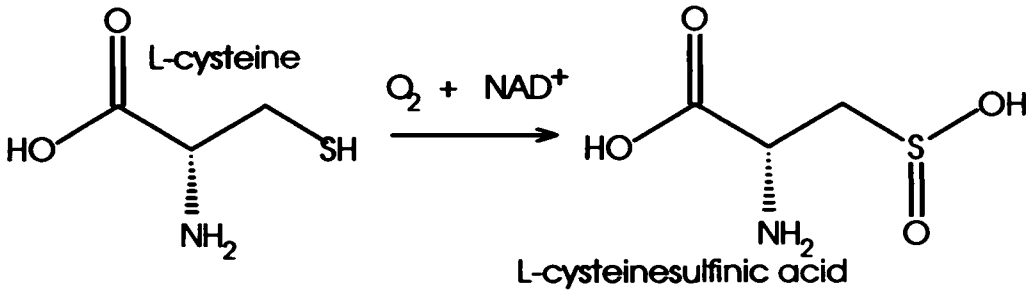


Figure 5.14: Conversion of L-Cysteine to L-cysteinesulfinic acid

5.3.2 Functionality Assay Using NMR

The NMR functionality assay of CDO was dependant on the very different chemical differences between L-cysteine and L-cysteinesulfinic acid. The ^1H NMR experiment was performed in D_2O so only the hydrogens attached to aliphatic carbons[3] appear in the spectra. The greatest expected proton shift for substrate conversion was that for the carbon adjacent to the sulfhydryl group.

Sulphur, oxygen, carbon, and hydrogen have electronegativities of 2.4, 3.5, 2.5 and 2.2 respectively. Depending in the difference in magnitude of these values, a dipole moment will exist between two bound atoms. When the proton of the sulfhydryl group is replaced with three oxygen bonds, the electronegativity of the sulphur is increased due to the electron withdrawing effect of the oxygens. Do to the greater electronegativity of the sulfinic acid group, a NMR resonance shift is expected for the adjacent carbon.

As standards, NMR spectra were taken for L-cysteine, L-cysteinesulfinic acid, NAD^+ and NADH individually (Figure 5.15 and 5.16). There was an easily detectable shift between L-cysteine and L-cysteinesulfinic acid. The scales between the substrate and product do not match because the resonances are affected by concentration and pH. It is the relative shifts that are important.

[3]Aromatic protons are detected in NAD^+/NADH ^1H NMR spectra as well

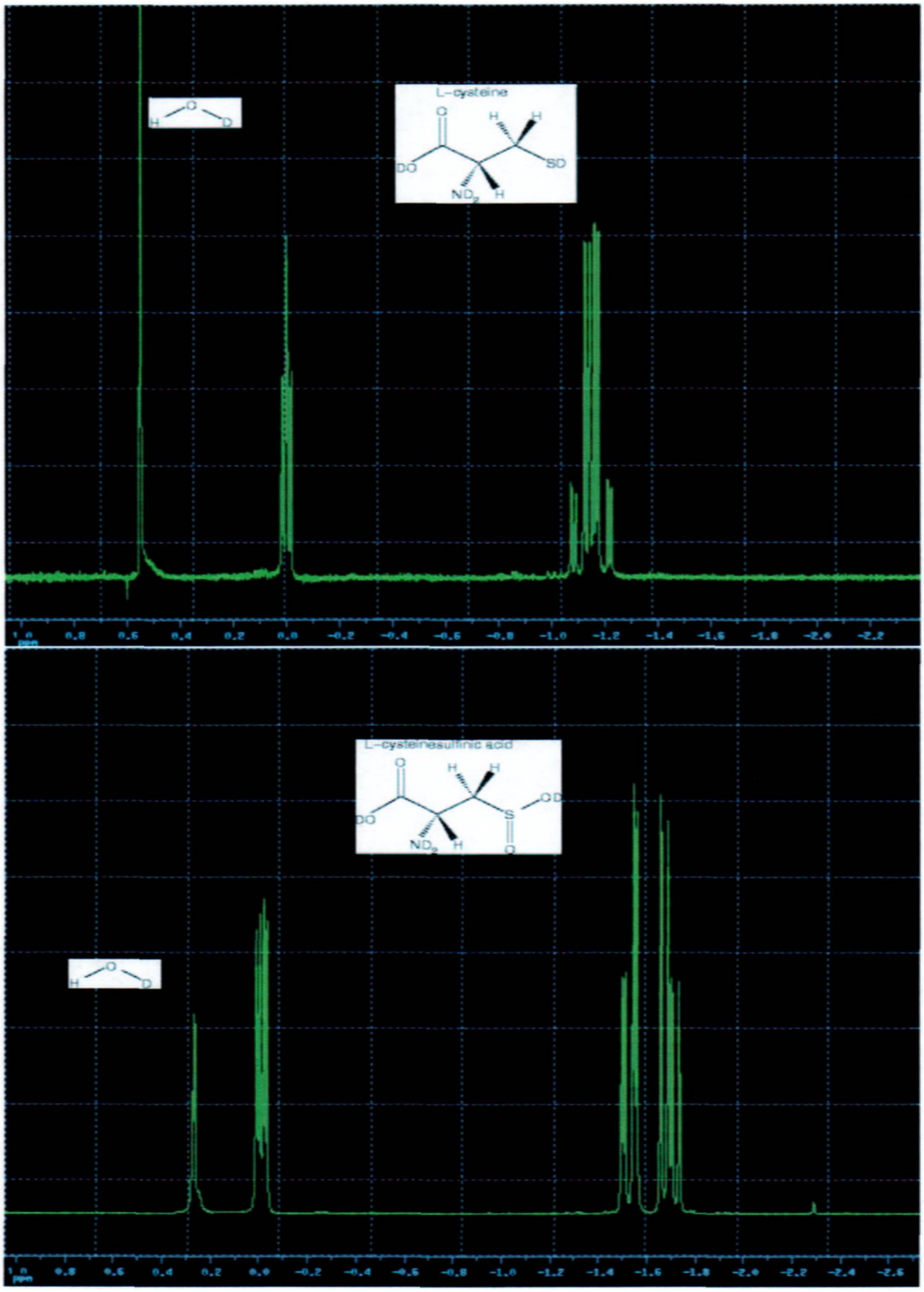


Figure 5.15: 1-D NMR of (a) cysteine and (b) cysteinesulfinic acid

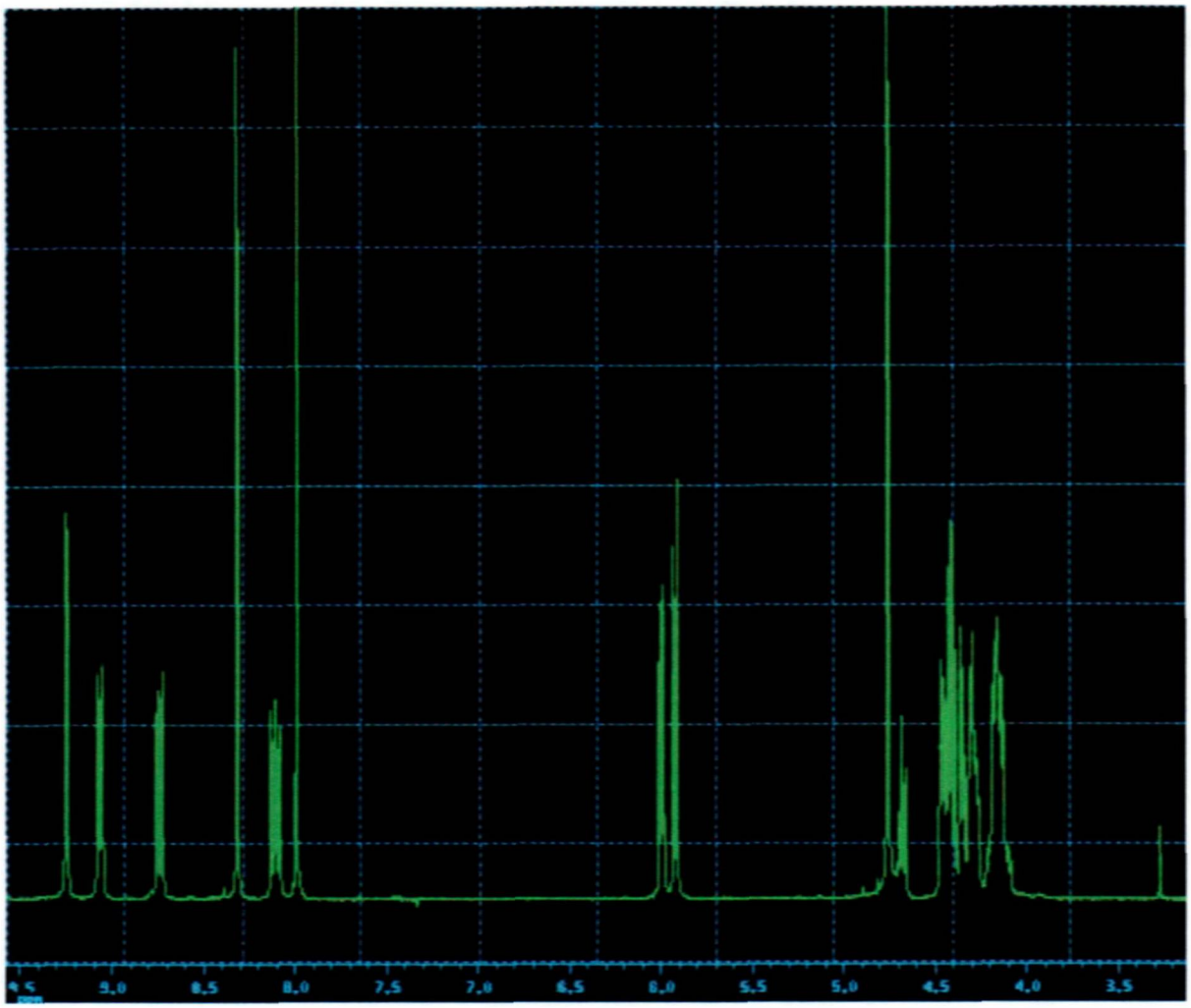


Figure 5.16: 1-D NMR spectra of NAD^+

Compounds that prevent the oxidation of L-cysteine to cystine were not added because they would complicate the NMR spectra. Differences in spectra between L-cysteine and cystine were not expected to vary appreciable due to the similar electronegativities. The singlet resonance that appears at resonances 0.48 ppm and 0.28 ppm represents trace hydrogen deuterium oxide (HDO) for L-cysteine and L-cysteinesulfinic respectively. The carbonyl carbon is not protonated and does not appear.

The cysteine used was stored under ambient conditions and likely contained some oxidative product from exposure to the oxygen in the air. The relative molar concentrations were thought to be insignificant. Although not visible on the reproduced plot, there are faintly detectable peaks between -0.8 and -1.0 ppm in the spectra of L-cysteine. This may be due to a cysteine-sulfinic/sulfonic acid group on the α-carbon.

Spectra were taken of the substrates in combination with and without the presence of enzyme (5.17) . Most of the spectra of L-cysteine is masked by the resonances from the cofactors. However, the resonances just above 3.0 ppm corresponding to protons on the α-carbon are clearly visible. There was no detectable decrease in these resonances with the addition of CDO. Also, there were no detectable increases in resonances between 3.1 and 3.5 ppm above baseline with CDO. The data indicates that CDO is not functional.

The single resonance at 3.2 ppm that disappears with the addition of CDO is likely a methanol or acetone contaminant. Neither of these compounds were used at any stage in the experiment and it is not clear how it appeared in one sample. The Wilmad NMR tubes used for the one experiment were likely not dried thoroughly.

Although these data indicate the recombinant CDO to be non-functional, the assays were by no means exhaustive; Buffers optimal for the stability of

Figure 5.17: 1-D NMR spectra of (a) cysteine, NAD^+ and (b) cysteine, NAD^+, and cysteine dioxygenase

the protein were not ideal for for its catalytic activity. Organic buffers had to be avoided to prevent contaminating signals. Literature indicates that CDO is more functional in a slightly basic environment and DLS experiments have shown that this protein will aggregate under these conditions. Iron is a required co-factor of CDO. The presence of $FeSO_4$ has also been shown to induce aggregation in DLS experiments[4]. All previous studies of endogenous CDO have been from HPLC/FPLC purified cell extracts. These may have included an associated protein that affects the activity of CDO. Indeed, some groups have reported an associated "protein-A". While these results were not what was hoped for, the apparent lack of activity may not necessarily be due to the choice of expression system used. It is also possible the prokaryotic expression system may not incorporate the required iron atom.

5.3.3 Crystallography

Several attempts were made to get CDO to crystallise. In initial purification schemes, to was found that crystals formed at ~4°C in elution fractions (Figure 5.18). Initial purification of CDO was done in 500mM-1M NaCl, 500mM imidazole with either Tris or phosphate buffer (pH ~7.4). These crystals were blue in colour. It was assumed that the colour was due to nickel from purification. It was also observed that, after elution, regions of the sephadex that bound CDO lost its blue hue.

A crystal was washed with H_2O and dissolved in 100% MeOH. The mixture was then blotted and probed with antibodies for CDO. The crystal tested positive for the protein. The crystal was not however given further consideration for X-ray analysis because most protein crystals are fragile and the crystals

[4]The data was not recorded due to the severity of the initial measurement and the limited availability of the instrument.

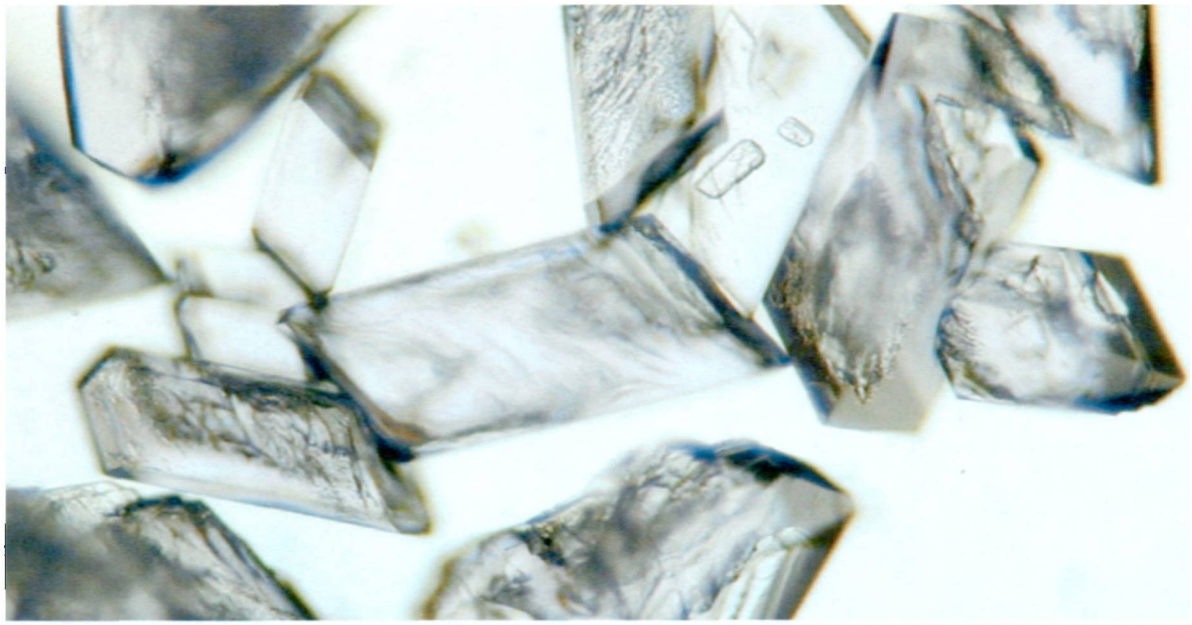


Figure 5.18: Crystals formed in elution fractions that contained CDO.

found were dense and sturdy. It was presumed that these were salt crystals and no beam time was allocated.

Further crystallisation assays were performed CDO under a variety of starting buffers using the Clear Strategy matrix. After a period of several days to weeks, the hanging drops were observed using a light microscope. Figure 5.19 lists a few typical results. CDO had a tendency to precipitate out of solution with a variety of morphologies. In one well, there was a portion of a CDO sample that appeared to be semi-crystalline (Figure 5.20). The remainder of the sample contained a lot of protein precipitate.

Attemps were made to focus on the pH, temperature, and starting concentration of CDO for crystallography. Soon afterwards however, DLS experiments showed that the protein aggregated under virtually any condition. This process might have been slowed slightly with the increased viscosity of the crystallisation solutions, however, it would not be possible to form a crystal with a non-disperse starting protein.

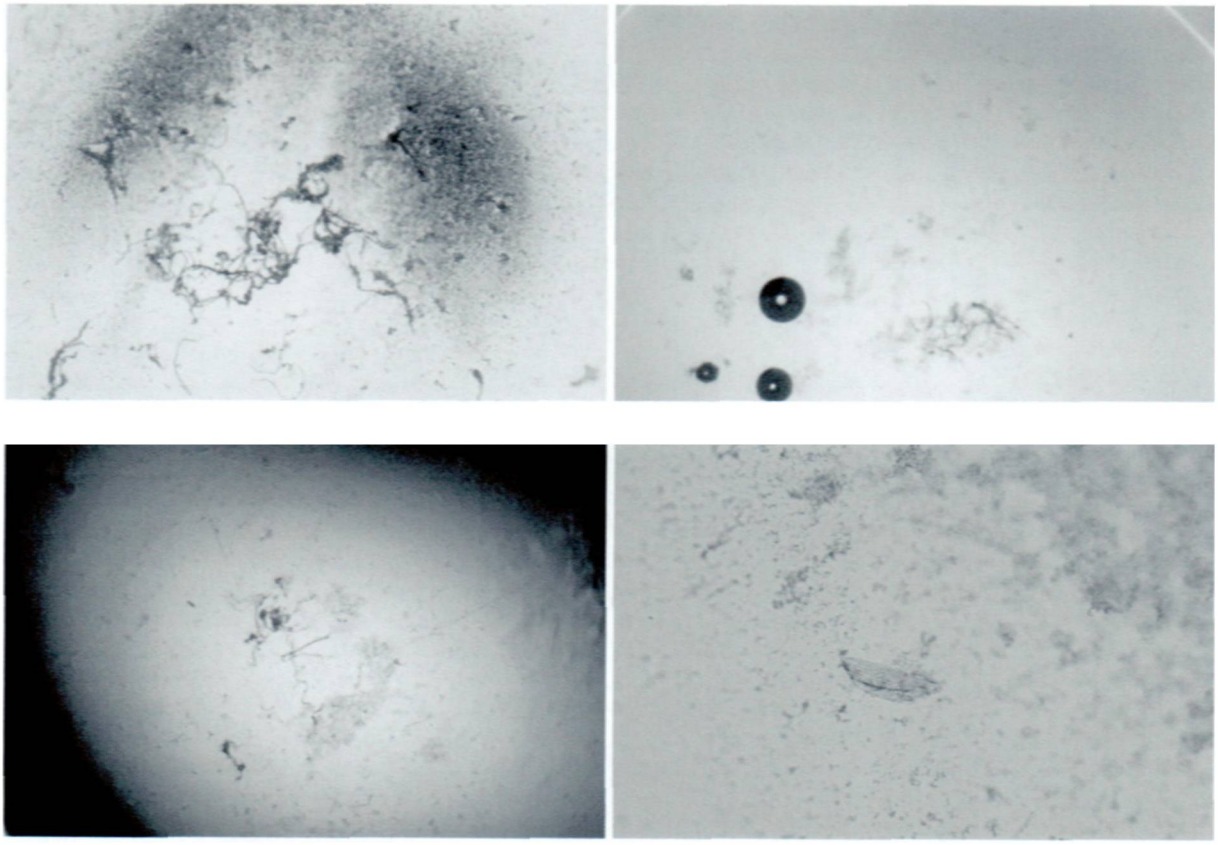


Figure 5.19: Results of protein crystallisation matrix screen

Figure 5.20: Close up of CDO in hanging drop.

All further crystallisation work had to be put on hold until a buffer was found where CDO was stable. By the time a buffer was found, it was not possible to resume the crystallography work due to time constraints.

5.3.4 Nuclear Magnetic Resonance

A preliminary concentrated and purified sample of cysteine dioxygenase in 200mM NaCl/100mM PO_4 was submitted for a ^1H as a feasibility test. This experiment was performed before conditions were found that prevented aggregation and the entire process from expression to concentration of purified sample had to be performed in one day. Although no structural information was expected, resonant signals were detected in the 6.5 to 8.5 ppm region (see Figure on page 135). These are signals associated with aromatic groups, structures not present in the buffer and only attributed CDO aromatic groups. This is the second non-theoretical result that indicated that this protein can be sufficiently concentrated to be assigned by NMR; the first came from circular dichroism data that indicated that there was enough secondary structure to allow sufficient separation of multi-dimensional NMR data.

CDO was later expressed in minimal media containing $^{15}NH_3Cl$ to incorporate an ^{15}N label along the entire protein backbone of the protein. The protein was purified under conditions shown to keep the protein mono-disperse in DLS experiments. The result of the heteronuclear single quantum coherence (HSQC) revealed a folded protein that has sufficient secondary structure to be considered for assignment (Figure 5.22); An unfolded/misfolded protein typically has all its resonant signals overlapping in one area of the spectra. The signal was very weak and required a 24hr integration time for the sample. There was only a signal representing approximately 50% of the protein

(See Table A.1 on page 177); this was partly due to to concentration of the sample and the pH. There was also a problem supplementing the sample with 10% D_2O which resulted in too swift a change in the ionic strength of the buffer. Stepwise changes in ionic strength and pH invariably lead to aggregation. However, it was possible to rescue this sample early and to obtain a signal by filtration through a 0.22μm filter.

Detection of a ^{1}H-^{15}N bond is not just dependant upon protonation of the nitrogen; the rate of disassociation is also important. Regions of a protein buried within the structure are often stabilised by hydrogen bonds. The rate of disassociation in these regions is very slow. For amine ligands and nitrogens along the amide backbone of the protein involved in hydrogen binding, the HSQC resonances are large. Circular dichroism results indicated that 50% of the peptide has ordered structure and NMR indicated a protein with roughly half the protein is internalised and folded. PHDsec, a program that predicts secondary characteristics of a protein, was run on the CDO sequence and the amino acid residues were shaded using TEXshade (Figure A.17, page 180); this was done to see if there was a correlation between amide containing residues (Table A.1, page 177) and probable α-helical/β-sheet regions. The recombinant protein was not used in the analysis because the fusion protein is not buried within the protein by virtue of the fact it can bind the NTA-sepharose and is not sterically hindered. Because the fusion protein is free in solution, it is also unlikely to show strong resonances in the HSQC as it is not internalised. Half of the amide containing residues reside within or adjacent to areas of predicted secondary structure.

A second HSQC spectra was performed on a ^{15}N labelled sample and several more resonances were observed (Figure 5.23). It was impossible to determine whether these weak resonances were signals of the remaining nitro-

gens or whether they are noise. It is also impossible to assign a structure without analysis of a ^{13}C labelled protein. However, given the concentration of the protein, the ionic strength of the buffer and pH, months of dedicated NMR time would be required to gather spectral data. Several further months of computation and analysis would be required to perform the assignment. Consideration also needs to be taken to determine the stability of CDO over this prolonged time period.

In summary, it has been shown that recombinant CDO has an ordered structure. It is possible to purify a labelled protein and sufficiently concentrate it for NMR analysis. Conditions have been found to keep the protein disperse while preserving its structure. It is possible to consider assigning a structure to CDO, however the resources and time to do so are non-trivial.

