Biophysical Characterisation of Domain 1 of Rat CD2

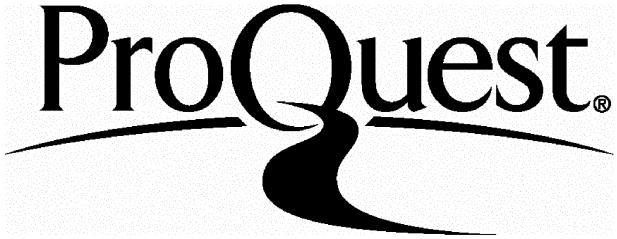
H. A Chen

Department of Biochemistry and Molecular Biology

University College

London

Thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy from the University of London



Abstract

Ho Ann Chen
University College London

Degree of Doctor of Philosophy
July 2000

CD2, a glycoprotein found on the T-lymphocyte, plays an important role in mediating the adhesion of T-lymphocytes to its accessory and target cells. The ligand-binding surface of CD2, which is located on the N-terminal domain of CD2 (CD2d1), has an unusually high proportion of charged residues. The ionic interactions of these charged residues are though to play a significant role in defining the ligand-specificity and binding affinity of CD2 with its ligands CD48 and CD58. The determination of the electrostatic properties of these proteins can therefore contribute to our understanding of the structure-activity relationships for these adhesion complexes.

In this thesis, the biophysical characterisation of the electrostatic properties of CD2d1 is described. The principal method used for the investigation is nuclear magnetic resonance (NMR) which permit the accurate determination of the ionisation constants of all the individual acidic residues. The characterisation of the binding interaction involved site-directed mutagenesis of these residues on the binding surface. In addition, the dynamic properties of CD2d1 are also investigated by NMR relaxation experiments. The significance of the finding are discussed

The pH titration of CD2d1 revealed a glutamate (Glu41) on the binding surface that has an anomalously high pK_a. This anomalous pK_a has an extensive effect on the chemical shift that suggests protein self-association mediated by this residue. This self-association was confirmed by relaxation analysis, and the CD2 was shown to dimerise with a very low affinity, but this dimerisation is nevertheless highly specific and has a pronounced effect on the relaxation parameters. The results indicated that CD2 dimerisation is maximal when Glu41 is protonated and Glu29 is deprotonated. The implication of the findings to the analysis of dynamics by NMR in discussed, and the significance of the observations to recent structural and functional analyses of rat CD2 interaction with CD48 is also examined.

Acknowledgements

I would like to first thank Dr. Paul C. Driscoll, my supervisor, for giving me the opportunity to pursue a scientific career at a time when it may have appeared to be the end of one. I am also indebted to his careful guidance and encouragement all through the years working in his laboratory, and grateful for his endless patience waiting for the long-promised completion of this thesis. On the subject of my thesis, I would also like to thank Dr Steve Matthews for his generosity in giving me the time to complete it when I should be working hard in his laboratory.

There are a number of people who have assisted me in my project that I must thank: first off is Dr Mark McAlister who helped to get me started in the wet lab, then Dr Ben Davis for guidance on matter NMR, and Dr Søren Kristensen for his scripts for the relaxation analysis. A big thank you must go to Dr Mark Pfuhl who contributed significantly in many aspects of my project, not least in helping me make some sense of the relaxation analysis. Other people who must be mentioned include Mark Domin at the School of Pharmacy who provided the mass spectroscopy analysis, Dr R Wallis at Oxford University who showed me how to use the AUC machine, Dr Simon Davis who provided a DNA construct, and Dr Sue Jones and Dr Roman Laskowski who provided assistance in bioinformatic analysis.

A few other individuals who helped to make my time studying for the PhD an enjoyable one should also be mentioned: Matt Bottomley and Paola Lo Surdo for their boundless good-humour and cheery disposition, Andrew Sankar as well Mark McAlister (again) and wife for their parties and outings, and all the members in the Paul Driscoll group who provided cheerful companionship. A special thank you is reserved for Geoff Kelly for the discussions, the beer and the opportunities to vent off steam when things get tough.

No word of thanks, however, could adequately express the profound gratitude I feel towards my mother for sticking by me all through the difficult years. Without her steadfast and quiet support I would not have been able to complete my PhD. This thesis is dedicated to her.

Contents

Chapter 2 - Methods – Molecular biology, protein expression and purification and analytical techniques

Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

Tables

Chapter2

Chapter3

Chapter5

Chapter6

Abbreviations

APC	antigen-presenting cell
ASA	accessible surface area
ATP	adenine triphosphate
AUC	analytical ultracentrifugation
CAM	cell adhesion molecule
CD	cluster of differentiation
CD2AP	CD2-associated protein
CD2BP	CD2-binding protein
CD2d1	CD2 domain 1
CHO	Chinese hampster ovary
COSY	correlation spectroscopy
CPMG	Carr-Purcell-Meiboom-Gill
CSA	chemical shift anisotropy
CT-HSQC	constant time heteronuclear single quantum coherence
CTLA	cytotoxic T-lymphocyte antigen
CW	continuous wave
DMSO	dimethylsulphoxide
DNA	deoxyribonucleic acid
DTT	dithiothreitol
EDTA	ethylenediamine tetraacetic acid
EEO	electroendosmosis
FID	free induction decay
FT	Fourier transform
GlyCAM	glycosylation-dependent cell adhesion molecule
GST	glutathione S-transferase
HIV	human immunodeficiency virus
HLA	human leukocyte antigen
HSQC	heteronuclear single quantum coherence
HTLV-1	human T lymphocyte virus type-1
ICAM	intercellular cell adhesion molecule
IEF	isoelectric focusing
IgSF	immunoglobulin superfamily
IL	interleukin
INEPT	insensitive nuclei enhancement by polarisation transfer
IPTG	isopropyl-thiogalactoside
ITAM	immunoreceptor tyrosine-based activation motif
IR	infra red
LAT	linker for activation of T-cell
LB	Luria-Bertani
LFA	leucocyte function antigen
LMP	low-melting point
MALDI-TOF	matrix-assisted desorption ionisation-time of flight
MAPK	mitogen-activated protein kinase
MHC	major histocompatibility complex
mAB	monoclonal antibody

mRNA	messenger ribonucleic acid
NCAM	neural cell adhesion molecule
NK	natural killer
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
NOESY	nuclear Overhauser effect spectroscopy
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PFG	pulsed field gradients
PFGLED	pulsed field gradients longitudinal encode-decode
PMSF	phenylmethyl sulphonyl fluoride
pH	potenz hydrogen
PTK	protein tyrosine kinase
RF	radio-frequency
RBS	ribosome binding site
RNA	ribonucleic acid
SAR	structure-activity relationship
SDM	site-directed mutagenesis
SEC	size exclusion chromatography
SDS	sodium dedocyl sulphate
SHC	Src homology collagen-like
SLAM	signalling lymphocyte activation molecule
SLP	Sh2 domain-containing leukocyte phosphoprotein
SMAC	supramolecular activation cluster
SPI	selective population inversion
SW	spectral width
T_C	cytotoxic T-cell
T_H	helper T-cell
TCR	T-cell receptor
TOCSY	total correlation spectroscopy
tRNA	transfer ribonucleic acid
TROSY	transverse relaxation optimised spectroscopy
UV	ultra-violet
ZAP	ζ-associated protein

Chapter 1

CD2 - the T-cell antigen

CD2, a cell adhesion molecule, is found ubiquitously in all T-lymphocytes, the majority of NK cells, and all except the most immature thymocytes (Springer *et al.*, 1987). It is absent in all other cells, with the exception of B-lymphocytes in mice, splenic macrophages in rat and sheep, and a subset of B-lymphocytes in pigs (Barclay *et al.*, 1993; Sinkora *et al.*, 1998). CD2 was also known as leukocyte function antigen-2 (LFA-2), as well as T11/Leu-5/OKT1/Tp50 and OX-34 T-cell surface antigen. The cluster of differentiation (CD) designation, originally designed for the identification of leukocyte membrane proteins or protein complexes whose functions have yet to be identified in human, is based on the interactions with specific monoclonal antibodies that has become a useful and systematic classification of cell surface antigens in human as well as other species (Barclay *et al.*, 1993).

The appearance of CD2 begins early in thymic ontogeny (Sen *et al.*, 1989; Yagita *et al.*, 1989). The progenitors of T-lymphocytes are produced in the bone marrow or foetal liver; from there they migrate to the thymus where they can develop and mature. These progenitor cells are largely devoid of most of the cell-surface molecules that characterised mature T-lymphocytes. The expression of CD2 on the

cell-surface can be detected within 13-15 days of gestation, and the level of expression then increases until it reaches the amount that may be found on resting T-cells of about 14,000 molecules. From this point onwards, CD2 remains expressed throughout the entire lifespans of all the cells and their progenies in the T-cell lineage. The expression level of CD2 can be further increased upon exposure of T-cells to antigen (Abraham *et al.*, 1994), and up to 250,000 CD2 molecules may be found in these activated T-cells. The high level of expression also persists in memory T-cells (Springer *et al.*, 1987).

The ubiquity of CD2 on T-cells and its absence in nearly all other cells make it a useful marker for the identification for T-cells. Indeed, in the early days, the T-cells were identified as well as purified by using the ability of T-cells to form "rosettes" or aggregations with sheep erythrocytes (Brain *et al.*, 1970; Jondal *et al.*, 1972; Lay *et al.*, 1971). This rosetting property of T-cells has since been shown to be mediated by CD2. There are many other cell surface antigens that are found almost exclusively on T-cells, including CD6, CD7 and CDw121a as well as the T-cell receptor (TCR). While the TCR, together with its associated CD3 (the signalling component of TCR complex), can undoubtedly be considered the definitive T-cell molecules, the TCR is structurally variable and the CD3 complex does not appear until relatively late in T-cell ontogeny. The CD2 may therefore be more useful as a general identifying marker throughout the T-cell development; in this, and in view of its historical role, the CD2 can justifiably lay claim as the T-cell antigen.

Biological context

CD2 is one of best characterised of all the molecules involved in cell adhesion and cell-cell recognition (reviewed by Bierer *et al.*, 1989; Davis *et al.*, 1998b; Davis & van der Merwe, 1996; Moingeon *et al.*, 1989a; Springer, 1990; Springer *et al.*, 1987). The CD2 and its ligand CD58 are the first cell surface molecules shown to be involved in heterophilic adhesion. It mediates the adhesion interactions of T-lymphocytes with the antigen-presenting cells (APC) (Shaw *et al.*, 1986), such as those between helper T (T_H) lymphocytes and B lymphocytes, and

between cytotoxic T (T_C) lymphocytes and target cells. It can also mediate the interactions between thymocytes and thymic epithelial cells (Vollger *et al.*, 1987). It stabilises the cell/cell contact (Moingeon *et al.*, 1989b) and may facilitate the dynamic binding interaction of TCR with the peptide-major histocompatibility complex (MHC) (Dustin *et al.*, 1998; van der Merwe *et al.*, 1995). It contributes to the antigen-specific response of both T_C and T_H lymphocytes (Springer, 1990) as well as facilitating the cytotoxic function of NK cells (Moingeon *et al.*, 1992; Siliciano *et al.*, 1985). It can reverse T-cell anergy (Boussiotis *et al.*, 1994) and may also play a role in the thymic ontogeny of T-cells. Its importance in the immune response can be seen in transplantation studies in which inhibition of CD2 by monoclonal antibodies (mAbs) or modulation of CD2 co-stimulatory pathway can significantly promote the survival of allografts (Qin *et al.*, 1994; Sultan *et al.*, 1997).

However, despite being one of most intensely studied molecules on the T-cells, the precise function of CD2 is still the subject of some debate (Davis & van der Merwe, 1996). CD2 may be involved in the transmembrane signalling during T-cell activation (Bierer *et al.*, 1989; Moingeon *et al.*, 1989a; Sunder-Plassmann & Reinherz, 1998), however, some degree of uncertainty surround this signalling role and its significance in the functioning of T-cell. This and other aspects of the roles played by CD2 in the immune response will be described in greater detail later.

Cell-Cell Adhesion

The adhesion of leukocytes to other cells types is of immense importance in the immune response (Klein & Horejsi 1997; Springer, 1990). For example, the leukocytes must attach to stromal cells in the bone marrow and in the thymus in order to receive proper differentiation signals, to endothelial cells as well as to the extracellular matrix in order to enter and navigate around in tissues, and to adhere to other cells such as APCs, macrophages and B cells for it to survey for foreign antigens and elicit an immune response. These cell adhesion events are mediated by pairs of adhesion molecules and a wide variety of cell adhesion molecules can be found on the cell surface of leukocytes (Barclay *et al.*, 1993). Five principal families of cell adhesion molecules (CAMs) may be found in mammalian cells – integrins, cadherins, selectins, mucins and the immunoglobulin superfamily (IgSF) to which

CD2 belongs (van der Merwe & Barclay, 1994). The most abundant of these are the immunoglobulins, with ~40% of the cell surface molecules containing at least one immunoglobulin domain. These CAMs mediate the recognition between cell types, direct the circulation and migration of lymphocytes, and in many cases, can also elicit stimulatory or negative signal in the cells that affect the activity of cells.

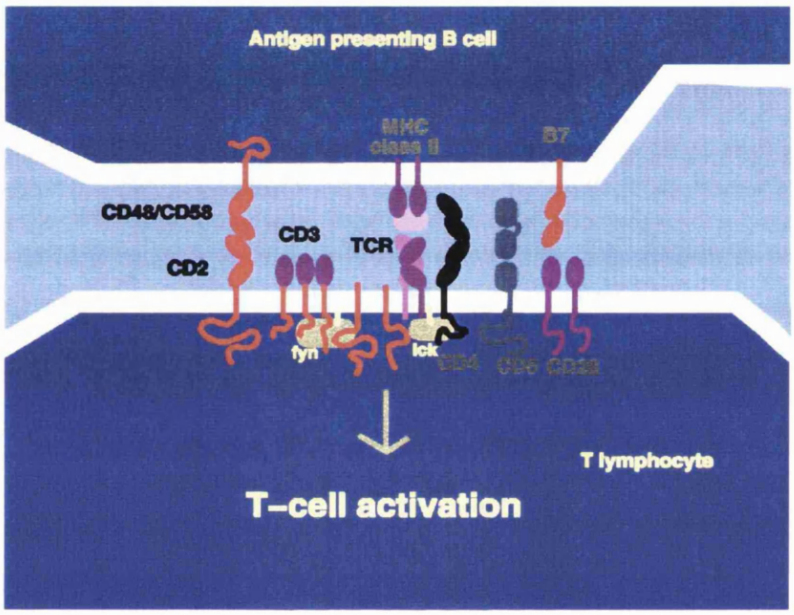


Fig 1.1 Schematic diagram of some of the adhesion molecules present on the adhesion contact zone between T-cell and antigen presenting cell.

In T-cell and APC, the adhesion molecules include, among many others, the β2 integrin LFA-1 (CD11a/CD18), CD2, CD5 CD6, CTLA-4 (cytotoxic T lymphocyte antigen-4) and counter-receptors of these molecules (Fig. 1.1). The TCR, together with the MHC, CD4 and CD8, could also be considered as adhesion molecules as they contribute to specific contact between T cells and APCs. While some cell/cell adhesion interactions can be strong, stable and long-lasting, such as those mediated by cadherins between cells in organs and epithelium, for leukocytes, such long-lasting and stable interactions are largely absent (Barclay *et al.*, 1993). Their adhesion interactions may be generally weaker, transient and reversible (van der Merwe & Barclay, 1994). For example, the rolling of the leukocytes along the endothelium as mediated by selectins requires very weak interaction so that the bonds can be broken very easily (Hogg & Berlin, 1995). Stronger associations, however,

may be provided by the adhesion interactions of a number of molecules such as LFA-1 which binds intercellular CAM (ICAM).

The interactions of these molecules are well-controlled, with the adhesion regulated by changes in the types and amount of adhesion molecules expressed, as well as by changes in distribution and avidity of the receptors (Abraham *et al.*, 1994). Different sets of CAMs have been identified to be involved at different stages of cell adhesion and in different immune response. For example, during inflammation, the expression of ICAM-I (CD54), a ligand for LFA-1 normally found only on B and dendritic cells, is induced in the capillary endothelial cells in inflamed area, thereby allowing lymphocytes and monocytes to adhere and move into inflamed tissues. The differences in expression of the types and levels of CAMs also depend on the activation states of the T-cells. The level of expression of CD2, CD27 and CD44, for instance, may be up-regulated on T-cell activation. The binding interaction of CD2 or LFA-1 can also further regulate the expression of CD2 as well as the activity of other adhesion molecules (Kivens *et al.*, 1998; Lin *et al.*, 1995; Porter & Hogg, 1997; Vankooyk *et al.*, 1989). The regulation of the CAMs also extends to their binding affinity. For example, LFA-1 as well as human CD2 can increase their avidity for ligands upon the triggering of the TCR in a phenomenon known as 'inside-out' signalling in which the intracellular signals are manifested on the cell-surface (Springer, 1990). Such changes in the avidity and expression level of adhesion molecules are vital to the functioning of the leukocytes. For example, in T-cells, a mobile circulatory state is necessary for efficient immune surveillance, but upon antigen recognition, the cells will need to switch to a more adherent and motile state requiring different interactions from their CAMs.

The amount of CAMs on the cell surface may be down-regulated by reducing the expression level (Haque *et al.*, 1998), or by a process called 'shedding' whereby undesired adhesion receptors are detached from cell surface by proteases (Springer, 1990). The shedded receptors can also inhibit cell adhesion by competing with counter-receptors. Other anti-adhesion forces such as those from the negatively-charged polysialic acid and sialic acid on glycoproteins and glycolipids may also interfere with the cell adhesion (Bell *et al.*, 1984; Rutishauser *et al.*, 1988).

T-cell Activation and the immune response

The immune system is a highly complex system involving many cell types and complex interactions between large number of molecules. The T-cells, together with macrophages, are the principal cell types involved in a category of immune responses called cell-mediated immunity (Klein & Horejsi, 1997). Such cell-mediated processes may include graft rejection, hypersensitivity reactions and defence against malignant cells and many viruses. The T-cells can kill their target cells, secrete cytokines to provide accessory signals for B-cells to promote the antibodies response, as well as enhancing the functions of other cells such as macrophages.

The T-cell initiates its immune defence response when it detects foreign antigens that are presented as small peptides on the surface of antigen-presenting cells (APCs). These APCs can be virtually all cell types; some, such as the macrophages and B-lymphocytes, are specialised APCs whose interactions with T-cells are of particular significance. The foreign peptides are displayed on APC cell surface by the MHCs to be recognised by the TCRs on T-cells.

The TCR is a heterodimeric molecule that is composed of either α and β, or γ and δ polypeptide chains. It is always associated with CD3, a complex constituted from different transmembrane surface polypeptides, the γ, δ, ε, and η chains, as well as the ζ chain. The TCR forms the recognition unit of the receptor while CD3/ζ chain provides the signalling function. The $\alpha\beta$ TCR is also coexpressed with their co-receptors CD4 or CD8 molecules which help the T cells to bind to APCs and determine the class of MHC molecules recognised by T cells. Some CD2 have been shown to associate with the ζ chain of the TCR/CD3 complex (Brown *et al.*, 1989).

The T-cell activation is triggered by the ligation of the TCR with the peptide-MHC molecule assemblies on the surface of the APC. The earliest event of T-cell activation is the increase in cytoplasmic protein tyrosine kinase (PTK) activity which occurs within seconds of stimulation. The PTKs implicated in TCR signalling include $p59^{fyn}$, which is associated with cytoplasmic domain of CD3 but may also become associated with CD2 (Gassmann *et al.*, 1994; Lin *et al.*, 1998); $p56^{lck}$ which binds to CD4 or CD8 and possibly CD2 (Bell *et al.*, 1996); and ζ-associated protein

(ZAP)-70 which become associated with the ζ-chain following the phosphorylation of immunoreceptor tyrosine-based activation motifs (ITAMs), the signalling motifs of TCR/CD3 (MarieCardine *et al.*, 1996). The phosphorylation by these PTKs triggers a variety of downstream events such as the activation of phospholipase C-γ1, stimulation of fluxes of cytoplasmic calcium and the activation of Ras and the MAPK (mitogen-activated protein kinase) cascade (Ueno *et al*, 2000; and reviewed in Cantrell, 1996; van Leeuwen & Samelson, 1999). The cascade of biochemical signalling events eventually culminates in T-cell activation and proliferation (Samelson *et al.*, 1990). An increase in mRNA synthesis can be observed to occur within the first hour and cell proliferation can begin in ~24 hours with the activated T cells undergoing a sequence of phenotypic changes and differentiation into effector and memory cells.

However, the signals transmitted from the TCR alone may not be sufficient for such cell proliferation, and the full T-cells activation may require at least an additional co-stimulatory signal provided by the ligation of other accessory receptors (Bierer & Burakoff, 1989). These accessory receptors can further enhance the adhesion between the T-cell and APC as well as amplifying the signal transmitted by TCR such that a cellular response can be achieved at low number of triggered TCR (Lanzavecchia *et al.*, 1999). In the absence of certain co-stimulatory signals, the T-cell can enter into a state of anergy in which it is refractory to stimulation by antigen (Schwartz, 1990). One of the most important co-stimulatory signals is provided by the adhesion protein CD28 when bound to the B7 family of cell surface molecules, CD80 and CD86 (Chambers & Allison, 1999; June *et al.*, 1994; Linsley & Ledbetter, 1993). Alternative co-stimulation signals may also be supplied, amongst others, by the LFA-1 interaction with CD54, SLAM (signalling lymphocyte activation molecule), and possibly CD2 interaction to its counter-receptors (Bierer *et al.*, 1988; Watts & DeBenedette, 1999). Some interactions of these receptors, however, only lead to partial activation, and other interactions may lead to anergy of T cell or negative regulation of the T-cell activation, while others may reverse the anergy of the T cells. The signalling events in T-cell activation are therefore the product of complex interactions between large number of different molecules.

The activation of T-cell also requires the sustained TCR engagement and signalling that last for many minutes or hours. This involves the polarised

reorganisation of the cytoskeleton and the aggregation of the receptors in a process known as 'capping' (Dustin *et al.*, 1998; Shaw & Dustin, 1997). The aggregation of the receptors and activation of the T-cells can also be induced with antigens and mitogens that can cross-link the different receptors. During this aggregation events, the receptors cluster at one pole of the cell surface at the point of contact with APC, creating a zone of specialized cell-cell adhesion called the immunological synapse or supramolecular activation clusters (SMACs) (Dustin *et al.*, 1998; Grakoui *et al.*, 1999; Monks *et al.*, 1998). Such clustering of receptors at the cell-cell interface would results in the accumulation of tyrosine kinases that are associated with or recruited by these receptors. For example, the recruitment and clustering of CD2 and CD28 with the TCR onto the T cell contact may bring together $p56^{lck}$ and $p59^{fyn}$ with their respective activating ligands. Many of the signalling components such as Lck and LAT (linker for activation of T-cells) are concentrated in 'rafts', the membrane microdomains that are enriched in cholesterol, and the triggered TCRs are recruited into these kinase-rich environment of rafts (Simons & Ikonen, 1997; Yashiro-Ohtani *et al*; 2000). The receptors were also segregated according to size, therefore CD45 which is larger and has phosphatase activity is excluded from the central cluster of receptors (Leupin *et al*; 2000). Such segregation may have important implications for the regulation of phosphorylation events during T-cell activation. The adhesion and aggregation event of CAMs during T-cell activation will be described in greater details later.

CD2 - a signalling molecule?

Much interest has been focused on the possible roles played by CD2 as a signalling molecule (for reviews, see Bierer & Burakoff, 1989; Davis & van der Merwe, 1996; Dustin & Springer, 1991; Moingeon *et al.*, 1989a). CD2 is suggested to generate specific co-stimulatory signal required for T-cell activation as modulation of the cell-surface expression of CD2 and its ligand CD58 can regulate the antigen-specific responses of various subsets of T-cells (Dustin & Springer, 1991). Its interaction can also reverse T-cell anergy (Boussiotis *et al.*, 1994). The CD2 may associate with the TCR/CD3 complex together with CD4/CD8 and CD5 (Beyers *et al.*, 1992; Carmo *et al.*, 1999), and may be involved in bi-directional costimulatory

signals during T-T and T-B cells signalling. In addition, since CD2 appears early in T cell ontogeny, its stimulation may be crucial for the development of cells that have not fully develop their TCR-CD3 complex.

Several lines of evidence exist that suggest a signalling role by CD2. Anti-CD2 monoclonal antibodies (mAbs) and soluble CD58 have been demonstrated to co-stimulate T-cell proliferation in concert with anti-TCR/CD3 mAb stimulation. (Bierer & Burakoff, 1989; Clark *et al.*, 1988). Furthermore, combination of anti-CD2 mAbs by themselves can induce proliferation of resting human and rat T-cells (Beyers *et al.*, 1989; Meuer *et al.*, 1984a). For optimal mAb-mediated activation, two antibodies are necessary, one directed against conventional CD2 epitopes, and the second against an activation-associated neo-epitope of CD2, termed the CD2R which has been mapped to the flexible linker region between the two extracellular domains (Li *et al.*, 1996).

The stimulation of human T cells through the CD2 can lead to tyrosine kinase activation, expression of interluekin-2 (IL-2) receptor and synthesis of IL-2 and interferon-γ (IFN-γ) (Targan *et al.*, 1995). CD2 may also provide costimulatory signal for the production of IL-4 and IL5 by memory T-cell (Peng *et al.*, 1997). CD2 has been shown to regulate the responsiveness of activated T-cells to IL-12 as anti-CD2 and anti-CD58 mAbs can specifically inhibit IL-12 induced proliferation and IFN-γ production by activated T-cells (Gollob *et al.*, 1995; Gollob & Ritz, 1996). CD2 can also participate in bi-directional signalling by regulating the induction of monocyte IL-1β mRNA (McAllister & Ellis, 1996).

The signalling role of CD2 may be indicated by the presence of several proline-rich SH3 binding regions in its cytoplasmic domain (Chang *et al.*, 1990; Nishizawa *et al.*, 1998; Ren *et al.*, 1993). These regions may be bound by proteins such as the Src tyrosine kinase $p59^{fyn}$ and $p56^{lck}$ (Bell *et al.*, 1996; Gassmann *et al.*, 1994) as well as several adaptor proteins (Dustin *et al.*, 1998; Li *et al.*, 1998; Nishizawa *et al.*, 1998). The presence of this cytoplasmic domain is required for optimal CD2-mediated augmentation of antigen-triggered T-cell responses and IL-2 production, and it may also regulate CD2 avidity (Chang *et al.*, 1990; Hahn *et al.*, 1992b; He *et al.*, 1988). Deletions of certain conserved C-terminal sequences result in only partial signal transduction (Bierer *et al.*, 1990; Moingeon *et al.*, 1989b).

There is therefore considerable evidence that suggests possible CD2 involvement in signalling, however, the precise nature as well as the significance of this signalling role is still the subject of some controversy. It has been suggested that CD2 and CD3 (i.e. TCR signalling) signalling share the same pathway in the earliest signalling events (Ley *et al.*, 1991; Yang *et al.*, 1986). For example, both CD2 and CD3 mediated signalling can produce phosphorylation on the same set of molecules such as phospholipase Cγ-1 (Kanner *et al.*, 1992), and CD2 can also enhance the CD3-mediated phosphorylation of Cbl and Shc (Umehara *et al.*, 1999). Evidence, however, suggests that CD2 signalling may proceed along quite distinct pathways from TCR/CD3, with signalling pathways that may be antigen-dependent or antigen-independent (Sen *et al.*, 1992). Although CD2, CD3 and CD28 activate many of the same signalling molecules, they differ in their capacity to induce tyrosine phosphorylation of proteins (Hutchcroft *et al.*, 1998). The phosphorylation sites may also be different in CD2 and CD3 activated responses (leGouvello *et al.*, 1998) and CD2 has been shown to trigger the phosphorylation of a unique 72 KDa substrate not seen in other modes of activation (Targan *et al.*, 1995). Secretion of cytokines in T_H cell by TCR/CD3-mediated and CD2-mediated signal also appears to involve different pathways (Semnani *et al.*, 1998). Engagement of CD3 but not CD2 can induce the marked tyrosine-phosphorylation of the ZAP-70 (Howe & Weiss, 1995), while signalling via CD2 but not CD3 can enhance human T lymphotropic virus type-1 (HTLV-1) replication (Guyot *et al.*, 1997). Engagement of CD2 and/or CD28 can also differentially activate protein tyrosine phosphatases. CD2 has been shown to associate with CD45 which has phosphatase activity (Schraven *et al.*, 1990) and forms functional complex involving CD45 and $p56^{lck}$ (MarieCardine *et al.*, 1996). It may also associate with CD5 which is in turn associated with tyrosine phosphatase SHP-1 (Carmo *et al.*, 1999). CD2-mediated signalling, in contrast to TCR-mediated signalling, may proceed along a pathway that is independent of the $p56^{lck}$ and occur instead via Jun kinase (Sunder-Plassmann & Reinherz, 1998). CD2-dependent pathway has been shown, according to the activation state of peripheral blood T cells, to direct them either toward proliferation or toward apoptosis (Dumont *et al.*, 1998).

The significance of the signalling role of CD2, however, has been questioned (Davis & van der Merwe, 1996). Activation by anti-CD2 mAbs need not suggest a signalling role as antibodies against a substantial number of cell-surface molecules

can activates the T cells (Barclay *et al.*, 1993). The activation by CD2 requires the coexpression of TCR on T-cells (Alcover *et al.*, 1988), or CD16/ζ or Fcε receptor γ complex on NK cells (Spruyt *et al.*, 1991). CD2 therefore cannot transduce signal on its own and the signalling effects are suggested to be an indirect result of the aggregation with TCR and other receptors (Davis & van der Merwe, 1996). Moreover, anti-CD2, while capable of augmenting initial T-cell proliferation, fails to prevent T-cells from undergoing apoptosis, suggesting that CD2, in contrast to CD28, does not provide complete co-stimulatory signal (Yashiro *et al.*, 1998).

More disturbingly, studies using 'knock-out' mice showed only mild abnormalities in T cell development with no apparent ill-effect in CD2-deficient mice (Killeen *et al.*, 1992), suggesting that CD2 may not be essential. However, studies on these CD2-deficient mice also showed that interaction of CD2 with CD48 influences both positive selection and T cell activation (Teh *et al.*, 1997). Further studies on CD2-deficient mice indicate that CD2 can enhanced T-cell activation at low antigen concentration but is unnecessary at high antigen concentration (Bachmann *et al.*, 1999). It is therefore argued that CD2 functions primarily as an accessory molecule that can mediate adhesion and facilitate TCR interaction, thereby lowering the threshold necessary for T-cell activation, rather than as a co-stimulatory molecule which supplies signals that augment TCR-mediated signalling (Bachmann *et al.*, 1999). It is however possible that the role played by CD2 has become less important in mice only recently in evolution, and the studies on mice therefore cannot be extrapolated to human. Human CD2 has a higher affinity for its ligand CD58, and CD58 is also much more widely expressed than mouse CD48; furthermore, mouse T-cells are less dependent than human T cells on CD2-ligand interactions (van der Merwe, 1999). It is also likely that there are other molecules that play similar role or have similar properties to CD2 (Dustin *et al.*, 1998), and recent research suggests that CD2 and CD28 indeed share some functions in facilitating interaction between T-cell and the APC (Green *et al*; 2000). The absence of CD2 may also be compensated in T-cell, albeit imperfectly, by the selection of TCR with higher affinity for antigen (Killeen *et al.*, 1992). The role of CD2 is therefore postulated to be for increasing the size of the repertoire of immune response (Davis & van der Merwe, 1996).

CD2 adhesion in T cell activation

A number of cell adhesion molecules are involved in the T-cell activation and these adhesion interactions help stabilise the TCR binding and increase the strength of the stimulus transmitted to the cells (Collins *et al.*, 1994). While the complexity of adhesion interactions during T-cell activation preclude a clear understanding of such events at present, enough evidence has accumulated to suggest a probable sequence of events.

A number of binding interactions, including that of CD2 to its counter-receptors, may precede the TCR-peptide-MHC binding interaction. The initial long-range contact between T-cell and APC is likely to be mediated by larger molecules such as LFA-1 (Brown & Shaw, 1999). Such initial contact by LFA-1 has been demonstrated using real-time imaging of T-cell in contact with glass-supported lipid bilayer containing ICAM-1 and MHC (Grakoui *et al.*, 1999). This contact region, consisting of cluster of LFA-1 molecules, may facilitate the low-affinity and close-range interactions of TCRs with their peptide-MHC ligands as well as those of other smaller adhesion molecules. Binding by CD4 may then stop the migration of T-cell. The closer-range interactions occurs at the outer rim of the initial contact region, and over the next five minutes following the initial engagement and TCR triggering, the engaged MHC-peptide moved to the centre of the junction to form a central cluster (Grakoui *et al.*, 1999). The receptors are segregated at the contact 'capping' region according to size (Dustin *et al.*, 1998; Monks *et al.*, 1998), in which TCR and CD28 are redistributed into the central zone, while larger molecules such as LFA-1 are redistributed to the outer rims of the contact zone. Interestingly, CD2 is segregated into a distinct "inner" adhesion ring surrounding the central cluster (Leupin *et al*; 2000). It may therefore function as a filter that prevent molecules larger than the dimension of ligand-bound CD2 being transported into the central cluster, creating a gasket-like tight seal between the opposing membranes at the immunological synapse. When a minimum threshold density of peptide-MHC complex is reached within the mature synapse, maximum proliferation of T-cells is triggered.

Surprisingly, the experimental data from Dustin and co-workers suggest that CD2 or CD28 may not be necessary for the TCR interaction and segregation of receptors (Grakoui *et al.*, 1999). This is unexpected as the dimension of the

extracellular portion of CD2 is thought to be particularly important in TCR interaction (Davis & van der Merwe, 1996; Wild *et al.*, 1999a). The dimension of membrane separation on CD2 interaction with its ligands, as extrapolated from the dimeric crystal structures (~14-21 nm between membrane), is similar to that of TCR-MHC complex (~15 nm) (Dustin *et al.*, 1998; Garboczi *et al.*, 1996; Garcia *et al.*, 1996; van der Merwe *et al.*, 1995). CD2 binding to its ligands can therefore separate the cells at the T-cell contact surface at a distance that is optimal for TCR-peptide-MHC interaction. It may also exclude larger molecules such as integrins, CD43 and CD45 that may interfere with such interaction. It should be noted, however, that the experiments were performed on a glass-supported planar lipid bilayer, and adhesion pattern may be different in the uneven, crenulated and highly deformable surface of APC (Malissen, 1999). Experiments with CD2/CD58 and LFA-1/ICAM interactions revealed that low affinity interactions are strongly dependent on membrane alignment (Dustin, 1998; Dustin *et al.*, 1997a), and such alignment may be easier to achieved on a glass-supported planar lipid bilayer than the uneven surface of APC.

It has been demonstrated that T cells can detect very low level of specific peptide-MHC ligands on cells (Demotz *et al.*, 1990; Valitutti *et al.*, 1995); moreover, a substantial proportion of TCRs appeared to be ligated by this small number of peptide-MHC (Valitutti *et al.*, 1995). One interpretation of these observations is that the contact zone is highly dynamic, and many peptide-MHC and TCR molecules need to diffuse rapidly in and out of the zone of contact between a T cell and APC in order for the TCR to encounter a specific peptide-MHC complex. Such rapid association and reforming of receptors at the contact area has been demonstrated for CD2 interaction with CD58 (Dustin *et al.*, 1996), and the rapid diffusion of the TCR may be facilitated by the very fast binding kinetics of CD2. The serial engagement of the TCR by the peptide-MHC complex has also been suggested to occur during the movement of the TCR from the outer ring into the central cluster during the formation of the immunological synapse (Grakoui *et al.*, 1999).

TCR has only a low binding affinity for peptide-MHC complex with a high off-rate (Corr *et al.*, 1994; Matsui *et al.*, 1994). However, if the TCR recognises its antigen, the TCR-mediated signals lead to changes that further stabilise the conjugate formation between the T cell and the APC. As mentioned earlier, the expression of CD2 and LFA-1 can be increased on TCR activation which can enhance adhesion

between the cells (Koyasu *et al.*, 1990). The coligation of CD2 with ligands on opposing cells at the contact surface can also recruit other CD2 molecules to this region of cell-cell contact thereby further increasing the adhesion strength (Abraham *et al.*, 1994). The movement of the receptors into the region of contact may be mediated by lateral mobility of the adhesion molecules (Dustin *et al.*, 1997b), and this redistribution of CD2 can occur without the presence of its cytoplasmic tail and is also independent of cytoskeletal rearrangement event (Koyasu *et al.*, 1990; Li *et al.*, 1996). While there is no absolute requirement for the CD2 cytoplasmic tail in CD2-mediated cell adhesion and ligand-induced CD2 reorganisation, the presence of cytoplasmic tail is nevertheless necessary for mediating the central patterning of CD2 at the immunological synapse and effecting the T-cell polarisation by its association with adaptor proteins such as CD2-associated protein (CD2AP) (Dustin *et al.*, 1998), and CD2-binding protein (CD2BP1) (Li *et al.*, 1998). The adaptor proteins are molecules that possess no intrinsic enzymatic function, but can mediate protein-protein interactions critical for regulating signalling events following TCR activation (Peterson *et al.*, 1998). Through the recruitment of these cytosolic adaptor proteins, the CD2 may be attached to the cytoskeleton, and the establishment of the cytoskeletal polarisation allows further regulation of the adhesion (Dustin *et al.*, 1998; Nishizawa *et al.*, 1998). The cytoplasmic domain is also bound by other adaptor proteins such as CD2BP2 which may play a role in mediating signalling (Freund *et al.*, 1999; Nishizawa *et al.*, 1998). Many adaptor proteins such as Grb2 (growth factor receptor binding protein 2), LAT, SH2 domain-containing leukocyte phosphoprotein (SLP)-76, Shc and Cbl have been identified to be important in T-cell activation (reviewed by Peterson *et al.*, 1998).

Within minutes of TCR activation, the CD2 in this contact zone also showed a marked increased in avidity in its binding interaction (Hahn *et al.*, 1992b). However, such increase in avidity is not due to an increase in CD2 cell-surface expression which can take hours, and it can also be induced by incubating T cell with phorbol ester or anti-CD3 mAb. This change in CD2 avidity coincides with the emergence of neo-epitope after TCR activation, leading to a suggestion of conformational changes (Meuer *et al.*, 1984b). However, any such changes is likely to be limited to changes in the hinge region between the domains as conformation changes are limited to loops and shifts in quaternary organisations in other molecules

Alternatively, the change in avidity has been suggested to be due to the effect of cross-linking and aggregation of the receptors (Li *et al.*, 1996), or the attachment of the receptors to the cytoskeletons (Davis *et al.*, 1998b). The TCR activation also stimulates the integrin-mediated adhesion; this increase in cell-cell interaction may then further reinforce the CD2-mediated adhesion.

Structure and binding analysis of CD2

Structure of CD2

CD2 is a 50-60 KDa glycoprotein comprising of two extracellular immunoglobulin superfamily (IgSF) domains and an intracellular C-terminal domain linked by a putative transmembrane helix. The structure of the CD2 extracellular portion (Fig. 1.2), the ectodomain, has been determined by X-ray crystallography (Bodian *et al.*, 1994; Jones *et al.*, 1992), while the solution structures of N-terminal domain of CD2 (CD2d1) has been determined by NMR spectroscopy (Driscoll *et al.*, 1991; Withka *et al.*, 1993). The N-terminal domain of CD2 (CD2d1), the focus of this investigation, was demonstrated to be solely responsible for the adhesion function CD2 (Hahn *et al.*, 1992a; Peterson & Seed, 1987; Sayre *et al.*, 1989).

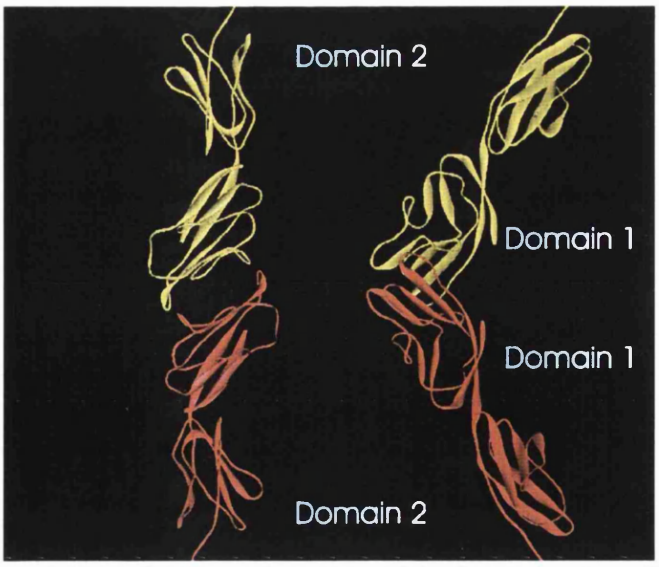


Fig. 1.2 Two orthogonal views of the crystal structure of the 2-domain extracellular portion of rat CD2 (PDB ID: 1HNG) in a 'head-to-head' dimeric orientation. Each monomer is indicated by a different colour.

CD2 ectodomain - From the structures of CD2 as determined by X-ray crystallography and NMR spectroscopy, both extracellular domains are shown to be of IgSF fold - the N-terminal domain CD2d1 is of V-set fold while the membrane-proximal domain is of C2-set. The V-type fold of domain 1 (CD2d1) consists of 9 stranded β-sheet, with five antiparallel strand GFCC'C'' forming one sheet and four antiparallel strands forming the ABED sheet. The A strand, however, can also be considered to be shared by both sheets. The intrachain disulphide bond that normally holds together the two sheets and stabilises the structure in V-set immunoglobulin is absent in CD2d1. Two cysteines, however, can be inserted at the predicted positions (I18 and V78) to form a disulphide bond (Gray *et al.*, 1993). There are differences in CD2 to the standard V-type fold in its loop configuration: its DE loop is very short, while the BC loop is longer. The β-strands B, D and E are also truncated (Jones *et al.*, 1992), with the short B strand being due to the presence of Pro19 which result in a kink at the beginning of the BC loop.

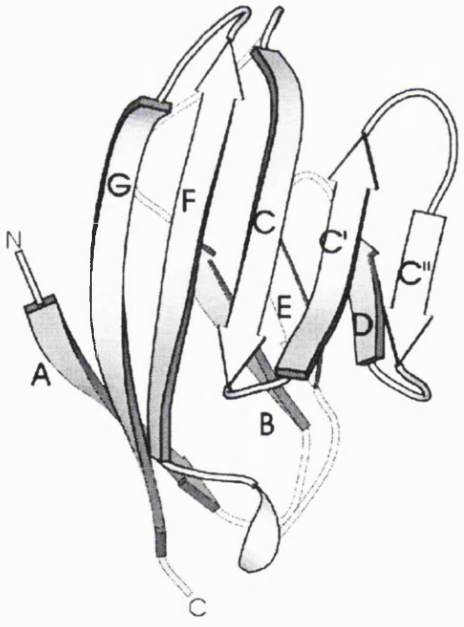


Fig. 1.3 Ribbon diagram of the N-terminal domain of CD2.

The C2 fold of the second domain is consisted of 7 stranded β-sheet. This second domain of CD2 may be involved in CD2 function by interacting with adjacent molecules. The domain 1 and 2 are linked by a highly conserved flexible linker region which allow up to 20° orientational freedom for domain 1 relative to domain

2. This may be important in the binding interaction by allowing the CD2d1 to be orientated such that the GFCC'C'' face form a platform nearly parallel to the cell surface. The linker may therefore enhance binding by maximising exposure. The stalk linking the extracellular portion to the membrane, for unknown reason, is also highly conserved.

There are four Asn-linked glycans in rat CD2, with three on the N-terminal domain CD2d1 (Jones *et al.*, 1992; Williams & Barclay, 1988). The adhesion surface, however, is free from glycosylation site and the binding of rat CD2 to CD48 is independent of glycosylation (Davis *et al.*, 1995a; van der Merwe *et al.*, 1993a; van der Merwe *et al.*, 1993b). It was previously thought that the glycosylation found on human CD2d1 may contribute to ligand binding (Recny *et al.*, 1992), however, structural studies have shown that this N-linked glycan is located opposite to the ligand binding site and therefore not directly involved in binding (Wyss *et al.*, 1995). It was also suggested that the N-glycan is crucial for stabilising the protein fold and counterbalancing an unfavourable clustering of positive charges (Wyss *et al.*, 1995), however, this interpretation is disputed (Davis & van der Merwe, 1996). The N-glycan attached to the membrane proximal domain was proposed to have a function in maintaining the upright position of the extracellular portion of CD2, orientating the binding surface of CD2 in such a manner as to promote *trans*-interaction of CD2 with CD48 instead of *cis*-interaction (Dustin *et al.*, 1996). The glycans may also prevent non-specific protein-protein interaction during receptor clustering, protect against proteases and play a general role in controlling the assembly and stabilisation of the complexes in the immunological synapse (Rudd *et al.*, 1999).

Cytoplasmic domain - While the N-terminal domain is responsible for adhesion, the cytoplasmic domain is responsible for attachment to cytoskeleton and signalling. The cytoplasmic tail contains 116 amino acids and is largely unstructured. It is highly basic (pI~12), contains large number of prolines (~20%), and possesses the most highly conserved sequences among all CD2 homologues (Fig. 1.4) (Clayton *et al.*, 1987; Tavernor *et al.*, 1994). The cytoplasmic tail contains several functional regions which may be important in mediating signal triggered by extracellular stimulus (He *et al.*, 1988; Kivens *et al.*, 1998; Moingeon *et al.*, 1989b). As mentioned earlier, the cytoplasmic domain may be bound by several adaptor proteins;

for example, the KGPPLPRPRV sequence at the C-terminus can be bound by the SH3 domains of CD2BP1 and CD2AP (Dustin *et al.*, 1998; Li *et al.*, 1998). The CD2AP is associated with CD2-triggered cytoskeletal rearrangements. The CD2BP1 can recruit cytosolic protein tyrosine phosphatase (PTP)-PEST to the cytoplasmic tail and is associated with regulation of adhesion by CD2 and integrins (Li *et al.*, 1998), it may also induce clustering of CD2. The cytoplasmic domain of human CD2 also contains two PPPGHR segments that are bound by CD2BP2 proline-binding modules, the glycine-tyrosine-phenylalanine (GYF) domain which regulates protein-protein interaction (Freund *et al.*, 1999; Nishizawa *et al.*, 1998). CD2BP2 is involved in IL-2 production and Ca^{2+} flux. The C-terminus may also become associated with the SH3 domain of p56lck and $p59^{fyn}$ (Gassmann *et al.*, 1994; Lin *et al.*, 1998). There is, however, no tyrosine present that can mediate interaction with SH2 domain-containing signalling elements. The C-terminal asparagine appeared to be important for the upregulation of the CD2 avidity as substitution of this residue to alanine or aspartic acid completely abrogates CD2 capacity for upregulation (Hahn *et al.*, 1992b).

CD2 Ligands

A number of CAMs have been identified as the ligands for CD2. The major ligands such as human CD58 and rat CD48 are members of the CD2 family of cell surface molecules (Fig. 1.4). Other members of this family include SLAM, CD84, Ly-9 and 2B4 (Nakajima & Colonna, 2000).

CD58 - CD58 is the first ligand identified for human CD2 (Hunig, 1985; Selvaraj *et al.*, 1987) and is also known as LFA-3. While CD2 is restricted to T cells, the CD58 is widely distributed in haematopoietic and various non-haematopoietic cells. There is an alternative phospholipid-linked form of CD58 (Seed, 1987) which arises from alternate splicing of the mRNA (Wallich *et al.*, 1998). No CD58, however, has been detected in rodents. Both CD2, and CD58 have similar structure (Driscoll *et al.*, 1991; Ikemizu *et al.*, 1999; Jones *et al.*, 1992; Sun *et al.*, 1999; Wang *et al.*, 1999) and CD58 has also been demonstrated to be involved in signalling (Webb *et al.*,

1990). The CD58 in sheep erythrocytes can bind to human CD2, producing the 'rosetting' of sheep erythrocytes to human T-cells described earlier.

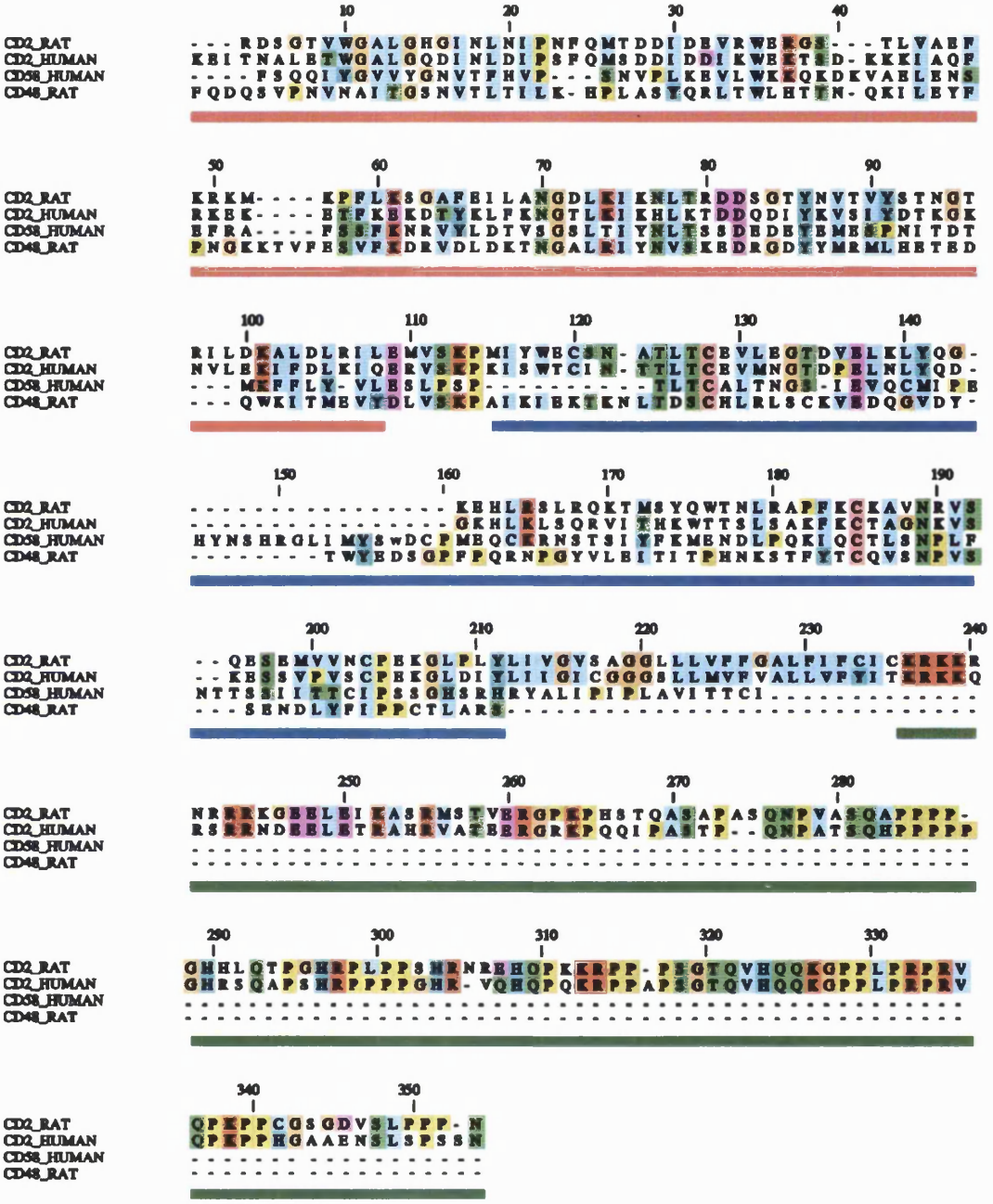


Fig. 1.4 Sequence alignment of full-length rat CD2 (Williams et al., 1987) and human CD2 (Diamond et al., 1988), and the extracellular domains of rat CD48 (Killeen et al., 1988) and human CD58 (Seed, 1987). The domain 1 is underlined in red, domain 2 in blue, and the cytoplasmic domain in green. Consensus polar residues are indicated in green, consensus acidic residues in pink, conserved basic residues in red and hydrophobic residues in blue, and conserved cysteines in magenta. Glycines are shown in orange, histidines and tyrosines cyan and prolines yellow.

CD48 - The physiological binding partner of CD2 has been shown to be CD48 in rodents (Kato *et al.*, 1992; van der Merwe *et al.*, 1993b). CD48 is also known as

MRC OX45, BCM1 and Blast-1 (Killeen *et al.*, 1988; Wong *et al.*, 1990). In rodents, CD48 is the only known ligand for CD2. The CD48 is similar in the extracellular structure to CD2, but lacks the transmembrane and cytoplasmic regions. It is anchored to the membrane via a C-terminal attached glycosyl phosphatidylinositol (GPI). In human, the CD48 binds CD2 only with very low affinity (Arulanandam *et al.*, 1993a; Sandrin *et al.*, 1993), and any such interaction may therefore be not physiologically relevant. While CD58 is widely expressed in haematopoietic and non-haematopoietic cells, the CD48 expression is limited to haematopoietic cells and endothelium (Barclay *et al.*, 1993). An alternative ligand for mouse and human CD48 has since been identified on NK and T-cell (Latchman *et al.*, 1998). In rodents, 2B4, a glycoprotein, has been shown to bind CD48 with a six to nine times higher affinity than CD2 (Brown *et al.*, 1998).

The higher binding affinity of CD48 for 2B4 in rodents and the much weaker interaction of rat CD48 with CD2 when compared to human CD2/CD58 interaction raise the question if rat CD48 is an actual ligand for CD2 *in vivo*. A number of arguments for CD48 as a physiological ligand for CD2 have been put forward (van der Merwe *et al.*, 1993b). There is as yet no equivalent of human CD58 identified in rodents (Kato *et al.*, 1992; van der Merwe *et al.*, 1993b). CD2, CD48 and CD58 have similar genetic structures (Killeen *et al.*, 1988; Wong *et al.*, 1990) and distributions (Springer *et al.*, 1987), and they exhibit similar biological activities, with both the expressions of CD58 and CD48 increased on T-cell activation. Furthermore, CD48 can mediate rosetting, and murine and rat CD2 can only bind cell that express CD48, binding that can be blocked by anti-CD48 mAbs.

CD59 - CD59, one of major complement-regulatory proteins, has been suggested to be a ligand for CD2 (Deckert *et al.*, 1992; Hahn *et al.*, 1992a). However, doubts have been raised as to whether CD59 is an actual ligand for CD2 under physiological conditions. CD59 is structurally different from CD2 (Kieffer *et al.*, 1994); moreover, the result is not replicated with no binding detected on surface plasmon resonance (SPR) studies (van der Merwe *et al.*, 1994). Nevertheless, biological role has been proposed for the interaction between CD2 and CD59 (Naderi *et al.*, 1999).

Unidentified ligands - Apart from those already identified, there may also be ligands for CD2 that have yet to be identified. For example in the murine fibroblast, no CD48 can be detected, however, murine CD2 can bind fibroblast and such binding is not block by anti-CD48 mAb (Abraham *et al.*, 1994), suggesting at least one alternative ligand for CD2. Suggested alternative ligands for CD2 include sulphated carbohydrates and lectins (O'Flynn *et al.*, 1985; Parish *et al.*, 1988; Walzel *et al.*, 2000; Warren *et al.*, 1996).

Mapping of Ligand-binding site

The mapping of CD2 ligand-binding site was first performed on hCD2 by Peterson & Seed (1987). Using mutagenesis and mAb binding, three separate binding epitopes were identified, with only regions 1 and 2 affecting adhesion interaction. The region 1 and 2 were later shown to be located on the same surface when the structure was solved (Driscoll *et al.*, 1991), while the region 3 was determined to lie between domain 1 and 2, and this region 3 forms the neo-epitope which appears after initial adhesion event mediated by region 1 and 2. A number of further mutagenesis studies then mapped out the binding sites (Arulanandam *et al.*, 1994; Arulanandam *et al.*, 1993b; Hahn *et al.*, 1992a; Miller *et al.*, 1993; Recny *et al.*, 1990; Sayre *et al.*, 1989; Somoza *et al.*, 1993), all showing the human CD2 adhesion surface to be situated on the GFCC'C" strands of CD2.

Similar results were obtained by ligand-mapping analyses on human CD58 (Arulanandam *et al.*, 1994; Dengler *et al.*, 1992; Miller *et al.*, 1993; Osborn *et al.*, 1995; Sun *et al.*, 1999), as well as for rat CD2 (Davis *et al.*, 1998a; Davis *et al.*, 1995b; van der Merwe *et al.*, 1995) and rat CD48 (van der Merwe *et al.*, 1995). In binding studies using NMR spectroscopy, the residues on rat CD2 whose resonance were perturbed on CD48 binding are again mapped onto the GFCC'C'' face of the molecule (McAlister *et al.*, 1996). These residues form a binding 'patch' which extends across an area that corresponds closely with that which mediates the 'head-to-head' contact seen in the crystal structures of rat and human CD2 (Bodian *et al.*, 1994; Jones *et al.*, 1992).

The concordance of the binding surface with the dimer-interface, suggests an evolutionary relationship between homophilic and heterophilic interactions of CD2.

Both CD2 and its ligands belong to the immunoglobulin superfamily (IgSF). These immunoglobulins may have derived from a single common ancestor whose original function was to mediate cell-cell contact by the apposition of identical immunoglobulins between two cells. Subsequent diversification of the immunoglobulins leads to the generation of multiple immunoglobulin domains and the interaction with non-identical members of IgSF. CD2 and its ligands have very similar sequence and structures (Killeen *et al.*, 1988; Williams *et al.*, 1987; Wong *et al.*, 1990), and their heterophilic interactions are therefore suggested to have evolved from such homophilic interaction (Williams & Barclay, 1988). Such homophilic interaction of CD2 dimer can be observed in the crystal structures of rat and human CD2 (Bodian *et al.*, 1994; Jones *et al.*, 1992). The two CD2 molecules interact in a manner which is comparable to the standard orthogonal mode of β-sheet packing, and such interaction has been proposed as the model for the physiological heterophilic association of CD2 with its ligands. This mode of interaction has recently been confirmed in human CD2 by the crystal structure of CD2/CD58 heterodimer which showed heterophilic interaction between CD2 and CD58 in the same orientation and involving the same binding surface as in CD2 dimer crystal structures (Wang *et al.*, 1999).

While the residues forming the hydrophobic core are largely conserved with >50% of core residues conserved (such as the invariant tryptophan W32 in rat CD2), most of the surface residues are not well conserved. However, there are two hydrophobic residues that are conserved in all species - Tyr81 and Phe49, and these may be important in CD2 binding interaction (Fig. 1.5). Phe49 in particular is unusually exposed for a hydrophobic residue. Mutation of these residues and Leu38 disrupts binding by CD2 (Davis *et al.*, 1998a; Hahn *et al.*, 1992a; Peterson & Seed, 1987). The binding surface of CD2 is highly charged, with 35% and 70% of the residues on the binding surfaces of rat and human CD2 respectively are charged, compared with ~19% for the average solvent-exposed protein surface (Miller *et al.*, 1987). In rat CD2, the alanine mutants on Asp28, Glu29 and Arg31 are found to disrupt binding (Davis *et al.*, 1998a), which may point their importance in CD2 adhesion interaction.

Adhesion interaction of CD2

The interaction of CD2 with its ligands differs significantly from the well-characterised high-affinity protein-protein interactions of many soluble proteins and antigen-antibody complexes. The adhesion interaction is transient but tightly regulated, weak yet highly specific, and it involves the simultaneous engagement of large number of molecules (Davis *et al.*, 1998a; Davis *et al.*, 1998b; Dustin & Springer, 1991; van der Merwe & Barclay, 1994). This type of intermolecular association, which is characteristic of many cell-adhesion molecules, has been suggested to be essential to maintain the transient, reversible and highly dynamic interaction between T-lymphocytes and antigen-presenting target cells (Davis *et al.*, 1998a; Davis *et al.*, 1998b; van der Merwe & Barclay, 1994). The ligand binding characteristics of CD2 has been investigated by a number of studies in recent years in an effort to understand the physicochemical basis of cell-cell adhesion and recognition at the membrane surface (Davis *et al.*, 1998a; Dustin, 1997; Dustin *et al.*, 1997a; Pierres *et al.*, 1996; Pierres *et al.*, 1997; Silkowski *et al.*, 1997).

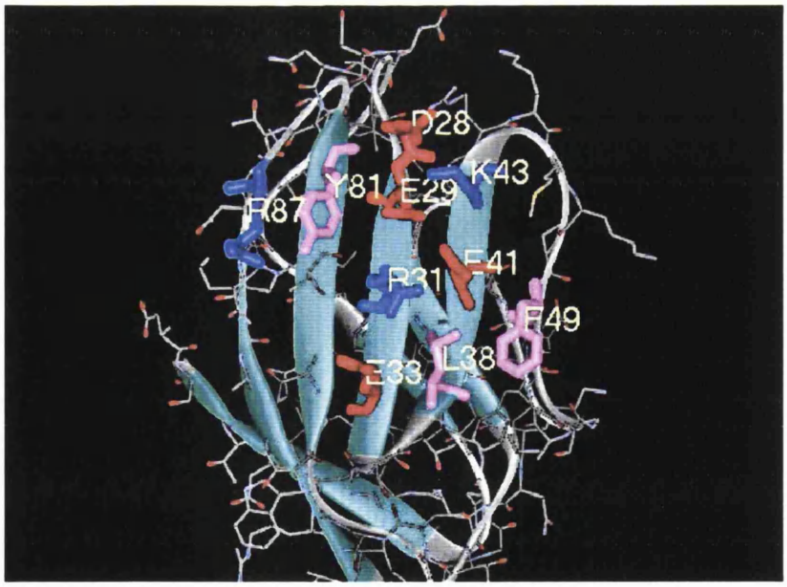


Fig. 1.5 Residues on the binding surface of CD2d1. Red indicates acidic residue, blue indicates basic residues. Three hydrophobic residues, L38, F49 and Y81, which are thought to be important in the binding interaction of CD2, are indicated in pink.

The low affinity and high specificity binding characteristics of CD2 have been attributed to the unusual features of its ligand binding surface (Davis *et al.*, 1998a). The binding surface of both rat and human CD2 is relatively flat and highly

charged (Fig. 1.5). This flat surface of CD2 suggests that the binding of CD2 relies less on surface-shape complementarity but rather on the spatial pattern of charges on the binding surface (Davis *et al.*, 1998a; Ikemizu *et al.*, 1999; Nuallain & Grant, 1998; Wang *et al.*, 1999).

A number of studies on the binding kinetics of CD2 had been performed (Dustin *et al.*, 1989; Sayre *et al.*, 1989; van der Merwe *et al.*, 1994; van der Merwe *et al.*, 1993b). Using surface plasmon resonance technology, the dissociation constants (K_D) for CD2 interactions with its ligands were determined to be 10-20 μM for human CD2/CD58 (van der Merwe *et al.*, 1994). For rat CD2/CD48 interaction, however, the affinity is approximately four times lower than that of human CD2/CD58 interaction, with a K_D value determined to be 60-90 μM, (van der Merwe *et al.*, 1993a). Other studies has yielded a K_D of ~1 μM for human CD2/CD58 interaction (Sayre *et al.*, 1989; Dustin, 1997). The k_{on} for human CD2/CD58 is around 10^5 M^{-1} s^{-1} (Davis *et al.*, 1998a), a value close to that for other high-affinity protein-protein interaction such as mAb binding to protein antigen. However, the k_{off} is very rapid, for human CD2, a $k_{off} \geq 4$ s^{-1} has been determined (van der Merwe *et al.*, 1994). In rat CD2/CD48 interaction, the off-rate is exceptionally fast, with the k_{off} determined to be 7.8 s^{-1} in studies using glass beads (Pierres *et al.*, 1996). This rate is more than 10,000 times greater than most protein-antigen/antibody interaction (Davis *et al.*, 1992). As CD2 is membrane-tethered and its interaction is limited by the planar surface of the cell membrane, a more useful measurement of the CD2 affinity and kinetic would be its two dimensional (2D) Kd. This is determined to be 1.1 and 47 molecules/μm^2 for CD2/CD58 and CD2/CD48 interaction respectively (Dustin *et al.*, 1996; Dustin *et al.*, 1997a). CD2 may be present on the cell surface at a density of 110 molecules/μm^2, the 2D affinity is therefore more than adequate for human CD2/CD58 interaction, while for rat CD2/CD48, the 2D affinity appears to be barely adequate.

This weak ligand interaction of CD2 can nevertheless be highly specific. The binding site of human and rat CD2 showed close similarities but little cross-species binding to CD48/58. Human CD2, for example, does not bind rat CD48, and rat CD2 does not associate with human CD58 (Davis *et al.*, 1998b; Somoza *et al.*, 1993). There is, however, some cross-species reactivity. The murine CD2 binds to both

murine and rat CD48 and vice versa (Wong *et al.*, 1990), and, as mentioned earlier, human CD2 interact with human and sheep CD58 in the same manner, even though sequence conservation of the extracellular domain is low (~47%) (Arulanandam *et al.*, 1994; Arulanandam *et al.*, 1993b; Somoza *et al.*, 1993).

The charged residues are thought to be important in the binding interaction of CD2. Charge complementary mutagenesis studies of rat CD2 whereby the charged surface residues on the binding surface are mutated to the opposite charge (glutamate to lysine and vice versa) result in loss of CD2 binding to CD48, but the binding capability of CD2 is restored if the complementary charged residues on CD48 are similarly mutated (van der Merwe *et al.*, 1995), suggesting that electrostatic complementarity may be of critical importance to the binding interaction of CD2. These charged residues, however, were found to contribute little to the binding energy due to the unfavourable cost of desolvation (Davis *et al.*, 1998a). It was therefore proposed that these residues play a dominant role in CD2 ligand recognition by ensuring that the binding is both weak and highly specific.

The charge-charge interaction is particularly evident in the crystal structure of CD2 in complex with CD58 (Wang *et al.*, 1999). This heterodimeric structure showed the contact surface between the two molecules to be relatively small, and is largely populated by salt-bridges and hydrogen-bonding interactions. It is therefore suggested that, for CD2/CD58 interaction, the hydrophobic van der Waals contacts play only a minor role in the binding interaction, while the electrostatic interactions between the charged residues provide the major portion of the binding energy. This interpretation is in contrast to that from the study of CD2/CD48 interaction in which the main binding energy is suggested to derived from hydrophobic interactions while the charged residues provide only the specificity of interaction (Davis *et al.*, 1998a).

Whatever the reason, the fast kinetics and low binding affinity may be an important feature of many adhesion molecules. Low affinity binding and very fast kinetics can also be seen the binding of CD80 to CD28 and CTLA-4 (van der Merwe *et al.*, 1997), CD62L to glycosylation-dependent CAM-1 (GlyCAM-1) (Nicholson *et al.*, 1998), as well as CD8αα to HLA-A2 (Gao *et al.*, 1997). The CD2 interaction with its ligands may therefore serve as a useful model for transient but highly regulated interaction on T-cell, and provide insight into how such interactions are regulated.

Knowledge of the properties of binding site can also have practical therapeutic applications. For example, understanding of the binding interaction of CD2 can be used for developing novel tumour therapy techniques by designing bispecific antibodies that can target tumour antigens as well as CD2 without affecting its binding interaction (Wild *et al.*, 1999b). CD2 mAbs have also been shown to induce unresponsiveness in T-cell and tolerance to allografts *in vivo* (Guckel *et al.*, 1991), and combining anti-CD2 and CD48 mAbs can allow indefinite survival of allografts by affecting CD3 driven proliferation of $CD4^+$ cells. (Qin *et al.*, 1994). Knowledge of the binding site can therefore aid in the design of peptide or small molecules that could block the CD2/CD58 pathway, thereby providing a new immunosuppressive agent. The binding interaction of CD2 is weak, and is therefore easily blocked by something that can bind more strongly.

Scope and aims of the thesis

Adhesion interaction between leukocytes and other cells are typically transient and highly regulated. It is essential to understand how the regulation of such transient interactions is achieved, and there is also a need to understand the relationship between molecular structure and biological function, in this case, the relationship between the residues on the binding site and adhesion interaction. Electrostatic properties of a protein may be important in the functioning of a protein. Charged residues in particular are thought to play a pivotal role in the binding interaction of adhesion molecules such as CD2. In this thesis, the electrostatic properties of the charged residues are examined by the determination of the ionisation constants (pK_a) of the charged residues, in particular, those of the carboxylates. The pK_as are determined using heteronuclear NMR spectroscopy by following the pH-dependent chemical shift changes of the resonances which also allow the electrostatic effects due to the ionisations of the charge groups to be monitored.

In order to produce the protein suitable for the NMR analysis, the thesis describes the initial optimisation studies in a bid to improve protein expression such that the protein can be expressed to the high level necessary for the isotopic labelling of the protein. This includes the construction of a number of expression plasmids of

the first domain of CD2 (CD2d1), and the development of the expression and purification protocols. The thesis also describes the use of isotopically-labelled protein for the complete assignment of the ^{1}H, ^{15}N and ^{13}C resonances of CD2d1 that is necessary for further NMR analysis. It further describes the generation of the various mutants of CD2d1 as well as their use in the studies of residues on the binding surface of CD2d1.

In addition to the assessment of electrostatic properties of CD2 by pH titration, the CD2d1 was also used for other biophysical analysis, in particular, the investigation of the NMR relaxation parameters of CD2d1 in an effort to understand the dynamic properties of the protein. A detailed analysis of the relaxation parameter at varying conditions was also undertaken in order to assess the molecular origin of strongly enhanced transverse relaxation rates that is present for some residues on CD2d1. The implications of the findings to the analysis of internal dynamics of protein in general are discussed. The protein-protein interaction of CD2 was also examined by NMR in order to assess the effect of the charged residues on the binding properties of the molecules.

References

Abraham, D., Bougharios, G., Tulip, G., Sumner, H. & Olsen, I. (1994). Regulation of CD2-Mediated Heterotypic Interactions of Murine T- Lymphocytes. *Cellular Immunology* **156**(2), 342-356.

Alcover, A., Alberini, C., Acuto, O., Clayton, L. K., Transy, C., Spagnoli, G. C., Moingeon, P., Lopez, P. & Reinherz, E. L. (1988). Interdependence of CD3-Ti and CD2 Activation Pathways in Human Lymphocytes-T. *Embo Journal* **7**(7), 1973-1977.

Arulanandam, A. R. N., Kister, A., McGregor, M. J., Wyss, D. F., Wagner, G. & Reinherz, E. L. (1994). Interaction Between Human CD2 and CD58 Involves the Major Beta-Sheet Surface of Each of Their Respective Adhesion Domains. *Journal of Experimental Medicine* **180**(5), 1861-1871.

Arulanandam, A. R. N., Moingeon, P., Concino, M., Recny, M. A., Kato, K., Yagita, H., Koyasu, S. & Reinherz, E. L. (1993a). A Soluble Multimeric Recombinant CD2 Protein Identifies CD48 As a Low Affinity Ligand For Human CD2 - Divergence of CD2 Ligands During the Evolution of Man and Mouse. *Journal of Immunology* **150**(8 Pt2), A 187-A 187.

Arulanandam, A. R. N., Withka, J. M., Wyss, D. F., Wagner, G., Kister, A., Pallai, P., Recny, M. A. & Reinherz, E. L. (1993b). The CD58 (LFA-3) Binding-Site Is a Localized and Highly-Charged Surface-Area On the AGFCC'C" Face of the Human CD2 Adhesion Domain. *Proceedings of the National Academy of Sciences of the United States of America* **90**(24), 11613-11617.

Bachmann, M. E., Barner, M. & Kopf, M. (1999). CD2 Sets Quantitative Thresholds in T Cell Activation. *Journal of Experimental Medicine* **190**(10), 1383-1391.

Barclay, A. N., Birkeland, M. L., Brown, M. H., Beyers, A. D., Davis, S. J., Somoza, C. & Williams, A. F. (1993). *The Leucocyte Antigen FactsBook*, Academic Press.

Bell, G. I., Dembo, M. & Bongrand, P. (1984). Cell-Adhesion - Competition Between Nonspecific Repulsion and Specific Bonding. *Biophysical Journal* **45**(6), 1051-1064.

Bell, G. M., Fargnoli, J., Bolen, J. B., Kish, L. & Imboden, J. B. (1996). The SH3 domain of p56(lck) binds to proline-rich sequences in the cytoplasmic domain of CD2. *Journal of Experimental Medicine* **183**(1), 169-178.

Beyers, A. D., Barclay, A. N., Law, D. A., He, Q. & Williams, A. F. (1989). Activation of Lymphocytes-T Via Monoclonal-Antibodies Against Rat- Cell Surface-Antigens With Particular Reference to CD2 Antigen. *Immunological Reviews* **111**, 59-77.

Beyers, A. D., Spruyt, L. L. & Williams, A. F. (1992). Molecular Associations Between the Lymphocyte-T Antigen Receptor Complex and the Surface Antigen-CD2, Antigen-CD4, or Antigen-CD8 and Antigen-CD5. *Proceedings of the National Academy of Sciences of the United States of America* **89**(7), 2945-2949.

Bierer, B. E., Barbosa, J., Herrmann, S. & Burakoff, S. J. (1988). Interaction of CD2 With Its Ligand, LFA-3, in Human T-Cell Proliferation. *Journal of Immunology* **140**(10), 3358-3363.

Bierer, B. E., Bogart, R. E. & Burakoff, S. J. (1990). Partial Deletions of the Cytoplasmic Domain of CD2 Result in a Partial Defect in Signal Transduction. *Journal of Immunology* **144**(3), 785-789.

Bierer, B. E. & Burakoff, S. J. (1989). Lymphocyte-T Activation - the Biology and Function of CD2 and CD4. *Immunological Reviews* **111**, 267-294.

Bierer, B. E., Sleckman, B. P., Ratnofsky, S. E. & Burakoff, S. J. (1989). The Biologic Roles of CD2, CD4, and CD8 in T-Cell Activation. *Annual Review of Immunology* **7**, 579-599.

Bodian, D. L., Jones, E. Y., Harlos, K., Stuart, D. I. & Davis, S. J. (1994). Crystal-Structure of the Extracellular Region of the Human Cell- Adhesion Molecule CD2 At 2.5-Angstrom Resolution. *Structure* **2**(8), 755-766.

Boussiotis, V. A., Freeman, G. J., Griffin, J. D., Gray, G. S., Gribben, J. G. & Nadler, L. M. (1994). CD2 Is Involved in Maintenance and Reversal of Human Alloantigen- Specific Clonal Anergy. *Journal of Experimental Medicine* **180**(5), 1665-1673.

Brain, P., Gordon, J. & Willetts, W., A. (1970). Rosette formation by peripheral lymphocytes. *Clinical Experimental Immunology* **6**, 681.

Brown, M. H., Boles, K., vanderMerwe, P. A., Kumar, V., Mathew, P. A. & Barclay, A. N. (1998). 2B4, the natural killer and T cell immunoglobulin superfamily surface protein, is a ligand for CD48. *Journal of Experimental Medicine* **188**(11), 2083-2090.

Brown, M. H., Cantrell, D. A., Brattsand, G., Crumpton, M. J. & Gullberg, M. (1989). The CD2 Antigen Associates With the T-Cell Antigen Receptor CD3 Antigen Complex On the Surface of Human Lymphocytes-T. *Nature* **339**(6225), 551-553.

Brown, M. J. & Shaw, S. (1999). T-cell activation: Interplay at the interface. *Current Biology* **9**(1), R26 (4 pages).

Cantrell, D. (1996). T cell antigen receptor signal transduction pathways. *Annual Review of Immunology* **14**, 259-274.

Carmo, A. M., Castro, M. A. A. & Arosa, F. A. (1999). CD2 and CD3 associate independently with CD5 and differentially regulate signaling through CD5 in Jurkat T cells. *Journal of Immunology* **163**(8), 4238-4245.

Chambers, C. A. & Allison, J. P. (1999). Costimulatory regulation of T cell function. *Current Opinion in Cell Biology* **11**(2), 203-210.

Chang, H. C., Moingeon, P., Pedersen, R., Lucich, J., Stebbins, C. & Reinherz, E. L. (1990). Involvement of the PPPGHR Motif in T-Cell Activation Via CD2. *Journal of Experimental Medicine* **172**(1), 351-355.

Clark, S. J., Law, D. A., Paterson, D. J., Puklavec, M. & Williams, A. F. (1988). Activation of Rat Lymphocytes-T By Anti-CD2 Monoclonal-Antibodies. *Journal of Experimental Medicine* **167**(6), 1861-1872.

Clayton, L. K., Sayre, P. H., Novotny, J. & Reinherz, E. L. (1987). Murine and Human T11 (CD2) Cdna Sequences Suggest a Common Signal Transduction Mechanism. *European Journal of Immunology* **17**(9), 1367-1370.

Collins, T. L., Kassner, P. D., Bierer, B. E. & Burakoff, S. J. (1994). Adhesion Receptors in Lymphocyte activation. *Current Opinion in Immunology* **6**, 385-393.

Corr, M., Slanetz, A. E., Boyd, L. F., Jelonek, M. T., Khilko, S., Alramadi, B. K., Kim, Y. S., Maher, S. E., Bothwell, A. L. M. & Margulies, D. H. (1994). T-Cell Receptor-MHC Class-I Peptide Interactions - Affinity, Kinetics, and Specificity. *Science* **265**(5174), 946-949.

Davis, S. J., Davies, E. A., Barclay, A. N., Daenke, S., Bodian, D. L., Jones, E. Y., Stuart, D. I., Butters, T. D., Dwek, R. A. & van der Merwe, P. A. (1995a). Ligand-Binding By the Immunoglobulin Superfamily Recognition Molecule CD2 Is Glycosylation-Independent. *Journal of Biological Chemistry* **270**(1), 369-375.

Davis, S. J., Davies, E. A., Tucknott, M. G., Jones, E. Y. & van der Merwe, P. A. (1998a). The role of charged residues mediating low affinity protein-protein recognition at the cell surface by CD2. *Proceedings of the National Academy of Sciences of the United States of America* **95**(10), 5490-5494.

Davis, S. J., Davies, E. A. & van der Merwe, P. A. (1995b). Mutational Analysis of the Epitopes Recognized By Anti-(Rat CD2) and Anti-(Rat CD48) Monoclonal-Antibodies. *Biochemical Society Transactions* **23**(1), 188-194.

Davis, S. J., Ikemizu, S., Wild, M. K. & van der Merwe, P. A. (1998b). CD2 and the nature of protein interactions mediating cell-cell recognition. *Immunological Reviews* **163**, 217-236.

Davis, S. J., Schockmel, G. A., Somoza, C., Buck, D. W., Healey, D. G., Rieber, E. P., Reiter, C. & Williams, A. F. (1992). Antibody and HIV-1 Gp120 Recognition of CD4 Undermines the Concept of Mimicry Between Antibodies and Receptors. *Nature* **358**(6381), 76-79.

Davis, S. J. & van der Merwe, P. A. (1996). The Structure and Ligand Interactions of CD2 - Implications For T- Cell Function. *Immunology Today* **17**(4), 177-187.

Davis, S. J. & vanderMerwe, P. A. (1996). CD2: An exception to the immunoglobulin superfamily concept? *Science* **273**(5279), 1241-1242.

Deckert, M., Kubar, J., Zoccola, D., Bernardpomier, G., Angelisova, P., Horejsi, V. & Bernard, A. (1992). CD59 Molecule - a 2nd Ligand For CD2 in T-Cell Adhesion. *European Journal of Immunology* **22**(11), 2943-2947.

Demotz, S., Grey, H. M. & Sette, A. (1990). The Minimal Number of Class-Ii MHC Antigen Complexes Needed For T- Cell Activation. *Science* **249**(4972), 1028-1030.

Dengler, T. J., Hoffmann, J. C., Knolle, P., Albertwolf, M., Roux, M., Wallich, R. & Meuer, S. C. (1992). Structural and Functional Epitopes of the Human Adhesion Receptor- Cd58 (Lfa-3). *European Journal of Immunology* **22**(11), 2809-2817.

Diamond, D. J., Clayton, L. K., Sayre, P. H. & Reinherz, E. L. (1988). Exon Intron Organization and Sequence Comparison of Human and Murine T11 (CD2) Genes. *Proceedings of the National Academy of Sciences of the United States of America* **85**(5), 1615-1619.

Driscoll, P. C., Cyster, J. G., Campbell, I. D. & Williams, A. F. (1991). Structure of Domain-1 of Rat Lymphocyte-T CD2 Antigen. *Nature* **353**(6346), 762-765.

Dumont, C., Deas, O., Mollereau, B., Hebib, C., GiovinoBarry, V., Bernard, A., Hirsch, F., Charpentier, B. & Senik, A. (1998). Potent apoptotic signaling and subsequent unresponsiveness induced by a single CD2 mAb (BTI-322) in activated human peripheral T cells. *Journal of Immunology* **160**(8), 3797-3804.

Dustin, M. L. (1997). Adhesive bond dynamics in contacts between T lymphocytes and glass-supported planar bilayers reconstituted with the immunoglobulin- related adhesion molecule CD58. *Journal of Biological Chemistry* **272**(25), 15782-15788.

Dustin, M. L. (1998). Making a little affinity go a long way: A topological view of LFA-1 regulation. *Cell Adhesion and Communication* **6**(2-3), 255-262.

Dustin, M. L., Ferguson, L. M., Chan, P. Y., Springer, T. A. & Golan, D. E. (1996). Visualization of CD2 Interaction With LFA-3 and Determination of the 2-Dimensional Dissociation-Constant For Adhesion Receptors in a Contact Area. *Journal of Cell Biology* **132**(3), 465-474.

Dustin, M. L., Golan, D. E., Zhu, D. M., Miller, J. M., Meier, W., Davies, E. A. & van der Merwe, P. A. (1997a). Low affinity interaction of human or rat T cell adhesion molecule CD2 with its

ligand aligns adhering membranes to achieve high physiological affinity. *Journal of Biological Chemistry* **272**(49), 30889-30898.

Dustin, M. L., Olive, D. & Springer, T. A. (1989). Correlation of CD2 Binding and Functional-Properties of Multimeric and Monomeric Lymphocyte Function-Associated Antigen-3. *Journal of Experimental Medicine* **169**(2), 503-517.

Dustin, M. L., Olszowy, M. W., Holdorf, A. D., Li, J., Bromley, S., Desai, N., Widder, P., Rosenberger, F., van der Merwe, P. A., Allen, P. M. & Shaw, A. S. (1998). A novel adaptor protein orchestrates receptor patterning and cytoskeletal polarity in T-cell contacts. *Cell* **94**(5), 667-677.

Dustin, M. L. & Springer, T. A. (1991). Role of Lymphocyte Adhesion Receptors in Transient Interactions and Cell Locomotion. *Annual Review of Immunology* **9**, 27-66.

Dustin, M. L., White, R., Suri, A., Springer, T. A. & SelvaRaj, P. (1997b). Principles governing the rapid organization of adhesion molecules in cell-cell contact areas. *Molecular Biology of the Cell* **8**(SS), 1176.

Freund, C., Dotsch, V., Nishizawa, K., Reinherz, E. L. & Wagner, G. (1999). The GYF domain is a novel structural fold that is involved in lymphoid signaling through proline-rich sequences. *Nature Structural Biology* **6**(7), 656-660.

Gao, G. F., Tormo, J., Gerth, U. C., Wyer, J. R., McMichael, A. J., Stuart, D. I., Bell, J. I., Jones, E. Y. & Jakobsen, B. K. (1997). Crystal structure of the complex between human CD8 alpha alpha and HLA-A2. *Nature* **387**(6633), 630-634.

Garboczi, D. N., Ghosh, P., Utz, U., Fan, Q. R., Biddison, W. E. & Wiley, D. C. (1996). Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. *Nature* **384**(6605), 134-141.

Garcia, K. C., Degano, M., Stanfield, R. L., Brunmark, A., Jackson, M. R., Peterson, P. A., Teyton, L. & Wilson, I. A. (1996). An alpha beta T cell receptor structure at 2.5 angstrom and its orientation in the TCR-MHC complex. *Science* **274**(5285), 209-219.

Gassmann, M., Amrein, K. E., Flint, N. A., Schraven, B. & Burn, P. (1994). Identification of a Signaling Complex Involving CD2, Zeta-Chain and P59(Fyn) in T-Lymphocytes. *European Journal of Immunology* **24**(1), 139-144.

Gollob, J. A., Li, J., Reinherz, E. L. & Ritz, J. (1995). CD2 Regulates Responsiveness of Activated T-Cells to Interleukin-12. *Journal of Experimental Medicine* **182**(3), 721-731.

Gollob, J. A. & Ritz, J. (1996). CD2-CD58 interaction and the control of T-cell interleukin-12 responsiveness - Adhesion molecules link innate and acquired immunity. *Annals of the New York Academy of Sciences* **795**, 71-81.

Grakoui, A., Bromley, S. K., Sumen, C., Davis, M. M., Shaw, A. S., Allen, P. M. & Dustin, M. L. (1999). The immunological synapse: A molecular machine controlling T cell activation. *Science* **285**(5425), 221-227.

Gray, F., Cyster, J. G., Willis, A. C., Barclay, A. N. & Williams, A. F. (1993). Structural-Analysis of the CD2 T-Lymphocyte Antigen By Site-Directed Mutagenesis to Introduce a Disulfide Bond Into Domain-1. *Protein Engineering* **6**(8), 965-970.

Green, J.M.; Karpitskiy, V.; Kimzey, S. L.; Shaw, A. S. (2000) Coordinate regulation of T cell activation by CD2 and CD28. *J Immunol* **164**(7):3591-3595.

Guckel, B., Berek, C., Lutz, M., Altevogt, P., Schirrmacher, V. & Kyewski, B. A. (1991). Anti-CD2 Antibodies Induce T-Cell Unresponsiveness Invivo. *Journal of Experimental Medicine* **174**(5), 957-967.

Guyot, D. J., Newbound, G. C. & Lairmore, M. D. (1997). Signaling via the CD2 receptor enhances HTLV-1 replication in T lymphocytes. *Virology* **234**(1), 123-129.

Hahn, W. C., Menu, E., Bothwell, A. L. M., Sims, P. J. & Bierer, B. E. (1992a). Overlapping But Nonidentical Binding-Sites On CD2 For CD58 and a 2nd Ligand CD59. *Science* **256**(5065), 1805-1807.

Hahn, W. C., Rosenstein, Y., Calvo, V., Burakoff, S. J. & Bierer, B. E. (1992b). A Distinct Cytoplasmic Domain of CD2 Regulates Ligand Avidity and T- Cell Responsiveness to Antigen. *Proceedings of the National Academy of Sciences of the United States of America* **89**(15), 7179-7183.

Haque, S., Dumon, H., Haque, A. & Kasper, L. H. (1998). Alteration of intracellular calcium flux and impairment of nuclear Factor-AT translocation in T cells during acute Toxoplasma gondii infection in mice. *Journal of Immunology* **161**(12), 6812-6818.

He, Q., Beyers, A. D., Barclay, A. N. & Williams, A. F. (1988). A Role in Transmembrane Signaling For the Cytoplasmic Domain of the CD2 Lymphocyte-T Surface-Antigen. *Cell* **54**(7), 979-984.

Hogg, N. & Berlin, C. (1995). Structure and Function of Adhesion Receptors in Leukocyte Trafficking. *Immunology Today* **16**(7), 327-330.

Howe, L. R. & Weiss, A. (1995). Multiple Kinases Mediate T-Cell-Receptor Signaling. *Trends in Biochemical Sciences* **20**(2), 59-64.

Hunig, T. (1985). The Cell-Surface Molecule Recognized By the Erythrocyte Receptor of Lymphocytes-T - Identification and Partial Characterization Using a Monoclonal-Antibody. *Journal of Experimental Medicine* **162**(3), 890-901.

Hutchcroft, J. E., Slavik, J. M., Lin, H. M., Watanabe, T. & Bierer, B. E. (1998). Uncoupling activation-dependent HS1 phosphorylation from nuclear factor of activated T cells transcriptional activation in Jurkat T cells: Differential signaling through CD3 and the costimulatory receptors CD2 and CD28. *Journal of Immunology* **161**(9), 4506-4512.

Ikemizu, S., Sparks, L. M., VanderMerwe, P. A., Harlos, K., Stuart, D. I., Jones, E. Y. & Davis, S. J. (1999). Crystal structure of the CD2-binding domain of CD58 (lymphocyte function-associated antigen 3) at 1.8-angstrom resolution. *Proceedings of the National Academy of Sciences of the United States of America* **96**(8), 4289-4294.

Jondal, M., Homl, G. & Wigsell, H. (1972). Surface markers on himan T and B lymphocytes: a large population of lymphocytes forming nonimmune rosettes with sheep red blood cells. *Journal of experimental medicine* **136**, 207.

Jones, E. Y., Davis, S. J., Williams, A. F., Harlos, K. & Stuart, D. I. (1992). Crystal-Structure At 2.8-Angstrom Resolution of a Soluble Form of the Cell-Adhesion Molecule CD2. *Nature* **360**(6401), 232-239.

June, C. H., Bluestone, J. A., Nadler, L. M. & Thompson, C. B. (1994). The B7 and Cd28 Receptor Families. *Immunology Today* **15**(7), 321-331.

Kanner, S. B., Damle, N. K., Blake, J., Aruffo, A. & Ledbetter, J. A. (1992). CD2/LFA-3 Ligation Induces Phospholipase-C-Gamma-1 Tyrosine Phosphorylation and Regulates CD3 Signaling. *Journal of Immunology* **148**(7), 2023-2029.

Kato, K., Koyanagi, M., Okada, H., Takanashi, T., Wong, Y. W., Williams, A. F., Okumura, K. & Yagita, H. (1992). CD48 Is a Counter-Receptor For Mouse CD2 and Is Involved in T-Cell Activation. *Journal of Experimental Medicine* **176**(5), 1241-1249.

Kieffer, B., Driscoll, P. C., Campbell, I. D., Willis, A. C., Vandermerwe, P. A. & Davis, S. J. (1994). 3-Dimensional Solution Structure of the Extracellular Region of the Complement Regulatory Protein CD59, a New Cell-Surface Protein Domain Related to Snake-Venom Neurotoxins. *Biochemistry* **33**(15), 4471-4482.

Killeen, N., Moessner, R., Arvieux, J., Willis, A. & Williams, A. F. (1988). The MRC OX-45 Antigen of Rat Leukocytes and Endothelium Is in a Subset of the Immunoglobulin Superfamily With CD2, LFA-3 and Carcinoembryonic Antigens. *EMBO Journal* **7**(10), 3087-3091.

Killeen, N., Stuart, S. G. & Littman, D. R. (1992). Development and Function of T-Cells in Mice With a Disrupted CD2 Gene. *EMBO Journal* **11**(12), 4329-4336.

Kivens, W. J., Hunt, S. W., Mobley, J. L., Zell, T., Dell, C. L., Bierer, B. E. & Shimizu, Y. (1998). Identification of a proline-rich sequence in the CD2 cytoplasmic domain critical for regulation of integrin-mediated adhesion and activation of phosphoinositide 3-kinase. *Molecular and Cellular Biology* **18**(9), 5291-5307.

Klein, J. & Horejsi, V. (1997). *Immunology*. Second edit, Blackwell Science Ltd.

Koyasu, S., Lawton, T., Novick, D., Recny, M. A., Siliciano, R. F., Wallner, B. P. & Reinherz, E. L. (1990). Role of Interaction of CD2 Molecules With Lymphocyte Function- Associated Antigen-3 in T-Cell Recognition of Nominal Antigen. *Proceedings of the National Academy of Sciences of the United States of America* **87**(7), 2603-2607.

Lanzavecchia, A., Iezzi, G. & Viola, A. (1999). From TCR engagement to T cell activation: A kinetic view of T cell behavior. *Cell* **96**(1), 1-4.

Latchman, Y., McKay, P. F. & Reiser, H. (1998). Cutting edge: Identification of the 2B4 molecule as a counter- receptor for CD48. *Journal of Immunology* **161**(11), 5809-5812.

Lay, W. H., Mendes, N., F, Bianco, C. & Nussenzweig, V. (1971). Binding of sheep red blood cells to a lrage population of human lymphocytes. *Nature* **230**, 531.

leGouvello, S., Manceau, V. & Sobel, A. (1998). Serine 16 of stathmin as a cytosolic target for Ca2+/calmodulin- dependent kinase II after CD2 triggering of human T lymphocytes. *Journal of Immunology* **161**(3), 1113-1122.

Leupin, O; Zaru, R; Laroche, T; Muller, S; Valitutti, S (2000). Exclusion of CD45 from the T-cell receptor signaling area in antigen-stimulated T lymphocytes. *Current Biologyl* **10**(5):277-280.

Ley, S. C., Davies, A. A., Druker, B. & Crumpton, M. J. (1991). The T-Cell Receptor CD3 Complex and CD2 Stimulate the Tyrosine Phosphorylation of Indistinguishable Patterns of Polypeptides in the Human T-Leukemic Cell-Line Jurkat. *European Journal of Immunology* **21**(9), 2203-2209.

Li, J., Nishizawa, K., An, W. Q., Hussey, R. E., Lialios, F. E., Salgia, R., SunderPlassmann, R. & Reinherz, E. L. (1998). A cdc15-like adaptor protein (CD2BP1) interacts with the CD2 cytoplasmic domain and regulates CD2-triggered adhesion. *Embo Journal* **17**(24), 7320-7336.

Li, J., Smolyar, A., Sunderplassmann, R. & Reinherz, E. L. (1996). Ligand-Induced Conformational Change Within the CD2 Ectodomain Accompanies Receptor Clustering - Implication For Molecular Lattice Formation. *Journal of Molecular Biology* **263**(2), 209-226.

Lin, H. M., Hutchcroft, J. E., Andoniou, C. E., Kamoun, M., Band, H. & Bierer, B. E. (1998). Association of p59(fyn) with the T lymphocyte costimulatory receptor CD2 - Binding of the Fyn Src homology (SH) 3 domain is regulated by the Fyn SH2 domain. *Journal of Biological Chemistry* **273**(31), 19914-19921.

Lin, J. X., Yon, R. W., Chavin, K. D., Qin, L. H., Woodward, J., Ding, Y. Z., Yagita, H. & Bromberg, J. S. (1995). Anti-CD2 Monoclonal Antibody-Induced Receptor Changes - Down-Modulation of Cell-Surface CD2. *Transplantation* **59**(8), 1162-1171.

Linsley, P. S. & Ledbetter, J. A. (1993). The Role of the CD28 Receptor During T-Cell Responses to Antigen. *Annual Review of Immunology* **11**, 191-212.

Malissen, B. (1999). Immunology - Dancing the immunological two-step. *Science* **285**(5425), 207-208.

MarieCardine, A., Fischer, S., Gorvel, J. P. & MaridonneauParini, I. (1996). Recruitment of activated p56(lck) on endosomes of CD2-triggered T cells, colocalization with ZAP-70. *Journal of Biological Chemistry* **271**(34), 20734-20739.

Matsui, K., Boniface, J. J., Steffner, P., Reay, P. A. & Davis, M. M. (1994). Kinetics of T-Cell Receptor-Binding to Peptide I-E(K) Complexes - Correlation of the Dissociation Rate With T-Cell Responsiveness. *Proceedings of the National Academy of Sciences of the United States of America* **91**(26), 12862-12866.

McAlister, M. S. B., Mott, H. R., van der Merwe, P. A., Campbell, I. D., Davis, S. J. & Driscoll, P. C. (1996). NMR Analysis of Interacting Soluble Forms of the Cell-Cell Recognition Molecules CD2 and CD48. *Biochemistry* **35**(19), 5982-5991.

McAllister, P. T. & Ellis, T. M. (1996). CD2 regulates T cell-dependent induction of monocyte IL-1 beta mRNA during anti-CD3 mitogenesis. *Cellular Immunology* **170**(1), 120-126.

Meuer, S. C., Hussey, R. E., Cantrell, D. A., Hodgdon, J. C., Schlossman, S. F., Smith, K. A. & Reinherz, E. L. (1984a). Triggering of the T3-Ti Antigen Receptor Complex Results in Clonal T- Cell Proliferation Through an Interleukin 2-Dependent Autocrine Pathway. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences* **81**(5), 1509-1513.

Meuer, S. C., Hussey, R. E., Fabbi, M., Fox, D., Acuto, O., Fitzgerald, K. A., Hodgdon, J. C., Protentis, J. P., Schlossman, S. F. & Reinherz, E. L. (1984b). An Alternative Pathway of T-Cell Activation - a Functional-Role For the 50 Kd T11 Sheep Erythrocyte Receptor Protein. *Cell* **36**(4), 897-906.

Miller, G. T., Hochman, P. S., Meier, W., Tizard, R., Bixler, S. A., Rosa, M. D. & Wallner, B. P. (1993). Specific Interaction of Lymphocyte Function-Associated Antigen-3 With CD2 Can Inhibit T-Cell Responses. *Journal of Experimental Medicine* **178**(1), 211-222.

Miller, S., Janin, J., Lesk, A. M. & Chothia, C. (1987). Interior and Surface of Monomeric Proteins. *Journal of Molecular Biology* **196**(3), 641-656.

Moingeon, P., Chang, H. C., Sayre, P. H., Clayton, L. K., Alcover, A., Gardner, P. & Reinherz, E. L. (1989a). The Structural Biology of CD2. *Immunological Reviews* **111**, 111-144.

Moingeon, P., Chang, H. C., Wallner, B. P., Stebbins, C., Frey, A. Z. & Reinherz, E. L. (1989b). CD2-Mediated Adhesion Facilitates Lymphocyte-T Antigen Recognition Function. *Nature* **339**(6222), 312-314.

Moingeon, P., Lucich, J. L., McConkey, D. J., Letourneur, F., Malissen, B., Kochan, J., Chang, H. C., Rodewald, H. R. & Reinherz, E. L. (1992). CD3-Zeta Dependence of the CD2 Pathway of Activation in Lymphocytes-T and Natural-Killer-Cells. *Proceedings of the National Academy of Sciences of the United States of America* **89**(4), 1492-1496.

Monks, C. R. F., Freiberg, B. A., Kupfer, H., Sciaky, N. & Kupfer, A. (1998). Three-dimensional segregation of supramolecular activation clusters in T cells. *Nature* **395**(6697), 82-86.

Naderi, S., Hormann, P., Seiter, S., Tilgen, W., Abken, H. & Reinhold, U. (1999). CD2-mediated CD59 stimulation in keratinocytes results in secretion of IL-1 alpha, IL-6, and GM-CSF: Implications for the interaction of keratinocytes with intraepidermal T lymphocytes. *International Journal of Molecular Medicine* **3**(6), 609-614.

Nakajima, H. & Colonna, M. (2000). 2B4: an NK cell activating receptor with unique specificity and signal transduction mechanism. *Human Immunology* **61**(1):39-43.

Nicholson, M. W., Barclay, A. N., Singer, M. S., Rosen, S. D. & vanderMerwe, P. A. (1998). Affinity and kinetic analysis of L-selectin (CD62L) binding to glycosylation-dependent cell-adhesion molecule-1. *Journal of Biological Chemistry* **273**(2), 763-770.

Nishizawa, K., Freund, C., Li, J., Wagner, G. & Reinherz, E. L. (1998). Identification of a proline-binding motif regulating CD2-triggered T lymphocyte activation. *Proceedings of the National Academy of Sciences of the United States of America* **95**(25), 14897-14902.

Nuallain, B. O. & Grant, G. H. (1998). A computer model of the LFA-3/CD2 complex. *Biochemical Society Transactions* **26**(1), S44.

O'Flynn, K., Krensky, A. M., Beverley, P. C. L., Burakoff, S. J. & Linch, D. C. (1985). Phytohemagglutinin Activation of T-Cells Through the Sheep Red Blood- Cell Receptor. *Nature* **313**(6004), 686-687.

Osborn, L., Day, E. S., Miller, G. T., Karpusas, M., Tizard, R., Meuer, S. C. & Hochman, P. S. (1995). Amino-Acid-Residues Required For Binding of Lymphocyte Function- Associated Antigen-3 (CD58) to Its Counter-Receptor CD2. *Journal of Experimental Medicine* **181**(1), 429-434.

Parish, C. R., McPhun, V. & Warren, H. S. (1988). Is a Natural Ligand of the Lymphocyte-T Cd2 Molecule a Sulfated Carbohydrate. *Journal of Immunology* **141**(10), 3498-3504.

Peng, X. H., Kasran, A., Bullens, D. & Ceuppens, J. L. (1997). Ligation of CD2 provides a strong helper signal for the production of the type 2 cytokines interleukin-4 and -5 by memory T cells. *Cellular Immunology* **181**(1), 76-85.

Peterson, A. & Seed, B. (1987). Monoclonal-Antibody and Ligand-Binding Sites of the T-Cell Erythrocyte Receptor (CD2). *Nature* **329**(6142), 842-846.

Peterson, E. J., Clements, J. L., Fang, N. & Koretzky, G. A. (1998). Adaptor proteins in lymphocyte antigen-receptor signaling. *Current Opinion in Immunology* **10**(3), 337-344.

Pierres, A., Benoliel, A. M., Bongrand, P. & van der Merwe, P. A. (1996). Determination of the lifetime and force dependence of interactions of single bonds between surface-attached CD2 and CD48 adhesion molecules. *Proceedings of the National Academy of Sciences of the United States of America* **93**(26), 15114-15118.

Pierres, A., Benoliel, A. M., Bongrand, P. & van der Merwe, P. A. (1997). The dependence of the association rate of surface-attached adhesion molecules CD2 and CD48 on separation distance. *Febs Letters* **403**(3), 239-244.

Porter, J. C. & Hogg, N. (1997). Integrin cross talk: Activation of lymphocyte function-associated antigen-1 on human T cells alters alpha 4 beta 1- and alpha 5 beta 1- mediated function. *Journal of Cell Biology* **138**(6), 1437-1447.

Qin, L. H., Chavin, K. D., Lin, J. X., Yagita, H. & Bromberg, J. S. (1994). Anti-CD2 Receptor and Anti-CD2 Ligand (CD48) Antibodies Synergize to Prolong Allograft Survival. *Journal of Experimental Medicine* **179**(1), 341-346.

Recny, M. A., Luther, M. A., Knoppers, M. H., Neidhardt, E. A., Khandekar, S. S., Concino, M. F., Schimke, P. A., Francis, M. A., Moebius, U., Reinhold, B. B., Reinhold, V. N. & Reinherz, E. L. (1992). N-Glycosylation Is Required For Human CD2 Immunoadhesion Functions. *Journal of Biological Chemistry* **267**(31), 22428-22434.

Recny, M. A., Neidhardt, E. A., Sayre, P. H., Ciardelli, T. L. & Reinherz, E. L. (1990). Structural and Functional-Characterization of the CD2 Immunoadhesion Domain - Evidence For Inclusion of CD2 in an Alpha-Beta-Protein Folding Class. *Journal of Biological Chemistry* **265**(15), 8542-8549.

Ren, R. B., Mayer, B. J., Cicchetti, P. & Baltimore, D. (1993). Identification of a 10-Amino Acid Proline-Rich Sh3 Binding-Site. *Science* **259**(5098), 1157-1161.

Rudd, P. M., Wormald, M. R. & Dwek, R. A. (1999). Glycosylation and the immune system. *Trends in Glycoscience and Glycotechnology* **11**(57), 1-21.

Rutishauser, U., Acheson, A., Hall, A. K., Mann, D. M. & Sunshine, J. (1988). The Neural Cell-Adhesion Molecule (NCAM) As a Regulator of Cell-Cell Interactions. *Science* **240**(4848), 53-57.

Samelson, L. E., Fletcher, M. C., Ledbetter, J. A. & June, C. H. (1990). Activation of Tyrosine Phosphorylation in Human T-Cells Via the CD2 Pathway - Regulation By the CD45 Tyrosine Phosphatase. *Journal of Immunology* **145**(8), 2448-2454.

Sandrin, M. S., Mouhtouris, E., Vaughan, H. A., Warren, H. S. & Parish, C. R. (1993). CD48 Is a Low-Affinity Ligand For Human CD2. *Journal of Immunology* **151**(9), 4606-4613.

Sayre, P. H., Hussey, R. E., Chang, H. C., Ciardelli, T. L. & Reinherz, E. L. (1989). Structural and Binding Analysis of a 2-Domain Extracellular CD2 Molecule. *Journal of Experimental Medicine* **169**(3), 995-1009.

Schraven, B., Samstag, Y., Altevogt, P. & Meuer, S. C. (1990). Association of CD2 and CD45 On Human Lymphocytes-T. *Nature* **345**(6270), 71-74.

Schwartz, R. H. (1990). A Cell-Culture Model For Lymphocyte-T Clonal Anergy. *Science* **248**(4961), 1349-1356.

Seed, B. (1987). An LFA-3 cDNA Encodes a Phospholipid-Linked Membrane-Protein Homologous to Its Receptor CD2. *Nature* **329**(6142), 840-842.

Selvaraj, P., Plunkett, M. L., Dustin, M., Sanders, M. E., Shaw, S. & Springer, T. A. (1987). The Lymphocyte-T Glycoprotein CD2 Binds the Cell-Surface Ligand LFA-3. *Nature* **326**(6111), 400-403.

Semnani, R. T., Svoboda, K., Khoshnood, B. & vanSeventer, G. A. (1998). Acquisition of interleukin-5 secretion by human naive T-helper cells is regulated by distinct signals from both the T-cell receptor CD3 complex and CD2. *Scandinavian Journal of Immunology* **47**(5), 436-443.

Sen, J., Arceci, R. J., Jones, W. & Burakoff, S. J. (1989). Expression and Ontogeny of Murine CD2. *European Journal of Immunology* **19**(7), 1297-1302.

Sen, J., Bossu, P., Burakoff, S. J. & Abbas, A. K. (1992). T-Cell Surface Molecules Regulating Noncognate Lymphocyte-B Activation - Role of CD2 and LFA-1. *Journal of Immunology* **148**(4), 1037-1042.

Shaw, A. S. & Dustin, M. L. (1997). Making the T cell receptor go the distance: A topological view of T cell activation. *Immunity* **6**(4), 361-369.

Shaw, S., Luce, G. E. G., Quinones, R., Gress, R. E., Springer, T. A. & Sanders, M. E. (1986). 2 Antigen-Independent Adhesion Pathways Used By Human Cytotoxic T- Cell Clones. *Nature* **323**(6085), 262-264.

Siliciano, R. F., Pratt, J. C., Schmidt, R. E., Ritz, J. & Reinherz, E. L. (1985). Activation of Cytolytic Lymphocyte-T and Natural-Killer Cell-Function Through the T11 Sheep Erythrocyte Binding-Protein. *Nature* **317**(6036), 428-430.

Silkowski, H., Davis, S. J., Barclay, A. N., Rowe, A. J., Harding, S. E. & Byron, O. (1997). Characterisation of the low affinity interaction between rat cell adhesion molecules CD2 and CD48 by analytical ultracentrifugation. *European Biophysics Journal With Biophysics Letters* **25**(5-6), 455-462.

Simons, K. & Ikonen, E. (1997). Functional rafts in cell membranes. *Nature* **387**(6633), 569-572.

Sinkora, J., Rehakova, Z., Sinkora, M., Cukrowska, B., TlaskalovaHogenova, H., Bianchi, A. T. J. & DeGeus, B. (1998). Expression of CD2 on porcine B lymphocytes [Full text available, price (Pounds)7.64]. *Immunology* **95**(3), 443-449.

Somoza, C., Driscoll, P. C., Cyster, J. G. & Williams, A. F. (1993). Mutational Analysis of the CD2/CD58 Interaction - the Binding-Site For CD58 Lies On One Face of the 1st Domain of Human CD2. *Journal of Experimental Medicine* **178**(2), 549-558.

Springer, T. A. (1990). Adhesion Receptors of the Immune-System. *Nature* **346**(6283), 425-434.

Springer, T. A., Dustin, M. L., Kishimoto, T. K. & Marlin, S. D. (1987). The Lymphocyte Function-Associated LFA-1, CD2, and LFA-3 Molecules - Cell-Adhesion Receptors of the Immune-System. *Annual Review of Immunology* **5**, 223-252.

Spruyt, L. L., Glennie, M. J., Beyers, A. D. & Williams, A. F. (1991). Signal Transduction By the CD2 Antigen in T-Cells and Natural-Killer- Cells - Requirement For Expression of a Functional T-Cell Receptor or Binding of Antibody-Fc to the Fc-Receptor, Fc-Gamma-Riiia (CD16). *Journal of Experimental Medicine* **174**(6), 1407-1415.

Sultan, P., Schechner, J. S., McNiff, J. M., Hochman, P. S., Hughes, C. C. W., Lorber, M. I., Askenase, P. W. & Pober, J. S. (1997). Blockade of CD2-LFA-3 interactions protects human skin allografts in immunodeficient mouse/human chimeras. *Nature Biotechnology* **15**(8), 759-762.

Sun, Z. Y. J., Dotsch, V., Kim, M., Li, J., Reinherz, E. L. & Wagner, G. (1999). Functional glycan-free adhesion domain of human cell surface receptor CD58: design, production and NMR studies. *Embo Journal* **18**(11), 2941-2949.

Sunder-Plassmann, R. & Reinherz, E. L. (1998). A p56(lck)-independent pathway of CD2 signaling involves Jun kinase. *Journal of Biological Chemistry* **273**(37), 24249-24257.

Targan, S. R., Deem, R. L., Liu, M., Wang, S. & Nel, A. (1995). Definition of a Lamina Propria T-Cell Responsive State - Enhanced Cytokine Responsiveness of T-Cells Stimulated Through the Cd2 Pathway. *Journal of Immunology* **154**(2), 664-675.

Tavernor, A. S., Kydd, J. H., Bodian, D. L., Jones, E. Y., Stuart, D. I., Davis, S. J. & Butcher, G. W. (1994). Expression Cloning of an Equine T-Lymphocyte Glycoprotein CD2 Cdna Structure-Based Analysis of Conserved Sequence Elements. *European Journal of Biochemistry* **219**(3), 969-976.

Teh, S. J., Killeen, N., Tarakhovsky, A., Littman, D. R. & Teh, H. S. (1997). CD2 regulates the positive selection and function of antigen-specific CD4(-)CD8(+) T cells. *Blood* **89**(4), 1308-1318.

Ueno, H; Matsuda, S; Katamura, K; Mayumi, M; Koyasu, S (2000). ZAP-70 is required for calcium mobilization but is dispensable for mitogen-activated protein kinase (MAPK) superfamily activation induced via CD2 in human T cells.. *Eur J Immunol* **30**(1), 78-86.

Umehara, H., Inoue, H., Yoneda, O., Gouda, S. & Domae, N. (1999). CD2 crosslinking enhances tyrosine phosphorylation of CBL, resulting in increased association with CRK in NK3.3 cell line. *Faseb Journal* **13**(5 Pt2 SS), A963.

Valitutti, S., Muller, S., Cella, M., Padovan, E. & Lanzavecchia, A. (1995). Serial Triggering of Many T-Cell Receptors By a Few Peptide-Mhc Complexes. *Nature* **375**(6527), 148-151.

van der Merwe, P. A. (1999). A Subtle Role for CD2 in T Cell Antigen Recognition. *Journal of Experimental Medicine* **190**(10), 1371-1374.

van der Merwe, P. A. & Barclay, A. N. (1994). Transient intermolecular adhesion: the importance of weak protein-protein interactions. *Trends in Biochemical Science* **19**, 354-358.

van der Merwe, P. A., Barclay, A. N., Mason, D. W., Davies, E. A., Morgan, B. P., Tone, M., Krishnam, A. K. C., Ianelli, C. & Davis, S. J. (1994). Human Cell-Adhesion Molecule CD2 Binds CD58 (LFA-3) With a Very-Low Affinity and an Extremely Fast Dissociation Rate But Does Not Bind CD48 or CD59. *Biochemistry* **33**(33), 10149-10160.

van der Merwe, P. A., Bodian, D. L., Daenke, S., Linsley, P. & Davis, S. J. (1997). CD80 (B7-1) binds both CD28 and CTLA-4 with a low affinity and very fast kinetics. *Journal of Experimental Medicine* **185**(3), 393-403.

van der Merwe, P. A., Brown, M. H., Davis, S. J. & Barclay, A. N. (1993a). Affinity and Kinetic-Analysis of the Interaction of the Cell-Adhesion Molecules Rat CD2 and CD48. *EMBO Journal* **12**(13), 4945-4954.

van der Merwe, P. A., McNamee, P. N., Davies, E. A., Barclay, A. N. & Davis, S. J. (1995). Topology of the CD2-CD48 Cell-Adhesion Molecule Complex - Implications For Antigen Recognition By T-Cells. *Current Biology* **5**(1), 74-84.

van der Merwe, P. A., McPherson, D. C., Brown, M. H., Barclay, A. N., Cyster, J. G., Williams, A. F. & Davis, S. J. (1993b). The NH2-Terminal Domain of Rat CD2 Binds Rat CD48 With a Low-Affinity and Binding Does Not Require Glycosylation of CD2. *European Journal of Immunology* **23**(6), 1373-1377.

van Leeuwen, J. E. M. & Samelson, L. E. (1999). T cell antigen-receptor signal transduction. *Current Opinion in Immunology* **11**(3), 242-248.

Vankooyk, Y., Kemenade, P. V., Weder, P., Kuijpers, T. W. & Figdor, C. G. (1989). Enhancement of LFA-1-Mediated Cell-Adhesion By Triggering Through CD2 or CD3 On Lymphocytes-T. *Nature* **342**(6251), 811-813.

Vollger, L. W., Tuck, D. T., Springer, T. A., Haynes, B. F. & Singer, K. H. (1987). Thymocyte Binding to Human Thymic Epithelial-Cells Is Inhibited By Monoclonal-Antibodies to CD-2 and LFA-3 Antigens. *Journal of Immunology* **138**(2), 358-363.

Wallich, R., Brenner, C., Brand, Y., Roux, M., Reister, M. & Meuer, S. (1998). Gene structure, promoter characterization, and basis for alternative mRNA splicing of the human CD58 gene. *Journal of Immunology* **160**(6), 2862-2871.

Walzel H., Blach M., Hirabayashi J., Kasai K.I., Brock J. (2000). nvolvement of CD2 and CD3 in galectin-1 induced signaling in human Jurkat T-cells. *Glycobiology* **10**(2):131-140.

Wang, J., Smolyar, A., Tan, K. M., Liu, J., Kim, M. Y., Sun, Z. J., Wagner, G. & Reinherz, E. L. (1999). Structure of a heterophilic adhesion complex between the human CD2 and CD58 (LFA-3) counterreceptors. *Cell* **97**(6), 791-803.

Warren, H. S., Altin, J. G., Waldron, J. C., Kinnear, B. F. & Parish, C. R. (1996). A Carbohydrate Structure Associated With CD15 (Lewis(X)) On Myeloid Cells Is a Novel Ligand For Human CD2. *Journal of Immunology* **156**(8), 2866-2873.

Watts, T. H. & DeBenedette, M. A. (1999). T cell co-stimulatory molecules other than CD28. *Current Opinion in Immunology* **11**(3), 286-293.

Webb, D. S. A., Shimizu, Y., Vanseventer, G. A., Shaw, S. & Gerrard, T. L. (1990). LFA-3, CD44, and CD45 - Physiological Triggers of Human Monocyte Tnf and Il-1 Release. *Science* **249**(4974), 1295-1297.

Wild, M. K., Cambiaggi, A., Brown, M. H., Davies, E. A., Ohno, H., Saito, T. & vanderMerwe, P. A. (1999a). Dependence of T cell antigen recognition on the dimensions of an accessory receptor-ligand complex. *Journal of Experimental Medicine* **190**(1), 31-41.

Wild, M. K., Strittmatter, W., Matzku, S., Schraven, B. & Meuer, S. C. (1999b). Tumor therapy with bispecific antibody: The targeting and triggering steps can be separated employing a CD2-based strategy. *Journal of Immunology* **163**(4), 2064-2072.

Williams, A. F. & Barclay, A. N. (1988). The Immunoglobulin Superfamily - Domains For Cell-Surface Recognition. *Annual Review of Immunology* **6**, 381-405.

Williams, A. F., Barclay, A. N., Clark, S. J., Paterson, D. J. & Willis, A. C. (1987). Similarities in Sequences and Cellular Expression Between Rat CD2 and CD4 Antigens. *Journal of Experimental Medicine* **165**(2), 368-380.

Wilson, I. A. & Stanfield, R. L. (1993). Antibody-Antigen Interactions. *Current Opinion in Structural Biology* **3**(1), 113-118.

Withka, J. M., Wyss, D. F., Wagner, G., Arulanandam, A. R. N., Reinherz, E. L. & Recny, M. A. (1993). Structure of the Glycosylated Adhesion Domain of Human T-Lymphocyte Glycoprotein CD2. *Structure* **1**(1), 69-81.

Wong, Y. W., Williams, A. F., Kingsmore, S. F. & Seldin, M. F. (1990). Structure, Expression, and Genetic-Linkage of the Mouse BCM1 (Ox45 or Blast-1) Antigen - Evidence For Genetic Duplication Giving Rise to the Bcm1 Region On Mouse Chromosome-1 and the CD2/Lfa3 Region On Mouse Chromosome-3. *Journal of Experimental Medicine* **171**(6), 2115-2130.

Wyss, D. F., Choi, J. S., Li, J., Knoppers, M. H., Willis, K. J., Arulanandam, A. R. N., Smolyar, A., Reinherz, E. L. & Wagner, G. (1995). Conformation and Function of the N-Linked Glycan in the Adhesion Domain of Human CD2. *Science* **269**(5228), 1273-1278.

Yagita, H., Asakawa, J., Tansyo, S., Nakamura, T., Habu, S. & Okumura, K. (1989). Expression and Function of CD2 During Murine Thymocyte Ontogeny. *European Journal of Immunology* **19**(12), 2211-2217.

Yang, S. Y., Chouaib, S. & Dupont, B. (1986). A Common Pathway For Lymphocyte-T Activation Involving Both the CD3- Ti Complex and CD2 Sheep Erythrocyte Receptor Determinants. *Journal of Immunology* **137**(4), 1097-1100.

Yashiro, Y., Tai, X. G., Toyooka, K., Park, C. S., Abe, R., Hamaoka, T., Kobayashi, M., Neben, S. & Fujiwara, H. (1998). A fundamental difference in the capacity to induce proliferation of naive T cells between CD28 and other co-stimulatory molecules. *European Journal of Immunology* **28**(3), 926-935.

Yashiro-Ohtani, Y; Zhou, X.Y.; Toyo-Oka, K; Tai, X.G.; Park, C.S.; Hamaoka, T; Abe, R; Miyake, K; Fujiwara, H (2000). Non-CD28 costimulatory molecules present in T cell rafts induce T cell costimulation by enhancing the association of TCR with rafts. *Journal of Immunology* **164**(3), 1251-1259.

Chapter 2

Methods - Molecular biology, protein purification and analytical techniques

This chapter describes the methods used in the processes directed at the production of protein samples as well as some of the techniques used for the analysis of these proteins. These include the various protocols used for the generation of the various recombinant plasmid DNA constructs, methods for the expression and purification of proteins, and a variety of biophysical and analytical techniques.

Overview of PCR cloning

For the purpose of optimising the gene expression of CD2d1, it was necessary to express the protein using different DNA constructs in order to examine the effect of promoters, fusion partners and codon usage. Polymerase chain reaction (PCR) cloning is a very flexible way of generating these different recombinant gene constructs and mutants. In essence, the process involved the use of PCR to introduce various elements to the 5'- and 3'-ends of the target gene which allow the products of

the PCR to be cloned into the vector of choice. This is achieved by the incorporation of suitable restriction sites or other extra DNA sequences into the 5' non-hybridising ends of the primers used in the PCR. On amplification using these primers, the designed elements will be incorporated into the amplified DNA fragment, thereby allowing the PCR-amplified product to be further manipulated for insertion into the chosen expression vectors.

The PCR was performed using the pGEX-2T GST-fusion CD2d1 plasmid construct (Driscoll *et al.*, 1991) as the template. After PCR, the amplified DNA product can be purified either by using Qiaquick PCR purification kit (Qiagen) or by gel purification to remove the primers, buffers and enzymes used in the PCR. The purified PCR products were digested with suitable restriction enzymes in an appropriate buffer. This is followed by purification by agarose gel electrophoresis to remove unwanted DNA fragments, enzymes and various contaminants before ligating the purified DNA into a vector similarly digested and purified. The ligation product was then transformed into competent *E. coli* cells and plated onto selective agar. A parallel control ligation reaction of the digested vector without insert was performed. If this control reaction showed few background self-ligation products, the colonies are picked directly for DNA preparation; and if there was a high background, the transformed colonies were screened for the presence of ligated insert in the vectors. Plasmid DNA containing the recombinant gene was purified by DNA midiprep. The DNA insert was analysed by restriction digest and its sequence confirmed by DNA sequencing.

In all cases, small-scale centrifugation of 5 ml or less was done on a bench-top Micro Centaur microcentrifuge (MSE). For volume of between 10-100 ml and centrifugation at speed less than 4000g, Heraeus Labofuge 400R centrifuge (rotor type 8179) was used. Centrifugation at higher speed and/or larger volume were done on a Sorvall RC-5B centrifuge using SS34 for volume of 100 ml or less, or GS-3 rotor for volume greater than 100 ml. The liquid culture was incubated in a shaking incubator and the agar plates in an Economy incubator (both from Gallenkamp). The expression vector pET21b vector was obtained from Novagen and pT7-7 vector was a gift from Dr Gregg Siegal.

Primer design

All the oligonucleotides (Table 2.1) were designed according to the following general rules: The sequence of the forward primers has the same sequence as the sense strand, while the reverse primers were complementary to the sense strand. Oligonucleotide sequence of 18-20 bases complementary to the template is sufficient for effective annealing to the template DNA in most cases. The desired hexanucleotide restriction sites and other genetic elements were introduced to the 5'-end of the complementary sequence. No more than two G's or C's were used at the 3' end to avoid problem of mispriming. A start (ATG) codon was added as appropriate immediately preceding the 5' end of the coding sequence and a stop codon (TAA) after the coding sequence; in the case of fusion to a C-terminal His-tag, no stop codon needed to be introduced. A restriction site was introduced adjoining the start or stop codons or coding sequence. In the case of *NdeI*, the restriction site CATATG incorporates the start codon. A further eight extra bases were added to allow for efficient restriction digest with 6 or more of these bases being G or C at the 5' end forming the GC clamp. The oligonucleotides were synthesized commercially (Genosys, Perkin-Elmer or Pharmacia).

Table 2.1 *List of primers used for PCR cloning*

Primers	***Oligonucleotide sequence***
pT7 CGT	5'-GCCGGCCCCCATATGCGTGACAGTGGGACCGTCTG
pT7 AGA	5'-GCCGGCCCCATATGAGAGACAGTGGGACCGT
Reverse	5'-CCCCGGTGGGATCCTTACTCGAGAAATCCTCAAGTC

PCR (polymerase chain reaction)

The following were added to a 0.5 ml microfuge tube: 10 µl of 10X reaction buffer, 1 µl each of 15 µM forward and reverse primers, 1 µl of template DNA (~0.01 µg), 5µl of 2 mM dNTP mix, 8 µl of 25 mM $MgCl_2$ solution, and water to make up to 100 µl of reaction mix. The microfuge tube was placed in a thermocycler (OmniGene, HYBAID) and heated to 95 °C. 1 µl of thermostable DNA polymerase (*Taq*, *Tli*, or *Pfu*) was added immediately and a drop of mineral oil was used to

overlay the reaction mix. The PCR was then performed according to the following cycling scheme: a denaturation step at 94 °C for 1 minute, an annealing step at 55 °C for 1 minute, and an extension step at 72 °C for 1 minute. This cycling scheme was repeated 29 times followed by a final cycle with an extension step of 5 minutes to fill in any uncompleted polymerisation. Most of the parameters used in the PCR can be varied to optimise the reaction. For example the Mg^{2+}, dNTP or template concentration can be changed to improve efficiency or reduce error rate, and the annealing and the extension time can be varied depending on the primers, template or polymerase used. *Taq* DNA polymerase was only used in cases such as colony screening whereby fidelity of the DNA polymerisation reaction is of little importance, while the high-fidelity polymerases such as *Tli* or *Pfu* which have proof-reading capability were used for the PCR cloning to minimise undesirable mutations.

Restriction digest

The PCR products and the vector DNA were digested using the appropriate restriction enzymes. An excess of the restriction enzyme was used to ensure complete digest. Into a 1.5 ml eppendorf tube, 1 μg of the DNA was added together with 2 μl of the appropriate buffer and 1 μl of restriction enzymes. Water was added to make up to 20 μl. The reaction mix was allowed to incubate for 1-2 hour at 37°C. In cases where the DNA needed to be digest with two restriction enzymes, a buffer suitable for both enzymes were chosen and both restriction enzymes were added to the reaction mix. If a suitable buffer could not be found, after digestion with the first enzyme, the DNA was precipitated by adding 0.5 volume 5 M ammonium acetate and 2 volume isopropanol followed by centrifugation for 10 minutes. The pellet was washed with 80% ethanol, centrifuged and the supernatant removed. The air-dried DNA was then dissolved in water and then digested with the second enzyme.

DNA purification by agarose gel electrophoresis

All digested DNA were purified by horizontal agarose gel-electrophoresis before ligation. Low-melting-point (LMP) agarose was used for DNA purification as it contains fewer contaminants that can interfere with the ligation reaction. The gel

was first cast in an agarose gel electrophoresis apparatus (Horizon 11.14, GIBCO BRL, Life Technologies) with 50 ml of 2% Type 1-A low electroendosmosis (EEO) agarose (Sigma) in TAE buffer (40 mM Tris base, 1 mM EDTA, adjusted to pH 8.0 with glacial acetic acid). When the gel has set, the comb was positioned above the gel, leaving a small gap between the gel surface and comb. 100 ml of 1% ultra-pure LMP agarose (GIBCO BRL) was then poured on top and allowed to set. This method helped the LMP agarose to set and provided a firm base for the handling of the LMP agarose gel. The gel was submerged in TAE buffer and the samples loaded into the wells with gel loading buffer (0.04% bromophenol blue, 0.04% xylene cyanol, 2.5% Ficoll type 400 in water) and the electrophoresis conducted at 100 – 150 V for 30 - 60 minutes. The gel was stained with ethidium bromide solution (10 µl of a 10 mg/ml solution) for 10-15 minutes and the resolved DNA bands viewed on a long-wave UV transilluminator. DNA band of the correct size was excised from the gel and purified using Geneclean III DNA gel purification kit (BIO 101 Inc.).

Ligation

The ligation mix contained the following: ~ 100 ng of digested vector DNA; equimolar, or 2 or 3X molar concentration of digested DNA insert; and water to make up to 18 µl. There should not be too high a concentration of DNA in the ligation mix, as high DNA concentration will promote intermolecular reaction rather than the intramolecular reaction necessary for the circularisation of the plasmid. The ligation mixture was heated at 45 °C for 5 minutes to melt any annealed cohesive ends, and then cooled on ice. 2 µl of 10X ligase buffer (0.3 M Tris-HCl pH7.8, 0.1 M $MgCl_2$, 0.1 M DTT, 5 mM ATP) and 0.5 µl of ligase were then added. The ligation mixture was incubated for 2-4 hours at 14 °C or overnight at 4 °C. The ligase was inactivated by heating at 65 °C for 10 minutes and then allowed to cool on ice. 5-10 µl of the ligation mix was used to transform competent cells.

Preparation of competent cells

A number of protocols exist for the preparation of competent cells (Hanahan *et al.*, 1991). This protocol by Inoue *et al.*, (1990) gives highly competent cells with

transformation efficiency greater than 10^8 cpu/μg of supercoiled DNA. Cells strains that are suitable for this method include DH5α and JM109 (Table 2.2).

Table 2.2 *Cell strains used for cloning and protein expression*

Cell strains	Genotype
DH5α	*supE44 ΔlacU169 (ϕ80 lacZΔM15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1*
JM109	*recA1 supE44 endA1 hsdR17 gyrA96 relA1 thi Δ(lac-proAB)* F'*[traD36 proAB*$^+$ *lacI*^{q} *lacZΔM15]*
BL21(DE3)	*F*$^-$ *hsdS dcm gal ompT* (λcIts857 *ind*1 *S*am7 *nin5 lacUV5*-T7 gene 1)

A few colonies were picked from a freshly-streaked plate and inoculated into 100 ml of SOB medium (2 g tryptone, 0.5 g yeast extract, 0.05 g NaCl, 2.5 mM KCl, 10 mM $MgSO_4$, pH 7.0) in a 500ml baffled polycarbonate flask (Fisher). The cells were incubated at 18 °C with vigorous shaking for 30-36 hours. When the density of the cell culture reached an OD_{600} measurement of 0.6, the cells were chilled on ice for 10 minutes, and then pelleted by centrifugation at 1500-2000 g. The supernatant was decanted, the cells gently resuspended in 30 ml of ice-cold transformation buffer TB (10 mM Pipes, 15 mM $CaCl_2$, 250 mM KCl, 55 mM $MnCl_2$, pH 6.7) and incubated on ice for 10 minutes. Centrifugation was repeated at 1000 g, and the pelleted cells resuspended in 8 ml of TB. At this point the cells should be noticeably easier to pellet and resuspend. Dimethylsulphoxide (DMSO) was added to a final concentration of 7% and the cell suspension incubated on ice for a further 10 minutes. The competent cells were then aliquoted into 2.5 ml cryovials, snap-froze and stored in liquid nitrogen.

For the expression strain BL21, this protocol did not appear to produce significant improvement in its transformation efficiency. Therefore in the preparation of competent cells for this strain, the cells can be grown at 37°C.

Transformation

The frozen competent cells were allowed to defrost slowly on ice for 30 minutes. DNA solution or ligation mixture (maximum 10 μl) was placed in a 1.5 ml microcentrifuge tube, and 200 μl of the competent was then gently pippetted into the

microcentrifuge tube and mixed by slow swirling of the pipette tips. The cells were left to incubate on ice for 30-60 minutes before a heat-shock treatment at 42 °C for 30 - 120 seconds (the length of heat-shock time used depends on the cell strains). The tubes were then returned to ice to cool for 1 minute. 0.8 ml of pre-warmed SOC medium was added to the cells and incubated at 37 °C for 30 minutes. The cells were then spread onto the appropriate selective LB agar plate and allowed to incubate overnight at 37 °C.

Colony screening

Two different methods of screening for recombinant DNA were used for the detection of ligated inserts:

1) **Colony PCR** - Colonies were picked from the plate and each suspended in 11 µl of 1X PCR buffer in separate 0.5 ml microfuge tubes. A fresh plate was spotted with 1 µl of this suspension. The rest of the cell suspension are then heated in a thermocycler to 95 °C and 10µl of the 1x PCR reaction mixture including *Taq* DNA polymerase and the PCR was performed as described previously. The presence of the gene insert should produce a PCR product of the correct molecular size. In order to avoid false negative, one of the oligonucleotide used should prime to the vector while the other should prime to the insert. 5 µl of the reaction mixture was used for gel-electrophoresis and the PCR product can be visualised on long-wave UV-transilluminator after staining with ethidium bromide.

2) **Quick screen** - Colonies were picked from a plate and the cells dispersed in 100 µl of Quick Screen Buffer (6% glycerol, 1% SDS, 10 µg/µl Ribonuclease A, 0.002% bromophenol blue, 1X TAE) in 1.5 ml microcentrifuge tubes. The suspensions were left at room temperature for 5 minutes or until the cells had lysed. The tubes were then spun at top speed for 15 minutes. 30-40 µl of each sample was loaded dry (i.e. no buffer in well) into the wells of an agarose gel. The electrophoresis was run for 1-2 minute at 300 V. The wells were then washed out with buffer and the gel electrophoresis ran at 100-150 V for 30 minutes. The gel can then be stained and viewed on an UV-transilluminator. That vector with an insert would show a slightly larger apparent molecular mass.

DNA preparation (Midiprep)

A single colony was picked from the plate and used to inoculate 100 ml of LB medium (10 g/L tryptone, 5 g/L yeast extract, 0.5 g/L NaCl) containing ampicillin as the selective antibiotics. After incubation overnight at 37 °C, the cells were harvested by centrifugation and the plasmid DNA were prepared either using QIAfilter Plasmid Midi kit (Qiagen) or by the ammonium acetate method described below. This method utilised the differential precipitation by ammonium acetate and is a cheap alternative to Qiagen Midiprep kit for the preparation of clean DNA:

The pelleted cells were resuspended in 10 ml of Solution I (50 mM Glucose, 10 mM EDTA, 25 mM Tris HCl, pH 8.0). For lysis of the cells, 20 ml of Solution II (2 ml 10% SDS. 0.4 ml 10 M NaOH, 17.6 ml water) was added, mixed gently and left on ice for 5 minutes. After cell lysis, 15 ml of 7.5 M ammonium acetate was added to neutralise the solution, resulting in the precipitation the cell debris and chromosomal DNA. The precipitation was removed by centrifugation at 20,000 g for 10 minutes. 27 ml of isopropanol was then added to the supernatant and a further centrifugation step at 20,000 g for 10 minutes pelleted both DNA and proteins. The pellet was dissolved in 2 ml of 2M ammonium acetate. After 5 minutes on ice, the sample was centrifuged to pellet to remove any precipitated proteins. Equal volume of isopropanol was then added and centrifuged for 10 minutes at 20,000 g to pellet the DNA. The DNA pellet was dissolved with 0.5 ml of RNase solution (0.1 mg/ml in TE buffer) and incubated at 37 °C for 10 minutes. The DNA was then precipitated by addition of 0.5 ml of 7.5 M ammonium acetate and 0.6 ml of isopropanol. After centrifugation at 20,000g, the pellet was washed with 80% ethanol. The precipitated DNA was left to air-dry and then dissolved in 200 µl of TE buffer.

DNA Sequencing

For DNA sequencing, the DNA template for the sequencing reaction was prepared and the sequencing reaction performed as described using either the Sequenase Version 2.0 DNA sequencing kit (USB) or the CircumVent™ Thermal Cycle Dideoxy DNA Sequencing Kit (New England BioLabs). [α-^{35}S] dATP (1250 Ci/mmol) was used as the radioactive label with 5 µCi used per reaction when using the Sequenase kit or 10 µCi for the cycle sequencing. The primers used were T7

promoter primer for the forward reactions or T7 terminator primer (both from Novagen) for the reverse sequencing reactions. When the sequencing reaction has completed, the DNA fragments were resolved in a sequencing gel.

The sequencing gel was prepared largely as described in Sambrook *et al.* (1989). Two cleaned sequencing plates were assembled separated by two spacers at the side edges, with the inner surface of shorter plate siliconised with trimethylsilane solution (BDH) beforehand. The edges were sealed with tapes to form a mould. TBE buffer (10X concentration - 108 g/L Tris-base, 55 g/L boric acid, 7.4 g/L $Na_2EDTA.2H2O$) and the polyacrylamide gel mixture (26 ml 8.5 M urea, 10 ml of 30% acrylamide solution, 5 ml of TBE, 9 ml of H2O) were prepared. The polymerisation of the gel mix was initiated with 300 µl of 10% APS and 50 µl of TEMED and pippetted directly into the mould. The comb was placed into the mould with the serrated edge upward forming the well. The mould was allowed to rest horizontally with a slight upward tilt until the polymerisation has completed. When the gel has set, the tape and comb were removed, the well rinsed with water and the plates assembled onto the DNA sequencing gel electrophoresis apparatus (Sequencing System Model S2, Life Technology). Both the top and bottom tanks were filled with 500 ml of 1X TBE buffer. The comb was inserted into the well with the serrated edge down, with the teeth of the comb just fractionally penetrating the gel surface, forming a series of small wells. These wells were rinsed and the gel pre-ran at 1250 V for 30 minutes before the DNA sample can be loaded. The DNA samples from the sequencing reaction were heated to 85 °C for 2 minutes and 2 µl loaded into the well. The electrophoresis was done at a constant power of 65 W for 2-4 hours. When the electrophoresis was completed, the gel plates were removed from the apparatus and the plates prised apart. The gel was transferred to a large piece of chromatography paper (Whatman), then overlaid with Saran wrap (Dow) and placed in gel-dryer (Model 583, Bio-Rad). Once dried, the dried gel was placed in film cassette (Genetic Research Instrumentation Ltd.) together with an RX 35x44 autoradiograph (Kodak) for 12 - 24 hours of exposure. The autoradiograph was developed using Compact X2 film developer (X-ograph Ltd.) to reveal the ladders of DNA bands.

Site-directed mutagenesis

The CD2 mutants were constructed using a modified method of site-directed mutagenesis by Ito *et al.* (1991). This method utilises four oligonucleotides - one mutagenic primer (oligo III in Figure 2.2) which is specific for each mutagenesis reaction, and 3 invariant ones for all the different mutant constructs - T7 promoter primer (oligo I), pT7 terminator primer (oligo II), and a reverse primer which destroys both *Xho*I and *Bam*HI restriction sites (oligo IV - the 'knock-out' primer). In this modified method, the mutant oligo serves as the forward primer while the 'knock-out' primer serves as reverse primer. The modification allows the DNA construct to be inserted between NdeI and XhoI or BamHI restriction sites of the expression vector pET21b. A list of the mutant oligonucleotides is listed in Table 2.3.

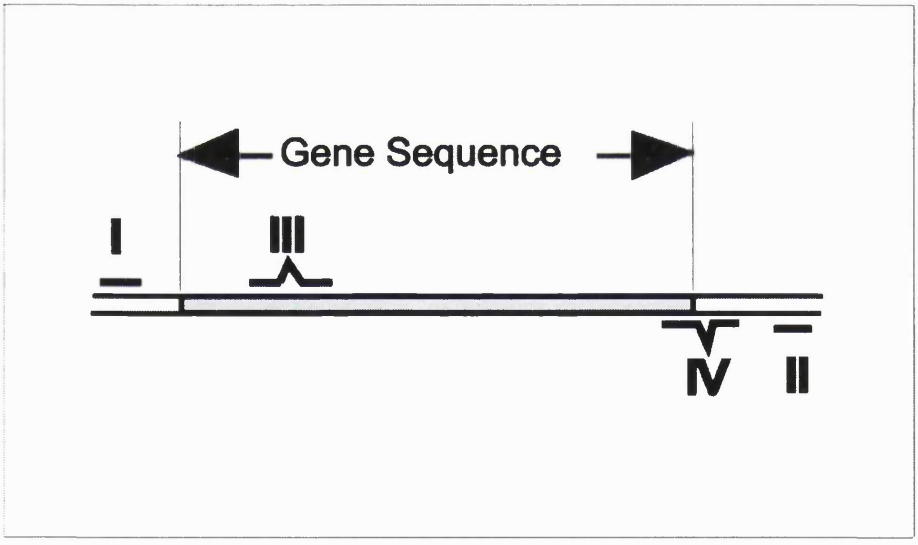


Fig 3.2 Schematic diagram of oligonucleotides used in the mutagenesis reaction. Oligo I - T7 promoter primer, oligo II- T7 terminator primer, oligo III - mutagenic primer, oligo IV - 'knock-out' primer.

There are two PCR steps in this method. In the first step, PCR was performed in 2 tubes with different set of primers, one using oligos 1 and IV and the other oligos II and III. The two DNA amplification products generated were then purified by agarose gel electrophoresis to remove the template DNA and primers that might otherwise interfere with the subsequent round of PCR. For this second PCR,

1 µl of each purified PCR product was added to a single tube using oligos I and II. The resultant PCR product was purified using Qiaquick PCR purification kit (Qiagen) and then digested using the appropriate restriction enzymes. The digested DNA was gel-purified and ligated into the expression vector pET21b. DNA strands without the mutation will have its restriction site 'knocked-out' and cannot be digested, while the restriction site on the mutated strand remained unaltered and can therefore be digested, thus only the mutant DNA which have the suitable digest cohesive ends can be ligated into the vector. This method is an inexpensive yet highly efficient method of generating mutants with >80% of the constructs sequenced containing the desired mutation.

Table 2.3 List of oligonucleotides used in site-directed mutagenesis of CD2.

Primers	***Oligonucleotide sequence***
T7 promoter	5'-TAATACGACTCACTATAGGG
T7 terminator	5'-GCTAGTTATTGCTCAGCGG
'knock-out' primer	5'-GAGCTCGAATTCGGATTCTTATTCGAGAATCCTCAAG
D28N	5'-GACTGATGATATTAACGAAGTGCGATGGGAGAG
E29D	5'-GACTGATGATATTGACAACGTGCGATGGGAGAG
E29Q	5'-GACTGATGATATTGACCAGGTGCGATGGGAGAG
E33Q	5'-GATGAGGTGCGATGGCAGCGTGGGAGCACCCTGGTT
E41Q	5'-AGCACCCTGGTTGCCCAGTTTAAAAGGAAGATGA
F21L	5'-CTGAACATCCCTAACCTGCAAATGACTGATGATATTG
F42Y	5'-CTTGGTTGCCGAAGACAAAAGGAAGATGAAGC
F49A	5'-AACGTAAGATGAAGCCTTACTTGAAATCGGGAGCAT
F49Y	5'- AACGTAAGATGAAGCCTTACTTGAAATCGGGAGCAT
F55Y	5'-GAAATCGGGAGCATACGAGATCTTAGCAAATGG
K43A	5'-GAATTTGCTCGTAAGATGAAGCCTTTTTTGAAA
L38A	5'-GAGAGGGGGAGCACCGCTGTTGCCGAGTTTAAAAG
R31A	5'-GGTGGCTTGGGAACGTGGGAGCACCCTGGTTGCCGA
Y81A	5'-CTATAATGTAACGGTAGCTAGCACAAATGGGAC

Heterologous protein expression in *E. coli*

Heterologous proteins can be expressed to a very high level in recombinant *Escherichia coli*, which make *E. coli* an ideal host for the production of large quantity of protein necessary for biophysical studies. However, for the purpose of studies of protein using heteronuclear NMR, it is necessary to produce proteins labelled with NMR-sensitive isotopes such as ^{13}C, ^{15}N, and ^{19}F. This requires the use of minimal or defined media so that only the desired isotope is uniformly incorporated into the protein at a high level. A number of protocols for protein expression were therefore developed to produce different isotopically labelled and unlabelled proteins in *E. coli*. All the protocols for expression, however, shared common procedures such as the transformation of the *E. coli* expression strains, growth in a suitable medium and induction of protein expression. At its simplest is the expression of unlabelled protein in rich media, with modifications and variations for isotopic labelling of proteins.

The expression vectors used were pT7, pET21b, pGEX-2T and pTrcHisB. The antibiotics of choice was carbenicillin disodium (Melford Laboratory) added to a concentration of 50 μg/ml in all the media used, for BL21(DE3) pLysS expression strain, 34 μg/ml chloramphenicol was also used. Other antibiotics were used as appropriate, for example 30 μg/ml kanamycin was used for selection of the PUBS plasmid which contains the *argU* gene. The protein was induced by the derepression of *lacI* by isopropyl thiogalactoside (IPTG, Melford Laboratories, Ltd.).