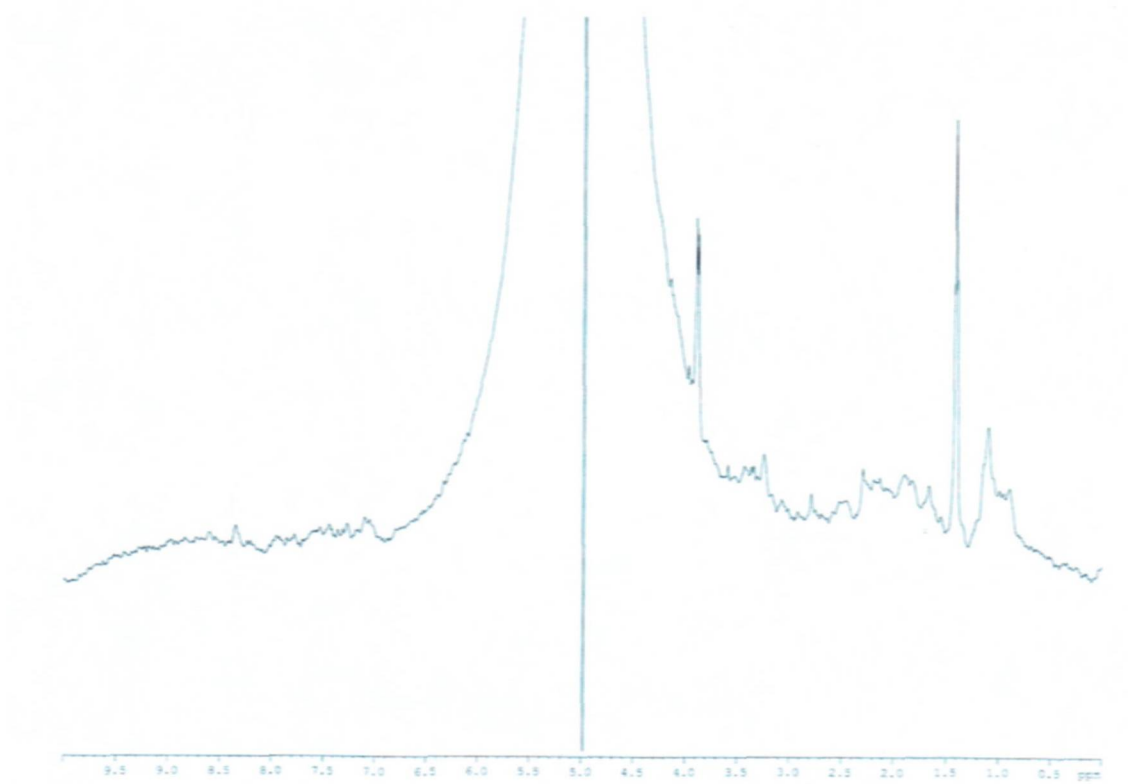


Figure 5.21: A preliminary 1-D spectra of cysteine dioxygenase

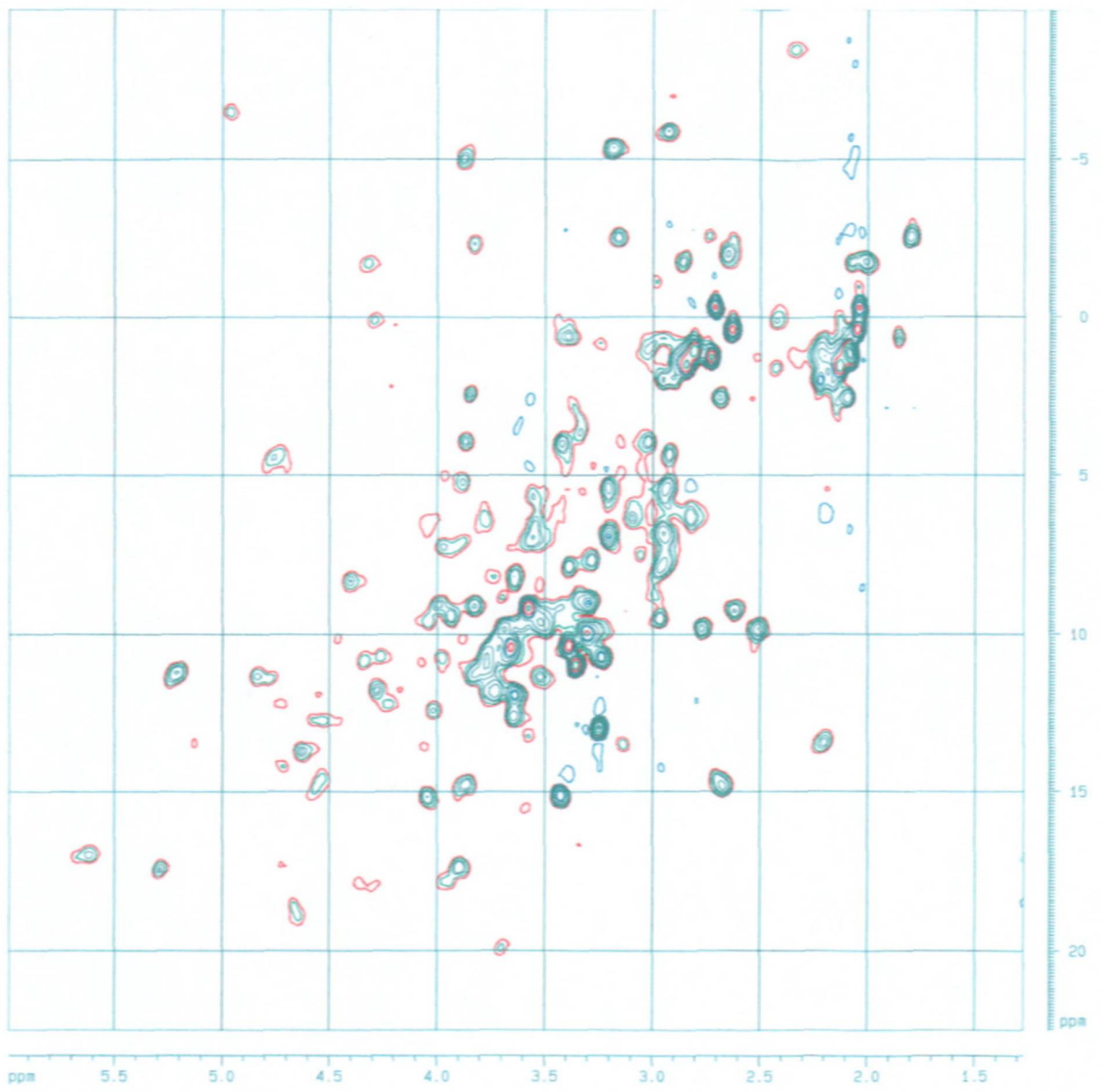


Figure 5.22: 2-D HSQC ^{1}H-^{15}N spectra of ^{15}N labelled cysteine dioxygenase in 320mM NaF, 1mM Glycine, 10mM PO_4^{-3}, 400μM EDTA pH 7.8

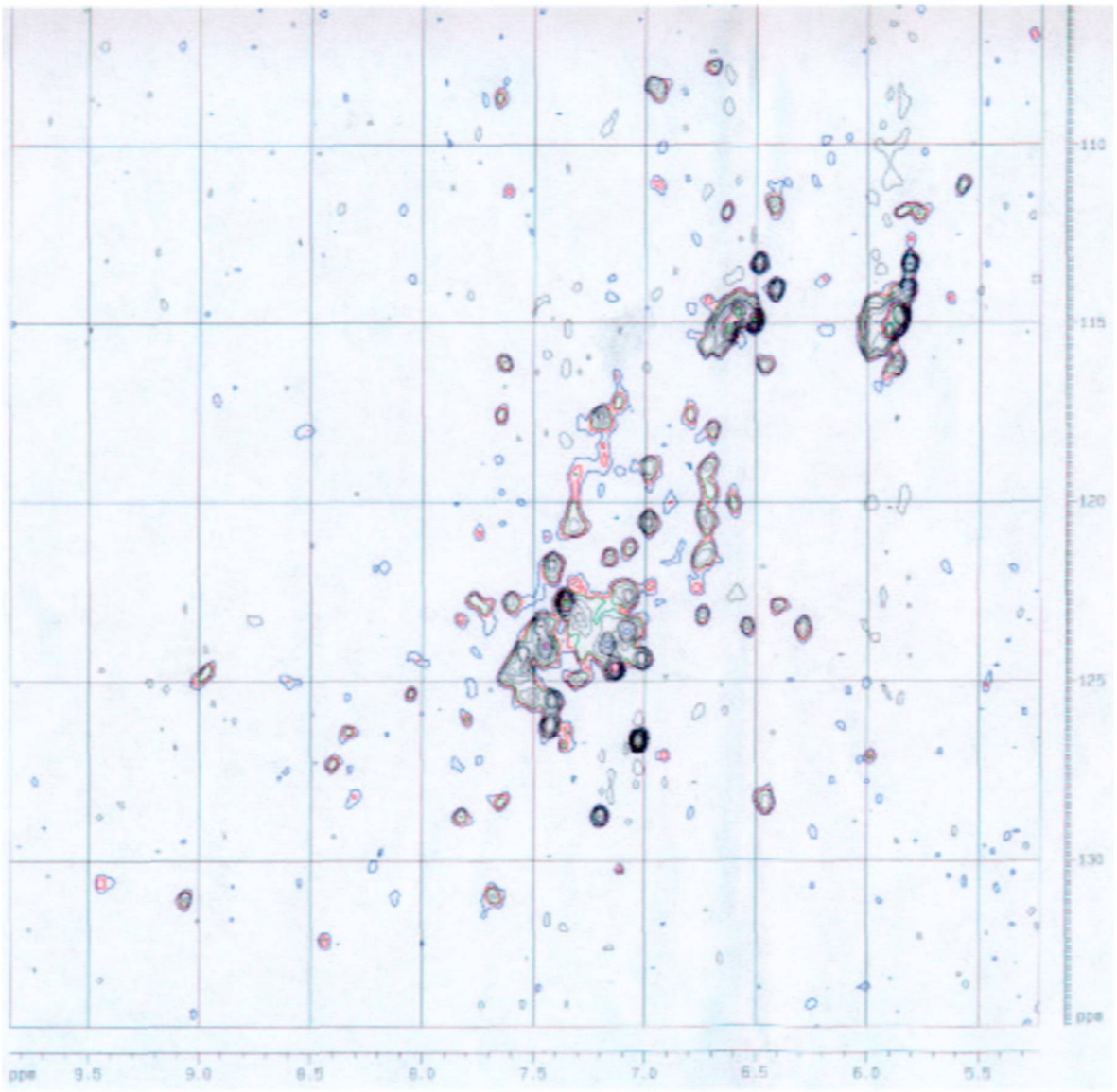


Figure 5.23: 2-D HSQC ^{1}H-^{15}N spectra of ^{15}N labelled cysteine dioxygenase in 320mM NaF, 1mM Glycine, 10mM PO_4^{-3}, 400μM EDTA pH 7.8 at increased sensitivity and integration time.

Chapter 6

Conclusions

A recombinant human CDO has been successfully expressed and highly purified. Experimental data has indicated that conditions exist that can keep the protein stable. Protein instability is always an impediment for further research and this discovery opens up new avenues for future research on the protein. It was not possible to fully take advantage of the light scattering data to get a form factor for CDO because it was not possible to precisely measure and regulate variables as temperature and viscosity. Small angle x-ray scattering (SAXS) would have been more appropriate for topographical measurements.

Circular dichroism experiments have show the protein to have a discrete secondary structure that parallels what would be expected from sequence predictions. Functionality tests of recombinant CDO have shown the protein to be non-functional. It could be argued that the prokaryotic expression system used does not incorporate Fe^{++}; however, EDTA has been shown to quench the reaction with the endogenous CDO. This indicates that a disassociation constant exists and the iron can bind reversibly. It is also possible that buffers chosen for the stability of CDO were not optimal for its function. Functionality

problems are more likely to be due to a regulatory issue than with misfolding or fusion protein addition. To expound on this issue a little further, several programmes exist to predict possible modifications sites on a protein. There are several sites on CDO that have been identified as potential glycosylation sites and regulatory domains. The sequence (Figure A.18) in the appendix illustrates the types of domains as well as their locations .

Experimental data has shown that the secondary structures exist in the same percentages and it is assumed that their distributions are in the same locations as well. The predicted glycosylation sites are generally located more toward tail end of the final transcript and their predicted probability is low (Figure A.1, A.4, and A.5 page 144). This indicated that at least the first 133 residues are unmodified. Potential kinase phosphorylation sites are more evenly distributed about the protein; sites that affect protein activity. It seems more probable that eukaryotic kinases were absent in the prokaryotic expression system.

Cysteine dioxygenases have been discovered in *schistosoma japonicum* and *drosophila melanogaster* that contain an additional nine and twenty upstream residues respectively. Also, some of the most conserved regions of the protein are stretches that lie between secondary structures; random coil domains. There is a high probability that these domains are external regions. The less conserved sites are domains with predicted post-translational sites on the human protein. Figure 5.4 (page 112) and the following sequence illustrates these contentions.

```
                                   α1
1  .........MEQTEVLKPRTLADLIRILHQLFAGD  cydx_human
1  .........MERTELLKPRTLADLIRILHELFAGD  cydx_rat
1  .........MEHTEVVVPKTLDELIAILHQIFESD  cydx_gmir
1  ...................MISFVQLVVQIREIFQQK  cydx_cele
1  MSMYTHQSNNELIPLKTVSTLNDLIKTIRIIFNQK  cydx_sjap
```

```
                α2                  α3
27  EVNVEEVQAIMEAYESDPIEWAMYAKFDQYRYTRN  cydx_human
27  EVNVEEVQAVLEAYESNPAEWALYAKFDQYRYTRN  cydx_rat
27  NINVEEVQQIMESYESNPQDWNKFAKFDQYRYTRN  cydx_gmir
19  LIDVDEVMKLMASYKSNANEWRRFAIFDMNKYTRN  cydx_cele
36  EINVNEIHKILNDFQCDFTEWQKYIYFNKTHYTRN  cydx_sjap

                β1
                 →                  α4
62  LVDQGNGKFNLMILCWGEGHGSSIHDHTNSHCFLK  cydx_human
62  LVDQGNGKFNLMILCWGEGHGSSIHDHTDSHCFLK  cydx_rat
62  LVDEGNGKFNLMILCWGEGHGSSIHDHTDSHCFMK  cydx_gmir
54  LVDVGNGKYNLMILCWGPGMASSVHDHTDAHCFVK  cydx_cele
71  LIDEGNGRYNLFLLCWSEDQGTRIHDHSGAHCFVK  cydx_sjap

                α4                          α5
97  MLQGNLKETLFAWPD....KKSN........EMVK  cydx_human
97  LLQGNLKETLFDWPD....KKSN........EMIK  cydx_rat
97  LLQGQLKETLFEWPE....GEQN.......GEMVQ  cydx_gmir
89  ILDGELTETKYAWP.....RKRH.......VPLDI  cydx_cele
106 LIKGCIKETIFEWPKYFTVEKSNYSINQIDLPLTV  cydx_sjap

                                β2          β3
                                 →           →
                α5
120 KSERVLRENQCAYINDSIGLHRVENISHTEPAVSL  cydx_human
120 KSERTLRENQCAYINDSIGLHRVENVSHTEPAVSL  cydx_rat
121 KSRDPDGKQGCLHK.....................  cydx_gmir
112 SENKTYGMNGVSYMNDELGLHRMENLSHSNGAVSL  cydx_cele
141 KSVSEMRPGDVTYMHDKIGIHRLHNPSTTETAITL  cydx_sjap

                                    β4
                                     →
155 HLYSPPFDTCHAFDQRTGHKNKVTMTFHSKFGIRT  cydx_human
155 HLYSPPFDTCHAFDQRTGHKNKVTMTFHSKFGIRT  cydx_rat
135 ......................................  cydx_gmir
147 HLYIPPYSTCNAFDERTGKKTQCTVTFYSKYGKKV  cydx_cele
176 HLYFPPYTNSMIFEESTSRMKKMDVTFHSKFGKQI  cydx_sjap

190 PNATSGSLENN  cydx_human
190 PFTTSGSLENN  cydx_rat
135 ...........  cydx_gmir
182 DYRGSKNGN..  cydx_cele
211 PQ.........  cydx_sjap

    X non conserved
    X similar
    X conserved
    X all match
```

Attempts were made to approach solving the structure quantitatively. Crystallography and NMR attempts were hampered by the stability of the protein. Once a suitable buffer was developed, time constraints prevented the resumption of crystallography experiments. The NMR assays were more successful.

Initial ^{1}H NMR spectra suggested that the protein could be concentrated to the point where it could be detected. Two dimensional NMR of ^{15}N labelled CDO confirmed the circular dichroism experiments that indicated a protein that is folded (as opposed to misfolded or aggregated). Further HSQC experiments yielded even more data. It was not possible to fully assign the 3D structure of the protein. Due to the concentration, an amount of NMR time needed to integrate the signal was not made available. The time needed to achieve this could have been minimised with a larger magnet and a cryoprobe but both options were unavailable. The structure of this protein could likely be done within a few months if the resources become available.

Chapter 7

Acknowledgements

- I would like to thank Ruedy Allemann and the Chemistry Department for after-hours access and generous use of the NMR facilities, the mass spectroscopy facilities and the Jasco circular dichroism instrument. I would also like to acknowledge Dr. Taylor for ground-breaking performance in pedantics and matriarchy.
- I would like to thank Eva Hyde and Yuan Gao for use of their NMR instrument for labelled CDO and their assistance in acquiring protein data.
- Special thanks to Trevor Rutherford, Mike O'Donnell, David Miller, Rakesh Bargota, John Sanvoisin and the rest of the Chemistry gang for our dialogues in protein chemistry.
- Thank you to Steve Young for allowing me to use his Biacore instrument and for being very approachable.
- A special thank you to the CRC centre and David Oupickey for access to light scattering instrumentation and after-hours access to the laboratory.

- I would also like to thank John Fox for his assistance on nucleotide synthesis, sequencing, and other biochemical techniques. A behind the scenes technical powerhouse with no reservations in assisting any laboratory problems.

- I would like to acknowledge Nick Fitch's early work in the sub-cloning of the CDO gene.

- I would like to especially thank Adrian Williams with the Department of Clinical Neuroscience for generously supporting this research and my stay in England. I will be forever grateful and hope that someday I will be able to return something to my Alma Matter.

- Lastly, but not least, I would like to especially thank David Ramsden as an advisor, friend, and for directing my research. It has been a refreshing change to work with someone who is supportive and enthusiastic about aggressively pursuing disparate avenues of research. I will miss working in his laboratory.

Appendix A

Sequence Specific Details of CDO

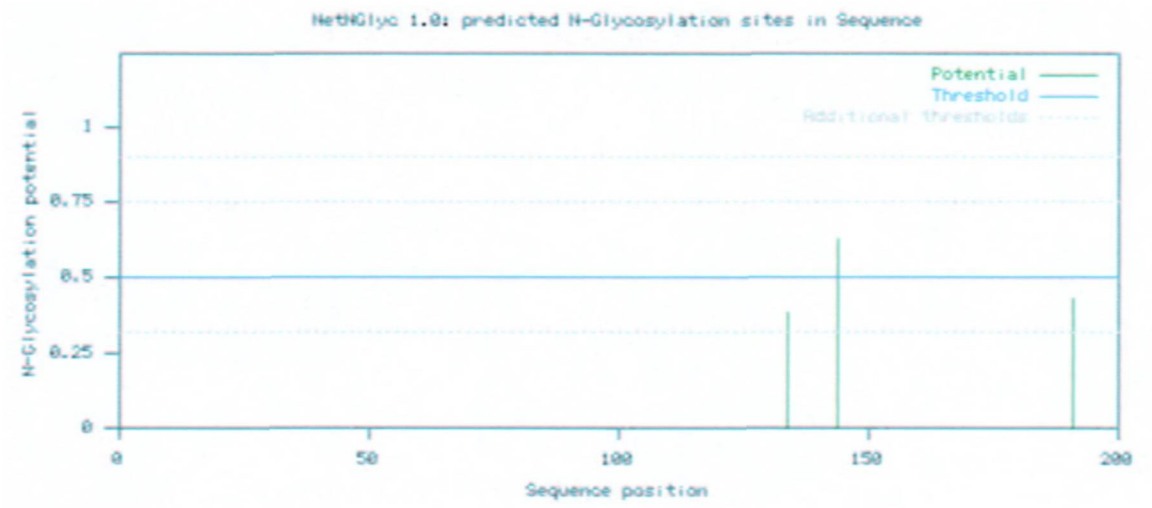


Figure A.1: Predicted N-Glycosylation of CDO

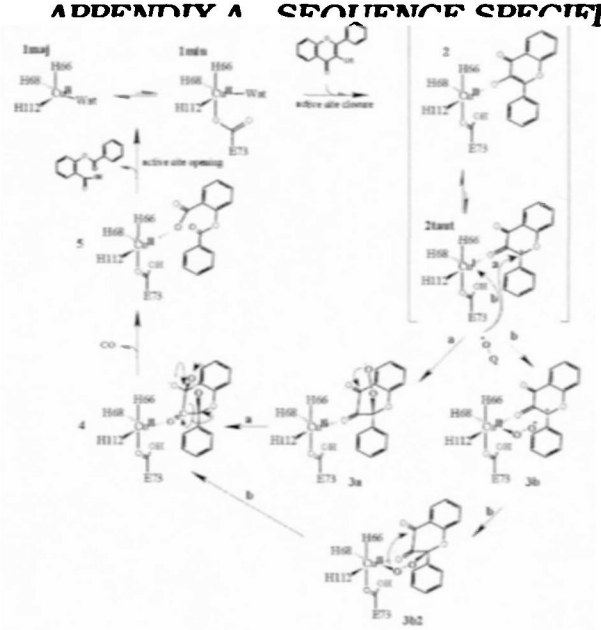


Fig. 4. Proposed reaction mechanism for 2,3QD-mediated dioxygenation of flavonols.

Compared with KMP, the natural substrate QUE possesses an additional hydroxyl group at the 3′ position of the B-ring. Two conformations are therefore possible for QUE while preserving an extended electronic delocalization: *cis*-QUE, in which the 3′-OH group is *cis* to the O-heteroatom of the C-ring, and *trans*-QUE, in which the 3′-OH group is on the opposite side. In the x-ray structure of quercetin (30), crystal contacts induce QUE to adopt a *cis* conformation. In the 2,3QD-QUE complex, QUE is instead observed in the *trans* conformation only (Fig. 2*A*). In this geometry, the 3′OH group fills a void formed by the backbone atoms of Asn^{74} and the side chain of Glu^{73}. It is not hydrogen bonded to any atom. Unfavorable interactions between the 3′OH group and the side chains of Met^{123} and Thr^{53} would exist if the *cis* conformation was adopted.

Discussion

Formation of the E·S Complex and Role of Glu^{73}. A possible mechanism for 2,3QD-mediated dioxygenation of flavonols is shown in Fig. 4. It combines the information from the present and previous crystallographic studies (6, 29), EPR investigations (10), and bio-mimetic studies (12–14). The initial step of the catalytic cycle requires complexation of the flavonol substrate to the copper ion. On substrate binding, the heterogeneous distorted tetrahedral (major)/distorted trigonal bipyramidal (minor) native copper (structures **1maj** and **1min** in Fig. 4, respectively) gains structural order, changing to a square pyramidal geometry represented by structure **2**. Concomitantly, the portion of the linker peptide located in front of the active site entrance becomes ordered and assists, via the van der Waals interaction between Pro^{164} and the substrate molecule, the stabilization of the loaded substrate.

The formation of the E·S complex likely entails the loss of the 3OH flavonol proton. Although deprotonation of the 3OH group may result from the lowering of its pKa on its complexation to the copper ion or simply from exchange to water, the 1,000-fold decrease in activity on a Glu^{73}→Gln mutation (L. M. Kooter, personal communication) and the proper positioning of the side chain of Glu^{73} strongly support a role for this residue in substrate deprotonation. The observation of a close interaction between the $Glu_{Oε2}{}^{73}$ atom and the O3 atom of the flavonol (2.43 Å in 2,3QD·KMP and 2.66 Å in 2,3QD·QUE) suggests in addition that Glu^{73} might be bound to the copper ion retaining the 3OH proton. In mechanistic terms, the retention of the 3OH proton in the vicinity of the copper center could be advantageous for protonating the leaving depside product (step 5 → 1). Additionally, a Glu residue bound to the metal is an effective way to allow, if needed, small adjustments in coordination geometries.

Dioxygen Attack. The reaction catalyzed by 2,3QD is a spin-forbidden process. The triplet ground state O_2 reacts with a singlet ground state flavonol to form singlet ground state products (depside and CO). Such a reaction requires some form of activation. Similarly to what is observed in the case of the Fe^{3+}-dependent intradiol-type catechol dioxygenases, EPR (10) and x-ray absorption spectroscopy (15) experiments provide evidence that the copper center of 2,3QD in complex with its flavonol substrates retains a formal high oxidation state. Normally, complexes of Cu^{2+}, even when coordinatively unsaturated, do not bind O_2 (36). This result is not surprising because good ligands for high oxidation state metal ions are usually relatively strong bases, and O_2 is not. The spin-forbiddeness of the reaction can, nonetheless, be circumvented by allowing the existence of a flavonoxy radical-Cu^+ valence tautomer (**2taut**), arising from the flow of one electron from the flavonol into the copper. The radical center of this complex can react with the diradical O_2 (route **a** in Fig. 4). The pyramidalization of the flavonol C2 atom, independently observed in both E·S complexes, is consistent with the stabilization of the unpaired electron on this atom. Analysis of the shape of the active site cavity shows that room is available for a dioxygen molecule to directly attack the flavonol C2 atom from the copper side. O_2 binding would then produce the peroxide **3a**, which, after nucleophilic attack on the C4 atom, would generate the endoperoxide **4**. In these peroxidic intermediates, the C2 atom is sp^3 hybridized. The pyramidalization observed in the bound substrates indicates that, apart from likely radical stabilization, the protein structure also preorganizes a conformation of the E·S complex favorable for subsequent intermediates.