Expression of unlabelled protein

For each protein expression, the competent cells were transformed fresh with the vector containing the target gene. The cells were plated onto selective agar and incubated overnight at 37 °C. A few colonies were picked from the plate, and inoculated into 2 X 500 ml of LB in 2 litre baffled flasks. The cell cultures were incubated at 37 °C with vigorous shaking, and when the cell culture reached an optical density at OD_{600} of 1-1.5, IPTG was added to a final concentration of 0.1

mM. A further addition of IPTG that doubled its concentration was administered after 30 minutes. The cells were harvested 4 hours after induction.

$^{15}N/^{13}C$ labelling of protein

A colony of cells was picked from a freshly transformed plate and inoculated into 5 ml of LB medium. The culture was incubated at 37 °C until the cell density reached an OD_{600} reading of 0.6 - 1. 0.2 ml of the culture was then transferred to 100 ml of minimal media (MM in Table 3.2 in Chapter 3) which is a modification of M9 media supplemented with vitamins and micronutrients (Neidhardt *et al.*, 1974). The cells were incubated with shaking at 37 °C overnight and collected by centrifugation the next day. The pellet was resuspended in 2X 500 ml of MM, with $(^{15}NH_4)_2SO_4$ and ^{13}C glucose (or unlabelled glucose for ^{15}N labelling only) used as the nitrogen and carbon source respectively. The cell culture was incubated at 37 °C with vigorous shaking until the OD_{600} reached ~0.4, the cells were then induced initially with 0.025 mM IPTG, followed by another additions of 0.05 mM IPTG after an hour and a further addition at 0.1mM an hour later. Alternatively the cells can be induced at OD600 at ~0.8-1.0 with 0.1 mM IPTG. The cells were harvested 6 hours after the initial induction.

^{19}F labelling of CD2

For the fluorine labelling of protein, the procedure was essentially the same as that of the $^{15}N/^{13}C$ isotopic labelling of protein. The significant difference being that no labelled compound was used until the induction stage. The culture was allowed to grow until the cell density reached an OD_{600} measurement of 0.8-1.0, at which time the aromatic amino acids tyrosine, tryptophan and phenylalanine or their respective fluorinated analogues (all obtained from Sigma) were added to a final concentration of 50 mg/L. For ^{19}F-labelling of phenylalanine, phenylalanine was substituted by *o*-fluoro-$_{DL}$-phenylalanine; for the labelling of tryptophan, 5-fluoro-$_{DL}$-tryptophan was used. The inhibition of the synthesis may be achieved by the addition of glyphosate (N,N-bis-phosphonomethyl-glycine, Sigma) to a final concentration of 1 g/L (Kim *et al.*, 1990), or by feedback inhibition with the addition of 1mM – 4 mM of each

aromatic residues (but minus the amino acid to be labelled which was substituted by its fluorinated analogue used at a concentration of 36 mg/L) into the culture media (Brooks & Benisek, 1994). The culture was induced with 0.1 mM IPTG with more IPTG added after an hour to a final concentration of 0.2 mM. The cells were harvested 4-6 hours after induction. This method can give high levels of labelling, but if, for structural reasons, a lower amount of labelling is deemed desirable, a fraction of the labelled amino-acid added may be replaced by unlabelled one. However, it should be noted that the unlabelled amino acid is preferentially incorporated into the protein, therefore a smaller proportion of unlabelled amino acid should be used.

Protocols for CD2d1 purification

For the purpose of our studies, it is essential that the protein sample used should be sufficiently homogeneous (>95 % purity). A sensible protein purification strategy would be to remove as much contaminants as possible in the fewest steps. A common approach is to remove most of the unwanted proteins in the first purification step. The use of affinity chromatography with tags and fusion partners have been particularly successful in this respect. Different constructs of CD2d1 with and without affinity tag were used for the expression of protein that required different purification methods. However, all the purification procedures shared common features which will be described in this section. Apart from the first step in purification whereby the methods vary according the presence or absence of affinity tag or fusion partner, the procedures were essentially the same for all the constructs.

A litre of cell culture was induced and the desired heterologous protein expressed in *E. coli* as described earlier. The cells were pelleted by centrifugation at 5000 g (Sorvall RC-5B centrifuge, GS3 rotor) for 10 minutes and the supernatant decanted. The cells were resuspended in a buffer appropriate for each purification method, and lysed by passing the cell suspension twice through a chilled French press (American Instrument Company) at high pressure (1000 psi). The cell debris was then pelleted by centrifugation at 12,000 g for 30 minutes (Sorvall RC-5B centrifuge, SS34 rotor). The supernatant containing the CD2d1 was used for the first

purification step which varies for the different constructs and will be described separately later. After this first step, all the purification steps were done on the FPLC (Fast Protein Liquid Chromatography) apparatus (Pharmacia), and all the buffers used contained 0.5 mM EDTA and 0.1 mM PMSF. The eluate collected from the first step were pooled together and loaded onto a Superose 10 size-exclusion column previously equilibrated with 20 mM phosphate pH 5.4. The protein was eluted at a flow rate of 1.2 ml/min. The desired fractions were pooled and loaded onto an S-10 cation exchange column (Pharmacia). The protein was then eluted stepwise with pH - elution at pH 6.5, 7.5 and 8.5 with the protein eluting at pH7.5. Due to a propensity for CD2d1 to precipitate at high concentration, the protein were concentrated slowly using Centriprep 3 (Amicon) at ~400-800g (Heraeus Labofuge 400R centrifuge, rotor type 8179). This purification scheme is also applicable to most CD2d1 mutant constructs, however, due to a tendency for some mutant proteins (e.g. K43A) to precipitate at lower pH, it may be preferable to use the His-tag version of the proteins for their purifications which will be described later.

At every step of the purification process, the protein samples were monitored with discontinuous tricine-SDS PAGE. Protein concentration was quantified using UV-spectrometry (λ =280 nm) with extinction coefficient ε_{280} = 13940 M^{-1} cm^{-1} as determined from the method of Gill & von Hippel (1989). The purity of the protein samples used in NMR studies are assessed to be >99% pure by discontinuous SDS-PAGE.

Purification of native CD2

The cells were pelleted as describe before and then resuspended in 40 ml of pre-chilled buffer (50 mM sodium malonate, pH 5.2, 0.5 mM EDTA and 0.1 mM PMSF). The cells were then lysed and the cell debris pelleted as described previously. The cell lysate was then loaded onto a SP-sepharose column previously equilibrated with malonate buffer and then washed at a flow rate of 1.5 mil/min for 40 minutes. The protein was eluted with a NaCl concentration gradient with CD2d1 eluting at 0.1-0.2 M salt concentration. The fractions containing CD2d1 were then further purified by gel filtration and ion-exchange chromatography as described earlier.

Purification of GST-fusion CD2

The pelleted cells were resuspended in 40 ml of ice-cold PBS (140 mM NaCl, 2.7 mM KCl, 10 mM Na2HPO4, 1.8 mM KH2PO4 pH7.3) and lysed. After centrifugation, the supernatant was diluted with ice-cold PBS to 100 ml. 2 ml of 50% glutathione sepharose 4B slurry of beads (Pharmacia) in PBS was added to the supernatant and incubated at 4 °C for 1 hour. The beads were pelleted by centrifugation at 500g for 5 minute, the supernatant decanted, and the beads resuspended in 10 ml of PBS. This washing step was repeated two more times. The protein was eluted by adding 1 ml of GEB (10mM reduced glutathione, 50 mM Tris-HCl pH8.0) to the pelleted beads and incubating for 10 minutes. The elution step was repeated 3 times. The CD2 was cleaved from the GST moiety of the eluted fusion protein by incubating the eluent with 10 units of thrombin (Boehringer-Mannheim) for 2 hours at room temperature. The cleaved CD2 protein was then separated from the GST during the gel-filtration chromatography step.

Purification of His-tag CD2

The cells were lysed in Binding Buffer (20 mM Tris-HCl pH7.9, 5 mM imidazole, 0.5M NaCl). After centrifugation, a column containing Ni-chelate His-Bind resin (Novagen) was prepared and the cell lysate was passed through the column by gravity flow at a rate of 10 column volumes per hour. The column was then washed with 10 column volumes of Binding Buffer and 6 volumes of Wash Buffer (20 mM Tris-HCl pH 7.9, 60 mM Imidazole, 0.5 mM NaCl). The bound protein was eluted with 6 volumes of Elute Buffer (20 mM tris-HCl pH 7.9, 1M imidazole, 0.5 M NaCl). The His-tag may be removed at this stage by incubating the eluate with 5mg of carboxypeptidase (Boehringer Mannheim) for 1 hour at room temperature. The protein samples may then be further purified by gel filtration and cation-exchange chromatography. For some of the mutants, the His-tag may be left uncleaved, however, there is a difference in the binding properties of proteins with and without uncleaved His-tag in the cation exchange chromatography. For proteins with a His-tag, the gel filtration was performed at buffer pH of 6.5. The proteins were then loaded onto the S-10 column at pH 6.5, washed at 7.5 with the proteins eluted at pH 8.5. If necessary, the His-tag may then be removed by carboxypeptidase

and a further gel-filtration step can be used to remove the cleaved residues and proteases. Alternatively, these proteins may be purified using the same protocols used for native protein, the difference being that all the buffers are kept at pH 6.5 or higher.

Protein analytical techniques

Tricine SDS PAGE (Schagger & von Jagow, 1987)

The tricine SDS-PAGE was performed using the Mini-protean II gel electrophoresis apparatus (BIORAD). The gel plates were assembled onto the clamp as instructed and placed in the casting stand. 12% separating gel was prepared with the TEMED and APS added last to initiate polymerisation (Table 2.4). 3.7 ml was pippetted immediately in between the glass plates. 4% stacking gel was prepared and 1 ml pippetted directly onto the top of the separating gel. The comb was inserted and the gel allowed to set. When the gel has set, the clamps were assembled to form upper buffer chamber and place in the gel tank. Both the upper and lower chamber were filled with 1X electrophoresis buffer (10X electrophoresis buffer - 121 g/L Tris base (1 M), 179 g/L Tricine (1 M), 10 g/L SDS (1% w/v)). The protein sample was boiled in Solubilising Buffer (50 mM Tris HCl, 4% SDS, 2 mM EDTA, 1% β-mercaptoethanol, 1% glycerol, 0.1% bromophenol blue) for 5-10 minutes before being loaded into the wells. The electrophoresis was performed at 150 volt for 60 minutes.

After electrophoresis, the gel was the removed from the plates and the resolved protein bands visualised by incubating with Staining Solution (45% methanol, 10% acetic acid and 0.25% Coomassie Brilliant Blue R-250 in water) for 30 minutes. The gel was then destained by incubating with Destaining Solution (20% methanol and 7% acetic acid in water) until the staining of the non-protein components was removed.

Table 2.4 Acrylamide gel mix for tricine SDS-PAGE

	12% Separating Gel	4% Stacking Gel
Acrylamide solution[a]	7.2 ml	1 ml
Gel buffer [b]	10 ml	4 ml
Glycerol (50% w/v)	6.7 ml	-
H_2O	4.1 ml	6.95 ml
Ammonium persulphate (APS)	100 µl	40 µl
TEMED	25 µl	10 µl

[a] Acylamide solution – 100 ml H_20, 48 g acrylamide, 1.5 g bisacrylamide

[b] Gel buffer – 200 ml H_20, 72.66 g Tris base (3M) pH to 8.45 with HCl, 0.6 g SDS

Quantitation of protein in gel

The protein can be quantitated by the gel elution method (Fenner, 1975). Protein bands were resolved in using SDS-PAGE and stained with Coomassie Brilliant Blue R-250 solution as described above. The resolved band was excised, macerated and the dye eluted by incubating the gel slice in 5 ml of 25% pyridine in water at room temperature overnight. Quantification of the dye eluted was done at an absorbance of 605 nm in the range of 0.05 and 1.5 absorbance units. The percentage of total cell protein from *E. coli* lysates was estimated using this method by comparing background proteins to the band bearing CD2d1 protein.

The protein can also be quantitated using densitometry software. The image of the dried gel was acquired using the UVIdoc software in a UVItec gel documentation system and analysed using the UVIBand software. A slice of the gel containing the protein band was quantified and compared to the equivalent slice in gel without the expressed protein, and the difference between the two was taken as the amount of protein expressed. The amount of expressed protein is presented as a percentage of the total protein expressed.

Isoelectric Focusing

The isoelectric point of the protein was determined by using the Model 111 Mini IEF cell (Bio-Rad). The polyacrylamide gel containing Bio-lyte 3/10

ampholytes was prepared and placed into the IEF cell as instructed. The protein samples were desalted before loading onto the gel. The focusing was done at constant voltage initially set at 100 V, after 15 minutes the voltage was increased to 200 V for 10 minutes, followed by final step at 450 V for 50 minutes. The protein bands were visualised by first staining with 0.05% crocein scarlet, 27% isopropanol, 10% acetic acid, 0.04% Coomassie brilliant blue R-250, 0.5% CuSO4 in water, following by destaining with 12% isopropanol, 7% acetic acid, 0.5% CuSO4 in water, and a final destain using 25% isopropanol, 7% acetic acid. The stained gel was allowed to air-dry overnight at room temperature.

Mass Spectrometry

The mass spectrometry analyses were conducted using electrospray ionisation and MALDI-TOF (matrix-assisted desorption ionisation - time of flight) mass spectrometry. The samples were prepared to a protein concentration of 10 pmol/μl and desalted by dialysis. The analyses were performed by Dr. Mark Domin at the School of Pharmacy, University of London.

Analytical size-exclusion chromatography

Size exclusion chromatography (SEC) is widely used as a protein purification tools, and in recent years, it is also used as an analytical method for the determination of biomolecular weight and oligomerisation state of a biomolecule. The method does not have the accuracy or rigour of other methods such as analytical ultracentrifugation, the reasons being that hydrodynamic properties may vary between proteins and in different condition, and some protein may have a tendency to interact with or adsorb to the column matrix. However, it has the advantage of being quick and easy to use, and it can resolve presence of sub-populations of molecular species in physiological condition. The technique relies on the separation of proteins by size when moved through a column packed with porous bead of cross-linked polymers such as dextran or agarose. The pore size of the beads is large enough to allow small molecules to enter, but small enough to exclude the larger molecules. In this way larger molecules flow more rapidly through the column and emerge earlier

because a smaller solvent volume is accessible to them, while smaller molecules are retained on the column longer.

The SEC experiments were performed on BioCAD SPRINT Perfusion Chromatography System workstation (PerSeptive Biosystems) using a pre-packed silica (SiO_2) column (Bio-Sil SEC 125-5 column, Bio-Rad). The column was first rinsed with 40 ml of water followed by 10% methanol, and cleaned by three 100 µl injection of 20% DMSO with 10% methanol as the mobile phase. The column was then equilibrated with buffer at the appropriate pH (20 mM potassium phosphate, 150 mM NaCl). 100 µl of the sample was injected each time and the protein eluted at a flow rate of 0.5 ml/min. The protein was monitored by UV absorbance at a wavelength of 280 nm. Calibration standards (Bio-Rad) were run for all the varying buffer conditions in the experiments done.

Analytical ultracentrifugation

Analytical ultracentrifugation (AUC) is a method of characterisation of biomolecules and it is a particularly useful method for the determination of biomolecular association properties of protein (reviewed by Hensley, 1996; Rivas & Minton, 1993). This involved the use of the sedimentation equilibrium (SE) experiment in which the sample is place in an experimental cell, and centrifuge at a speed that allows an equilibrium distribution of protein to be established across the whole cell. The sedimentation equilibrium condition is attained when the protein sedimentation is equally opposed by diffusion. This resultant equilibrium distribution is dependent on the molecular weight of the protein and can be detected by absorbance optics.

The AUC was performed using an An60-Ti rotor in a Beckman Optima XL-A Analytical ultracentrifuge equipped with Xenon light-source absorbance optics. The proteins samples used were dialysed against AUC buffer (20 mM phosphate, pH 5.7) overnight. The protein samples were prepared at three different concentrations (100 µM, 50 µM, and 20 µM) and 110 µl of each was then loaded into the sample cells of an Epon charcoal-filled six-channel centrepiece. The three corresponding reference cells were loaded with 120 µl aliquots of the dialysate. The SE AUC experiments were performed at 25 °C at operating speed of 10-20,00 rpm, and ran for a period of

time (~12-24 hours) to allow the solute to reach equilibrium. The data were analysed using Microcal Origin (Version 3.78).

References

Brooks, B. & Benisek, W. F. (1994). Mechanism of the Reaction Catalyzed By Delta(5)-3-Ketosteroid Isomerase of Comamonas (Pseudomonas) Testosteroni - Kinetic- Properties of a Modified Enzyme in Which Tyrosine-14 Is Replaced By 3-Fluorotyrosine. *Biochemistry* **33**(9), 2682-2687.

Driscoll, P. C., Cyster, J. G., Campbell, I. D. & Williams, A. F. (1991). Structure of Domain-1 of Rat Lymphocyte-T CD2 Antigen. *Nature* **353**(6346), 762-765.

Fenner, C. (1975). Quantitation of Coomassie Blue Stained Proteins in Polyacrylamide Gels Based on Analyses of Eluted Dye. *Analytical Biochemistry* **63**, 595-602.

Gill, S. C. & von Hippel, P. H. (1989). Calculation of Protein Extinction Coefficients From Amino-Acid Sequence Data. *Analytical Biochemistry* **182**(2), 319-326.

Hanahan, D., Jessee, J. & Bloom, F. R. (1991). Plasmid Transformation of *Escherichia-Coli* and Other Bacteria. *Methods in Enzymology* **204**, 63-113.

Hensley, P. (1996). Defining the structure and stability of macromolecular assemblies in solution: The re-emergence of analytical ultracentrifugation as a practical tool. *Structure* **4**(4), 367-373.

Inoue, H., Nojima, H. & Okayama, H. (1990). High-Efficiency Transformation of *Escherichia-Coli* With Plasmids. *Gene* **96**(1), 23-28.

Ito, W., Ishiguro, H. & Kurosawa, Y. (1991). A general method for introducing a series of mutations into clones DNA using the polymerase chain reaction. *Gene* **102**, 67-70.

Kim, H. W., Perez, J. A., Ferguson, S. J. & Campbell, I. D. (1990). The Specific Incorporation of Labeled Aromatic-Amino-Acids Into Proteins Through Growth of Bacteria in the Presence of Glyphosate - Application to Fluorotryptophan Labeling to the H+-ATPase of *Escherichia-Coli* and NMR-Studies. *FEBS Letters* **272**(1-2), 34-36.

Neidhardt, F. C., Bloch, P. L. & Smith, D. F. (1974). Culture medium for enterobacteria. *Journal of Bacteriology* **119**(3), 736-747.

Rivas, G. & Minton, A. P. (1993). New Developments in the Study of Biomolecular Associations Via Sedimentation Equilibrium. *Trends in Biochemical Sciences* **18**(8), 284-287.

Sambrook, J., Fritsh, E. J. & Maniatis, J. (1989). *Molecular Cloning - A laboratory manual.* Second edit, Cold Spring Harbor Laboratory Press.

Schagger, H. & Vonjagow, G. (1987). Tricine Sodium Dodecyl-Sulfate Polyacrylamide-Gel Electrophoresis For the Separation of Proteins in the Range From 1-KDa to 100-KDa. *Analytical Biochemistry* **166**(2), 368-379.

Chapter 3

Optimisation of CD2d1 Expression

The biophysical studies of protein often require the production of large amount of the protein under analysis. For studies by NMR spectroscopy in particular, for a moderately sized protein, this may entail the use of 10 - 30 mg of highly purified protein. Furthermore, there is a need to isotopically label the protein with NMR-sensitive nuclei, and such labelling requires the use of costly isotopically labelled reagents. An expression system that allows for the high level expression of heterologous protein as well as the isotopic labelling within reasonable cost is therefore highly desirable. CD2 has been expressed in various cell-types and expression systems, for example on mammalian cell-lines such as Chinese hamster ovary (CHO) cell (Arulanandam *et al.*, 1994) and *Pichia pastoris* (Shiomi *et al.*, 1998). These expression systems allow a fully glycosylated CD2 to be expressed, but unfortunately these expression system fails to meet the criteria of cost. *Escherichia coli*, however, has been demonstrated to be successful in the high level expression of heterologous protein and is particularly cost-effective in the isotopic labelling of expressed protein. It was also found that it was possible to produce a soluble form of CD2d1 in *E. coli*, it was therefore chosen as the expression system for the over-production of isotopically labelled CD2d1.

CD2d1 was first expressed in *E. coli* as a fusion to the enzyme GST (glutathione-S-transferase) of *Schistosoma japonicum* (Driscoll *et al.*, 1991). *E. coli* however lacks the complex post-translation modification mechanism required for the glycosylation of proteins. Nevertheless it was found that the unglycosylated rat CD2 produced from E.coli has the same structural integrity as that of the glycosylated one (Driscoll *et al.*, 1991; Jones *et al.*, 1992; Withka *et al.*, 1993), and unlike human CD2, it does not require the glycans for stability or solubility. Furthermore, the presence of glycosylation is not required for the adhesion function of CD2, studies of protein expressed in *E.coli* therefore can be biologically relevant. The *E. coli* expression system produces a yield of fusion protein of in the region of 40 mg/L in rich media. In minimal media, and after cleavage to remove the GST moiety, a final yield of 5-10 mg/L of purified CD2d1 can be obtained (Crawford, 1994). In the interest of minimising the cost of production of isotopically labelled protein, a number of strategies for the optimisation of protein expression in *E. coli* were explored.

Heterologous Protein Expression in *E. coli*

Escherichia coli, a Gram-negative enterobacterium, has been one of most intensely studied organism. It is routinely used in molecular biology for the manipulation of genetic material as well as the production of heterologous proteins. Its popularity as a host for protein expression is due to a number of factors: it is well-understood; there exist a large number of commercially available cell strains and expression vectors that can cater to a wide variety of requirements, such as the need for protein fusion partners or secretion of target protein; it is relatively easy to manipulate; it is also cost effective, can express proteins to a high level and the entire process of protein expression can be completed in a comparatively short time.

A number of problems, however, may be encountered when attempting to express the desired heterologous protein to a high level. The protein may express poorly, or the protein may be insoluble or degraded. Not all the rules that govern the high level expression of protein in *E. coli* are known, and optimisation of the protein expression is therefore often a process of trial and error. Nevertheless, a number of

factors are known to be important for the high-level production of heterologous protein in *E. coli*. These include the structural features of the gene sequence, the stability and translational efficiency of mRNA, the ease of folding and degradation of protein by host cell proteases, codon usage differences, and the toxicity of the protein to the host (Makrides, 1996). Some empirical "rules" exist that can help guide the optimisation of protein expression in *E. coli*, these include the choice of transcription elements promoters which regulate the transcription of the gene, translational elements, fusion partners and fermentation conditions. However these "rules" are only very rough guide to the problem of heterologous protein expression in *E. coli* and each protein may require extensive investigation in order to optimise its yield. The use of *E. coli* expression system for high-level production of heterologous proteins and the optimisation of protein expression in *E. coli* can be found in a number of reviews (Baneyx, 1999; Gold, 1990; Hannig & Makrides, 1998; Makrides, 1996; Weickert *et al.*, 1996). Some of the factors involved in high-level protein expression in *E. coli* are described below.

Transcription elements

Transcription control is one of the most important elements in the expression of protein, and the expression level may be increased 1000-fold in uninduced cell to maximal induction in a well-regulated transcription system (Balbas & Bolivar, 1990). The transcription initiation event is the first and a rate-limiting step in gene expression. This initiation is mediated by the promoter sequence which may be regulated by variety of mechanisms. The most commonly used native promoters used in *E. coli* expression systems are *trp*, *lacUV5*, *lpp*, T7, λPL and λPR. Also frequently used are a number of hybrid promoters such as *tac* and *trc* which were constructed from different promoters to improve protein expression. For all practical purposes, it is not possible to examine the effect of transcription elements/promoters on protein expression apart from using commercially available vectors with promoters already optimised for expression and control. In the protein expression studies for CD2, four different expression vectors were tested for their effects on expression: pT7-7, pET-21b, pTrcHisB, and pGEX-2T, and the description of the promoters will be largely limited to those used.

The pTrcHisB and pGEX-2T transcription systems (Smith & Johnson, 1988) use the *trc* and *tac* promoters respectively (Amann *et al.*, 1983; Deboer *et al.*, 1983). Both *trc* and *tac* promoters are constructed from *lac* and *trp* promoters with only a single base difference between them. For protein expression, the *lac* promoter used is usually the *lacUV5* mutant which increases the efficiency of transcription from the promoter. Another useful property of this mutant is its insensitivity to catabolite repression by glucose, thereby allowing the target genes under the control of a *lac*-type promoter to be expressed in a minimal media in which glucose is used as the sole carbon source (Winnacker, 1987). A suitable promoter also requires the protein expression to be well regulated; the *lac* promoter is controlled by the gene product from the *lacI* gene which binds to the *lac* operator as a tetramer, thereby blocking the initiation of transcription. This *lac* repressor can be released from the DNA by the binding of IPTG (isopropyl-thiogalactoside) or by thermal-induction in the case of temperature-sensitive mutant. Tighter regulation can be afforded by the presence in the plasmid of extra copies of *lacI* gene (found in pET-21b) or its over-expressing $lacI^{Q}$ variant (found in pGEX-2T and pTrcHisB).

Both pT7-7 and pET-21b use the T7 promoter system which is driven by T7 RNA polymerase whose rate of polymerisation is five times faster than that of *E. coli* RNA polymerase (Studier & Moffatt, 1986). The expression of protein under the control of T7 promoter therefore requires cells with a chromosomal copy of T7 RNA polymerase gene such as BL21(DE3) (a lysogen of bacteriophage λDE3), or alternatively the T7 RNA polymerase gene may be introduced via transfection by bacteriophage such as λCE6. In this case, Bl21(DE3) was used as the expression strain. All four expression systems are inducible by IPTG; the difference between the T7 and *lac* promoter systems being that the T7 system is not directly induced by IPTG but via the induction of T7 RNA polymerase under the control of a *lacUV5* promoter.

The main difference between the pT7-7 and pET-21b vectors lies in the degree of regulation of protein expression. In addition to the regulation of the T7 RNA polymerase by *lacUV5* promoter, the T7 promoter of pET-21b is also fused to a *lac* operator which allows the promoter to be regulated by *lac* repressor. The protein expression using pET-21b vector is therefore doubly repressed, and any background expression is reduced significantly. Such tighter transcription control can be

beneficial as it prevent the selection of 'expression-down' mutation of the host-cell transcription/translation machinery, minimise plasmid loss, and avoid accumulation of mutations in the recombinant protein products themselves, particularly if those products are inherently toxic to the host.

Other transcriptional features that may influence protein synthesis are amino acid sequences which can result in attenuation or premature termination of the transcription. Anti-terminator elements such as *rrn*B are therefore often included in the expression system to ensure complete transcription of long messages (Li *et al.*, 1984). However, the presence of long stretches of A's or U's in the transcripts can nevertheless result in the slippage of the *E.coli* RNA polymerase during elongation (Wagner *et al.*, 1990). The same phenomenon has been demonstrated for T7 RNA polymerase for long stretches of U's (Macdonald *et al.*, 1993), phenomenon that can result in poor expression and premature termination of translation. While stretches of T's may be seen in CD2d1 DNA sequence it appear not to be long enough to cause any serious problem in protein expression, this aspect of transcription was therefore not explored in the expression optimisation studies.

The degradation of the product of transcription, the mRNA, can also affect the expression of the protein. The rate of the degradation can be affected by factors such as the rate of translation and mRNA secondary structures at the 5' and 3' ends. Transcription terminators placed after the coding sequence can substantially enhance the stability of the mRNA by forming strong hair-pin structure. An efficient transcription terminator can also improve protein synthesis by preventing transcription promoted by strong promoters from continuing into the replication region thereby destabilising the plasmids. Transcription terminators can also be placed upstream of the promoter, this has the effect of preventing transcription across the promoter region which can interfere with its function, as well as minimising background transcription.

Translation elements

A problem frequently encountered in protein expression is one that arises from the coding sequence of the gene itself. This includes the problem of translation

initiation and codon usage. The efficiency of the translation initiation is strongly affect by the presence of Shine-Dalgarno (SD) sequence and start codon that form the ribosome binding site (RBS), as well as sequences preceding and following the start codon (Gold & Stormo, 1990). A set of general rule for efficient translation initiation has been compiled (Gold, 1988; Stormo, 1986); however, no universally effective translation-initiation consensus sequence has been identified. Some sequence elements known as 'translational enhancers' upstream and downstream of the start codon are known to improve heterologous protein expression (Makrides, 1996), however, the mechanism of action is poorly understood and do not have universal application in *E. coli*. In pTrcHis B vector, the bacteriophage T7 gene ϕ10 translation enhancer is included in the N-terminal fusion peptide (Olins *et al.*, 1988).

The often unpredictable and inefficient translation initiation of eukaryotic mRNAs on bacterial ribosome can be obviated by the use of an N-terminal fusion protein such as glutathione S-transferase (GST) as in pGEX-2T vector. The GST is known to be able to express to a high level and translation initiation is therefore not an issue. This same problem of translation initiation is tackled in the pTrcHisB expression vector by using a dicistronic gene structure (Ito & Kurosawa, 1992; Schoner *et al.*, 1986). In this system, a mini-cistron which encodes a short peptide with favourable translation initiation DNA sequence is used to facilitate ribosome binding, while the target gene may be inserted in a second cistron that follows the first cistron. The dicistronic structure is designed in such a way that the translation of the second cistron begins before the ribosome can dissociate from the RNA after the translation of the first cistron has terminated.

For the expression of native protein without a fusion partner at the N-terminal region, however, some considerations about the translation initiation and codon usage are necessary. The presence of strong mRNA secondary structure at the 5' end can significantly reduced protein expression by occluding the SD sequence and the start codon (Gheysen *et al.*, 1982; Iserentent & Fiers, 1980), and reduction of such mRNA secondary structures can improve expression 100-fold (Balbas & Bolivar, 1990). The mRNA secondary structure, however, can be strongly influenced by the initial coding sequence of the gene. A gene with high GC content at the translation initiation site will produce more double-stranded helical structures that can reduce translation initiation. The codon usage pattern at the translation initiation region may

be partly related to mRNA secondary structure formation at the 5'end. Such mRNA secondary structure can be reduced by substituting G and C with A and T wherever possible. The second codon (after the start codon ATG) has also been demonstrated to have significant effect on protein synthesis (Cantrell *et al.*, 1991; Looman *et al.*, 1987), with GCTA and AAAA being preferred as the initiating sequence in highly expressed protein (Stormo, 1986). The initial coding sequence can be easily restructured during PCR cloning by the appropriate design of the oligonucleotides.

A different codon usage problem is encountered for the translation of the bulk of the protein. The genetic code is degenerate, and through evolution, different organisms have acquired different preferences in codon usage. This codon bias can be correlated with the abundance of their cognate tRNAs which may reflect the need by the cell to maximise the efficiency of the translation (Ikemura, 1982; Kurland, 1987). Unfortunately for the expression of eukaryotic gene in *E. coli*, eukaryotes exhibit very different codon usage pattern from *E. coli* (Zhang *et al.*, 1991). As a consequence, gene sequences can contain large number of codons rarely used in *E.coli.* These rare codons can have a number of deleterious effect on the protein expression, such as low level of protein synthesis and/or accumulation, premature termination of translation, frameshifting, as well as misincorporation of the amino acids (reviewed by Kane, 1995; Kurland & Gallant, 1996). In particular, AGG and AGA, the rarest codons in *E. coli,* have been demonstrated to have a pronounced effect in the quantity and quality of the expressed protein (Calderone *et al.*, 1996; Chen & Inouye, 1994; Day *et al.*, 1996; Forman *et al.*, 1998; Rosenberg *et al.*, 1993; Spanjaard & Vanduin, 1988). Such effect on protein expression can be observed most clearly for proteins in which these rare codons are present in tandem or in cluster (Ivanov *et al.*, 1992; Rosenberg *et al.*, 1993; Spanjaard & Vanduin, 1988). Two strategies can be employed to overcome this problem. The gene sequence may be systematically altered, either by removing specific rare codons by site-directed mutagenesis (SDM), or chemically synthesising the entire gene using the set of codons preferred by *E. coli* (Rangwala *et al.*, 1992). Alternatively, the cognate tRNAs for the rare codons can be co-expressed with the target gene in order to alter the population of the tRNA species. Such approach has been successfully attempted using *argU* (*dnaY*) gene which express the cognate tRNA for AGG/AGA (Brinkmann *et al.*, 1989). Other rare codons that have been shown to affect the

expression of heterologous proteins include AUA, CUA, CGA/CGG and CCC. The deleterious effect due to the rare codons can also be seen for more common codons if the fractional amount of such codons is high in a gene (Bula & Wilcox, 1996; Kane, 1995; Kurland & Gallant, 1996).

The peptide chain is released from the ribosome at the chain termination signal which can be encoded by three different stop codons. These stop codons differ in their efficacy of translation termination, and a strong bias toward UAA is seen in gene expressed at high level. A common strategy for efficient termination of translation is the use of tandem termination codons.

Post-translational events

A number of post-translational events can affect the expression level of the protein. Proteolysis, for example, can strongly influence the degree of accumulation of the expressed protein (reviewed by Enfors, 1992). Some of the structural determinants for protein instability have been defined, for example, in the 'N-terminal rule', certain amino acids such as arginine, lysine, leucine, and the aromatic residues at the N-terminus can result in the rapid degradation of the protein (Tobias *et al.*, 1991). This 'N-end rule', however, need to be considered in conjunction with the effect of the second residue on the removal of the N-terminal methionine (Hirel *et al.*, 1989). The removal of the N-terminal methionine is efficient when the second residue is small and uncharged (such as alanine), but not if the second residue is bulky and charged (e.g. arginine). The second residue becomes the N-terminal residue when the N-terminal methionine is removed. A leucine is therefore considered undesirable as a second residue as its N-terminal methionine can be removed (albeit not as efficiently as for glycine or alanine) and may therefore be preferentially degraded according to the 'N-end rule' once the N-terminal methionine is removed. Protein degradation due to other proteases may be avoided by using host cells deficient in proteases (for example BL21 which is deficient in OmpT); however, some proteases-deficient strains such as the Lon^- mutants have their own drawbacks such as poor growth or formation of filamentations.

Other factors such as the misfolding of protein and presence of unstructured sequences may also increase the degradation. These may be overcome by expression

in protease-deficient cell strains, secretion into the periplasmic space or into the culture media, or by driving the expressed protein into inclusion body. The formation of inclusion body can be beneficial as it can simplify the purification process, prevent degradation, improve expression and side-step any problem of protein toxicity. This strategy, however, depends crucially on how well the protein can be refolded (reviewed in Lilie *et al.*, 1998). Other remedies include growing the cell at low temperature, coexpression of chaperones such as GroEL and GroES, as well as using fusion partners which can improve the solubility and stability of the protein by perhaps functioning *in vivo* as covalently-linked 'chaperones', thereby helping the protein to fold properly.

Fermentation conditions

Optimising the growth and induction conditions can also significantly improve the protein production by maximising the efficiency of the system. The growth rate of the cells can be affected by physical factors such as temperature and aeration rates, as well as chemical factors such as media composition and pH. The growth rate, however, influences the plasmid stability - a high growth rate tends to reduce the copy number of the plasmid which can lead to segregational instability whereby the plasmid may be lost from the cells. Moreover, optimal culture condition for protein expression is not necessary compatible with optimal growth rate. It is therefore useful to moderate the growth rate for optimal expression of the protein. It is generally undesirable to limit the nutrients available, for example, the depletion of the nitrogen source can induce the 'stringent response' which increases the synthesis of proteases. The modulation of the cell growth during induction by other methods such as lowering the incubation temperature is therefore preferable.

A higher yield of the protein can be obtained from a higher cell-density, it is therefore desirable to find a suitable condition for obtaining maximal final cell-density before harvesting the cells. Factors such as nutrients, temperature, pH and induction condition can affect the final cell-density obtained.

Overview of Optimisation of CD2d1 Expression

As previously described, large number of factors can contribute to a high-level and efficient expression of the protein. In the case of CD2d1, it is not possible to examine all the possible variables, therefore only those that can be investigated within reasonable time and effort were considered. CD2d1 is soluble, with no significant folding or proteolysis problem, and there is also no sign of toxicity, all of which suggests that it should be possible to express CD2d1 to a high level in *E. coli.* A number of factors that may influence the expression level are investigated: a) promoter/expression vector, b) translation initiation region/codon usage, c) growth and induction conditions.

Expression vectors

As described earlier, four different vectors - pT7-7, pET21b, pTrcHisB, and pGEX2T were used for the expression studies. The CD2d1 in pGEX-2T had previously been constructed (Driscoll *et al.*, 1991), while the pTrcHisB-CD2 construct is a gift from Dr. Simon Davis. The CD2d1 gene therefore only needed to be cloned into the pT7-7 and pET21b vectors. The pET21b vector also allows a C-terminal His-tag version of the protein to be constructed which may be useful for the purification of the protein. The various cloned CD2d1 constructs were transformed into BL21(DE3) and tested for expression.

Codon usage

The gene from CD2d1 begins with a codon coding for arginine, AGA. AGA, one of rarest codon, is known to have deleterious effect in protein expression, and presence of rare codon nearer to the 5' end can have a stronger effect on the translation and protein quality (Calderone *et al.*, 1996; Day *et al.*, 1996). It would therefore be advantageous to remove this rare codon and replaced with a more favourable one such as CGT. However, as mentioned earlier, the codon usage pattern is different at the translation initiation region (Bulmer, 1988). Rarer codons may be found more frequently at the translation initiation region and codons with

higher AT content may be preferred, in which case the rarer codon AGA would be preferable to the more common CGT. In order to investigate the effect of this codon usage, 2 constructs were prepared with the two different codons as the second codon. It is also possible to create more conserved mutants at the translation initiation sites to reduce the GC content, however, an examination of the sequence does not reveal any further modification that may be necessary.

CD2d1 has two AGA and three AGG codons which represent 5% of total codons. These rare codons are dispersed throughout the coding sequence and not clustered, they are therefore not expected to have a significant effect on the expression level. Nevertheless it is possible that the presence of these rare codons may affect the quality of the protein produced due to misincorporation. For comparison, CD2d1 was therefore co-expressed together with the *argU* gene which would increase the population of the $tRNA^{arg(AGG/AGA)}$ species and can block the misincorporation of amino acid at the AGA/AGG sites as well as improving protein expression (Forman *et al.*, 1998; Brinkmann *et al.*, 1989). The PUBS plasmid (a gift from Dr. John Hinks) was used for coexpression of the *argU* gene (Brinkmann *et al.*, 1989).

Fermentation conditions and protein induction

Various aspects of the fermentation as well as the induction condition were examined. As the isotopic labelling of CD2d1 requires the use of minimal media, three different minimal media were tested as culture media. In addition to the usual M9 medium (Sambrook *et al.*, 1989), the MOPS (Neidhardt *et al.*, 1974) and MJ (Jansson *et al.*, 1996) minimal media were also tested for cells growth and expression. The induction of protein expression can have significant effect on the cell growth and inductions at varying time were therefore compared.

Results and Discussion

Generation of recombinant gene constructs

The CD2d1 was cloned into the pT7-7 and pET-21b expression vector using the PCR cloning protocol as described in Chapter 2. The CD2d1 construct in pGEX-2T was used as the template for the PCR cloning. Three oligonucleotides were used for the generation of all the constructs. The oligonucleotides were designed such that the PCR products may be ligated into both pT7-7 and pET-21b vectors, as well as allowing the gene to be cloned as in the form of native protein in pT7-7/pET21b, or be fused to a His-tag in pET21b. Two forward oligonucleotides were synthesized, one with AGA as the second codon and the other with CGT. The second codon is preceded by the *Nde*I restriction site which allows the PCR products to be cloned into both pT7-7 and pET-21b vectors. The CD2d1 gene coding sequence ends with the nucleotide sequence CTAGAG which conveniently can be changed into CTCGAG by incorporating this mutation in the reverse primer. This conserved mutation forms the *Xho*I restriction site which enables the gene to be ligated to the *Xho*I site in the pET-21b vector, thereby fusing the gene to a C-terminal His-tag. The reverse primer also contains a stop codon (TAA), followed by a *Bam*HI restriction site that allows the PCR product to be ligated into both pT7-7 and pET-21b. In this way, the three primers synthesized permit six different constructs to be made: pT7-AGA, pT7-CGT, pET-AGA, pET-CGT, pET-AGA/His, and pET-CGT/His. The CD2d1 was initially cloned into pET-21b vector only as His-tag fusion protein and into pT7 vector as a native protein. However, it was found that the presence of a His-tag may influence the expression level, the CD2d1 was therefore sub-cloned from pT7-CD2d1 constructs into pET21b as native protein (Fig. 3.1). All the recombinant gene constructs were sequenced to ensure the correctness of the plasmid construction and genetic sequence (Fig. 3.7).

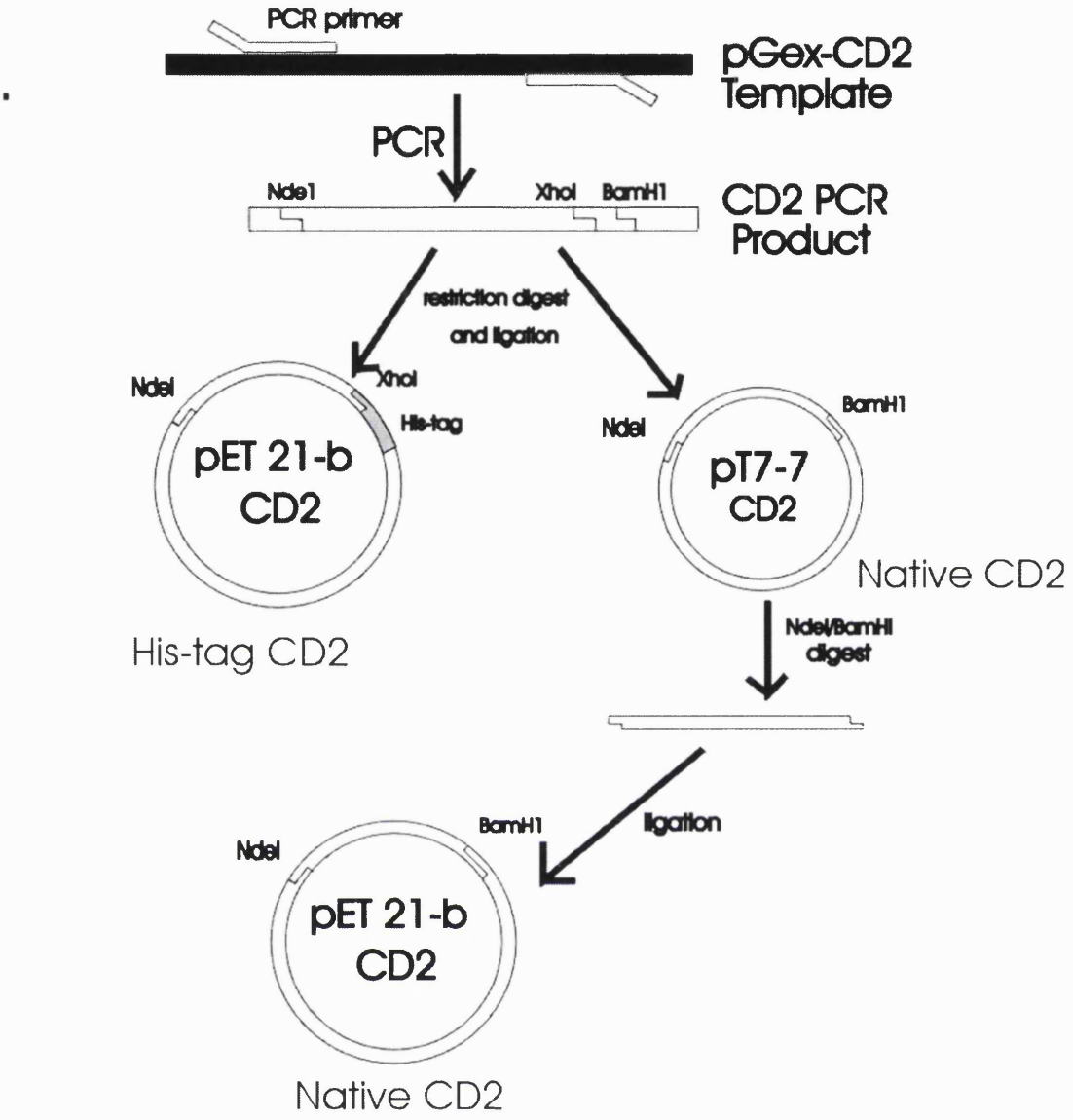


Fig. 3.1 Overview of the construction of the CD2d1 expression vectors.

Comparison of protein expression

The different CD2d1 expression vector constructs were transformed into BL21(DE3) cells and were cultured in rich media (LB) as well as minimal media (M9). The protein was induced with 0.1 mM IPTG. The results of the expression tests using the various constructs were shown in Table 3.1. It should be noted at this point that the results listed are values from densitometry measurement of protein gels (Fig. 3.2). In general the protein quantitation at higher protein concentration using this method is less reliable (Fenner, 1975), and highly expressed protein may therefore be under-estimated. The values were not corrected for such effect, but

nevertheless it can give a good indication of the relative performance, and repeat of the quantifications for some of the expression tests in SDS-PAGE gel using gel elution method produced a broad agreement with the densitometric measurement. The experimental conditions were not optimised for protein expression.

Table 3.1. Results of the expression studies using different CD2d1 constructs.

	Protein yield (% of total cell protein)	
Expression plasmid construct	**Rich Media**	**Minimal media**
PT7 CCT	5.7	6.3
PT7 AGA	10.0	9.1
pET CGT His-tag	12.7	7.5
pET AGA His-tag	16.2	10.3
pET CGT	15.7	9.8
pET AGA	16.6	13.7
GST-fusion	9.9	7.7
pTrc-HisB	3.5	3.2

The data presented is the fractional amount of CD2d1 expressed four hours after induction using 0.1 mM IPTG. Each value was obtained as the average of 2 separate set of data from densitometric analysis, with the exception of pTrc-His A construct where only 1 set of data was used due to lack of expression in other expression test.

Choice of vectors - The results in Table 3.1 showed that the pET CD2 construct produced the highest expression level in both rich and minimal media. Although the pT7 construct can express as well as the pET construct, the result was not always reproducible. This was especially noticeable for cell cultures that have been left overnight when markedly less accumulation of protein may be observed. This unreliability may be a result of the lack of tight control, and the increased metabolic load from higher amount of protein expressed constitutively may lead to segragational instability of plasmids. Very poor expression was obtained with pTrcHisB construct, with some expression tests showing no discernible protein expression. This may be due to the presence of $lacI^{Q}$ in the vector which over-expresses the *lac* inhibitor, a greater amount of IPTG is therefore required for the same level of induction. However, attempt at induction using higher IPTG concentration failed to show significant improvement. In the data shown in Table

3.1, the pGEX2T CD2 construct produced a lower expression level of the GCT-fusion protein, however, the yield can be increased significantly to the level achieved by the pET expression vector with longer period of incubation, especially if the culture has been left overnight. However, the GST moiety of the protein represents more than two-third of the bulk of the protein, the final yield of CD2d1 is therefore correspondingly a third lower. This result suggests that the pET21b should be the expression vector of choice for CD2d1 as it allows for high level of protein expression as well as tight regulation of the protein expression.

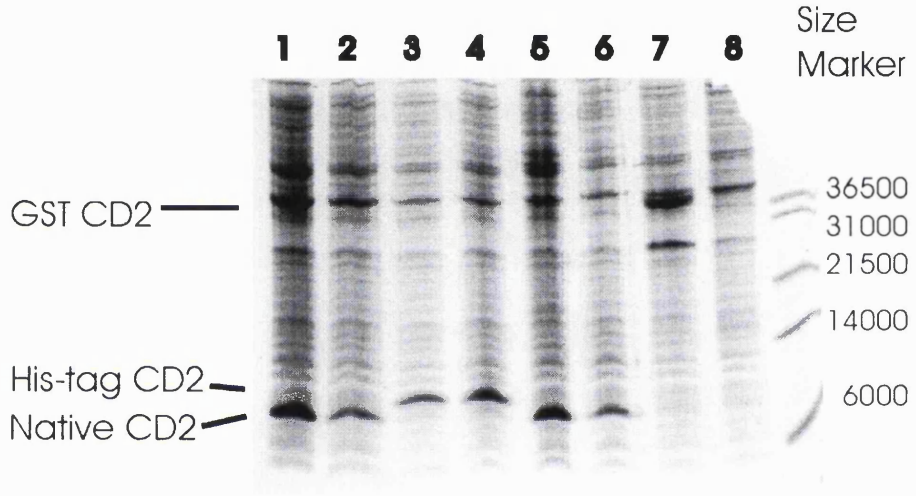


***Figure 3.2**: Protein expression test in M9 minimal media. Lane 1 – pT7 AGA, lane 2 – pT7 CGT, lane3 – pETAGA/His, lane 4 – pET CGT/His, lane5 pET AGA, lane 6 – pET CGT, lane 7 – pGEX-2T, lane8 pTrcHisA. The positions of the expressed heterologous proteins are indicated. The protein expression was induced at OD600 of 0.6 with 0.1 mM IPTG.*

Effect of His-tag - there is a surprising effect from the attachment of a C-terminal His-Tag to CD2d1. Expressions of proteins with His-Tag were generally lower in both the AGA and CGT constructs with the expression level at ~ 80% of the native CD2d1. This is likely to be due to the 'hungry codon syndrome' in which large number of the same codons, especially if they are present in consecutive sequence as can found in the His-tag, can deplete the charged t-RNA species, thereby affecting the rate of translation in the process (Kurland & Gallant, 1996).

Effect of second codon - The effect of the codon usage and mRNA secondary structure were examined in this study using DNA constructs containing CGT or AGA as the second codon. In general, the percentage yield of the protein expressed with

CGT as the second codon position is lower that those with AGA, with the pET-CGT construct producing ~75% less protein than that for the pET-AGA construct. This lower protein yield appears to be the consequence of a slower rate of protein synthesis.

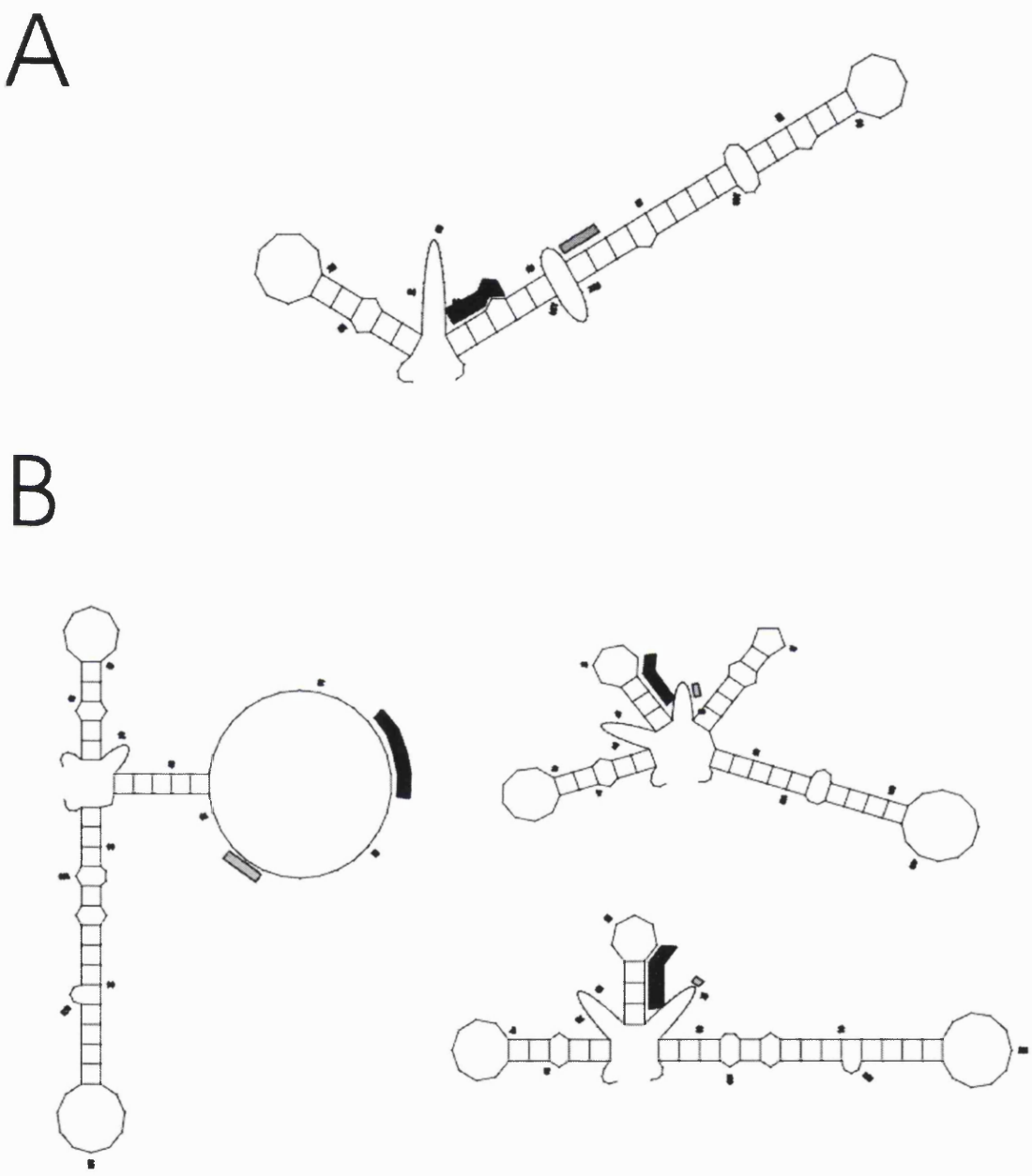


Figure 3.3: Secondary mRNA structures of translation initiation translation regions CD2d1 constructs - A) CD2-CGT B) CD2-AGA. Black bar indicates Shine-Dalgarno (SD) sequence, grey bar indicates start codon ATG. The optimal and sub-optimal structures were calculated using the MFold algorithm (Jaeger et al., 1990; Zuker, 1989) with structures within 2 Kcal/mole of the minimum free energy shown. The secondary structures were displayed using the Plotfold programme. Both programmes can be found in the GCG package of programmes.

The observations can be rationalised as the result of an increase in secondary structure at the translation initiation region (Fig.3.3). The formation of strong helical double-stranded structure at the translation initiation region can obscure the SD sequence and start codon and prevent efficient ribosome binding and translation initiation (Jaeger *et al.*, 1990). An easily accessible start codon in a single-stranded region in particular is important for efficient translation (Iserentent & Fiers, 1980). The rate of protein synthesis may be determined in part by the speed of decoding the mRNA by ribosome which may be in turn dependent on the abundance of tRNA species (Berg & Kurland, 1997), and the presence of AGA/AGG within the first 25 codons has been suggested to have a regulatory role and can limit protein synthesis under stress condition (Chen & Inouye, 1994). However, for CDd1, the translation initiation process appears to be the rate-limiting process with the abundance of particular tRNA species being of secondary importance. The optimal structures for the sequence folds of the translation initiation region for pET-AGA and pET-CGT constructs are shown in Fig. 3.3. Only one optimal structure was obtained for pET-CGT within 2 Kcal/mol of the minimum free energy while five were obtained for pET-AGA (only three shown in Fig3.3). The RBS and start codon can be observed to be sequestered in the pET-CGT construct, while in the pET-AGA construct, the SD sequence and ATG are partially or completely exposed. The pET-AGA construct therefore permits more efficient translation initiation and this is reflected in the greater amount of accumulated protein.

Effect of induction of protein expression on growth - The time at which the protein synthesis is induced can have a significant effect on the final yield of the protein. As can be seen in the Fig.3.4, for protein expression in rich media, the final cell density was higher for cells containing vectors using *lac*-based promoter (pGEX2T and pTrcHis B), while expression with vectors containing T7-based promoter (pT7-7 and pET21b) produced a lower final cell density. For T7-based promoters synthesis, the greater rate of transcription from T7 RNA polymerase dominates the protein synthesis machinery with the synthesis of host-cell proteins severely affected. The cells therefore would not show further significant growth beyond a single doubling in cell density following induction. The *lac*-based

promoter appears to be less deleterious to the host cell and can permit the further continuing growth of cells.

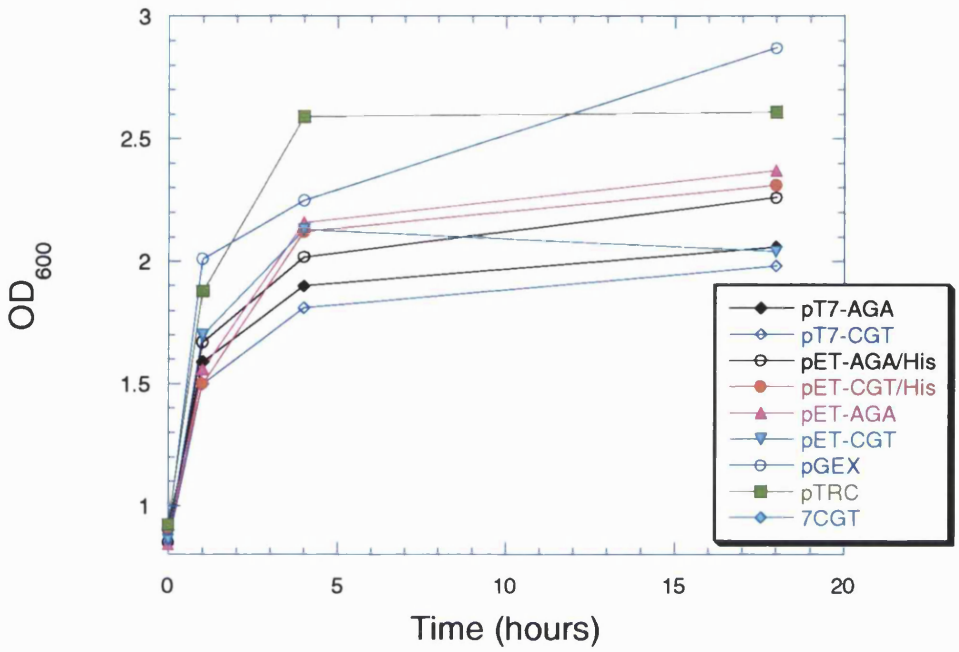


Fig. 3.4 Graph of cell density against time after induction with 0.1 mM IPTG for cell culture in rich media.

The effect on cell growth by induction can be more clearly seen in expression in minimal media. As can be seen in Fig 3.5, the final cell density depended on the cell density at which the cells were induced. Early induction resulted in significantly lower final cell density than a later induction irrespective of the promoters used. This is perhaps a reflection of the fact that the cells are more sensitive to the culture and induction condition in minimal media due to the greater metabolic load. The proportion of CD2d1 expressed also did not appear to be affected by the time of induction. The results therefore suggest that a later induction is preferable to early induction in minimal media as a higher final cell density would obviously produce a higher yield of protein. Other observations also suggest that a strong induction with higher concentration of IPTG having a stronger effect on the growth of cells. An alternative strategy for the induction was therefore developed. In this method the induction begins early at OD_{600} at 0.3-0.4 with very low concentration of IPTG of

0.025 mM added, further addition of IPTG can then be applied for the next two hours, with the amount of IPTG added doubling every hour. This allows the cells to be mildly induced without affecting the growth rate at the exponential growth phase, but towards the later stage of the exponential growth phase the cells can be more strongly induced.

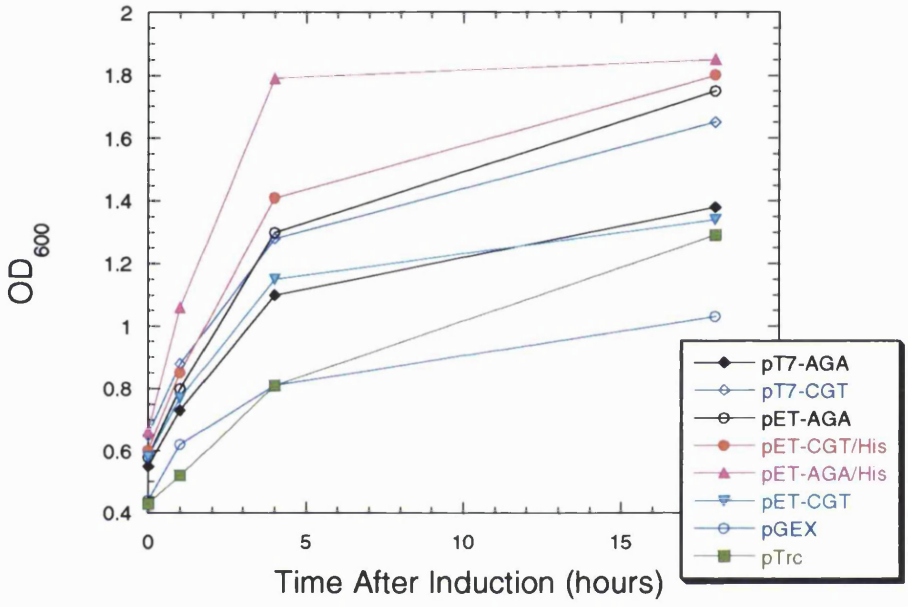


Fig. 3.5 The graph of cell growth after induction with IPTG in minimal media.

Choice of Minimal Media - The three different media used in this study were M9, MJ and MOPS (Table 3.2). The cell growth in MJ medium proved to be slow with a very long lag phase, possibly as a result of the higher amount of micronutrients used at a level which may be inhibitory for cell growth. Even though the cell growth was vigorous at the exponential phase, this growth medium was not used for further tests. Cells grown in MOPS medium showed faster growth than those in M9 as can be seen in Fig. 3.6, with greater final cell density due to induction at higher cell density. The cells grown in MOPS medium, however, appeared to show poorer expression level for CD2d1. This may be due to some unidentified inhibitory effect on protein expression by constituents of the medium, or more probably, the faster rate of growth resulted in greater segregational instability of the plasmids at the higher growth rate. Such segregational instability, however, can be avoided by a careful maintenance of

the selective pressure. MOPS medium would therefore appear to be the minimal medium most favourable for the cell growth, however, it was also found that performance close to that of MOPS may be achieved in M9 medium when supplemented with vitamins. In view of the reduced level of protein expression in MOPS medium and the uncertain reason for such reduction, a minimal medium which is essentially M9 supplemented with micronutrients from the MOPS medium and vitamins from MJ medium was devised and used for the expression of CD2d1.

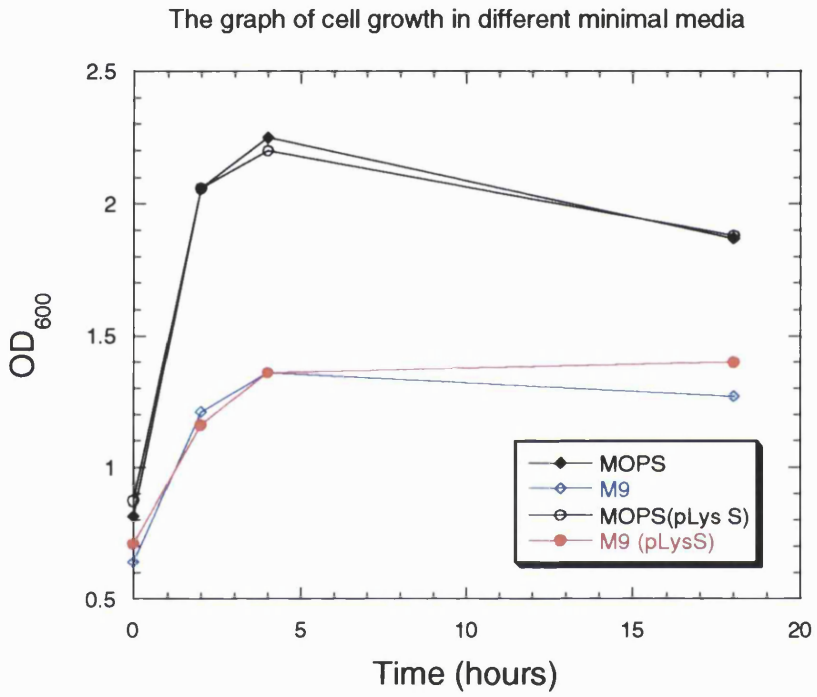


Fig 3.6 Growth of the cells in different minimal media after induction.

As mentioned earlier, the segregational instability of the plasmid may be increased at a high growth rate which requires the careful maintenance of the selective pressure. This may be partly achieved by using carbenicillin, an analog of ampicillin with a bulky head group that renders it less hydrolysable by β-lactamase. β-lactamase may also be released into the culture media, therefore in order to ensure that little β-lactamase is carried over into the fresh media, the overnight culture was spun down and the media decanted before the cells were transferred to the fresh media. The overnight culture was grown in minimal media at a culture volume such that when the cells are spun down and inoculated into fresh media, the initial cell density in the fresh media is high enough for the cells to be induced after at most two

doubling period of the cells. This is to ensure that at the time of induction, most, if not all, of the cells carry the plasmid with the target gene.

Table 3.2: Minimal media used for protein expression.

	M9[a]	MOPS[b]	MJ[c]	MM[d]
Na_2HPO_4	6			6
K2HPO4		0.132 mM	9	
KH_2PO_4	3		6	3
MOPS [e]		40 mM		
Tricine [f]		4 mM		
$(NH_4)_2SO_4$	1		2.5	1
NH4Cl		9.52 mM		
NaCl	0.5	50 mM		0.5
Glucose	2	2	2	2
$MgSO_4$	0.5	0.5 mM	1	0.5
$CaCl_2$	0.01	0.5 µM		0.5 µM
$FeSO_4$ [g]		0.01 mM		0.01 mM
K_2SO_4		0.276 mM		
Sodium Citrate [I]			0.5	
Trace elements				1 ml
Vitamins			0.65 ml	1 ml
Thiamine			0.07	

Trace elements and Vitamins at 1000X concentration

Trace Elements (1000x)	MJ	MOPS	Vitamins (1000x)	
$FeCl_3.6H_2O$	16.2	-	Pantothenic acid	0.5
$ZnSO_4.7H_2O$	2.4	10 µM	Choline Chloride	0.5
$CoCl_2.8H_2O$	4.2	30 µM	Folic acid	0.5
(NH4)6 MnO4.$2H_2O$	4.2	3 µM	Myo-inositol	1.0
$CuSO_4.5H_2O$	4.8	10 µM	Nicotinamide	0.5
H_3BO_3	1.2	400 µM	Pyridoxal HCl	0.5
$MnSO_4$	3.0	80 µM	Riboflavin	0.05
HCl	30 ml	-	Thiamine HCl	0.5
			Biotin	0.8
H_2O	570 ml	1000 ml	H2O	1000 ml

All the units in the table are in grams unless stated otherwise. [a] *The $MgSO_4$ is autoclaved separately added to the medium before use.* [b] *The medium used is filtered sterilised c The medium without $MgSO_4$ and vitamins is adjusted to pH 6.6 before autoclaving. The $MgSO_4$ is autoclaved separately and added together with vitamins to the medium after autoclave. The citrate is used as a metal chelater and is known not to be transported into E.coli.* [d] *The MM is the minimal media devised for the expression of CD2d1 which is a modification of the M9 with added vitamin supplements and micronutrients.* [e] *The MOPS buffer is prepared as 1M stock solution and pH to 7.4 with KOH.* [f] *The tricine is prepared as 1M solution pH to 7.4 with KOH.* [g] *The $FeSO_4$ is prepared fresh on day of use.*

The stability of the plasmid can also be increased by a tightly regulated expression such that any background expression is kept to a minimum. The BL21(DE3)pLysS cell strain was therefore also tested in minimal media (Fig. 3.6). The pLysS produces lysozyme which has an inhibitory effect on T7 RNA polymerase, and can therefore reduce background expression. No advantage, however, can be observed in the use of pLysS cells as a high degree of control can be achieved with pET21b expression vector alone. Moreover, the cells produce a protein, possibly the lysozyme, which co-purified with CD2d1 and proved difficult to remove. These cells were therefore not used further.

As mentioned earlier, an optimal growth rate may be incompatible with optimal protein expression. The growth rate can be modulated by lowering the temperature to 30°C after induction. Maximum amount of CD2d1 expressed was reached after 4-6 hours and the cells were then harvested.

Purification Schemes of CD2d1

Different purification schemes were necessary for the different proteins expressed using the different DNA constructs. The GST fusion protein was purified using standard protocol as described in Chapter 2 without the need to develop a separate purification strategy. Similarly the His-tag proteins were purified using an established protocol for metal-chelate resins, however, a considerable amount of contaminants was present after the metal-chelate purification step. The C-terminal His-tag was removed by incubation with carboxypeptidase followed by a purification step using gel-filtration chromatography. A further cation exchange was necessary to remove most of the contaminants. From analysis using SDS-PAGE and mass spectrometry, it was found that the six histidines of the His-tag cannot be efficiently removed, possibly as a result of steric hindrance from the His-tag being situated too close to the bulk of the protein. The purified protein contained at least an extra histidine attached. Since the CD2d1 will be used for the study of ionisation of a residue with a pK_a in the range of 6-7, the extra histidines may interfere with such analysis, the His-tag CD2d1 construct was therefore not used for further studies.

For the native CD2d1, it was necessary to develop a custom purification scheme. A useful property of CD2d1 that can be exploited for its purification is its

high pI value which is ~8.2 as determined by isoelectric focusing. Most of *E.coli* protein has low pI and large amount of proteins precipitate at pH value below 5.0, therefore in principle, large amount of *E.coli* proteins can be separated from CD2d1 by lowering the pH down to ~ pH 4.0-4.5. In this case, due to concern that CD2d1 may co-precipitate with the *E.coli* protein, cation exchange chromatography, in which most of the *E. coli* will not be bound, is used. The cells were lysed in malonate at pH value ~5.2, and the supernatant loaded onto a SP column after centrifugation to remove cell debris. The proteins were then eluted with a salt gradient. As will be seen in a later chapter, the CD2d1 in fact has a greater tendency to precipitate between pH 5.0 and pH 6.0 , therefore using precipitation of bacterial proteins at a lower pH may be useful. After the ion-exchange chromatography, the eluate was buffer-exchanged into phosphate buffer and the proteins separated by size (as well as from the salt used for elution during cation exchange) using size-exclusion chromatography. CD2d1 obtained after this stage was reasonably pure, with only two bands of contaminants visible. These contaminants can be removed by another cation exchange step, in which the proteins were eluted stepwise at pH 6.5, 7.5 and 8.5, with the CD2d1 eluting at pH 7.5.

The purified CD2d1 is stable, and is not significantly affected by proteases, however, after storage at 4°C for long period, a small amount of proteolysis can nevertheless be observed. This can be remedied by the addition of 0.5 mM EDTA and 0.1 mM PMSF.

Quantity vs. Quality

The protein expression was done at a relatively low IPTG concentration 0.1-0.2 mM. It was found that increasing the IPTG concentration can significantly increase the amount of protein expressed, however, the protein appeared to be less homogeneous at such high expression level. For example, during purification using the cation exchange chromatography, significant amount of CD2d1 were eluted isocratically in the wash, an effect not seen for protein expressed using a lower concentration of IPTG. The CD2d1 that was eluted in the wash did not bind well to the cation exchange resin, suggesting a difference in the overall charge status of the

protein. Isoelectric focusing also showed protein bands with different pI from CD2d1 purified from GST fusion protein.

The purified protein samples were analysed using mass spectrometry. As expected, the purified native CD2d1 contained an N-terminal methionine residue (relative molecular mass of 11272.5 from the electrospray ionisation mass spectrometry data compared to 11273 calculated for CD2d1 with an N-terminal methionine). From the MADLDI-TOF mass spectrometry, the protein appeared to be consisted of a single species (Fig. 3.7). More accurate analysis, however, may be obtained from electrospray ionisation mass spectrometry (EIMS). In this analysis, the CD2d1 was expressed in cells containing the PUBS plasmid which allowed the cognate tRNA for the rare codons AGA/AGG to be co-expressed. Purified protein samples were obtained and analysis of the protein samples using EIMS suggests that when CD2d1 is coexpressed with the *argU* gene, less heterogeneity, though not significantly so, of the protein may be obtained (Fig. 3.8). Even then, the data nevertheless suggest there are still some degree heterogeneity of the protein which is of uncertain origin. It was not possible to determine the origin of the heterogeneity in this limited analysis and the analysis may also have been affected by some amount of proteolysis observed as the degradation products can interfere with the analysis of data.

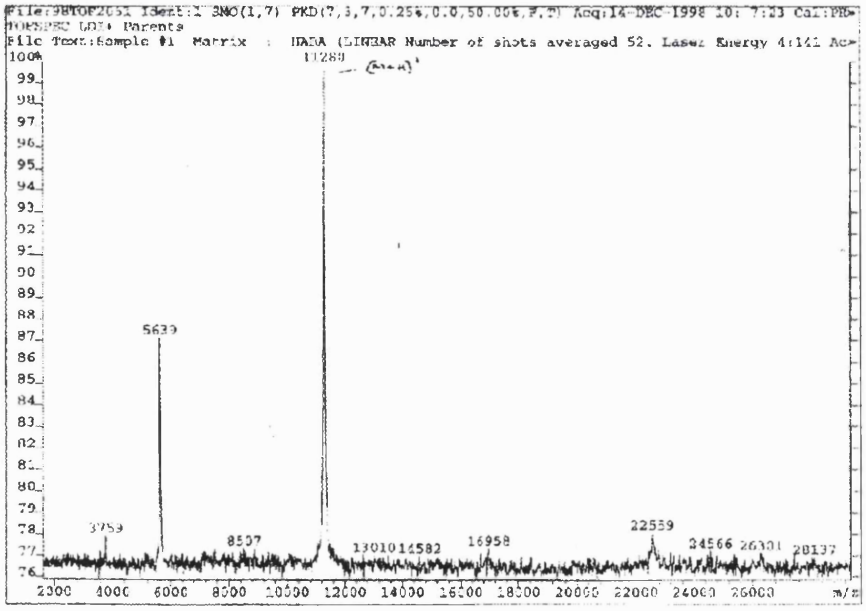


Figure 3.7: Spectrum obtained for CD2d1 using MAlDI-TOF mass spectrometry.

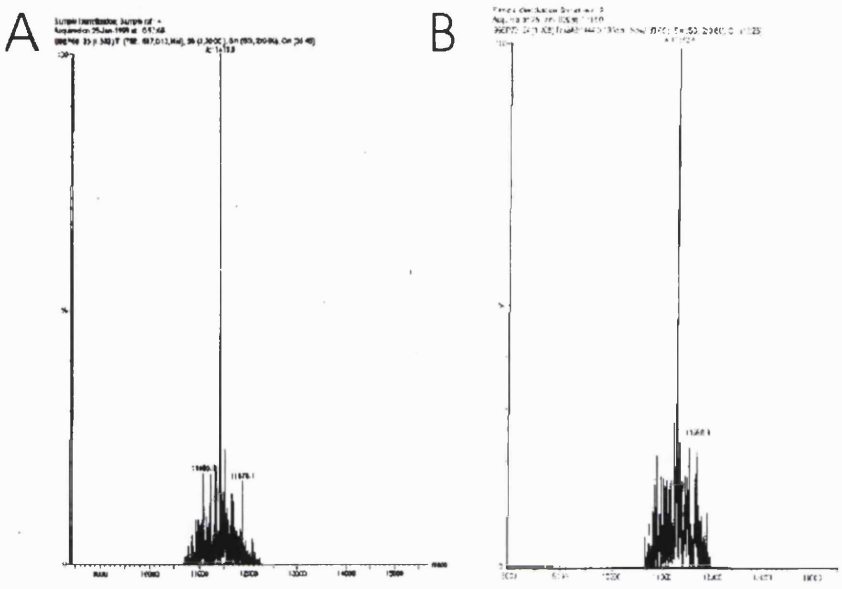


Fig. 3.8 Analyses using electrospray ionisation mass spectrometry. A - protein co-expressed with PUBS plasmid containing argU gene. B - CD2d1 expressed without PUBS.

The misincorporation of the amino acid may be due to non-standard pairing of the bases at the 'wobble' position which is the third base position of the codon, for example, leucine (UUG/UUA) may be incorporated in place of phenylalanine (UUC) (Parker & Precup, 1986). The mispairing may also occur at the second position as was demonstrated by the misincorporation of lysine (AAA) for the rare arginine codon (AGA) (Calderone *et al.*, 1996; Day *et al.*, 1996; Forman *et al.*, 1998; McPherson, 1988). However, the N-terminal sequencing of the purified CD2d1 did not reveal any misincorporation of the Arg1 position coded by the rare codon AGA. Furthermore the misincorporation of lysine for arginine does not change the overall charge of the protein and therefore should not affect the elution behaviour of the protein. There may be other possible misincorporations of other amino acids which may alter the overall charge of the protein, or possible post-translation events such as modification of the protein or deamination of asparagines, or, as will be discussed in a later chapter, other CD2d1-specific problems such protein folding. However, these issues associated with protein expression are beyond the scope of this thesis. While there are interesting issues that may be worth exploring further, for the purpose of the work within this thesis, the protein produced was of reasonable purity and produced excellent NMR spectra with little signs of inhomogeneity, these issues were therefore not pursue further.

CD2 mutants - construction and purification

A number of mutants were constructed using the method described in Chapter2. For each mutant, a native form and a His-tag form of the protein were constructed. All the mutants were sequenced to ascertain the presence of the mutations (Fig. 3.9 and Fig. 3.10).

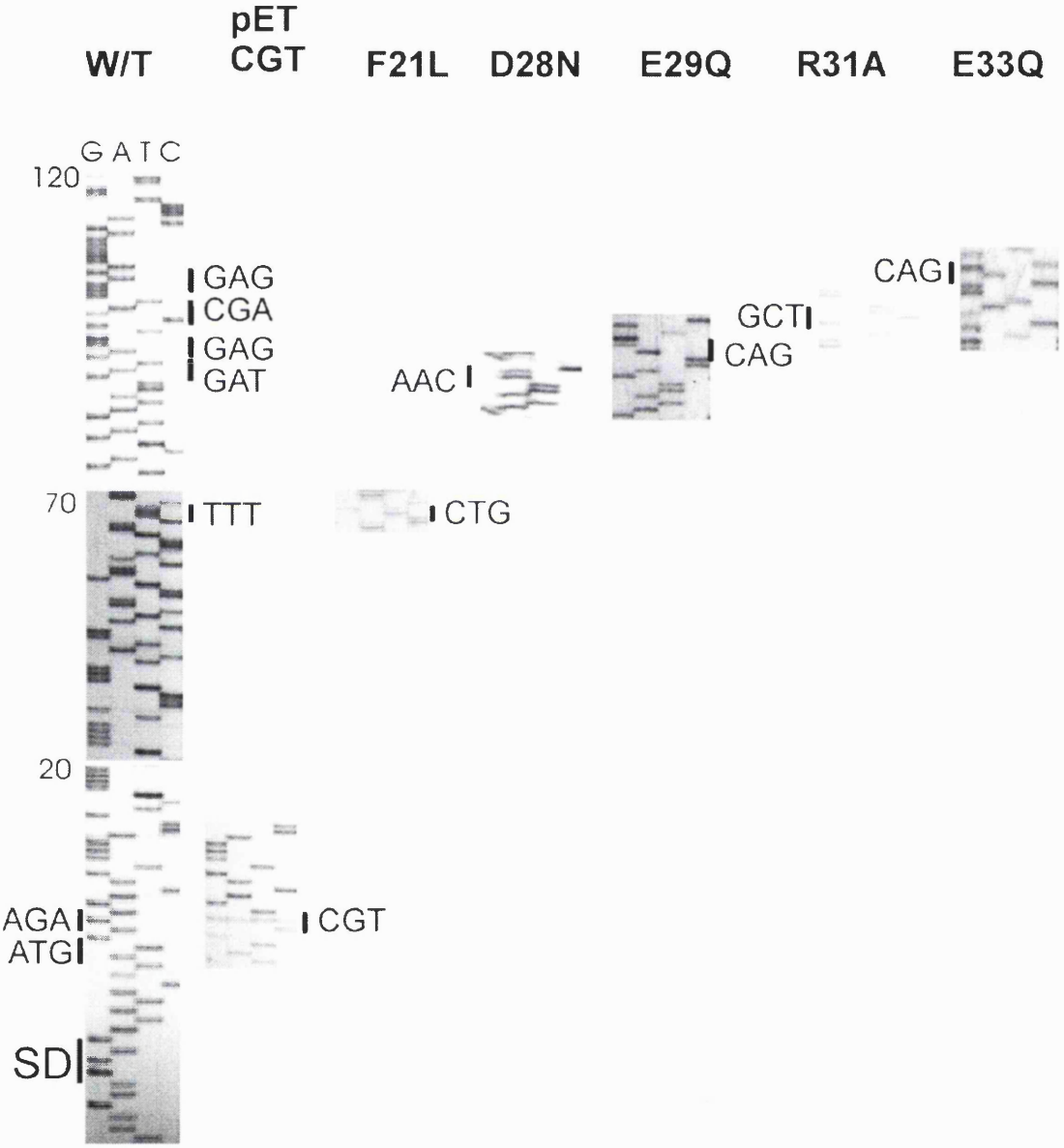


Fig. 3.9 The result of the sequencing of the mutants constructed. The codons in the wild-type and the corresponding mutation in the mutants are indicated. The SD denotes the Shine-Dalgarno sequence. The lanes in the sequencing gel were loaded in the order of GATC.

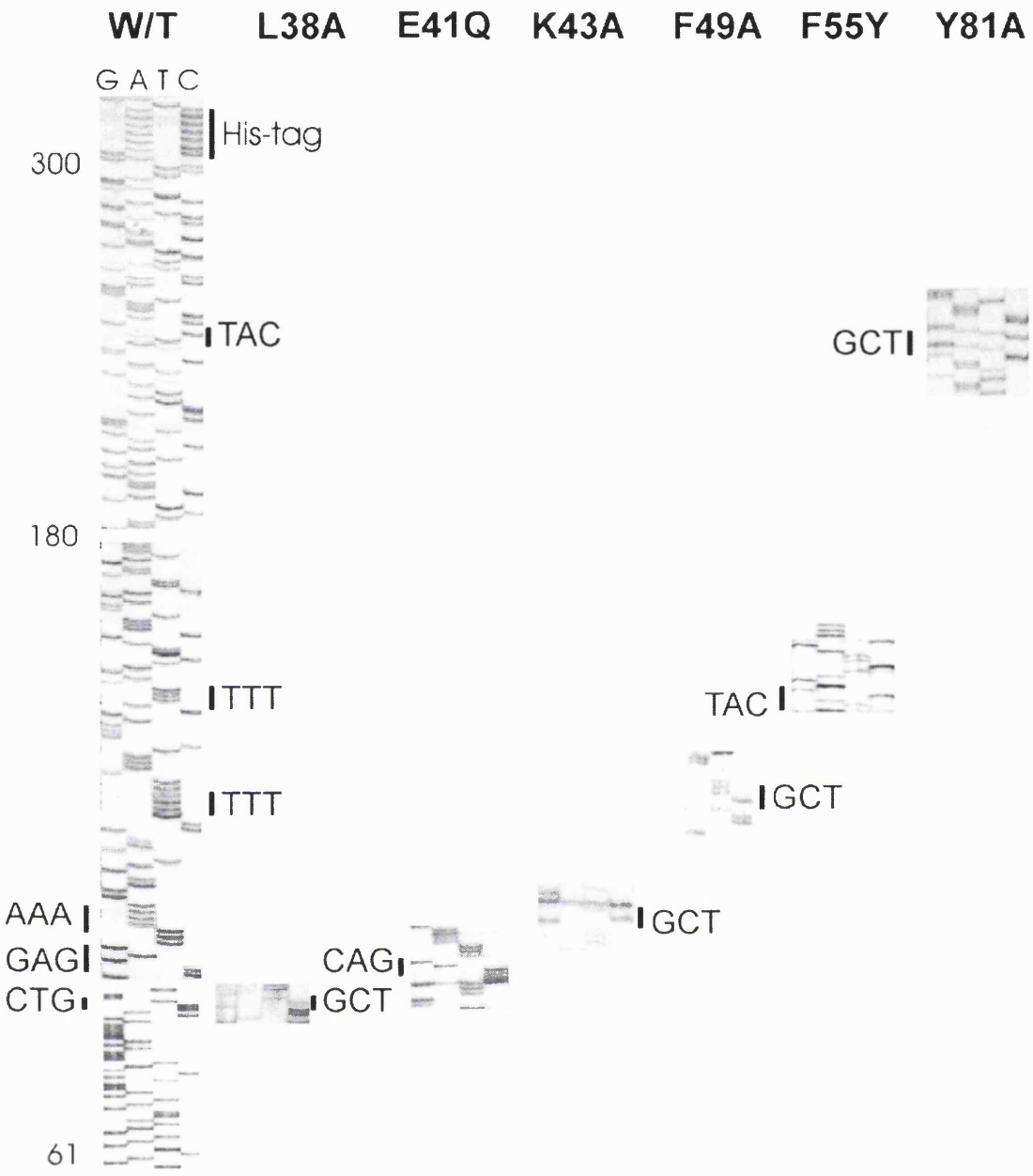


Fig. 3.10 Sequencing of the mutants constructed. Not shown are the F42Y and F49Y mutants. The codons in the wild-type and mutants are indicated.

The mutants were expressed and purified according to the method developed for wild-type CD2. However, it was found that the mutants have very different aggregational properties. Many mutants (e.g. K43A, R31A, L38A, Y81A) showed pronounced tendency to precipitate, particularly at pH values lower than 6, while some are highly soluble (e.g. E29Q). The variations in the propensity for aggregation in these mutants also demonstrate that avoidance of aggregation is crucial for a high yield of purified CD2d1 and its mutants. For example, E29Q which showed the lowest tendency to precipitate also produced the highest yield (~40 mg/L), while

K43A and Y81A which precipitate very strongly gave a yield of purified proteins of 1 mg/L or less. The purification method developed for the wild-type CD2d1 may therefore be unsuitable for some mutants which showed a tendency to precipitate at low pH. In cases such as these, the use of the His-tag fusion protein may be a more desirable option as the pH of the buffers used can kept above 6 at all time. The proteins can be purified using metal-chelate affinity chromatography followed by a size-exclusion chromatography step. The His-tag should not be removed for the next step which is the cation exchange chromatography. The protein is loaded onto the S-10 column at pH 6.5, then wash with pH7.5, and eluted at pH 8.5. For these mutants, the insertion of extra glycines between the protein and the His-tag may be desirable as it will permit the efficient removal of the His-tag if so wished.

Conclusion

The expression studies indicated that pET21b should be the preferred choice as the expression vector. The translation initiation region has the greater influence on the protein expression with the rare codon AGA preferred over the more common CGT as the second codon. For the protein expression, maximal yield may be achieve from a combination of the highest possible percentage yield of the target protein and the greatest final cell density. A protocol for CD2d1 expression and purification was established using a formulated culture medium.

It was found that the CD2d1 has a tendency to precipitate at high concentration, particularly between pH 5 and 6. The key to achieving a high level of production of purified CD2d1 lies in the avoidance of the precipitation event. Typical yield of 15-20 mg/L of protein may be obtained in minimal media, with up to 35-40 mg/L of protein possible if careful steps are taken to avoid protein precipitation and if the protein is expressed strongly. Some sacrifice of the quality of the protein, however, may accompany the higher expression level of the protein.

References

Amann, E., Brosius, J. & Ptashne, M. (1983). Vectors Bearing a Hybrid *Trp-Lac* Promoter Useful For Regulated Expression of Cloned Genes in *Escherichia-Coli*. *Gene* **25**(2-3), 167-178.

Arulanandam, A. R. N., Kister, A., McGregor, M. J., Wyss, D. F., Wagner, G. & Reinherz, E. L. (1994). Interaction Between Human CD2 and CD58 Involves the Major Beta-Sheet Surface of Each of Their Respective Adhesion Domains. *Journal of Experimental Medicine* **180**(5), 1861-1871.

Balbas, P. & Bolivar, F. (1990). Design and Construction of Expression Plasmid Vectors in *Escherichia- Coli*. *Methods in Enzymology* **185**, 14-37.

Baneyx, F. (1999). Recombinant protein expression in *Escherichia coli*. *Current Opinion in Biotechnology* **10**(5), 411-421.

Berg, O. G. & Kurland, C. G. (1997). Growth rate-optimised tRNA abundance and codon usage. *Journal of Molecular Biology* **270**(4), 544-550.

Brinkmann, U., Mattes, R. E. & Buckel, P. (1989). High-Level Expression of Recombinant Genes in Escherichia-Coli Is Dependent On the Availability of the *dnaY* Gene-Product. *Gene* **85**(1), 109-114.

Bula, C. & Wilcox, K. W. (1996). Negative effect of sequential serine codons on expression of foreign genes in *Escherichia coli*. *Protein Expression and Purification* **7**(1), 92-103.

Bulmer, M. (1988). Codon Usage and Intragenic Position. *Journal of Theoretical Biology* **133**(1), 67-71.

Calderone, T. L., Stevens, R. D. & Oas, T. G. (1996). High-level misincorporation of lysine for arginine at AGA codons in a fusion protein expressed in *Escherichia coli*. *Journal of Molecular Biology* **262**(4), 407-412.

Cantrell, A. S., Burgett, S. G., Cook, J. A., Smith, M. C. & Hsiung, H. M. (1991). Effects of 2nd-Codon Mutations On Expression of the Insulin-Like Growth Factor-Ii-Encoding Gene in *Escherichia-Coli*. *Gene* **98**(2), 217-223.

Chen, G. F. T. & Inouye, M. (1994). Role of the AGA/AGG Codons, the Rarest Codons in Global Gene- Expression in *Escherichia-Coli*. *Genes & Development* **8**(21), 2641-2652.

Crawford, D. A. (1994). Structure and Dynamics of a Cell Adhesion Protein, PhD Thesis, Hertford College, University of Oxford.

Day, A. J., Aplin, R. T. & Willis, A. C. (1996). Overexpression, purification, and refolding of link module from human TSG-6 in *Escherichia coli*: Effect of temperature, media, and mutagenesis on lysine misincorporation at arginine AGA codons. *Protein Expression and Purification* **8**(1), 1-16.

Deboer, H. A., Comstock, L. J. & Vasser, M. (1983). The *Tac* Promoter - a Functional Hybrid Derived From the *Trp* and *Lac* Promoters. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences* **80**(1), 21-25.

Driscoll, P. C., Cyster, J. G., Campbell, I. D. & Williams, A. F. (1991). Structure of Domain-1 of Rat Lymphocyte-T CD2 Antigen. *Nature* **353**(6346), 762-765.

Enfors, S. O. (1992). Control of *in vivo* Proteolysis in the Production of Recombinant Proteins. *Trends in Biotechnology* **10**(9), 310-315.

Fenner, C. (1975). Quantitation of Coomassie Blue Stained Proteins in Polyacrylamide Gels Based on Analyses of Eluted Dye. *Analytical Biochemistry* **63**, 595-602.

Forman, M. D., Stack, R. F., Masters, P. S., Hauer, C. R. & Baxter, S. M. (1998). High level, context dependent misincorporation of lysine for arginine in *Saccharomyces cerevisiae* a1 homeodomain expressed in Escherichia coli. *Protein Science* **7**(2), 500-503.

Gheysen, D., Iserentant, D., Derom, C. & Fiers, W. (1982). Systematic Alteration of the Nucleotide-Sequence Preceding the Translation Initiation Codon and the Effects On Bacterial Expression of the Cloned Sv40 Small-T Antigen Gene. *Gene* **17**(1), 55-63.

Gold, L. (1988). Posttranscriptional Regulatory Mechanisms in *Escherichia-Coli*. *Annual Review of Biochemistry* **57**, 199-233.

Gold, L. (1990). Expression of Heterologous Proteins in *Escherichia-Coli*. *Methods in Enzymology* **185**, 11-14.

Gold, L. & Stormo, G. D. (1990). High-Level Translation Initiation. *Methods in Enzymology* **185**, 89-93.

Hannig, G. & Makrides, S. C. (1998). Strategies for optimizing heterologous protein expression in *Escherichia coli*. *Trends in Biotechnology* **16**(2), 54-60.

Hirel, P. H., Schmitter, J. M., Dessen, P., Fayat, G. & Blanquet, S. (1989). Extent of N-Terminal Methionine Excision From *Escherichia-Coli* Proteins Is Governed By the Side-Chain Length of the Penultimate Amino-Acid. *Proceedings of the National Academy of Sciences of the United States of America* **86**(21), 8247-8251.

Ikemura, T. (1982). Correlation Between the Abundance of Yeast Transfer-RNAs and the Occurrence of the Respective Codons in Protein Genes - Differences in Synonymous Codon Choice Patterns of Yeast and *Escherichia-Coli* With Reference to the Abundance of Isoaccepting Transfer-RNAs. *Journal of Molecular Biology* **158**(4), 573-597.

Iserentent, D. & Fiers, W. (1980). Secondary Structure of mRNA and Efficiency of Translation Initiation. *Gene* **9**, 1-12.

Ito, W. & Kurosawa, Y. (1992). Development of a Prokaryotic Expression Vector That Exploits Dicistronic Gene Organization. *Gene* **118**(1), 87-91.

Ivanov, I., Alexandrova, R., Dragulev, B., Saraffova, A. & Abouhaidar, M. G. (1992). Effect of Tandemly Repeated AGG Triplets On the Translation of Cat Messenger-RNA in *Escherichia-Coli*. *FEBS Letters* **307**(2), 173-176.

Jaeger, J. A., Turner, D. H. & Zuker, M. (1990). Predicting Optimal and Suboptimal Secondary Structure For RNA. *Methods in Enzymology* **183**, 281-306.

Jansson, M., Li, Y. C., Jendeberg, L., Anderson, S., Montelione, G. T. & Nilsson, B. (1996). High-level production of uniformly N-15- and C-13-enriched fusion proteins in *Escherichia coli*. *Journal of Biomolecular NMR* **7**(2), 131-141.

Jones, E. Y., Davis, S. J., Williams, A. F., Harlos, K. & Stuart, D. I. (1992). Crystal-Structure At 2.8-Angstrom Resolution of a Soluble Form of the Cell-Adhesion Molecule CD2. *Nature* **360**(6401), 232-239.

Kane, J. F. (1995). Effects of Rare Codon Clusters On High-Level Expression of Heterologous Proteins in *Escherichia-Coli*. *Current Opinion in Biotechnology* **6**(5), 494-500.

Kurland, C. & Gallant, J. (1996). Errors of heterologous protein expression. *Current Opinion in Biotechnology* **7**(5), 489-493.

Kurland, C. G. (1987). Strategies For Efficiency and Accuracy in Gene-Expression. *Trends in Biochemical Sciences* **12**(4), 126-128.

Li, S. C., Squires, C. L. & Squires, C. (1984). Antitermination of *Escherichia-Coli* Ribosomal-RNA Transcription Is Caused By a Control Region Segment Containing Lambda Nut-Like Sequences. *Cell* **38**(3), 851-860.

Lilie, H., Schwarz, E. & Rudolph, R. (1998). Advances in refolding of proteins produced in *E-coli*. *Current Opinion in Biotechnology* **9**(5), 497-501.

Looman, A. C., Bodlaender, J., Comstock, L. J., Eaton, D., Jhurani, P., Deboer, H. A. & Vanknippenberg, P. H. (1987). Influence of the Codon Following the Aug Initiation Codon On the Expression of a Modified LacZ Gene in *Escherichia-Coli*. *EMBO Journal* **6**(8), 2489-2492.

Macdonald, L. E., Zhou, Y. W. & McAllister, W. T. (1993). Termination and Slippage By Bacteriophage-T7 RNA-Polymerase. *Journal of Molecular Biology* **232**(4), 1030-1047.

Makrides, S. C. (1996). Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiological Reviews* **60**(3), 512.

McPherson, D. T. (1988). Codon Preference Reflects Mistranslational Constraints - a Proposal. *Nucleic Acids Research* **16**(9), 4111-4120.

Neidhardt, F. C., Bloch, P. L. & Smith, D. F. (1974). Culture medium for enterobacteria. *Journal of Bacteriology* **119**(3), 736-747.

Olins, P. O., Devine, C. S., Rangwala, S. H. & Kavka, K. S. (1988). The T7 Phage Gene 10 Leader RNA, a Ribosome-Binding Site That Dramatically Enhances the Expression of Foreign Genes in *Escherichia- Coli*. *Gene* **73**(1), 227-235.

Parker, J. & Precup, J. (1986). Mistranslation During Phenylalanine Starvation. *Molecular & General Genetics* **204**(1), 70-74.

Rangwala, S. H., Finn, R. F., Smith, C. E., Berberich, S. A., Salsgiver, W. J., Stallings, W. C., Glover, G. I. & Olins, P. O. (1992). High-Level Production of Active HIV-1 Protease in *Escherichia-Coli*. *Gene* **122**(2), 263-269.

Rosenberg, A. H., Goldman, E., Dunn, J. J., Studier, F. W. & Zubay, G. (1993). Effects of Consecutive AGG Codons On Translation in *Escherichia-Coli*, Demonstrated With a Versatile Codon Test System. *Journal of Bacteriology* **175**(3), 716-722.

Sambrook, J., Fritsh, E. J. & Maniatis, J. (1989). *Molecular Cloning - A laboratory manual.* Second edit, Cold Spring Harbor Laboratory Press.

Schoner, B. E., Belagaje, R. M. & Schoner, R. G. (1986). Translation of a Synthetic 2-Cistron Messenger-RNA in *Escherichia- Coli. Proceedings of the National Academy of Sciences of the United States of America* **83**(22), 8506-8510.

Shiomi, N., Tomioka, K. & Katoh, S. (1998). Production and purification of soluble human CD2 secreted from recombinant Pichia pastoris. *Process Biochemistry* **33**(4), 377-383.

Smith, D. B. & Johnson, K. S. (1988). Single-Step Purification of Polypeptides Expressed in *Escherichia- Coli* As Fusions With Glutathione S-Transferase. *Gene* **67**(1), 31-40.

Spanjaard, R. A. & Vanduin, J. (1988). Translation of the Sequence AGG-AGG Yields 50-Percent Ribosomal Frameshift. *Proceedings of the National Academy of Sciences of the United States of America* **85**(21), 7967-7971.

Stormo, G. D. (1986). *Maximizing protein expression* (Reznikoff, W. & Gold, L., Eds.), Butterworth, Stoneham, Massachusetts.

Studier, F. W. & Moffatt, B. A. (1986). Use of Bacteriophage-T7 RNA-Polymerase to Direct Selective High-Level Expression of Cloned Genes. *Journal of Molecular Biology* **189**(1), 113-130.

Tobias, J. W., Shrader, T. E., Rocap, G. & Varshavsky, A. (1991). The N-End Rule in Bacteria. *Science* **254**(5036), 1374-1377.

Wagner, L. A., Weiss, R. B., Driscoll, R., Dunn, D. S. & Gesteland, R. F. (1990). Transcriptional Slippage Occurs During Elongation At Runs of Adenine or Thymine in *Escherichia-Coli. Nucleic Acids Research* **18**(12), 3529-3535.

Weickert, M. J., Doherty, D. H., Best, E. A. & Olins, P. O. (1996). Optimization of heterologous protein production in Escherichia coli. *Current Opinion in Biotechnology* **7**(5), 494-499.

Winnacker, E.-L. (1987). *From Genes to Clones - Introduction to Gene Technology*, VCH, Weinheim.

Withka, J. M., Wyss, D. F., Wagner, G., Arulanandam, A. R. N., Reinherz, E. L. & Recny, M. A. (1993). Structure of the Glycosylated Adhesion Domain of Human T-Lymphocyte Glycoprotein CD2. *Structure* **1**(1), 69-81.

Zhang, S. P., Zubay, G. & Goldman, E. (1991). Low-Usage Codons in *Escherichia-Coli*, Yeast, Fruit-Fly and Primates. *Gene* **105**(1), 61-72.

Zuker, M. (1989). On Finding All Suboptimal Foldings of an RNA Molecule. *Science* **244**(4900), 48-52.

Chapter 4

Theory and Methods of NMR and Resonance Assignments of CD2d1

The first direct observations of the nuclear magnetic resonance (NMR) phenomenon in bulk materials were made more half a century ago by two American research groups led by physicists Felix Bloch (Bloch *et al.,* 1946) and Edward Purcell (Purcell *et al.,* 1946). Since then the field of NMR research has progressed and diversified considerably, with NMR spectroscopy becoming an invaluable spectroscopic technique in a wide variety of disciplines and applications; from the rapid determination of molecular structure of small organic molecules in chemistry, to the non-invasive imaging of a patient's internal organs in medicine. In the biological sciences, NMR allows the biochemical pathways of metabolites to be monitored, and the human cerebral processes to be visualised. Based on the pioneering work by Richard Ernst on pulsed Fourier transform NMR spectroscopy (Ernst *et al.,* 1990) and Kurt Wüthrich on its application to proteins (Wuthrich 1986), NMR has also emerged as the only credible alternative to X-ray crystallography for the determination of high resolution three-dimensional structures of macromolecules. Such high resolution structures of protein are now obtained routinely using NMR.

With recent innovations in experimental approaches (Pervushin *et al.,* 1997; Tjandra & Bax 1997) and the continuing advance in instrumentation, the enormous progress that has occurred in the field of macromolecule NMR research of the past two decades is continuing unabated.

In this chapter, some of the basic principles and techniques of NMR spectroscopy are introduced. However, the limitations of this thesis dictate that only a small area of NMR of relevance to its application in biomolecular analysis can be dealt with. The description of NMR is therefore confined to a brief introduction to some concepts that may be useful for the understanding of the NMR experiments used. A particular NMR experiment, the heteronuclear single quantum coherence (HSQC) experiment will be described in greater detail as it illustrates some of the principles involved in many other NMR experiments, but descriptions of all other experiments are necessarily brief. Comprehensive treatment of the NMR phenomenon and its practical applications, however, can be found in a number of monographs (Abragam 1961; Cavanagh *et al.,* 1996; Derome 1987; Ernst *et al.,* 1990; Evans 1995; Freeman 1998). For a historical perspective of the development of NMR as a spectroscopic technique, see Grant & Harris (1996).

Theoretical basis of the NMR phenomenon

The phenomenon of the NMR spectroscopy is a consequence of the properties of the nuclear spin, a form of innate angular momentum and an intrinsic quantum-mechanical property of a nucleus that does not have an equivalent in classical mechanics. The nuclei of interest in protein NMR, such as ^{1}H, ^{15}N and ^{13}C, have a spin quantum number I of ½. A spin ½ nucleus, with a magnetic moment μ, can be viewed as a tiny bar magnet that, when placed in a static magnetic field B_0, aligns itself to the magnetic field in two spin states, i.e. in a parallel (α-state) or anti-parallel (β-state) orientation, with the parallel orientation having a lower energy state. At thermal equilibrium, there will be a difference in population between the two states, with a small surplus in the lower energy state according to the Boltzmann distribution. There is therefore a net component aligned parallel with the magnetic

field in the α state. In a magnetic field, the spin will precess with its axis of rotation moving around the direction of the field (the z axis in Fig. 4.1) in the manner of a gyroscope. At equilibrium, all the precessing spins have a component of the magnetic moment in the z axis, while on the x-y plane the phase of precession is random. Consequently, from the view of bulk properties, there is a net magnetisation M_0 along the z axis, but no net magnetisation on the x-y plane (Fig. 4.1). This net magnetisation can be manipulated by the application of an oscillating radio-frequency (RF) field B_1, and the entire basis of NMR experiments revolves around how this bulk magnetisation vector M_0 is manipulated by the judicious application of pulses of the RF field B_1.

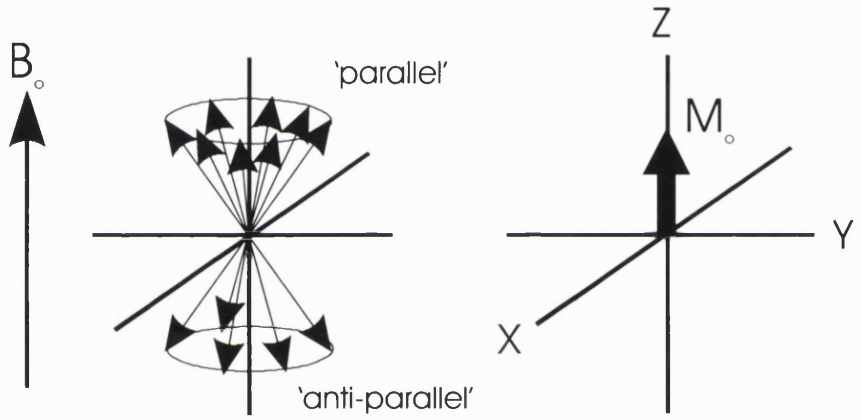


Fig 4.1. *At thermal equilibrium, a slight excess of individual nuclear magnetic moments (spin ½), represented by vectors, is orientated parallel to the applied field, so that the sample becomes magnetised in that direction with a net magnetisation M_0. The lack of phase coherence means that there is no net magnetisation in the xy-plane.*

It should be noted at this point that this, the vector model description of NMR, is based on the classical formalism and is only applicable to a bulk manifestation of the nuclear magnetisation. The physical reality of the NMR phenomenon on the level of individual nuclei can be better understood from a quantum mechanical viewpoint using density matrix theory (Abragam 1961). Other alternative approaches for the description of NMR include the product operator formalism (Sorensen *et al.*, 1983) which is based on density matrix theory, and method for tracing coherence pathways (Bodenhausen *et al.*, 1984). A detailed account of such approaches and the description of individual pulse sequence using

the product operator formalism, however, are beyond the scope of this thesis and will not be attempted.

The alternating RF field B_1 is arranged such that the applied field is in the *x-y* plane, perpendicular to the static field B_0. The application of a pulse of the B_1 RF produces a torque acting on the sample magnetisation, driving the sample magnetisation around the B_1 field vector. The length of the pulse determines how far the sample magnetisation will rotate round. The pulse length can be set at a value such that the magnetisation is rotated through 90° ($\pi/2$ radian) or 180° (π radian). When the magnetisation is rotated through 90° onto the *x-y* plane, there is an equal population of spins in the α and β spin states; while after 180° rotation, there is a complete inversion of population in the two spin states, with a greater population of spin now in the excited β state. Many NMR experiments are built up from a combination these so-called $\pi/2$ and π pulses (Fig. 4.2).

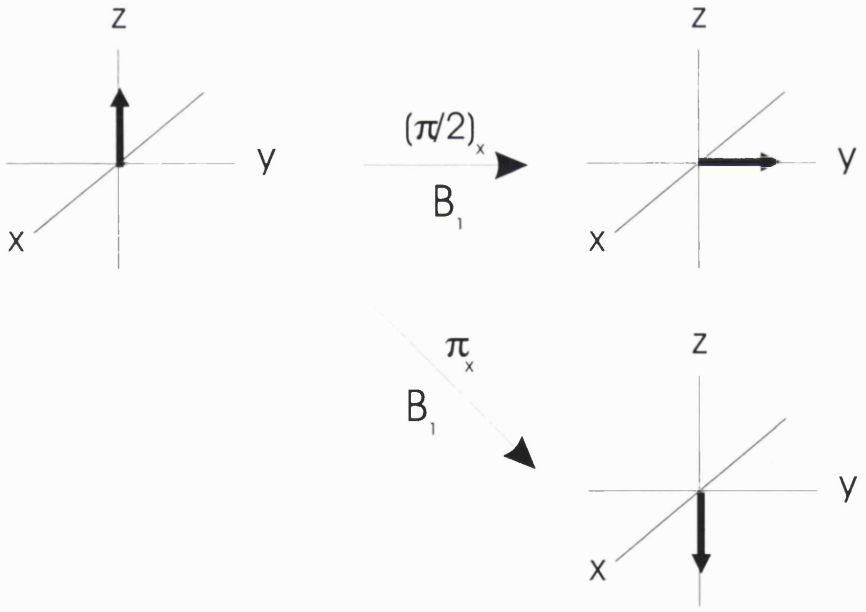


Fig. 4.2 *When the RF field is switched on, the sample magnetisation is rotated round. A 90° pulse will leave the sample magnetisation on the x-y plane while a 180° pulse invert the magnetisation. The subscript denotes the direction in which the pulse is applied. In this example the pulse is applied along the x-axis.*

In the presence of a static magnetic field B_0, the nuclear spin precesses round the axis at a rate called the Larmor frequency. This precession at Larmor frequency ω can generate an NMR signal by inducing an alternating current in the receiver coil.

The NMR signal is detected on the *x-y* plane, and in the absence of an RF pulse, no signal is detected because there is no net magnetisation on the *x-y* plane. However, when the net magnetisation is tipped over onto the x-y plane by an RF pulse, a signal is generated as the magnetisation precesses. In Fig. 4.2, the net magnetisation would be rotated round the *z* axis, however, the representation of this rotating vector could be greatly simplify by choosing a set of coordinates which also rotates round at frequency ω, therefore the net magnetisation in this 'rotating frame' would appeared static.

The Larmor frequency ω is dependent on the strength of the applied field B_0 and on the intrinsic properties of the nucleus reflected in the gyromagnetic ratio γ:

$$\omega = -\gamma B_0 \qquad [4.1]$$

This Larmor frequency of a nuclear spin can be related to its chemical environment by its chemical shift. In a protein, each of many hundreds of nuclei is situated in a slightly different chemical environment, and the magnetic field experienced by each nucleus is modified by the local magnetic field B_{loc}. The precession frequency of each nuclei therefore varies, with $\omega = -\gamma(B_0 + B_{loc})$. This variation gives rise to distinct chemical shift for each nuclei, making it possible to study the molecular structure by NMR. For the protons in a diamagnetic protein, the chemical shift may range over 10-12 parts per million (ppm).

The signal detected from an NMR experiment is called the free induction decay (FID) which is a combination of the sinusoidal function, the signal from the precessing nuclei, and an exponential decay. The exponential decay is the result of relaxation processes which return the nuclei from the excited state back to equilibrium. This concept will be dealt with in greater detail in Chapter 6. The FID is normally recorded using quadrature detection whereby the radio frequency signal is detected by a phase-sensitive detector as two orthogonal signals, one a sine function and the other a cosine function. Both are subjected to Fourier transformation in which the time domain data (i.e. the FID which is acquired as a function of time) is converted to the frequency domain that produces the recognisable 1D NMR spectrum and recombined. A simple 1D experiment can therefore be performed with a single

pulse and recording the FID. During the processing of the data, the appearance of the spectrum may be improved and the resolution enhanced by the application of appropriate mathematical algorithms. It should be mentioned that the development of Fourier transform (FT) NMR by Ernst and Anderson (Ernst & Andersen 1966) is one of the landmarks in the history of NMR. FT NMR is much faster than the older continuous wave (CW) method, thereby allowing the signal-to-noise ratio to be greatly improved in a given time. It also allows versatile complex pulse sequences to be used on macromolecules, without which the structural studies of proteins by NMR would not be possible.

Multi-dimensional NMR

The development of multi-dimensional NMR is crucial to the application of NMR to studies of protein. The first step in any biomolecular NMR studies is to assign the resonance frequency of particular nuclei. However, even in a small protein, there are hundreds of proton resonances which inevitably leads to severe overlap in the spectrum, making assignment impossible in one-dimensional (1D) spectroscopy. This problem of spectral overlap can be alleviated by dispersing the spectrum into a second dimension. Resonances that were previously overlapped may be resolved, thereby aiding the assignment of the resonance.

However, the conventional proton-based homonuclear NMR methods are limited by the size of the protein; at a higher molecular mass (>10kDa), the shorter proton relaxation times and the larger number of resonances lead to reduced magnetisation transfer efficiency, greater chemical shift overlap and degeneracy, and spin diffusion effects. These problems can be circumvented by making use of heteronuclear couplings in ^{15}N- and ^{13}C-labelled proteins and further increasing the dimensionality. Even in a protein that is only enriched in ^{15}N or ^{13}C, the conventional homonuclear two dimensional (2D) experiments can be edited with respect to the ^{15}N or ^{13}C chemical shift, thereby simplifying the spectrum considerably. For a protein that is uniformly enriched in both in ^{15}N and ^{13}C, by using ^{13}C-edited and triple resonance NMR techniques, the spin system identification

and sequential assignment can be dramatically simplified. Heteronuclear NMR techniques also provide additional parameters such the heteronuclear scalar coupling constants and chemical shifts which are useful in the structural and biophysical analysis of protein. NMR methods making use of the ^{15}N and ^{13}C labelling are called triple resonance techniques as the ^{1}H resonances are always recorded together with the ^{15}N and ^{13}C resonances. Such methodology allows the study and determination of protein structure up to ~30 kDa. At higher molecular mass, however, the transverse relaxation time decreases with increase in molecular mass and correlation time, resulting in broadened peaks. This may be overcome by per-, selective or fractional deuteration of the protein, which can simplify the spectrum by 'diluting' the protons present, and enhance the NMR spectrum by significantly reducing the line-width by increasing the relaxation times (Sattler & Fesik 1996). Such advances in heteronuclear multidimensional NMR techniques have made possible the assignment of resonances and structural determination of larger proteins with increasing precision and accuracy, and a large number of structures of proteins and protein complexes with molecular mass 10-20 kDa, with some of 30 kDa and over, have been reported in the past few years (Garrett *et al.,* 1999; Kelly *et al.,* 1999). Backbone resonance assignments of a multimeric protein with molecular weight in excess of 60 kDa have also been reported (Shan *et al.,* 1996).

The multidimensional NMR experiments is made possible by a number of developments in the last twenty years or so: the progress in recombinant gene technology that permits ^{13}C, ^{15}N and ^{2}H isotopically-labelled protein to be produced routinely by heterologous expression, the development of complex pulse sequences to transfer magnetisation between scalar *J* (through bond) and dipolar (through space) coupled spins, the availability of fast computing power and better spectrometers with improved radio frequency electronics, and the increased magnetic field strength as well as the development of commercially available hardware such as pulsed field gradients. Current research to further extend the capability of NMR is still proceeding apace, with novel applications found for protein NMR such as Structure-Activity Relationships (SAR) by NMR (Shuker *et al.,* 1996), and new methodology such as the transverse relaxation optimised spectroscopy (TROSY) experiments (Pervushin *et al.,* 1997) and application of liquid crystal technology (Tjandra & Bax 1997) being introduced. Together with the prospect of even higher magnetic field

strength spectrometers becoming available, the future of protein NMR spectroscopy looks to be highly promising.

2D NMR

The multidimensional experiments may be understood by first examining how 2D NMR experiment are constructed. The first papers published on 2D NMR were by Ernst and his co-workers (Aue *et al.,* 1976; Kumar *et al.,* 1975; Muller *et al.,* 1975), inspired by the original idea of Jean Jeener (Jeener 1971). The basic concept is simple - in a basic 2D experiment, two pulses are separated by a time delay t_1 which is steadily incremented when the pulse sequence is repeated during the course of the experiment, and the subsequent Fourier transform of not only the FIDs (time period t_2) after the second pulse, but also with respect to the time variable t_1. The signal detected is therefore a function of two time variables and the resultant spectra would appear as diagonal and cross-peaks in a two-dimensional plot generated as a function of the two time periods.

The two time variables t_1 and t_2 are detected differently (Fig. 4.3). In a normal 1D NMR experiment, the familiar sinusoidal decay of the FID is recorded directly as a series of data points separated by the dwell time, a duration which is set to be the reciprocal of the spectral width (SW) in order to satisfy the Nyquist criterion. These data points are therefore sampled at 1/SW intervals following excitation, with the total number of points collected N chosen to suit the required digital resolution and data size. In a 2D experiment, this is the acquisition time t_2.

Notionally, this same set of data can be acquired indirectly - an N number of experiments are performed, but with only one data point per experiment recorded, and in each successive experiment, the time (the t_1 time delay) is steadily incremented by a 1/SW second delay before acquiring the single data point. The sum of these discrete data points from this rather more time-consuming set of experiments is equivalent to the directly acquired FID (Fig. 4.3). In reality, of course, the experiments are recorded not as a series of points, but as a series of 1D experiments which, when Fourier transformed with respect to t_1, produces the second dimension in a 2D frequency plot. This, in a 2D experiment, is essentially how the incremented delay t_1 can generate another dimension, i.e. the indirect dimension that forms the

basis of multidimensional NMR. How this t_1 time delay is inserted in a pulse sequence and used to carry the information about a spin during a 2D NMR experiment is described below.

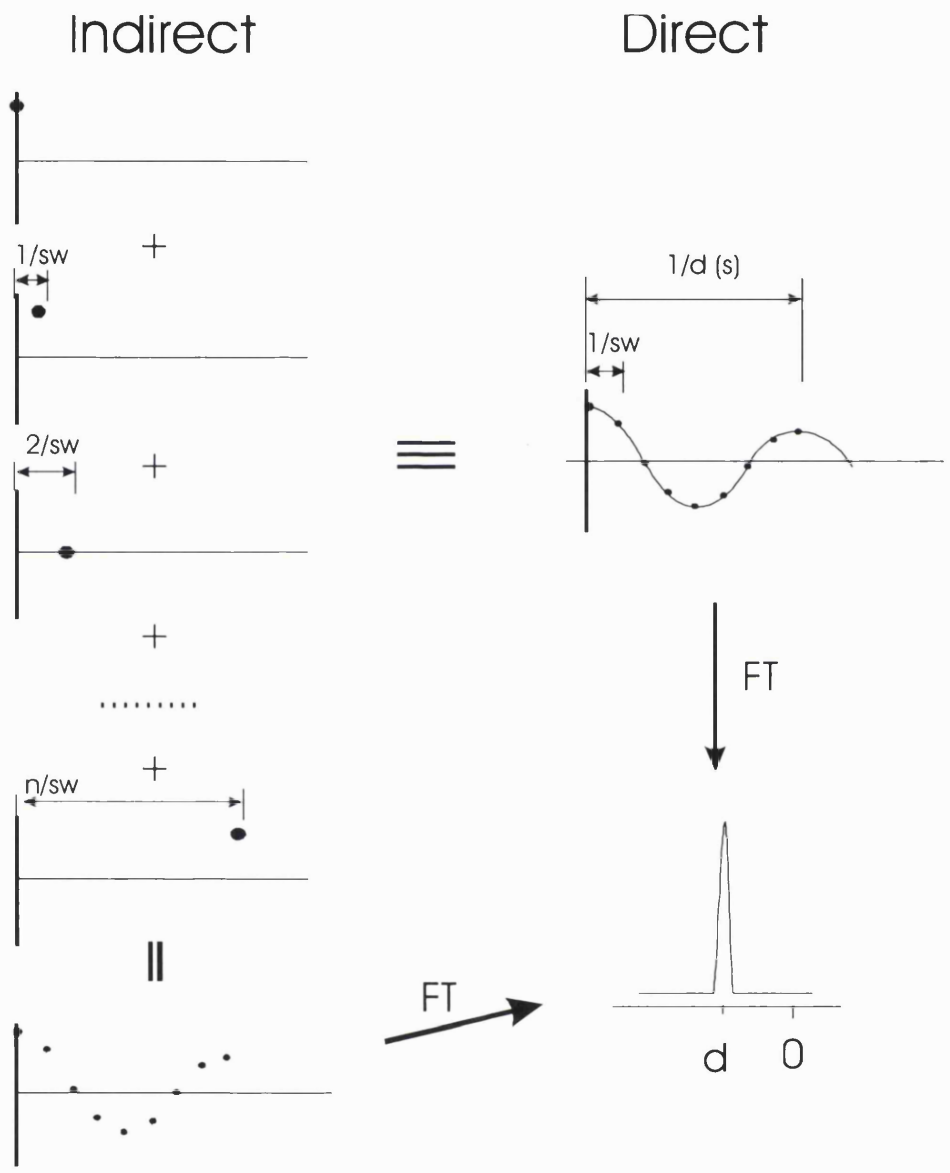


***Figure 4.3**: Direct and indirect method for the acquisition of time domain signal. For directly detected dimensions of multidimensional experiments a series of points are collected during a single experiment, while in indirectly detected dimensions a single point is collected in each of a series of experiments.*

In a simple 1D 'pulse-and-collect' NMR experiment, there are two time period, i.e. the preparation and acquisition time. In the basic 2D NMR scheme, however, there are four time periods: preparation, evolution, mixing and detection (Fig. 4.4). The preparation period brings the sample to a state of thermal equilibrium

followed by the application of a pulse to bring the magnetisation onto the *x*-*y* plane. During the evolution period (the time period t_1), the spins are allowed to precess or 'evolve' for a time t_1, and the spins are 'labelled' according to the precessing frequency - by changing the period t_1, the phases of the spin at the beginning of the mixing period are altered as a function of its frequency, thereby carrying over information about the spin. During the mixing period, the 'coherence' is transferred from one spin to another and correlation between two spins is achieved. The design of the mixing period determines, to a large extend, just what the particular 2D experiment achieves. The detection or acquisition period (time period t_2) simply records each FIDs. Each successive FID records the history of the evolution of the spin system as a function of t_1. The Fourier transformation of each t_2 dimension yields a set of directly detected 1D spectra, and the intensities of the resonances are modulated sinusoidally as a function of the t_1 duration as a result of phase changes at the end of t_1. The t_1 dimension can then be further Fourier transformed to produce a 2D spectrum $S(\omega_1,\omega_2)$ in which a 'cross-peak' appears at the intersection of the two nuclear signals.

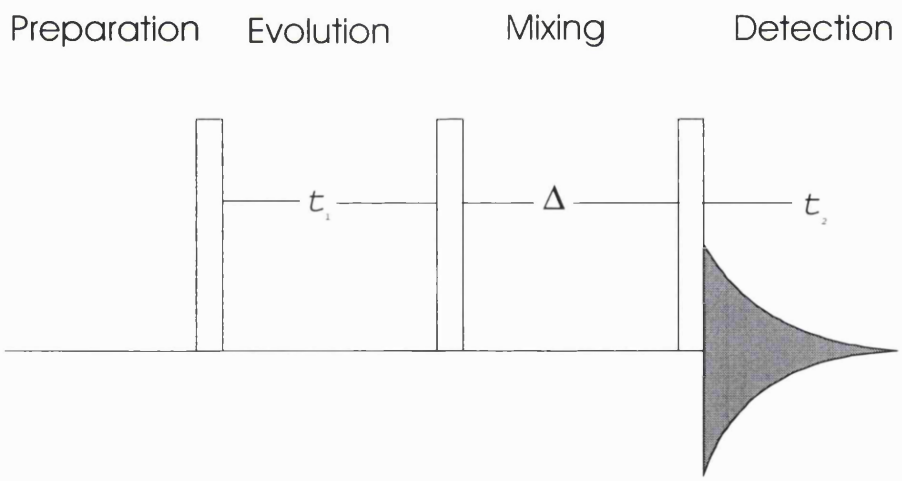


***Figure 4.4**: The basic steps in a 2D NMR experiment.*

The individual 2D NMR experiments are distinguished by the applied excitation and mixing procedures. A classic 2D experiment is the correlated spectroscopy (COSY) experiment, the homonuclear version of Jan Jeener's original experiment. A simple COSY experiment has the pulse sequence of $(90°)_x$ - t_1 - $(90°)_x$ - Acquisition (t_2), i.e. two 90° pulses applied along the *x* axis separated by a

time delay t_1. Scalar coupled protons are correlated in this experiment and the proton resonances are dispersed into two dimensions, forming diagonal peaks and cross-peaks. Other important homonuclear 2D NMR experiments include total correlation spectroscopy (TOCSY) which gives correlations to all the protons in an extended coupling network, and nuclear Overhauser effect spectroscopy (NOESY) that yields through space connectivities of protons by dipolar coupling. 2D NMR can also be used to resolve the proton resonances according to the chemical shifts of the directly bonded heteronuclei such as ^{15}N and ^{13}C as in the HSQC experiment. The HSQC is analogous to the COSY sequence but correlates two different nuclei. How this is done will be described in greater detail later.

Higher Dimensional NMR

The dimensionality of the NMR experiments can be increased by the appropriate combination of 2D NMR experiments (for reviews see Clore & Gronenborn 1991; Oschkinat *et al.,* 1994). A three-dimensional (3D) experiment is constructed by joining together two 2D pulse schemes, but with the detection period of the first and the preparation period of the second omitted. This results in a pulse train comprising of two independently incremented evolution periods (t_1 and t_2), two mixing periods (Δ_1 and Δ_2), and a detection period t_3. The subsequent 3D Fourier transformation, with respect to the time variable t_1, t_2 and t_3, produces a 3D frequency domain NMR spectrum. A four-dimensional experiment can be constructed in similar manner. A multidimensional NMR experiment will have the general form:

Preparation - (Evolution - Mixing)$_n$ - Detection

The multidimensional experiments can therefore be understood as concatenation of 2D experiments. The experiments can thus be named either as combination of the underlying 2D techniques such as NOESY-HSQC and TOCSY-NOESY, or after the participating nuclei as in HNCACB, CBCA(CO)NH. A CBCA(CO)NH experiment, for example, gives correlations between the amides (both the ^1H and ^{15}N atom) of a residue and the Cα and Cβ atoms of the preceding

residue for every amino acid where such correlation is possible in a protein. Nuclei given in parenthesis are not detected but are involved in coherence transfer pathway.

^{1}H-^{15}N HSQC

In this section, the 2D NMR experiment HSQC (Bodenhausen & Ruben 1980) is described in some detail as it is relatively simple and illustrates some of the principles involved in heteronuclear *J*-correlation experiments. Briefly, in a heteronuclear two-spin AX system (A is the proton and X the ^{15}N or ^{13}C atom), the magnetisation on proton is first transferred to the X nucleus by an INEPT step so that the polarisation of the X nucleus is enhanced. After a time delay, the magnetisation is then transferred back again by another INEPT step to the proton so that proton magnetisation can be detected. HSQC is a highly sensitive experiment which is first performed on a ^{1}H-^{15}N system but can also be extended to the ^{1}H-^{13}C system. The basic scheme of HSQC is as follows:

INEPT - time delay t_1 - reverse INEPT - Acquisition time t_2

INEPT

INEPT (Insensitive Nuclei Enhancement by Polarisation Transfer) is one of the most important building blocks of heteronuclear multidimensional NMR experiments (Morris & Freeman 1979). As its name indicates, the function is to transfer magnetisation from a sensitive nucleus such as proton to an insensitive one (such as the ^{13}C or ^{15}N atom) by polarisation transfer. This sensitivity enhancement step is necessary as insensitive nucleus generates a much weaker signal than the sensitive nucleus. The detection sensitivity of a nucleus depends on the gyromagnetic ratio γ. A high gyromagnetic ratio gives greater polarisation and in general, the signal obtainable from a nucleus with a gyromagnetic ratio γ is proportional to γ^3. A nucleus with a low γ would produce an extremely small NMR signal. For example, the gyromagnetic ratio of ^{13}C is about 4 times lower than that of

protons, the intensities of the carbon signal is therefore 64 times lower than that of proton, and thus more difficult to detect.

The INEPT building block is given by the pulse scheme in the following figure:

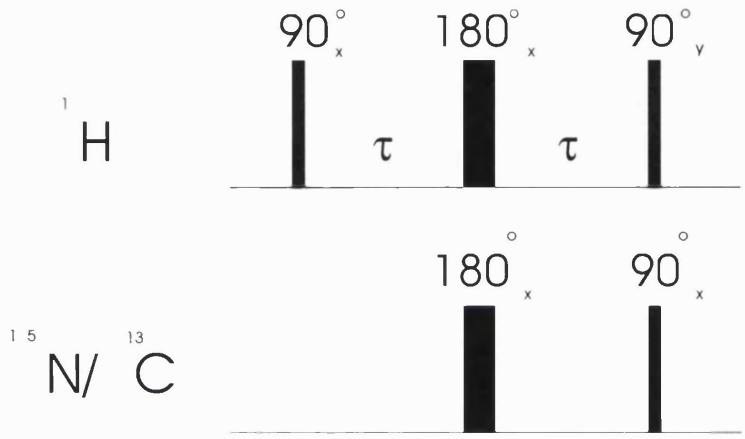


Fig. 4.5 *Diagrammatic representation of the pulses acting on the sensitive and insensitive nuclei. The proton is the sensitive nuclei, while the insensitive nuclei may be ^{15}N or ^{13}C. The thin bars represent 90° pulse while the thick bars represent 180° pulse. The time delay between pulses is τ. For an AX system with coupling constant J, the time delay τ is set optimally to 1/(4J).*

The scheme can be seen as a combination of the spin echo and selective population inversion (SPI). Spin echo was accidentally discovered by Erwin Hahn when he applied two successive pulses separated by a short duration, and detected a signal (the echo) at a time when no pulse was applied. The phenomenon can be understood as follows: the first pulse ($90°_x$) tips the sample magnetisation over onto the $+y$ axis, however, due to the inhomogeneity of the field, the isochromats (the individual vectors that contributes to the net magnetisation) eventually fan out as they precess at slightly different rates. On the application of a second pulse along the x axis ($180°_x$) after time τ, all the isochromats are rotated by 180° onto the $-y$ axis, the result being that those on the leading edge (the faster ones) now find themselves lagging behind and those lagging behind (the slow one) are now on the leading edge. However, all the isochromats still precess at their original rates, therefore after a further time τ, all the isochromats again converge (i.e. they are refocused), thereby regenerating the signal and thus the echo. In addition to the inhomogeneity, the chemical shift can be refocused in an analogous manner. Chemical shifts with varying frequencies (therefore precessing a different rate) are refocused at the peak of

the echo. This useful property of refocusing the chemical shift by spin echo allows the nuclear magnetism to be manipulated independent of the chemical shifts, and such a sequence is therefore frequently used in pulse sequences. This spin echo sequence is seen in the first two pulses on the sensitive nucleus (the proton) in the INEPT sequence (Fig 4.5) and the inhomogeneity and chemical shift are refocused at the end of the second τ interval. The refocused chemical shift permits the following step of population inversion for all the proton chemical shifts without the undesirable selectivity of SPI which will be described later.

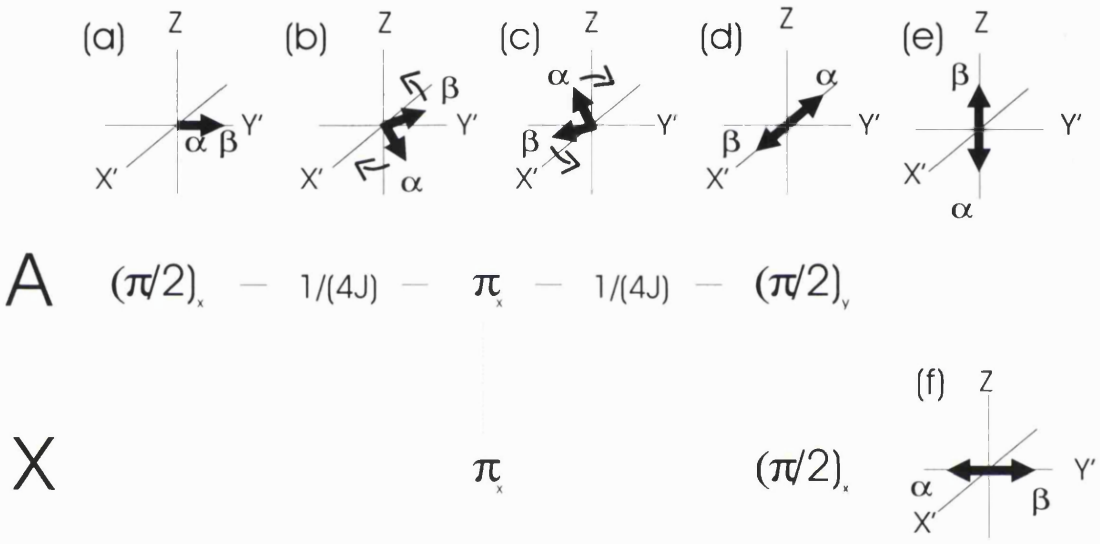


Fig. 4.6 *Vector diagram of the INEPT experiment (Derome 1987). The x' and y' axis denotes the rotating frame. (a) The sensitive nuclei's spin vectors* α *and* β*, which arise as a result of the coupling to the attached insensitive nuclei, are flipped onto the y-axis by a* $\pi/2$ *pulse. (b) Precession of the vectors for a time* τ *which is set at 1/(4J) such that the two vectors subtend an angle of* $\pi/2$ *radians. (c) There are two* π *pulses - the first, applied to* ^1H*, rotate the vectors by* π *radian (i.e. onto their mirror image location), the second* π *pulse, applied to the insensitive nuclei X, interchanges the* α *and* β *labels of the vectors. (d) Free precession of the vectors for 1/(4J) interval carries them onto the x axis and antiphase alignment. (e) A* $(\pi/2)_y$ *pulse rotates the vectors along the* $\pm$ *z axis. (f) Magnetisation vectors for the insensitive nuclei are rotated by the X* $\pi/2$ *pulse onto the y axis for detection.*

Note, however, that there is a simultaneous 180° pulse on the insensitive nucleus. The simultaneous pulses refocus the chemical shifts but not the heteronuclear scalar couplings, the reason being that the 180° pulse on the insensitive nucleus inverts its population, those that were previously in the higher energy β state are now in the lower energy α state and vice versa, with the result that the precessing vectors of the protons attached to these nuclei also exchange their rates of precession.

In Fig. 4.6, the α and β vectors of the protons interchange their labels and precess towards the *x* axis rather than refocus onto the *y* axis, therefore, the precession due to chemical shift evolution can be refocused while the precession due to scalar coupling evolution is not. For a heteronucleus with coupling constant *J*, the time interval τ for the free precession is set at optimally at $1/(4J)$, such that the vectors are brought into antiphase alignment on the *x* axis after the second τ time interval. This antiphase alignment is then rotated onto the *z* axis by a $90°_y$ pulse. The last $90°_x$ pulse on the insensitive nuclei may be applied simultaneously (though not necessary so) with this $90°_y$ pulse in order to create a transverse antiphase magnetisation that is observable.

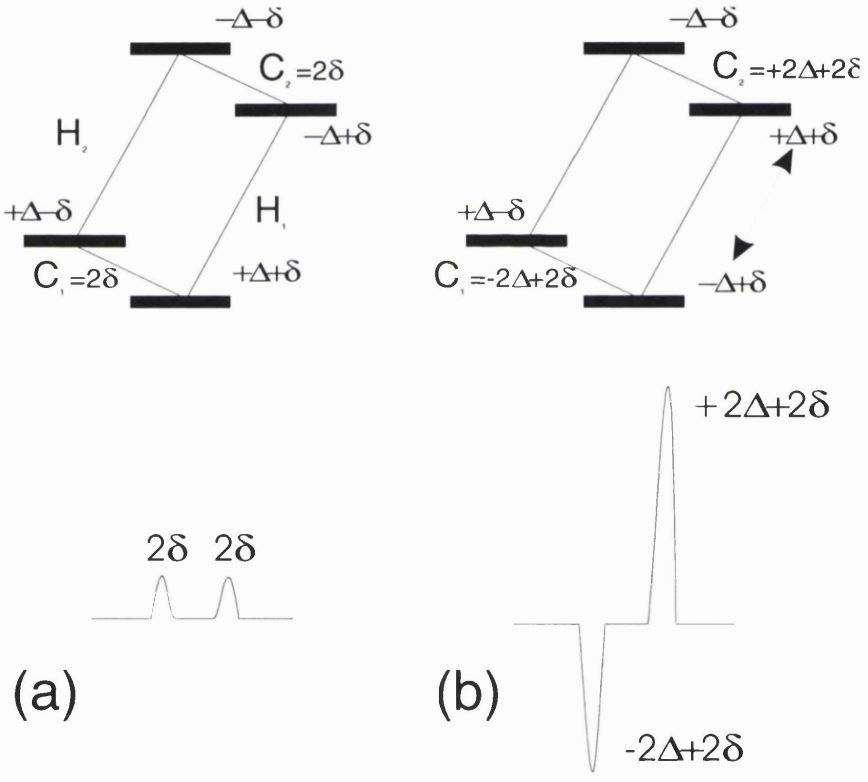


Fig. 4.7 *Energy-level diagram for a heteronuclear AX system showing the populations and transition levels for a ^1H-^{13}C system and the resulting intensities of the carbon signals (Freeman 1997). (a) At Boltzmann equilibrium. (b) After selective proton inversion showing perturbed populations and intensities.*

The antiphase alignment on the *z* axis is important as this is the stage at which the polarisation is transferred from the sensitive nucleus to the insensitive one. This process can be understood with reference to the energies and populations of a nuclear system. In a heteronuclear two-spin AX system such as ^{13}C-^1H, the energy diagram

together with the spectrum that would be obtained is shown in Fig. 4.7. At equilibrium, the spin population difference across the proton transition is 2Δ, while the population difference across the carbon transition is 2δ. Selective inversion of the proton population can be represented by the change in sign of Δ of the two energy levels across the H_1 transition. We can see that the resultant population differences would be $-2\Delta+2\delta$ across the C_1 transition and $+2\Delta+2\delta$ across the C_2 transition, a change of 2Δ for each transition. Given that the polarisation of proton is four times that of carbon ($\Delta=4\delta$), the C_1 and C_2 polarisation would be enhanced by 5 and 3 times respectively. If we sample the carbon magnetisation of carbon after the population inversion, doublet components with relative intensities of +5 and -3 (the antiphase doublet in Fig. 4.7) would be obtained. In INEPT, when the two vectors are aligned in antiphase orientation along the z axis, one proton transition would be in Boltzmann equilibrium, while in the other, population inversion would be achieved, thereby enhancing population difference across the carbon transition. This is the process of polarisation transfer. In SPI, this polarisation is achieved by selective excitation of particular frequency, and is therefore lacking in generality and undesirable for most experiments. In INEPT, the chemical shifts are refocused, therefore all the proton transitions can be inverted regardless of the chemical shifts and hence much more useful experimentally.

The gain in sensitivity in the INEPT step is the ratio of the two gyromagnetic ratio γ_H/γ_X where γ_H and γ_X is the gyromagnetic ratio of proton and X nucleus respectively. The INEPT step is particularly useful for sensitivity enhancement of nuclei such as ^{15}N which has a long T_1 and negative γ. For ^{15}N the signals are enhanced by a factor of 10 (Freeman 1997).

The INEPT sequence can be equally well carried out in the reverse, by applying the initial pulse sequence in the insensitive nucleus channel and observing the signals of the sensitive nucleus. The detection properties for the proton is more favourable than that of the insensitive nuclei as the signal intensity in the spectrum depends on the gyromagnetic ratio γ and the resonance frequency of the observed nucleus. It is therefore preferable to transfer the magnetisation back to the proton. The pulse sequence is identical to the INEPT except that the roles of the ^1H and $^{13}C/^{15}N$ channels are reversed.

HSQC pulse sequence

The basic 2D HSQC pulse sequence is shown in Fig. 4.8. The first part of the HSQC pulse sequence is identical to the INEPT experiment and, as described previously, the two ^{15}N magnetisation vectors are left on the +y and -y axis after this first INEPT step. In order to produce the second dimension, an incremented time delay t_1 then follows. A 180° pulse is placed in the centre of this evolution period t_1. This pulse has the effect of decoupling the ^{15}N heteronuclear splitting due to the attached proton in the F_1 dimension by refocusing the divergence produced by this J_{NH} coupling effect. The t_1 period is incremented by a few to tens of microseconds depending on the desired spectral width in F_1, starting from zero and reaching up to a few milliseconds. The two $90°_x$ pulses after t_1 in the ^{15}N channel rotates the antiphase magnetisation back onto the z axis and transfer the ^{15}N antiphase magnetisation into ^1H antiphase magnetisation. The correlation is achieved between the two nuclei in this crucial 'COSY' step. After two τ time periods (with the 180° pulses in between), the original antiphase magnetisation is refocused to an in-phase magnetisation which is detected under ^{15}N decoupling condition in the t_2 dimension.

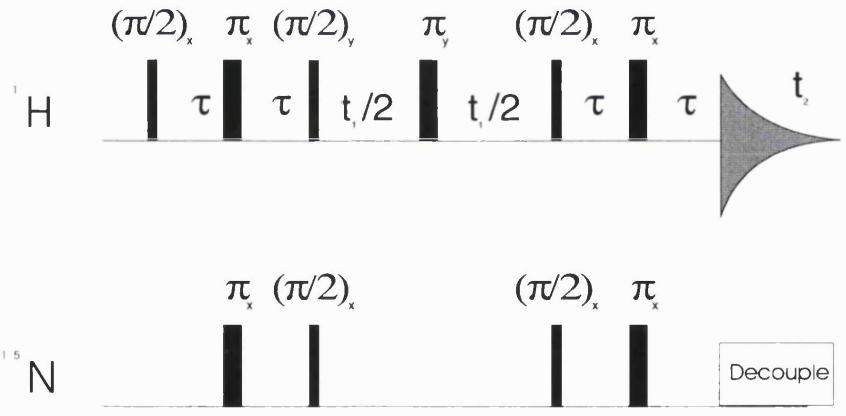


Fig. 4.8 Basic scheme for an HSQC experiment.

This 'round-trip' of polarisation transfer back onto the amide proton can further increase the improvement in the sensitivity. The improvement in sensitivity over detection on the insensitive nuclei X without the INEPT steps is predicted to be $(\gamma_H/\gamma_X)^{5/2}$ which is approximately 32 for ^{13}C and 300 for ^{15}N (Freeman 1997).

A number of modifications are made in the actual HSQC pulse sequence used. In the original pulse scheme, a number of pulse cycling steps (i.e. repeating the experiment with different phases of the pulse) were necessary for spectral editing or isotope filtration (i.e. selecting for signals from protons directly attached to the ^{15}N nucleus), removal of artefacts due to quadrature detection in the indirect dimension, and removal of artefacts due to pulse imperfection as the HSQC requires pulses of precisely 90° and 180° making it sensitive to small error. The phase cycling has now been largely replaced by pulse-field gradients which can be used to remove unwanted water signal, select for the desired coherence transfer pathway and reduce experimental artefacts to a minimum. The advantage of pulsed-field gradients lies in the shortened time necessary for the experiments by reducing the number of phase cycles needed. In the ^{1}H-^{15}N HSQC, the pulsed-field gradients are also used to enhance the sensitivity of the experiment by recording pure absorption HSQC spectra with coherence transfer selection achieved using gradients (Kay *et al.,* 1992).

The pulsed-field gradients can act as a filter by selecting magnetisation with the desired properties at the same time rejecting all other magnetisation components (Kay 1995; Keeler *et al.,* 1994). The application of a magnetic field gradient imparts a phase shift on the magnetisation, the amplitude of which is dependent on the position of the resonating nuclei in the magnetic field, the duration and strength of the field gradient, and the γ of the nucleus in question. A second field gradient that produce the same magnitude in phase change of transverse magnetisation is then applied some time later in the pulse sequence to refocus the magnetisation. However, the phase shift associated with "unwanted" components of magnetisation are not refocused, and can therefore be eliminated from the spectra. Alternatively, the magnetisation of interest may be left intact while the gradient is used to suppress unwanted signals. In the example shown in Fig. 4.9, two gradients are applied in opposite senses (GS1) around the 180° pulse. This dephases the coherences that may be generated before the refocusing pulse as well as any coherences that may be generated by the refocusing pulse. A second field gradient (GS2) is applied before the last 180° pulse to refocus the desired magnetisation.