The existence of an active flavonoxy radical-Cu^+ complex **2taut** allows, in principle, also to hypothesize a reaction mechanism in which O_2 is activated by coordination to the Cu^+ ion (route **b**) (14). An *a priori* prediction of whether this complex possesses the reducing equivalents to do so is difficult. However, from a geometrical point of view a binding site for O_2 is available in the E·S complex. Differently from what was generally assumed from the chelating behavior of flavonols, the protein matrix constrains the flavonol to bind to the metal with only its O3 atom. As a consequence, the coordination position *trans* to His^{68} might be used by the dioxygen molecule to directly bind to the metal (**3b**). The crystal structure of 2,3QD complexed with kojic acid (29) shows that an oxygen atom can indeed occupy this sixth coordination position, producing a pseudooctahedral copper site. Owing to this result, we have tried to model the relevant structures along the reaction path b: **3b** and **3b2**. We find that, although formation of **3b2** from **3b** is geometrically feasible, intramolecular nucleophilic attack of the oxygen atom, bound to the copper on the C4 flavonol atom required to generate the endoperoxide **4** from **3b2**, seems more difficult. The model indicates that the attacking oxygen atom would be at about 3.5–4.0 Å from the C4 atom and not above the carbonyl plane, which represents the most likely direction of attack. In contrast, such an arrangement seems favored in **3a**.

On the basis of geometrical considerations, we therefore conclude that the mechanism of substrate activation (route **a**), which assumes the formation of the endoperoxide **4** from **3a**, is favored. Experiments of radical-initiated oxygenation of flavonols support a mechanism of substrate activation (37). In the presence of free radicals such as TEMPO (2,2,6,6-tetramethyl-

Figure A.2. Proposed mechanism of quercetin 2,3-dioxygenase [104]

BIOCHEMISTRY

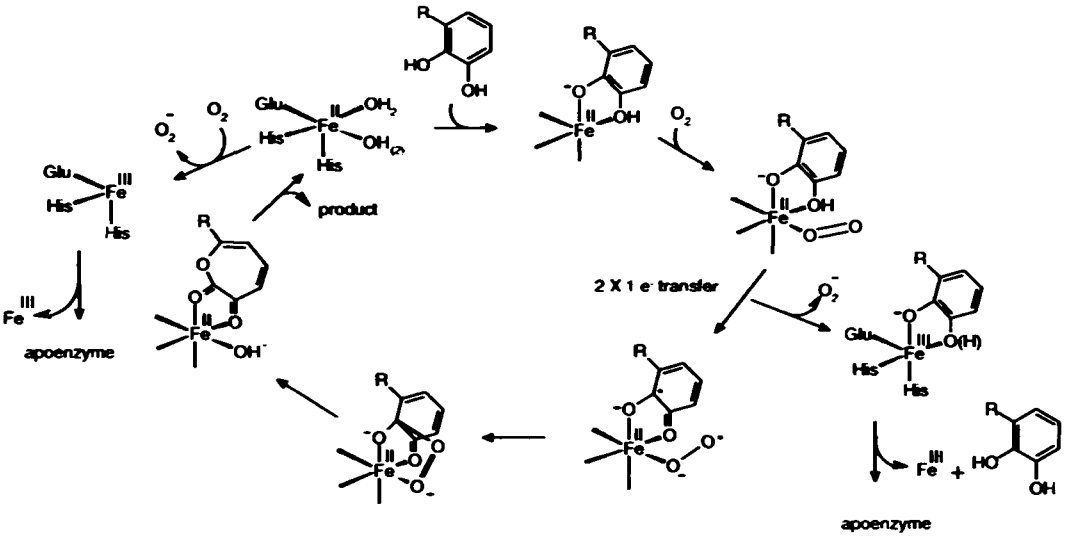


FIG. 6. **General mechanism of inactivation of extradiol dioxygenases.** Only those intermediates for which some experimental evidence exists are shown. The exact step at which superoxide dissociates from the ternary complex has not been determined. The ligands in the ferric form of the enzyme are unknown.

3-chlorocatechol, respectively. In contrast, oxygraph assays yielded a partition coefficient of 11 ± 2.

DISCUSSION

DHBD is typical of extradiol-type dioxygenases in that it is subject to inactivation during the steady-state cleavage reaction (35). The present analysis indicates that this inactivation in DHBD requires the formation of the EAO_2 ternary complex. In particular, the rates of inactivation of EA, AEA, and EP (j_2, j_4, and j_5 in Fig. 2) are negligible with respect to the rate of inactivation during steady-state turnover. Thus, DHBD is not inactivated by chelation of the active site Fe(II) by catecholic substrates (19). Although free DHBD is subject to significant inactivation by O_2, the apparent rate constant of this inactivation is significantly lower than the rate constant of inactivation by the preferred catecholic substrate of the enzyme, DHB. The current study does not rule out the possibility that the AEA and EP forms are unstable in the presence of O_2. However, given the DHBD high K_m value for O_2 (1.3 mM (35)), high K_{iA} for DHB (3 mM (35)), and high K_i values for HOPDA (~ 3 mM), such inactivation seems unlikely to be significant under the conditions studied.

Further analysis of the mechanism-based inactivation of DHBD revealed that it is similar in nature to the O_2-dependent inactivation of DHBD in the absence of catecholic substrate, arising principally from the oxidation of the active site Fe(II) to Fe(III). Thus, EPR and absorption spectroscopy data demonstrate the formation of Fe(III) in samples of inactivated enzyme, and anaerobic incubation of the inactivated enzyme with Fe(II) and DTT restored the activity. The activity was partially restored upon incubation of desalted samples of inactivated DHBD with DTT alone, indicating that some of the oxidized Fe(III) remained bound to the protein. Although no association constants of an extradiol enzyme for Fe(III) and Fe(II) have been reported, the apparently higher affinity of DHBD for Fe(II) than for Fe(III) is consistent with the crystallographic data of DHBD from *Pseudomonas* sp. KKS102, in which a more intense electron density was observed at the active site when the iron was reduced (55). Moreover, the oxidation of the active site Fe(II) of C23O by H_2O_2 resulted in the immediate release of Fe(III) (57).

The present studies suggest that the mechanism-based inactivation of DHBD does not involve covalent modification, as judged by a lack of change to the molecular mass of DHBD inactivated in a number of ways. Moreover, DHBD was readily reactivated in cells in the absence of protein synthesis. Thus, inactivation does not involve hydroxylation of an active site residue as observed in the O_2-dependent inactivation of an α-ketoglutarate-dependent oxygenase (58), which like DHBD has a catalytically essential mononuclear iron bound to the enzyme by a 2-histidine 1-carboxylate facial triad (59). These results also demonstrate that although 3-chlorocatechol is a very potent mechanism-based inactivator and that the DHBD-catalyzed cleavage of 3-chlorocatechol produces an acyl halide, the inactivation does not involve covalent modification by the acyl chloride as has been proposed for C23O (20). Indeed, one study reported that the inactivation of C23O by 3-chlorocatechol also involves oxidation of the active site Fe(II) (60).

A straightforward explanation of the mechanism-based inactivation of DHBD involves the dissociation of superoxide from the EAO_2 ternary complex. In a proposed catalytic mechanism, formation of the EAO_2 ternary complex is followed by successive electron transfer steps from the Fe(II) to the bound O_2 and from the bound catecholate to the iron. C-O bond formation at C-2 in the resulting semiquinone-Fe(II)-superoxide intermediate yields an iron-alkylperoxo intermediate that undergoes a Criegee rearrangement (3). Mechanism-based inactivation could arise from dissociation of the bound superoxide before electron transfer from the catecholate to the iron or before C-O bond formation between the bound superoxide and semiquinone. Thus, catecholic substrates that slow either step either through steric or electronic factors would be good mechanism-based inactivators. For example, electron transfer between 3-chlorocatechol and Fe(III) would be slower than between 3-methylcatechol and Fe(III) due to the expected higher reduction potential of a catechol with an electron-withdrawing substituent. Consistent with this hypothesis, catechols with electron-withdrawing substituents were not cleaved in a model extradiol cleavage reaction (61).

Failure to detect superoxide in the inactivation of DHBD by 3-chlorocatechol seems to be due to the rapid reaction of superoxide with the catechol, possibly before their diffusion from the active site channel of the enzyme. The current studies with xanthine oxidase demonstrate that 3-chlorocatechol is highly reactive with superoxide, consistent with the known role of catechols as superoxide scavengers (62, 63). Moreover, ferric

Figure A.3: Proposed mechanism of 2,3-dihydroxybiphenyl 1,2-dioxygenase [105]

NetNGlyc 1.0 Prediction Results

Asn-Xaa-Ser/Thr sequons in the sequence output below are highlighted in blue. Asparagines predicted to be N-glycosylated are highlighted in red. For further details on the output format, click here.

```
# Predictions for N-Glycosylation sites in 1 sequence

-----------------------------------------------------------------------------

Warning:   This sequence may not contain a signal peptide !!

       Proteins without signal peptides are unlikely to be exposed to the
       N-glycosylation machinery and thus may not be glycosylated (in vivo)
       even though they contain potential motifs.
  # SignalP-NN euk predictions   (SignalP Output explanation)
  # name        Cmax  pos ?  Ymax  pos ?  Smax  pos ?  Smean ?
  Sequence              0.297  25 N  0.170  25 N  0.388  24 N  0.163 N
-----------------------------------------------------------------------------


Name:  Sequence          Length:  200
MEQTEVLKPRTLADLIRILHQLFAGDEVNVEEVQAIMEAYESDPIEWAMYAKFDQYRYTRNLVDQGNGKFNLMILCWGEG
HGSSIHDHTNSHCFLKMLQGNLKETLFAWPDKKSNEMVKKSERVLRENQCAYINDSIGLHRVENISHTEPAVSLHLYSPP
FDTCHAFDQRTGHKNKVTMTFHSKFGIRTPNATSGSLENN
..                                                                            ..
..                                                                            N..
..                                                                            ..

(Threshold=0.5)
-----------------------------------------------------------------------------
SeqName      Position  Potential   Jury    NGlyc
                                 agreement result
-----------------------------------------------------------------------------
Sequence     134 NDSI   0.3820     (8/9)
Sequence     144 NISH   0.6256     (6/9)   +
Sequence     191 NATS   0.4323     (7/9)
-----------------------------------------------------------------------------
```

Figure A.4: Predicted N-glycosylation sites on human CDO using NetNGlyc

NetOGlyc 2.0 Prediction Results

```
Name:  Sequence          Length:  200
MEQTEVLKPRTLADLIRILEQLFAGDEVNVEEVQAIMEAYESDPIEWAMYAKFDQYRYTRNLVQGGNGKFNLMILCMGEG
EGSSIEDETNSECFLKMLQGNLKETLFAMPDKKSNEMVKKSERVLRENQCAYINDSIGLERVENISETEPAVSLELYSPP
FDTCHAFDQRTGEKNKVTMTFESKFGIRTPNATSGSLENN
.............................................. .. . ..... ......                 .......
.............................................  .    . . .   .....                 ...... .
.............................T.....  .
```

Name	Residue	No.	Potential	Threshold	Assignment
Sequence	Thr	4	0.2848	0.5393	
Sequence	Thr	11	0.0088	0.6214	
Sequence	Thr	59	0.0084	0.6213	
Sequence	Thr	89	0.1981	0.6759	
Sequence	Thr	105	0.0085	0.6017	
Sequence	Thr	148	0.2550	0.6480	
Sequence	Thr	163	0.0097	0.6299	
Sequence	Thr	171	0.0563	0.5796	
Sequence	Thr	178	0.1020	0.6251	
Sequence	Thr	180	0.0096	0.6477	
Sequence	Thr	189	0.2316	0.6097	.
Sequence	Thr	193	0.6103	0.5679	T

Name	Residue	No.	Potential	Threshold	Assignment
Sequence	Ser	42	0.0218	0.5507	
Sequence	Ser	83	0.0336	0.6164	
Sequence	Ser	84	0.0007	0.6636	
Sequence	Ser	91	0.0008	0.6157	
Sequence	Ser	114	0.0011	0.5259	
Sequence	Ser	121	0.0013	0.5601	
Sequence	Ser	136	0.0004	0.6524	
Sequence	Ser	146	0.0176	0.6309	
Sequence	Ser	153	0.0874	0.6965	
Sequence	Ser	158	0.1661	0.6319	
Sequence	Ser	183	0.0629	0.6313	
Sequence	Ser	194	0.0060	0.5832	
Sequence	Ser	196	0.3918	0.5866	

Figure A.5: Predicted O-glycosylation sites on human CDO using NetOGlyc

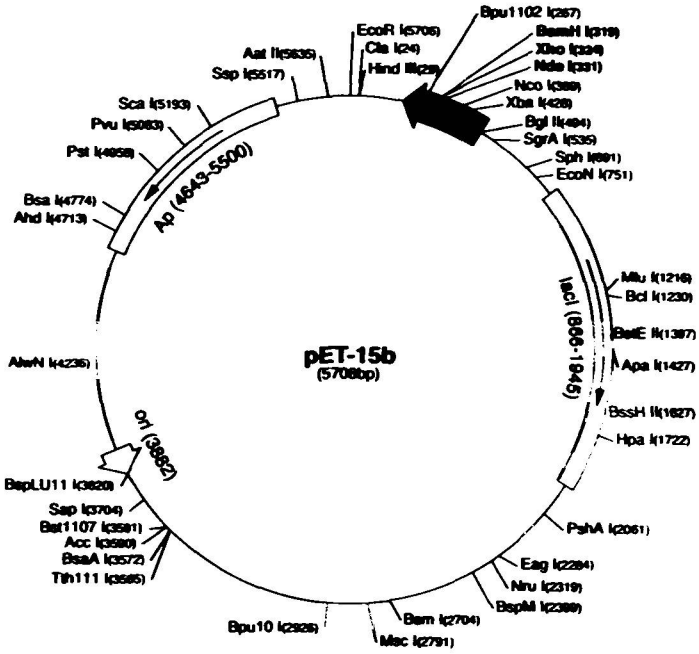
Novagen

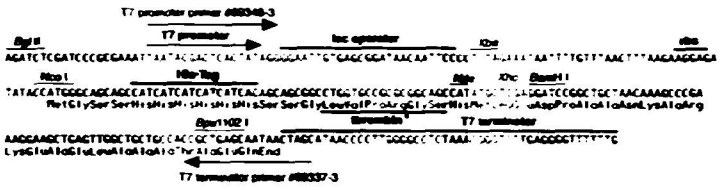
pET-15b Vector

TB045 5/99

The pET-15b vector (Cat. No. 69661-3) carries an N-terminal His•Tag® sequence followed by a thrombin site and three cloning sites. Unique sites are shown on the circle map. Note that the sequence is numbered by the pBR322 convention, so the T7 expression region is reversed on the circular map. The cloning/expression region of the coding strand transcribed by T7 RNA polymerase is shown, below.

pET-15b sequence landmarks	
T7 promoter	463-479
T7 transcription start	452
His•Tag coding sequence	362-380
Multiple cloning sites (*Nde* I - *Bam*H I)	319-335
T7 terminator	213-259
lacI coding sequence	(866-1945)
pBR322 origin	3882
bla coding sequence	4643-5500





pET-15b cloning/expression region

Figure A.6: Novagen pET15b vector

```
The following information has been received by the server:
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
```

```
reference predict_h16487 (May 29, 2002 06:22:58)
reference pred_h16487 (May 29, 2002 05:43:03)
PPhdr from: anon@pugh.bip.bham.ac.uk
PPhdr resp: HTML
PPhdr orig: HTML
PPhdr want: ASCII
PPhdr password(###)
prediction of: - default prediction of: - PHDsec PHDacc PHDhtm ProSite S
return msf format
ret store
# default: single protein sequence
MEQTEVLKPR TLADLIRILH QLFAGDEVNV EEVQAIMEAY ESDPIEWAMY
AKFDQYRYTR NLVDQGNGKF NLMILCWGEG HGSSIHDHTN SHCFLKMLQG
NLKETLFAWP DKKSNEMVKK SERVLRENQC AYINDSIGLH RVENISHTEP
AVSLHLYSPP FDTCHAFDQR TGHKNKVTMT FHSKFGIRTP NATSGSLENN
```

```
Result of PROSITE search (Amos Bairoch):
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

please quote: A Bairoch, P Bucher & K Hofmann: The PROSITE database,
its status in 1997. Nucl. Acids Res., 1997, 25, 217-221.
```

```
--------------------------------------------------------
Pattern-ID: ASN_GLYCOSYLATION PS00001 PDOC00001
Pattern-DE: N-glycosylation site
Pattern:    N[^P][ST][^P]
   134      NDSI
   144      NISH
```

```
   191       NATS

Pattern-ID: PKC_PHOSPHO_SITE PS00005 PDOC00005
Pattern-DE: Protein kinase C phosphorylation site
Pattern:    [ST].[RK]
   121       SER

Pattern-ID: CK2_PHOSPHO_SITE PS00006 PDOC00006
Pattern-DE: Casein kinase II phosphorylation site
Pattern:    [ST].{2}[DE]
   11        TLAD
   84        SIHD
   146       SHTE

Pattern-ID: MYRISTYL PS00008 PDOC00008
Pattern-DE: N-myristoylation site
Pattern:    G[^EDRKHPFYW].{2}[STAGCN][^P]
   195       GSLENN

Pattern-ID: LUM_BINDING PS00693 PDOC00581
Pattern-DE: Riboflavin synthase alpha chain family signature
Pattern:    [LIVMF].{5}G[STADNQ][KREQIYW]VN[LIVM]E
   19        LHQLFAGDEVNVE
```

```
Result of ProDom domain search (Sonnhammer; Corpet, Gouzy, Kahn):
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

- please quote: ELL Sonnhammer & D Kahn, Prot. Sci., 1994, 3, 482-492
```

```
--- ------------------------------------------------------------
--- Results from running BLAST against PRODOM domains
---
--- PLEASE quote:
```

```
---       F Corpet, J Gouzy, D Kahn (1998).  The ProDom database
---       of protein domain families. Nucleic Ac Res 26:323-326.
---
--- BEGIN of BLASTP output
BLASTP 1.4.7 [16-Oct-94] [Build 12:52:03 Oct 30 1994]

Reference:  Altschul, Stephen F., Warren Gish, Webb Miller, Eugene W. My
and David J. Lipman (1990).  Basic local alignment search tool.  J. Mol.
215:403-10.

Query=  prot (#) ppOld, default: single protein sequence
     /home/phd/server/work/predict_h16487
        (200 letters)

Database:  prodom_00_1
           174,952 sequences; 19,895,393 total letters.
Searching.................................................done

                                                                    Sma
                                                                      S
                                                              High  Prob
Sequences producing High-scoring Segment Pairs:              Score  P(N)

 PD016803 p2000.1 (3) CYDX(2) Q20893(1)  // DIOXYGENASE C...   796  1.6e
 PD094785 p2000.1 (1) O32085_BACSU // YUBC PROTEIN              58  5.9e



>PD016803 p2000.1 (3) CYDX(2) Q20893(1)  // DIOXYGENASE CYSTEINE CDO
  OXIDOREDUCTASE IRON CDO-I PUTATIVE
  Length = 180

 Score = 796 (366.3 bits), Expect = 1.6e-106, P = 1.6e-106
 Identities = 143/176 (81%), Positives = 154/176 (87%)

Query:    15 LIRILHQLFAGDEVNVEEVQAIMEAYESDPIEWAMYAKFDQYRYTRNLVDQGNGKFNLM
             LIR LH+LF GDEVNV+EVQ IMEAYES+P EW  YAKFDQYRYTRNLVDQGNGK+NLM
Sbjct:     1 LIRQLHELFQGDEVNVDEVQKIMEAYESNPNEWRRYAKFDQYRYTRNLVDQGNGKYNLM

Query:    75 LCWGEGHGSSIHDHTNSHCFLKMLQGNLKETLFAWPDKKSNEMVKKSERVLRENQCAYI
             LCWG GHGSSIHDHT+SHCFLK+L GNLKET +AWPDKK N M KK  R   EN CAY+
Sbjct:    61 LCWGPGHGSSIHDHTDSHCFLKILDGNLKETKYAWPDKKHNPMDKKENRTYGENGCAYM

Query:   135 DSIGLHRVENISHTEPAVSLHLYSPPFDTCHAFDQRTGHKNKVTMTFHSKFGIRTP 19
```

```
             D IGLHR+ENISHT PAVSLHLYSPP+DTCHAFDQRTGHKNK TMTFHSK+G RTP
Sbjct:   121 DEIGLHRMENISHTNPAVSLHLYSPPYDTCHAFDQRTGHKNKCTMTFHSKYGKRTP 17


>PD094785 p2000.1 (1) O32085_BACSU // YUBC PROTEIN
  Length = 161

 Score = 58 (26.7 bits), Expect = 5.9e-05, Sum P(3) = 5.9e-05
 Identities = 10/23 (43%), Positives = 16/23 (69%)

Query:   148 TEPAVSLHLYSPPFDTCHAFDQR 170
             +E  VSLH+YSPP +    F+++
Sbjct:   132 SERMVSLHVYSPPLEDMTVFEEQ 154

 Score = 43 (19.8 bits), Expect = 5.9e-05, Sum P(3) = 5.9e-05
 Identities = 8/41 (19%), Positives = 21/41 (51%)

Query:    67 NGKFNLMILCWGEGHGSSIHDHTNSHCFLKMLQGNLKETLF 107
             N +  ++++       +++HDH  S     +L+G L  +++
Sbjct:    56 NNELEIIVINIPPNKETTVHDHGQSIGCAMVLEGKLLNSIY 96

 Score = 39 (17.9 bits), Expect = 5.9e-05, Sum P(3) = 5.9e-05
 Identities = 7/14 (50%), Positives = 9/14 (64%)

Query:    54 DQYRYTRNLVDQGN 67
             DQY Y RN + + N
Sbjct:    44 DQYAYGRNAIYRNN 57


Parameters:
  E=0.001
  B=500

  V=500
  -ctxfactor=1.00

  Query                          -----  As Used  -----     -----  Computed
  Frame  MatID Matrix name       Lambda    K       H       Lambda    K
   +0      0   BLOSUM62          0.319   0.134   0.403     same    same

  Query
  Frame  MatID  Length  Eff.Length   E     S W   T  X     E2  S2
   +0      0      200       200    0.0010 86 3  11 22    0.24 32
```

```
Statistics:
  Query          Expected          Observed          HSPs       HSPs
  Frame  MatID  High Score       High Score       Reportable  Reported
   +0      0    62 (28.5 bits)  796 (366.3 bits)        4          4

  Query         Neighborhd  Word      Excluded    Failed   Successful  C
  Frame  MatID   Words      Hits        Hits    Extensions Extensions  E
   +0      0      5558     8373478     1785644     6571796    16037

  Database:  prodom_00_1
    Release date:  unknown
    Posted date:  5:56 PM EDT Jun 21, 2000
  # of letters in database:  19,895,393
  # of sequences in database:  174,952
  # of database sequences satisfying E:  2
  No. of states in DFA:  565 (56 KB)
  Total size of DFA:  116 KB (128 KB)
  Time to generate neighborhood:  0.01u 0.00s 0.01t  Real: 00:00:00
  Time to search database:  11.43u 0.05s 11.48t  Real: 00:00:11
  Total cpu time:  11.45u 0.07s 11.52t  Real: 00:00:11
--- END of BLASTP output
--- ------------------------------------------------------------
---
--- Again: these results were obtained based on the domain data-
--- base collected by Daniel Kahn and his coworkers in Toulouse.
---
--- PLEASE quote:
---       F Corpet, J Gouzy, D Kahn (1998).  The ProDom database
---       of protein domain families. Nucleic Ac Res 26:323-326.
---
--- The general WWW page is on:
----      ---------------------------------------
---       http://www.toulouse.inra.fr/prodom.html
----      ---------------------------------------
---
--- For WWW graphic interfaces to PRODOM, in particular for your
--- protein family, follow the following links (each line is ONE
--- single link for your protein!!):
---
http://www.toulouse.inra.fr/prodom/cgi-bin/ReqProdomII.pl?id_dom1=PD0168
http://www.toulouse.inra.fr/prodom/cgi-bin/ReqProdomII.pl?id_dom2=PD0168
http://www.toulouse.inra.fr/prodom/cgi-bin/ReqProdomII.pl?id_dom1=PD0947
http://www.toulouse.inra.fr/prodom/cgi-bin/ReqProdomII.pl?id_dom2=PD0947
```

```
---
--- NOTE: if you want to use the link, make sure the entire line
---       is pasted as URL into your browser!
---
--- END of PRODOM
--- ------------------------------------------------------------
```

The alignment that has been used as input to the network is:
~~~~~~

```
---
--- Version of database searched for alignment:
--- SWISS-PROT release 39.0 (5/00) with 85 249 proteins
---