In this HSQC experiment, the use of pulsed-field gradients reduces the number of phase cycle steps to two. The first 180° pulse in the INEPT step is phase

cycled in order to remove artefacts that can arise as a result of pulse imperfection. Other modifications in this HSQC also include a sensitivity enhancement step which is a modification of the INEPT step in which the two orthogonal in-phase proton magnetisation components are refocused and recorded (Palmer *et al.*, 1992). A soft pulse is also found in the first INEPT step during the water flip-back pulse used for water suppression and small enhancement in sensitivity by the minimisation of cross saturation effects (Grzesiek & Bax 1993). An example of a HSQC pulse sequence used is shown in Fig 4.9.

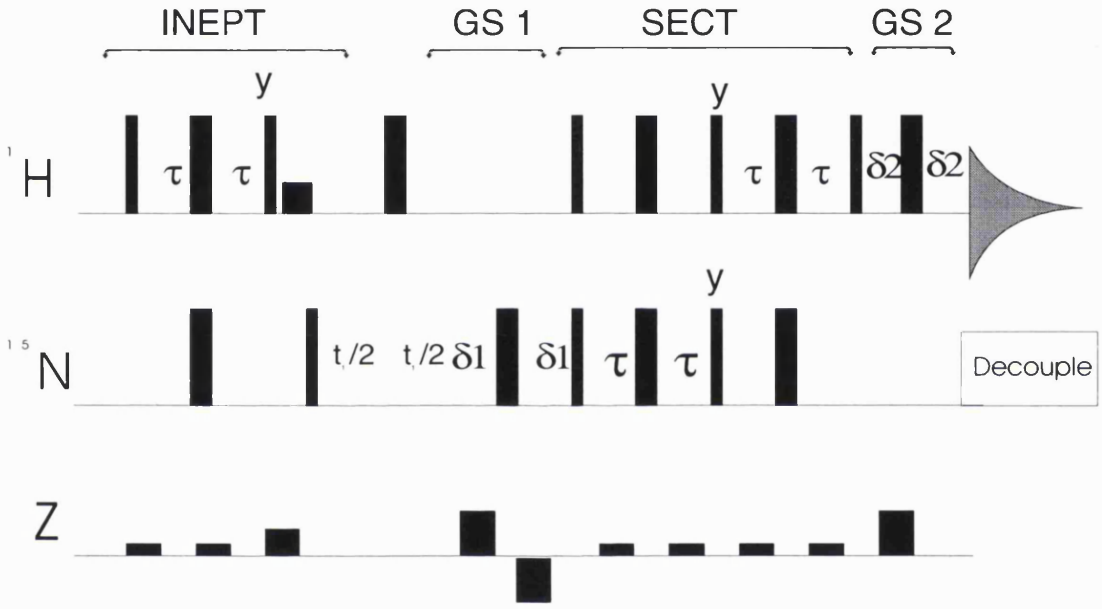


Fig. 4.9 An example of the HSQC pulse sequence used. The delays τ, δ1 and δ2 are set to $1/(4J_{NH})$=2.3, 2.3 and 0.3 ms respectively. Two rectangular gradient selection pulses GS1 and GS2 are used. Soft gradient pulses are used for water suppression, removal of the effects of radiation damping and pulse imperfection. A sensitivity enhanced coherence transfer step (SECT), which is a modification of the INEPT step, refocuses the two in-phase orthogonal magnetisations. The two pulse field gradients applied during the first evolution period is set at 1.25 ms at 30 G/cm and -30 G/cm, and the last gradient is set at 0.25 ms (at a tenth of the first gradient to take account of the difference in gyromagnetic ratio of ^{1}H and ^{15}N) at 30 G/cm. The soft gradients are set at 0.1 ms and for the water flip-back it is set at 2 ms.

Variations of the HSQC experiment that were used included one in which the initial "sampling delay" during the evolution time is set to be half the dwell time, a change that helps to remove base-line distortion in the F_1 dimension (Bax *et al.*, 1991). This also has the advantage of distinguishing the arginine side chain NHε resonances from those of the backbone amides because they appear as negative peaks as resonances that are folded an odd number of times show a phase inversion.

Resonance Assignment of CD2d1

The ^{1}H resonance assignments were originally obtained for the structure determination of rat CD2d1 (Driscoll *et al.,* 1991). However, for NMR studies of CD2d1 aimed at analysing the details of protein-protein interaction at high resolution, such specific residues properties as the side chain pK_a values of titration residues and backbone and side chain dynamics can only be analysed using heteronuclear experiments. It was therefore necessary to extend the original, limited assignment of CD2d1 to a heteronuclear resonance assignment incorporating as many ^{13}C and ^{15}N resonances as possible. The resonance assignment was therefore made from scratch independent from the original assignment.

In this section, the NMR experiments used for the acquisition of the data is briefly introduced. The data acquired from CD2d1 are used to illustrate the experiments and the results discussed.

NMR experiments for the assignment were performed on a 1.5 mM uniformly $^{15}N/^{13}C$ enriched sample of CD2 at a temperature of 25°C in 90% H_2O/10% D_2O in the presence of 20 mM phosphate buffer pH 5.0. Measurement were performed at a proton frequency of 600 MHz on a Varian UnityPlus spectrometer equipped with a 5 mm triple resonance/shielded Z-gradient probe and three RF channels. All the spectra were processed with nmrPipe (Delaglio *et al.,* 1995) or Felix version 2.3 (BIOSYM Inc. San Diego, CA), and analysed with XEASY (Bartels *et al.,* 1995). ^{1}H chemical shifts were referenced to the water signal at 4.774 ppm, and ^{13}C and ^{15}N were referenced indirectly to the water resonance using the $^{13}C/^{1}H$ and $^{15}N/^{1}H$ frequency ratios of $0.2514495192+(T-300)*1.04e^{-9}$ and $0.1013290513+(T-300)*2.7e^{-10}$ respectively, where T is the temperature in Kelvin (Edison *et al.,* 1994). Errors were estimated to be 0.02 ppm for ^{1}H, 0.2 ppm for ^{13}C and 0.1 ppm for ^{15}N chemical shifts.

Overview

The backbone ^{1}H, ^{15}N and ^{13}C resonances of CD2d1 were assigned from the basic set of triple resonance experiments, namely the HNCACB and CACB(CO)NH

spectra. Peaks were first picked in the HSQC spectrum, and the $^{13}C\alpha$, $^{13}C\beta$ and the amide proton and ^{15}N resonances were sequentially assigned by an analysis the HNCACB and CBCA(CO)NH spectra. The assignment of the backbone and some of the side-chain carbonyl groups was obtained from HNCO spectrum. With the assignment of Cα Cβ and carbonyl carbon resonances, the assignment for Hα can be extracted from the CBCACO(CA)HA spectrum. The CBCACO(CA)HA together and HNCO can also be used to provide confirmation of the sequential assignment. The assignment for Hα resonances, together with the assignment for Cα and Cβ, allows the backbone assignment to be extended into the side chains using 3D ^{1}H-^{13}C HCCH-TOCSY, HCCH-COSY and 2D ^{1}H-^{13}C CT-HSQC experiments. The triple resonance pulse sequences used were essentially as proposed by (Kay *et al.,* 1990) and reviewed by (Clore & Gronenborn 1998). Aromatic side chains resonances were assigned from adapted ^{13}C-CT-HSQC pulse sequences and connected to the backbone resonances via (HB)CB(CGCD)HD and (HB)CB(CGCDCE)HE experiments (Yamazaki *et al.,* 1993).

One of the objectives of this assignment was to obtain the side chain assignment for the glutamate carboxyl resonances needed for the pH titration study. Side chain carbonyl carbons were assigned using 2D and 3D versions of H(C)CO and HC(C)CO-TOCSY spectra (modified by Kay *et al.,*). However, assignments at pH 5.0 proved to be difficult to obtain due to exchange processes that resulted in severe peak broadening. The resonance assignment was therefore repeated at pH 3.0 and pH 7.0 where the peak broadening is less severe. As the chemical shifts of the CD2d1 resonances change quite significantly with pH, a smaller set of experiments comprising of ^{1}H-^{15}N HSQC, CBCA(CO)NH, HNCACB, HCCHCOSY, HC(C)CO was performed. This yielded a rather limited set of assignments at pH 3 as well as pH 7. The assignment of side chain carboxylates were then obtained at pH5 by tracing the change in chemical shifts through a pH titration.

Sequential Assignment

A number of triple resonance experiments can be used for sequential assignment without recourse to the proton-proton NOE. The ^{15}N- and ^{13}C-labelled protein allows one-bond and two-bond heteronuclear couplings to be used for

sequential assignment by 3D heteronuclear correlation experiments which provide sequential connectivities between a residue with its preceding residue. A number of independent pathways for linking the resonances of one residue with its preceding residue are therefore available, allowing a nucleus to be correlated with different nuclei using different experiments. This redundancy of information, together with the excellent resolution with decreased spectral overlap and good sensitivity, makes the backbone assignment straightforward and unambiguous.

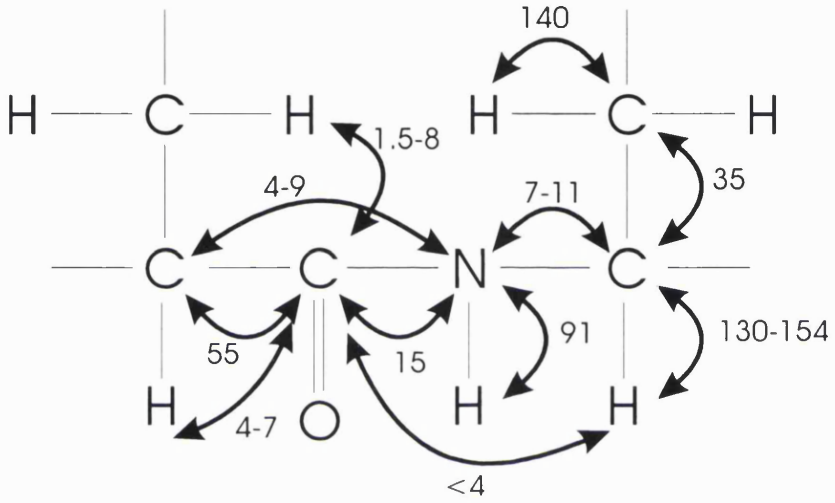
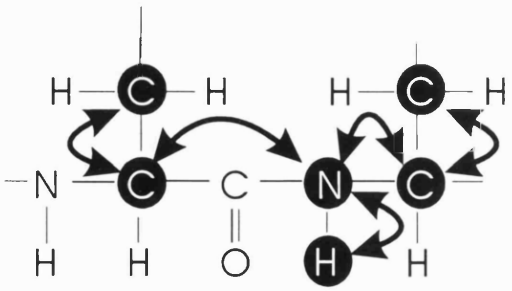
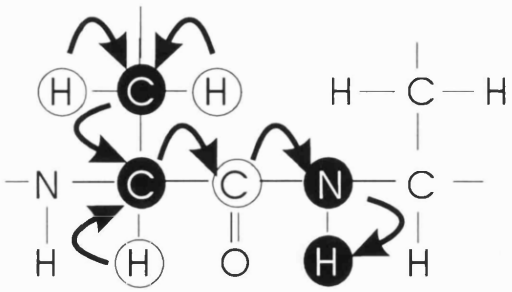


Fig.4.10 A dipeptide segment with summary of the approximate values of the heteronuclear J coupling constants (in Hz) important in the 3D experiments used for the assignment procedure in isotopically enriched protein.

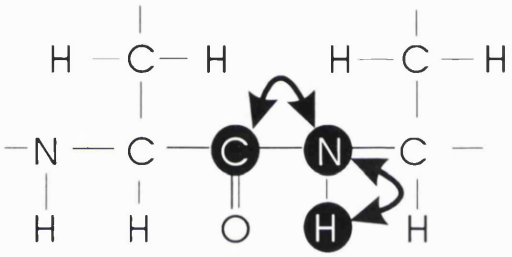
The effectiveness of the 3D heteronuclear experiments relies on the large one-bond scalar heteronuclear couplings along the polypeptide chain that allow the efficient transfer of magnetisation from the highly sensitive ^{1}H nuclei, and making it less sensitive to wide linewidths associated with larger proteins than methods which rely on small proton couplings, thereby increasing the sensitivity of the experiment. The heteronuclear one-bond coupling for the backbone amide $^{1}J_{NH}$ is ~91 Hz, and for $^{1}J_{CH}$ ~125-160 Hz (Fig 4.10). The magnetisation transfer process in a 3D heteronuclear experiment always begins and ends in protons due to the need for enhancement of polarisation of the insensitive nuclei and the high sensitivity of the protons for detection. All the 3D experiments used here therefore start with an INEPT step and end with an INEPT-like sequence.



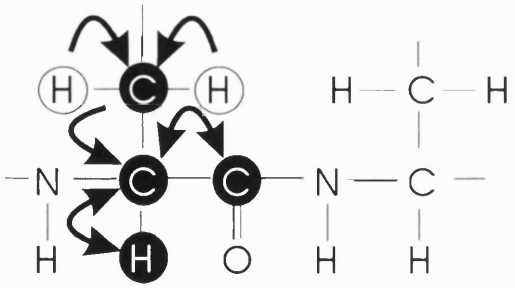
HNCACB



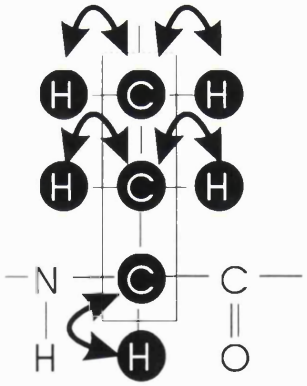
CBCA(CO)NH



HNCO



CBCACO(CA)HA



HCCH-TOCSY

Fig. 4.11 Schematic diagrams of the nuclei that are correlated in the multidimensional NMR experiments used in the assignment of CD2. Nuclei for which the chemical shift is measured are marked by solid circles. Nuclei involved in the magnetisation transfer pathway but not observed are marked with open circles. Magnetisation transfer is marked by curved lines and the direction of the transfer is marked by arrows. The nuclei involved in C-C TOCSY transfer are enclosed by rectangle.

The sets of triple resonance experiments that have been developed for sequential assignment of resonances include, for example, HNCA and HN(CO)CA, HNCO and HN(CA)CO (Kay *et al.*, 1990). The experiments used for the assignment

of CD2d1 are CBCA(CO)NH and HNCACB (Fig. 4.11). Both experiments correlate the amide NH to the Cα and Cβ signals. The HNCACB spectra provides inter- and intra-residue connectivities while CBCA(CO)NH spectra, which only provides inter-residue connectivities, allows the inter- and intra-residue resonances to be discriminated. Furthermore, information on amino acid types can be obtained from ^{13}C chemical shift data by the combination of the chemical shift of $^{13}C\alpha$ and $^{13}C\beta$. In particular, the distinctive chemical shifts of alanine, threonine, serine and glycine residues provide easy identification of these amino acid types from the CBCA(CO)NH and HNCACB spectra. With the alignment of sequentially connected spin system with the known amino acid sequence, the sequential assignment can be determined with relative ease. A small number of experiments are generally sufficient for the complete sequence specific assignment and in cases where there is ambiguity, the sequential assignment can be confirmed by combination of other sets of experiments using other kind of correlations but involving at least one shared nuclei. However, it was found that CBCA(CO)NH and HNCACB were sufficient for the unambiguous and complete sequential assignment of CD2d1.

HSQC

The process of the assignment first begins with peak-picking in the spectrum from the ^{1}H-^{15}N HSQC experiment (Fig.4.12). This experiment gives the backbone amide resonances of all the residues in a protein, with proline which lacks an amide proton as the sole exception. Some of the side chain NH groups such as those of asparagine, glutamine, arginine and tryptophan are also detected. Both asparagine and glutamine residue produce four peaks for each residue, two each from the two protons in the NH_2 groups, while a further two peaks comes from the NHD formed as a result of exchange of the proton for deuterium with the D_2O used as lock signal. Only 10% D_2O is used and the signal from NHD produced is therefore proportionally less intense.

In the ^{1}H -^{15}N HSQC spectra for CD2d1, the peaks are generally sharp, well-dispersed and with little overlap. The ^{1}H-^{15}N spectra also contain a few weak cross-peaks. The glycines amide cross peaks are situated at the top left region of the 2D spectra, while the asparagine and glutamine side-chain lies on the top left. Two

tryptophan side-chain cross peaks can be observed at their characteristic positions in the bottom left of the spectra, while the C-terminus residue (Glu99) is revealed as a strong peak situated near the bottom right with a characteristically large upfield shift with increasing pH. The backbone resonances showed little significant overlap apart from a cluster of 5 peaks at the proton chemical shift of ~9 ppm, and ~129 ppm in the ^{15}N chemical shift.

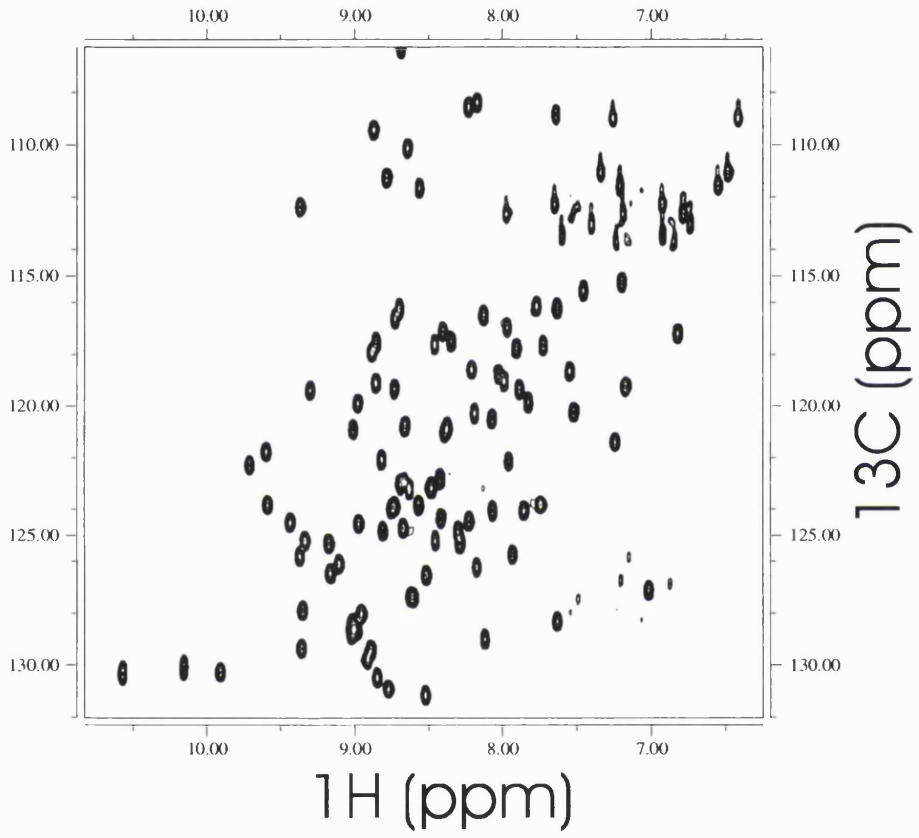


Fig. 4.12 Spectrum of the ^{1}H-^{15}N HSQC experiment of CD2d1.

CBCA(CO)NH

This experiment records the correlation between the amide signal and the Cα and Cβ resonances of the preceding residue (Grzesiek & Bax 1992). The experiment utilises the large ^{1}J coupling between protons and the ^{13}C nuclei to transfer the magnetisation. Magnetisation from proton is transferred to Cβ and Cα by an INEPT step, and the magnetisation on Cβ is then transferred to Cα by a COSY mixing pulse. Inter-residue correlations are established by transferring coherence via the

intervening ^{13}CO spin. For detection, the magnetisation is transferred to the amide protons.

Two peaks for every residues are observed in this experiment (Fig 4.13), with the exception of glycine which has only the Cα carbon, and residues that precede proline as proline lacks an amide proton therefore cannot produce any peaks in this experiment. The correlation on the side-chains of the glutamines and asparagines can also be observed (the amide signal to the Cα and Cβ in asparagines, and Cβ and Cγ in glutamines) in this experiment, it is therefore possible to assign their resonances. The projection of this 3D spectrum onto the 2D ^{1}H-^{15}N plane should yield a spectrum nearly identical to that of the ^{1}H-^{15}N HSQC experiment. However, some of the side-chain resonances such as those of tryptophans and arginines are absent.

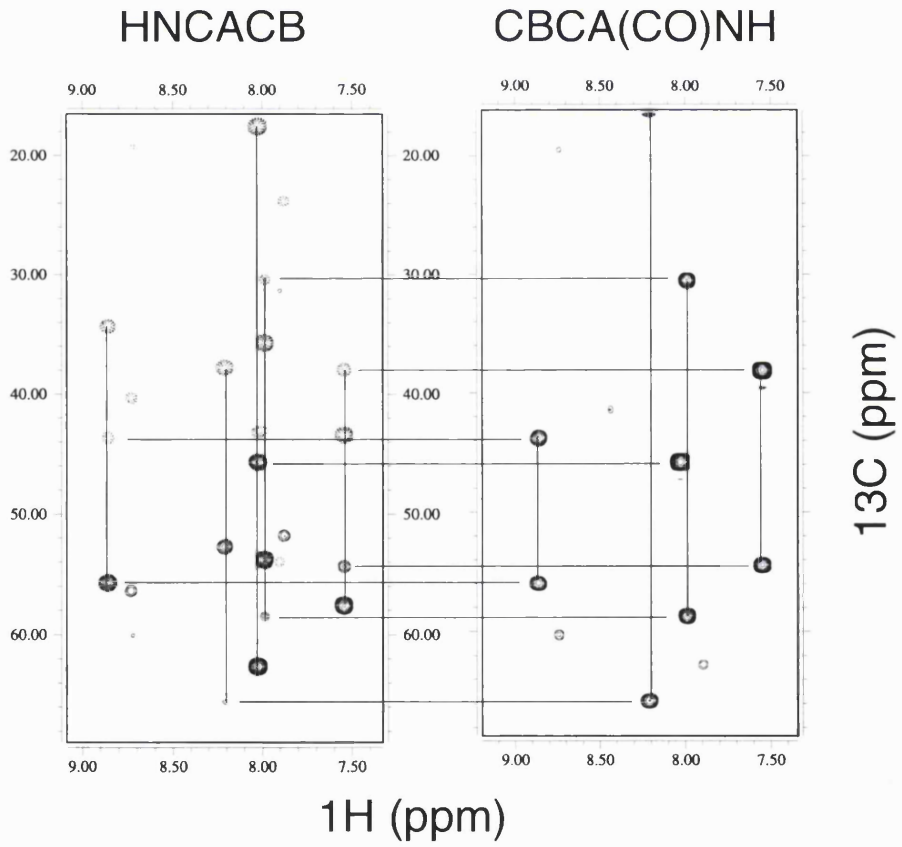


Fig. 4.13 Corresponding slices of HNCACB and CBCA(CO)NH spectra of CD2d1. Note that the inter-residue peaks are weaker in HNCACB spectrum. The CBCA(CO)NH spectrum contains only inter-residue resonances, while HNCACB spectrum contains both inter- and intra-residue resonances.

HNCACB

This experiment provides connectivities between the amide of a residue with its Cβ and Cα atoms as well as those of its preceding residue (Wittekind & Mueller 1993; Yamazaki *et al.,* 1994). Together with the CBCA(CO)NH data it allows the inter- and intra-residue resonances to be distinguished, with the inter-residue connectivities easily identified using CBCA(CO)NH. The HNCACB pulse sequence utilises the 'out-and-back' coherence transfer, whereby the amide proton magnetisation is the point of the initial excitation as well as the point of detection during t_3 in a manner analogous to the ^{1}H-^{15}N HSQC pulse sequence. After the transfer of magnetisation from the amide proton, the magnetisation of the ^{15}N nuclei is transferred to the Cα via the one-bond $^{13}C\alpha$-^{15}N *J* coupling (7-11 Hz) which is relatively inefficient process, particularly in larger proteins for which the $^{13}C\alpha$ line width can be significantly larger than the *J* coupling of ~11 Hz. While this magnetisation is evolving, the magnetisation can also be transferred via the two-bond $^{13}C\alpha$-^{15}N *J* coupling (7-9 Hz) to the Cα of the preceding residue, thereby establishing inter- and intra-residue connectivities at the same time. This inter-residue correlation however is weaker and the peaks for the preceding residue are therefore generally weaker than intra-residue signals in HNCACB experiment. However, as the ranges of the coupling constant of the one-bond and two-bond $^{13}C\alpha$-^{15}N can overlap, the intensity of the signal is not always reliable as a mean of distinguishing the inter- and intra-residue connectivities. It is possible to obtain the same correlation using a different experiment called CBCANH, however, CBCANH is more sensitive to protein line-width problem and therefore less useful for larger proteins.

Apart from the differences in their chemical shift, the Cα and Cβ resonances can also be differentiated as the Cα resonances appear as positive peaks while the Cβ resonances appear negative in the HNCACB spectrum (Fig. 4.14). This inversion of the Cβ peaks is the result of the transfer of magnetisation from the Cα to Cβ in which an additional evolution period and a $90°_x$ pulse put the two peaks into antiphase orientation. An overlay of the HNCACB and CBCA(CO)NH therefore allows easy identification of the inter- and intra-residues connectivities as well as Cα

and Cβ. In this experiment, correlation of the arginine side chain Nε and NεH to its Cδ and Cγ is also observed.

Nearly all the resonances of CD2d1 that are expected to be observed in these two experiments can be observed. The exceptions being the Cα of Arg87 and inter-residue Cβ of Ile88 in Thr89 strip. However, it is possible to observe these resonances as inter-residue peaks in the case of Arg87 and intra-residue peak in the case of Ile88. The Cβ of Trp32 is ~2 ppm outside the normal value expected for the tryptophan but this may be a result of ring current effect from the tryptophan or Tyr76 which is situated nearby.

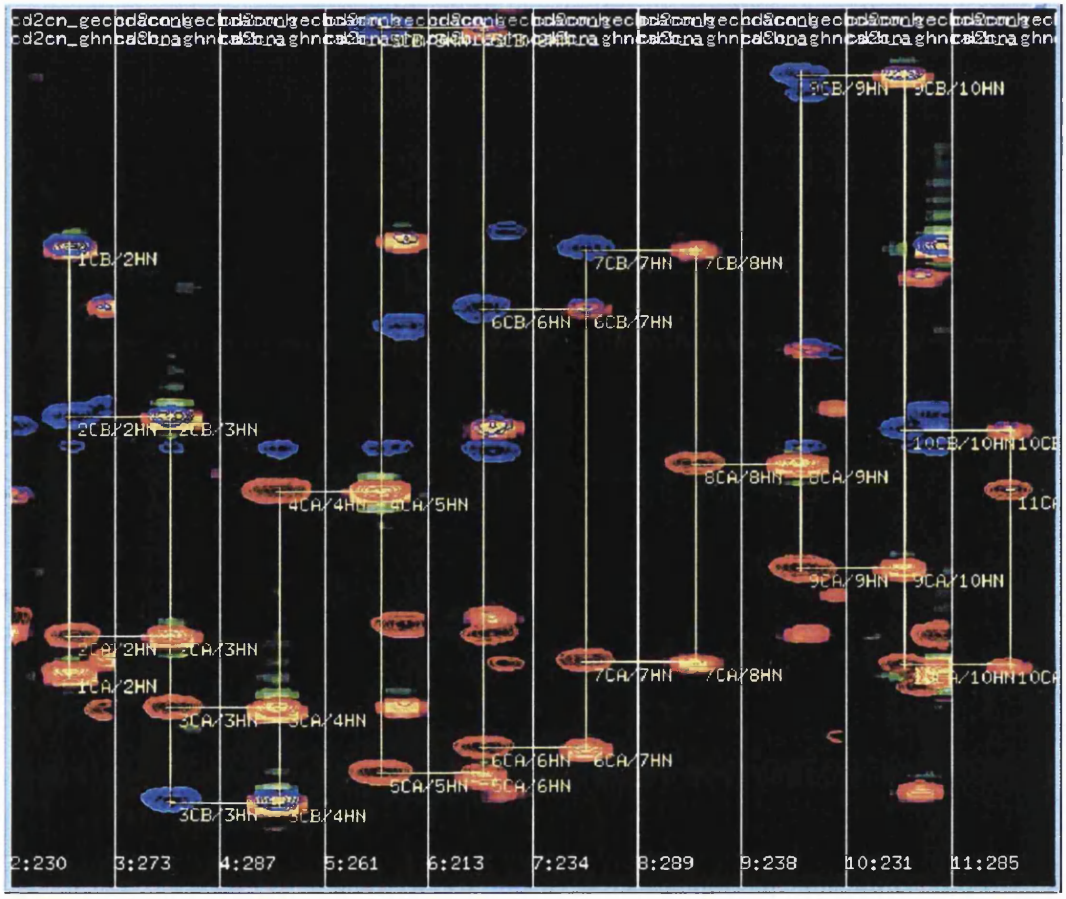


Fig. 4.14 Overlay strips of spectra of CBCA(CO)NH and HNCACB from CD2d1. The HNCACB is in contour, and CBCA(CO)NH is in intensity plot (yellow). The resonances have been assigned and the residues 2-11 are shown in this figure together with the connectivities. For the HNCACB, the red contours indicate positive peaks from the Cα nuclei while blue contours indicate negative peaks from the Cβ nuclei. Note the distinctive chemical shifts of serine (Ser3), threonine (Thr4), alanine (Ala9) and glycine (Gly4 and Gly8).

HNCO

This experiment correlates the amide ^{1}H and ^{15}N chemical shift of one amino acid with the carbonyl (C') chemical shift of the preceding residue via the one-bond $^{1}J_{NC'}$ coupling (~15 Hz), thereby providing sequential connectivity (Kay *et al.,* 1990). In conjunction with the CBCACO(CA)HA experiment which provides intra-residue connectivities, the HNCO can be used for independent assessment of the correctness of the sequential assignment. As in the HSQC and HNCACB experiments, the magnetisation is originated and detected at the amide protons. Apart from the backbone C' to NH connectivities, this experiment also provides correlations for the side chains of asparagine and glutamine (Fig. 4.15). In addition, the arginine guanidino group which has a coupling similar to the carboxyamide groups also give correlations between the $^{15}N\varepsilon$, $^{15}N\varepsilon H$ and $^{13}C\zeta$ chemical shifts as the Nε-$^{13}C\zeta$=NH group has similar coupling constant to the N-^{13}C=O group.

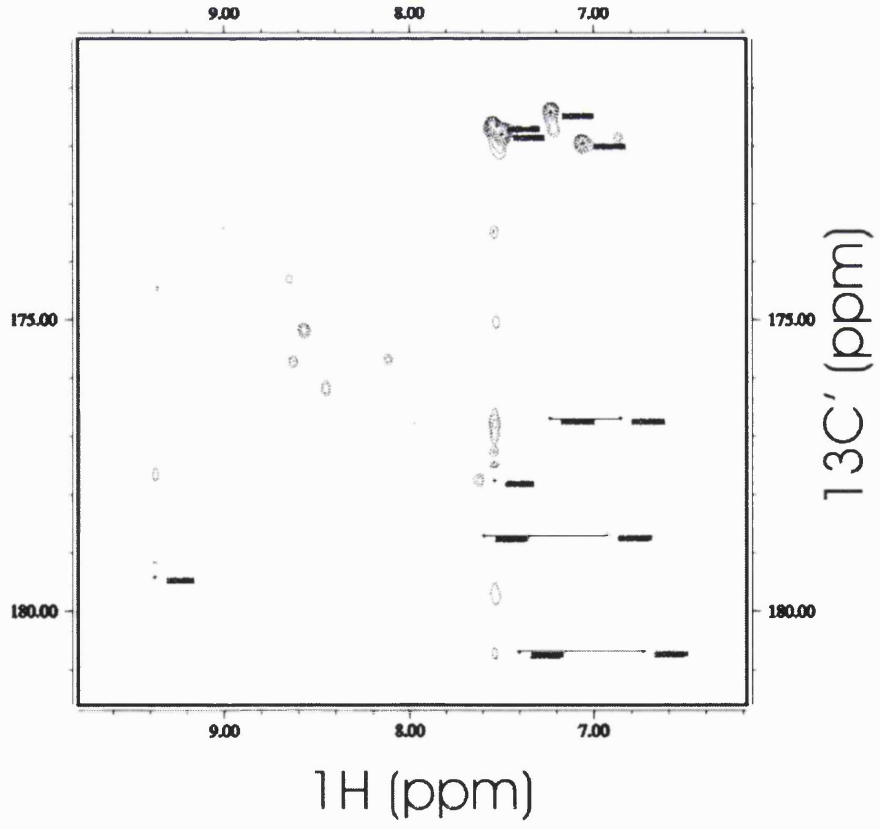


Fig. 4.15 A 2D slice of the CD2d1 HNCO experiment which shows peaks from the backbone as well as arginine, glutamine and asparagine side-chains. Glutamine and asparagine produces two side-chain peak per residue (shown on bottom right). Arginine side chain peaks are folded and appear as negative peaks shown in dotted lines.

CBCACO(CA)HA

This experiment correlates the Cβ and Cα chemical shift with the carbonyl (C') and Hα resonances (Kay 1993). The path of the magnetisation transfer is as follows:

$$H\beta/\alpha \xrightarrow{J_{CH}} C\beta/\alpha(t_1) \xrightarrow{J_{C\alpha C\beta}} C\alpha \xrightarrow{J_{C\alpha C'}} C'(t_2) \xrightarrow{J_{C\alpha C'}} C\beta\alpha \xrightarrow{J_{CH}} H\alpha(t_3)$$

The active couplings involved in each magnetisation transfer step are indicated above each arrow. In contrast to the experiments previously described, the Hα resonances instead of amide proton chemical shifts are detected during acquisition. The pulsed field gradients used in this experiment allow the spectra to be obtained on samples dissolved in H_2O instead of D_2O. It should be possible to obtain the assignment for the carboxyl groups of the acidic residues in this experiment, however, it was found that only the rather weak aspartate side chain carboxyl resonances were discernible, while most of the glutamate carboxyl resonances (the C-terminal Glu99 being the exception) were too weak to be observed (Fig. 4.16). Glycines appeared as negative peaks and the intensities are weak due to inefficient coherence transfer to the $C\alpha H_2$ group.

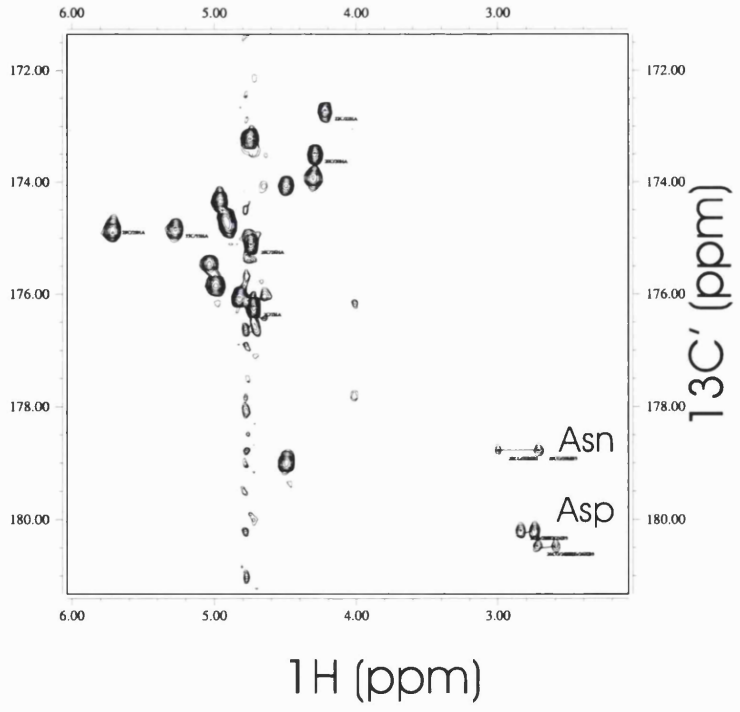


Fig.4.16 A 2D slice of the CBCACO(CA)NH experiment. The cross peaks from the backbone (strong peaks) and asparagine and aspartate side chains (weak peaks at bottom right) are shown.

Side chain resonance assignments

Once the α-protons and the α-carbon have been assigned, the assignment can then be extended to the side chains. In a double-labelled protein, the α-proton and α-carbon nucleus can be correlated with the side chain aliphatic protons and carbon nuclei of the side chain using 3D HCCH-TOCSY or HCCH-COSY experiments. For CD2d1, HCCH-TOCSY which gives correlation to all the side-chain aliphatic protons and carbon nuclei of a particular amino acid was used. 2D CT-HSQC was also used to aid the assignment using HCCH-TOCSY. Some of the resonances of the side chain functional groups such as the carboxyl groups were incompletely assigned, and more experiments which provide the necessary correlations were therefore performed in order to complete the assignments. A further set of experiments were conducted to assign the aromatic groups.

HCCH-TOCSY

Once the assignment for the Cα, Cβ and Hα has been obtained, the HCCH-TOCSY spectra can be used to confirm and obtain complete ^{1}H and ^{13}C assignment of the side chains. The HCCH-TOCSY experiment permits the tracing of the ^{13}C-^{13}C scalar connectivity in a manner analogous to the homonuclear TOCSY experiment, but with greater enhanced sensitivity as it employs the large $^{1}J_{CC}$ couplings (~140 Hz) during the isotropic mixing period rather than the small $^{3}J_{HH}$ couplings (0-10 Hz) (Bax *et al.,* 1990). The pathway of the transfer of magnetisation is as follows:

$$^{1}H \rightarrow {}^{13}C \rightarrow {}^{13}C \rightarrow {}^{1}H$$

After the evolution period t_1, ^{1}H magnetisation is transferred to its directly bonded ^{13}C nucleus via a refocused INEPT step. The ^{13}C magnetisation evolves during t_2 under the influence of its ^{13}C chemical shift, and the coherence is then forced to spread throughout the system by isotropic mixing during a DIPSI-3 composite pulse mixing sequence. Finally the ^{13}C magnetisation is transferred back to the attached proton by a reverse refocused INEPT and is detected during t_3. In HCCH-COSY a COSY-type mixing transfer step is placed between the two INEPT steps instead of the isotropic mixing used in HCCH-TOCSY.

In principle, in a 3D HCCH-TOCSY spectrum, all the protons of a side chain can be found in one carbon plane, making the identification of spin system easier. The signals, however, are not symmetrical. In practice, not all the peaks are necessarily present in one plane as the spectral width of ^{13}C is too large for the efficient isotropic mixing on which the intensities of the cross-peaks depends. The isotropic mixing depend on the rate at which magnetisation flows through a *J*-coupled spin system during the mixing period which is in turn dependent on the topology of the spin system, the size of the pertinent *J* couplings and the relaxation rate of the protons. Nevertheless, complete spin systems can often be traced in other side-chain carbon planes.

The spectrum of the HCCH-TOCSY is multiply folded in the ^{13}C dimension with data outside the spectral range folded back in. This folding is necessary to improve resolution. The folding is possible as the chemical shifts of protons attached to different carbon atoms has distinct chemical shifts ranges, therefore the peaks from different carbon atoms do not overlap significantly and the pattern of diagonal peaks and cross peaks observed can be used for the identification of the carbon chemical shifts and whether the carbon planes are folded. Where there is some ambiguity, folded and unfolded CT-HSQC spectra can be used to aid identification. In Fig. 4.17B, all the diagonal peaks lies downfield to its cross-peaks, indicating that they are Hα resonances, while in 4.17A, the diagonal peaks arise from both Hβ and Hα atoms. Many residue types can also be identified from their characteristics of the spin system. Some of the peaks appeared to be weak as the coupling constant is set at the optimum for a three bond coupling. The intensity of the peaks is dependent on a number of factors such as the residue type, the mixing time, the line width and flexibility of the side chains.

A few proton resonances in CDd1 are substantially upfield shifted, for example, β-proton resonance of Leu63 at -1.16 ppm and the β and γ proton resonances of Val78 at -0.32 and -0.38 ppm respectively. This is likely to be due to these aliphatic proton being located immediately above aromatic rings, and the increased shielding due to the circulating π electrons from the ring current in the aromatic residues shift these peaks upfield.

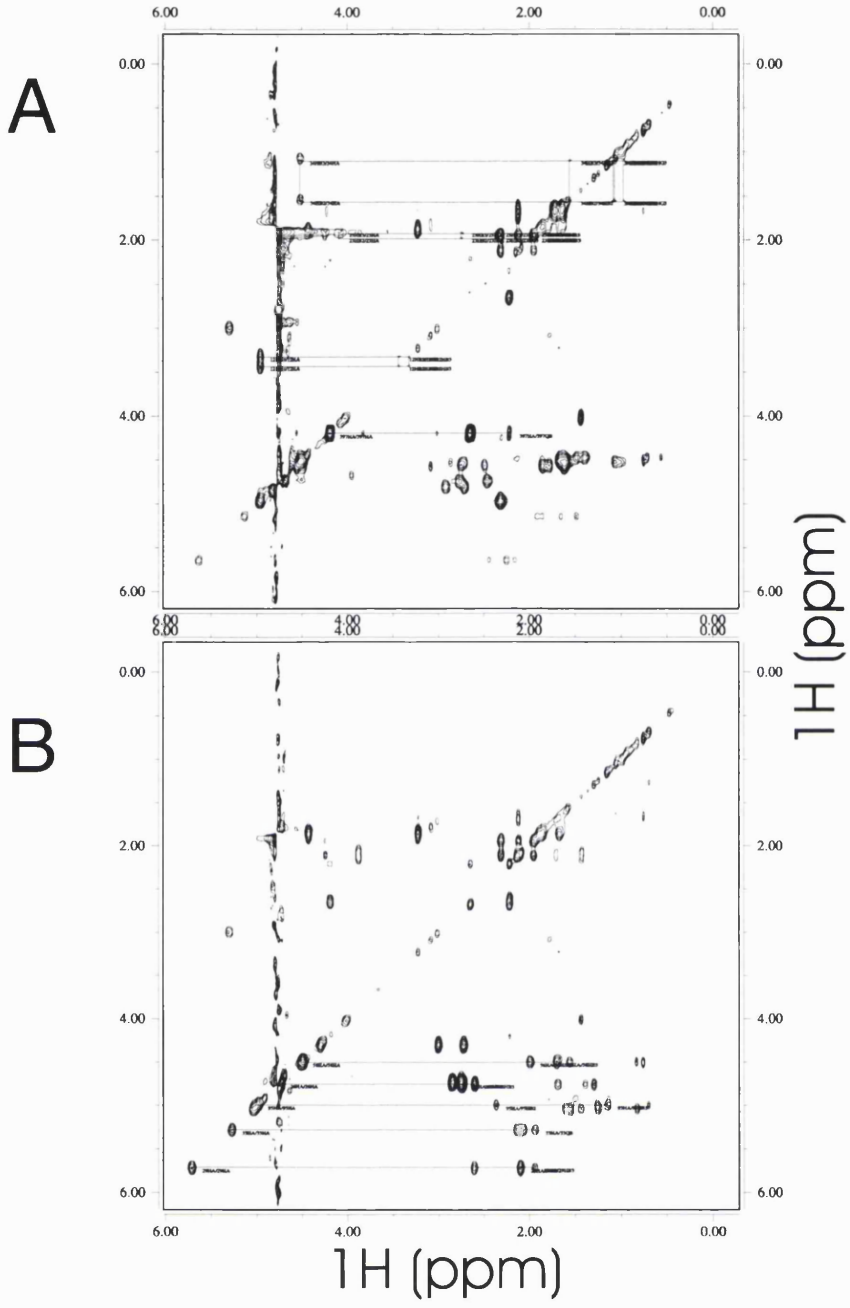


Fig. 4.17 Two slices of the CD2d1 3D HCCH-TOCSY spectrum. The approximate carbon chemical shifts can be inferred from the position of the diagonal peaks of the protons. (A) Slice showing Hβ diagonal peaks and peaks from their spin systems, a single Hα diagonal peaks lies at downfield position. (b) Slice showing Hα diagonal peaks which lie downfield to other peaks in their spin systems.

^{1}H-^{13}C CT-HSQC

This HSQC experiment provides correlation between the aliphatic carbon and proton (Vuister & Bax 1992) and the constant-time (CT) version of the experiment significantly enhances the resolution of the spectrum. The constant-time version differs from the generalised HSQC experiment in the way the heteronuclear single

quantum coherence evolves during the time period t_1 between the two INEPT segments. It's called constant time because the entire evolution period is held fixed, while the 180° pulse which is normally positioned in the middle now shifts in each increment. The time preceding the 180° pulse is set to be T/2 + t_1/2, while that which follows is set at T/2 - t_1/2 where T is the total constant-time period. When T is set to be $1/J_{cc}$, the sign of the ^{13}C magnetisation of carbons attached to an odd number of aliphatic carbon will be opposite to those with even number. This property is useful for providing information about the number of aliphatic carbons attached any particular nucleus (Fig. 4.18). For example, the Cβ of methionine has 2 directly attached carbons and appeared as negative peaks, while Cγ which has only 1 carbon attached appeared as positive peak. The Cε of methionine which is not observed in HCCH-TOCSY can also be observed in CT-HSQC. A CT-HSQC folded in the same way as the HCCH-TOCSY and an unfolded spectrum were recorded for CD2d1.

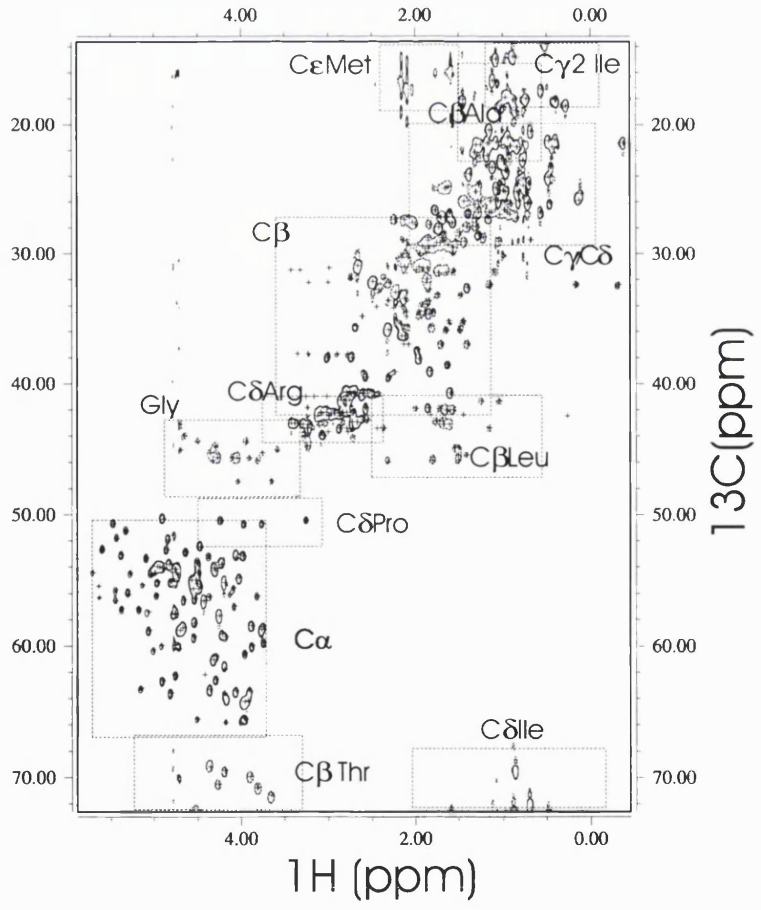


Fig. 4.18 The unfolded ^1H-^{13}C CT-HSQC spectrum of CD2d1 with the approximate areas where particular atom types may be found shown. Note, however, the Cδ resonances of isoleucines are folded. The positive peaks are in solid line while the negative peaks are shown in dotted line.

Assignments of the Aromatic Protons

The assignment for the $^{1}H\delta$ and $^{1}H\varepsilon$ of the aromatic groups are assigned using (HB)CB(CGCD)HD and (HB)CB(CGCDCE)HE pulse sequences. These 2D experiments provide correlations between the side chain $^{13}C\beta$ and the $^{1}H\delta$ and $^{1}H\varepsilon$ of the aromatic ring by transfer of magnetisation via scalar couplings (Yamazaki *et al.*, 1993). The (HB)CB(CGCDCE)HE experiment provides both Hε and Hδ peaks while (HB)CB(CGCD)HD experiment provides only Hδ peaks, and the combination of these data thereby allow scope for discrimination between the two sets of aromatic ^{1}H signals. These experiments are sensitive and side-step some of the traditional problems associated with the assignment of the aromatic side chains. For example, in TOCSY or COSY spectra, it is not possible to get unambiguous residue-specific assignment as there is no proton on the Cγ atom of the residue to provide connectivities, and it is not easy to transfer magnetisation using HCCH-TOCSY/COSY as there is too wide a gap in the chemical shift between Cβ and Cγ resonances.

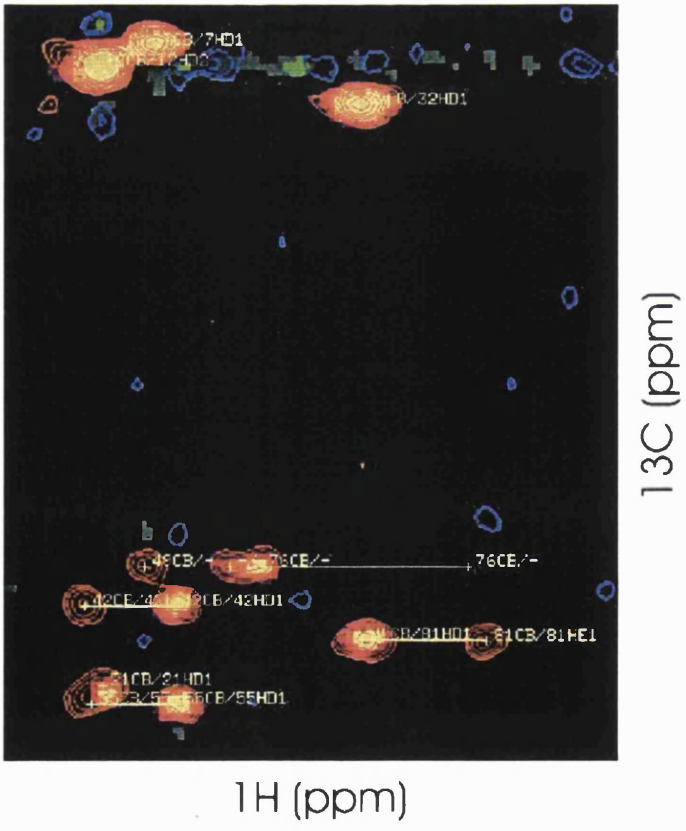


Fig. 4.19 Overlay of (HB)CB(CGCD)HD (intensity plot) and (HB)CB(CGCDCE)HE (contour plot) spectra. The intensity plot gives the Hδ peaks of the aromatic residues, while the contour plots gives the Hδ and Hε resonances.

The (HB)CB(CGCD)HD and (HB)CB(CGCDCE)HE experiments are limited to the assignment for Hδ and Hε of tyrosine and phenylalanine, and Hδ of tryptophan and histidine (Fig 4.19). Nonetheless, having unambiguously assigned these resonances, it is possible to further extend the assignment using homonuclear TOCSY and COSY, or CT-HSQC experiments (Fig. 4.20). The is, however, some degree of degeneracy of the resonances, and a few of the phenylalanine resonances therefore remain unassigned for CD2d1.

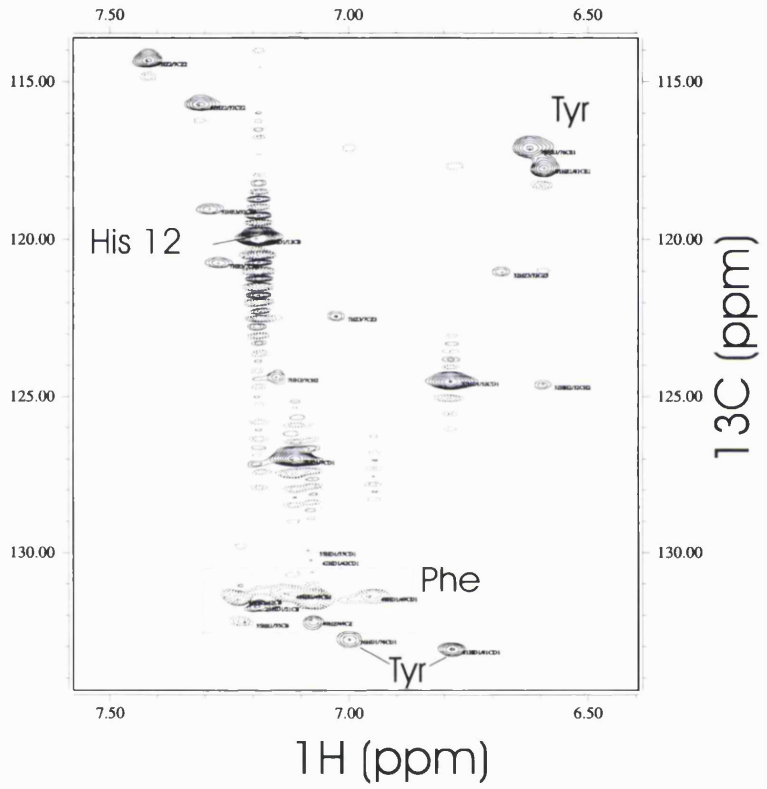


Fig. 4.20 ^{13}C-CT-HSQC spectrum of the aromatic groups of rat CD2d1. A very strong signal from histidine Hδ can be observed.

Assignment of Carboxyl Carbon

The side-chain carboxyl resonances of CD2d1 were assigned using a combination of 2D and 3D H(C)CO spectra, as well as 2D H(C)CO-TOCSY experiments connecting Cγ/δ carboxylate in Asp/Glu directly with the α-proton (Zhang & Gmeiner 1996). As mentioned earlier, the severe exchange broadening of the peaks made unfeasible the assignment of some of the side-chain carboxyl carbon resonances at pH 5. The backbone assignment was therefore repeated at pH 3 and 7

using the HNCACB and CBCACONH pair of experiments. The main purposes of the assignment at low pH was to determine the chemical shifts of the glutamates, particularly that of Glu29 and Glu41 and this was achieved with the analysis of the H(C)CO-TOCSY spectra at different pH value. The H(C)CO-TOCSY spectrum provides correlation of the carboxyl carbon with Cα and Cβ signals (Fig. 4.21). At low pH the cross peaks for Glu29 and Glu41 are weak but nevertheless it was possible determine their chemical shifts. A number of different TOCSY mixing times were tried with 6.5 ms chosen as the most suitable. The chemical shifts of Glu29 and Glu41 carboxyl carbon at pH 5 was then obtained by tracing the movement of peaks in a pH titration series using the 2D HC(C)O experiment.

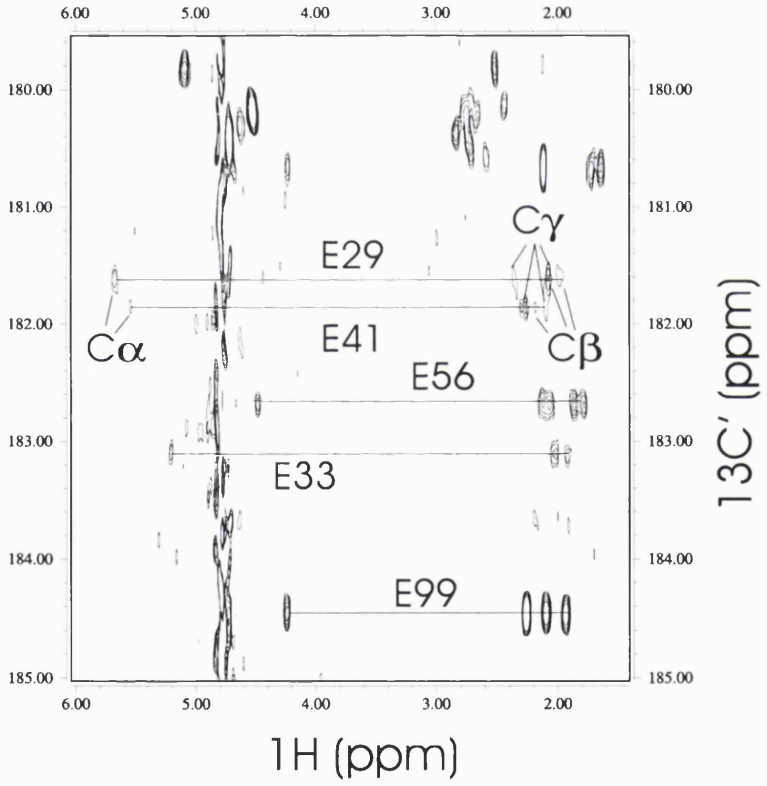


Fig. 4.21 The H(C)CO-TOCSY spectrum showing the connectivities of the Cα, Cβ/Cγ with the carbonyl resonances.

A selected region of the 2D $^{1}H,^{13}C$ H(C)CO spectra of CD2d1 at three different pH values is shown in Figure 4.22. The resonance frequencies of most of the carboxylates lie within the expected range. For most protonated aspartates and glutamates, the carboxyl ^{13}C resonances are located between 176-178 ppm and 178-

180 ppm respectively, while the signals from that of the ionised forms are ~3-5ppm downfield as expected (Jeng & Dyson 1996; Oda *et al.,* 1994; Richarz & Wuthrich 1978). The β-CH_2 protons of aspartates and γ-CH_2 glutamates lie between 3.0 - 2.7 ppm and 2.6 -2.2 ppm respectively with an expected upfield shift of 0.2-0.3 ppm upon deprotonation (Bundi & Wuthrich 1979). In contrast to this general pattern, at pH 5.0 an unusual upfield δ-carboxyl chemical shift is seen for Glu41, at a position more typical of a protonated glutamate. Similar but less pronounced anomalies are also seen for Glu29, Asp28 and Asp72. These observations suggested the presence of at least one residue with an anomalously high pK_a value. Even though Glu41 and Glu29 are surface residues, each of them exhibited two non-degenerate resonances for their Cγ protons, suggesting a lack of side-chain mobility. The characteristics of the pK_a values for CD2d1 are described in full in Chapter 5.

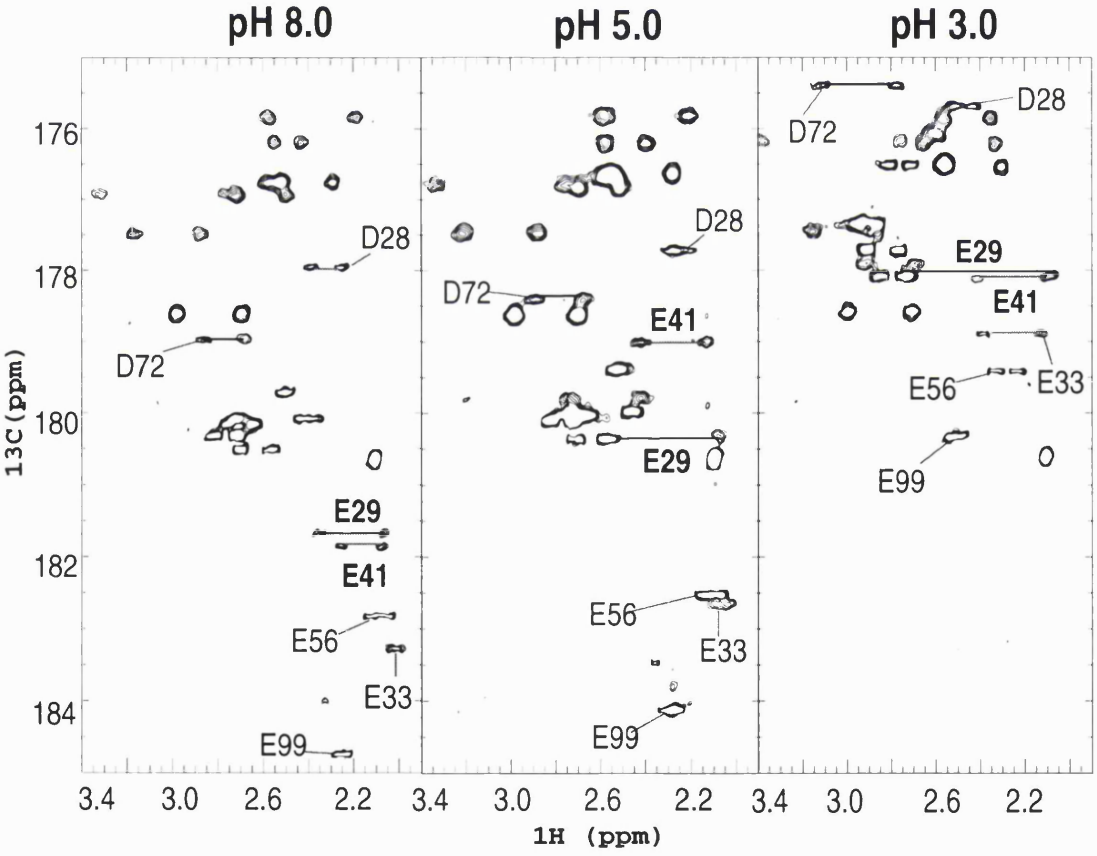


Fig. 4.22 A selected region of the 2D H(C)CO spectrum for CD2d1 at 3 different pH values.

Secondary Structures from the Chemical Shift

The secondary structure formation in a protein has been shown to induce very specific and highly significant chemical shifts changes for all the amino acid residues (Wishart & Sykes 1994). The α-proton, for instance, tends to be shifted upfield on helix formation while a downfield shift of its chemical shift tends to suggests β-sheet formation. The chemical shifts of carbon and proton can therefore provide some indication of secondary structure and methods for such analysis have been proposed and used to predict the secondary structure of proteins (Pastore & Saudek 1990; Wishart *et al.,* 1992). In this simple analysis, the chemical shift dispersion is calculated as the deviation in chemical shifts from the 'random coil' value and plotted against residue number (Fig. 4.23). The 'random coil' values used in this analysis are obtained from Wishart & Sykes (1994).

The CD2d1 structures as previously determined by NMR (Driscoll *et al.,* 1991) and crystallography (Jones *et al.,* 1992) has shown the protein to be composed entirely of β-sheets. The is directly reflected in chemical shifts dispersion of assigned resonances shown in Fig. 4.23. Particularly good agreement is shown by the carbonyl carbon chemical shifts in which stretches of residues with negative values are indicative of presence of β-sheet structures. Reasonably good agreement is also obtained from the Hα where the β-sheets may be indicated by regions of residues with positive values, while for Cα (negative values taken as indication of β-sheet structures), the agreement is patchy. This analysis therefore to some extend indicates the general correctness of the assignments. While it is possible to further analyse the data using other method such as chemical shift index (Wishart & Sykes, 1994), including using more chemical shift data (e.g. ^{15}N) in order to eliminate chemical shift noise, the structure of CD2d1 is known and little advantage can be gained from further analysis. It is possible to use chemical shifts to generate structure restraints and for 3D structural refinement (Wishart& Nip, 1998), however, this is beyond the scope of the thesis.

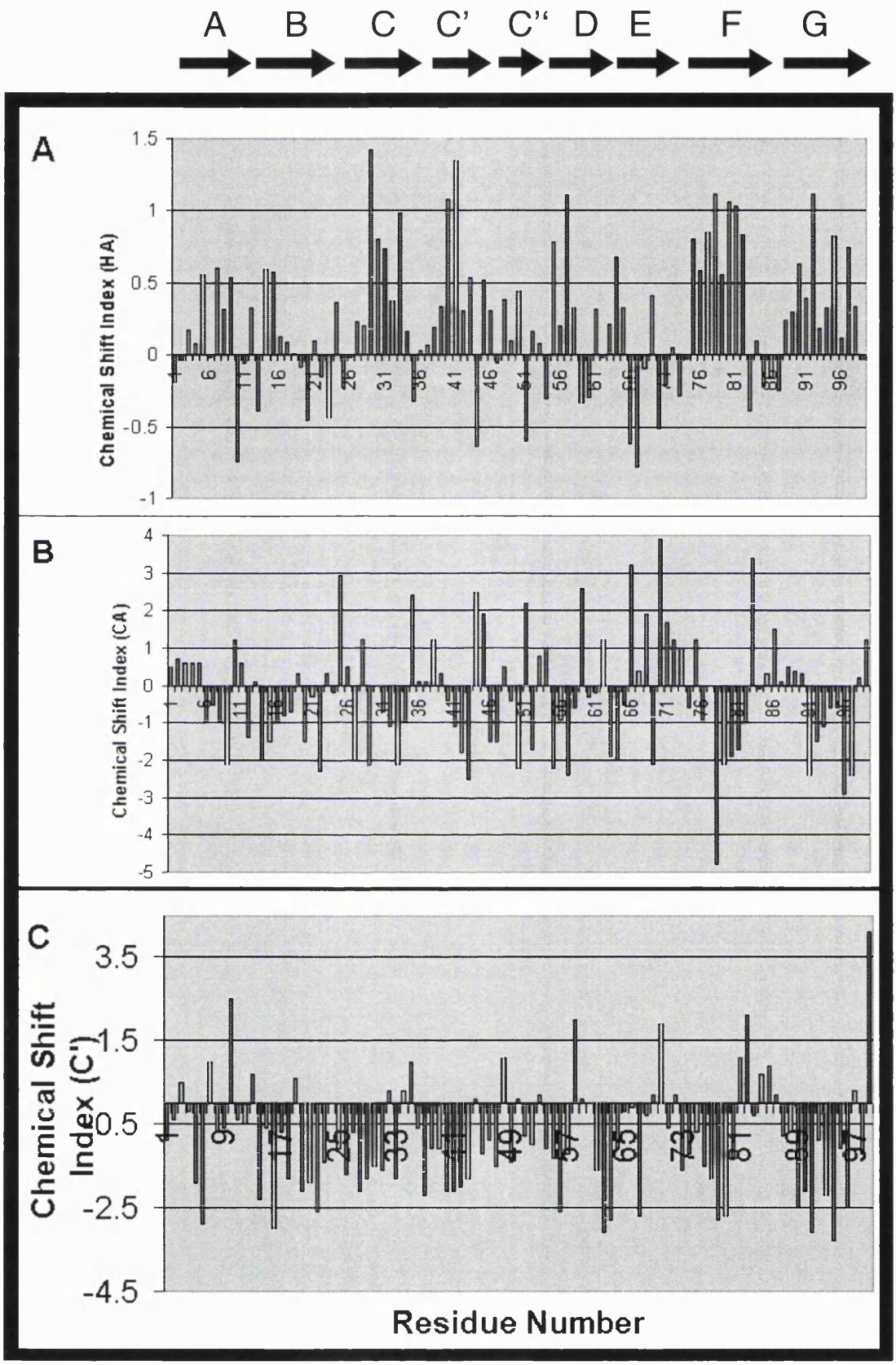


Fig. 4.23 Chemical shift dispersion (Δδ) of A) α-proton, B) α-^{13}C atom and C) backbone ^{13}C carbonyl resonances for CD2d1. The secondary structure as derived from the NMR and crystal structures of CD2d1 is shown on top.

Resonance Assignment of the Mutants

A number of mutants of CD2d1 were created in order to study the effects of the mutations on the biophysical properties of the protein. For a number of these mutants, such as E41Q, E29Q, E33Q and E28N, the backbone amide resonances of the mutants did not show significant changes in chemical shifts at low pH (pH 2-5), with the exception of a few resonances that correspond to residues near to the mutation sites as well as the mutated residues themselves (Fig. 4.24). The assignments of the majority of the resonances for these CD2d1 mutants are therefore relatively straightforward. In cases where ambiguity arose, especially in crowded regions of the spectrum, the resonances can also be identified by their distinctive titration profiles by following the chemical shift variation in resonances in ^{1}H-^{15}N HSQC experiments. As an example, in the E41Q mutant, most the resonance remain little changed at low pH, and have titration curves that closely resembled that of the wild-type CD2 in the range pH 2 - 5. The Gln41 substitution in this residue showed a change in its titration curves, however, at low pH its chemical shift is virtually identical to that of the Glu41 in wild-type CD2d1.

For some other mutants, however, there were greater changes for some of the chemical shifts of backbone resonances. This was particularly true of the mutants involving the aromatic and aliphatic residues such as Y81A, F49A and L38A. Even though only a limited number of the resonances showed significant change in chemical shifts, attempts were made to ascertain the assignments by using 3D ^{15}N-NOESY-HSQC spectra, an experiment which gives sequential connectivities by providing NH_i-$NH_{i\pm1}$ NOEs in ^{15}N-labelled protein sample. However, some of these mutants also proved to be difficult to purify, with the protein suffering from heavy precipitation during the purification. It was found not to be possible to get concentrated samples of these mutants. With protein sample of concentration 0.2 mM or less, the application of 3D spectroscopy to obtain assignments were unsuccessful due to low signal-to-noise ratio. The assignments therefore relied on comparison of the chemical shifts and pH titration curves obtained in 2D ^{15}N-^{1}H HSQC spectra with those of the wild-type protein. Where there are uncertainties in the assignments the resonances are left unannotated.

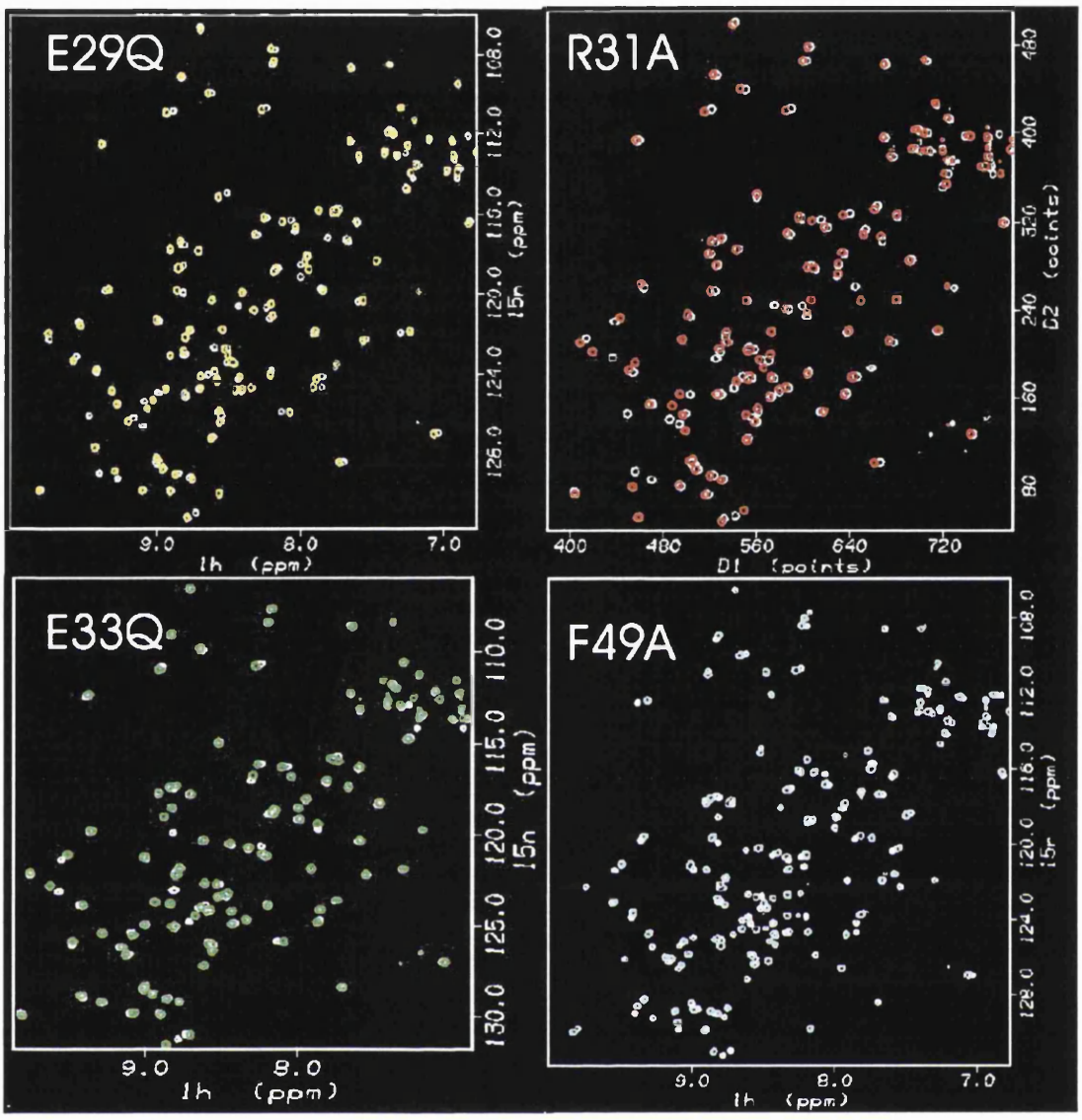


Fig. 4.24 A selection of overlay ^{1}H-^{15}N HSQC spectra of different mutants of CD2d1 at low pH. The peaks in white is of wild-type CD2d1, the yellow is from E29Q mutant, the red R31A mutant, the green E33Q mutant and the blue F49A mutant.

Summary of Resonance Assignments

An essentially complete resonance assignments were obtained for ^{1}H, ^{15}N and ^{13}C resonances based on a uniformly double isotope-labelled protein sample of rat CD2d1. These include most of the aromatic ^{13}C and ^{1}H as well as $^{15}N/^{13}C$ assignments of arginine side chains. Not assigned are some of the side chain resonances in lysine, histidine and phenylalanine residues. Assignments are deposited with the BioMagResBank (http://www.bmrb.wisc.edu/), accession number 4109

(Chen *et al.,* 1998). A more limited additional backbone and partial side chain assignments were provided at pH values of 3.0 and 7.5. A list of the resonance assignments of wild-type rat CD2d1 at pH 5.0 can be found in the Appendix.

References

Abragam, A. 1961. *Principles of Nuclear Magnetism* Oxfod University Press

Aue, W. P., Bartholdi, E., Ernst, R. R. 1976. Two-Dimensional spectroscopy. Appllcation to nuclear magnetic resonance. *Journal of Chemical Physics* 64:2229-2246

Bartels, C., Xia, T. H., Billeter, M., Guntert, P., Wuthrich, K. 1995. The Program XEASY For Computer-Supported NMR Spectral-Analysis of Biological Macromolecules. *Journal of Biomolecular NMR* 6:1-10

Bax, A., Clore, G. M., Gronenborn, A. M. 1990. H-1-H-1 Correlation Via Isotropic Mixing of C-13 Magnetization, a New 3-Dimensional Approach For Assigning H-1 and C-13 Spectra of C-13- Enriched Proteins. *Journal of Magnetic Resonance* 88:425-431

Bax, A., Ikura, M., Kay, L. E., Zhu, G. 1991. Removal of F1-Base-Line Distortion and Optimization of Folding in Multidimensional NMR-Spectra. *Journal of Magnetic Resonance* 91:174-178

Bloch, F., Hansen, W. W., Packard, M. 1946. Nuclear Induction. *Physical Review* 69:127

Bodenhausen, G., Kogler, H., Ernst, R. R. 1984. Selection of Coherence-Transfer Pathways in NMR Pulse Experiments. *Journal of Magnetic Resonance* 58:370-388

Bodenhausen, G., Ruben, D. J. 1980. Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Lett.* 69:185-189

Bundi, A., Wuthrich, K. 1979. 1H-NMR Parameters of the Common Amino Acid Residues Measured in Aqueous Solutions of the Linear Tetrapepetide H-Gly-Gly-X-L-ALA -OH. *Biopolymers* 18:285 - 297

Cavanagh, J., Fairbrother, W. J., Palmer III, A. G., Skelton, N. J. 1996. *Protein NMR spectroscopy - Principles and Practice* London: Academic Press

Chen, H. A., Pfuhl, M., Davis, B., Driscoll, P. C. 1998. Sequence specific 1H, 13C and 15N resonance assignment of rat CD2 domain 1. *Journal of Biomolecular NMR* 12:457-458

Clore, G. M., Gronenborn, A. M. 1991. Applications of 3-Dimensional and 4-Dimensional Heteronuclear NMR - Spectroscopy to Protein-Structure Determination. *Progress in Nuclear Magnetic Resonance Spectroscopy* 23:43-92

Clore, G. M., Gronenborn, A. M. 1998. Determining the structures of large proteins and protein complexes by NMR. *Trends in Biotechnology* 16:22-34

Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., Bax, A. 1995. NMRpipe - a Multidimensional Spectral Processing System Based On Unix Pipes. *Journal of Biomolecular NMR* 6:277-293

Derome, A. E. 1987. *Modern NMR Techniques for Chemistry Research* Pergamon Press Ltd.

Driscoll, P. C., Cyster, J. G., Campbell, I. D., Williams, A. F. 1991. Structure of Domain-1 of Rat Lymphocyte-T CD2 Antigen. *Nature* 353:762-765

Edison, A. S., Abildgaard, F., Westler, W. M., Mooberry, E. S., Markley, J. L. 1994. Practical Introduction to Theory and Implementation of Multinuclear, Multidimensional Nuclear-Magnetic-Resonance Experiments. *Methods in Enzymology* 239:3-79

Ernst, R. R., Andersen, W. A. 1966. Application of Fourier Transform Spectroscopy to Magnetic Resonance. *Review of Scientific Instruments* 37:93

Ernst, R. R., Bodenhausen, G., Wokaun, A. 1990. *Principles of Nuclear Magnetic Resonance in One and Two Dimensions.* Paperback ed. Oxford, Clarendon Press

Evans, J. N. S. 1995. *Biomolecular NMR spectroscopy* Oxford, Oxford University Press

Freeman, R. 1997. *A Handbook of Nuclear Magnetic Resonance.* 2nd ed. Longman

Freeman, R. 1998. *Spin Choreography - Basic Steps in High Resolution NMR* Oxford, Oxford University Press

Garrett, D. S., Seok, D. I., Peterkofsky, A., Gronenborn, A. M., Clore, G. M. Garrett DS, 1999. Solution structure of the 40,000 Mr phosphoryl transfer complex between the N-terminal domain of enzyme I and HPr.. *Nature Structure Biology* 6(2):166-173

Grant, D. M., Harris, R. K. 1996. Encyclopaedia of NMR spectroscopy Vol. 1 *Historical Perspectives.* John Wiley & Sons

Grzesiek, S., Bax, A. 1992. Correlating Backbone Amide and Side-Chain Resonances in Larger Proteins By Multiple Relayed Triple Resonance NMR. *Journal of the American Chemical Society* 114:6291-6293

Grzesiek, S., Bax, A. 1993. The Importance of Not Saturating H2O in Protein NMR - Application to Sensitivity Enhancement and NOE Measurements. *Journal of the American Chemical Society* 115:12593-12594

Jeener, J. 1971. Homonuclear Shift Correlated Spectroscopy. In *AMPERE Summer School.* Basko Polje, Yugoslavia

Jeng, M. F., Dyson, H. J. 1996. Direct Measurement of the Aspartic-Acid-26 pK(a) For Reduced Escherichia-Coli Thioredoxin By C-13 NMR. *Biochemistry* 35:1-6

Jones, E. Y., Davis, S. J., Williams, A. F., Harlos, K., Stuart, D. I. 1992. Crystal-Structure At 2.8-Angstrom Resolution of a Soluble Form of the Cell-Adhesion Molecule CD2. *Nature* 360:232-239

Kay, L. E. 1993. Pulsed-Field Gradient-Enhanced 3-Dimensional NMR Experiment For Correlating C-13-Alpha-Beta, C-13', and H-1-Alpha Chemical-Shifts in Uniformly C-13-Labeled Proteins Dissolved in H2O. *Journal of the American Chemical Society* 115:2055-2057

Kay, L. E. 1995. Field Gradient Techniques in NMR-Spectroscopy. *Current Opinion in Structural Biology* 5:674-681

Kay, L. E., Ikura, M., Tschudin, R., Bax, A. 1990. 3-Dimensional Triple-Resonance NMR-Spectroscopy of Isotopically Enriched Proteins. *Journal of Magnetic Resonance* 89:496-514

Kay, L. E., Keifer, P., Saarinen, T. 1992. Pure Absorption Gradient Enhanced Heteronuclear Single Quantum Correlation Spectroscopy With Improved Sensitivity. *Journal of the American Chemical Society* 114:10663-10665

Keeler, J., Clowes, R. T., Davis, A. L., Laue, E. D. 1994. Pulsed-Field Gradients - Theory and Practice. *Methods in Enzymology* 239:145-207

Kelly, G., Prasannan, S., Daniell, S., Fleming, K., Frankel, G., S. Matthews. 1999. Structure of the cell-adhesion fragment of intimin from enteropathogenic Escherichia coli. *Nature Structural Biology* 6:313-318

Kumar, A., Welti, D., Ernst, R. R. 1975. NMR Fourier Zeugmatography. *Journal of Magnetic Resonance* 18:69-83

Morris, G. A., Freeman, R. 1979. Enhancement of Nuclear Magnetic Resonance Signals by Polarisation Transfer. *Journal of American Chemical Society* 101:760-762

Muller, L., Kumar, A., Ernst, R. R. 1975. Two-dimensional carbon-13 NMR spectroscopy. *Journal of Chemical physics* 63:5490-5491

Oda, Y., Yamazaki, T., Nagayama, K., Kanaya, S., Kuroda, Y., Nakamura, H. 1994. Individual Ionization-Constants of All the Carboxyl Groups in Ribonuclease H1 From Escherichia-Coli Determined By NMR. *Biochemistry* 33:5275-5284

Oschkinat, H., Muller, T., Dieckmann, T. 1994. Protein-Structure Determination With 3-Dimensional and 4-Dimensional NMR-Spectroscopy. *Angewandte Chemie-International Edition in English* 33:277-293

Palmer, A. G., Cavanagh, J., Byrd, R. A., Rance, M. 1992. Sensitivity Improvement in 3-Dimensional Heteronuclear Correlation NMR-Spectroscopy. *Journal of Magnetic Resonance* 96:416-424

Pastore, A., Saudek, V. 1990. The Relationship Between Chemical-Shift and Secondary Structure in Proteins. *Journal of Magnetic Resonance* 90:165-176

Pervushin, K., Riek, R., Wider, G., Wuthrich, K. 1997. Attenuated T-2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proceedings of the National Academy of Sciences of the United States of America* 94:12366-12371

Purcell, E. M., Torrey, H. C., Pound, R. V. 1946. Resonance Absorption by Nuclear Magnetic Moments in a solid. *Physical Review* 69:37-38

Richarz, R., Wuthrich, K. 1978. High-field 13C Nuclear Magnetic Resonance Studies at 90.5 MHz of the Basic Pancreatic Trypsin Inhibitor. *Biochemistry* 17:2263 -2269

Sattler, M., Fesik, S. W. 1996. Use of deuterium labeling in NMR: Overcoming a sizeable problem. *Structure* 4:1245-1249

Shan, X., Gardner, K. H., Muhandiram, D. R., Rao, N. S., Arrowsmith, C. H., Kay, L. E. 1996. Assignment of N-15, C-13(alpha), C-13(beta), and HN resonances in an N-15, C-13, H-2 labelled 64 kDa trp repressor-operator complex using triple-resonance NMR spectroscopy and H-2-decoupling. *Journal of the American Chemical Society* 118:6570-6579

Shuker, S. B., Hajduk, P. J., Meadows, R. P., Fesik, S. W. 1996. Discovering high-affinity ligands for proteins: SAR by NMR. *Science* 274:1531-1534

Sorensen, O. W., Eich, G. W., Levitt, M. H., Bodenhausen, G., Ernst, R. R. 1983. Product Operator-Formalism For the Description of NMR Pulse Experiments. *Progress in Nuclear Magnetic Resonance Spectroscopy* 16:163-192

Tjandra, N., Bax, A. 1997. Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* 278:1111-1114

Vuister, G. W., Bax, A. 1992. Resolution Enhancement and Spectral Editing of Uniformly C-13-Enriched Proteins By Homonuclear Broad-Band C-13 Decoupling. *Journal of Magnetic Resonance* 98:428-435

Wishart, D. S., Nip, A. M. 1998. Protein chemical shift analysis: a practical guide. *Biochem Cell Biol* 1998;76(2-3):153-163

Wishart, D. S., Sykes, B. D. 1994. Chemical-Shifts As a Tool For Structure Determination. *Methods in Enzymology* 239:363-392

Wishart, D. S., Sykes, B. D., Richards, F. M. 1992. The Chemical-Shift Index - a Fast and Simple Method For the Assignment of Protein Secondary Structure Through NMR-Spectroscopy. *Biochemistry* 31:1647-1651

Wittekind, M., Mueller, L. 1993. HNCACB, a High-Sensitivity 3d NMR Experiment to Correlate Amide- Proton and Nitrogen Resonances With the Alpha-Carbon and Beta-Carbon Resonances in Proteins. *Journal of Magnetic Resonance Series B* 101:201-205

Wuthrich, K. 1986. *NMR of proteins and nucleic acids* Wiley

Yamazaki, T., Formankay, J. D., Kay, L. E. 1993. 2-Dimensional NMR Experiments For Correlating C-13-Beta and H-1- Delta/Epsilon Chemical-Shifts of Aromatic Residues in C-13-Labeled Proteins Via Scalar Couplings. *Journal of the American Chemical Society* 115:11054-11055

Yamazaki, T., Lee, W., Arrowsmith, C. H., Muhandiram, D. R., Kay, L. E. 1994. A Suite of Triple-Resonance NMR Experiments For the Backbone Assignment of N-15, C-13, H-2 Labeled Proteins With High-Sensitivity. *Journal of the American Chemical Society* 116:11655-11666

Zhang, W. X., Gmeiner, W. H. 1996. Improved 3D gd-HCACO and gd-(H)CACO-TOCSY experiments for isotopically enriched proteins dissolved in H2O. *Journal of Biomolecular NMR* 7:247-250

Chapter 5

Determination of pK_a value

Electrostatic effect dominates the activities and properties of protein and electrostatic interaction plays an important role in a wide range of biologically important protein functions (Nakamura, 1996; Perutz, 1978). It is fundamental in the catalytic mechanism of enzymes (Fersht, 1985), and the interplay of the electrostatic forces may strongly affect the stability of the proteins (Perutz, 1978). Long-range electrostatic interactions can also act as a steering force in the guidance of substrates or ligands toward enzyme actives sites or other binding sites and stabilising the complex formed between enzymes and ligands (Russell & Fersht, 1987; Wade *et al.*, 1998). A detailed knowledge of the pH-dependent behaviour of protein can therefore contribute significantly to the understanding of protein function. The acidity constant (pK_a) which reports on the charge state of an ionisable group, is a sensitive probe of the local electrostatic potential, and the pK_a values of catalytically and functionally important residues can have important implications for the functioning of the proteins. The determination of pK_a is therefore a subject of considerable interest to both biochemist and biophysicist.

As previously described in Chapter 1, the characteristic low affinity and high specificity binding behaviour of CD2 have been attributed to its unusually flat and highly charged ligand binding surface (Davis *et al.*, 1998). Electrostatic complementarity is postulated to be of critical importance to the binding interaction of rat CD2 and the charged residues play a dominant role in CD2 ligand recognition by ensuring that the binding strength is of the appropriate magnitude (i.e. not excessively strong) whilst maintaining high specificity (Davis *et al.*, 1998). Moreover, the crystal structure of the complex between human CD2 and CD58 showed the contact surface consists exclusively of complex network of charged groups (Wang *et al.*, 1999). Electrostatic interactions are therefore crucial to the function of adhesion molecules and a detailed understanding of the electrostatic properties of CD2 would provide invaluable insight into the high specificity and low affinity binding characteristics of adhesion molecules. The roles played by electrostatics in the protein-protein interactions of antibodies have been widely studied (Gibas *et al.*, 1997; Novotny & Sharp, 1992). However, the binding properties of antibody-antigen complexes are significantly different from that of the low-affinity adhesion molecules such as CD2. The binding interactions of CD2 to its ligands are 10^4 to 10^5 fold weaker than binding by antibodies. Surface shape complementarity and hydrophobic interaction play significant roles in the antibody-antigen interaction, while charge-charge interaction predominates in the ligand binding interaction of CD2 with little shape complementarity between interaction molecules (Davis *et al.*, 1998). Few detailed studies of the electrostatic properties of adhesion molecules had been undertaken, and in this study, the pH-dependent chemical shift changes of the ^{1}H, ^{13}C, ^{15}N and ^{19}F resonances were examined in order to provide a starting point for a comprehensive evaluation of the pH-dependent properties of rat CD2d1 and the charge status on the ligand-binding surface.

Determination of ionisation constant by NMR

The classical titration curve relates the changes in the ionisation states of potential hydrogen binding sites or groups in protein to the changes in pH, and so defines the sum of the individual changes in site occupancy of a protein as a function

of pH. It is difficult to identify and determine in such a titration curve the individual pK_a of any particular charged residue in a protein. Therefore, in order to determine the individual pK_a (defined as the pH at which an ionisable group is considered half-dissociated), a titration curve that primarily reflects the ionisation event of a particular residue is necessary. A number of methods have been used for the determination of the pK_a, such as enzyme kinetics (Brocklehurst, 1994; Lin *et al.*, 1992), spectrometric methods such as UV, infra-red or Raman spectroscopy (Li *et al.*, 1993; Nozaki & Tanford, 1967; Szaraz *et al.*, 1994; Takahashi & Creighton, 1996), X-ray crystallography (Berisio *et al.*, 1999) and theoretical modelling of electrostatic interactions (Warshel & Papazyan, 1998).

The best method to date for the determination of the pK_a of individual titration group, however, is by NMR spectroscopy in which the changes in chemical shifts of NMR resonances are monitored as a function of pH. The chemical shift of a particular resonance reflects in part the electrostatic charge field environment of each nuclei, and the protonation state of nearest ionisable group may be the dominant contributor. This chemical shift may come from the averaging of the NMR frequencies from the protonated and deprotonated states in rapid equilibrium, and the chemical shift therefore reflects the relative population of the protonated and deprotonated forms. Tracing the shift in resonance with pH therefore allows the ionisation state of a particular charged residue to be directly monitored and the pK_a to be determined (Fig 5.1). In the early days of protein NMR, only the well-resolved imidazole proton resonances permitted the ionisation equilibria of histidines to be studied (Bradbury & Scheraga, 1966; Cohen *et al.*, 1970; Meadows *et al.*, 1967). With the advent of multi-dimensional NMR and isotopic labelling of protein, however, the chemical shift changes of specific ^{1}H, ^{15}N, or ^{13}C resonances can be monitored, and the individual microscopic pK_a of other ionisable groups can therefore be determined. Determination of pK_a by NMR is now well-established (for a selection of such studies see Bartik *et al.*, 1993; Ebina & Wuthrich, 1984; Forman-Kay *et al.*, 1992; Joshi *et al.*, 1997; Khare *et al.*, 1997; Kohda *et al.*, 1991; Kuhlman *et al.*, 1999; Perez-Canadillas *et al.*, 1998; Schaller & Robertson, 1995; Singer & Forman-Kay, 1997; Sorensen & Led, 1994; Tishmack *et al.*, 1997). For ^{13}C-NMR in particular, the chemical shift changes are dominated by the ionisation state of the carboxyl groups, and with the benefit of a large change in the carboxyl carbon

chemical shift (~3 ppm) with pH, the acid dissociation equilibria of carboxylic acids can be determined with great precision (Chivers *et al.*, 1997; Jeng & Dyson, 1996; March *et al.*, 1982; Oda *et al.*, 1994; Qin *et al.*, 1996; Richarz & Wuthrich, 1978).

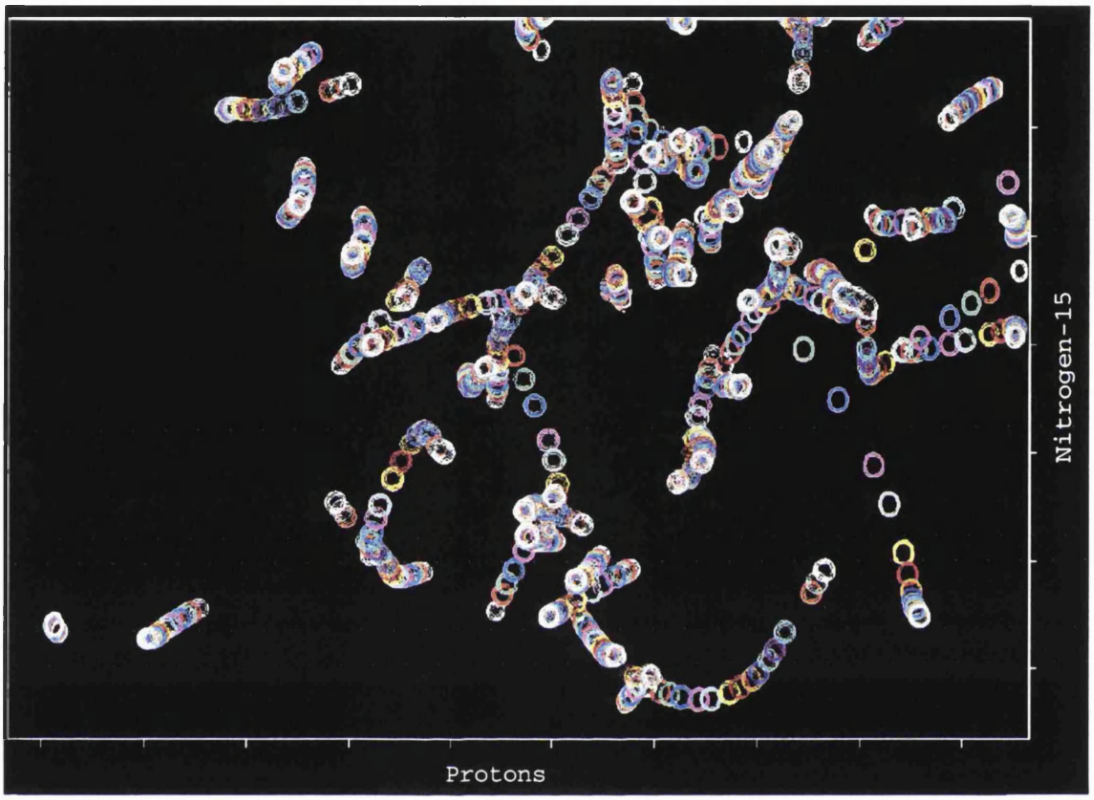


Fig. 5.1 *pH dependent chemical shift changes of the amide resonance from a ^{1}H-^{15}N HSQC experiment. Each pH titration point is represent by a different colour, and the change in chemical shift can therefore be trace as a function of pH and the pK_a determined.*

In contrast to the carboxyl carbon resonances, the backbone amide signals, are susceptible to multiple concurrent influences. The amide proton and nitrogen chemical shifts are very sensitive to changes in the electronic environment, a possible reflection of the fact that the amide N-H bond is more easily polarisable than the C-H bond (de Dios *et al.*, 1993). This allows the amide proton and nitrogen to sense the electrostatic influence of other ionisable groups which can operate over quite long range (Loewenthal *et al.*, 1993; Russell & Fersht, 1987), as well as the perturbation that may arise from conformational changes. The dominant influences on the chemical shift need not necessarily be that of the nearest charged group, and the pK_a determined from backbone amide chemical shift subjected to multiple influences may not correspond to that of the ionisable group. Accurate determination of pK_a from the ^{15}N-^{1}H correlation experiment is consequently difficult and identification of

influences of specific ionisable group haphazard. Nevertheless, with a thorough and rigorous analysis of their titration behaviours, useful information about the physical properties and interactions of the protein molecule may be obtained. By monitoring the individual pK_a of the charged groups and assessing the effect that these pK_as have on other nuclei, it is possible to evaluate the effect of pH on the structure of a protein such as unfolding (Tan *et al.*, 1995), conformation changes (Kohda *et al.*, 1991) and other physical interactions not readily available from the pH titration of the carboxyl carbon. The electrostatics in the protein may directly reflect the physico-chemical nature of the protein, analysis of the ionisation the pK_a can therefore also reveal much about the energetics and biophysical processes in the protein.

Experimental procedures

Sample preparation

Isotopically-labelled recombinant rat CD2d1 (residue 1-99) was expressed and purified as described in the Chapter 2. Unless stated otherwise, all protein samples used for the pH titration were at a concentration of 1.2 mM in 20 mM potassium phosphate with 10% D_2O, and contained 0.5 mM EDTA and 0.1 mM PMSF. The protein samples used for the titration were initially at pH 7.5 and the pH lowered or raised by the appropriate addition of small aliquots of 0.2 mM HCl or NaOH. The pH was measured before and after each titration point and the average of the two values used. The pH showed little variation at most pH titration points (±0.05 pH unit). However, above pH 8.5 where the solution is not well buffered by phosphate, the pH sometimes differed significantly before and after the experiment (up to 0.4 pH unit difference) due to an undetermined slow process. As a result of this variability, titration points at pH higher than 8.5 were not used for pK_a determination. For experiments in D_2O solution, the protein was lyophilised and then resuspended in 99.96% D_2O and the pH adjusted with aliquots of NaOD. The pH of protein sample quoted in D_2O is not corrected for the deuterium effect on the glass electrode.

NMR Spectroscopy

The spectra were recorded on a 600 MHz Varian UnityPlus NMR spectrometer equipped with a 5 mm triple resonance probehead from Nalorac Cryogenics Corporation (Martinez, CA), and the data were processed using Felix version 2.3 (BIOSYM Inc. San Diego, CA) and nmrPipe (Delaglio *et al.*, 1995). Chemical shift for ^{1}H, ^{15}N and ^{13}C were referenced to TSP = 0 ppm, as proposed by Wishart *et al.* (1995). All NMR experiments were performed at a temperature of 25°C. Chemical shifts of backbone and side-chain amide resonances were followed during pH titration via two-dimensional (2D) ^{15}N-^{1}H heteronuclear single quantum coherence (HSQC) experiments (Bax *et al.*, 1990; Norwood *et al.*, 1990). The chemical shifts of carbon resonances were followed during pH titration via constant time 2D HSQC experiment (Vuister & Bax, 1992). For the carbonyl resonances a modified 2D H(C)CO experiment was used (Zhang & Gmeiner, 1996). However, due to severe line broadening, in particular of Glu29 and Glu41 signals, a non-constant time version of the experiment was adopted for the titration. The ^{15}N-^{1}H HSQC experiments were recorded with 2048 real points in F2 and 256 complex points in F1 with spectral widths of 10000 Hz and 1700 Hz in F2 and F1 respectively. For the 2D H(C)CO experiments, the spectral widths were 8000 and 2000 Hz in F2 and F1 respectively, with 2048 real points in F2 and 60 complex points in F1. The number of transients recorded per t_1 increment ranged from 32 up to 400 to compensate for the extreme line broadening of some resonances which occurs in the range of pH 4.5 - 7.5. The 2D constant time HSQC were recorded with 2048 real points in the acquisition dimension (F2) and 80 complex points in the indirect dimension (F1). Spectral widths were 10000 and 4000 Hz in F2 and F1 respectively. The constant time delay was 13.6 ms and 16 scans were summed per F1 increment.

Curve-fitting Procedures

The pK_a values of the carboxyl groups were determined by following the NMR chemical shift changes as a function of pH. The titration data for carbon, nitrogen and proton chemical shifts (~1000 in all) were analysed using Mathematica (Wolfram Research) and Kaleidograph (Synergy Software). For data that shows a single titration event, the curve-fitting was performed using a non-linear least squares

method according to the equation derived from the Henderson-Hasselbalch (Edsel & Wyman, 1958):

$$\delta = \frac{\delta_{acid} + \delta_{base} 10^{pH-pK_a}}{1+10^{pH-pK_a}} \qquad [5.1]$$

where δ is the instantaneous chemical shift in ppm, δ_{acid} is the chemical shift of the protonated form and δ_{base} is the chemical shift of the deprotonated form. Equation 6.1 can also be expanded to fit 2 pK_as,

$$\delta = \frac{\delta_0 + \delta_1 10^{pH-pK_1} + \delta_2 10^{2pH-pK_1-pK_2}}{1+10^{pH-pK_1} + 10^{2pH-pK_1-pK_2}} \qquad [5.2]$$

or n number of pK_as with the general equation

$$\delta = \frac{\sum_{i=0}^{n} (\delta_i 10^{ipH-\sum_0^i pK_i})}{\sum_{i=0}^{n} (10^{ipH-\sum_0^i pK_i})} \qquad [5.3]$$

where δ_i are the chemical shifts at the beginning and at the end of the titration, as well as other transition point chemical shifts in-between the n number of pK_as. This general equation was based on the assumption that the protonation equilibria were non-interacting (Forman-Kay *et al.*, 1992), i.e. each titration curve was treated as the simple sum of all individual titrating events experienced by an atom. For titration data that indicated multiple protonation events and therefore did not yield a good fit using the simple relationship of a single titrating event, the number of pK_a values used to fit the curves was increased. In all cases, no more than five pK_a values were required to adequately fit the curves. In principle it is possible to pre-select the number of pK_as used to fit the curve based upon the number of charged residues present around the target atom in the three dimensional structure. However in this analysis an unbiased method for assessing the number of pK_as required to fit a titration curve by using the F-statistic (Kinney, 1997) was used. The chi-square (χ^2) values obtained from each fit were used to calculate F-values for each change in the

number of degrees of freedom associated with allowing additional titration events. The total number of degrees of freedom in the fit was assumed to be the number of titration data points (pH values), reduced by three degrees of freedom for each pK_a used in the fit (corresponding to one pK_a value and two end-point chemical shifts). The significance of the resulting F-value was then evaluated based on a normal distribution. A value of $p < 0.01$ was taken as indication that the F-value is significant and therefore the use of an increased number of pK_a values in the curve-fitting was considered justified.

Structural Analysis of CD2

The crystal structure of rat CD2 (PDB accession code 1HNG) (Jones *et al.*, 1992) was used as the template for the structural analysis. The solvent accessible surface area (ASA) of the residues in CD2 was calculated using the NACCESS program with a probe size of 1.4 Å (Hubbard *et al.*, 1991). The accessible surface area calculated is expressed as a percentage of a standard value which is the ASA calculated from a residue X in a tripeptide Ala-X-Ala in a standard extended state. The hydrogen bond network was analysed using HBPLUS (McDonald & Thornton, 1994). These and other tools used in the analysis of CD2d1 protein binding surface are available in Protein-Protein Interaction Server (Jones & Thornton) and SURFNET computer programs (Laskowski, 1995).

Theoretical Modelling

Molecular simulation of electrostatics for the prediction of the protonation equilibria of ionisable groups is widely used but with varying degree of success. For CD2d1, the electrostatic calculations were performed using the MEAD (microscopic electrostatics with atomic details) computer program (Bashford & Gerwert, 1992). Detailed description of the formalism for this method can be found elsewhere (Bashford & Gerwert, 1992; Tishmack *et al.*, 1997) and only a brief outline of the formalism used will be described.

This model treats the protein as a relatively low dielectric medium with embedded charges surrounded by a high dielectric solvent, and the boundary between dielectric regions is a complex one defined by the coordinates and radii of the atoms.

The electrostatic potential in and around molecule in solvent is assumed to be governed by the linearised Poisson-Boltzmann equation:

$$\nabla[\varepsilon(r)\nabla\phi(r)] - \kappa^2(r)\varepsilon(r)\phi(r) = -4\pi\rho(r) \qquad [5.4]$$

where $\phi(r)$ is the electrostatic potential at distance r from a point charge ρ, ε is the dielectric constant, κ is the Boltzmann term which accounts for ionic strength effects. This allows the dielectric geometries with the details of the protein structure and solvent accessibility that define a complex boundary between the interior low dielectric and surrounding high dielectric solvent to be determined. The charge-charge interactions, interactions with non-titrating charges and dipoles, and energetics of full or partial burial of titrating charges in the low dielectric environment can also be calculated.

In this model, the pK_a of an ionisable group in the protein can be treated as the result of electrostatic interaction between charged residues, each possessing an intrinsic pK_a (pK_{int}), i.e. the pK_a of an ionisable site in the protein if all other ionisable groups were assumed to be neutral. The pK_{int} of each ionisable group is first calculated using the following equation:

$$pK_{int} = pK_{mod} - (2.303k_BT)^{-1}(\Delta G_{Born} + \Delta G_{back}) \qquad [5.5]$$

where ΔG_{born} is the energy term that arise from the interaction of the titrating charges with the polarisation that these charges themselves induced, ΔG_{back} the interaction of the titrating charges with the background of non-titrating charges of the protein or model compound, and pK_{mod} the pK_a of model compounds. The calculated pK_{int}, together with a calculated matrix of site-site interactions, allowed the titration curve on each site to be determined in which the titration curve is a Boltzmann-weighted sum over the 2^N possible protonation sites. The pK_a is the mid-point of this titration curve. Unless stated otherwise, the analysis was based on the crystal structure of using the coordinates of only the first domain of CD2 (CD2d1) (Jones *et al.*, 1992). Coordinates for the mutants were generated using the MidasPlus molecular modelling software (Ferrin *et al.*, 1988).

Results

pH titration and curve-fitting of the H(C)CO data

H(C)CO data were recorded at a total of 31 titration points between pH 2.0 and 10.0 in order to adequately sample the titration behaviour of the acidic side-chains. Significant broadening of some of the cross-peaks (Glu41 and Glu29, Glu33) was observed at pH 4 - 7. All the peaks that showed severe broadening are from residues situated on the ligand binding surface of CD2. It is also noteworthy that the cross-peaks for these particular residues were relatively weak at all pH values, a possible reflection of a relatively rigid conformation of the side-chains, with local correlation times being dominated by the overall tumbling of the protein. More severe broadening occurred at and near the pK_a values of these residues, indicating chemical exchange events in the milli- to micro-second timescale and possible retarded kinetics of protonation of the acidic group linked to such chemical exchange events (McIntosh *et al.*, 1996; Plesniak *et al.*, 1996).

CD2d1 has five glutamic acid residues (Glu29, Glu33, Glu41, Glu56, Glu99) and eight aspartic acid residues (Asp2, Asp25, Asp26, Asp28, Asp62, Asp71, Asp72, Asp94) as well as the C-terminal backbone carboxylate group. The pH-dependence of the chemical shifts of the carboxyl ^{13}C resonances for all of these carboxyl groups is shown in Figures 5.2 and 5.3. Most of the carboxyl groups showed essentially monophasic titration behaviour. A few, however, exhibited minor inflexions on the titration curves and slight deviations from the monophasic model due to influences from nearby ionisable groups (Shrager *et al.*, 1972). For example, the data for C-terminal residue Glu99 can be curve-fitted to two pK_as which can be attributed to the influence of both its side chain and C-terminal backbone carboxylate groups. The most pronounced deviation from the monophasic model can be seen in the titration curves Glu29 and Glu41, both of which showed a distinct biphasic titration behaviour (Fig. 5.2). The proton chemical shift changes of E29 and E41 also revealed their close relationship, with both showing large movements in chemical shifts in the titration range of pH 5.5 and 8 (Fig. 5.5).

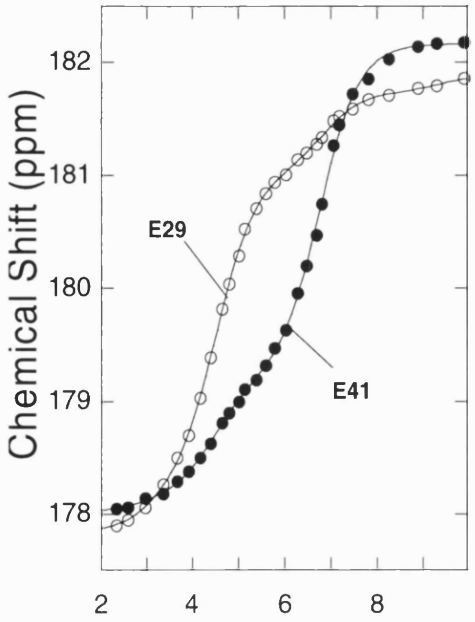


Fig. 5.2 The 13Cδ titration curves of E41 and E29 of CD2d1 showing the characteristic biphasic curve.

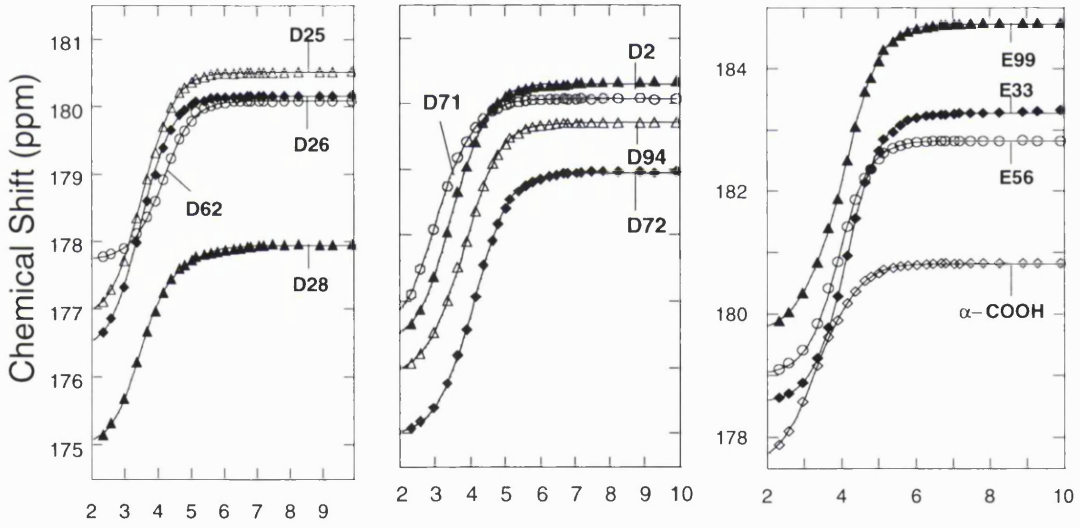


Fig. 5.3 The titration curves of ^{13}C carboxyl carbon of the carboxylate groups of rat CD2d1.

pK_a of the carboxyl groups

The pK_as of the carboxyl groups as determined by the carboxyl ^{13}C chemical shifts are tabulated in Table 5.1 and the location of some of the charged residues on CD2d1 are shown in Fig. 5.4. The majority of the carboxyl groups, with the notable exception of Glu41, have a pK_a that do not deviate substantially from that obtained from model compounds. For aspartates, the average observed pK_a value is *ca.* 3.7 which is comparable to the standard quoted value of 3.9-4.0 (Creighton, 1993). This

average value, however, is higher than the average value of 2.7 obtained from a survey of measured pK_a of aspartates in other proteins (Antosiewicz *et al.*, 1996). For glutamate residues, the average pK_a of 4.66 is skewed by the pK_a of Glu41 which is abnormally high (see below), but nevertheless is still close to the standard value of 4.3-4.4 (Creighton, 1993). The average pK_a for glutamates in CD2d1 excluding that of Glu41 is 4.14.

***Table 1**: pKa Values of Acidic Residues of CD2d1 Determined from HCCO Experiments*

Residue	pK_a (1.2 mM)[a]	Δδ (ppm)	pK_a (0.1 mM)[b]	H^{β}/H^{γ} [c]	Solvent accessibility ($Å^2$) [d]	
					Non-Polar	Polar
D2	3.55 ±0.04	3.77	3.48 ±0.02	3.46	42.53	79.24
D25	3.53 ±0.02	3.68	3.50 ±0.02	3.45, 3.61	40.05	78.26
D26	3.58 ±0.02	3.55	3.57 ±0.02	3.60, 3.55	28.22	66.28
D28	3.57 ±0.06	2.69	3.53 ±0.03	3.61, 3.96	4.34	25.11
D62	4.15 ±0.02	2.33	4.02 ±0.02	4.20, 4.19	0.00	12.61
D71	3.18 ±0.04	3.31	3.30 ±0.03	3.69, 3.01	37.66	82.82
D72	4.14 ±0.05	3.83	3.97 ±0.01	4.06, 3.82	0.00	15.78
D94	3.87 ±0.04	3.65	3.80 ±0.02	3.79	8.43	35.55
E29	4.42 ±0.04	3.03	4.39 ±0.05	4.66	0.57	23.90
	6.87 ±0.07	0.82	6.60 ±0.14			
E33	4.16 ±0.02	4.68	4.02 ±0.01	4.23	7.45	50.67
E41	6.73 ±0.05	3.00	6.36 ±0.02	6.88, 6.87	1.91	18.73
	4.34 ±0.06	1.10	4.19 ±0.07			
E56	3.92 ±0.01	3.80	3.74 ±0.02	3.99	28.11	40.63
E99	4.25 ±0.04	4.74	4.03 ±0.01	3.95	77.20	99.11 [e]
α-COOH	3.11 ±0.05	2.30	- [f]	3.49	-	-

[a] *pK_a values determined from the carboxyl carbon of the acidic residues at a protein concentration of 1.2 mM. If the curves showed more than one pK_a, the dominant one is shown first. The errors are derived from the fitting of titration curves.* [b] *pK_a values determined at a protein concentration of 0.1 mM.* [c] *pK_a values determined from the two intra-residue protons (H^{β}/H^{γ}) nearest to the carboxyl group. Only 1 pK_a given if the resonance is degenerate or if one of the pK_a cannot be determined. Only one pK_a reflecting the ionisation of its side-chain is shown.* [d] *Solvent accessible area (ASA) in square Angstroms is calculated using the program Naccess using a probe size of 1.4 Å. Cα is considered part of the non-polar aliphatic side-chain. The polar side-chain include all oxygens and nitrogens but not the carbons.* [e] *This value includes the C-terminus carboxylate.* [f] *Value cannot be determined due to extremely weak peaks.*

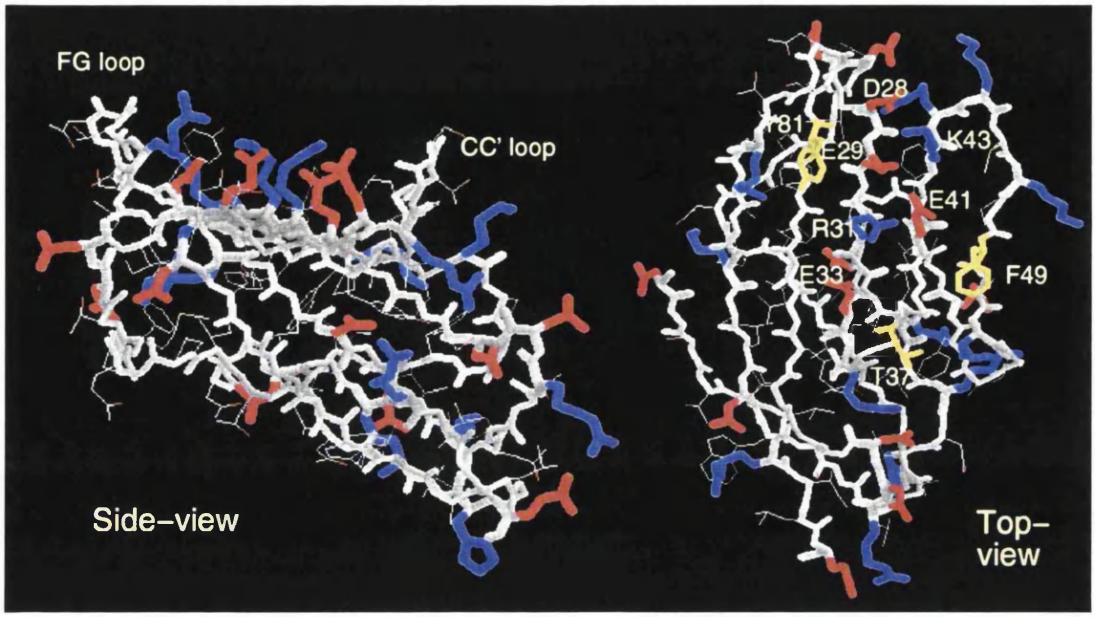


Fig. 5.4 *The charged groups on the surface of CD2d1. Large number of charges cluster on the binding surface. Note that the only histidine in CD2 is located on the opposite side to the binding surface (bottom in the side view of the molecule with the binding surface facing upwards).*

It should be noted here that the pK_a value of model compound generally quoted (Creighton, 1993; Nozaki & Tanford, 1967) is usually only applicable to denatured protein in the presence of high concentration of salt or denaturants (Tan *et al.*, 1995). It reflects the pK_a of a particular amino acid in aqueous solution, while measured pK_a in protein reflects the added change (ΔpKa) caused by moving from an aqueous medium into folded protein (Tozawa *et al.*, 1996), and this deviation from the pK_a value of model compound reflects the specific environment of the ionisation group. The lower average experimental pK_a value for aspartates in protein therefore suggests that the deprotonated form of aspartate is favoured in a folded protein (Antosiewicz *et al.*, 1996). As noted earlier, the average pK_a of aspartates in CD2 does not show significant deviation from the model compound, however, the average pK_a excluding those of Asp62 and Asp72 which are well-buried (therefore should be elevated) is 3.55. Although the value is still somewhat higher than the quoted average experimental values of 2.7 (Antosiewicz *et al.*, 1996), the result nevertheless lend some credibility to the suggestion that the environment of protein stabilises to some extent the ionised form of aspartates. It should be noted, however, the pI of CD2 as determined from isoelectric focusing is ~8.2, this lowering of the pK_a may therefore also be a result of the influences from the net positive charge on the molecule.

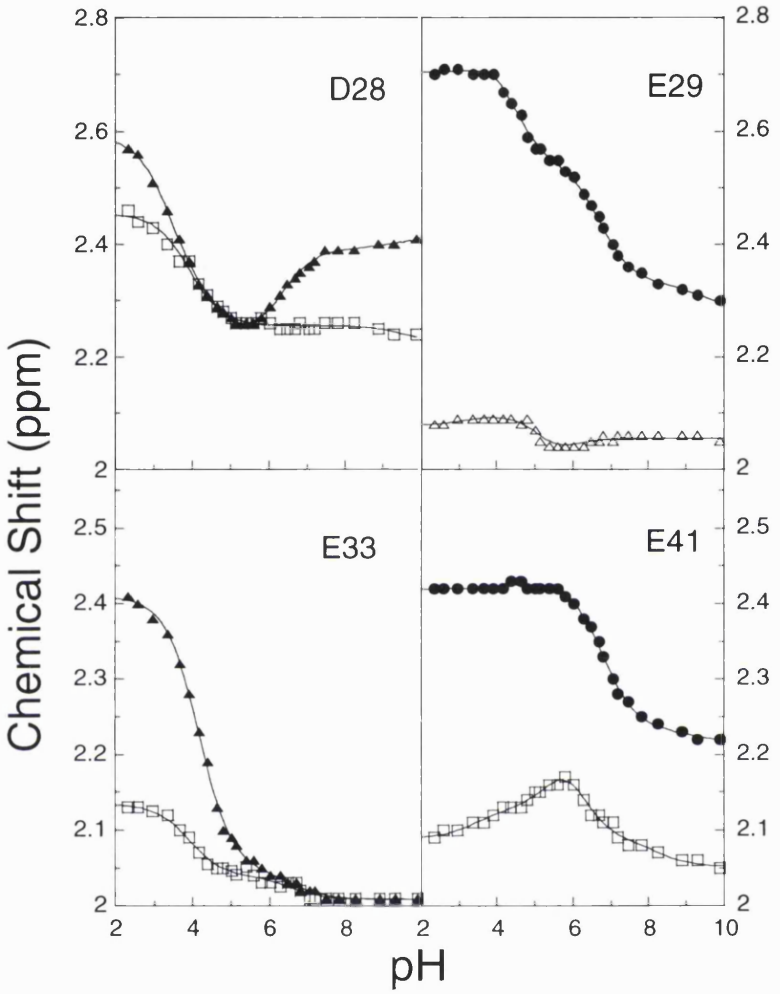


Fig. 5.5 *A selection of titration curves obtained from the resonances of aliphatic protons nearest to its intra-residue carboxyl groups, i.e. for aspartate the Hβ and for glutamates Hγ.*

The most notable feature of the pK_a measured is the abnormally high pK_a of Glu41. There also appears to be a reciprocal relationship between Glu41 and Glu29 in the pK_as obtained as well as in the amplitudes of the chemical shift change of each of the two ionisation events detected. The unusually high pK_a of Glu41 is determined to be 6.73 ± 0.05 with an associated Δδ amplitude of 3.0 ppm. The data for Glu41 also indicated a more minor shift variation of 1.1 ppm corresponding to an ionisation with pK_a of 4.34 ± 0.06. The pK_a of Glu29 is determined to be 4.42 ± 0.04 associated with a major Δδ of 3.0 ppm and an additional minor Δδ shift of 0.8 ppm corresponding to a pK_a of 6.87 ± 0.07. This reciprocal titration pattern of Glu41 and

Glu29 is characteristic of two carboxyl groups coupled in their ionisations (McIntosh *et al.*, 1996; Oda *et al.*, 1994). An examination of the rat CD2 crystal structure reveals that the C^{δ} atoms of these two residues are separated by a distance of 5.58 Å. It is therefore likely that the two carboxyl groups interact strongly in the manner of a dicarboxylic acid (Perrin & Thoburn, 1992).

The pK_a of the ionisable groups can also be determined from chemical shift variations of their immediately adjacent intra-residue protons. Most of the values correspond to those obtained from the carboxyl groups (agreement to within 0.1 pH unit), with the exceptions of a few that lie in close proximity to other carboxyl groups such as Glu29, Glu41 and Glu99 (Table 5.1). The protons are susceptible to multiple influences that give rise to more complex titration curves. As can be seen from the pH-dependent chemical shift changes of Glu29 $C^{\gamma}H$ (Fig. 5.4), a slightly greater influence is exerted by the ionisation of Glu41 ($\Delta\delta$ =0.21 ppm) than that from its own intra-residue carboxyl group ($\Delta\delta$ = 0.18 ppm). This is at variance with the suggestion that the chemical shifts showed greater influence from through-bond effect of the ionisation of its side chain ionisation than through-space effect of nearby ionisable residues (Forsyth *et al.*, 1998). Caution should therefore be exercised when determining the pK_a from the protons of closely interacting groups.

Deuterium isotope effect

The H(C)CO titration result indicated that Glu41 should be mostly protonated and Glu29 deprotonated at pH ~5.5. The protonation state of the glutamate can be revealed by examining the deuterium isotope effect on the carboxyl carbon chemical shift. For a protonated carboxyl ^{13}C resonance, the chemical shift is shifted upfield when the solvent is changed from H_2O to D_2O. The analysis of these isotope shifts is complicated by the fact that the pH meter reading is lowered in D_2O, but at the same time acids are weaker in D_2O - one effect may therefore be compensated by the other. For studies in the acidic range, the pH reading can be taken as a good approximation for pD (Bundi & Wuthrich, 1979a). Nevertheless, it does not preclude the possibility of error in the pH reading and slight shift in derived pK_a values (Bundi & Wuthrich, 1979a; Led & Petersen, 1979). In the present case whereby the resonances of both Glu29 and Glu41 are very sensitive to the precise H^+ and D^+ concentration, a degree

of uncertainty enters the analysis of the deuterium isotope shift effect. The $C^{\gamma}H$ proton chemical shift, on the other hand, should be less affected by the deuterium isotope effect and it was therefore used a reference for monitoring the pH. The H(C)CO spectra of CD2 in 90% H_2O and 99% D_2O showed good alignment of $C^{\gamma}H$ proton chemical shift (Fig. 5.6) at pH 5.5. In the D_2O spectrum, an upfield shift of 0.2 ppm was observed for the Glu41 carboxyl ^{13}C resonance, in line with values obtained in similar studies (Joshi *et al.*, 1997; Ladner *et al.*, 1975; Wang *et al.*, 1996; Yamazaki *et al.*, 1994). The Glu29 carboxyl, however, also showed a significant shift of 0.1 ppm, a possible reflection of the uncertainty of the pH values coupled with slight shift in pK_a, or it could suggest the oscillation of proton/deuteron between these two carboxyl groups (Perrin & Thoburn, 1992).

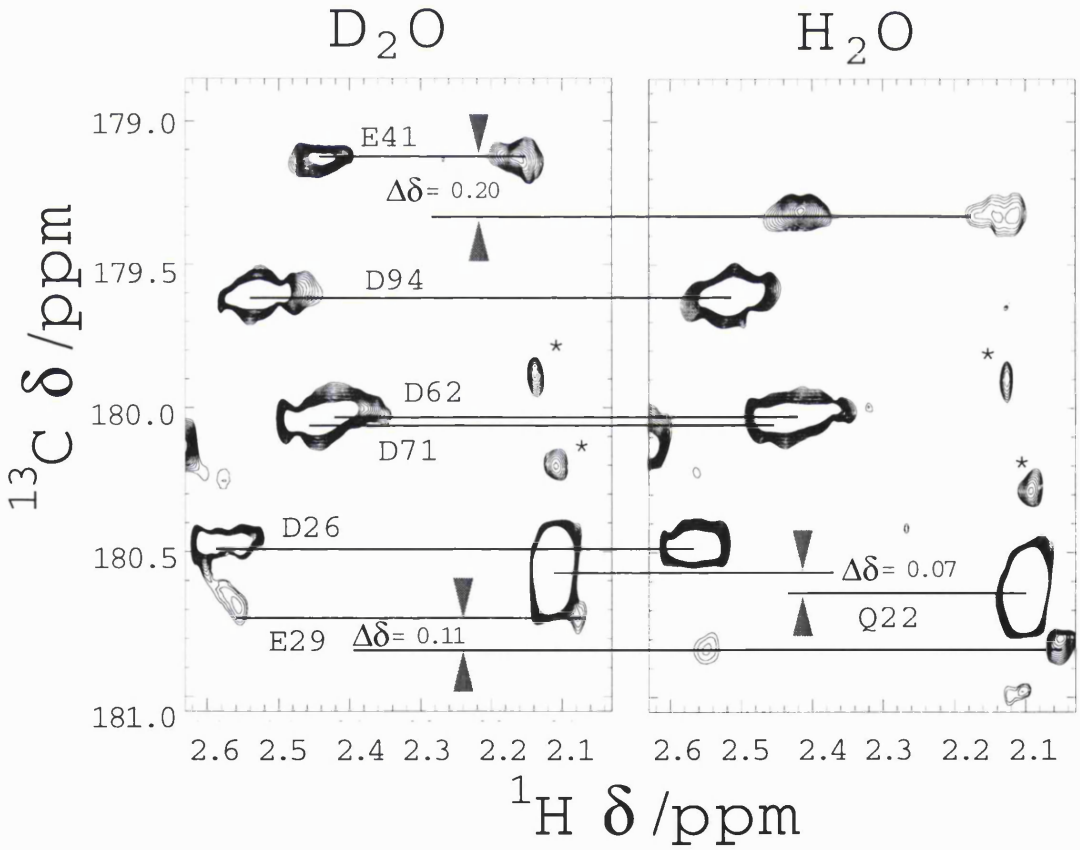


Fig. 5.6 *Comparison of HCCO spectra obtained in H_2O and D_2O demonstrating the shift of peaks due to the deuterium isotope effect. At the pH selected, Glu41 should be protonated while Glu29 deprotonated. The resonances of Glu41 therefore would be shifted in D_2O compared with those in H_2O.*

pH titration monitored by ^{15}N-^{1}H HSQC

The first indication of the unusual pH dependence behaviour of an acidic residue in CD2 was initially observed from data collected from ^{15}N-^{1}H correlation spectra. As described earlier, the titration curves for ^{15}N resonances can exhibit complex titration patterns, and care must be taken in the analysis of these shift variations as the interpretation of the titration curves of the backbone amides is not straightforward. The dominant pK_a as reflected by the backbone amide ^{15}N or ^{1}H chemical shift variation need not necessarily belong to the intra-residue side-chain. For example, for rat CD2d1 the titration curves of Glu29 backbone amide showed the effect of ionisation of Asp28, Glu29 and Glu41 carboxyl groups, with Asp28 exerting the greatest influence in the chemical shift changes (Fig. 5.7). The titration curves of the amide resonances therefore cannot readily be used to probe the ionisation equilibria unless the pK_a can be unambiguously assigned to a particular residue. Nevertheless, having accurately established the pK_a of all the acidic groups via the carboxyl ^{13}C shifts, it is possible to correlate the pattern of titration behaviour of the backbone amides with that of the side-chain carboxyl groups and analyse the effect of the ionisable groups on the chemical shift changes of the backbone amide. The 2D ^{1}H-^{15}N HSQC experiment is simple, relatively rapid and economical method of studying the ionisation equilibria in a protein. We have therefore adopted this method to monitor the effect of salts, buffering conditions and mutations.

A selection of the CD2d1 pK_a values determined from chemical shift variation in the 2D ^{15}N-^{1}H HSQC titrations is tabulated in Table 5.2. As can be noted, most of the pK_as as determined from the backbone amide agree with those obtained from the carboxyl carbon to within 0.2 pH unit, however, a few may vary by as much as 0.5 pH units for those carboxylates in close proximity to other carboxylates.

The amide cross peaks of many residues were observed to shift significantly in the pH range of 5.5 - 7.5 yielding apparent pK_a values in the range of 6.3 - 7.1. Since electrostatic effects may be quite long-range, only those amide resonances which showed shifts of 0.05 ppm in the ^{1}H dimension and/or 0.5 ppm in the ^{15}N dimension are considered to be significant (Forman-Kay *et al.*, 1992; Khare *et al.*, 1997). The residues that exhibit evidence of a pK_a within this range are: Asp2, Gly8,

Table 5.2 *pK_a Values Determined from the ^{15}N HSQC Experiments of Selected Residues.*

Residues (atoms)[a]	Ionisable groups [b]	pK_a values [c]					
		No NaCl	100 mM NaCl	300 mM NaCl	Bis-tris	E41Q	E29Q
D2 (N)	D2	3.6	3.7.	3.5	3.5	3.3	3.2
	α-NH2	7.1	7.4	- [d]	7.00	6.9	7.2
H12 (N)	H12	6.8	6.9	6.9	6.7	6.5	6.8
	E99	4.3	4.1	4.1	4.4	4.3	3.6
I18 (NH)	D62	4.4	4.3	4.3	4.2	4.1	4.0
D25 (N)	D25	3.7	3.6	3.5	3.6	3.4	3.3
D26 (N)	D26	3.7	3.7	3.5	3.6	3.5	3.1
D28 (NH)	D28	3.4	3.4	3.3	3.3	3.5	3.2
	E41	7.1	6.9	6.9	7.0	-	5.0
E29 (NH)	E41	6.6	6.6	6.6	6.8	-	5.0
E29 (N)	E41	6.6	5.8	5.2	6.4	-	4.9
	D28	3.4	3.1	3.3	3.5	3.1	2.9
E33 (N)	E33	4.3	4.2	4.1	4.1	4.3	3.6
E41 NH	E41	6.9	6.8	6.7	6.8	-	5.1
	E29	4.9	4.9	4.6	4.8	4.6	-
K43 (N)	E41	6.7	6.6	6.8	6.7	-	5.1
	E29	4.9	4.8	4.7	4.5	4.4	-
S52 (N)	E41	6.8	6.7	6.7	6.7	-	-
E56 (NH)	E56	4.0	4.0	4.0	3.9	3.7	3.8
N60 (NH)	D62	4.4	4.3	4.0	4.2	4.1	4.0
D62 (NH)	D62	4.2	4.2	4.0	4.1	4.0	4.0
T69 (NH)	D72	4.3	4.3	4.3	4.1	4.4	3.6
R70 (N)	D71	3.1	3.3	2.9	3.4	3.0	3.1
D71 (NH)	D72	4.5	4.5	4.5	4.3	4.2	4.2
D72 (N)	D72	4.5	4.5	4.6	4.2	4.2	4.1
T79 (N)	E41	6.7	6.6	6.6	6.7	-	-
D94 (N)	D94	4.1	4.0	4.0	4.0	3.8	3.7
E99 (N)	E99	4.6	4.9	5.0	4.7	4.4	4.4
	α-COOH	3.3	3.5	3.3	3.4	3.1	3.1

[a] *Selected residues whose pK_a values are listed in the Table. The nucleus from which the pK_a is determined is in the bracket.* [b] *The ionisable groups identified to be responsible for the pK_as. If more than one pK_as are present, only those with changes >0.05 ppm for protons and >0.5 ppm for ^{15}N chemical shifts are listed.* [c] *pK_a values determined for wildtype CD2d1 at varying salt and buffer conditions, and for the E41Q mutant. All the experiments were done in 20 mM phosphate except for one in 10 mM Bis-tris. Effect of salt are monitored from no added salt to 300 mM added salt in phosphate buffer.* [d] *pK_a value not determined due to increased amide proton exchange at higher salt concentration.*

Leu10, His12, Gly13, Asp28, Glu29, Val30, Arg31, Trp32, Gly35, Thr37, Leu38, Val39, Ala40, Glu41, Phe42, Lys43, Arg44, Met46, Lys47, Phe49, Leu50, Ser52,

Gly53, Ala54, Phe55, Asn67, Leu68, Val78, Thr79, Val80, Ser82, Thr83 (Fig. 5.8). Apart from Gly8, Leu10, His12, Gly13, Asn67, and Leu68 which are located around His12 in the tertiary structure of CD2d1, and Asp2 which is close to the N-terminal amino group, most of the remainder of this group can be mapped to the F, C, C' and C" β-strands, their connecting loops and the C"D turn (Fig. 5.8). Particularly large effects are seen for Arg31, Thr37, Leu38, Glu41, Lys43, Arg44, Met46 and Ser52. Most of this subset lie near to Glu41 with the exception of Ser52 which is located at ~13 Å away at the β-turn between strands C" and D and the only member of this set which is not directly on the ligand-binding surface.

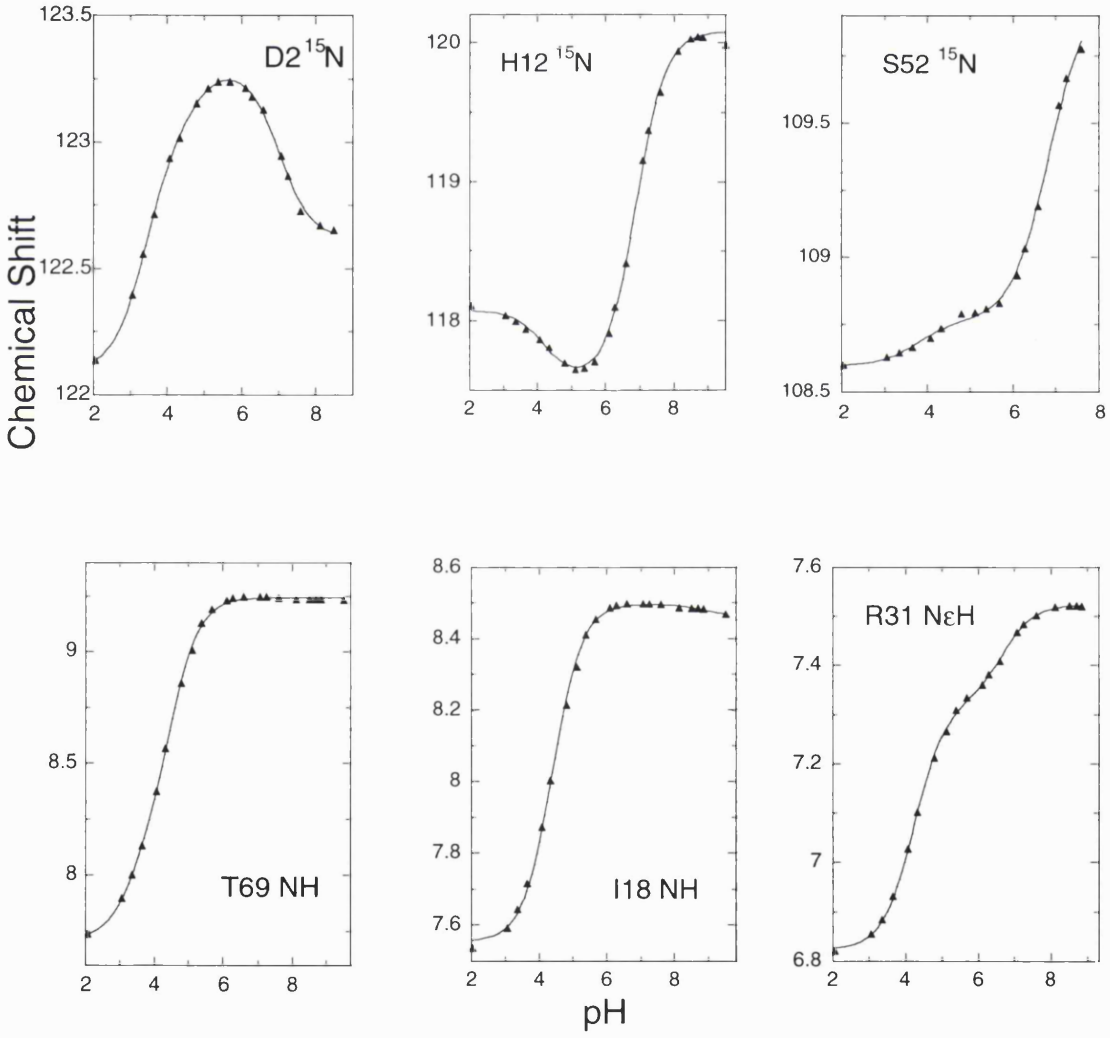


Fig. 5.7 A selection of titration curves obtained from the change in chemical shift of the backbone and sidechain amide resonances in ^{1}H-^{15}N HSQC experiment

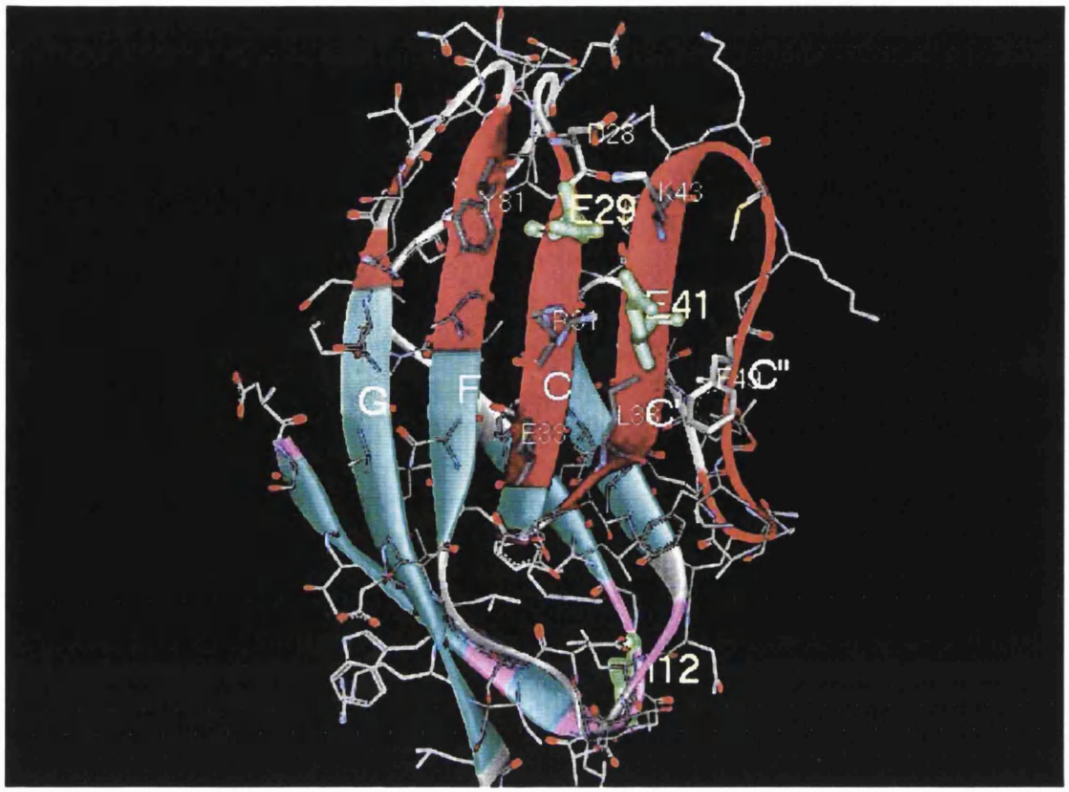


Fig. 5.8 *Residues on CD2d1 showing significant chemical shift changes in their backbone amide and/or side-chain aliphatic carbon resonances between pH 5.5 and 7.5. Residues coloured red showed the effect of Glu41 ionisation, residues coloured pink showed the effect of the ionisations of His12 and the N-terminal amino group.*

The pK_a of the only histidine present on CD2d1 (His12) was determined to be 6.8, within the normal range expected for histidines (Creighton, 1993). The pK_a of the N-terminal backbone amino group was determined to be 7.1 from the ^{15}N chemical shift variation of Asp2 backbone amide (Fig. 5.7). Both His12 and the N-terminus are situated on the ABED β-sheet face of CD2d1 (Fig.5.4). These ionisations are therefore unlikely to be responsible for the chemical shift changes observed for the large number of residues on the GFCCC'C" side of CD2d1, leaving Glu41 as the only likely candidate responsible for the observed effects.

Above pH 8.5, other titration events are observable for the ^{15}N and protons titration curves. This may be due to the titration of additional ionisable groups and/or other physico-chemical events linked to the titration of these groups. There are seven lysine residues, seven arginines and two tyrosines, but no cysteines. However, due to inconsistent pH measurement above pH 8.5, the pK_as of these residues were not determined in this study.

Apart from the aforementioned extensive influence of the abnormal pK_a for Glu41, the most striking pH dependent change in chemical shift is the very large downfield shift with increasing pH of the amide proton (>1ppm) observed for Thr69 and Ile18. Such significant downfield shift is normally associated with the formation of backbone amide hydrogen bond to a deprotonated side-chain carboxyl group (Bundi & Wuthrich, 1979b; Ebina & Wuthrich, 1984; Szyperski *et al.*, 1994). The pK_a determined from the amide protons of Thr69 and Ile18 should accord with that of their respective hydrogen bond acceptors, which on the basis of the CD2 crystal structure are Asp72 and Asp62 respectively. The magnitude of the chemical shift perturbation should also directly reflect the extent of population the hydrogen bond or the hydrogen bond length (Bundi & Wuthrich, 1979b; Wagner *et al.*, 1983). However, some small discrepancies in the pK_as determined from the amide protons and those from the carboxyl carbons are observed which may be a result of influences from other charged groups on the amide chemical shift. A smaller, but still significant, downfield shift (0.2-0.5 ppm) indicative of transient hydrogen bonding with nearby carboxylate groups can be observed for many other backbone amide cross peaks (Asn67, Gly4, Thr83, Arg44, Val39, Thr81, Phe42) as well as side-chain $^{1}H/^{15}N$ resonances (Arg31, Arg70, Arg90, Asn17, Asn60) with the proton shift of the Arg31 H^{ε} signal being of most interest. Its titration showed a biphasic curve corresponding to pK_a values of 6.68 ($\Delta\delta$=0.2 ppm) and 4.24 ($\Delta\delta$=0.5 ppm) which respectively reflects the ionisations of Glu41 and Glu29 (Fig. 5.7). Arg31 is situated next to Glu29 and Glu41. This result can therefore be interpreted as the formation of bifurcated or rapidly interchanging hydrogen bonds between Arg31 H^{ε} and both carboxyl groups of Glu29 and Glu41 (Szyperski *et al.*, 1994). Similar bifurcated hydrogen bonding interactions may also exist, as evidenced from the titration curves, for Arg44 H^{ε} to Asp25 and Asp28, and transient hydrogen bonding interaction can also be observed for Arg70 H^{ε} with the C-terminus residue Glu99 as the hydrogen bond acceptor.

Other interesting observations of the pH dependent behaviour of the backbone amide resonances include the pronounced exchange broadening behaviour for Leu38 and Val80 that occurred between pH 5 and 7. A few other peaks such as Thr37 and Thr86 also showed observable pH-dependent broadening (Fig.5.9). This

phenomenon suggests a chemical exchange event on a millisecond to microsecond time scale between pH 5 and 7 and this exchange process will be examined in greater details in Chapter 6.