--- ------------------------------------------------------------
--- MAXHOM multiple sequence alignment
--- ------------------------------------------------------------
---
--- MAXHOM ALIGNMENT HEADER: ABBREVIATIONS FOR SUMMARY
--- ID            : identifier of aligned (homologous) protein
--- STRID         : PDB identifier (only for known structures)
--- IDE           : percentage of pairwise sequence identity
--- WSIM          : percentage of weighted similarity
--- LALI          : number of residues aligned
--- NGAP          : number of insertions and deletions (indels)
--- LGAP          : number of residues in all indels
--- LSEQ2         : length of aligned sequence
--- ACCNUM        : SwissProt accession number
--- OMIM          : OMIM (Online Mendelian Inheritance in Man) ID
--- NAME          : one-line description of aligned protein
---
--- MAXHOM ALIGNMENT HEADER: SUMMARY
ID          STRID  IDE WSIM LALI NGAP LGAP LSEQ2 ACCNUM OMIM    NAME
cydx_human         100  100  200    0    0   200 Q16878 603943  CYSTEINE
cydx_rat            92   95  200    0    0   200 P21816         CYSTEINE
```
~~~~~~

```
---
--- MAXHOM ALIGNMENT: IN MSF FORMAT
MSF of: /home/phd/server/work/predict_h16487.hsspFilter from:    1 to:
 /home/phd/server/work/predict_h16487.msfRet  MSF:  200  Type: P 29-May-


 Name: predict_h1640    Len:    200  Check: 8892  Weight:  1.00
 Name: cydx_human       Len:    200  Check: 8892  Weight:  1.00
 Name: cydx_rat         Len:    200  Check: 9388  Weight:  1.00

//


               1                                                   50
predict_h1640  MEQTEVLKPR TLADLIRILH QLFAGDEVNV EEVQAIMEAY ESDPIEWAMY
cydx_human     MEQTEVLKPR TLADLIRILH QLFAGDEVNV EEVQAIMEAY ESDPIEWAMY
cydx_rat       MERTELLKPR TLADLIRILH ELFAGDEVNV EEVQAVLEAY ESNPAEWALY

               51                                                 100
predict_h1640  AKFDQYRYTR NLVDQGNGKF NLMILCWGEG HGSSIHDHTN SHCFLKMLQG
cydx_human     AKFDQYRYTR NLVDQGNGKF NLMILCWGEG HGSSIHDHTN SHCFLKMLQG
cydx_rat       AKFDQYRYTR NLVDQGNGKF NLMILCWGEG HGSSIHDHTD SHCFLKLLQG

               101                                                150
predict_h1640  NLKETLFAWP DKKSNEMVKK SERVLRENQC AYINDSIGLH RVENISHTEP
cydx_human     NLKETLFAWP DKKSNEMVKK SERVLRENQC AYINDSIGLH RVENISHTEP
cydx_rat       NLKETLFDWP DKKSNEMIKK SERTLRENQC AYINDSIGLH RVENVSHTEP

               151                                                200
predict_h1640  AVSLHLYSPP FDTCHAFDQR TGHKNKVTMT FHSKFGIRTP NATSGSLENN
cydx_human     AVSLHLYSPP FDTCHAFDQR TGHKNKVTMT FHSKFGIRTP NATSGSLENN
cydx_rat       AVSLHLYSPP FDTCHAFDQR TGHKNKVTMT FHSKFGIRTP FTTSGSLENN
```

```
Result of COILS prediction (Andrei Lupas):
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

A Lupas: Methods in Enzymology, 1996, 266, 513-525.
```

version 2.2: Rob B. Russell & Andrei N. Lupas, 1999

---

no coiled-coil above probability 0.5

```
 PHD: Profile fed neural network systems from HeiDelberg
 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

 Prediction of:
secondary structure,        by PHDsec
solvent accessibility,      by PHDacc
and helical transmembrane regions,     by PHDhtm

 Author:
Burkhard Rost
    EMBL, 69012 Heidelberg, Germany
    Internet: Rost@EMBL-Heidelberg.DE

 All rights reserved.


 The network systems are described in:

 PHDsec:    B Rost & C Sander: JMB, 1993, 232, 584-599.
  B Rost & C Sander: Proteins, 1994, 19, 55-72.
 PHDacc:   B Rost & C Sander: Proteins, 1994, 20, 216-226.
 PHDhtm:   B Rost et al.:     Prot. Science, 1995, 4, 521-533.


The resulting network (PHD) prediction is:
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
```

```
 PHD: Profile fed neural network systems from HeiDelberg
 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

 Prediction of:
secondary structure,        by PHDsec
solvent accessibility,      by PHDacc
and helical transmembrane regions,      by PHDhtm

 Author:
Burkhard Rost
    EMBL, 69012 Heidelberg, Germany
    Internet: Rost@EMBL-Heidelberg.DE

 All rights reserved.


 The network systems are described in:

 PHDsec:    B Rost & C Sander: JMB, 1993, 232, 584-599.
  B Rost & C Sander: Proteins, 1994, 19, 55-72.
 PHDacc:   B Rost & C Sander: Proteins, 1994, 20, 216-226.
 PHDhtm:   B Rost et al.:     Prot. Science, 1995, 4, 521-533.


 Some statistics
 ~~~~~~~~~~~~~~~

 Percentage of amino acids:
+--------------+--------+--------+--------+--------+--------+
| AA:          |   L    |   E    |   N    |   S    |   T    |
| % of AA:     |   8.5  |   8.0  |   7.0  |   6.5  |   6.0  |
+--------------+--------+--------+--------+--------+--------+
| AA:          |   K    |   H    |   G    |   A    |   V    |
| % of AA:     |   6.0  |   5.5  |   5.5  |   5.5  |   5.0  |
+--------------+--------+--------+--------+--------+--------+
| AA:          |   I    |   D    |   R    |   F    |   Q    |
| % of AA:     |   5.0  |   5.0  |   4.5  |   4.5  |   4.0  |
```

```
+--------------+--------+--------+--------+--------+--------+
| AA:          |   P    |   M    |   Y    |   C    |   W    |
| % of AA:     |   3.5  |   3.5  |   3.0  |   2.0  |   1.5  |
+--------------+--------+--------+--------+--------+--------+

Percentage of secondary structure predicted:
+--------------+--------+--------+--------+
| SecStr:      |   H    |   E    |   L    |
| % Predicted: |  40.5  |  16.5  |  43.0  |
+--------------+--------+--------+--------+

According to the following classes:
   all-alpha:   %H>45 and %E< 5; all-beta : %H<5 and %E>45
   alpha-beta : %H>30 and %E>20; mixed:    rest,
this means that the predicted class is:          mixed class



PHD output for your protein
~~~~~~~~~~~~~~~~~~~~~~~~~~~

Wed May 29 06:23:53 2002
Jury on:       10    different architectures (version   5.94_317 ).
Note: differently trained architectures, i.e., different versions can
result in different predictions.



About the protein
~~~~~~~~~~~~~~~~~

HEADER     /home/phd/server/work/predict_h16487.fas
COMPND
SOURCE
AUTHOR
SEQLENGTH   200
NCHAIN        1 chain(s) in predict_h16487 data set
NALIGN        1
(=number of aligned sequences in HSSP file)



WARNING
~~~~~~~
```

```
Expected accuracy is about 72% if, and only if, the alignment contain
sufficient information.  For your sequence there was no homologue in
the current version of Swissprot detected.  This implies that the
expected accuracy is about 6-10 percentage points lower !


Abbreviations: PHDsec
~~~~~~~~~~~~~~~~~~~~~

sequence:
   AA : amino acid sequence
secondary structure:
   HEL: H=helix, E=extended (sheet), blank=other (loop)
   PHD: Profile network prediction HeiDelberg
   Rel: Reliability index of prediction (0-9)
detail:
   prH: 'probability' for assigning helix
   prE: 'probability' for assigning strand
   prL: 'probability' for assigning loop
        note: the 'probabilites' are scaled to the interval 0-9, e.g.,
              prH=5 means, that the first output node is 0.5-0.6
subset:
   SUB: a subset of the prediction, for all residues with an expected
        average accuracy > 82% (tables in header)
        note: for this subset the following symbols are used:
     L: is loop (for which above " " is used)
   " ": means that no prediction is made for this residue, as the
        reliability is:  Rel < 5

Abbreviations: PHDacc
~~~~~~~~~~~~~~~~~~~~~

   SS : secondary structure
   HEL: H=helix, E=extended (sheet), blank=other (loop)
solvent accessibility:
   3st: relative solvent accessibility (acc) in 3 states:
        b = 0-9%, i = 9-36%, e = 36-100%.
   PHD: Profile network prediction HeiDelberg
   Rel: Reliability index of prediction (0-9)
   O_3: observed relative acc. in 3 states: B, I, E
        note: for convenience a blank is used intermediate (i).
   P_3: predicted relative accessibility in 3 states
```

```
   10st:relative accessibility in 10 states:
        = n corresponds to a relative acc. of n*n %
subset:
   SUB: a subset of the prediction, for all residues with an expected
        average correlation > 0.69 (tables in header)
        note: for this subset the following symbols are used:
   "I": is intermediate (for which above " " is used)
   ".": means that no prediction is made for this residue, as the
        reliability is: Rel < 4


protein:       predict        length      200


                 ....,....1....,....2....,....3....,....4....,....5....
        AA      |MEQTEVLKPRTLADLIRILHQLFAGDEVNVEEVQAIMEAYESDPIEWAMYAKFI
        PHD sec |          HHHHHHHHHHHHH        HHHHHHHHHHH   HHHHHHHHHHH
        Rel sec |9765434751499999999971688522799999999972677999999988
detail:
        prH sec |011121002569999999987411112279999999997421799999988
        prE sec |010112310000000000000000001331000000000000000000000000
        prL sec |977665676430000000000157886440000000000157820000000112
subset: SUB sec |LLLL...LL..HHHHHHHHHHH.LLLL..HHHHHHHHHHH.LLHHHHHHHHHHH
accessibility
3st:    P_3 acc |eeebebbeeebbbebbbbbbbbbeeebbbeebebbbeb eeeebebbbbbebe
10st:   PHD acc |986070067600070000000000777000760600070379670600000607
        Rel acc |330031432103317809740421013204319076536020202265606231
subset: SUB acc |......b.......bb.bbb.b.......b..b.bbb.b.......bbb.b...


                 ....,....7....,....8....,....9....,....10...,....11...
        AA      |NLVDQGNGKFNLMILCWGEGHGSSIHDHTNSHCFLKMLQGNLKETLFAWPDKKS
        PHD sec |HHH      EEEEEEEE            HHHHHHHHHHHHHHHHHHH     H
        Rel sec |763258872568999843789765124761589999998868886753177881
detail:
        prH sec |765321001000000000110011222125789989998878887765311005
        prE sec |011100013678998863100112321000000000000000001122200000
        prL sec |113568885210000126789866446774200000001121100112488884
subset: SUB sec |HH..LLLL.EEEEEEE..LLLLLL...LL.HHHHHHHHHHHHHHHHH..LLLL.
accessibility
3st:    P_3 acc |ebbeeeeeebbbbbbbbbee bbbb e beb bbbebbeeebeebbbb eeeee
10st:   PHD acc |600779776000000000795000048306050007007760770000579977
```

```
         Rel acc |14221310222869682341100240201020669332122242142001323(
subset:  SUB acc |.b.........bbbbb..e.....b.......bbb.......e..b........


                  ....,....13...,....14...,....15...,....16...,....17...
         AA      |SERVLRENQCAYINDSIGLHRVENISHTEPAVSLHLYSPPFDTCHAFDQRTGHH
         PHD sec |HHHHHHHHHHHHEEE      EEEEE         EEEEE            EE
         Rel sec |99998764441120334425641348999633774136886633231225679
detail:
         prH sec |99998876654433211000001000000110000000001122321111110(
         prE sec |00000000013444132237764521000135776432000122124532000(
         prL sec |00001123321111556652224368889653112467887755554346778
subset:  SUB sec |HHHHHHH..............EE...LLLLL..EE...LLLLL......LLLLL
accessibility
3st:     P_3 acc |beebbeebebbbbbebbbbb bebbeee ebbbbbb bee ebb bbee eeee
10st:    PHD acc |07600670600000700000406007775900000030673700500765976
         Rel acc |11215240053160024211061231101036351400010115062210021
subset:  SUB acc |....b.e..b..b...b....b.........b.b.b.......b.b.......e


                  ....,....19...,....20...,....21...,....22...,....23...
         AA      |FHSKFGIRTPNATSGSLENN|
         PHD sec |EE    EEE           |
         Rel sec |94356514357886787899|
detail:
         prH sec |00000000000011100000|
         prE sec |86321256621000000100|
         prL sec |03677742368887788899|
subset:  SUB sec |E..LLL...LLLLLLLLLLL|
accessibility
3st:     P_3 acc |bbbebbbebeebeeeb eee|
10st:    PHD acc |00060006067067704799|
         Rel acc |30422032112101100335|
subset:  SUB acc |..b................e|
```

```
The resulting prediction of globularity is:
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
```

```
---
--- GLOBE: prediction of protein globularity
---
--- nexp =    93    (number of predicted exposed residues)
--- nfit =    90    (number of expected exposed residues
--- diff =     3.00 (difference nexp-nfit)
--- =====> your protein appears as compact, as a globular domain
---
---
--- GLOBE: further explanations preliminaryily in:
---        http://www.columbia.edu/~rost/Papers/98globe.html
---
--- END of GLOBE
```

Result of ASP prediction(Malin Young, Kent Kirshenbaum, Stefan Highsmith
~~~~~~

Kirshenbaum K, Young M and Highsmith S.
Prot. Sci.(1999) 8:1806-1815.
Young M, Kirshenbaum K, Dill KA and Highsmith S.
Prot. Sci.(1999) 8:1752-1764.

Ambivalent Sequence Predictor (ASP v1.0) mmy

Parameters:
Window size : 5
Min mu dPr : 9
Z-score cutoff : -1.75

Mean dPr score=13.583, Standard deviation=2.903

```
                 ....,....1....,....2....,....3....,....4....,....5....
        AA      |MEQTEVLKPRTLADLIRILHQLFAGDEVNVEEVQAIMEAYESDPIEWAMYAKFI
        prH sec |011121002569999999998741111227999999999742179999999887
        prE sec |010112310000000000000000001331000000000000000000000000(
```
~~~~~~

```
prL sec |97766567643000000000015788644000000000015782000000011
ASP sec |..................................................

         ....,....7....,....8....,....9....,....10...,....11...
AA      |NLVDQGNGKFNLMILCWGEGHGSSIHDHTNSHCFLKMLQGNLKETLFAWPDKKS
prH sec |76532100100000000011001122212578998999887888776531100
prE sec |01110001367899886310011232100000000000000000112220000
prL sec |11356888521000012678986644677420000000112110011248888
ASP sec |.......................SS...........................

         ....,....13...,....14...,....15...,....16...,....17...
AA      |SERVLRENQCAYINDSIGLHRVENISHTEPAVSLHLYSPPFDTCHAFDQRTGHH
prH sec |99998876654433211000010000001100000000011223211111100
prE sec |00000000013444132237764521000135776432000122124532000
prL sec |00001123321111556652224368889653112467887755554346778
ASP sec |..........SSSSS............................SSSSS......

         ....,....19...,....20...,....21...,....22...,....23...
AA      |FHSKFGIRTPNATSGSLENN|
prH sec |00000000000011100000|
prE sec |86321256621000000100|
prL sec |03677742368887788899|
ASP sec |....................|
```

Please note: ASP was designed to identify the location of conformational switches in amino acid sequences. It is NOT designed to predict whether a given sequence does or does not contain a switch. For best results, ASP should be used on sequences of length >150 amino acids with >10 sequence homologues in the SWISS-PROT data bank.
ASP has been validated against a set of globular proteins and may not be generally applicable. Please see Young et al., Protein Science 8(9):1852-64. 1999. for details and for how best to interpret this output. We consider ASP to be experimental at this time, and would appreciate any feedback from our users.

```
-----------------------------------------------------------------------
- PredictProtein (PP): News 2000
-----------------------------------------------------------------------
-
- PP home:
  New York              http://cubic.bioc.columbia.edu/predictprotein/
-
- PP mirrors:

  Australia Sydney        http://molmod.angis.org.au/predictprotein/
  Germany EMBL            http://www.embl-heidelberg.de/predictprotein/
  China CBI,Peking        http://www.cbi.pku.edu.cn/predictprotein/
  China Inst. Microbiol.  http://micronet.im.ac.cn/predictprotein/
  England EBI             http://www.ebi.ac.uk/~rost/predictprotein/
  India CDFD   http://www.cdfd.org.in/~www/pp/predictprotein/
  India Pune              http://202.41.70.33/predictprotein/
  Iran Tehran             http://www.ibc.ut.ac.ir/predictprotein/
  Israel Beer-Sheva       http://www.cs.bgu.ac.il/~dfischer/predictprote
  Italy Rome              http://obelix.bio.uniroma2.it/www/predictprote
  Mexico Cuernavaca UNAM  http://www.ibt.unam.mx/paginas/lorenzo/predict
  Netherlands CMBI        http://www.cmbi.kun.nl/bioinf/predictprotein/
  Russia Puschino         http://mirror.protres.ru/predictprotein/
  Singapore               http://embl.bic.nus.edu.sg/predictprotein/
  Spain CNB               http://www.es.embnet.org/Services/MolBio/Predi
  Switzerland Glaxo       http://www.gwer.ch/tools/predictprotein/
  USA San Diego SDSC      http://www.sdsc.edu/predictprotein/

-
- Tools to post-process PP results:
-
- Generate a PostScript (or GIF, or TIFF):
  ESPript (New York)  http://cubic.bioc.columbia.edu/cgi/pp/nph-ESPript_
  ESPript (Toulouse)  http://www-pgm1.ipbs.fr:8080/ESPript
-
-----------------------------------------------------------------------
```

```
\feature{top}{1}{12..22}{box:$\alpha $1}{}
\feature{top}{1}{30..40}{box:$\alpha $2}{}
\feature{top}{1}{44..62}{box:$\alpha $3}{}
\feature{top}{1}{70..76}{-->}{$\beta $1}
\feature{top}{1}{91..107}{box:$\alpha $4}{}
\feature{top}{1}{116..127}{box:$\alpha $5}{}
\feature{top}{1}{140..141}{-->}{$\beta $2}
\feature{top}{1}{153..154}{-->}{$\beta $3}
\feature{top}{1}{177..181}{-->}{$\beta $4}
```

Figure A.7: Predicted secondary features by TEXshade based upon PhD sequence analysis of CDO in preceding data.

```
!!AA_MULTIPLE_ALIGNMENT 1.0
PileUp of: @/home/pugh/anon/.seqlab-pugh/pileup_244.list

 Symbol comparison table: GenRunData:blosum62.cmp  CompCheck: 1102

                    GapWeight: 8
               GapLengthWeight: 2

 pileup_244.msf  MSF: 221  Type: P  May 28, 2002 16:53  Check: 6608 ..

 Name: cydx_human       Len:   221  Check: 9529  Weight:  1.00
 Name: cydx_rat         Len:   221  Check: 9956  Weight:  1.00
 Name: cydx_gmir        Len:   221  Check: 6058  Weight:  1.00
 Name: cydx_cele        Len:   221  Check: 5305  Weight:  1.00
 Name: cydx_sjap        Len:   221  Check: 5760  Weight:  1.00

//

             1                                                  50
cydx_human   ~~~~~~~~~M EQTEVLKPRT LADLIRILHQ LFAGDEVNVE EVQAIMEAYE
  cydx_rat   ~~~~~~~~~M ERTELLKPRT LADLIRILHE LFAGDEVNVE EVQAVLEAYE
 cydx_gmir   ~~~~~~~~~M EHTEVVVPKT LDELIAILHQ IFESDNINVE EVQQIMESYE
 cydx_cele   ~~~~~~~~~~ ~~~~~~~MIS FVQLVVQIRE IFQQKLIDVD EVMKLMASYK
 cydx_sjap   MSMYTHQSNN ELIPLKTVST LNDLIKTIRI IFNQKEINVN EIHKILNDFQ

             51                                                 100
cydx_human   SDPIEWAMYA KFDQYRYTRN LVDQGNGKFN LMILCWGEGH GSSIHDHTNS
  cydx_rat   SNPAEWALYA KFDQYRYTRN LVDQGNGKFN LMILCWGEGH GSSIHDHTDS
 cydx_gmir   SNPQDWNKFA KFDQYRYTRN LVDEGNGKFN LMILCWGEGH GSSIHDHTDS
 cydx_cele   SNANEWRRFA IFDMNKYTRN LVDVGNGKYN LMILCWGPGM ASSVHDHTDA
 cydx_sjap   CDFTEWQKYI YFNKTHYTRN LIDEGNGRYN LFLLCWSEDQ GTRIHDHSGA
```

```
            101                                                150
cydx_human  HCFLKMLQGN LKETLFAWPD ....KKSN.. ......EMVK KSERVLRENQ
  cydx_rat  HCFLKLLQGN LKETLFDWPD ....KKSN.. ......EMIK KSERTLRENQ
 cydx_gmir  HCFMKLLQGQ LKETLFEWPE ....GEQN.. .....GEMVQ KSRDPDGKQG
 cydx_cele  HCFVKILDGE LTETKYAWP. ....RKRH.. .....VPLDI SENKTYGMNG
 cydx_sjap  HCFVKLIKGC IKETIFEWPK YFTVEKSNYS INQIDLPLTV KSVSEMRPGD

            151                                                200
cydx_human  CAYINDSIGL HRVENISHTE PAVSLHLYSP PFDTCHAFDQ RTGHKNKVTM
  cydx_rat  CAYINDSIGL HRVENVSHTE PAVSLHLYSP PFDTCHAFDQ RTGHKNKVTM
 cydx_gmir  CLHK------ ---------- ---------- ---------- ----------
 cydx_cele  VSYMNDELGL HRMENLSHSN GAVSLHLYIP PYSTCNAFDE RTGKKTQCTV
 cydx_sjap  VTYMHDKIGI HRLHNPSTTE TAITLHLYFP PYTNSMIFEE STSRMKKMDV

            201                 221
cydx_human  TFHSKFGIRT PNATSGSLEN N
  cydx_rat  TFHSKFGIRT PFTTSGSLEN N
 cydx_gmir  ---------- ---------- -
 cydx_cele  TFYSKYGKKV DYRGSKNGN- -
 cydx_sjap  TFHSKFGKQI PQ-------- -
```