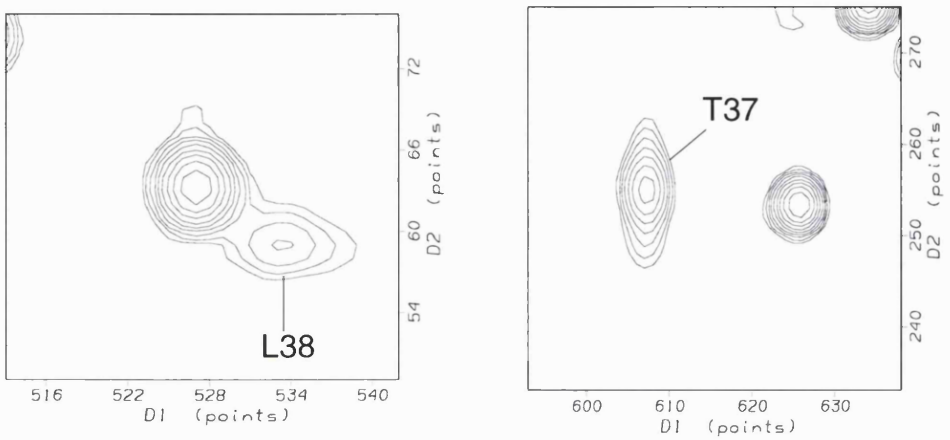


Fig. 5.9 Exchange broadened peaks of Thr37 and Leu38 at pH 5. L38 broadens in the ^{1}H dimension while T37 shows broadening in the ^{15}N dimension.

Effect of salt concentration and buffering condition

The pK_a values can be strongly influenced by the ionic strength as well as the buffering condition (Matthew *et al.*, 1985; Russell & Fersht, 1987) and larger perturbations of pK_a values are expected to be observed at low ionic strength (Abe *et al.*, 1995; Tan *et al.*, 1995). Increasing the ionic strength of the solvent results in the screening of charges which arise from the formation of Debye-Hückle atmosphere of oppositely-charged ions about a solvent-exposed charged group, thereby reducing charge-charge interaction. This screening effect is particularly effective for long range electrostatic interaction (Gilson & Honig, 1988), but less so for short-range charge-charge interaction. The screening should have the effect of shifting the pK_a towards that of its intrinsic value for residues involved in long-range interaction. The ^{15}N-^{1}H HSQC pH titration of CD2d1 was therefore repeated at higher salt concentrations (100mM and 300 mM) in order to ascertain that the abnormal pK_a is not a product of the specific buffering condition.

With increasing salt concentration, most proton resonances showed a systematic downfield shift of 0.02-0.04 ppm from no added NaCl to 300 mM NaCl, while the ^{15}N resonance showed a systematic shift of ~0.05 ppm with a few showing

greater changes in chemical shifts of up to 0.4 ppm (e.g. Lys45). With few exceptions, the titration patterns of all the amides are little changed at higher salt concentration, with only slight shift in the observed pK_as (Table 5.2). One of the largest shifts in pK_a is seen in the changes for the resonances of Asn60 which reports on the pK_a of Asp62. The pK_a dropped from 4.4 at low ionic strength to 4.0 at 300 mM NaCl (Fig.5.10). Interestingly, the pK_a is only slightly altered for Ile18 whose backbone amide directly forms a hydrogen bond with Asp62. This change in pK_a, as obtained from the chemical shift variation for the backbone amide of Asn60, therefore does not necessarily reflect a change in the protonation equilibrium of Asp62. Rather this effect is an indication that the chemical shifts of backbone amides are the product of multiple influences, and the contributions from longer range electrostatic effects are screened at higher salt concentration. This is most obviously seen for Glu29 which showed an apparent large drop in pK_a in part of its ^{15}N titration curve reflecting the ionisation of Glu41 (Fig 5.10). Such a large change is not seen in other titration curves, it is therefore most likely a result of the screening of the electrostatic effect of Glu41 on the backbone amide.

The presence of salt can weaken electrostatic repulsion between like charges (Tan *et al.*, 1995). The anomalous pK_a of Glu41, however, remained unaffected at the higher ionic strength conditions, suggesting that the short-range electrostatic interaction such as that which exists between Glu29 and Glu41 is insensitive to ionic strength. Furthermore the Glu41 and Glu29 have only limited solvent exposure, the effect of salts is therefore limited by the exclusion of water and the screening effect by salts is reduced (Jackson & Fersht, 1993). Nearly all the titration curves influenced by the ionisation of Glu41 are little changed, even those corresponding to nuclei some distance away where long-range electrostatic interaction may be expected to be effectively screened. For example, the magnitude of the chemical shift variation for the NH of Ser52 at ~13 Å away was little changed in the higher salt concentration, and showed only a slight drop of the pK_a reported of 0.1 units. This long-range effect of Glu41 is therefore unlikely to be electrostatic in origin.

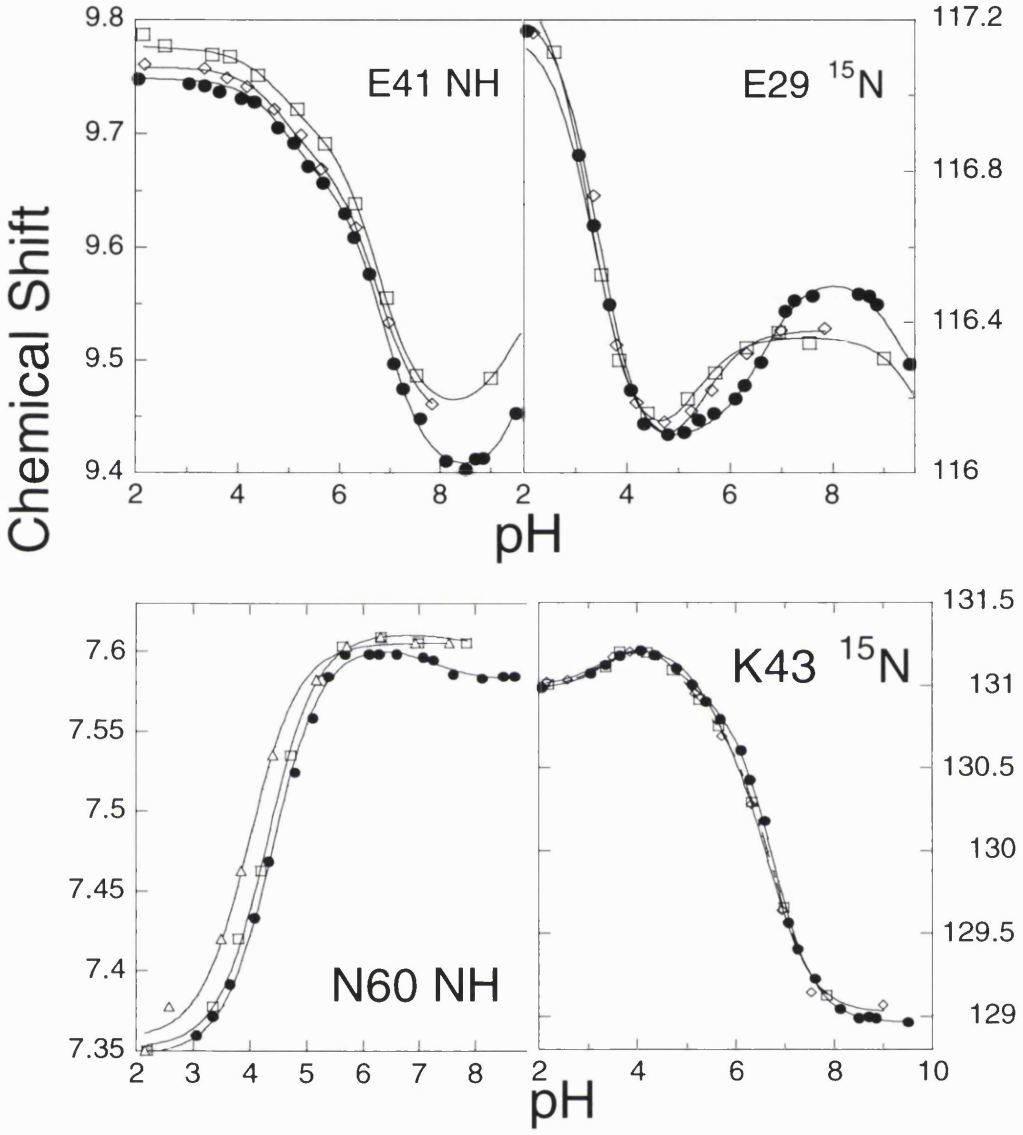


Fig. 5.10 *A selection of titration curves obtained from the ^{1}H-^{15}N HSQC experiment examining the effect of different salt concentration are used. Filled circle - no extra salt added, diamond - 100 mM NaCl added, open square - 300 mM NaCl added).*

The pH titration experiment was also repeated in Bis-tris buffer as the binding of counter-ion has been shown to influence the electrostatic interaction in a protein (Russell & Fersht, 1987). For example, in ribonuclease A, the pK_as of charged groups determined in water was shown to be different from those performed in phosphate buffer, a possible reflection of the binding of phosphate ion (Baker &

Kintanar, 1996). Likewise, the observed pK_a of His-146β in carbonmonoxyhaemoglobin is 7.85 in Bis-tris buffer but 7.1 in phosphate buffer (Matthew *et al.*, 1985). In CD2d1, however, the pK_as of most the acidic residue appeared to be largely unchanged with only slight shifts observed (Table 5.2). In particular, no significant difference is observed for the pK_a of Glu41 which remained essentially unperturbed at 6.7, indicating that the unusual acidity constant of this residue is not the result of specific ion binding effects or an artefact of the buffering condition.

pH titration curves from the aliphatic side-chains

The titration behaviour of the aliphatic side-chain resonances for wildtype CD2 was examined using ^{13}C-^{1}H correlation experiments. The titration curves of the aliphatic side-chain can showed similar complex titration behaviour, albeit to a lesser extent, to the ^{15}N titration curves as can seen in the titration curves of the C_β atoms of Asp28 and Leu38 (Fig. 5.11). In general, larger changes in chemical shifts were seen for ^{13}C resonances than for proton, most of the data used are therefore from the ^{13}C chemical shift which would give more precise curve-fitting. As was the case with the amide titration, a large number of aliphatic side-chains showed slight shift between pH 5.5 and 7.5 which may be due to the long-range nature of electrostatic effect. A cut-off of 0.15 ppm was therefore adopted as an indication of significance, and only those resonances that showed such significant shift within this pH range were examined in detailed. Residues whose aliphatic carbon showed significant changes in chemical shift are: Arg1, Gly11, His12, Asp28, Glu29, Arg31, Glu33, Arg34, Thr37, Leu38, Ala40, Glu41, Phe42, Lys43, Arg44, Lys45, Met46, Pro48, Leu50, Lys51, Ser52, Ile57, Thr79, Tyr81, Thr86, Ile88. This result closely parallel that obtained form the amide titrations, and most of these aliphatic side-chains can be mapped onto the GFCC'C'' β-sheet and the connecting loops which form the ligand-binding surface of CD2 (Fig. 5.6). A few of these shift effects can be ascribed to the influence of the protonation of His12 (Gly11, His12) and the N-terminal NH2 (Arg1), with the rest attributable to the titration of Glu41. However, these residues are some distance away from Glu41, thereby ruling out electrostatic effect as the basis of these changes. There is no simple relationship between the extent of

chemical shift and distance, for example Cβ of Ser 52 which is 15.3 Å away from Glu41 has a chemical shift change of 0.23 ppm, for Cβ of Asp28 at 9.67 Å away showed a shift of 0.13 ppm, while the Cβ of Lys45 showed a large change of 0.61 ppm at 12.07 Å away. Other atoms which showed very large shift are Cγ of Glu29 ($\Delta\delta$ = 0.68 ppm), Cβ of Met46 ($\Delta\delta$ = 0.57 ppm), Cβ of Thr37 ($\Delta\delta$ = 0.57 ppm). Some of the large changes may be a result of the ring current effect of nearby aromatic group Tyr81 and Phe49.

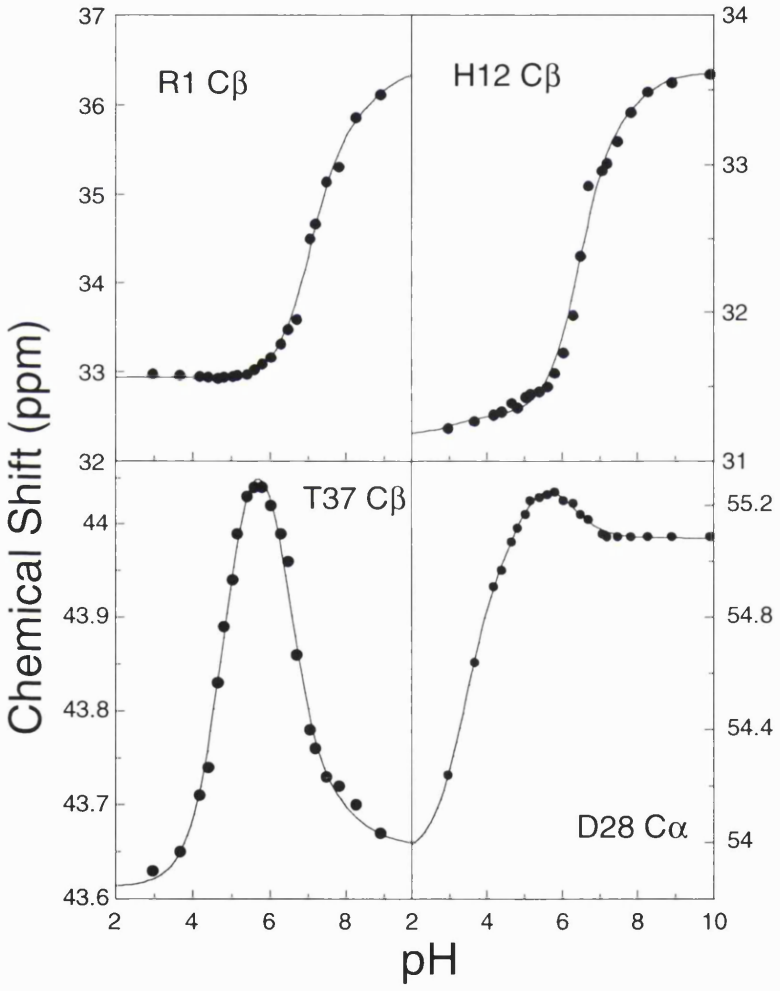


Fig. 5.11 *A selection of the titration curves obtained from the chemical shift changes of the aliphatic carbon in ^{1}H-^{13}C CT-HSQC experiments.*

In principle, a more accurate pK_a could be obtained from atoms nearest to the titrating group, i.e. Cβ for aspartates and Cγ for glutamates. This, however, appeared

to be true only for acidic groups not in close proximity with other carboxylates. For example, the pK_as obtained from the Cβ of Asp25, Asp26 and Asp71 showed significant deviations (up to 0.6 unit) from the pK_as determined from the carboxyl carbon, again demonstrating that determination of pK_a from aliphatic carbons, as were the case of aliphatic protons described earlier, requires some degree of caution.

Table 5.3 *pK_a values as determined from the aliphatic carbon chemical shift.*

Reporter[a]	Ionisable group[b]	Cα[c]	Cβ[c]	Cγ[c]
R1	α-NH2	6.72	7.04	-
	D2 (3.55)	3.60	-	-
D2	D2 (3.55)	3.24		
H12	H12	-	6.54	
D25	D25 (3.53)	3.74	3.88	-
D26	D26 (3.58)	3.52	3.98	-
D28	D28 (3.57)	3.70	3.77	-
	E41 (6.73)	6.70	7.04	-
E29	E41 (6.73)	6.95	6.96	6.67
	E29 (4.42)	4.67	4.44	4.40
E33	E33 (4.16)	4.86	4.21	-
E41	E41 (6.73)	6.72	6.74	6.84
	E29 (4.42)	3.66	4.27	4.88
E56	E56 (3.92)	-	3.93	4.04
D62	D62 (4.15)	-	4.15	-
D71	D71 (3.18)	3.85	3.78	-
D72	D72 (4.14)	4.04	-	-
D94	D94 (3.87)	3.93	3.89	-
E99	E99 (4.25)	3.79	3.92	4.07

[a] Selected residues whose pK_a values as determined from the aliphatic carbons are listed in the Table.
[b] The ionisable groups identified to be responsible for the pK_as with the pK_a as determined from the carboxyl carbon listed in brackets.. The pK_a values as determined the carboxyl carbons are listed in brackets. *[c] pK_a values determined from the aliphatic carbons. If more than one pK_as are present, only those considered to be of significance are listed. For those resonances that overlap resulting in ambiguity in their assignments, the pK_as are not determined.*

From the titration curves of the Cβ resonances, the pK_a of the His12 and N-terminal amino group was also determined to be 6.54 and 7.04 respectively. The pK_a of the N-terminal amino group (7.04) compares well with that obtained from ^{1}H-^{15}N HSQC (7.1), but for H12, the pK_a of 6.54 showed greater discrepancy (6.8 from the amide nitrogen chemical shift).

^{19}F NMR studies

Apart from the ^{13}C, ^{15}N and ^{1}H resonance normally used for NMR analysis of protein, fluorine (^{19}F) NMR can also be used to examine the biophysical properties of the protein. The aromatic residues in a protein can be labelled with ^{19}F by expressing the protein in media containing fluorinated analogue of the aromatic residue. In this set of experiment, 4-fluorophenylalanine is used and the resonances that arise from the ^{19}F-labelled phenylalanine residue were examined. Four peaks can be observed in the ^{19}F spectra of CD2d1, each correspond to one of the four phenylalanines residues (Phe21, Phe42, Phe49, Phe55). Phe49 was of particular interest as it lies on the binding surface and has been implicated as a residue that plays an important role in the binding interaction of CD2 (Davis *et al.*, 1998).

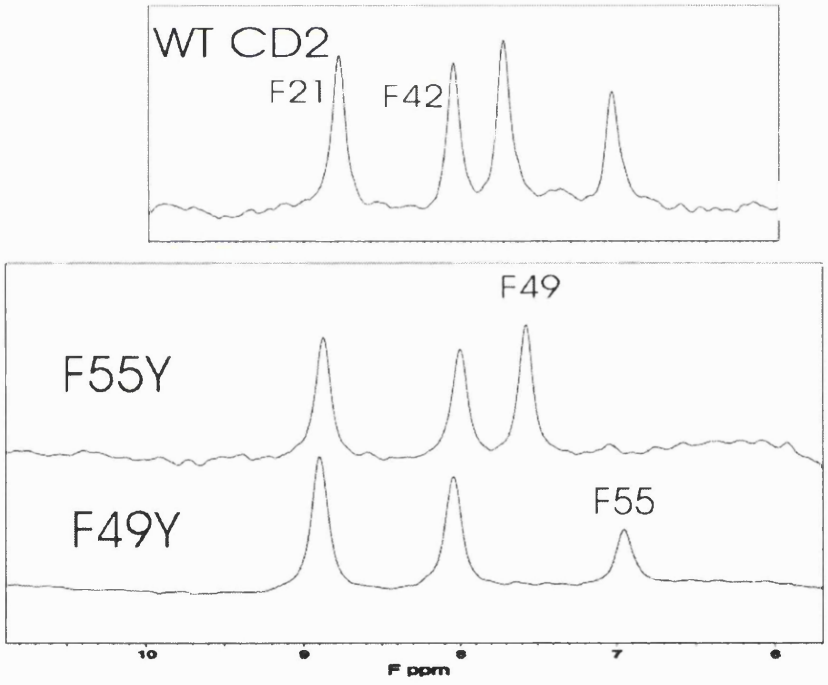


Fig. 5.12 *Identification of the ^{19}F resonances of CD2d1 using the mutant constructs. Not shown is the spectrum of F21L which suffered from severe rolling baseline as a result of the low protein concentration used that produced a very weak peak.*

In order to identify the ^{19}F resonances, tyrosine mutants of some of phenylalanine residues in CD2 residue were constructed, and mutant proteins were expressed labelled with fluorophenylalanine. Comparison of the spectra of the

mutants with that of the wild-type CD2d1 allows the resonances to be easily identified (Fig. 5.12). For the F42Y mutant, possibly as a result of introducing a hydroxyl group into the hydrophobic core thereby causing protein instability, heavy precipitation of the protein was observed. As a result, it was not possible to obtain a protein sample in sufficient concentration for NMR studies. A further mutant F21L was constructed with leucine in place of phenylalanine as leucine is considered to be a 'safe' substitution for phenylalanine (Bordo & Argos, 1991). Only a low concentration of this mutant protein was obtained, nevertheless, it is possible to identify the last two unassigned peaks.

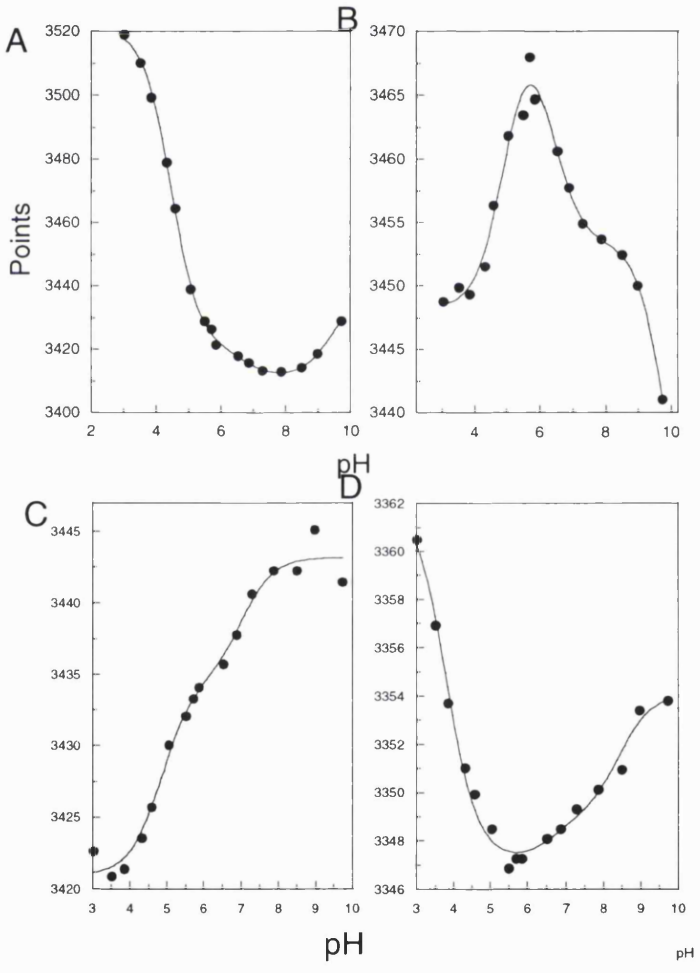


Fig. 5.13 *Titration curves of the ^{19}F-labelled protein. A - Phe55 B- Phe49 C - Phe42 D - Phe21.*

All the titration curves showed the effect of ionisation event with a pK_a value between 6-7 (Fig. 5.13). In the light of extensive effect of Glu41 on the chemical

shift changes of ^{1}H, ^{15}N and ^{13}C resonances, this is not surprising as three of the phenylalanines (Phe42, Phe49, Phe55 which showed a pK_a of 6.8, 6.4 and 6.8 respectively) lie at or near the regions that exhibited strong influence from the ionisation of Glu41. Phe21 which lies some distance away also showed a small influece of a titrating group with pK_a in such range. It is possible that this residue is reporting on the ionsation of the N-terminal amino group. These ^{19}F resonances also report on the ionisation of other residues, for Phe55, a particular large pH-dependent chemical shift change is seen at low pH with a pK_a determined to be 4.4. This is likely to be the pK_a of Asp72 (pK_a = 4.14 from carboxyl carbon) which is situated closeby. Also observed in the ^{19}F spectra is the pH dependent exchange broadening behaviour of Phe59 which will be described later in Chapter 6.

pK_a determination at low protein concentration

The results obtained from the pH-dependent ^{1}H ^{13}C, ^{15}N chemical shift changes suggested possible self-association of the protein. In order to determine whether the anomalous pK_a is an artefact of protein-protein interaction, the 2D ^{1}H-^{13}C H(C)CO pH titration was repeated at a concentration of 0.1 mM. The pK_a have been shown to be significantly perturbed on the binding of ligand (Yu & Fesik, 1994) and upon protein-protein interaction (Kato *et al.*, 1993). In addition, pH titration at high protein concentration can also influence the pK_a by an increase in electrostatic screening (Kesvatera *et al.*, 1996). At lower protein concentration with less effect from screening, any interaction between charged groups should be stronger and perturbation of pK_a values due to electrostatic interaction more pronounced. Obtaining the pK_a values at low protein concentration can therefore provide additional information about the protein.

For CD2d1, some shifts in the pK_a values were observed at the lower concentration. For example, the pK_a of Asp28 increased by 0.10 pH unit while that of Glu56 dropped by 0.18 unit (Table 5.1). Interestingly, the largest change in pK_a is seen in Glu41, down from 6.74 to 6.36 (Fig. 5.14). This change is inconsistent with the expectation that the pK_a value should be shifted towards its intrinsic pK_a at a higher rather lower protein concentration as a result of greater electrostatic screening effect at higher protein concentration. Significantly too, the pH-dependent peak

broadening seen in the titration at higher protein concentration is largely absent. These observed concentration dependent phenomena implicate a pH-dependent protein-protein association process occurring at higher protein concentration.

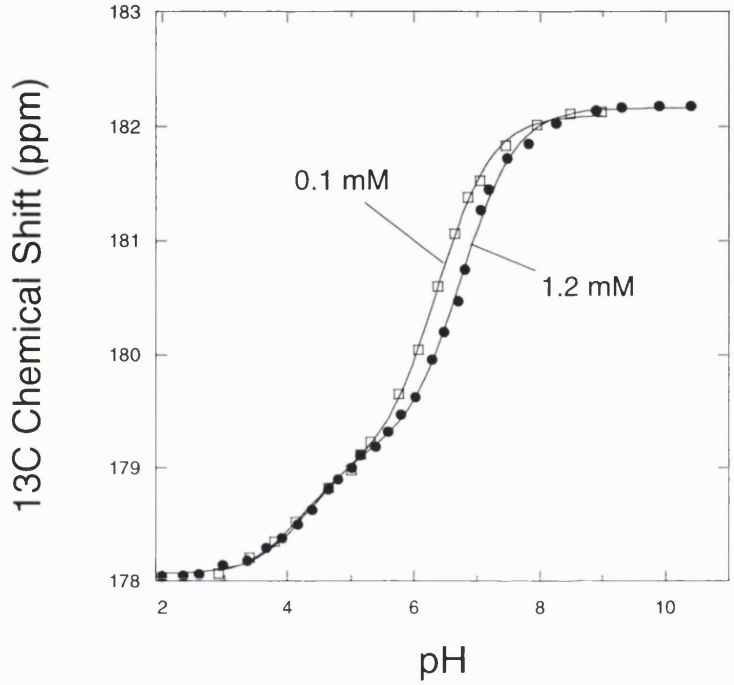


Fig. 5.14 *Titration curve of Glu41 obtained from HCCO experiments at protein concentration of 0.1 and 1.2 mM.*

Effects on pK_a upon CD2/CD48 interaction

The observation of the change in pK_a values at different protein concentrations prompted a further series of pH titrations using CD2d1 in the presence of CD48 to examine the effect on the ionisations of the carboxylates on CD2/CD48 interaction. A concentration of 0.2 mM of CD2d1 was used together with an approximate equimolar concentration of CD48 added. Many of the peaks of CD2d1 showed severe broadening as a result of interaction with CD48 at pH values between 4 and 8. The CD48 used in this study contain 2 domains and is glycosylated. The greatly increased apparent molecular weight of the CD2 when in complex with CD48 results in peak broadening. Even with greatly increase number of transients recorded, it was difficult to discern many of the peaks. With decreasing pH,

however, the peaks began to sharpen, indicating lower binding affinity at lower pH. This decrease in broadening starts at ~ pH 6.0 and below pH 4.0, no broadening could be observed, indicating complete loss of binding of CD2 for CD48.

The broadening of the peaks makes the tracing of the peaks with varying pH very difficult. For Glu41 and Lys43, extremely weak peaks may be observed and these were used to determine the pK_a of Glu41 (Fig.5.15). The result showed an apparent drop in pK_a to 6.3.

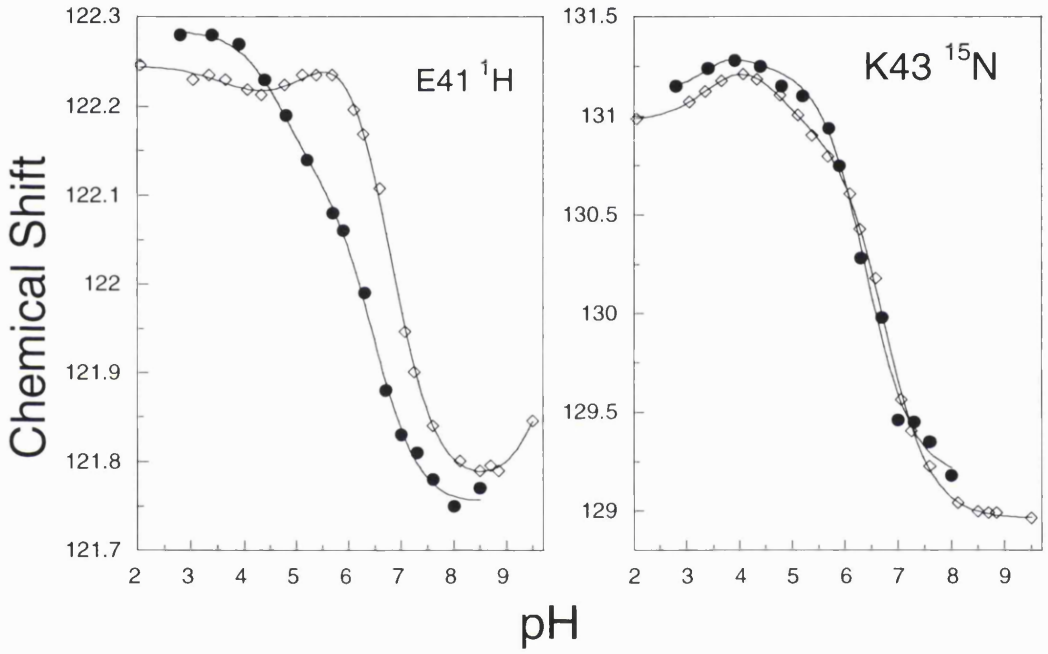


Fig. 5.15 *Titration curve of selected residues in CD2d1 on interaction with rat CD48. Open diamond – CD2d1 titration in the absence of CD48; filled circle – titration in the presence of CD48.*

Mutant studies

Mutagenesis can be used for a systematic investigation of the effect of charge-charge interaction in protein (Loewenthal *et al.*, 1993; Russell & Fersht, 1987). The basic strategy is to remove by mutation a charged group on the binding surface that may be involved in the interaction with other charged groups and/or participate in the binding interaction of CD2. The change in pK_a may then be directly related to the electrostatic interaction. As there is a dependence of the pK_a on the protein

concentration that suggest self-association, and the self-association states of the mutants are not known *a priori*, the pK_as obtained from these mutants will be compared to those obtained at low protein concentration in which any presumed self-association of the protein is minimal. Any decrease in the pK_a in the mutant is therefore the minimum that may be expected without any extra contribution that may arise from possible protein self-association.

A number of mutants were created as described earlier in Chapter 2, including most of the charged groups on the binding surface as well as a few hydrophobic residues thought to be important in CD2 binding interaction (Davis *et al.*, 1998). The residues of particular interest are Glu41 in view of its anomalous pK_a, and Glu29 by reason of its close relationship with Glu41. Both Glu41 and Glu29 were mutated to the glutamine (E41Q and E29Q), a change that removes the charge while maintaining the overall shape and bulk of the side chain. The chemical shifts did not show significant changes at low pH, and this can be taken as an indication that at low pH the protein fold and the disposition of the side chains are largely unchanged (Serrano *et al.*, 1990). The ^{15}N-^{1}H HSQC titration experiment were repeated for all the mutants. The titration curves for the majority of amide resonances for most mutants showed little change at low pH especially in the range of 2-4. For example, the formation of hydrogen bonding of the side-chain carboxylates with the backbone amides as indicated by the chemical shift changes of Ile18 and Thr69 are present in all the mutants, suggesting that the gross conformation of the backbone as well as the disposition of most of side-chains in all the mutants remain little changed. Most of the resonances of the mutants can therefore be identified and monitored.

For the analysis of the pK_a values of the mutants, a different approach for the determination of pK_a is taken. The value quoted for a particular residue is the average pK_a obtained from the titration curves of a number of nearby residues. Only residues within 7 Å of a particular charged residue with significant changes in chemical shifts ($\Delta\delta > 0.05$ for amide proton and $\Delta\delta > 0.5$ for amide ^{15}N shift) were used. A list of the pK_a values on the binding surface of CD2d1 obtained from the mutant is listed in Table 5.4. The changes in pK_a in the mutants relative to the wild type CD2d1 may indicate the extent of the contribution of the residue to the electrostatic interaction with Glu41. More pK_a values of E41Q and E29Q as determined from the individual amide resonance can also be found in Table 5.2.

Table 5.4 *pK_a values of the carboxyl groups on the binding surface of mutant CD2.*

Mutants	pK_a values			
	D28	E29	E33	E41
Wildtype	3.5	4.4	4.0	6.4
D28N	-	4.1	4.0	5.8
E29Q	3.2	-	3.6	5.0
E33Q	3.6	4.0	-	6.1
E41Q	3.5	4.5	4.3	-
R31A	3.6	4.4	4.4	6.9
K43A	4.1	4.8	4.4	6.8
L38A	3.4	4.6	3.9	5.8
F49A	3.4	4.7	3.9	5.9
Y81A	3.5	4.0	4.1	6.5

All the pK_a values in this table were obtained from the ^1H-^{15}N HSQC experiments. The values are the average values obtained from the titration curves of resonances that reflect the ionisation of particular carboxyl group. The pK_a values of the wild-type are ones determined at low protein concentration.

E29Q and E41Q

The titration curves of Glu29 and Glu41 indicate that these two residues are coupled in their titration. The removal of one of the charges should have considerable effect on the other. The most significant change observed in the titration of both the E29Q and E41Q mutants is the dramatic abolition of the large shift seen in the wild-type for many residues between pH 5.5 - 7.5 (Fig. 5.16). In particular, no significant pH-induced chemical shift changes can be detected in all the residues surrounding Gln41 or on the entire binding surface of the E41Q mutant in this pH range. In the E29Q mutant, there is a large drop in pK_a of Glu41 to 5.0, a fall of ~1.4 unit. This confirms the pivotal role played by Glu29 in stabilising the protonated state of Glu41. The absence of the extensive perturbation of chemical shift of residues on the binding surface of both the E29Q and E41Q mutants is particularly significant. It confirms that the extensive effect seen for large number of nuclei in wild-type CD2d1 is inextricably linked to the ionisation of Glu41.

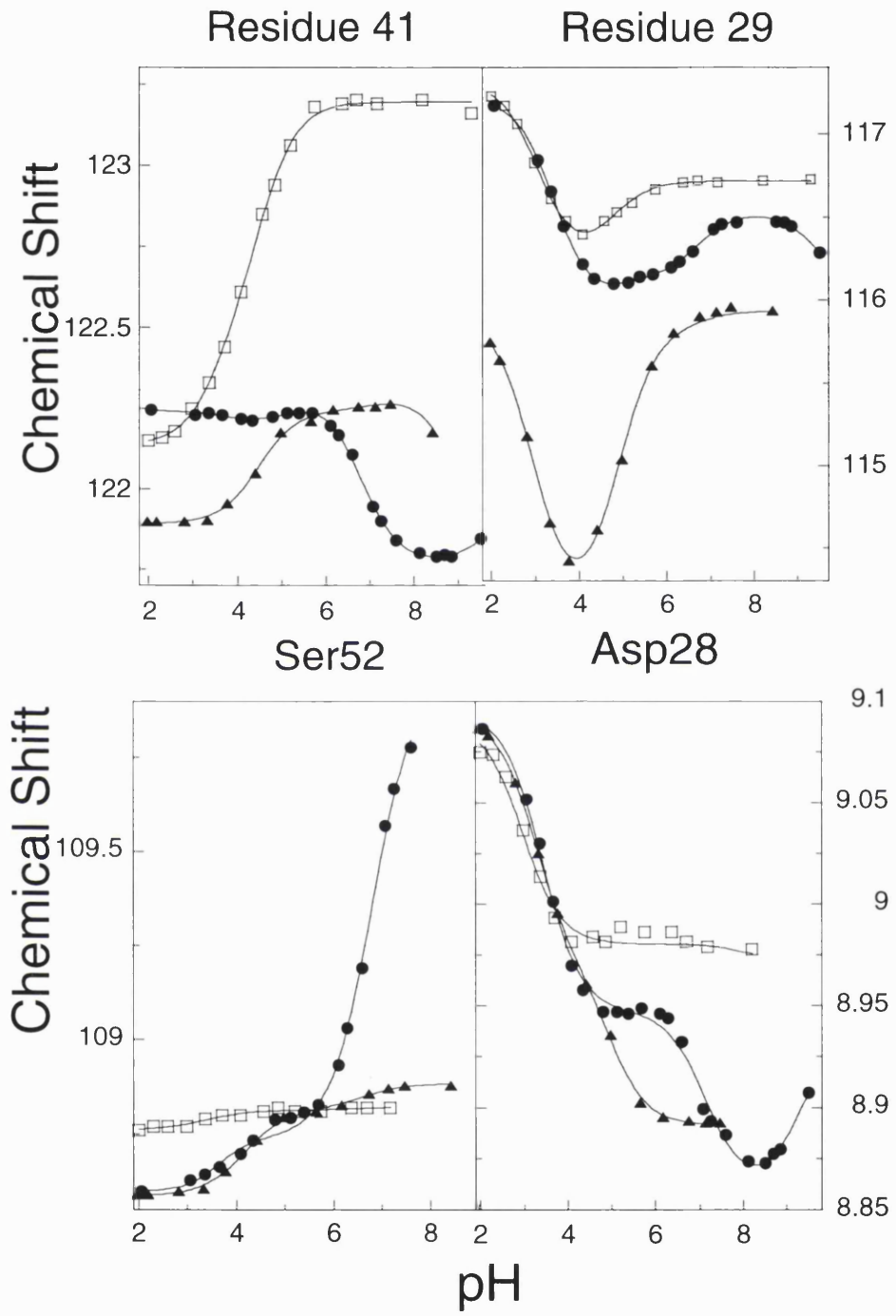


Fig. 5.16 *A selection of titration curves of E41Q and E29Q mutant. Filled circle- wild-type CD2d1, open square – E41Q, filled triangle – E29Q.*

Interestingly, the change in pK_a of Glu29 in E41Q mutant is surprising small, with only a small rise of 0.1 to 4.5 when a larger change may be expected. Given that in a coupled interaction, the mutual repulsion of the two carboxylates dictates that the increase in one should be accompanied by a concomitant and reciprocal

decrease in the other (McIntosh *et al.*, 1996), this small shift in pK_a of Glu29 suggests that the interaction between Glu41 and Glu29 involves other processes in addition to electrostatic interaction. A possibility is a conformation change of the side-chains of Glu41 and Glu29 as well as other surrounding residues in the wild-type CD2d1 associated with the changes in the protonation states of Glu41 and Glu29. Any positive interaction between a protonated Glu41 and deprotonated Glu29 should allow the side-chains of these two groups to move closer together, the shifts in their side-chains would simultaneously allows nearby residues greater spatial freedom for movement that changes the environment of Glu41 and Glu29.

This analysis assumes that the glutamine mutants do not showed significant difference in the dispositions of side-chains compared to the wild-type protein. However, a large downfield shift of Gln41 side-chain amide proton indicative of transient hydrogen bonding can be observed in the E41Q mutant. Curve-fitting of the chemical shift changes of the $N^{\varepsilon}H$ resonance yield a pK_a value of 4.31, a value that correspond better with that of Glu33, suggesting possible transient hydrogen bonding of Gln41 with Glu33 rather than Glu29. This can only occurs if the side-chain conformation of Gln41 in the E41Q mutant differs from that of the Glu41 in the wild-type protein. However, as described earlier, the margin of error of pK_a determined from the amide resonances can be substantial, it is therefore not possible to ascertain if any such conformational changes in the E41Q mutant has occurred.

E33Q and D28N

Two more carboxylates, E33 and Asp28, are located sufficiently close to Glu41 for them to exert some influence. The pK_as of the carboxylates were determined from D28N and E33Q mutant to assess the extent of their influence (Fig. 5.17). A particular large effect is seen from the D28N mutant with the pK_a of Glu41 dropping to 5.8 in D28N, a drop of ~0.6 pH unit (Table 5.4). A smaller drop is also seen for the E33Q mutant, with a decrease of ~0.3 for the Glu41. The result confirms the importance of the both carboxylates to the elevation of the anomalous pK_a of Glu41.

The greater contribution by Asp28 is intriguing. Asp28 lies slightly further away from Glu41 (a separation distance between the carboxyl carbons of 9.15 Å) then Glu33 (8.98 Å). Asp28 can be observed in the crystal structure to form a

hydrogen bond with Lys43 which is also hydrogen-bonded to Glu29. A lost of charge in this position for the D28N mutant may therefore result in some rearrangement of the electrostatic interactions of Lys43 to Glu29 and possibly Glu41, and the effect of the mutation may then be propagated to Glu41, either as a result of diminished interaction with Glu29 or greater interaction with Lys43.

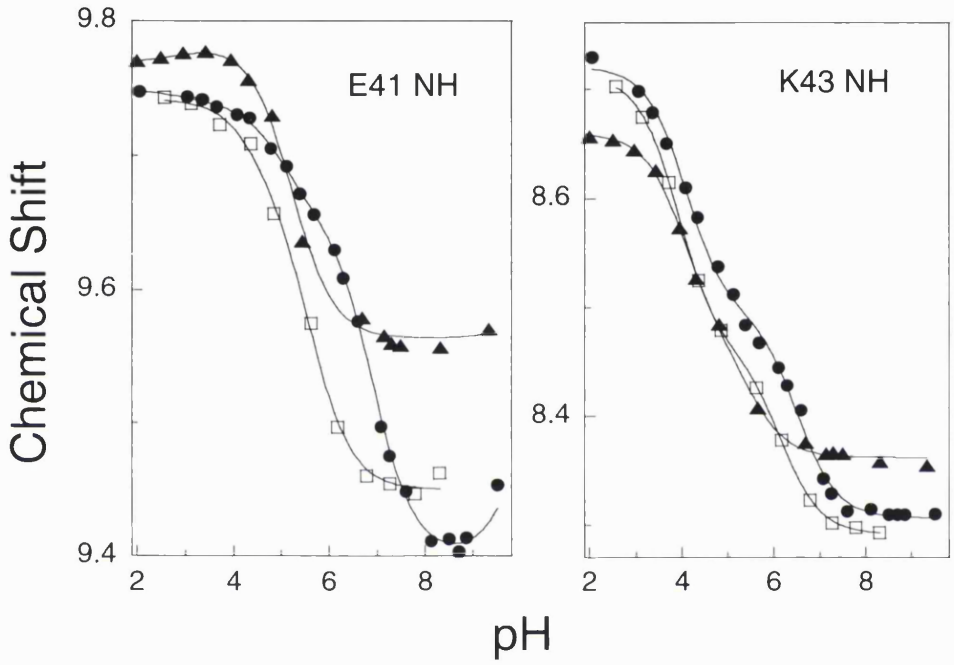


Fig. 5. 17 A selection of titration curves of E33Q and D28N mutants. Filled circle – wild-type CD2, open square – E33Q,, filled triangle – D28N.

The basic residues - R31A and K43A

Two basic residues are found near to Glu41 on the CD2 binding sites. The positively charged residues are expected to reduce the pK_a significantly. Arg31 in particular is situated very close to Glu41. However, no significant increase in the pK_a of Glu41 can be observed from the titrations of these two mutants. For the R31A mutant only a moderate increase of ~0.5 pK_a unit for Glu41 is observed. While the removal of Lys43 produces only an increase of ~0.4 for Glu41, its effect appears to be more extensive than Arg31. Lys43 lies in the middle of three acidic charges, Asp28, Glu29 and Glu41, with direct hydrogen-bonding observed between Asp28 and Glu29. Removing the charge on Lys43 also elevates the pK_a of Asp28 and

Glu29 by 0.6 and 0.4 respectively. While these changes are modest, cumulatively it suggests that the electrostatic influence of Lys43 is greater. Note that unlike the carboxylate mutants, this is maximum possible increase without taking into consideration any possible protein self-association effect.

A number of possibilities may explain these modest changes. In contrast to the carboxyamine mutants of the carboxylates, the alanine mutants of the basic residue can significantly alter the disposition of the side chain due to the lesser bulk of side chain of the alanine mutant. This may result in lower packing density and greater freedom of movement for other side chains. A small change in the structure can result in significant differences in electrostatic potential in the environment of charged groups. Interestingly, the titration curves of backbone amide resonances of Glu29 in the K43A mutant do not show the influence of Glu41 (Fig. 5.18), indicating possible alterations in the side-chain disposition. The complex interactions between the large number of charges on the binding surface also permit possible alternative electrostatic interactions with other groups. The small change in pK_a for Glu41 may be therefore be the result of a larger change in pK_a being offset by other changes.

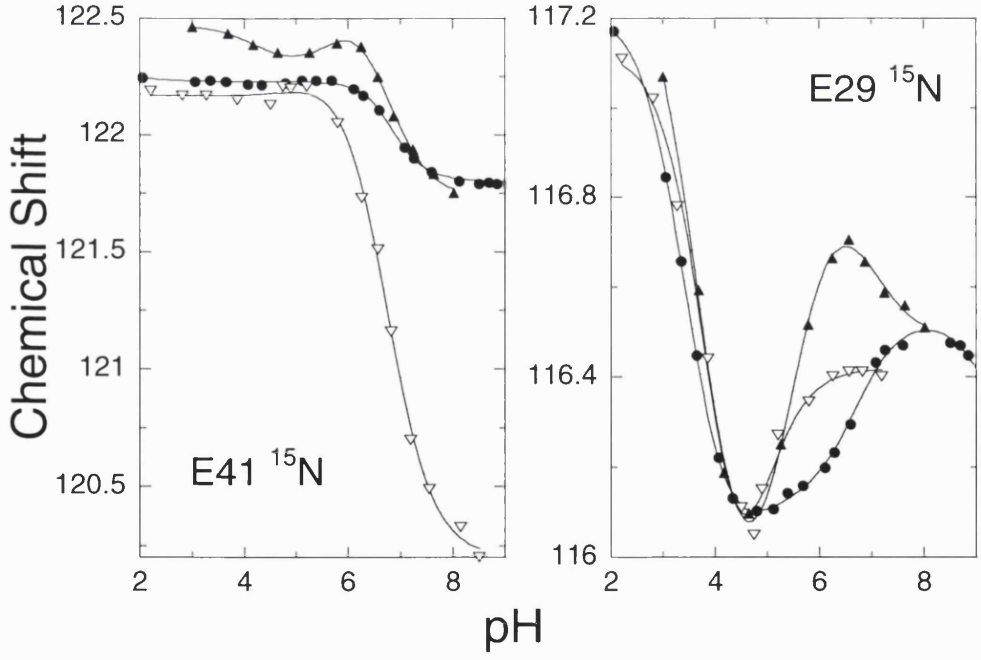


Fig. 5. 18 *A selection of titration curves of R31A and K43A mutants. Filled circle – wild-type CD2, open triangle – K43A, filled triangle – R31A.*

The hydrophobic residues- L38A, F49A and Y81A

A number of hydrophobic residues are thought to be important in the binding interaction of CD2. These residues, Leu38, Phe49 and Tyr81, are situated near to Glu41 and may play important role in the elevation of the pK_a of Glu41. Alanine mutants of each of these residues in CD2d1, i.e. L38A, F49A, Y81A, were constructed.

The side-chain of Glu29 is observed to form a hydrogen bond in the crystal structure with the hydroxyl group of Tyr81 that should stabilise a deprotonated Glu29. The hydroxyl may form strong hydrogen bond interaction with the Glu29 as hydrogen bonding between oxygen are known to among the strongest in protein, especially for those involving tyrosine hydroxyl groups (Khare *et al.*, 1997). In the Y81A mutant, the pK_a of Glu29 shows a small increase to 4.7.

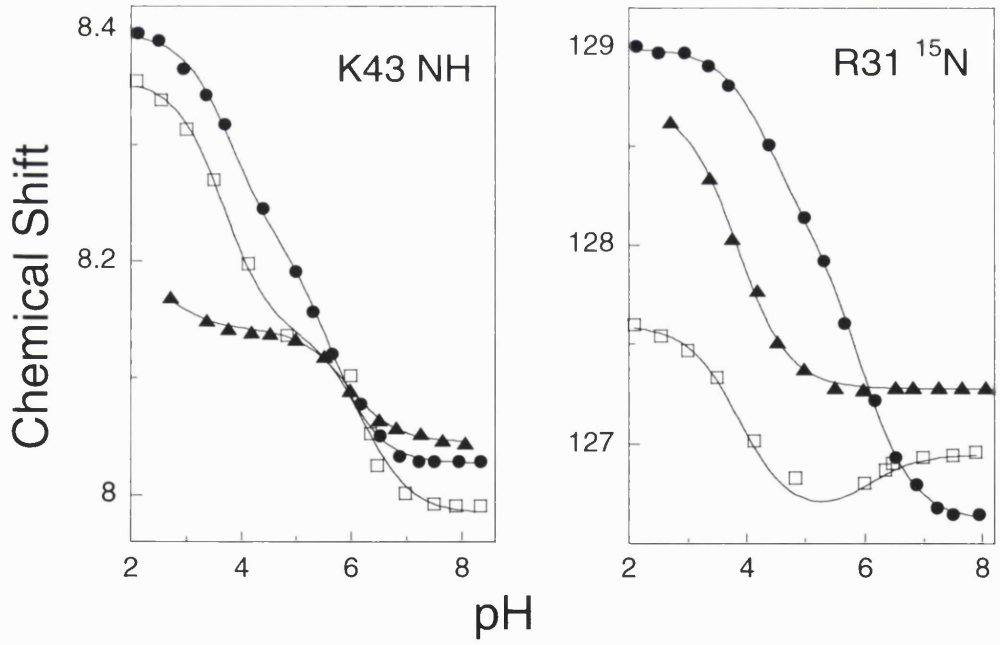


Fig. 5. 19 *A selection of titration curves of L38A, F49A and Y81A mutants. Filled circle – L38A, open square – Y81A; filled triangle - F49A.*

The mutations in both L38A and F49A resulted in a decrease in their respective pK_a of Glu41 to 5.8 and 5.9, a drop of 0.6 and 0.5 respectively. These hydrophobic residues therefore contribute to some extent to the elevation of pK_a in

Glu41. However, little change is seen for Glu41 in the Y81A mutant. This absence of significant change in the pK_a of Glu41 in Y81A is unexpected. The removal of a bulky groups on the binding site and near to Glu29/Glu41 appeared to have minimal effect when it is expected to disrupt the arrangement of the side-chains on the binding surface which should affect the binding capability and result in lowering of the pK_a. The lost of hydrophobic effect from the aromatic group may also be expected to result in lowering of the pK_a.

Electrostatic calculation

An alternative approach to the determination of pK_a is by theoretical modelling of the protonation equilibria using electrostatics calculations. Many approaches to the computation the pK_as have been devised (reviewed in Gilson, 1995; Juffer, 1998; Warshel & Papazyan, 1998). The various approaches to the improvement of calculation methods has led to a general decrease in deviation of the predicted pK_a from the experimental values. Greater agreement in the pK_a values, however, can be observed for aspartate than glutamate, while the pK_as of the basic residues tends to be over-estimated (Forsyth *et al.*, 1998). Comparison of the calculated pK_a values with that of the experimental can be found in a number of studies (Antosiewicz *et al.*, 1996; Dillet *et al.*, 1998; Forsyth *et al.*, 1998; Khare *et al.*, 1997; Tishmack *et al.*, 1997).

The main objective of the present study is to understand some of the factors that contributes to the unusual perturbation of the pK_a value of Glu41. Electrostatic modelling may yield some useful information on electrostatic interaction within the protein. In this present electrostatic calculation, both the crystal and NMR structures are tested. The crystal structure for CD2d1, from which coordinates for the mutant were also generated, is based of that of the CD2 ectodomain (Jones *et al.*, 1992). Two rotamers with different orientation of the NH_2 group of each glutamine mutant were used for the calculation and the average value used. For the analysis using the coordinates of the NMR structures (Driscoll *et al.*, 1991), a total of 18 NMR structures were used for the analysis and the calculated pK_a values from all the structures were then averaged.

Analysis based on crystal structure of CD2d1

A major disappointment in this analysis is the failure to predict the anomalous pK_a of Glu41, together with considerable overestimation (as can be seen for the pK_a of Asp62) as well as underestimation of the pK_a (as seen for Glu29 and Asp 72) of some of the residues (Table 5.5). The most glaring error being the predicted pK_a of Asp62 which at 7.57 is more than 3.5 units higher than the actual pK_a. The predicted pK_a values for Glu41 and Glu29 are 5.05 and 1.93 respectively, an underestimation by more than 2.5 units for both residues. Nevertheless, it is gratifying to note the elevation of pK_a of Glu41 relative to Glu29 with a difference of ~3 pK_a units between the two, a result which compares well with the 2.3 obtained from the experimental data. This further confirms the attribution of the abnormal pK_a to Glu41 rather than Glu29.

In order to understand the difference between the predicted and calculated value, it may be useful to examine the pK_{int} (Table 5.6). The pK_{int} measures the intrinsic pK of the charge group in a protein in the absence of other charges, the difference between pK_{int} and pK_a therefore gives a measure of the charge-charge interaction between residues. For Glu41, the predicted pK_{int} is 6.16, compared to the predicted pK_a of 5.05, a difference of ~1.1 that could be ascribed to charge-charge interactions. For Glu29, the predicted pK_a is 1.93, a drop of more than 4 units from pK_{int} of 6.06. Such a drop is directly reflected in the proportionate increase in pK_a (~4.6) from the pK_{int} of Lys43 to which Glu29 is observed to form a hydrogen bond in this crystal structure. Similarly increase can be seen for Arg31, a probable result of interaction with nearby Glu41. It can therefore be seen that in this analysis, the energetic contributions from the basic residues are grossly overestimated.

It is interesting to note the difference in the predicted pK_a for Asp62 an Asp72. Both Asp62 and Asp72 have little solvent accessibility and form hydrogen-bonds with the backbone amides. The predicted low pK_a of Asp72 may be due to presence of nearby Arg34, while the failure to take account of the hydrogen-bonding to the backbone amide may account for the excessively high predicted pK_a. In general, it is noted that those charged residues with greater solvent accessibility yield better predicted values, with the predicted pK_as within ~0.5 of the experimental values, while those that are buried yielded erroneous prediction.

Table 5.5 *The calculated theoretical pK_a values.*

	Expt [a]	Xtal[b]	NMR[c]	D28N[d]	E29Q[d]	E33Q[d]	R31A	K43A	F49A	Y81A
D2	3.55	3.08	-	2.99	2.98	3.19	3.075	3.08	3.08	3.11
D25	3.53	3.47	3.73	3.44	3.46	3.58	3.475	3.50	3.48	3.42
D26	3.58	4.21	4.58	3.95	4.01	3.55	4.217	4.20	4.20	3.95
D28	3.57	4.11	6.04	-	3.66	4.13	4.18	4.26	4.10	4.15
E29	4.42	1.93	4.97	1.89	-	2.13	2.80	5.88	1.97	2.40
E33	4.16	3.56	4.86	3.54	3.48	-	4.08	3.58	3.58	3.65
E41	6.73	5.05	5.11	5.14	2.90	5.64	6.08	4.22	4.97	4.89
E56	3.92	4.28	4.95	4.43	4.37	4.35	4.56	4.46	4.28	4.28
D62	4.15	7.57	7.68	7.48	7.34	7.44	7.58	7.50	7.55	7.70
D71	3.18	3.60	2.32	3.68	3.68	3.71	3.65	3.63	3.63	3.64
D72	4.14	2.50	5.85	2.57	2.54	2.19	2.61	2.50	2.48	2.36
D94	3.87	3.58	4.40	3.53	3.44	3.34	3.58	3.58	3.57	3.49
E99	4.25	3.81	4.44	3.92	3.80	3.88	3.79	3.80	3.83	3.88
αCOOH	3.11	3.43	3.41	3.51	3.50	3.51	3.48	3.51	3.48	3.48
R1	-	12.50	-	12.38	12.49	12.41	12.56	12.50	12.50	12.55
R31	-	13.88	14.95	13.84	13.21	12.74	-	13.83	13.83	14.00
R34	-	9.73	11.18	10.12	9.54	9.44	9.79	9.69	9.70	10.43
K43	-	12.89	11.77	10.91	10.64	12.35	13.00	-	12.90	12.58
R44	-	12.73	12.85	12.08	12.44	12.76	12.75	12.72	12.73	12.68
K45	-	10.83	10.21	10.48	10.89	10.80	10.87	11.18	10.85	10.78
K47	-	10.48	10.38	10.44	10.44	10.45	10.50	10.51	10.50	10.43
K51	-	10.46	9.43	10.40	10.34	10.40	10.55	10.48	10.60	10.27
K64	-	10.71	10.57	11.13	11.18	11.19	11.02	10.85	10.65	10.68
K66	-	11.00	10.63	10.85	10.85	10.68	11.12	11.08	10.97	10.91
R70	-	12.35	12.35	12.36	12.26	12.46	12.37	12.35	12.35	12.32
R87	-	12.73	12.77	12.73	12.59	12.78	12.83	12.75	12.71	12.60
K91	-	8.48	9.60	7.95	8.59	7.51	8.475	8.45	8.48	8.40
R96	-	12.57	12.29	12.48	12.54	12.37	12.59	12.61	12.57	12.55
Y76	-	17.78	16.01	17.78	17.82	17.82	17.85	17.63	17.73	17.78
Y81	-	9.83	10.65	9.62	8.80	7.70	10.25	10.05	9.83	10.95

[a] *The experimental pK_a values of carboxyl groups in CD2d1 as determined from the carboxyl carbon.* [b] *The pK_a values of all ionisable groups calculated from the X-ray crystallographic structure of CD2d1.* [c] *The values calculated from the NMR structures. The values cited are the averages of pK_a values determined from 18 individual conformer structures.* [d] *The values for the glutamine and asparagine mutants are averages values from two different conformers.*

Table 5.6 *Predicted pK_a with pK_{int} of selected residues*

	D28	E29	R31	E33	E41	K43	E56	Y81	K91
pK_{mod}	4.0	4.4	12.0	4.4	4.4	10.4	4.4	9.6	10.4
pK_{int}	5.83	6.06	9.99	4.70	6.16	8.28	6.16	10.79	8.08
pK_a	4.11	1.93	13.88	3.56	5.05	12.89	4.28	9.83	8.48

List of the values of pK_{mod} (pK_a of model compound), pK_{int} (intrinsic pK_a), and calculated pK_a of selected residues determined using the MEAD program.

It is also well know that the predicted pK_a values can differ by several pK_a units if different conformations of the protein are used (Bashford *et al.*, 1993). The assumption of structural rigidity implicit in the type of theoretical calculations used here can lead to exaggeration of the pK_a shifts because of the neglect of structural changes that may moderate the shifts (Bashford & Karplus, 1990). Conformational fluctuations in proteins may arise as a result of the variation in side-chain torsion angles of titrating amino acids as well as in relatively unrestrained regions of backbone such as in the loops and termini of the proteins. Prediction can therefore be more accurate if different conformers are taken into account in the pK_a prediction (Zhou & Vijayakumar, 1997). Furthermore, the pH can have a considerable influence on the structure of the protein and therefore may also needs to be taken into consideration.

Apart from the pK_a values of the carboxyl groups, these analyses also yielded the predicted pK_a of the arginines, lysines, and tyrosines. As described earlier, the pK_a of Arg31 and Lys43 were elevated due to interaction with the carboxylates. Other noteworthy predicted pK_a values include those for Lys91 and Arg34 (9.7 and 8.0 respectively) which are unusually depressed, a possible reflection of their relatively low solvent exposure. However, as noted earlier, the predicted values for the basic residues may be liable to error. For the tyrosines, Tyr76 is buried within the interior of the protein and the very high pK_a value predicted is therefore an indication that, to all intends and purposes, the residue is expected to be completely protonated at all realistic pH values. For the surface-exposed Tyr81, a pK_a of 9.83 was predicted, which is little changed from that of model compounds (~9.7) (Creighton, 1993). From the experiment data, a pK_a of ~9 is observed from the aliphatic as well as backbone amide of residue around the Tyr81, however, as noted earlier, no pK_a

value could be determined with any certainty due to variable pH reading of the protein samples at high pH values. Validation of the predicted pK_as for these residues must therefore be deferred until experimental conditions can be found in which the ionisation constants could be determined with greater confidence.

Predication based on NMR structures

As mentioned earlier, a more accurate result may be obtained if different conformations of the protein are taken into account, therefore pK_a predictions based on an ensemble of NMR structures or a single properly averaged structure may yield a better agreement with experimental values (van Vlijmen *et al.*, 1998, Zhou & Vijayakumar, 1997). A number of studies have indeed noted that calculations of pK_a values based on NMR structures can produce more accurate results than calculations based on single crystal structure (Antosiewicz *et al.*, 1996; Dillet *et al.*, 1998; Forsyth *et al.*, 1998). Such observations have been attributed to a more realistic sampling of possible conformations that can be obtained from NMR, and it is possible that the NMR structures can yield more accurate results simply because they are solution structures and pK_as are measured in solution (Antosiewicz *et al.*, 1996). In cases where the crystal structures produced better agreement with experimental pK_a values, for some surface residues, NMR structures can nevertheless produce better predictions; possibly because in the crystal structure, the side chain conformations of these residues are susceptible to distortion by crystal packing effects (Khare *et al.*, 1997). Furthermore, in the case of CD2, the crystal structure is that of a dimer, and the conformation of the surface residues may not reflect that of the monomeric CD2d1.

In this analysis using the solution structure of NMR, some of the pK_a values obtained showed better correspondence with the experimental values (e.g Glu29), while others are significantly worse (e.g. Asp28, Glu56). It is therefore not possible to determine if this NMR structure of CD2d1 is more representative of the actual structure in solution. It should be noted that the NMR structures of CD2d1 used are not of high resolution, and therefore likely to be susceptible to greater error. In this calculation using multiple NMR conformers, some of the structures produced noticeably better agreement with the experimental values, while a few yielded extremely poor predicted values. The correspondence or otherwise of the calculated

pK_a values suggests the possibility of refining the NMR structures by selecting for those conformers most consistent with the ionisation properties of the proteins (Dillet *et al.*, 1998). Those conformers that showed gross errors in the calculated pK_a values can be eliminated, while those that produce better correspondence can be used as model for further refinement of the structure. Such a method, however, may require a more accurate predictive algorithm than the one currently employed.

Predictions based on the mutants

As mentioned earlier, in the pK_a prediction, the electrostatic interactions from the basic residues are grossly overestimated. This overestimation is reflected in the analyses of the mutants. In the absence of electrostatic interactions between the basic residues and Glu29/Glu41, the pK_a of acidic residues should increase. This is shown by the increase in pK_a of Glu41 to 6.08 in R31A mutant and an increase to 5.88 for Glu29 in K43A mutant. Although the predicted values appeared to be better compared to the predictions based on the wild-type CD2d1, the pK_a of Glu41 in the R31A mutant is increased by ~1 unit, while the pK_a of Glu29 is increased by ~4 units in the K43A mutant. Such significant increase is not reflected in the results obtained experimentally. The calculated increases are comparable to those obtained from the analysis of pK_{int}-pK_a in the wild-type. The change in pK_a of Glu41 in the R31A mutant is relatively small, however, this is likely to be masked by a decreased interaction with Glu29, leading to a lower pK_a. The extent of interaction between Arg31 and Glu41 can be seen in the E29Q mutant in which the pK_{int}-pK_a of Glu41 is ~3, which roughly correspond to the pK_{int}-pK_a of ~ -4 observed for Arg31 in the wild-type CD2. For the E33Q and D28N mutants the model failed to predict the drop in pK_a of Glu41,

Little change was observed for the pK_as of Glu41 and Glu29 in the F49A mutant. For the Y81A mutant, slight increase in the pK_a of Glu29 was obtained. This is in variance with the experimental results in which the pK_a of Glu41 drops to 5.9 in the F49A mutant, and the pK_a of Glu29 drops to 3.9 in the Y81A mutant. For the mutants of these aliphatic residues, the predictions are therefore unsatisfactory, a probable reflection of the fact that significant changes to the disposition of the side-chains may have occurred in such mutants.

Discussion

Anomalous pK_a of Glu41

The analysis of the protonation equilibria for rat CD2d1 revealed the presence of a glutamate Glu41 with an unusually high pK_a and the unexpectedly extensive influence that this has on a large number of residues on one extended surface of the protein. The pK_a of the Glu41 is determined to be 6.72 ± 0.02, unusually elevated for a glutamate, particularly so for one that resides on the surface of a protein. The attribution of this pK_a to Glu41 is supported by studies using deuterium isotope effect that demonstrate Glu41 to be protonated at pH 5.5. Further confirmation comes from site-directed mutagenesis whereby mutation of Glu41 to glutamine completely abolished the backbone amide chemical shift changes previously seen for a large number of residues in wild type CD2 between pH 5.5 and 7.5, as well as from theoretical modelling which showed the pK_a of Glu41 to be elevated relative to Glu29. Independent confirmation of the attribution of the abnormal pK_a values is necessary as the interpretation of the data from pH titration for closely interacting charged residues is not straightforward and requires some careful consideration, as can be attested by the numerous studies on the controversial attribution of pK_a values in thioredoxin (Andersen *et al.*, 1997; Chivers *et al.*, 1997; Dillet *et al.*, 1998; Forman-Kay *et al.*, 1992; Jeng & Dyson, 1996; Li *et al.*, 1993; Qin *et al.*, 1996; Takahashi & Creighton, 1996; Wilson *et al.*, 1995).

Unusual pK_a values have often been observed for many residues types; these normally reside within an active site and are implicated in catalytic function (Fersht, 1985; Nakamura, 1996). They can also be found at ligand binding sites (Yu & Fesik, 1994) as well as at points of protein-protein contact (Andersen *et al.*, 1997). Proteins with anomalous pK_a values includes thioredoxin (Chivers *et al.*, 1997; Dillet *et al.*, 1998; Jeng & Dyson, 1996; Qin *et al.*, 1996), lysozyme (Inoue *et al.*, 1992), ribonuclease (Oda *et al.*, 1994), cyclophilin (Yu & Fesik, 1994), HIV protease (Wang *et al.*, 1996; Yamazaki *et al.*, 1994), turkey ovomucoid third domain (OMTKY3) (Schaller & Robertson, 1995, Forsyth *et al.*, 1998), bacteriorhodopsin (Metz *et al.*, 1992), F_1-ATPase β subunit (Tozawa *et al.*, 1996) and many others. Such unusual

pK_a values may be the product of a number of factors, such as charge-charge interaction between ionisable groups (Cederholm *et al.*, 1991), the effect of solvent and buffer conditions, and the desolvation penalty incurred in moving a charged residue from water upon its burial (Honig & Nicholls, 1995; Matthew, 1985). However, for CD2d1, no specific ion effect from the buffer used was observed, the deviation of pK_a of Glu41 from the intrinsic pK_a of a glutamate is therefore a consequence of its microscopic physico-chemical environment. Some of the factors that may affect the pK_a are examined below:

Charge-charge interaction between the acidic residues

The ligand-binding surface of CD2d1 is centred on a cluster of charged residues including Glu41. These charged groups in close proximity are expected to interact strongly with each other (Bashford & Gerwert, 1992; Yang *et al.*, 1993). Glu29 in particular appears to be coupled to Glu41 in its ionisation, as indicated by the biphasic nature and the reciprocal relationship of their titration curves (Fig. 5.2) (Shrager *et al.*, 1972). Similar coupled titration behaviours have been observed in ribonuclease H1 (Oda *et al.*, 1994), and in particular, *Bacillus circulans* xylanase (McIntosh *et al.*, 1996) in which the titration behaviour of two glutamates within the active site bear remarkable similarity to that of Glu41 and Glu29 in CD2d1. The presence of a nearby acidic residue has been shown to significantly affect the pK_a (Davoodi *et al.*, 1995; Tishmack *et al.*, 1997). An examination of the crystal structure of rat CD2 revealed that the carboxylate O^{ε} atoms of Glu41 and Glu29 are separated by a distance of 2.92 Å, close enough for a hydrogen bond to be formed between a protonated Glu41 and deprotonated Glu29 (Barlow & Thornton, 1983). The non-degeneracy of the γ-CH_2 protons of Glu41 and Glu29 suggests that the side chains of Glu41 and Glu29 have only limited mobility. Examination of the crystal structure, however, does not reveal close-packing of the side-chains, this limited mobility is therefore likely to be a result of the side-chains being held in position by complex electrostatic interactions between the many charged groups. Such electrostatic interactions may restrict the movement of the side chains and overcome the unfavourable interaction between the glutamates, thereby preventing the two glutamates from rotating away from each other due to electrostatic repulsion.

Two other carboxyl groups, Glu33 and Asp28, are also present close enough to Glu41 to contribute to its usually elevated pK_a as longer range (>5Å) electrostatic interaction can play important role in shifting the pK_a value (Forsyth *et al.*, 1998). The studies by Oda *et al* (Oda *et al.*, 1994) demonstrated that the biphasic titration behaviour can be the product of strong electrostatic interactions and cooperative ionisation of all four ionisable carboxyl groups.

The importance of these carboxylate groups has been demonstrated in the pH titration studies of the mutants. Site-directed mutagenesis can be used to modify a particular site on the protein surface and the shifts in pK_a of the charged residues in the mutants can provide a measure of the change in electrostatic potential (Thomas *et al.*, 1985). However, the analysis of the mutants is complicated considerably by the observation of the pK_a shifts for wild-type CD2d1 at different protein concentrations, a possible consequence of protein self-association. The pK_a shifts observed in the mutant are therefore compared with the pK_a determined at low protein concentration, a concentration at which the protein can be assumed to be essentially monomeric. The E29Q mutant showed a large drop in pK_a of ~1.4 for Glu41, demonstrating the central role played by Glu29 in raising the pK_a of Glu41. Glu33 and Asp28 were also shown to contribute to the anomalous pK_a of Glu41 in studies on E33Q and D28N mutants with Asp28 in particular showing significant influence. The results therefore showed that the unusually elevated pK_a is the product of multiple concurrent influences from a number of nearby acidic residues.

Electrostatic interactions with the basic residues

There are two basic groups (Arg31 and Lys43) located immediately adjacent to Glu41 that could have a strong influence on the protonation of Glu41. Any interaction between these groups, however, would be expected to lower rather than raise the pK_a (Baker & Kintanar, 1996; Dillet *et al.*, 1998). However, only a moderate increase in pK_a of Glu41 is observed in the alanine mutants of Arg31 and Lys43 in which the positive charge is removed. This contrasts strongly with the results from theoretical modelling which showed considerable influence from the basic residues. This may be a reflection of the conformation changes in the side-chains as a result of the removal of bulky side-chains of lysine and arginine in the K43A and R31A mutants. Alternatively, the electrostatic interactions may be less

intense than what might be expected from an observation of the crystal structure. Whatever the reason, it is intriguing that in the case of rat CD2d1, the effect of the proximity of the acidic residues around Glu41 dominates over the effects of the positively charged side chains. This interaction between the charged residues on the binding surface will be discussed further later.