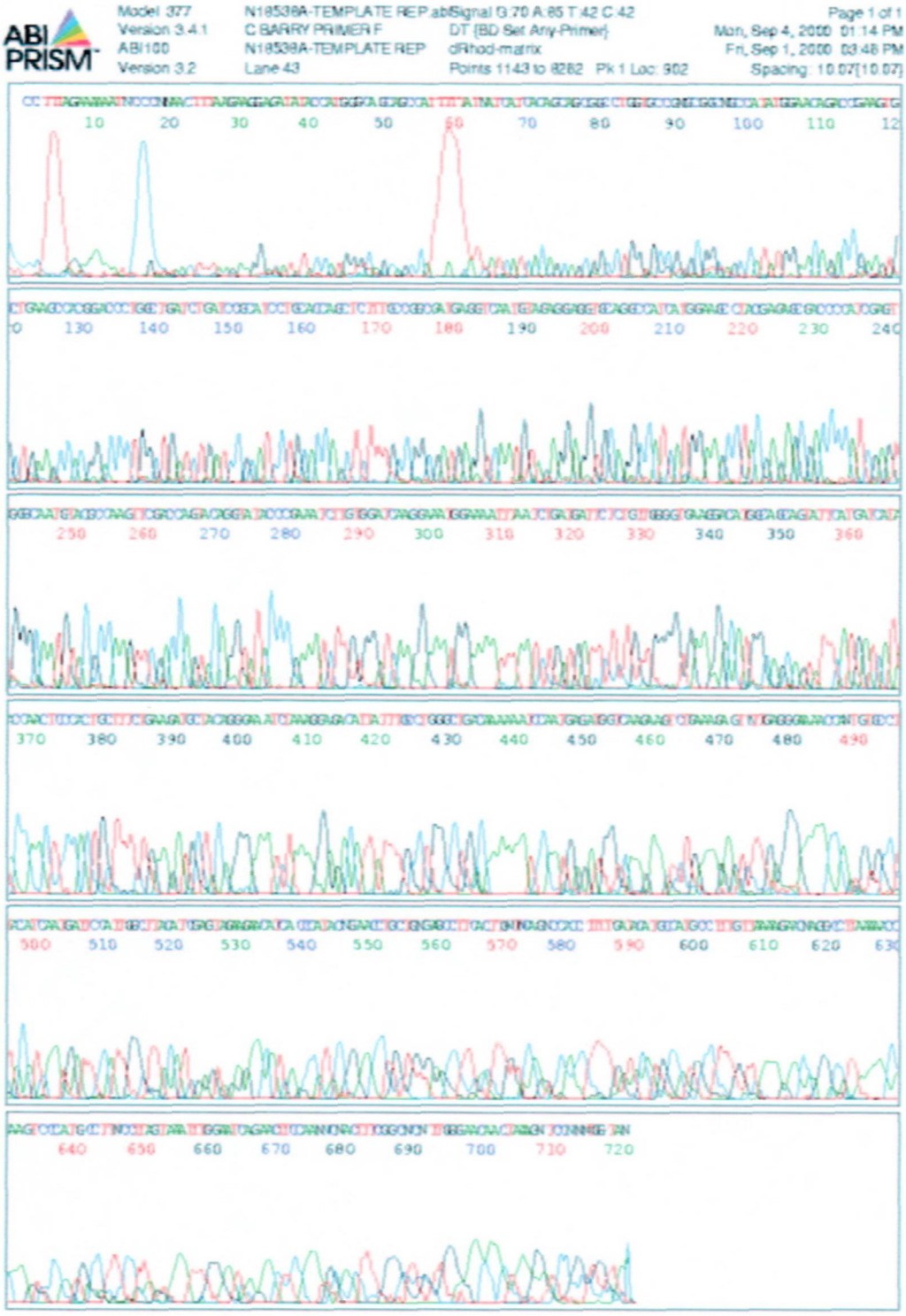


Figure A.8: Forward sequence of pET15b-hCDO construct

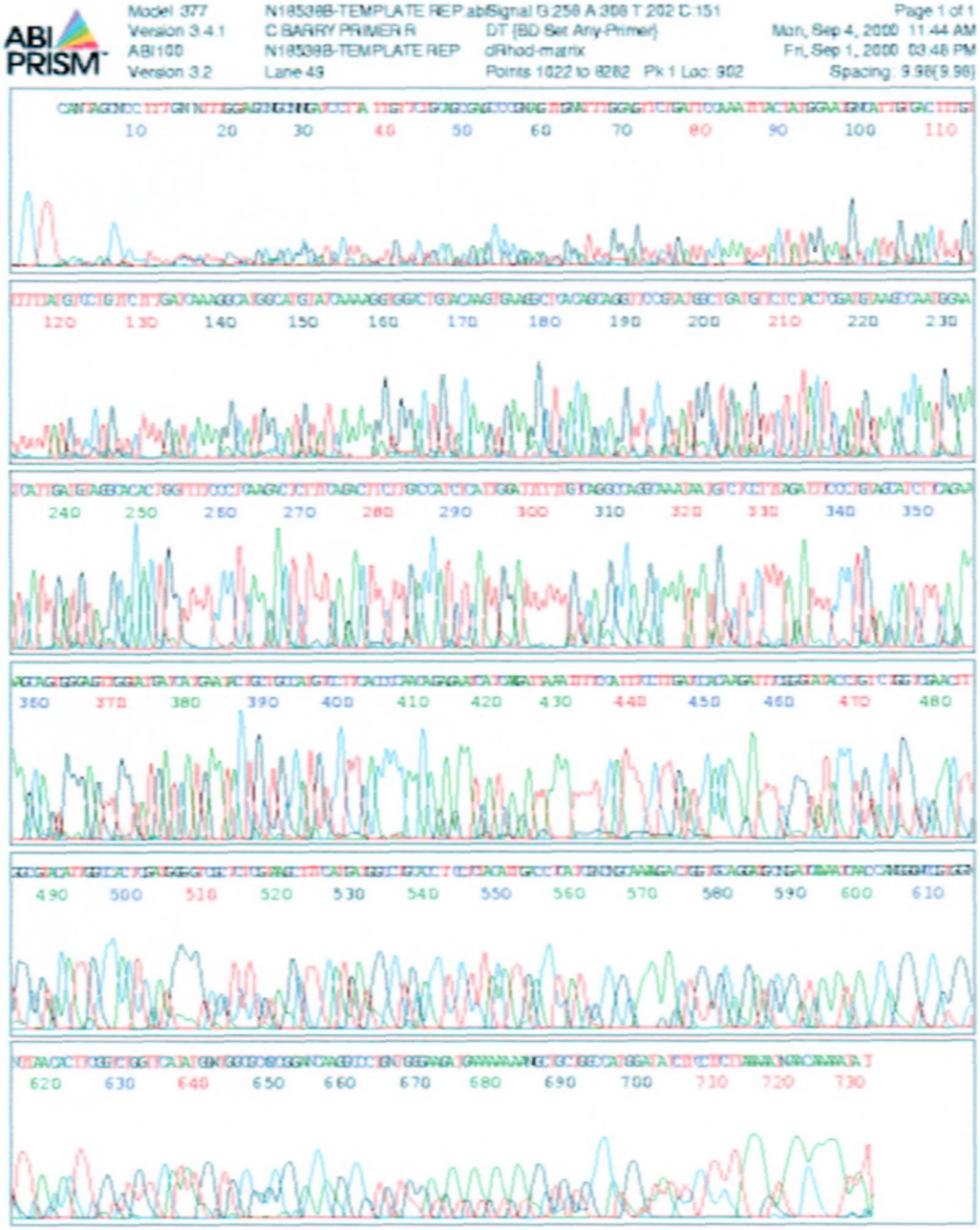


Figure A.9: Reverse sequence of pET15b-hCDO construct

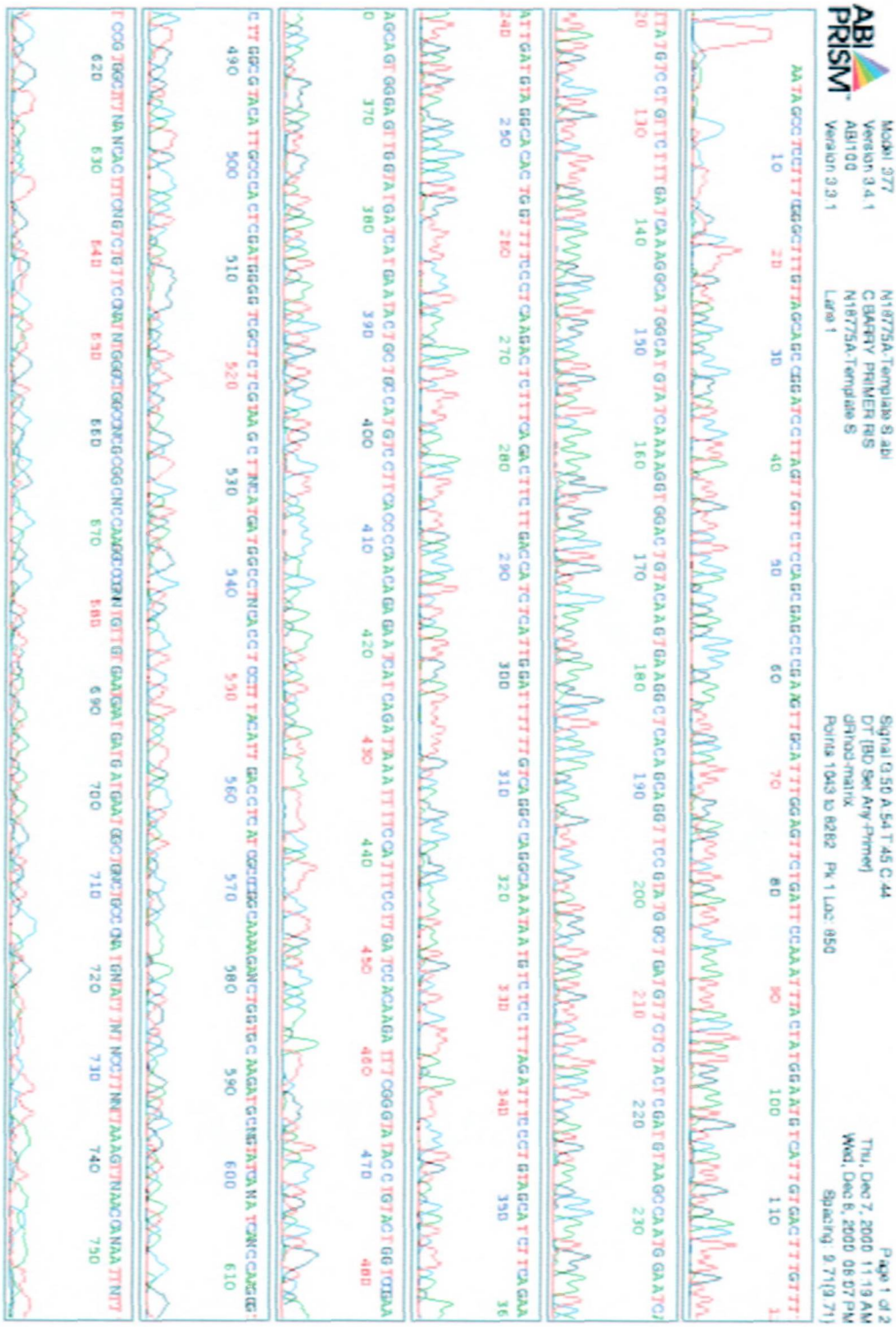


Figure A.10: sequence of pET15b-hCDO construct

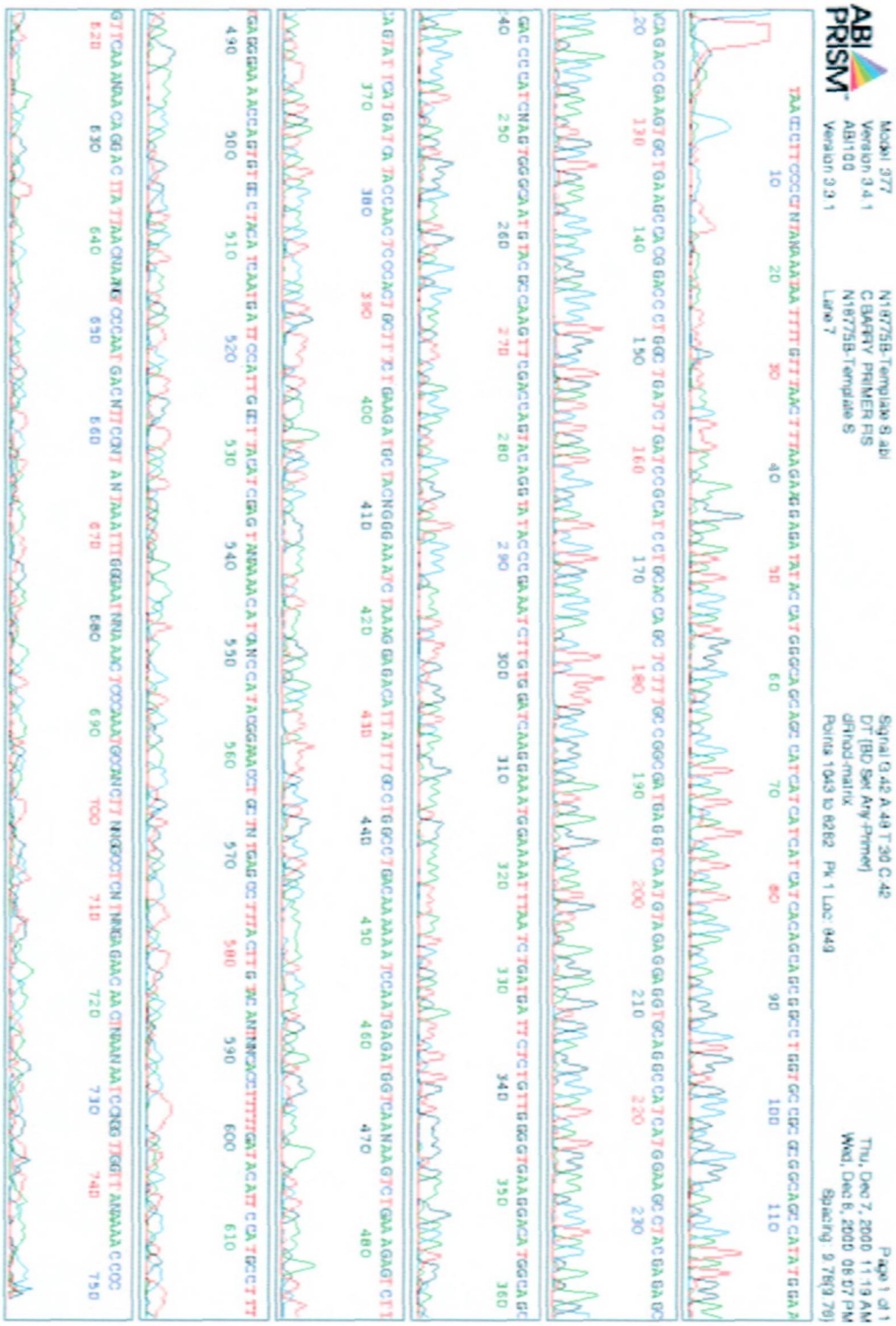


Figure A.11: sequence of pET15b-hCDO construct

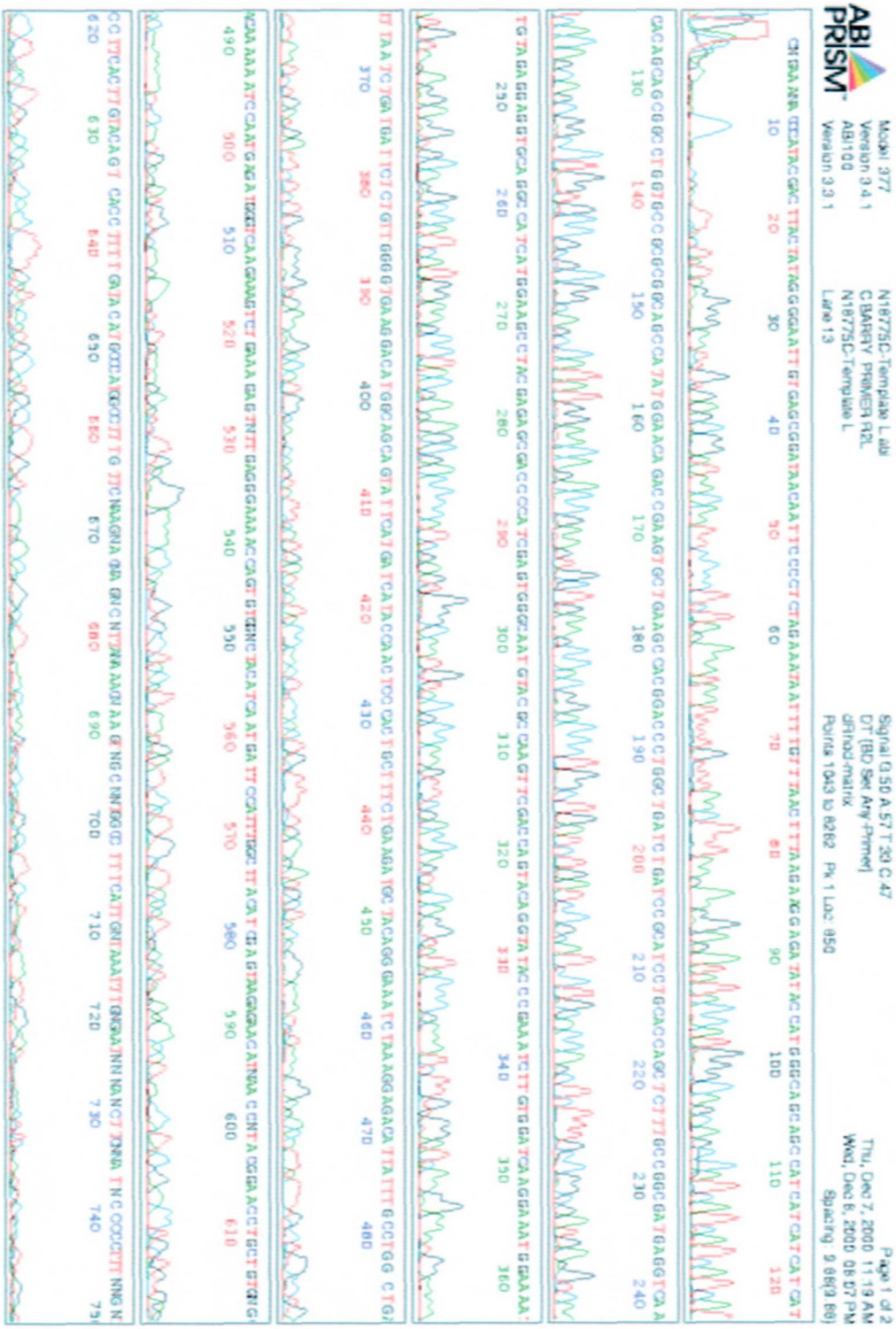


Figure A.12: sequence of pET15b-hCDO construct

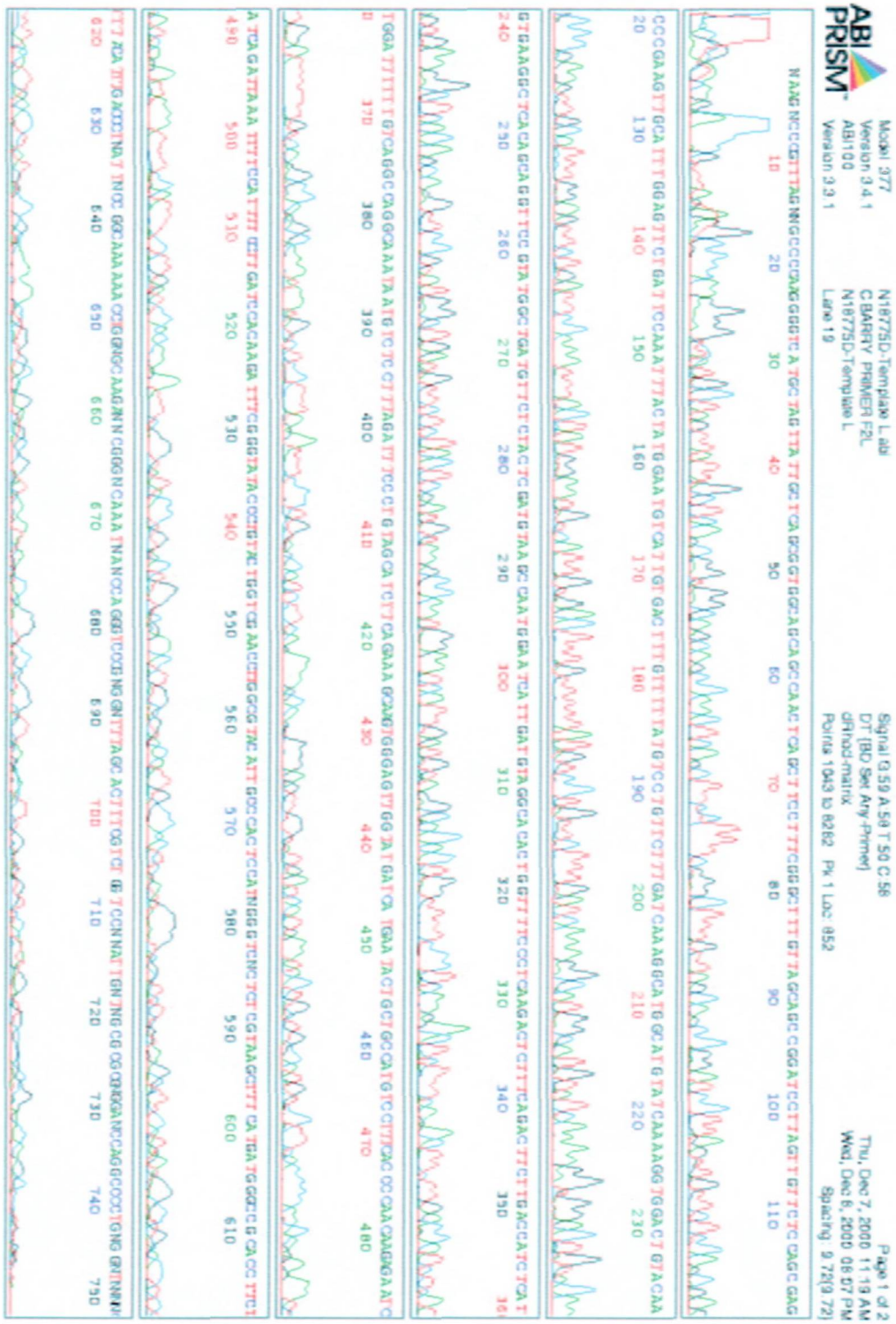


Figure A.13: sequence of pET15b-hCDO construct

```
                PRETTYBOX of: pileup_53.msf{*}  December 8, 2000 11:40:32.74

       hcdo_cdna   ~~ATGGAACA GACCGAAGTG CTGAAGCCAC GGACC  33
  forward_p_to_t1  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~  0
  reverse_p_to_t1  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ GGACC  5
  forward_p+to_t+  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~  0
reverse_p+_to_t+   ATATGGAACA GACCGAAGTG CTGAAGCCAC GGACC  35
  forward_p_to_t2  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~  0
  reverse_p_to_t2  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~  0
       Consensus   ---------- ---------- ---------- GGACC  35

       hcdo_cdna   CTGGCTGATC TGATCCGCAT CCTGCACCAG CTC..  66
  forward_p_to_t1  ~GGTTGATTT TGATCNGCAT CCTGCACCAG T..CT  32
  reverse_p_to_t1  CTGGCTGATC TGATCCGCAT CCTGCACCAG CTC..  38
  forward_p+to_t+  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~  0
reverse_p+_to_t+   CTGGCTGATC TGATCCGCAT CCTGCACCAG CTC..  68
  forward_p_to_t2  ~~GGCTGATC TGATCCGCAT CCTGCACCAG CTC..  31
  reverse_p_to_t2  ~~~~~~~TNT GATACNGCAT CTTGCACCAG NTCTT  28
       Consensus   CTGGCTGATC TGATCCGCAT CCTGCACCAG CTC--  70

       hcdo_cdna   TTTGCCGGCG ATGAGGTCAA TGTAGAGGAG GTGCA  101
  forward_p_to_t1  TTTGCNGNCG ATGAGGTCAA TGTAGAGGAG GTGCA  67
  reverse_p_to_t1  TTTGCCGGCG ATGAGGTCAA TGTAGAGGAG GTGCA  73
  forward_p+to_t+  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~  0
reverse_p+_to_t+   TTTGCCGGCG ATGAGGTCAA TGTAGAGGAG GTGCA  103
  forward_p_to_t2  TTTGCCGGCG ATGAGGTCAA TGTAGAGGAG GTGCA  66
  reverse_p_to_t2  TTGCCGGGCG ATGAGGTCAA TGTAAAGGAG GTGNA  63
       Consensus   TTTGCCGGCG ATGAGGTCAA TGTAGAGGAG GTGCA  105

       hcdo_cdna   GGCCATCATG GAAGCCTACG AGAGCGACCC CATCG  136
  forward_p_to_t1  GGCCATCATG AAAGCTTACG AGAGCGACCC CATCG  102
  reverse_p_to_t1  GGCCATCATG GAAGCCTACG AGAGCGACCC CATCG  108
  forward_p+to_t+  ~~~~~TCATG AAAGCTTACG AGAGNGACCC NATGG  30
reverse_p+_to_t+   GGCCATCATG GAAGCCTACG AGAGCGACCC CATCG  138
  forward_p_to_t2  GGCCATCATG GAAGCCTACG AGAGCGACCC CATCN  101
  reverse_p_to_t2  GGCCATCATG NAAGCTTACG AGAGCGACCC CATCG  98
       Consensus   GGCCATCATG GAAGCCTACG AGAGCGACCC CATCG  140

       hcdo_cdna   AGTGGGCAAT GTACGCCAAG TTCG.ACCAG TACA.  169
  forward_p_to_t1  AGTGGCCAAT GTACGCCAAG TTCG..ACCA GACA.  134
  reverse_p_to_t1  AGTGGGCAAT GTACGCCAAG TTCG.ACCAG TACA.  141
  forward_p+to_t+  AGTGGGCAAT GTACGCCAAG GTTCGACCAG TACAG  65
reverse_p+_to_t+   AGTGGGCAAT GTACGCCAAG TTCG.ACCAG TACA.  171
  forward_p_to_t2  AGTGGGCAAT GTACGCCAAG TTCG.ACCAG TACA.  134
  reverse_p_to_t2  AGTGGGCAAT GTACGCCAAG TTCGGACCAG TACA.  132
       Consensus   AGTGGGCAAT GTACGCCAAG TTCG-ACCAG TACA-  175

       hcdo_cdna   GGTATACCCG AAATCTTGTG GATCAAGGAA ATGGA  204
  forward_p_to_t1  GGTATACCCG AAATCTTGTG GATCAAGGAA ATGGA  169
  reverse_p_to_t1  GGTATACCCG AAATCTTGTG GATCAAGGAA ATGGA  176
  forward_p+to_t+  GGTATACCCG AAATCTTGTG GATCAAGGAA AATGG  100
reverse_p+_to_t+   GGTATACCCG AAATCTTGTG GATCAAGGAA ATGGA  206
  forward_p_to_t2  GGTATACCCG AAATCTTGTG GATCAAGGAA ATGGA  169
  reverse_p_to_t2  GGTATACCCG AAATCTTGTG GATCAAGGAA ATGGA  167
       Consensus   GGTATACCCG AAATCTTGTG GATCAAGGAA ATGGA  210

       hcdo_cdna   AAATTTAATC .TGATGATTC TCT..GTTGG GGTGA  236
  forward_p_to_t1  AAATTTAATC TTGATGATTC TCT..GTTGG GGTGA  202
  reverse_p_to_t1  AAATTTAATC .TGATGATTC TCT..GTTGG GGTGA  208
  forward_p+to_t+  AAAATTTAAT CTGATGATTC TCTTGTTGGG GTGAA  135
reverse_p+_to_t+   AAATTTAATC .TGATGATTC TCT..GTTGG GGTGA  238
  forward_p_to_t2  AAATTTAATC .TGATGATTC TCT..GTTGG GGTGA  201
  reverse_p_to_t2  AAATTTAATC .TGATGATTC TCT..GTTGG GGTGA  199
       Consensus   AAATTTAATC -TGATGATTC TCT--GTTGG GGTGA  245
```

Figure A.14: Alignments of construct sequencing with consensus sequence (1/3)

```
PRETTYBOX of: pileup_53.msf{*}  December 8, 2000 11:40:32.74

       hcdo_cdna  AGGACATGGC AGCAGTATTC ATGATCATAC CAACT 271
  forward_p_to_t1 AGGACATGGC AGCAGTATTC ATGATCATAC CAACT 237
  reverse_p_to_t1 AGGACATGGC AGCAGTATTC ATGATCATAC CAACT 243
  forward_p+to_t+ AGGACATGGC AGCAGTATTC ATGATCATAC CAACT 170
reverse_p+_to_t+  AGGACATGGC AGCAGTATTC ATGATCATAC CAACT 273
  forward_p_to_t2 AGGACATGGC AGCAGTATTC ATGATCATAC CAACT 236
  reverse_p_to_t2 AGGACATGGC AGCAGTATTC ATGATCATAC CAACT 234
        Consensus AGGACATGGC AGCAGTATTC ATGATCATAC CAACT 280

       hcdo_cdna  CCCAC.TGCT TTCTGAAGAT GCTACAGGGA AATCT 305
  forward_p_to_t1 CCCAC.TGCT TTCTGAAGAT GCTACAGGGA AATCT 271
  reverse_p_to_t1 CCCAC.TGCT TTCTGAAGAT GCTACAGGGA AATCT 277
  forward_p+to_t+ CCCACTTGCT TTCTGAAGAT GCTACAGGGA AATCT 205
reverse_p+_to_t+  CCCAC.TGCT TTCTGAAGAT GCTACAGGGA AATCT 307
  forward_p_to_t2 CCCAC.TGCT TTCTGAAGAT GCTACNGGGA AATCT 270
  reverse_p_to_t2 CCCAC.TGCT TTCTGAAGAT GCTACAGGGA AATCT 268
        Consensus CCCAC-TGCT TTCTGAAGAT GCTACAGGGA AATCT 315

       hcdo_cdna  AAAGGAGACA TTATTTGCCT GGCCTGACAA AAAAT 340
  forward_p_to_t1 AAAGGAGACA TTATTTGCCT GGCCTGACAA AAAAT 306
  reverse_p_to_t1 AAAGGAGACA TTATTTGCCT GGGCTGACAA AAAAT 312
  forward_p+to_t+ AAAGGAGACA TTATTTGCCT GGCCTGACAA AAAAT 240
reverse_p+_to_t+  AAAGGAGACA TTATTTGCCT GG.CTGACAA AAAAT 341
  forward_p_to_t2 AAAGGAGACA TTATTTGCCT GGCCTGACAA AAAAT 305
  reverse_p_to_t2 AAAGGAGACA TTATTTGCCT GGCCTGACAA AAAAT 303
        Consensus AAAGGAGACA TTATTTGCCT GGCCTGACAA AAAAT 350

       hcdo_cdna  CCAATGAGAT GG..TCAAGA AGTCTGAAAG AGTCT 373
  forward_p_to_t1 CCAATGAGAT GG..TCAAGA AGTCTGAAAG AGTCT 339
  reverse_p_to_t1 CCAATGAGAT GG..TCAAGA AGTCTGAAAG AGTNT 345
  forward_p+to_t+ CCAATGAGAT GG..TCAAGA AGTCTGAAAG AGTCT 273
reverse_p+_to_t+  CCAATGAGAT GGGTCAAGAA AGTCTGAAAG AGTNT 376
  forward_p_to_t2 CCAATGAGAT GG..TCAANA AGTCTGAAAG AGTCT 338
  reverse_p_to_t2 CCAATGAGAT GG..TCAAGA AGTCTGAAAG AGTCT 336
        Consensus CCAATGAGAT GG--TCAAGA AGTCTGAAAG AGTCT 385

       hcdo_cdna  TGAGGGAAAA CCAGTGT.GC CTACATCAAT GATTC 407
  forward_p_to_t1 TGAGGGAAAA CCAGTGT.GC CTACATCAAT GATTC 373
  reverse_p_to_t1 TGAGGGAAAA CCANTGT.GC CTACATCAAT GATTC 379
  forward_p+to_t+ TGAGGGAAAA CCAGTGT.GC CTACATCAAT GATTC 307
reverse_p+_to_t+  TGAGGGAAAA CCAGTGTGGN CTACATCAAT GATTC 411
  forward_p_to_t2 TGAGGGAAAA CCAGTGT.GC CTACATCAAT GATTC 372
  reverse_p_to_t2 TGAGGGAAAA CCAGTGT.GC CTACATCAAT GATTC 370
        Consensus TGAGGGAAAA CCAGTGT-GC CTACATCAAT GATTC 420

       hcdo_cdna  CA.TTGGCTT ACATCGAGTA G.AGAACATC AGCCA 440
  forward_p_to_t1 CA.TTGGCTT ACATCGAGTA G.AGAACATC AGCCA 406
  reverse_p_to_t1 CA.TTGGCTT ACATCGAGTA GAAGAACATC ACCCA 413
  forward_p+to_t+ CA.TTGGCTT ACATCGAGTA G.AGAACATC AGCCA 340
reverse_p+_to_t+  CATTTGGCTT ACATCGAGTA AGAGAACATN AACCN 446
  forward_p_to_t2 CA.TTGGCTT ACATCGAGTA NAAAACATCA NCCAT 406
  reverse_p_to_t2 CA.TTGGCTT ACATCGAGTA G.AGAACATC AGCCA 403
        Consensus CA-TTGGCTT ACATCGAGTA G-AGAACATC AGCCA 455

       hcdo_cdna  TACGGAACCT GCTGTGAGCC TTCACTT.GT ACAGT 474
  forward_p_to_t1 TACGGAACCT GCTGTGAGCC TTCACTT.GT ACAGT 440
  reverse_p_to_t1 TACNGAACCT GCTGNGAGCC TTCACTTGNT NCAGN 448
  forward_p+to_t+ TACGGAACCT GCTGTGAGCC TTCACTT.GT ACAGT 374
reverse_p+_to_t+  TACGGAACCT GCTGTGNGCC TTCACTT.GT ACAGT 480
  forward_p_to_t2 ACGGAAACCT GCTNTGAGCC TTTACTTGTA CANTN 441
  reverse_p_to_t2 TACGGAACCT GCTGTGAGCC TTCACTT.GT ACAGT 437
        Consensus TACGGAACCT GCTGTGAGCC TTCACTT-GT ACAGT 490
```

Figure A.15: Alignments of construct sequencing with consensus sequence (2/3)

```
PRETTYBOX of: pileup_53.msf{*}  December 8, 2000 11:40:32.74

       hcdo_cdna  CCACC.TTTT GATACATGCC ATGCCTTTGA TCAAA 508
  forward_p_to_t1 CCACC.TTTT GATACATGCC ATGCCTTTGA TCAAA 474
  reverse_p_to_t1 CCACC.TTTT GATACATGCC ATGCCTTTGT TAAAA 482
  forward_p+to_t+ CCACC.TTTT GATACATGCC ATGCCTTTGA TCAAA 408
reverse_p+_to_t+  CACCTTTTGA TACATGCCCA TGGCCTTTGT TCNAA 515
  forward_p_to_t2 NCACCTTTTT GATACATTCC ATGCCTTTGT TCAAA 476
  reverse_p_to_t2 CCACC.TTTT GATACATGCC ATGCCTTTGA TCAAA 471
        Consensus CCACC-TTTT GATACATGCC ATGCCTTTGA TCAAA 525

       hcdo_cdna  GAACAGGACA TAAAAACAAA G.TCACAATG ACATT 542
  forward_p_to_t1 GAACAGGACA TAAAAACAAA G.TCACAATG NCATT 508
  reverse_p_to_t1 GAACNAGGCC TTAAAAACCA A.GTCCCATG CCTTN 516
  forward_p+to_t+ GAACAGGACA TAAAAACAAA G.TCACAATG ACATT 442
reverse_p+_to_t+  GNACNAGNCN TTANAACNA AGTNGCNNTG GCCTT 550
  forward_p_to_t2 NAACAGGACT TATTAACNAA NGTCCCAATG ACNTT 511
  reverse_p_to_t2 GAACAGGACA TAAAAACAAA G.TCACAATG ACATT 505
        Consensus GAACAGGACA TAAAAACAAA G-TCACAATG ACATT 560

       hcdo_cdna  CCATAGTAAA TTTGG..AAT CAGAACTCCA AATGC 575
  forward_p_to_t1 CCATAGTAAA TTTGG..AAT CAGAACTCCA AATNC 541
  reverse_p_to_t1 CCTTAGTAAA TTTGG..AAT CAGAACTTCC AANNC 549
  forward_p+to_t+ CCATAGTAAA TTTGG..AAT CAGAACTCCA AATGC 475
reverse_p+_to_t+  TCATTGNTAA ATTTGNGAAT NNNANCTTCN NATNC 585
  forward_p_to_t2 CCNTANTAAA TTTGGGAATN NAAACTCCCA AATGC 546
  reverse_p_to_t2 CCATAGTAAA TTTGG..AAT CAGAACTCCA AATGC 538
        Consensus CCATAGTAAA TTTGG--AAT CAGAACTCCA AATGC 595

       hcdo_cdna  AACTTCGGGC TCGCTGGAGA ACAACTAA~~ ~~~~~ 603
  forward_p_to_t1 AACTNCGGGC TCGCTGCAGA ACAA.TAAGG ATCNN 575
  reverse_p_to_t1 NACTTTCGGC NCNTTGGGAA CAACTAAAGN TCCNN 584
  forward_p+to_t+ AACTTCGGGC TCGCTGGAGA ACAACTAAGG ATCCG 510
reverse_p+_to_t+  CCCCTTTNNG NTCGNTCG~~ ~~~~~~~~~~ ~~~~~ 603
  forward_p_to_t2 CANCTTNNGG CCTCNTNNGA GAACAACTNA ANAAT 581
  reverse_p_to_t2 AACTTCGGGC TCGCTGGAGA ACAACTAAGG ATCCG 573
        Consensus AACTTCGGGC TCGCTGGAGA ACAACTAAGG ATC-- 630

       hcdo_cdna  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~ 603
  forward_p_to_t1 GCNGCTACCA AANNCAAAGG NGCTANTG~~ ~~~~~ 603
  reverse_p_to_t1 NTGGTAN~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~ 591
  forward_p+to_t+ GCTGCTAACA AAGCCCGAAA GGAAGCTGAG TTGGC 545
reverse_p+_to_t+  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~ 603
  forward_p_to_t2 CCNGGTTGGT TANAAAACCC C~~~~~~~~~ ~~~~~ 602
  reverse_p_to_t2 GCTGCTAACA AAGCCCGAAA GGAGGCTATT ~~~~~ 603
        Consensus GC-GCTA-CA AA-CC--A-- GG---CT--- ----- 665

       hcdo_cdna  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~ 603
  forward_p_to_t1 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~ 603
  reverse_p_to_t1 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~ 591
  forward_p+to_t+ TGCTGCCACC GCTGAGCAAT AACTAGCATG ACCCC 580
reverse_p+_to_t+  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~ 603
  forward_p_to_t2 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~ 602
  reverse_p_to_t2 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~ 603
        Consensus ---------- ---------- ---------- ----- 700

       hcdo_cdna  ~~~~~~~~~~ ~~~~~~~~~~ ~~~603
  forward_p_to_t1 ~~~~~~~~~~ ~~~~~~~~~~ ~~~603
  reverse_p_to_t1 ~~~~~~~~~~ ~~~~~~~~~~ ~~~591
  forward_p+to_t+ TTGGGGCNNC TAAACGGGNC TTN603
reverse_p+_to_t+  ~~~~~~~~~~ ~~~~~~~~~~ ~~~603
  forward_p_to_t2 ~~~~~~~~~~ ~~~~~~~~~~ ~~~602
  reverse_p_to_t2 ~~~~~~~~~~ ~~~~~~~~~~ ~~~603
        Consensus ---------- ---------- ---723
```

Figure A.16: Alignments of construct sequencing with consensus sequence (3/3)

Amino Acid	No.	No. Nitrogen	Total N
Ala	11	1	11
Asx	0	1	0
Cys	4	1	4
Asp	10	1	10
Glu	16	1	16
Phe	9	1	9
Gly	14	1	14
His	18	3 (2x1°, 1x3°)	54
Ile	10	1	10
Lys	12	2 (1x1°, 1x2°)	24
Leu	18	1	18
Met	8	1	8
Asn	14	2 (1x1°, 1x2°)	28
Pro	8	1	8
Gln	8	2 (1x1°, 1x2°)	16
Arg	10	3 (1x1°, 2x2°)	30
Ser	18	1	18
Thr	12	1	12
Val	11	1	11
Trp	3	2 (2x2°)	6
Tyr	6	1	6
Glx	0	1	0
Total	220		313

Table A.1: Breakdown of nitrogen content of recombinant cysteine dioxygenase

SIGMA
3050 Spruce Street
Saint Louis, Missouri 63103 USA
Telephone 800-325-5832 • (314) 771-5765
Fax (314) 286-7828
email: techserv@sial.com
sigma-aldrich.com

ProductInformation

Color Markers
High Range, (M.W. 29,000-205,000), Product Code **C 3312**
Wide Range, (M.W. 6,500-205,000), Product Code **C 3437**
Low Range, (M.W. 6,500-45,000), Product Code **C 3187**

Product Description
Color Markers with high, low, and wide molecular weight ranges are specially designed for use as standards in the Laemmli SDS-PAGE system, the PhastSystem® electrophoresis work station, and in protein transfer to solid phase supports such as nitrocellulose, nylon, and PVDF membranes. Color Markers consist of mixtures of 5-8 proteins with each protein conjugated to a different dye. These proteins provide a visual monitor of protein migration during electrophoresis and/or protein transfer efficiency to a membrane.

Since the molecular weights of the proteins are altered by the attachment of dye, apparent molecular weights of the protein-dye conjugates are standardized using High, Low, and Wide Range SigmaMarkers (Product Codes M 3788, M 3913, and M 4038) on a 4-20% gradient gel in the Laemmli system. Molecular weights determined in this manner are available on lot specific Certificates of Analysis.

Each vial of Color Markers contains 500 µl of a mixture of the protein-dye conjugates in 62 mM Tris-HCl, pH 7.5, containing 2% SDS, 0.1 mM EDTA, 100 mM dithiothreitol, 4 M urea, 0.005% bromophenol blue, and 30% glycerol. Distribution of the various proteins is indicated in Table 1.

Notes:
1. When precise molecular weight determinations are required, use Sigma Biotinylated Standards (Product Code SDS-6B) or High, Low, and Wide Range SigmaMarkers (Product Codes M 3788, M 3913, and M 4038, respectively).
2. The apparent molecular weight of the myosin-dye conjugate cannot be accurately determined since its molecular weight lies outside the range of the standard curve used for molecular weight determination.
3. Color Markers may be stained with Coomassie® Blue, but equal intensity bands will not be obtained.

Disclaimer
This product is for laboratory research use only. Please consult the Material Safety Data Sheet for information regarding hazards and safe handling practices.

Storage/Stability
Aliquot vial contents and store at –70 °C or below. Repeated freezing and thawing is not recommended. By eliminating multiple freeze/thaw cycles, the effective lifetime of these products will be prolonged.

Preparation Instructions
Do not heat the Color Markers. Simply thaw to room temperature and load the gel. For use in a mini gel (10 cm x 10 cm), load 10 µl for visualization directly on the gel or 5 µl when transferring to a membrane. For use in a standard size gel (16 cm x 18 cm), load 20 µl for visualization directly on the gel or 10 µl when transferring to a membrane. For use in the PhastGel system, load 1-4 µl.

Table 1.
Distribution of Proteins in Color Markers

Proteins	Native Mol. Wt. of Subunit (kDa)	Color of Conjugate	High Range* (C 3312)	Wide Range (C 3437)	Low Range (C 3187)
Myosin, rabbit muscle	205	Blue	X	X	
β-Galactosidase, *E. coli*	116	Turquoise	X	X	
Albumin, bovine serum	66	Pink	X	X	
Ovalbumin, chicken egg	45	Yellow	X	X	X
Carbonic Anhydrase, bovine erythrocytes	29	Orange	X	X	X
Trypsin Inhibitor, soybean	20	Green		X	X
α-Lactalbumin, bovine milk	14.2	Purple		X	X
Aprotinin, bovine milk	6.5	Blue		X	X

*The light chain of myosin may appear as an additional blue band having a molecular weight approximately that of Trypsin Inhibitor from Soybean.

Reference
Laemmli, U.K., Nature, **227**, 680 (1970).

PhastGel is a registered trademark of Amersham Pharmacia Biotech Limited or its subsidiaries.

Coomassie is a registered trademark of Imperial Chemical Industries PLC.

MDS/MAM 6/03

Sigma brand products are sold through Sigma-Aldrich, Inc.
Sigma-Aldrich, Inc. warrants that its products conform to the information contained in this and other Sigma-Aldrich publications. Purchaser must determine the suitability of the product(s) for their particular use. Additional terms and conditions may apply. Please see reverse side of the invoice or packing slip.

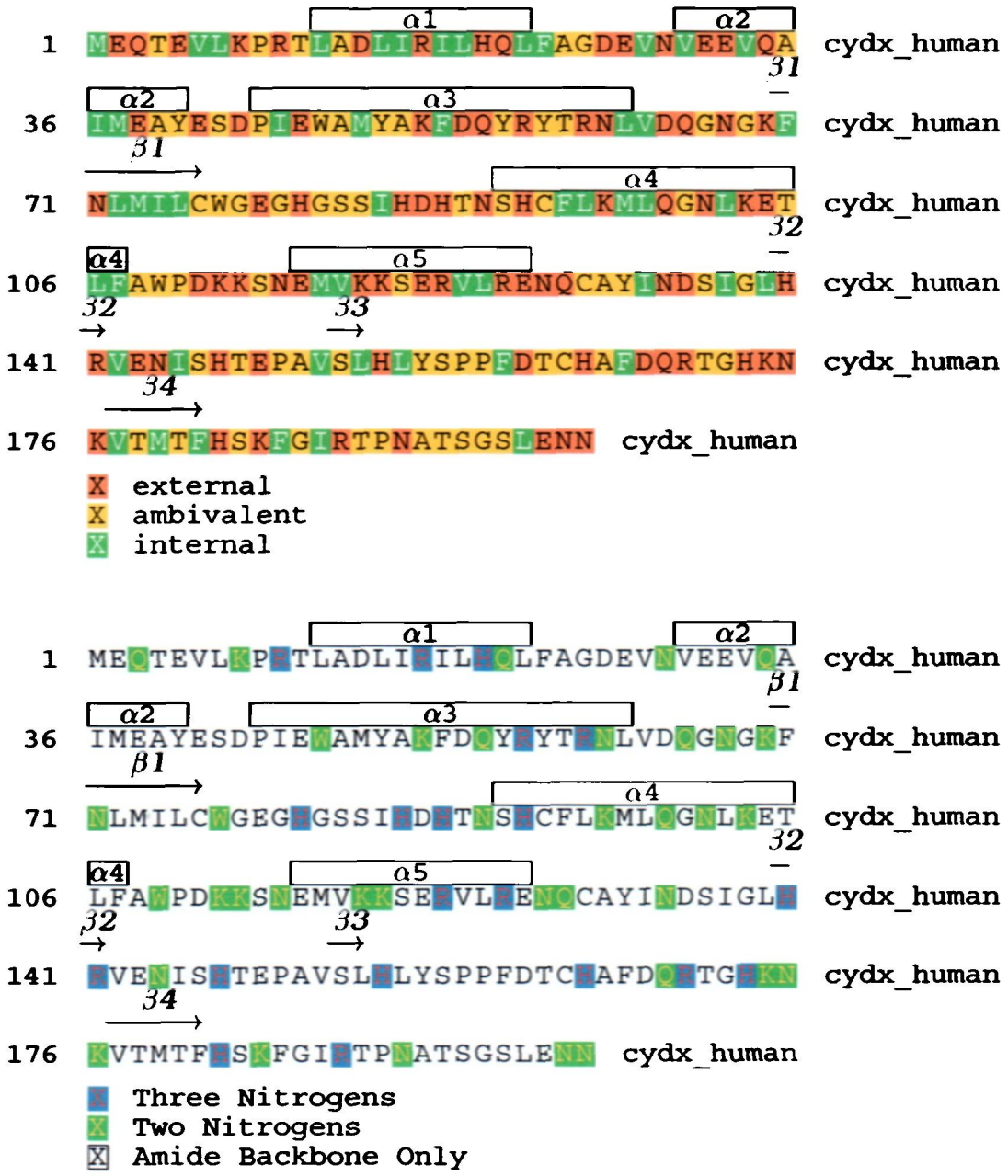


Figure A.17: Comparison of human CDO according to topography and nitrogen content with relative secondary structures

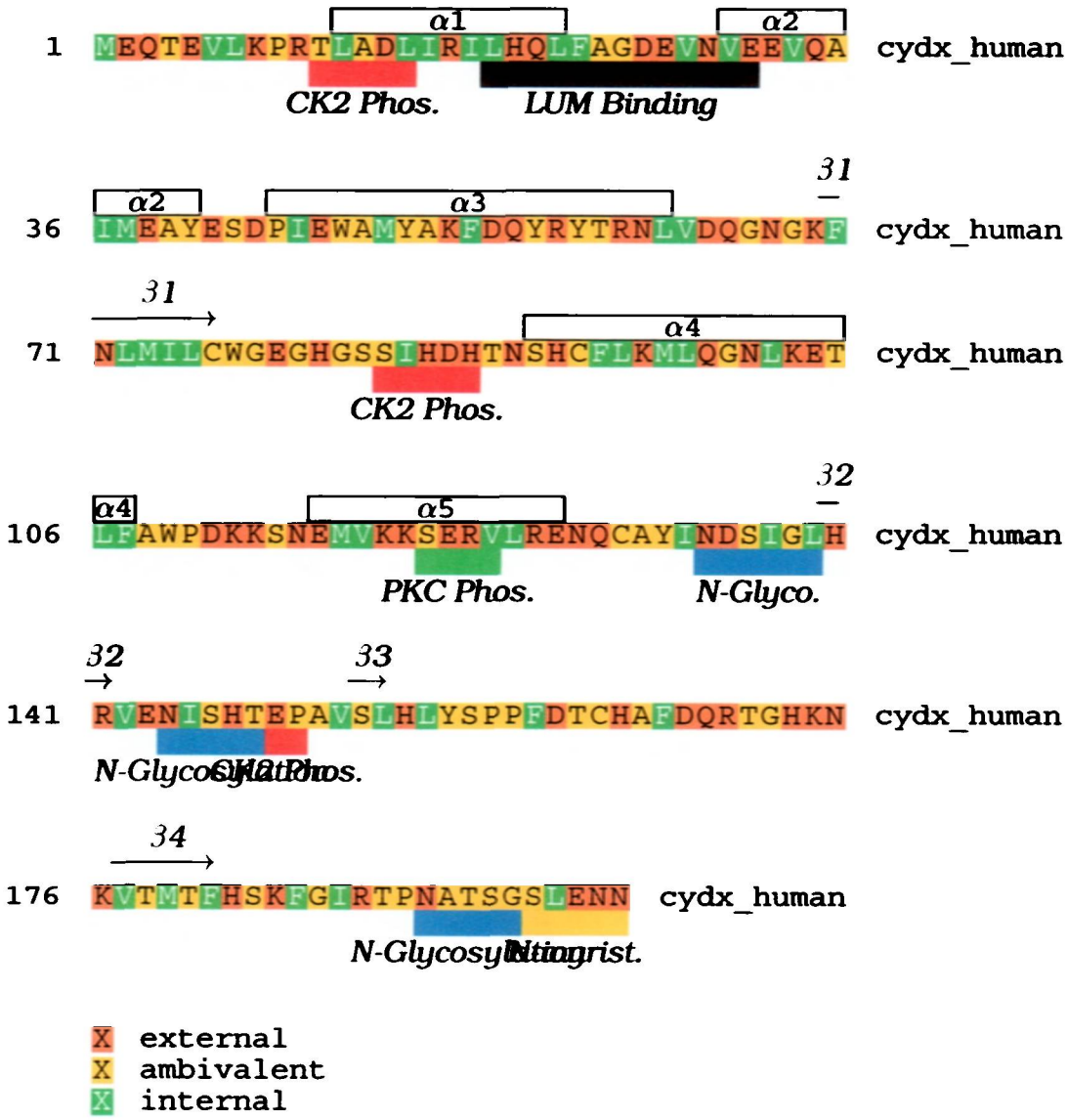


Figure A.18: Human CDO with predicted structural motifs and modification sites.

Protein Concentration & Solubility Assays Using Surface Plasmon Resonance

Christopher H. Barry, Adrian C. Williams and David B. Ramsden†

Department of Clinical Neuroscience, University of Birmingham, England

†Department of Medicine, University of Birmingham, England

Surface plasmon resonance is a responsive and non-destructive alternative for measuring the concentration of proteins in solution. Both the detection limits and resolution are unparallelled using non-destructive laboratory techniques. The method described in this experiment required minimal protein preparation and no chromophore to determine the concentration of a protein sample in solution. This method differentiates itself from other methods in that the analysis is performed on a planar interface with the sample; rather than a net analysis of the sample as a whole.

Spectroscopic and colorimetric methods have been devised to approximate protein concentration. Often these methods show a linear response to the concentration over a small range or require the destruction of the sample using a chromophore. It has already been established that there is a relation between the refractive index and the concentration of a sample[1]. Increments in these values differ widely for different proteins and must be experimentally determined [2]. Surface plasmon resonance (SPR) is a physical method that measures a quantity proportional to the refractive index and is a technology used in new chip based bio-technologies[3]. The only requirements for an accurate assay is a protein in solution and a sample of the buffer alone. SPR is a preferable measurement technique for microliter volumes or precious samples where accuracy and recovery are paramount.

Materials and Methods

An SPR device consists of a thin metal film deposited upon an optical substrate with an incident columnated light beam. The incident light, at a critical angle, induces a plasmon wave through the metal film. An analyte is exposed to the surface opposite the incident beam and optical substrate. Changes in the angular intensity of the reflected light are detected and are proportional to the refractive index of the sample, exposed to the planar surface.

A Biacore 3000 instrument was used for all measurements. An unmodified gold surface was used to detect changes in plasmon resonance (Biacore Pioneer Sensor Chip J1, 99-1000-01). Data analysis was evaluated from the raw

responses of Biacore unit. The instrument was programmed to inject 10 μL aliquots of sample, at a flow rate of 10 μL/min, through one micro-channel at 24°C.

Bovine serum albumin (BSA) and lysozyme were both purchased from Sigma (A-8022 and L-7651). Sodium chloride gradients were diluted with distilled and de-ionised water from a 20% stock NaCl solution. Solutions of BSA were prepared by preparing a concentrate in 150 mM NaCl and diluting it in the same solution. Samples were filtered through a 0.22 μm syringe-top filter prior to dilution. To prevent adsorption of lysozyme to the plasmon surface, washing steps between protein injections consisted of 150mM NaCl supplemented with 100mM L-cysteine. Samples were intended to be solubilised only and were not buffered to simplify measurements [2].

Results and Discussion

The output units on the Biacore instrument are in terms of resonance units (RU) as opposed to refractive index units. To confirm that there is a linear relation between the refractive index and RU, a sodium chloride gradient was measured and compared to the standard index of refraction values at the same concentration taken from a CRC Handbook. Figure 1 compares the linearity of the two methods. Sensorgraph readings (Figure 1A) are plotted as a function of time to illustrate the raw resonance reading as each sample of the gradient is injected automatically in turn. The change in RU with respect to concentration can be approximateted by 1,517.6 $\frac{\partial RU}{\partial c}$(mL/mg) using a linear trend line (Figure 1C); Several orders of magnitude greater than a $\frac{\partial n}{\partial c}$ of 0.00185 mL/mg over the same range.

The same method was applied for BSA. Figure 2A illustrates the measured changes in resonance units (RU) with the tabulated data. Due to small changes in baseline measurements between samples, changes from the baseline (ΔRU) were used to correlate protein concentration with RU in tabulated data. BSA shows a linear change in RU with protein concentration (Figure 2C).

The same method was applied to lysozyme. Initial results failed to bring the protein peak back to baseline. It was assumed that this was due to adsorption of free sulfhydryl groups of the protein to the gold surface of the sensor chip[4]. Supplementing the wash steps in the procedure with 100mM L-cysteine and 100mM NaCl was effective in bringing each measured sample to near baseline. Washing steps separate sample measurements. The additive in the lysozyme wash steps increase the refractive index for those intervals which explains the apparent negative RU at lysozyme concentrations of 0 and 250 μg/mL (Figure 2B, indicated by arrows). Lysozyme shows a nearly linear change in RU with concentration from 0 to 9 mg/mL. There was a drop in RU at 10 mg/mL. Because the RU is a reflection of concentration as opposed to absolute density, the apparent decrease in RU at 10mg/mL is due to a

decrease in the dispersion of the sample; Consistent with a protein dimerising at high concentrations.

A plot of the change in resonance (ΔRU) versus concentration, in mg/mL, between BSA and lysozyme differs substantially. Differences in response to concentration for BSA and lysozyme are nearly equal in magnitude to their individual molecular weights of the two species (Figure 2D). This indicates, while the refractive index is a function of density RI(ρ)[5], plasmon resonance is a function of density and individual mass, RU(ρ,m).

The results of this technique have been applied to two commonly available proteins and shows that approximate concentrations can be measured. Because the technique is a reflection of the number of particles in solution rather than the absolute mean in solution, it is possible to monitor aggregation of a samples. Changes in the refractive index for macromolecular systems have been cited in the literature and approximation of $\frac{\partial n}{dc}$ values cannot be approximated for individual monomer components as is practised in analytical ultracentrifugation[2]. SPR is an alternative or complementary technique to static or dynamic light scattering because the measurements do not require integration time, are not biased toward larger particle species, and are not influenced by flow. The technique could be adopted for use in tandem with spectroscopic analysis; It would be a valuable tool if the simultaneous measurement of two variables, concentration and aggregation state, could monitored in real-time during dialysis or other routine laboratory procedures.

References

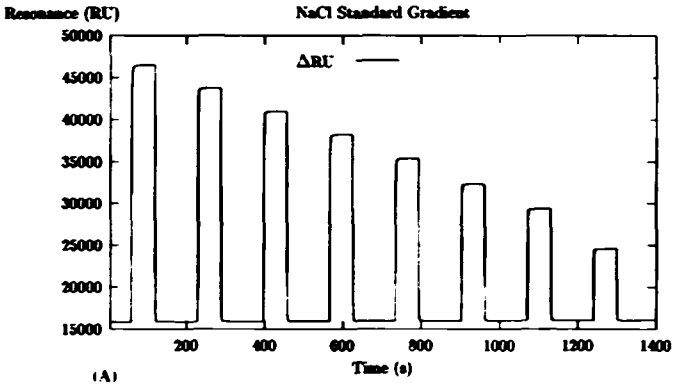
[1] C. Hitscherich, Jr, J. Kaplan, M. Allaman, J. Wiencek, and P. J. Loll. Static light scattering studies of ompf porin: implications for integral membrane protein crystallization. *Protein Sci*, 9(8):1559–66, August 2000.

[2] M.M. Ho K. Jumel and B. Bolgiano. Evaluation of meningoccal c oligosaccharide conjugate vaccines by size-exclusion chromatography/multi-angle laser light scattering. *Biotechnol. Appl. Biochemistry*, 36:219–226, 2002.

[3] M. Malmqvist. Biacore: an affinity biosensor system for characterization of biomolecular interactions. *Biochem Soc Trans*, 27(2):335–40, February 1999.

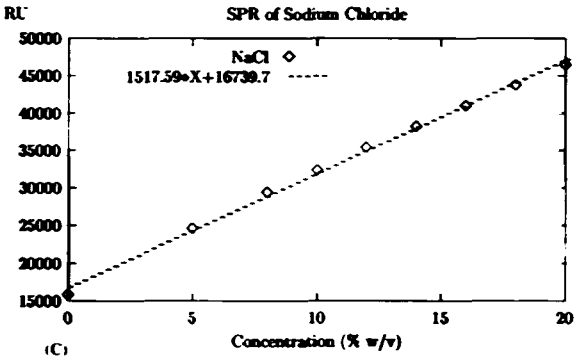
[4] A. Kuhnle, T. R. Linderoth, B. Hammer, and F. Besenbacher. Chiral recognition in dimerization of adsorbed cysteine observed by scanning tunnelling microscopy. 415(6874):891–3, Feb 21 2002.

[5] E. Stellwagen J. Babul. Measurement of protein concentration with interference optics. *Annal. Bioc.*, 28:216–221, 1969.



Conc.(%)	Resonance (RU)	Ref.Index (n)
0	15826.7	1.33054
5	24576.7	1.3418
8	29378.6	1.347
10	32364.2	1.3505
12	35369.4	1.3541
14	38229.1	1.3576
16	40983.6	1.3612
18	43757.5	1.3648
20	46483.1	1.3684

(B)



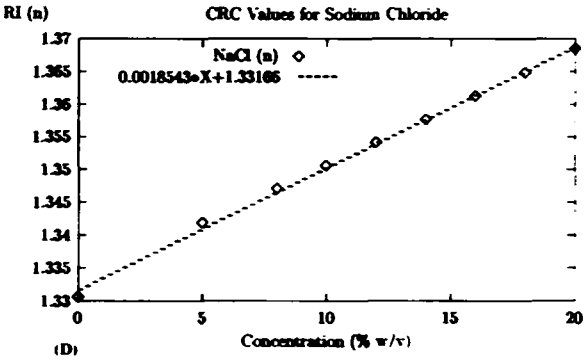
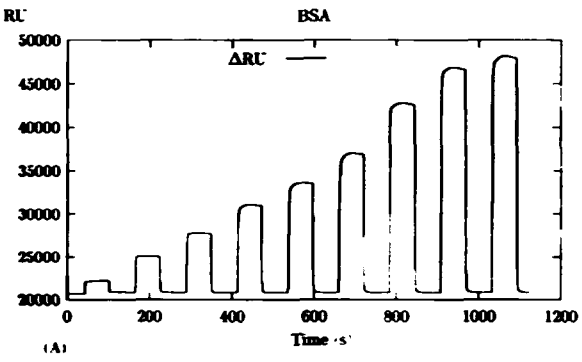
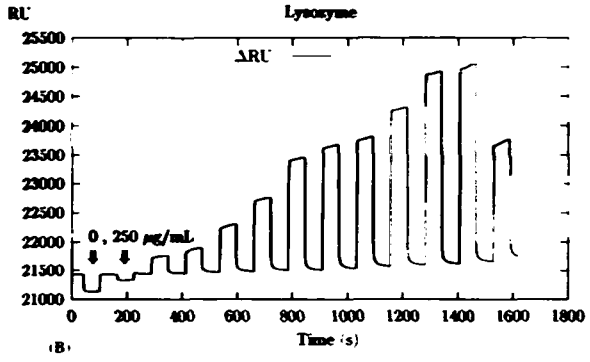
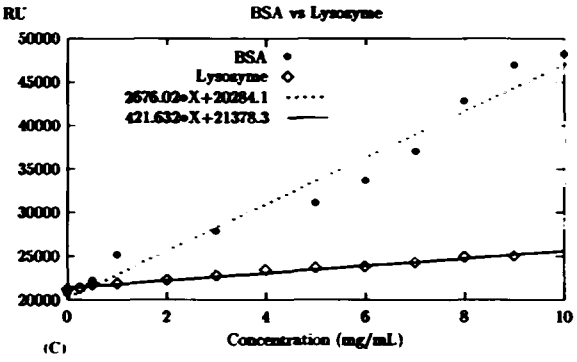


Figure 1: SPR sensorgraph of raw data for a series of 20% to 5% NaCl injections (A). Collected data was tabulated against known refractive indices at the same concentrations (B). To confirm the linearity of the two methods, the concentration gradient was then plotted against its resonant data (C) and its known refractive index (D).







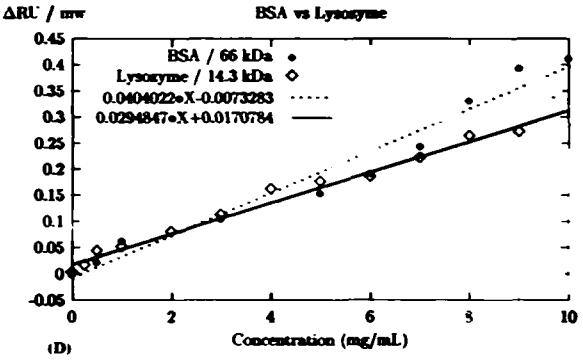


Figure 2: Sensorgraphs of BSA (A), and Lysozyme (B) raw data over a sample range of 0 mg/mL to 10 mg/mL. Data was plotted as a function of concentration (C) and molecular weight and concentration (D). Curves were fit using a first order polynomial and plotted using Gnuplot and LaTeX.

List of Figures

1.1 Reaction catalyzed by cysteine dioxygenase 11
1.2 Sanger Centre genome map with locus of human CDO gene . . . 11
1.3 End-products of L-cysteine involved in the anti-oxidant defence. . 13
1.4 Common mitochondrial sites documented a being impaired in different neurodegenerative pathologies. (Image taken from the KEGG database) 25
1.5 Two step process for biosynthesis of glutathione from L-cysteine. 30
1.6 Biochemically relevant roles of L-cysteine in intracellular homoeostasis. 34
1.7 Phylogenetic analysis of sequence data of human dioxygenases. . 43
1.8 Protein sequence homology of CDO 45

2.1 Representation of the aims of the study. 49

3.1 Predicted N-glycosylation of endogenous CDO using NetNGlyc 1.0. 51
3.2 pET15b (Novagen, Inc.) 52
3.3 Plasmid map of the pET15b-CDO construct 53
3.4 SOB growth medium and plating medium 53
3.5 PCR relevant mixtures. 55
3.6 GCG representation of predicted hydrophilicity of the SwissProt sequence entry of human cysteine dioxygenase. 60
3.7 A local alignment showing the location of the synthesised cysteine dioxygenase epitope with the human cysteine dioxygenase gene product. 60
3.8 Promoter to terminator product from colony PCR of pET15b-hCDO (colonies 1,3,5 respectively) in BL21(DE3) 62
3.9 (a) Western blot of CDO (2) with normal human(3), horse(4), and sheep sera(5) as controls.(b) Dot blots to CDO detected immunochemically. (Size standards are listed in the Appendix on page 178) 64

3.10 (a)SDS-PAGE and (b)Western blot analysis of human liver homogenate; flow through of homogenate of induced BL21(DE3); and affinity purified CDO **without using protease inhibitors**. (Size standards are listed in the Appendix on page 178). Four lanes are indicated in both gel and blot; starting with the size marker on the left. 65
3.11 Dot blot of CDO eluted from a nickel column. Gel overload of trypsinogen, CDO, and a lower concentration of CDO(Size standards are listed in the Appendix on page 178) 67
3.12 Dot blot of elute at pH 6.5 and pH 7.5. A 15% SDS-PAGE of CDO without β-mercaptoethanol (6.5), CDO with β-mercaptoethanol, flow-through from the column, and concentrated CDO 1 week with β-mercaptoethanol.(Size standards are listed in the Appendix on page 178) 67
3.13 Growth curve for BL21(DE3) in SOB media 68

4.1 Practical theory of SPR 72
4.2 Lysozyme. (Image taken from Protein Solutions, Inc.) 77
4.3 MALDI-TOF of CDO 80
4.4 Acquired SPR data. (Tabulated data listed on following pages) . . 83
4.5 Acquired SPR data. (Tabulated data listed on following pages) . . 84
4.6 SPR of NaCl Standard Gradient 85
4.7 BSA standard gradient 86
4.8 Lysozyme standard gradient 87
4.9 Measurement of time dependant aggregation of CDO using dynamic light scattering. 90
4.10 Time dependant measurements: (a) CDO in PBS (b) CDO in PBS with 0.01% β-mercaptoethanol (c) CDO in PBS with 0.01% SDS (d) CDO in PBS with 1mM Glycine 94
4.11 Cysteine Dioxygenase suspended in 25mM Glycine/200mM NaF 95
4.12 Influence of viscosity on particle size. Monodisperse CDO in NaF without (Red) and with 10% glycerol (Green) 95
4.13 Cysteine Dioxygenase in 1mM Glycine/1M NaCl/10mM Phosphate pH 6 96
4.14 Cysteine Dioxygenase in 1mM Glycine/1M NaCl/10mM Phosphate/excess iodoacetic acid pH 6 96
4.15 Cysteine Dioxygenase in 100mM PO_4/1mM Gly/200mM NaCl pH 7.5, 10mM PO_4/20mM Gly/100mM NaF pH 8.0, 10mM PO_4/20mM Gly/100mM NaF]
4.16 Effect of pH on CDO aggregation 98
4.17 Effect of pH on CDO aggregation 99
4.18 Photon correlation spectroscopy results of affinity purified CDO. . 100

5.1 Standard curves for pure peptide solutions containing 100% α-helix, β-sheet, and random coil. Image taken from Protein Solutions, Inc. 103
5.2 Schematic of the hanging drop method 105
5.3 Predicted secondary structure of cysteine dioxygenase aligned with its primary sequence. 110
5.4 Predicted secondary structural homology of CDO 112
5.5 Recombinant cysteine dioxygenase pH dependant changes in the optical rotation of light in 100mM NaF,20mM Gly, 10mM PO_4. (pH 8-blue, pH 7.4-green, red-pH7.0, and pH 5-yellow 114
5.6 Total loss of secondary structural information under conditions of protein aggregation of the recombinant CDO. (1.24mg/ml, 100mM NaF, 20mM Gly, 250nm $FeSO_4$ pH 7.4) 115
5.7 CDNN 2.1 interpretation of CD spectra using the 210nm-260nm method. 117
5.8 Predicted isoelectric point of CDO calculated using GCG Predictions based on primary sequence of the 220aa recombinant CDO. 118
5.9 CD spectrum and deconvolution of CDO at 200mM Na_2HPO_4 pH 7.0 119
5.10 f1f2 120
5.11 f3f4 121
5.12 f5f6 122
5.13 f7f8 123
5.14 Conversion of L-Cysteine to L-cysteinesulfinic acid 124
5.15 1-D NMR of (a) cysteine and (b) cysteinesulfinic acid 125
5.16 1-D NMR spectra of NAD^+ 126
5.17 1-D NMR spectra of (a) cysteine, NAD^+ and (b) cysteine, NAD^+, and cysteine dioxygenase 128
5.18 Crystals formed in elution fractions that contained CDO. 130
5.19 Results of protein crystallisation matrix screen 131
5.20 Close up of CDO in hanging drop. 132
5.21 A preliminary 1-D spectra of cysteine dioxygenase 135
5.22 2-D HSQC ^1H-^{15}N spectra of ^{15}N labelled cysteine dioxygenase in 320mM NaF, 1mM Glycine, 10mM PO_4^{-3}, 400μM EDTA pH 7.8 . . 136
5.23 2-D HSQC ^1H-^{15}N spectra of ^{15}N labelled cysteine dioxygenase in 320mM NaF, 1mM Glycine, 10mM PO_4^{-3}, 400μM EDTA pH 7.8 at increased sensitivity and integration time. 137

A.1 Predicted N-Glycosylation of CDO 144
A.2 Proposed mechanism of quercetin 2,3-dioxygenase[104] 145
A.3 Proposed mechanism of 2,3-dihydroxybiphenyl 1,2-dioxygenase[105] 146
A.4 Predicted N-glycosylation sites on human CDO using NetNGlyc . 147
A.5 Predicted O-glycosylation sites on human CDO using NetOGlyc . 148
A.6 Novagen pET15b vector 149

A.7 Predicted secondary features by TEXshade based upon PhD sequence analysis of CDO in preceding data. 166
A.8 Forward sequence of pET15b-hCDO construct 168
A.9 Reverse sequence of pET15b-hCDO construct 169
A.10 sequence of pET15b-hCDO construct 170
A.11 sequence of pET15b-hCDO construct 171
A.12 sequence of pET15b-hCDO construct 172
A.13 sequence of pET15b-hCDO construct 173
A.14 Alignments of construct sequencing with consensus sequence (1/3) 174
A.15 Alignments of construct sequencing with consensus sequence (2/3) 175
A.16 Alignments of construct sequencing with consensus sequence (3/3) 176
A.17 Comparison of human CDO according to topography and nitrogen content with relative secondary structures 180
A.18 Human CDO with predicted structural motifs and modification sites. 181

Bibliography

[1] S. Sakakibara, K. Yamaguchi, Y. Hosokawa, N. Kohashi, and I. Ueda. Purification and some properties of rat liver cysteine oxidase (cysteine dioxygenase). *Biochim Biophys Acta*, 422(2):273–9, Feb 13 1976.

[2] D. B. Ramsden, A. Kapadi, N. J. Fitch, M. J. Farmer, P. Bennett, and A. C. Williams. Human cysteine dioxygenase type i (cdo-i; ec 1.13.11.20): 5' flanking region and intron-exon structure of the gene. *Mol Pathol*, 50(5):269–71, October 1997.

[3] R. Janaky, V. Varga, A. Hermann, P. Saransaari, and S. S. Oja. Mechanisms of l-cysteine neurotoxicity. *Neurochem Res*, 25(9-10):1397–405, October 2000.

[4] D. M. Robertson R. L. Davis. *Textbook of Neuropathology, 2nd ed.* Williams and Wilkins.

[5] DC Nemo, GJ Gajdusek. Attempt to demonstrate virus interference in cell cultures persistently infected with the viruses of kuru and creutzfeldt-jacob disease. *Proc. Soc. Exp. Biol. Med.*, 163(2):171–6, 1980.

[6] M.P. Prusiner SB Bolton, D.C. McKinnley. Purification of prions from scrapie infected hamster brain... 218(4579):1309–11, 1982.

[7] Kiltzman R. Sorcery and science: Responses to kuru and other epidemics. *Western J. of Medicine*, 171(3):204–6, 1999.

[8] C. Weissmann and A. Aguzzi. Bovine spongiform encephalopathy and early onset variant creutzfeldt-jakob disease. *Curr Opin Neurobiol*, 7(5):695–700, October 1997.

[9] J. G. Fournier, F. Escaig-Haye, and V. Grigoriev. Ultrastructural localization of prion proteins: physiological and pathological implications. *Microsc Res Tech*, 50(1):76–88, Jul 1 2000.

[10] S. Capellari, S. I. Zaidi, C. B. Urig, G. Perry, M. A. Smith, and R. B. Petersen. Prion protein glycosylation is sensitive to redox change [published erratum appears in j biol chem 2000 apr 14;275(15):11538]. *J Biol Chem*, 274(49):34846–50, Dec 3 1999.

[11] W. E. Muller, J. Laplanche, H. Ushijima, and H. C. Schroder. Novel approaches in diagnosis and therapy of creutzfeldt-jakob disease [in process citation]. *Mech Ageing Dev*, 116(2-3):193–218, Jul 31 2000.

[12] A. R. White, S. J. Collins, F. Maher, M. F. Jobling, L. R. Stewart, J. M. Thyer, K. Beyreuther, C. L. Masters, and R. Cappai. Prion protein-deficient neurons reveal lower glutathione reductase activity and increased susceptibility to hydrogen peroxide toxicity. *Am J Pathol*, 155(5):1723–30, November 1999.

[13] M. Guentchev, T. Voigtlander, C. Haberler, M. H. Groschup, and H. Budka. Evidence for oxidative stress in experimental prion disease. *Neurobiol Dis*, 7(4):270–3, August 2000.

[14] S. I. Choi, W. K. Ju, E. K. Choi, J. Kim, H. Z. Lea, R. I. Carp, H. M. Wisniewski, and Y. S. Kim. Mitochondrial dysfunction induced by oxidative stress in the brains of hamsters infected with the 263 k scrapie agent. *Acta Neuropathol (Berl)*, 96(3):279–86, September 1998.

[15] A. Alves-Rodrigues, L. Gregori, and M. E. Figueiredo-Pereira. Ubiquitin, cellular inclusions and their role in neurodegeneration. *Trends Neurosci*, 21(12):516–20, December 1998.

[16] S.Smith. Private communication to d.b.ramsden. Private communication to D.B.Ramsden.

[17] A. J. Cooper, K. F. Sheu, J. R. Burke, W. J. Strittmatter, V. Gentile, G. Peluso, and J. P. Blass. Pathogenesis of inclusion bodies in (cag)n/qn-expansion diseases with special reference to the role of tissue transglutaminase and to selective vulnerability. *J Neurochem*, 72(3):889–99, March 1999.

[18] M. Munar-Ques, J. L. Pedrosa, T. Coelho, L. Gusmao, R. Seruca, A. Amorim, and J. Sequeiros. Two pairs of proven monozygotic twins discordant for familial amyloid neuropathy (fap) ttr met 30. *J Med Genet*, 36(8):629–32, August 1999.

[19] A. H. Schapira. Mitochondrial involvement in parkinson's disease, huntington's disease, hereditary spastic paraplegia and friedreich's ataxia [see comments]. *Biochim Biophys Acta*, 1410(2):159–70, Feb 9 1999.

[20] G. Bartzokis, J. Cummings, S. Perlman, D. B. Hance, and J. Mintz. Increased basal ganglia iron levels in huntington disease. *Arch Neurol*, 56(5):569–74, May 1999.

[21] J. Sian, D. T. Dexter, A. J. Lees, S. Daniel, Y. Agid, F. Javoy-Agid, P. Jenner, and C. D. Marsden. Alterations in glutathione levels in parkinson's disease and other neurodegenerative disorders affecting basal ganglia [see comments]. *Ann Neurol*, 36(3):348–55, September 1994.

[22] J. Busciglio, H. Hartmann, A. Lorenzo, C. Wong, K. Baumann, B. Sommer, M. Staufenbiel, and B. A. Yankner. Neuronal localization of presenilin-1 and association with amyloid plaques and neurofibrillary tangles in alzheimer's disease. *J Neurosci*, 17(13):5101–7, Jul 1 1997.

[23] S. Janciauskiene, H. Rubin, C. M. Lukacs, and H. T. Wright. Alzheimer's peptide abeta1-42 binds to two beta-sheets of alpha1-antichymotrypsin and transforms it from inhibitor to substrate. *J Biol Chem*, 273(43):28360–4, Oct 23 1998.

[24] K. Ogawa, T. Yamada, Y. Tsujioka, J. Taguchi, M. Takahashi, Y. Tsuboi, Y. Fujino, M. Nakajima, T. Yamamoto, H. Akatsu, S. Mitsui, and N. Yamaguchi. Localization of a novel type trypsin-like serine protease, neurosin, in brain tissues of alzheimer's disease and parkinson's disease [in process citation]. *Psychiatry Clin Neurosci*, 54(4):419–26, August 2000.

[25] P. Gomez-Ramos and M. A. Moran. Ultrastructural localization of butyrylcholinesterase in senile plaques in the brains of aged and alzheimer disease patients. *Mol Chem Neuropathol*, 30(3):161–73, April 1997.

[26] G. S. Withers, J. M. George, G. A. Banker, and D. F. Clayton. Delayed localization of synelfin (synuclein, nacp) to presynaptic terminals in cultured rat hippocampal neurons. *Brain Res Dev Brain Res*, 99(1):87–94, Mar 17 1997.

[27] R. D. Hollister, W. Kisiel, and B. T. Hyman. Immunohistochemical localization of tissue factor pathway inhibitor-1 (tfpi-1), a kunitz proteinase inhibitor, in alzheimer's disease. *Brain Res*, 728(1):13–9, Jul 22 1996.

[28] S. Silhol, A. Calenda, V. Jallageas, N. Mestre-Frances, M. Bellis, and N. Bons. beta-amyloid protein precursor in microcebus murinus: genotyping and brain localization. *Neurobiol Dis*, 3(3):169–82, 1996.

[29] N. Peress, E. Perillo, and S. Zucker. Localization of tissue inhibitor of matrix metalloproteinases in alzheimer's disease and normal brain. *J Neuropathol Exp Neurol*, 54(1):16–22, January 1995.

[30] J. H. Su, B. J. Cummings, and C. W. Cotman. Localization of heparan sulfate glycosaminoglycan and proteoglycan core protein in aged brain and alzheimer's disease. 51(4):801–13, December 1992.

[31] J. D. Adams, L. K. Klaidman, I. N. Odunze, H. C. Shen, and C. A. Miller. Alzheimer's and parkinson's disease. brain levels of glutathione, glutathione disulfide, and vitamin e. *Mol Chem Neuropathol*, 14(3):213–26, June 1991.

[32] N. W. Kowall and M. F. Beal. Glutamate-, glutaminase-, and taurine-immunoreactive neurons develop neurofibrillary tangles in alzheimer's disease [see comments]. *Ann Neurol*, 29(2):162–7, February 1991.

[33] R. B. Parsons, R. H. Waring, D. B. Ramsden, and A. C. Williams. In vitro effect of the cysteine metabolites homocysteic acid, homocysteine and cysteic acid upon human neuronal cell lines. 19(4-5):599–603, Aug-Oct 1998.

[34] M. T. Heafield, S. Fearn, G. B. Steventon, R. H. Waring, A. C. Williams, and S. G. Sturman. Plasma cysteine and sulphate levels in patients with motor neurone, parkinson's and alzheimer's disease. *Neurosci Lett*, 110(1-2):216–20, Mar 2 1990.

[35] A. Eisen and D. Calne. Amyotrophic lateral sclerosis, parkinson's disease and alzheimer's disease: phylogenetic disorders of the human neocortex sharing many characteristics. *Can J Neurol Sci*, 19(1 Suppl):117–23, February 1992.

[36] E. Andrasi, S. Igaz, Z. Molnar, and S. Mako. Disturbances of magnesium concentrations in various brain areas in alzheimer's disease [in process citation]. *Magnes Res*, 13(3):189–96, September 2000.

[37] J. Durlach. Magnesium depletion and pathogenesis of alzheimer's disease. *Magnes Res*, 3(3):217–8, September 1990.

[38] S. I. Dikalov, M. P. Vitek, K. R. Maples, and R. P. Mason. Amyloid beta peptides do not form peptide-derived free radicals spontaneously, but can enhance metal-catalyzed oxidation of hydroxylamines to nitroxides. *J Biol Chem*, 274(14):9392–9, Apr 2 1999.

[39] V. Rondeau, D. Commenges, H. Jacqmin-Gadda, and J. F. Dartigues. Relation between aluminum concentrations in drinking water and alzheimer's disease: an 8-year follow-up study. *Am J Epidemiol*, 152(1):59–66, Jul 1 2000.

[40] H. C. Powell, R. S. Garrett, A. Muehlenbachs, F. M. Brett, and I. L. Campbell. Crystalloid inclusions in brain macrophages and hemopoietic tissue in gfap-il3 mice resemble inclusions identified in multiple sclerosis. *Ultrastruct Pathol*, 23(5):285–97, Sep-Oct 1999.

[41] E. Karg, P. Klivenyi, I. Nemeth, K. Bencsik, S. Pinter, and L. Vecsei. Nonenzymatic antioxidants of blood in multiple sclerosis. *J Neurol*, 246(7):533–9, July 1999.

[42] V. Calabrese, R. Raffaele, E. Cosentino, and V. Rizza. Changes in cerebrospinal fluid levels of malondialdehyde and glutathione reductase activity in multiple sclerosis. *Int J Clin Pharmacol Res*, 14(4):119–23, 1994.

[43] V. Calabrese, R. Bella, D. Testa, F. Spadaro, A. Scrofani, V. Rizza, and G. Pennisi. Increased cerebrospinal fluid and plasma levels of ultraweak chemiluminescence are associated with changes in the thiol pool and lipid-soluble fluorescence in multiple sclerosis: the pathogenic role of oxidative stress. *Drugs Exp Clin Res*, 24(3):125–31, 1998.

[44] M. Gunnarsson, T. Stigbrand, and P. E. Jensen. Aberrant forms of alpha(2)-macroglobulin purified from patients with multiple sclerosis. *Clin Chim Acta*, 295(1-2):27–40, May 2000.

[45] R. Schroder, I. Nennesmo, and R. P. Linke. Amyloid in a multiple sclerosis lesion in clearly of alambda type [in process citation]. *Acta Neuropathol (Berl)*, 100(6):709–11, December 2000.

[46] S. M. LeVine. Iron deposits in multiple sclerosis and alzheimer's disease brains. *Brain Res*, 760(1-2):298–303, Jun 20 1997.

[47] C. W. McGrother, C. Dugmore, M. J. Phillips, N. T. Raymond, P. Garrick, and W. O. Baird. Multiple sclerosis, dental caries and fillings: a case-control study. *Br Dent J*, 187(5):261–4, Sep 11 1999.

[48] S. Ono, N. Shimizu, T. Imai, A. Mihori, and K. Nagao. Increased cystatin c immunoreactivity in the skin in amyotrophic lateral sclerosis [in process citation]. *Acta Neurol Scand*, 102(1):47–52, July 2000.

[49] H. Tohgi, T. Abe, K. Yamazaki, T. Murata, E. Ishizaki, and C. Isobe. Increase in oxidized no products and reduction in oxidized glutathione in cerebrospinal fluid from patients with sporadic form of amyotrophic lateral sclerosis. *Neurosci Lett*, 260(3):204–6, Feb 5 1999.

[50] O. A. Andreassen, A. Dedeoglu, P. Klivenyi, M. F. Beal, and A. I. Bush. N-acetyl-l-cysteine improves survival and preserves motor performance in an animal model of familial amyotrophic lateral sclerosis [in process citation]. 11(11):2491–3, Aug 3 2000.

[51] S. Malessa, P. N. Leigh, O. Bertel, E. Sluga, and O. Hornykiewicz. Amyotrophic lateral sclerosis: glutamate dehydrogenase and transmitter amino acids in the spinal cord. *J Neurol Neurosurg Psychiatry*, 54(11):984–8, November 1991.

[52] T. L. Perry, S. Hansen, and K. Jones. Brain glutamate deficiency in amyotrophic lateral sclerosis. 37(12):1845–8, December 1987.

[53] M. Neumann, S. Adler, O. Schluter, E. Kremmer, R. Benecke, and H. A. Kretzschmar. Alpha-synuclein accumulation in a case of neurodegeneration with brain iron accumulation type 1 (nbia-1, formerly hallervorden-spatz syndrome) with widespread cortical and brainstem-type lewy bodies [in process citation]. *Acta Neuropathol (Berl)*, 100(5):568–74, November 2000.

[54] Y. Saito, M. Kawai, K. Inoue, R. Sasaki, H. Arai, E. Nanba, S. Kuzuhara, Y. Ihara, I. Kanazawa, and S. Murayama. Widespread expression of alpha-synuclein and tau immunoreactivity in hallervorden-spatz syndrome with protracted clinical course [in process citation]. *J Neurol Sci*, 177(1):48–59, Aug 1 2000.

[55] T. L. Perry, M. G. Norman, V. W. Yong, S. Whiting, J. U. Crichton, S. Hansen, and S. J. Kish. Hallervorden-spatz disease: cysteine accumulation and cysteine dioxygenase deficiency in the globus pallidus. *Ann Neurol*, 18(4):482–9, October 1985.

[56] J. B. Winer, R. A. Hughes, M. J. Anderson, D. M. Jones, H. Kangro, and R. P. Watkins. A prospective study of acute idiopathic neuropathy. ii. antecedent events. *J Neurol Neurosurg Psychiatry*, 51(5):613–8, May 1988.

[57] F. Graus, C. Cordon-Cardo, and J. B. Posner. Neuronal antinuclear antibody in sensory neuronopathy from lung cancer. 35(4):538–43, April 1985.

[58] K. Kobayashi, M. Hayashi, Y. Fukutani, K. Miyazu, M. Shiozawa, F. Muramori, T. Aoki, and Y. Koshino. Kp1 expression of ghost pick bodies, amyloid p-positive astrocytes and selective nigral degeneration in early onset picks disease. *Clin Neuropathol*, 18(5):240–9, Sep-Oct 1999.

[59] A. Takeda, M. Hashimoto, M. Mallory, M. Sundsumo, L. Hansen, and E. Masliah. C-terminal alpha-synuclein immunoreactivity in structures other than lewy bodies in neurodegenerative disorders. *Acta Neuropathol (Berl)*, 99(3):296–304, March 2000.

[60] R. Castellani, M. A. Smith, P. L. Richey, R. Kalaria, P. Gambetti, and G. Perry. Evidence for oxidative stress in pick disease and corticobasal degeneration. *Brain Res*, 696(1-2):268–71, Oct 23 1995.

[61] O. Yasuhara, Y. Aimi, E. G. McGeer, and P. L. McGeer. Accumulation of amyloid precursor protein in brain lesions of patients with pick disease. *Neurosci Lett*, 171(1-2):63–6, Apr 25 1994.

[62] W. D. Ehmann, M. Alauddin, T. I. Hossain, and W. R. Markesbery. Brain trace elements in pick's disease. *Ann Neurol*, 15(1):102–4, January 1984.

[63] T. L. Perry, S. Hansen, and K. Jones. Brain amino acids and glutathione in progressive supranuclear palsy. 38(6):943–6, June 1988.

[64] D. S. Albers, S. J. Augood, L. C. Park, S. E. Browne, D. M. Martin, J. Adamson, M. Hutton, D. G. Standaert, J. P. Vonsattel, G. E. Gibson, and M. F. Beal. Frontal lobe dysfunction in progressive supranuclear palsy: evidence for oxidative stress and mitochondrial impairment. *J Neurochem*, 74(2):878–81, February 2000.

[65] R. H. Swerdlow, L. I. Golbe, J. K. Parks, D. S. Cassarino, D. R. Binder, A. E. Grawey, I. Litvan, , G. F. Wooten, and W. D. Parker. Mitochondrial dysfunction in cybrid lines expressing mitochondrial genes from patients with progressive supranuclear palsy. *J Neurochem*, 75(4):1681–4, October 2000.

[66] D. Caparros-Lefebvre and A. Elbaz. Possible relation of atypical parkinsonism in the french west indies with consumption of tropical plants: a case-control study. caribbean parkinsonism study group [see comments]. 354(9175):281–6, Jul 24 1999.

[67] K. Jellinger. *The Pathology of Parkinsonism.*

[68] R. K. Pearce, A. Owen, S. Daniel, P. Jenner, and C. D. Marsden. Alterations in the distribution of glutathione in the substantia nigra in parkinson's disease. *J Neural Transm*, 104(6-7):661–77, 1997.

[69] J. A. Molina, F. J. Jimenez-Jimenez, P. Gomez, C. Vargas, J. A. Navarro, M. Orti-Pareja, T. Gasalla, J. Benito-Leon, F. Bermejo, and J. Arenas. Decreased cerebrospinal fluid levels of neutral and basic amino acids in patients with parkinson's disease. *J Neurol Sci*, 150(2):123–7, Sep 10 1997.

[70] K. Arima, S. Hirai, N. Sunohara, K. Aoto, Y. Izumiyama, K. Ueda, K. Ikeda, and M. Kawai. Cellular co-localization of phosphorylated tau- and nacp/alpha-synuclein-epitopes in lewy bodies in sporadic parkinson's disease and in dementia with lewy bodies. *Brain Res*, 843(1-2):53–61, Oct 2 1999.

[71] R. J. Castellani, S. L. Siedlak, G. Perry, and M. A. Smith. Sequestration of iron by lewy bodies in parkinson's disease [in process citation]. *Acta Neuropathol (Berl)*, 100(2):111–4, August 2000.

[72] E. Kienzl, K. Jellinger, H. Stachelberger, and W. Linert. Iron as catalyst for oxidative stress in the pathogenesis of parkinson's disease? *Life Sci*, 65(18-19):1973–6, 1999.

[73] W. Kuhn, R. Winkel, D. Woitalla, S. Meves, H. Przuntek, and T. Muller. High prevalence of parkinsonism after occupational exposure to lead-sulfate batteries. 50(6):1885–6, June 1998.

[74] J. M. Gorell, B. A. Rybicki, C. Cole Johnson, and E. L. Peterson. Occupational metal exposures and the risk of parkinson's disease. 18(6):303–8, 1999.

[75] J. M. Gorell, B. A. Rybicki, C. C. Johnson, and E. L. Peterson. Smoking and parkinson's disease: a dose-response relationship [see comments]. 52(1):115–9, Jan 1 1999.

[76] A. Williams and R. Waring. The mptp tale: pathway to prevention of parkinson's disease? *Br J Hosp Med*, 49(10):716–9, May 19-Jun 1 1993.

[77] P. S. Spencer. Guam als/parkinsonism-dementia: a long-latency neurotoxic disorder caused by "slow toxin(s)" in food? *Can J Neurol Sci*, 14(3 Suppl):347–57, August 1987.

[78] J. Durlach, P. Bac, V. Durlach, A. Durlach, M. Bara, and A. Guiet-Bara. Are age-related neurodegenerative diseases linked with various types of magnesium depletion? *Magnes Res*, 10(4):339–53, December 1997.

[79] R. Wisnicka, A. Krzepilko, J. Wawryn, and T. Bilinski. Iron toxicity in yeast. *Acta Microbiol Pol*, 46(4):339–47, 1997.

[80] J. Liden and L.O.Ohman. Redox stabilization of iron and manganese in the +ii oxidation state by magnesium precipitates and some anionic polymers. implications for the oxygen based bleaching chemicals. *Journal of Pulp and Paper Science*, 5:J193–J199, May 1997.

[81] Yu A Berry RM Argyropoulos DS Y. Sun, Fenster M. The effect of metal ions on the reaction of hydrogen peroxide with kraft lignin. *Canadian Journal of Chemistry-Revue Canadienne de Chimie*, 77(5-6):667–675, May-Jun 1999.

[82] Robert Flanagan. What's your antedote? *Chemistry in Britain*, 38(2):29–31, 2002.

[83] J. M. McCord. The evolution of free radicals and oxidative stress. *Am J Med*, 108(8):652–9, Jun 1 2000.

[84] P. Wardman and L. P. Candeias. Fenton chemistry: an introduction. *Radiat Res*, 145(5):523–31, May 1996.

[85] D. R. Lloyd, P. L. Carmichael, and D. H. Phillips. Comparison of the formation of 8-hydroxy-2'-deoxyguanosine and single- and double-strand breaks in dna mediated by fenton reactions. *Chem Res Toxicol*, 11(5):420–7, May 1998.

[86] E. Y. Yang, S. X. Guo-Ross, and S. C. Bondy. The stabilization of ferrous iron by a toxic beta-amyloid fragment and by an aluminum salt. *Brain Res*, 839(2):221–6, Aug 28 1999.

[87] T. Chakraborti, S. Das, M. Mondal, S. Roychoudhury, and S. Chakraborti. Oxidant, mitochondria and calcium: an overview. *Cell Signal*, 11(2):77–85, February 1999.

[88] K. D. Held, F. C. Sylvester, K. L. Hopcia, and J. E. Biaglow. Role of fenton chemistry in thiol-induced toxicity and apoptosis. *Radiat Res*, 145(5):542–53, May 1996.

[89] J. E. Biaglow, K. D. Held, Y. Manevich, S. Tuttle, A. Kachur, and F. Uckun. Role of guanosine triphosphate in ferric ion-linked fenton chemistry. *Radiat Res*, 145(5):554–62, May 1996.

[90] A. Mukhopadhyay. Inclusion bodies and purification of proteins in biologically active forms. *Adv Biochem Eng Biotechnol*, 56:61–109, 1997.

[91] C. Wei, B. Tang, Y. Zhang, and K. Yang. Oxidative refolding of recombinant prochymosin. *Biochem J*, 340 (Pt 1):345–51, May 15 1999.

[92] M. Merad-Boudia, A. Nicole, D. Santiard-Baron, C. Saille, and I. Ceballos-Picot. Mitochondrial impairment as an early event in the process of apoptosis induced by glutathione depletion in neuronal cells: relevance to parkinson's disease. *Biochem Pharmacol*, 56(5):645–55, Sep 1 1998.

[93] Y.H. Kwon and M.H. Stipanuck. Cysteine regulates expression of cysteine dioxygenase and gamma-glutamylcysteine synthetase in cultured rat hepatocytes. *Am.J.Physiol.Endocrinol.Metab*, 280(5):E804–15, May 2001.

[94] M. Wulbeck J. Ribbe R.E. Kneusel, J. Crowe. *The Nucleic Acids Protocols Handbook*. Humana Press Inc., 1999.

[95] J. Dunlop. Substrate exchange properties of the high-affinity glutamate transporter eaat2. *J Neurosci Res*, 66(3):482–6, Nov 1 2001.

[96] A. Brand, D. Leibfritz, B. Hamprecht, and R. Dringen. Metabolism of cysteine in astroglial cells: synthesis of hypotaurine and taurine. *J Neurochem*, 71(2):827–32, August 1998.

[97] T. Fukuda, K. Ikejima, M. Hirose, Y. Takei, S. Watanabe, and N. Sato. Taurine preserves gap junctional intercellular communication in rat hepatocytes under oxidative stress. *J Gastroenterol*, 35(5):361–8, 2000.

[98] M. Palmi, G. T. Youmbi, F. Fusi, G. P. Sgaragli, H. B. Dixon, M. Frosini, and K. F. Tipton. Potentiation of mitochondrial ca2+ sequestration by taurine. *Biochem Pharmacol*, 58(7):1123–31, Oct 1 1999.

[99] M. B. O'Byrne and K. F. Tipton. Taurine-induced attenuation of mpp+ neurotoxicity in vitro: a possible role for the gaba(a) subclass of gaba receptors. *J Neurochem*, 74(5):2087–93, May 2000.

[100] H. Guizouarn, R. Motais, F. Garcia-Romeu, and F. Borgese. Cell volume regulation: the role of taurine loss in maintaining membrane potential and cell ph. *J Physiol (Lond)*, 523 Pt 1:147–54, Feb 15 2000.

[101] J. Moran, T. E. Maar, and H. Pasantes-Morales. Impaired cell volume regulation in taurine deficient cultured astrocytes. *Neurochem Res*, 19(4):415–20, April 1994.

[102] X. Shi, D. C. Flynn, D. W. Porter, S. S. Leonard, V. Vallyathan, and V. Castranova. Efficacy of taurine based compounds as hydroxyl radical scavengers in silica induced peroxidation. *Ann Clin Lab Sci*, 27(5):365–74, Sep-Oct 1997.

[103] B. Tadolini, G. Pintus, G. G. Pinna, F. Bennardini, and F. Franconi. Effects of taurine and hypotaurine on lipid peroxidation. *Biochem Biophys Res Commun*, 213(3):820–6, Aug 24 1995.

[104] K.H. Dijkstra B.W. Steiner, R.A. Kalk. Anaerobic enzyme substrate structures provide insight into the reaction mechansm of the copper dependant quercetin 2,3-dioxygenase. 99(26):16625–16630, 2002.

[105] L.D. Vaillancourt, FH Eltis. The mechanism-based inactivatio of 2,3-dihydroxybiphenyl 1,2-dioxygenase by catecholic substrates. *Journal of Biological Sciences*, 277(3):2019–2027, 2002.

[106] P. J. Bagley, L. L. Hirschberger, and M. H. Stipanuk. Evaluation and modification of an assay procedure for cysteine dioxygenase activity: high-performance liquid chromatography method for measurement of cysteine sulfinate and demonstration of physiological relevance of cysteine dioxygenase activity in cysteine catabolism. *Anal Biochem*, 227(1):40–8, May 1 1995.

[107] B. Maresca, A. M. Lambowitz, V. B. Kumar, G. A. Grant, G. S. Kobayashi, and G. Medoff. Role of cysteine in regulating morphogenesis and mitochondrial activity in the dimorphic fungus histoplasma capsulatum. *Proc Natl Acad Sci U S A*, 78(7):4596–600, July 1981.

[108] M. Sacco, B. Maresca, B. V. Kumar, G. S. Kobayashi, and G. Medoff. Temperature- and cyclic nucleotide-induced phase transitions of histoplasma capsulatum. *J Bacteriol*, 146(1):117–20, April 1981.

[109] V. Kumar, B. Maresca, M. Sacco, R. Goewert, G. S. Kobayashi, and G. Medoff. Purification and characterization of a cysteine dioxygenase from the yeast phase of histoplasma capsulatum. 22(4):762–8, Feb 15 1983.

[110] H. Li and G. Dryhurst. Irreversible inhibition of mitochondrial complex i by 7-(2-aminoethyl)-3,4-dihydro-5-hydroxy-2h-1,4-benzothiazine-3-carboxyli c acid (dhbt-1): a putative nigral endotoxin of relevance to parkinson's disease. *J Neurochem*, 69(4):1530–41, October 1997.

[111] J. N. Varghese. Development of neuraminidase inhibitors as anti-influenza virus drugs. *Drug Development Research*, 46:176–196, 1999.

[112] A. G. Tomasselli and R. L. Heinrikson. Targeting the hiv-protease in aids therapy: a current clinical perspective. *Biochim Biophys Acta*, 1477(1-2):189–214, Mar 7 2000.

[113] Kathryn Senior. Supercomputer-designed drug protects against chemotherapy toxcisity. *The Lancet*, 1:198, 2000.

[114] N. Fitch. Phd thesis at the university of birmingham.

[115] E.F. Fritsch J. Sambrook and T. Maniatin. *Molecular Cloning. A Laboratory Manual.* Cold Spring Harbor Laboratory Press, 2 edition, 1989.

[116] R. B. Parsons. Cysteine metabolism in the brain, liver and kidney.

[117] Incc. Nomadics. Spreeta evaluation module surface plasmon biosensor. Technical report, Naigproducts.com.

[118] V.P.Saxena and D.B.Wetlaufer. *PNAS*, 66:968, 1971.

[119] C.R.Cantor and P.R.Schimmel. *Biophysical Chemistry, Part 2 Techniques for the Study of Biological Structure and Function.* W.H.Freeman, 1980.

chemiluminescence are associated with changes in the thiol pool and lipid-soluble fluorescence in multiple sclerosis: the pathogenic role of
chemiluminescence are associated with changes in the thiol pool and