Hydrophobic interaction

The presence of a nearby hydrophobic group has also been demonstrated to have considerable effect in raising the pK_a values of acidic residues (Inoue *et al.*, 1992). Nearby aromatic groups in particular has been shown to perturb significantly the pK_a of carboxylates (Urry *et al.*, 1994). It is therefore interesting to note the presence of Phe49, Tyr81 and Leu38 near to Glu41 on the protein surface. The aromatic residues are found to be invariant in all species variants of CD2 sequenced to date (Tavernor *et al.*, 1994) and may play an important role in the function of CD2, either in the binding interaction with its ligand or in maintaining the structural integrity of the backbone and surface side-chains.

In the analysis, greater effects on the pK_a of Glu41 are demonstrated by Phe49 and Leu38, with a drop of 0.6 and 0.5 in the pK_a of Glu41 in F49A and L38A mutants respectively. However, Tyr81 has little influence on the pK_a of Glu41. The result therefore showed that the two nearby hydrophobic residues, Phe49 and Leu38, have a strong influence on the raising of the pK_a of Glu41.

Solvent Accessibility

Burial of charged groups within the low dielectric medium of the protein core can also significantly affect the pK_a (Langsetmo *et al.*, 1991; Varadarajan *et al.*, 1989). The energetic penalty of removing a charged residue from the solvent can result in an increase of pK_a for an acidic residue and correlation between solvent accessibility and the pK_a of histidines had been observed (Yu & Fesik, 1994). Such correlation between solvent accessibility and the pK_a of acidic residues may also be observed in the data presented. For example, among the aspartates, Asp62 and Asp72 showed both the lowest degree of solvent accessibility and the highest pK_a values - 4.15 and 4.14 respectively (Table 1) compared with an average of 3.55 of the rest. These values are not unusually elevated but it is interesting to note that both

Asp62 and Asp72 also form hydrogen bonds with the backbone amides of Ile18 and Thr69 respectively, as indicated by the very large downfield shift in the backbone amide chemical shift. This hydrogen bonding should have the effect of lowering the pK_a by stabilising the deprotonated form, as had been observed for the hydrogen-bonded carboxyl groups of other proteins (Baker & Kintanar, 1996; Joshi *et al.*, 1997; Schaller & Robertson, 1995). An interesting link between the magnitude of the deviation in pK_a from that of the model compound and the extent of the chemical shift changes of the backbone amide (and thus the extent of hydrogen bonding) has been previously described (Schaller & Robertson, 1995). In CD2d1, the absence of such lowering of pK_a as might be expected from the extent of chemical shift changes and hydrogen bonding observed suggests that any such effect is more than compensated by factors such as burial that favour the raising of pK_a. The side-chains of Glu41 and Glu29 are situated on the surface of the protein but with relatively low solvent accessibility (Table 1); this may partly account for the abnormally high pK_a of Glu41.

The formation of the electrostatic interaction between the two glutamates may also be the result of the burial as previous studies have shown that ion-pairs form strong interaction when partially buried (States & Karplus, 1987). However, it does not explain why the pK_a of Glu41 rather than that of Glu29 that should be elevated. There is a slightly lower solvent exposure for the polar group of Glu41 (Table 5.1), however, the difference is slight, and as such it is not possible to say if the relative solvent exposure is the main determinant. Another possible reason is the presence of the electrostatic influence from both Lys43 and Tyr81 to which Glu29 can be seen to be hydrogen-bonded in the crystal structure. While the results from the theoretical modelling are less than satisfactory, it does provide useful information about the electrostatic interaction on the binding surface. It is clear from the analysis that ionisation of Glu29 is strongly influenced by Lys43, with a large pK_{int}-pK_a value in wildtype CD2d1 due to the presence of Lys43, and a significant increase in its pK_a in the K43A mutant. Glu41 also have a slightly higher pK_{int}, a probably reflection of its lower solvent exposure. It is therefore likely that this preference for deprotonation of Glu29 over Glu41 is a combination of the slightly greater solvent exposure of Glu29 and electrostatic interaction with Lys43. The apparently normal pK_a of Glu29 is

therefore the product of partial burial of this charged residues counterbalanced by its charge-charge interactions with Glu41 and Lys43.

Electrostatic interactions on the binding surface

As mentioned earlier, Glu41 and Glu29 can be observed to lie in the middle of an extensive hydrogen-bonding network with hydrogen bonds and salt bridge seen for Glu29 to both Lys43 and Tyr81. In view of the lack of significant influence on the experimental pK_a values of Glu41 and Glu29 from the basic residues, the question arises as to whether the hydrogen-bonding network as seen in the crystal structure actually exists in solution condition. Hydrogen bond interactions between side chain amides and side chain carboxylates are thought to be scarce on protein surface because they do not contribute significantly to the protein stability (Szyperski *et al.*, 1994), and the cost of localising a pair of solvent-exposed charged groups on the surface largely offsets the interaction energy expected from formation of a defined salt bridge (Dao-Pin *et al.*, 1991). It is also important to note that the crystal structure of CD2 is that of a 'head-to-head' dimeric form in which the charges are buried within the binding surface (Jones *et al.*, 1992). The electrostatic environment within the buried surface may promote the formation of salt bridges in order to compensate for the unfavourable desolvation penalty incurred by the burial of charged residues during protein-protein interaction. Whether such salt bridges actually exist for monomeric CD2d1 in solution condition therefore could not be taken for granted. In addition, the resonances of many side chains showed pronounced pH-dependence behaviours that suggest changes of the side-chain conformations. For example, the H^{β} of Asp28 are degenerate between pH 5 - 7 that is suggestive of free rotation of the side chain (Fig. 5.5). Unusual titration curves may also be seen for other carboxylates on the binding surface, suggesting the presence of other influences apart from the electrostatic effect of the ionisable group such as conformational changes of the side chains. This is in contrast to the H^{β}/H^{γ} of most other carboxylates not on the binding surface which exhibited fairly standard pH titration curves. The conformation and physical environment of the side chains on the binding surface may therefore varies according to the solvent condition. The evidence for the presence of electrostatic interactions therefore needs to be examined.

In the case of Lys43, there are some indications of electrostatic interaction with nearby groups. The chemical shifts of the ε-protons of Lys43 is not degenerate as might be expected if the C^{ε} atom is in free rotation, perhaps indicating some degree of constraint on its movements arising from its interactions with nearby charged groups. For Arg31, there is indication of transient hydrogen bonding with Glu41 as can be observed from the pH dependent chemical shift changes of its H^{ε} atom. As will also be seen in a later chapter, from the analysis of the NOE experiments, the $N^{\varepsilon}H$ group of Arg31 is protected from exchange of proton which can be interpreted as the formation of hydrogen bonds of $N^{\varepsilon}H$ with the carboxylates or burial. However, there is no direct evidence of strong electrostatic interaction between the ionised groups. Arg31 is situated near to Glu41 (~4Å) in the crystal structure, but the absence of strong effect from the R31A mutant suggests that electrostatic interaction of Arg31 with Glu41 may be transient and weak. The theoretical modelling of the electrostatics on the binding surface of CD2d1 also showed considerable over-estimation of the influence of basic groups on the binding surface compare with the pK_a values obtained experimentally. One possible interpretation is that for CD2d1 in solution condition, the interaction between the opposite charges are less intense due to the freedom of rotation of the side-chains of the basic residues. Nevertheless, as mentioned earlier, the absence of observed close-packing of the side chains indicates that the electrostatic interaction between charged groups may be largely responsible for the relative rigidity of the side-chains of Glu41 and Glu29. The nature and strength of such interactions in solution, however, remain to be assessed fully.

Self-association of CD2d1?

The surprising result from the analysis of the backbone amide and aliphatic side chain chemical shifts changes is the extensive influence exerted by the anomalous pK_a of Glu41. Large number of residues, some more than 15 Å away from Glu41, showed significant changes in chemical shifts with an apparent pK_a that correspond to the ionisation constant of Glu41. In contrast, only a few nearby residues exhibited any significant influence from the protonation event on His12. When the residues affected by Glu41 are mapped onto the structure of CD2d1, they are found to lie in a contiguous patch which covers part of the F β-strand, the CC',

C'C" and FG loops and the nearly all the residues on the C, C' and C" β-strands. Although the electrostatic effect may be transmitted some distance through the low dielectric constant of the protein interior, the magnitude and the extent of Glu41 ionisation effects, as well as their insensitivity to screening by salt, suggests that some other property is responsible for this behaviour.

One possibility is the pH-induced conformation changes of the side chains, as proposed by Kohda *et al* (Kohda *et al.*, 1991) for the observed effect of a titrating residue on nuclei situated 15Å away in murine epidermal growth factor. Charge-charge interactions are often involved in conformational changes (Brown *et al.*, 1997; March *et al.*, 1982; Perutz, 1978), and any such conformational changes will be reflected in the chemical shift changes. The side-chains of two charged residues can be expected to move apart due to the strong electrostatic repulsion (Dillet *et al.*, 1998), a process that could clearly be coupled to local conformational changes. Such an effect may be indicated for CD2d1 by the pH-dependent degeneracy of the β-CH_2 proton resonances of Asp28 described earlier, suggesting possible rearrangement in the dispositions of other side chains in its vicinity in a process coupled to the protonation states of glutamates.

However, a more probable explanation for the extensive effect of the ionisation state of Glu41 may be self-association of CD2d1. A number of observations support this contention. The extensive changes in chemical shifts as seen in CD2d1 is often indicative of protein-protein interaction and also widely used to determined the binding sites in proteins (Otting, 1993). For example, it was used for the mapping of antigen-binding and domain-domain interaction sites in immunoglobulin G (IgG) (Kato *et al.*, 1993), as well as the CD48 binding sites in CD2d1 (McAlister *et al.*, 1996). It is therefore significant that this patch of residues exhibiting significant influence of the ionisation state of Glu41 showed correspondence with those residues whose resonances were perturbed upon CD2/CD48 binding (McAlister *et al.*, 1996), as well as with residues involved in a major intermolecular contact in the crystal structure (Jones *et al.*, 1992). It also corresponds to the "structural epitope" as identified by exhaustive mutagenesis and binding studies (Davis *et al.*, 1998; van der Merwe *et al.*, 1995).

Perturbation of pK_as as seen in Glu41 have been demonstrated in other proteins upon ligand binding as shown by the change in the pK_as on binding of

immunosuppressant ligands by cyclophilin and FK506-binding protein (Yu & Fesik, 1994), and the binding of substrates by aspartic proteases such as pepsin, rhizopuspepsin and HIV-1 protease (Lin *et al.*, 1992; Wang *et al.*, 1996). Similar perturbations of pK_as have also been demonstrated for protein-protein interactions, for example, in the interaction between the B fragment of Staphylococcal protein-A and the Fc fragment of mouse immunoglobulin-G (Kato *et al.*, 1993). Such changes in pK_a can be seen in the repeat of pH titration at a low protein concentration which sees the pK_a of Glu41 dropping from 6.73 ([CD2d1]=1.2 mM) to 6.34 ([CD2d1]=0.1 mM). The increased pK_a at higher protein concentration runs counter to the expectation that pK_a of Glu41 should decrease at this higher concentration (Kesvatera *et al.*, 1996). This would indicates that Glu41 and other glutamates on the binding site are in an environment that favours the protonated state, a situation that can arise as a result of the binding surface being buried upon self-association at higher protein concentration.

Moreover, some of the ^{15}N HSQC and ^{13}C HCCO cross-peaks exhibit pronounced pH-dependent exchange broadening behaviour at the higher protein concentration. The abolition of such exchange broadening behaviour at low protein concentration is suggestive of increased protein-protein interaction occurring at higher protein concentration. All the results therefore point towards a pH dependent self-association of CD2d1, a process that may be modulated by and dependent on the ionisation state of Glu41. Previous biophysical studies, however, did not reveal detectable dimer formation under physiological conditions (Silkowski *et al.*, 1997; van der Merwe *et al.*, 1994; Withka *et al.*, 1993), these observations that are highly suggestive of CD2d1 self-association therefore warrant further investigations. A series of further experiments were therefore conducted in order to examine and confirm this pH-dependent self-association, the results of which will be discussed in Chapter 6.

CD2/CD48 interaction

The abnormally high pK_a of Glu41 suggests that Glu41 may be protonated during the self-association of CD2d1, it also raises question about the protonation states of Glu41 when CD2 is in complex with CD48. A repeat titration showed an

apparent drop of pK_a down to 6.3. However, it is not certain if this pK_a is the pK_a of Glu41 when CD2d1 is in complex with CD48, or if it represents the pK_a of a very small amount of uncomplexed CD2d1. There are a few reasons for this uncertainty. The pK_a determined is similar to that determined for Glu41 at low protein concentration (pK_a = 6.36), the pK_a of Glu41 in the absence of self-association of CD2d1. In addition, when CD2d1 is in complex with CD48, the pK_a of Glu41 would be expected to change. It has been suggested that Glu41 forms a salt bridge with R31 of CD48, however, any such electrostatic interaction may be expected to significantly lower the pK_a of Glu41; however, if CD2/CD48 interaction resembles CD2d1 self-association, the pK_a would be expected to increase. The lack of any such changes in Glu41 could suggest that the very weak peaks used for the determination of pK_a may come from uncomplexed CD2d1. On the other hand, the Glu41 may contribute little to the binding energy (Davis *et al.*, 1998), in which case the absence of pK_a change therefore may not be unexpected. The observed lost of binding of CD2 to CD48 occurs beginning at pH values ~6. If a protonated Glu41 is preferred during CD2 binding to CD48, the lost of binding must be the result of other charged residues as Glu41 will be largely protonated at ~pH 6. The data is therefore inconclusive and the difficulty in interpreting the result is due largely to the use of the glycosylated, 2-domain CD48 which greatly increases the apparent relative molecular mass of the CD2 when in complex with CD48. The severe peak broadening rendered the observation of resonances as well as the analysis difficult. For a more accurate and detailed assessment of the ionisation state of charged residues on CD2d1 on CD48 binding, a more tractable CD48 may need to be produced.

Conclusion

The analysis suggests that the anomalous pK_a of Glu41 may be the a number of different factor: charge-charge interaction between the glutamates, in particular, the interaction between Glu29 and Glu41; limited solvent accessibility and presence of nearby hydrophobic groups such as Leu38 and Phe49; as well as possible protein self-association. The pK_a values determined at high protein concentration should be

regarded as that of a complexed CD2 in rapid equilibrium with the uncomplexed form, while the pK_a determined at low protein concentration could be regarded as the actual pK_a of CD2d1 in monomeric state. The analysis is therefore complicated by a number of factor, all of which may have considerable effect on the pK_a. For a more complete understanding of the electrostatic effect, it may be necessary to repeat the titration of the mutants at a suitably low protein concentration in order to eliminate the protein concentration effect that may arise from protein self-association. Alternatively, a prior determination of the dissociation constant for all the mutants should be done before a repeat of the pH titration.

References

Abe, Y., Ueda, T., Iwashita, H., Hashimoto, Y., Motoshima, H., Tanaka, Y. & Imoto, T. (1995). Effect of Salt Concentration On the pK(a) of Acidic Residues in Lysozyme. *Journal of Biochemistry* **118**(5), 946-952.

Andersen, J. F., Sanders, D. A. R., Gasdaska, J. R., Weichsel, A., Powis, G. & Montfort, W. R. (1997). Human thioredoxin homodimers: Regulation by pH, role of aspartate 60, and crystal structure of the aspartate 60->asparagine mutant. *Biochemistry* **36**(46), 13979-13988.

Antosiewicz, J., McCammon, J. A. & Gilson, M. K. (1996). The determinants of pK(a)s in proteins. *Biochemistry* **35**(24), 7819-7833.

Baker, W. R. & Kintanar, A. (1996). Characterization of the pH titration shifts of ribonuclease A by one- and two-dimensional nuclear magnetic resonance spectroscopy. *Archives of Biochemistry and Biophysics* **327**(1), 189-199.

Barlow, D. J. & Thornton, J. M. (1983). Ion-Pairs in Proteins. *Journal of Molecular Biology* **168**(4), 867-885.

Bartik, K., Dobson, C. M. & Redfield, C. (1993). H-1-NMR Analysis of Turkey Egg-White Lysozyme and Comparison With Hen Egg-White Lysozyme. *European Journal of Biochemistry* **215**(2), 255-266.

Bashford, D., Case, D. A., Dalvit, C., Tennant, L. & Wright, P. E. (1993). Electrostatic Calculations of Side-Chain pK(a) Values in Myoglobin and Comparison With NMR Data For Histidines. *Biochemistry* **32**(31), 8045-8056.

Bashford, D. & Gerwert, K. (1992). Electrostatic Calculations of the pKa Values of Ionizable Groups in Bacteriorhodopsin. *Journal of Molecular Biology* **224**(2), 473-486.

Bashford, D. & Karplus, M. (1990). pKas of Ionizable Groups in Proteins - Atomic Detail From a Continuum Electrostatic Model. *Biochemistry* **29**(44), 10219-10225.

Bax, A., Ikura, M., Kay, L. E., Torchia, D. A. & Tschudin, R. (1990). Comparison of Different Modes of 2-Dimensional Reverse-Correlation NMR For the Study of Proteins. *Journal of Magnetic Resonance* **86**(2), 304-318.

Berisio, R., Lamzin, V. S., Sica, F., Wilson, K. S., Zagari, A. & Mazzarella, L. (1999). Protein Titration in the Crystal State. *Journal of Molecular Biology* **292**, 845-854.

Bordo, D. & Argos, P. (1991). Suggestions For Safe Residue Substitutions in Site-Directed Mutagenesis. *Journal of Molecular Biology* **217**(4), 721-729.

Bradbury, J. H. & Scheraga, H. A. (1966). Structural Studies of Ribonuclease. XXIV. The Application of Nuclear Magnetic Resonance Spectroscopy to Distinguish Between Histidine Residues of Ribonuclease. *Journal of American Chemical Society* **88**, 4240.

Brocklehurst, K. (1994). A Sound Basis For pH-Dependent Kinetic-Studies On Enzymes. *Protein Engineering* **7**(3), 291-299.

Brown, L. S., Kamikubo, H., Zimanyi, L., Kataoka, M., Tokunaga, F., Verdegem, P., Lugtenburg, J., Lanyi, J. K. (1997). A local electrostatic change is the cause of the large-scale protein conformation shift in bacteriorhodopsin. *Proc Natl Acad Sci U S A; **94**(10):5040-5044.*

Bundi, A. & Wuthrich, K. (1979a). 1H-NMR Parameters of the Common Amino Acid Residues Measured in Aqueous Solutions of the Linear Tetrapepetide H-Gly-Gly-X-L-ALA -OH. *Biopolymers* **18**, 285 - 297.

Bundi, A. & Wuthrich, K. (1979b). Use of Amide 1H-NMR Titration Shifts for Studies of Polypeptide Conformation. *Biopolymers* **18**, 299-311.

Cederholm, M. T., Stuckey, J. A., Doscher, M. S. & Lee, L. (1991). Histidine pKa Shifts Accompanying the Inactivating Asp121-] Asn Substitution in a Semisynthetic Bovine Pancreatic Ribonuclease. *Proceedings of the National Academy of Sciences of the United States of America* **88**(18), 8116-8120.

Chivers, P. T., Prehoda, K. E., Volkman, B. F., Kim, B. M., Markley, J. L. & Raines, R. T. (1997). Microscopic pK(a) values of Escherichia coli thioredoxin. *Biochemistry* **36**(48), 14985-14991.

Cohen, J. S., Shrager, R. I., McNeel, M. & Schechter, A. N. (1970). Proton Magnetic Resonance Studies at 220 MHz of the Histidine Residues of Staphylococcal Nuclease. *Nature* **228**, 642 - 644.

Creighton, T. E. (1993). *Proteins - Structures and Molecular Properties*, W.H Freeman and Company.

Dao-Pin, S., Sauer, U., Nicholson, H. & Matthews, B. W. (1991). Contributions of Engineered Surface Salt Bridges to the Stability of T4 Lysozyme Determined By Directed Mutagenesis. *Biochemistry* **30**(29), 7142-7153.

Davis, S. J., Davies, E. A., Tucknott, M. G., Jones, E. Y. & van der Merwe, P. A. (1998). The role of charged residues mediating low affinity protein-protein recognition at the cell surface by CD2. *Proceedings of the National Academy of Sciences of the United States of America* **95**(10), 5490-5494.

Davoodi, J., Wakarchuk, W. W., Campbell, R. L., Carey, P. R. & Surewicz, W. K. (1995). Abnormally High pK(a) of an Active-Site Glutamic-Acid Residue in *Bacillus-Circulans* Xylanase - the Role of Electrostatic Interactions. *European Journal of Biochemistry* **232**(3), 839-843.

de Dios, A. C., Pearson, J. G. & Oldfield, E. (1993). Secondary and Tertiary Structural Effects On Protein NMR Chemical- Shifts - an ab initio Approach. *Science* **260**(5113), 1491-1496.

Dillet, V., Dyson, H. J. & Bashford, D. (1998). Calculations of electrostatic interactions and pK(a)s in the active site of *Escherichia coli* thioredoxin. *Biochemistry* **37**(28), 10298-10306.

Driscoll, P. C., Cyster, J. G., Campbell, I. D. & Williams, A. F. (1991). Structure of Domain-1 of Rat Lymphocyte-T CD2 Antigen. *Nature* **353**(6346), 762-765.

Ebina, S. & Wuthrich, K. (1984). Amide Proton Titration Shifts in Bull Seminal Inhibitor Iia By Two-Dimensional Correlated H-1 Nuclear Magnetic-Resonance (COSY) - Manifestation of Conformational Equilibria Involving Carboxylate Groups. *Journal of Molecular Biology* **179**(2), 283-288.

Edsel, J. T. & Wyman, J. (1958). *Biophysical Chemistry*, Academic Press, New York.

Ferrin, T. E., Huang, C. C., Jarvis, L. E. & Langridge, R. (1988). The MIDAS Display System. *Journal of Molecular Graphics* **6**(1), 13-27.

Fersht, A. (1985). *Enzyme Structure and Mechanism*, W.H. Freeman and Co.

Forman-Kay, J. D., Clore, G. M. & Gronenborn, A. M. (1992). Relationship Between Electrostatics and Redox Function in Human Thioredoxin - Characterization of pH Titration Shifts Using 2-Dimensional Homonuclear and Heteronuclear NMR. *Biochemistry* **31**(13), 3442-3452.

Forsyth, W. R., Gilson, M. K., Antosiewicz, J., Jaren, O. R. & Robertson, A. D. (1998). Theoretical and experimental analysis of ionization equilibria in ovomucoid third domain. *Biochemistry* **37**(24), 8643-8652.

Gibas, C. J., Subramaniam, S., McCammon, J. A., Braden, B. C. & Poljak, R. J. (1997). pH dependence of antibody/lysozyme complexation. *Biochemistry* **36**(50), 15599-15614.

Gilson, M. K. (1995). Theory of Electrostatic Interactions in Macromolecules. *Current Opinion in Structural Biology* **5**(2), 216-223.

Gilson, M. K. & Honig, B. H. (1988). Energetics of Charge-Charge interaction in Proteins. *Protein: Structure, Function & Genetics* **3**, 32-52.

Honig, B. & Nicholls, A. (1995). Classical Electrostatics in Biology and Chemistry. *Science* **268**(5214), 1144-1149.

Hubbard, S. J., Campbell, S. F. & Thornton, J. M. (1991). Molecular Recognition - Conformational-Analysis of Limited Proteolytic Sites and Serine Proteinase Protein Inhibitors. *Journal of Molecular Biology* **220**(2), 507-530.

Inoue, M., Yamada, H., Yasukochi, T., Kuroki, R., Miki, T., Horiuchi, T. & Imoto, T. (1992). Multiple Role of Hydrophobicity of Tryptophan-108 in Chicken Lysozyme - Structural Stability, Saccharide Binding Ability, and Abnormal pKa of Glutamic Acid-35. *Biochemistry* **31**(24), 5545-5553.

Jackson, S. E. & Fersht, A. R. (1993). Contribution of Long-Range Electrostatic Interactions to the Stabilization of the Catalytic Transition-State of the Serine- Protease Subtilisin Bpn'. *Biochemistry* **32**(50), 13909-13916.

Jeng, M. F. & Dyson, H. J. (1996). Direct Measurement of the Aspartic-Acid-26 pK(a) For Reduced *Escherichia-Coli* Thioredoxin By C-13 NMR. *Biochemistry* **35**(1), 1-6.

Jones, E. Y., Davis, S. J., Williams, A. F., Harlos, K. & Stuart, D. I. (1992). Crystal-Structure At 2.8-Angstrom Resolution of a Soluble Form of the Cell-Adhesion Molecule CD2. *Nature* **360**(6401), 232-239.

Jones, S. & Thornton, J. M. http://www.biochem.ucl.ac.uk/bsm/PP/server/ Protein-Protein Interaction Server. Department of Biochemistry and Molecular Biology, University College, London, England.

Joshi, M. D., Hedberg, A. & McIntosh, L. P. (1997). Complete measurement of the pK(a) values of the carboxyl and imidazole groups in Bacillus circulans xylanase. *Protein Science* **6**(12), 2667-2670.

Juffer, A. H. (1998). Theoretical Calculations of Acid-Dissociation Constants of Proteins. *Biochem. Cell Biol.* **76**, 198-209.

Kato, K., Gouda, H., Takaha, W., Yoshino, A., Matsunaga, C. & Arata, Y. (1993). C-13 NMR-Study of the Mode of Interaction in Solution of the B Fragment of Staphylococcal Protein-A and the Fc Fragments of Mouse Immunoglobulin-G. *Febs Letters* **328**(1-2), 49-54.

Kesvatera, T., Jonsson, B., Thulin, E. & Linse, S. (1996). Measurement and modelling of sequence-specific pK(a) values of lysine residues in calbindin D-9k. *Journal of Molecular Biology* **259**(4), 828-839.

Khare, D., Alexander, P., Antosiewicz, J., Bryan, P., Gilson, M. & Orban, J. (1997). pK(a) measurements from nuclear magnetic resonance for the B1 and B2 immunoglobulin G-binding domains of protein G: Comparison with calculated values for nuclear magnetic resonance and x-ray structures. *Biochemistry* **36**(12), 3580-3589.

Kinney, J. J. (1997). *Probability - an introduction with statistical applications*, John Wiley & Sons, Inc.

Kohda, D., Sawada, T. & Inagaki, F. (1991). Characterization of pH Titration Shifts For All the Nonlabile Proton Resonances in a Protein By 2-Dimensional NMR - the Case of Mouse Epidermal Growth-Factor. *Biochemistry* **30**(20), 4896-4900.

Kuhlman, B., Luisi, D. L., Young, P. & Raleigh, D. P. (1999). pKa values and the pH dependent stability of the N-terminal domain of L9 as probes of electrostatic interactions in the denatured state. Differentiation between local and nonlocal interactions. *Biochemistry* **38**(15), 4896-4903.

Ladner, H. K., Led, J. J. & Grant, D. M. (1975). Deuterium Isotope Effects on 13C Chemical Shifts in Amino Acids and Dipeptides. *Journal of Magnetic Resonance* **20**, 530-534.

Langsetmo, K., Fuchs, J. A. & Woodward, C. (1991). The Conserved, Buried Aspartic-Acid in Oxidized Escherichia-Coli Thioredoxin Has a pKa of 7.5 - Its Titration Produces a Related Shift in Global Stability. *Biochemistry* **30**(30), 7603-7609.

Laskowski, R. A. (1995). Surfnet - a Program For Visualizing Molecular-Surfaces, Cavities, and Intermolecular Interactions. *Journal of Molecular Graphics* **13**(5), 323 et seq.

Led, J. J. & Petersen, S. B. (1979). Deuterium Isotope Effects on Carbon-13 Chemical Shifts in Selected Amino Acids as Function of pH. *Journal of Magnetic Resonance* **33**, 603-617.

Li, H. M., Hanson, C., Fuchs, J. A., Woodward, C. & Thomas, G. J. (1993). Determination of the pKa Values of Active-Center Cysteines, Cysteines-32 and Cysteines-35, in *Escherichia-Coli* Thioredoxin By Raman-Spectroscopy. *Biochemistry* **32**(22), 5800-5808.

Lin, Y. Z., Fusek, M., Lin, X. L., Hartsuck, J. A., Kezdy, F. J. & Tang, J. (1992). pH-Dependence of Kinetic-Parameters of Pepsin, Rhizopuspepsin, and Their Active-Site Hydrogen-Bond Mutants. *Journal of Biological Chemistry* **267**(26), 18413-18418.

Loewenthal, R., Sancho, J., Reinikainen, T. & Fersht, A. R. (1993). Long-Range Surface-Charge Charge Interactions in Proteins - Comparison of Experimental Results With Calculations From a Theoretical Method. *Journal of Molecular Biology* **232**(2), 574-583.

March, K. L., Maskalick, D. G., England, R. D., Friend, S. H. & Gurd, F. R. N. (1982). Analysis of Electrostatic Interactions and Their Relationship to Conformation and Stability of Bovine Pancreatic Trypsin-Inhibitor. *Biochemistry* **21**(21), 5241-5251.

Matthew, J. B. (1985). Electrostatic Effects in Proteins. *Biophysical Journal* **47**(2), A20.

Matthew, J. B., Gurd, F. R. N., Garciamoreno, E. B., Flanagan, M. A., March, K. L. & Shire, S. J. (1985). pH-Dependent Processes in Proteins. *CRC Critical Reviews in Biochemistry* **18**(2), 91-197.

McAlister, M. S. B., Mott, H. R., van der Merwe, P. A., Campbell, I. D., Davis, S. J. & Driscoll, P. C. (1996). NMR Analysis of Interacting Soluble Forms of the Cell-Cell Recognition Molecules CD2 and CD48. *Biochemistry* **35**(19), 5982-5991.

McDonald, I. K. & Thornton, J. M. (1994). Satisfying Hydrogen-Bonding Potential in Proteins. *Journal of Molecular Biology* **238**(5), 777-793.

McIntosh, L. P., Hand, G., Johnson, P. E., Joshi, M. D., Korner, M., Plesniak, L. A., Ziser, L., Wakarchuk, W. W. & Withers, S. G. (1996). The pK(a) of the general acid/base carboxyl group of a glycosidase cycles during catalysis: A C-13-NMR study of Bacillus circuluns xylanase. *Biochemistry* **35**(31), 9958-9966.

Meadows, D. H., Markley, J. L., Cohen, J. S. & Jardetzky, O. (1967). Nuclear Magnetic Resonance Studies of the Structure and Binding Sites of Enzymes I. Histidine Residues. *Proceedings of National Academy of Sciences USA* **58**, 1307.

Metz, G., Siebert, F. & Engelhard, M. (1992). High-Resolution Solid-State C-13 NMR of Bacteriorhodopsin - Characterization of [4-C-13]Asp Resonances. *Biochemistry* **31**(2), 455-462.

Nakamura, H. (1996). Roles of electrostatic interaction in proteins. *Quarterly Reviews of Biophysics* **29**(1), 1-90.

Norwood, T. J., Boyd, J., Soffe, N. & Campbell, J. D. (1990). New Nuclear-Magnetic-Resonance Technique For Determining Long-Range Heteronuclear H-1-N-15 Correlations in Proteins. *Journal of the American Chemical Society* **112**(26), 9638-9640.

Novotny, J. & Sharp, K. (1992). Electrostatic Fields in Antibodies and Antibody Antigen Complexes. *Progress in Biophysics & Molecular Biology* **58**(3), 203-224.

Nozaki, Y. & Tanford, C. (1967). Examination of Titration Behaviour. *Methods in Enzymology* **84**, 714-734.

Oda, Y., Yamazaki, T., Nagayama, K., Kanaya, S., Kuroda, Y. & Nakamura, H. (1994). Individual Ionization-Constants of All the Carboxyl Groups in Ribonuclease H1 From *Escherichia-Coli* Determined By NMR. *Biochemistry* **33**(17), 5275-5284.

Otting, G. (1993). Experimental NMR Techniques For Studies of Protein Ligand Interactions. *Current Opinion in Structural Biology* **3**(5), 760-768.

Perez-Canadillas, J. M., Campos-Olivas, R., Lacadena, J., del Pozo, A. M., Gavilanes, J. G., Santoro, J., Rico, M. & Bruix, M. (1998). Characterization of pK(a) values and titration shifts in the cytotoxic ribonuclease alpha-sarcin by NMR. Relationship between electrostatic interactions, structure, and catalytic function. *Biochemistry* **37**(45), 15865-15876.

Perrin, C. L. & Thoburn, J. D. (1992). Symmetries of Hydrogen-Bonds in Monoanions of Dicarboxylic-Acids. *Journal of the American Chemical Society* **114**(22), 8559-8565.

Perutz, M. F. (1978). Electrostatic Effects in Protein. *Science* **29**, 1187-1191.

Plesniak, L. A., Connelly, G. P., Wakarchuk, W. W. & McIntosh, L. P. (1996). Characterization of a buried neutral histidine residue in Bacillus circulans xylanase: NMR assignments, pH titration, and hydrogen exchange. *Protein Science* **5**(11), 2319-2328.

Qin, J., Clore, G. M. & Gronenborn, A. M. (1996). Ionization equilibria for side-chain carboxyl groups in oxidized and reduced human thioredoxin and in the complex with its target peptide from the transcription factor NF kappa B. *Biochemistry* **35**(1), 7-13.

Richarz, R. & Wuthrich, K. (1978). High-field 13C Nuclear Magnetic Resonance Studies at 90.5 MHz of the Basic Pancreatic Trypsin Inhibitor. *Biochemistry* **17**(12), 2263 -2269.

Russell, A. J. & Fersht, A. R. (1987). Rational Modification of Enzyme Catalysis By Engineering Surface- Charge. *Nature* **328**(6130), 496-500.

Schaller, W. & Robertson, A. D. (1995). pH, Ionic-Strength, and Temperature Dependences of Ionization Equilibria For the Carboxyl Groups in Turkey Ovomucoid 3rd Domain. *Biochemistry* **34**(14), 4714-4723.

Serrano, L., Horovitz, A., Avron, B., Bycroft, M. & Fersht, A. R. (1990). Estimating the Contribution of Engineered Surface Electrostatic Interactions to Protein Stability By Using Double-Mutant Cycles. *Biochemistry* **29**(40), 9343-9352.

Shrager, R. I., Cohen, J. S., Heller, S. R., Sachs, D. H. & Schechter, A. N. (1972). Mathematical Models for Interacting Groups in Nuclear Magnetic Resonance Titration Curves. *Biochemistry* **11**, 541 - 547.

Silkowski, H., Davis, S. J., Barclay, A. N., Rowe, A. J., Harding, S. E. & Byron, O. (1997). Characterisation of the low affinity interaction between rat cell adhesion molecules CD2 and CD48 by analytical ultracentrifugation. *European Biophysics Journal With Biophysics Letters* **25**(5-6), 455-462.

Singer, A. U. & Forman-Kay, J. D. (1997). pH titration studies of an SH2 domain-phosphopeptide complex: Unusual histidine and phosphate pK(a) values. *Protein Science* **6**(9), 1910-1919.

Sorensen, M. D. & Led, J. J. (1994). Structural Details of Asp(B9) Human Insulin At Low pH From 2-Dimensional NMR Titration Studies. *Biochemistry* **33**(46), 13727-13733.

States, D. J. & Karplus, M. (1987). A Model For Electrostatic Effects in Proteins. *Journal of Molecular Biology* **197**(1), 122-130.

Szaraz, S., Oesterhelt, D. & Ormos, P. (1994). pH-Induced Structural-Changes in Bacteriorhodopsin Studied By Fourier-Transform Infrared-Spectroscopy. *Biophysical Journal* **67**(4), 1706-1712.

Szyperski, T., Antuch, W., Schick, M., Betz, A., Stone, S. R. & Wuthrich, K. (1994). Transient Hydrogen-Bonds Identified On the Surface of the NMR Solution Structure of Hirudin. *Biochemistry* **33**(31), 9303-9310.

Takahashi, N. & Creighton, T. E. (1996). On the reactivity and ionization of the active site cysteine residues of Escherichia coli thioredoxin. *Biochemistry* **35**(25), 8342-8353.

Tan, Y. J., Oliveberg, M., Davis, B. & Fersht, A. R. (1995). Perturbed pK(a)-Values in the Denatured States of Proteins. *Journal of Molecular Biology* **254**(5), 980-992.

Tavernor, A. S., Kydd, J. H., Bodian, D. L., Jones, E. Y., Stuart, D. I., Davis, S. J. & Butcher, G. W. (1994). Expression Cloning of an Equine T-Lymphocyte Glycoprotein CD2 cDNA Structure-Based Analysis of Conserved Sequence Elements. *European Journal of Biochemistry* **219**(3), 969-976.

Thomas, P. G., Russell, A. J. & Fersht, A. R. (1985). Tailoring the pH-Dependence of Enzyme Catalysis Using Protein Engineering. *Nature* **318**(6044), 375-376.

Tishmack, P. A., Bashford, D., Harms, E. & VanEtten, R. L. (1997). Use of H-1 NMR spectroscopy and computer simulations to analyze histidine pK(a) changes in a protein tyrosine phosphatase: Experimental and theoretical determination of electrostatic properties in a small protein. *Biochemistry* **36**(39), 11984-11994.

Tozawa, K., Ohbuchi, H., Yagi, H., Amano, T., Matsui, T., Yoshida, M. & Akutsu, H. (1996). Unusual pK(a) of the carboxylate at the putative catalytic position of thermophilic F-1-ATPase beta subunit determined by C-13-NMR. *Febs Letters* **397**(1), 122-126.

Urry, D. W., Gowda, D. C., Peng, S. Q., Parker, T. M., Jing, N. J. & Harris, R. D. (1994). Nanometric Design of Extraordinary Hydrophobic-Induced pKa Shifts For Aspartic-Acid - Relevance to Protein Mechanisms. *Biopolymers* **34**(7), 889-896.

van der Merwe, P. A., Brown, M. H., Davis, S. J. & Barclay, A. N. (1994). Affinity and Kinetic-Analysis of the Interaction of the Cell-Adhesion Molecule CD2 With Its Ligands CD48 and CD58. *Journal of Cellular Biochemistry*(S18C SIC), 257-257.

van der Merwe, P. A., McNamee, P. N., Davies, E. A., Barclay, A. N. & Davis, S. J. (1995). Topology of the CD2-CD48 Cell-Adhesion Molecule Complex - Implications For Antigen Recognition By T-Cells. *Current Biology* **5**(1), 74-84.

van Vlijmen, H. W., Schaefer, M., & Karplus M. (1998). Improving the accuracy of protein pKa calculations: conformational averaging versus the average structure. *Proteins*; **33**(2):145-158

Varadarajan, R., Lambright, D. G. & Boxer, S. G. (1989). Electrostatic Interactions in Wild-Type and Mutant Recombinant Human Myoglobins. *Biochemistry* **28**(9), 3771-3781.

Vuister, G. W. & Bax, A. (1992). Resolution Enhancement and Spectral Editing of Uniformly C-13-Enriched Proteins By Homonuclear Broad-Band C-13 Decoupling. *Journal of Magnetic Resonance* **98**(2), 428-435.

Wade, R. C., Gabdoulline, R. R., Ludemann, S. K. & Lounnas, V. (1998). Electrostatic steering and ionic tethering in enzyme-ligand binding: Insights from simulations. *Proceedings of the National Academy of Sciences of the United States of America* **95**(11), 5942-5949.

Wagner, G., Pardi, A. & Wuthrich, K. (1983). Hydrogen-Bond Length and H-1-NMR Chemical-Shifts in Proteins. *Journal of the American Chemical Society* **105**(18), 5948-5949.

Wang, J., Smolyar, A., Tan, K. M., Liu, J., Kim, M. Y., Sun, Z. J., Wagner, G. & Reinherz, E. L. (1999). Structure of a heterophilic adhesion complex between the human CD2 and CD58 (LFA-3) counterreceptors. *Cell* **97**(6), 791-803.

Wang, Y. X., Freedberg, D. I., Yamazaki, T., Wingfield, P. T., Stahl, S. J., Kaufman, J. D., Kiso, Y. & Torchia, D. A. (1996). Solution NMR evidence that the HIV-1 protease catalytic aspartyl groups have different ionization states in the complex formed with the asymmetric drug KNI-272. *Biochemistry* **35**(31), 9945-9950.

Warshel, A. & Papazyan, A. (1998). Electrostatic effects in macromolecules: fundamental concepts and practical modeling. *Current Opinion in Structural Biology* **8**(2), 211-217.

Wilson, N. A., Barbar, E., Fuchs, J. A. & Woodward, C. (1995). Aspartic-Acid 26 in Reduced *Escherichia-Coli* Thioredoxin Has a pK(a) Greater-Than-9. *Biochemistry* **34**(28), 8931-8939.

Wishart, D. S., Bigam, C. G., Yao, J., Abilgaard, F., Dyson, H. J., Oldfield, E., Markley, J. L. & Sykes, B. D. (1995). 1H, 13C and 15N Chemical Shift Referencing in Biomolecular NMR. *Journal of Biomolecular NMR* **6**, 135-140.

Withka, J. M., Wyss, D. F., Wagner, G., Arulanandam, A. R. N., Reinherz, E. L. & Recny, M. A. (1993). Structure of the Glycosylated Adhesion Domain of Human T-Lymphocyte Glycoprotein CD2. *Structure* **1**(1), 69-81.

Yamazaki, T., Nicholson, L. K., Torchia, D. A., Wingfield, P., Stahl, S. J., Kaufman, J. D., Eyermann, C. J., Hodge, C. N., Lam, P. Y. S., Ru, Y., Jadhav, P. K., Chang, C. H. & Weber, P. C. (1994). NMR and X-Ray Evidence That the HIV Protease Catalytic Aspartyl Groups Are Protonated in the Complex Formed By the Protease and a Nonpeptide Cyclic Urea-Based Inhibitor. *Journal of the American Chemical Society* **116**(23), 10791-10792.

Yang, A. S., Gunner, M. R., Sampogna, R., Sharp, K. & Honig, B. (1993). On the Calculation of pK(a)S in Proteins. *Proteins-Structure Function and Genetics* **15**(3), 252-265.

Yu, L. P. & Fesik, S. W. (1994). pH Titration of the Histidine-Residues of Cyclophilin and Fk506 Binding-Protein in the Absence and Presence of Immunosuppressant Ligands. *Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology* **1209**(1), 24-32.

Zhang, W. X. & Gmeiner, W. H. (1996). Improved 3D gd-HCACO and gd-(H)CACO-TOCSY experiments for isotopically enriched proteins dissolved in H2O. *Journal of Biomolecular NMR* **7**(3), 247-250.

Zhou, H. X. & Vijayakumar, M. (1997). Modeling of protein conformational fluctuations in pK(a) predictions. *Journal of Molecular Biology* **267**(4), 1002-1011.

Chapter 6

Molecular dynamics and monomer-dimer equilibria of CD2d1

As described in the previous chapter, the analyses of the ionisation equilibria, chemical shift changes and exchange broadening behaviour of CD2d1 suggest possible self-association of CD2d1. This observation is surprising as previous studies did not reveal any sign of self-association of CD2d1. Hydrodynamic studies on rat CD2 showed rat CD2 to be essentially monomeric under physiological conditions (Silkowski *et al.*, 1997), while for human CD2 the narrow linewidth and lack of detectable intermolecular NOEs was taken as evidence of lack of dimer formation (Withka *et al.*, 1993). Chemical shift changes, however, may be a more sensitive parameter for the detection of change in the electrostatic environment than NOEs (Folkers *et al.*, 1989). The extensive effect that was observed for Glu41 in CD2d1 on the resonances from the aliphatic side-chains and backbone amide may therefore be a more reliable indication of self-association where such interaction is weak.

In this chapter, this possible self-association of CD2d1 is investigated by a number of techniques, principally by NMR relaxation analysis. Relaxation analysis

was employed as it can yield, in addition to information about the association state, much useful information about the internal dynamics of the protein. The analysis of molecular dynamics may provide significant insight into the functioning of CD2d1 as the dynamic properties of adhesion molecules can affect its kinetic properties such as association and dissociation rates, and its thermodynamics properties such as binding equilibria. The knowledge of the molecular motions and flexibility within the folded polypeptide chain can therefore contribute to the understanding of the protein function. The binding affinity and association rate of human CD2d1, for example, have been suggested to be associated with the internal motions of the molecules (Wyss *et al.*, 1997). Previous relaxation analyses on the dynamic properties of CD2d1, however, have produced inconsistent results, in particular, the rotational correlation time τ_C value was observed to be variable (Crawford, 1994). In this study, a vigorous approach is therefore taken to investigate the dynamics in CD2d1 so as to provide a comprehensive picture of the molecular and internal motions of the protein.

Relaxation and Molecular Dynamics

The fact that proteins are not rigid and show varying degree of flexibility is now well recognised. This flexibility may be manifested in the mobility of the side-chains and segments of proteins, or even entire independently moving domains (Jardetzky, 1996). Such internal motions may play an important role in the functional activity of proteins (Karplus & McCammon, 1983). For example, the 'static' 3D structure of myoglobin as obtained from X-ray crystallography has been shown to be incompatible with its oxygen transport kinetics, and movement of its backbone and side-chain residues in close proximity to the active site are necessary for oxygen binding and dissociation. Such mobility in protein can be identified using crystallographic and NMR methods. In X-ray crystallography, only a crude guide to internal motions in protein by can be afforded by the crystallographic temperature factors (B-factors), although progress has been made in the field of X-ray crystallography that can provide better indication of structural dynamics. NMR, on the other hand, can provide a more detailed and quantitative study of both the magnitudes and time-scale of internal motions in proteins. For example, molecular

motions in the range of picosecond to millisecond range may be investigated by NMR relaxation measurements, while line-shape analysis, saturation transfer, backbone proton exchange can be used to investigate motion on the millisecond to seconds time scale (Jardetzky, 1996).

Relaxation is one of the fundamental aspects of magnetic resonance that has important influence on many aspects of NMR experiments. The relaxation rates determine the linewidths of resonances and some aspects of the design of NMR experiments. Relaxation properties such as cross-relaxation can give rise to the nuclear Overhauser effect (NOE) that is used to provide the through-space connectivities necessary for 3D molecular structure determination by NMR (Cavanagh *et al.*, 1996). In relaxation analysis, physical processes such as molecular motions and internal dynamic properties of the polypeptide backbone of the protein can be investigated by monitoring the spin-relaxation properties of specific nuclei. For example, different relaxation parameters of the amide ^{15}N atoms in a uniformly ^{15}N-labelled protein can be measured to give information about the mobility of all amide sites along the peptide chains. In this chapter, a brief description of the theory of relaxation is introduced, further in-depth treatment of the theoretical and practical aspects of relaxation can be found in Abragam, (1961) and Cavanagh *et al.*, (1996).

Theory of Spin Relaxation

As described earlier in Chapter 4, a nuclear spin system can be excited by the application of a RF pulse which can invert the population of the spin such that a greater population of the spin is present in the excited β state. This excited state, however, will eventually revert back to the original equilibrium state. In other forms of spectroscopy such as UV and IR spectroscopy, the principal mechanism for this is by spontaneous emission. However, the efficiency of this mechanism is proportional to the third power of the excitation frequency; for NMR transitions, at the frequency at which it operates, a probability of such event occurring is of the order of ~ 10^{-21} s^{-1} for a proton with precession frequency of 500 MHz (Cavanagh *et al.*, 1996). This process of spontaneous emission is therefore negligible for NMR, instead the excited

spin in NMR returns to the ground state by the process of spin relaxation. The relaxation times in solution NMR are quite long, and the NMR resonances are therefore much sharper compared with lines in UV absorption spectrum.

The dynamic properties of a protein may be separated into two broad categories: 1) the overall tumbling of the protein, 2) the faster motion of internal 'dynamic units'. These internal 'dynamic units' may include groups such as phenyl or methyl rotating about a C-C bond, or and the reorientation of the C-H or N-H vectors with respect to the overall molecular frame. The internal motions as revealed by the backbone amides reflect the concerted movements with their nearest bonded neighbour and therefore the larger scale motions of surrounding groups of atoms. The overall tumbling of the molecule in solution and the internal dynamics result in the motion of individual atoms. Many of these atoms have magnetically active nuclei and the motions of these nuclei result in the generation of fluctuating local magnetic fields in much the same way that movement of a bar magnet in space creates a time-varying magnetic field. If these magnetic fields fluctuate at frequencies corresponding to the resonance frequencies of nuclei in the system, then spin relaxation may restore the signal to its equilibrium state.

There are, however, many physical interactions that are capable of mediating spin relaxation, such as intramolecular magnetic dipolar, anisotropic chemical shift (CSA), quadrapolar and scalar coupling interactions. The analysis is therefore very complex for relaxation analysis of nuclei such as the proton whereby large number of relaxation pathways exists. Moreover, the many efficient relaxation pathways present in protons results in the rapid decay of the proton signals. The measurement of ^{1}H relaxation parameters is therefore difficult and little useful information may be obtained about molecular motion (Abragam, 1961). In the study that was undertaken, the relaxation of the amide ^{15}N nuclei was used. This has the advantage that the relaxation process is governed primarily by the strong dipolar interaction with a single attached proton, and by chemical shift anisotropy (CSA). Alternative relaxation pathways can therefore be assumed to be negligible. In the dipolar interaction of the NH bond, the ^{15}N atom experiences the local dipolar field of its directly bonded proton. The NH bond dynamics cause the local field to fluctuate which encourages relaxation. In the CSA interaction, relaxation is caused by

fluctuation in the local shielding of the ^{15}N atoms, and reflects the motion of the principal axes of ^{15}N shielding tensor with respect to the static magnetic field B_0.

The standard relaxation parameters most commonly measured in the analysis of protein dynamics are: 1) the longitudinal relaxation time T_1, 2) the transverse relaxation time T_2, and 3) the heteronuclear nuclear Overhauser effect (NOE). Since each nucleus lies in slightly differently environment, the T_1, T_2 and NOE values for each nucleus would be slightly different, and individual relaxation parameters can therefore be determined for each residue. These different relaxation parameters can reflect motions on different time-scale ranging from pico- and nanoseconds to about 1 millisecond. The T_1 is sensitive to the rotational diffusion of the protein, and motions faster than the rotational diffusion rate, i.e. motions on 10^8-10^{12} s^{-1} time-scale. The T_2 may be sensitive to similar motion as T_1, however, in addition, it is also sensitive to slower motions on the micro- to millisecond time-scale, as well as to chemical exchange between environments and non-equivalent chemical shifts that result in line-broadening. The NOE mainly reflects the fast motions of sub-nanosecond time-scale. Sometimes an additional parameter $T_{1\rho}$ (T_1 in the rotating frame) which is sensitive to motions on slower time-scales may also be measured.

Theoretical model

The simplest theoretical approach to spin relaxation is introduced by Felix Bloch (Bloch, 1946) with his phenomenologically formulated theory of spin relaxation which can be described by the following equation:

$$dM / dt = \gamma (M \times H) - [M_z - M_0]k / T_1 - (M_x + M_y j) / T_2 \qquad [6.1]$$

where the M is the magnetisation of a damped oscillating vector in a magnetic field. It has a component on the z axis (M_z) that decays exponentially to the equilibrium state (M_0) with a time constant T_1, while its transverse components (M_x or M_y) also decay exponentially to an equilibrium value of zero characterised by T_2. This classical model therefore assumes that the relaxation processes are of first-order and can be described in terms of the two relaxation times T_1 and T_2. The relaxation

parameters can also be measured as the relaxation rate constant R_1 and R_2 which are the reciprocal of T_1 and T_2 respectively. A fuller description of these relaxation parameters is given below.

T_1

The time constant T_1, the longitudinal or spin-lattice relaxation time, describes the return of magnetisation on the z axis back to equilibrium as a result of the interaction of the spin with its surrounding, i.e. the fluctuating local magnetic field. For the amide ^{15}N nuclei, the major contribution to T_1 comes from the magnetic dipole interaction between proximal nuclei. For a protein in solution, the T_1 is normally is the range of seconds.

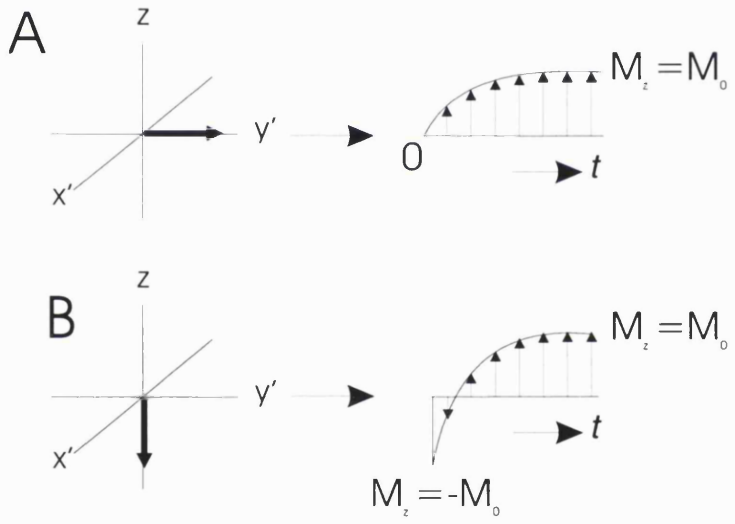


Fig. 6.1 Evolution with time of the longitudinal component M_z of the bulk magnetisation in the rotating frame. A - after a π/2 pulse. B - after a π pulse.

Consider Fig. 5.2 in which the bulk magnetisation M_0 has been tipped over by the application of a π or π/2 pulse, the magnetisation on the z-axis will eventually return to the equilibrium position. The Bloch equation for the description of this longitudinal relaxation is

$$\frac{dM_z}{dt} = -\frac{M_z - M_0}{T_1} \qquad [6.2]$$

Such that

$$M_z(t) = M_0(1 - exp(-t/T_1)) - M_z(0)\, exp(-t/T_1) \qquad [6.3]$$

in which $M_z(0)$ is the value of the component of the magnetisation along the z-axis at $t = 0$ (therefore $M_z(0) = 0$ after a $\pi/2$ pulse and $M_z(0) = -M_0$ after a π pulse).

The T_1 can therefore be measured by a simple 'inversion recovery' experiment. The z-magnetisation can be inverted with a π pulse and allowed to relax towards the $+z$ axis. The z-magnetisation, however, cannot be recorded. Therefore, after a time interval t, a $\pi/2$ pulse is applied which flips the z-magnetisation onto the x-y plane so that the FID can be recorded. By varying the interval t in a series of experiments, the intensity of the signal recorded would also vary. The intensity (I) of each peak (corresponding to a residue) in a series of spectra can then be measured and fitted as a function of the time to a single exponential equation derived from equation 6.3 to obtain the R_1 value:

$$I_t = I_0\,(1\text{-}2\exp\,(-R_1\, t)) \qquad [6.4]$$

where I_0 is the intensity at $t = 0$.

The pulse sequences used for the measurement of ^{15}N R_1, R_2, [^1H]-^{15}N NOE are all based on the ^1H-^{15}N HSQC experiment as described in Chapter 4 (Bodenhausen & Ruben, 1980). For the T_1 and T_2 measurements, a refocused INEPT is used for the first polarisation step such that after this step, the magnetisation is in-phase and not in the antiphase orientation as is normal for ^{15}N-^1H HSQC. For T_1 measurement, the relaxation time interval t can be incorporated into the ^1H-^{15}N HSQC pulse sequence as previously described (see Fig. 4.8) by its insertion before or after the incremented time delay t_1 (Barbato *et al.*, 1992; Kay *et al.*, 1989). In this version used which is gradient-enhanced, the time interval separate by two $\pi/2$ pulses on the ^{15}N nucleus is by necessity inserted before the time delay. The first $\pi/2$ pulse transfer the ^{15}N magnetisation to the z-axis followed by a relaxation delay. The length of this relaxation delay can be controlled by a number of repeats of ^1H π pulses on the proton which have the effect of removing the effects of cross-

correlation between the dipolar and CSA relaxation. The R_1 may be regarded as the simple sum of the relaxation rates due to dipolar interaction (DD) and CSA if these two terms do not interfere or correlate with each other. Therefore,

$$R_1 = R_1^{DD} + R_1^{CSA} \qquad [6.5]$$

The presence of cross-correlation can lead to non-exponential decay and error in the fitting of the curve. The second $\pi/2$ pulse rotates the magnetisation onto the *xy*-plane. The magnetisation is then transferred to the proton by a reverse INEPT for detection.

T_2

The T_2 is the transverse or spin-spin relaxation time which describes the lost of coherence of the *xy* components of the magnetisation through time. This T_2 relaxation arises as a result of linked relaxation of two nuclei in which the energy are mutually exchanged between spins (hence spin-spin relaxation), via a flip-flop mechanism. As described earlier in Chapter 4, the magnetisation may fan out as the isochromats precess at slightly different rate, the magnetisation components on the *x* or *y* axes, M_x and M_y, gradually lose coherence and diminish with time (Fig. 6.2). The description in Chapter 4 is, however, related to field inhomogeneity, this 'inauthentic' T_2 may be denoted as T_2^* to distinguish it from that due solely to fluctuations of local magnetic fields. Any process that contributes to the decay of the transverse magnetisation, including the return of the magnetisation to the *z*-axis (T_1), would also contribute to T_2. The T_2 is therefore almost always smaller than T_1 for protein, and T_1 can equal T_2 only in the extreme-narrowing limit. While T_1 determines the minimum time required for equilibrium to be restored, the T_2 determines the length of time that an FID can be observed.

The Bloch equations for the relaxation in the rotating frame *x'* and *y'* are

$$\frac{dM_{x'}}{dt} = -\frac{M_{x'}}{T_2} \qquad \text{and} \qquad \frac{dM_{y'}}{dt} = -\frac{M_{y'}}{T_2} \qquad [6.6]$$

giving

$$M_x(t) = M_x(0)\, exp(-t/T_2) \quad \text{and} \quad M_y(t) = M_y(0)\, exp(-t/T_2) \qquad [6.7]$$

in which $M_x(0)$ and $M_y(0)$ are the values of the transverse magnetisation at t=0.

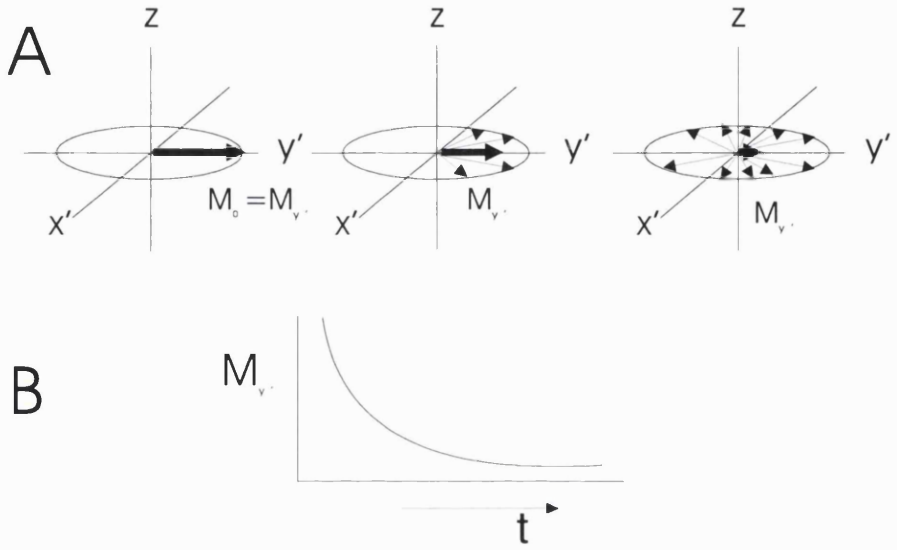


Fig. 6.2 Diagrammatic representation of transverse relaxation. A - a π/2 pulse rotates the bulk magnetisation M_0 onto the y' axis, the isochromats eventually fans outs due to spin-spin relaxation. B - diagram showing the exponential decay of the transverse decay of he transverse magnetisation component $M_{y'}$.

In order to measure the T_2, the spin echo sequence described in the Chapter 4 may be used to refocus the isochromats and removing the effect of field inhomogeneity. However, this spin echo has the disadvantage of accumulating errors due to imperfection of the pulses. Another version of the spin echo sequence is therefore applied for the R_2 measurement, the Carr-Purcell-Meiboom-Gill (CPMG) sequence that has the following pulse sequence:

$$(\pi/2)_x - \tau - \pi_y - 2\tau - \pi_y - 2\tau - \pi_y - \ldots.$$

The series of π_y pulses has the effect of refocusing the isochromats due to field inhomogeneity, as well as cancelling out any error that arises from imperfection of the pulse angle on an even number of the echoes. This sequence, however, cannot refocus effect due to the 'authentic' T_2 relaxation process. The refocused signal therefore shows a smaller amplitude from that of the original signal, and by

extending the CPMG pulse sequence, each successive spin echo would show decreasing amplitude. By varying the T_2 relaxation time delay, the decrease can be monitored by extracting the intensity of each peak in the series of ^{1}H -^{15}N HSQC spectra. The R_2 can then be derived by fitting the data to an equation derived from [6.7]:

$$I_t = I_0 \exp(-R_2 t) \qquad [6.8]$$

As in the T_1 experiment, the pulse sequence has a train of π pulses throughout the delay interval to suppress cross-correlation between dipolar and CSA relaxation mechanisms (Kay *et al.*, 1992b). These π pulses are applied at the height of the ^{15}N spin echoes. The time interval for T_2 measurement is inserted before the incremented time delay t_1.

However, the CPMG experiment can be systematically affected by chemical exchange during the spin-echo time delay, so that an extra exchange term may be necessary. This extra term, R_{ex}, is used to accommodate the chemical exchange and other pseudo-first-order processes that contribute to the decay of the transverse magnetisation, therefore,

$$R_2 = R_2^{DD} + R_2^{CSA} + R_{ex} \qquad [6.9]$$

This chemical exchange, as measured by the exchange term R_{ex}, can reflect the occurrence of a number of physical processes, its interpretation is therefore difficult and not entirely straightforward (Clore *et al.*, 1990a). Nevertheless, it is commonly taken to reflect the slower motions of segment of the polypeptide chain. However, these slower motions (or intermediate motion on the chemical shift time scale) are little understood. Slow motions can be associated with conformation transition and are described in vague terms such as fluctuations, breathing, conformational averaging, etc. The 'conformational exchange terms' in NMR relaxation studies are therefore localised to individual amide groups, however, the relationship between the exchange rate and the motions responsible for it remain ill-defined (Jardetzky, 1996).

NOE

In the heteronuclear NOE experiment, experiments with and without presaturation of the protons are performed. The saturation of the proton nuclei equalises the population difference of spins and enables the steady-state NOE condition to be observed. If two nuclei are close in space, the magnetisation can be transferred by dipolar coupling and the spin population difference of the ^{15}N nuclei can be enhanced. The NOE experiment can therefore be used to measure the saturation transfer of z-magnetisation between spins. The NOE enhancement can be quantified by comparing the measured intensities of the cross peaks in spectra with proton saturation (i.e. with NOE transfer) and without (no NOE transfer).

The [^1H]-^{15}N NOEs are therefore quantified using

$$NOE = I_{sat} / I_{unsat} \qquad [6.10]$$

where I_{sat} and I_{unsat} are the intensity measured in spectra recorded with and without initial ^1H saturation respectively. The cross-relaxation effect in NOE can be related by

$$NOE = 1 + (\sigma/R_1)(\gamma_H/\gamma_N) \qquad [6.11]$$

where σ is the cross-relaxation rate constant.

The sequence for measuring [^1H]-^{15}N NOE comprises of a train of 120° pulses prior to the first ^{15}N pulse, followed by production of transverse ^{15}N magnetisation, ^{15}N chemical shift evolution and reverse INEPT for magnetisation transfer to the amide proton for detection. Application of RF pulses for a long period of time is used for the saturation of the ^1H so that NOE enhancement can be established without perturbing the ^{15}N nucleus. The initial INEPT step used in the normal ^1H-^{15}N HSQC is absent as the experiment starts with the ^{15}N nucleus. The [^1H]-^{15}N NOE experiment is therefore the most error-prone experiment of the three relaxation experiments as the signals obtained are relatively weak due to the absence of the sensitivity enhancement step. The spectrum without NOE is recorded with identical parameters but without the proton saturation and the 120° pulses.

The rotational correlation time τ_c

The relaxation rates of the protein are strongly influenced by the overall tumbling of the molecule. The rotational correlation time τ_c, which is a measure of this molecular tumbling, corresponds roughly to the time interval between two successive reorientations or position changes of the molecules by rotational movement and is the approximate average time for the molecule to rotate by one radian. It may be calculated from a theoretical approach by the Stokes-Einstein relation which assume the molecule to be spherical and rigid:

$$\tau_c = \eta V/kT \qquad [6.12]$$

where η is the viscosity of the solvent, V is the volume of the sphere, k is the Boltzmann constant and T the temperature in Kelvin. Therefore, assuming all else being equal, the τ_c can be seen as proportional to the volume. In reality, there will be some degree of inaccuracy in the estimation of volume as the shape of protein is not necessarily spherical, and the shell of hydration in protein can also varies. The volume is estimated using average values for partial specific volume of a protein which is ~ 0.71-0.75 $cm^3\ g^{-1}$. For a globular protein, the volume can be regard roughly proportional to the relative molecular mass, consequently the τ_c can be a measure of the apparent relative molecular mass for a globular protein. Generally the τ_c is of the order of nanoseconds for biological macromolecules in aqueous solution (Cavanagh *et al.*, 1996).

The τ_c can also be obtained from the relaxation parameters as the calculated ratio of R_2/R_1 or T_1/T_2. For a molecule the size of protein, the τ_c can be scaled with the T_1; the slower a molecule moves, the greater is T_1. It also has a roughly inverse relationship with T_2, and a large T_2 implies a faster motion for the molecule. The effect of rapid motions (on the time scale of τ_e, the effective correlation time) may be ignored for the calculation of τ_c as, to a first approximation, these rapid motions lengthen both the T_1 and T_2 values by the same fraction (Clore *et al.*, 1990a; Kay *et al.*, 1989). For the analysis, however, residues exhibiting significant NOE effect are

excluded, as in such cases, the assumption that the very fast motion does not contribute significantly to T_1 is not valid (Barbato *et al.*, 1992).

The Lipari-Szabo Model-free Formalism

The central idea to the analysis of the NMR relaxation data is the spectral density function $J(\omega)$. This describes the frequency distribution function that profile the frequency content of the H-X bond motion, and therefore the power available from molecular motion at frequency ω to bring about relaxation. For the heteronuclear studies involving amide, the ^{15}N-^{1}H bond length is essentially fixed, therefore the bond motions refer to the rotational fluctuations with respect to the external field B_0. The rotational fluctuations of the NH bond vector whose orientation at time t can be related to its orientation at time $t+\Delta t$ by an auto-correlation function, and the Fourier transform of the auto-correlation function gives the spectral density function.

The relaxation parameters R_1, R_2 and NOE of the values can be given as linear combination of the spectral density $J(\omega)$ at five characteristic frequencies - $J(0)$, $J(\omega_H-\omega_N)$, $J(\omega_N)$, $J(\omega_H)$ and $J(\omega_H+\omega_N)$ (Abragam, 1961), such that

$$R_1 = d^2\{J(\omega_H-\omega_N) + 3\,J(\omega_N) + 6J(\omega_H+\omega_N)\} + c^2J(\omega_N) \qquad [6.13]$$

$$R_2 = \tfrac{1}{2}d^2\{4J(0) + J(\omega_H-\omega_N) + 3\,J(\omega_N) + 6(J\omega_H) + 6J(\omega_H+\omega_N)\} + 1/6C^2(3J(\omega_N) + 4J(0) + R_{ex} \qquad [6.14]$$

$$NOE = 1 + [\gamma_H/\gamma_N\, d^2\{6J(\omega_H+\omega_N) + J(\omega_N) + 6J(\omega_H-\omega_N)\}/R_1] \qquad [6.15]$$

where ω_H, ω_N are the respective Larmor frequencies of the ^{1}H and ^{15}N nuclei with gyromagnetic rations γ_H and γ_N, μ_0 is the permeability of free space, r_{NH} the internuclear ^{1}H-^{15}N distance (1.02 Å), and $\delta_{\|}$ and $\delta_{\perp}$ the parallel and perpendicular components of the axially symmetric ^{15}N chemical shift tensor whereby $\delta_{\|} - \delta_{\perp} = 160$ ppm. The d which measure the dipolar coupling interaction and c which measure the CSA are given by

$$d=[(\mu_0\gamma_H\gamma_N h)/8\pi^2)]\,\langle 1/r_{NH}^3\rangle \qquad c= (\omega_N/\sqrt{3})(\delta_{\|} - \delta_{\perp}) \qquad [6.16]$$

The spectral density does not contain information on internal motion slower than rotational diffusion, however, systematic increases in R_2 from exchange broadening provide qualitative evidence of motions on slower time-scale and is accounted for by the R_{ex} term in Equation [6.14].

In the equations above, five relaxation parameters are needed to determined the five unknown spectral density function (or six if the R_{ex} is large), however, only 3 (T_1, T_2, NOE) can be determined experimentally in this analysis. This under-determination of the experimental quantities is addressed by a number of approaches, of which the Lipari-Szabo formalism is the most widely used (Lipari & Szabo, 1982a; Lipari & Szabo, 1982b) and one used in this analysis.

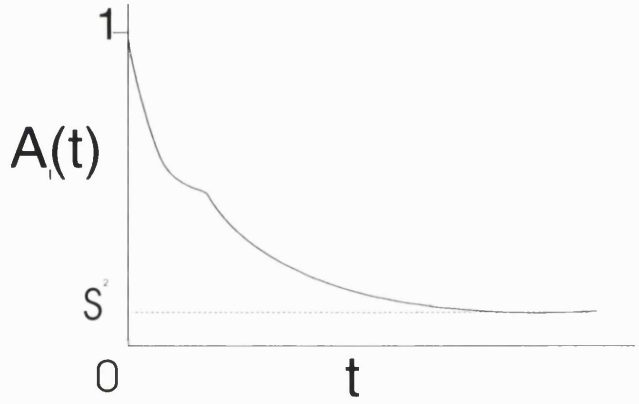


Fig. 6.3 The decay of the auto-correlation function A_I with time.

The Lipari-Szabo formalism assumes that the protein under investigation is tumbling isotropically and the overall motion of the molecule on a nanosecond time scale is superimposed with local motions on the picosecond time-scale. This formalism is so-called 'model-free' in that it gives no specific physical significance to the amplitude of the motion measured by the order parameter and no assumptions about the geometrical aspects for these motions are made. This model describes the internal motion by the square of a generalised order parameter (S^2). The S^2 represents the magnitude of an auto-correlation function for internal motion $A_I(t)$ at $t = \propto$ (Fig. 6.3) in which

$$A_I(t) = S^2 + (1 - S^2)\exp(-t/\tau_e) \qquad [6.16]$$

where τ_e is the effective correlation time of rapid internal motions of the NH bond vector. The S^2 is therefore the residual correlation of the amide bond vector with respect to the molecular frame, and large S^2 implied restricted motion of vector that yield large residual correlation.

The $A_I(t)$ is one of the components of the auto-correlation function for the reorientation of a bond vector $A(t)$ whereby

$$A(t) = A_O(t)\, A_I(t) \qquad [6.17]$$

in which $A_O(t)$ is the auto-correlation component associated with overall motion. It can be related to the molecular correlation time τ_c by

$$A_O(t) = \exp(-t/\tau_c) \qquad [6.18]$$

Combining equations [6.12] and [6.14],

$$A(t) = S^2 \exp(-t/\tau_c) + (1- S^2)\exp(-t/\tau) \qquad [6.19]$$

where $1/\tau = 1/\tau_c + 1/\tau_e$. Fourier cosine transform of [6.15] yield the corresponding spectral density function:

$$J(\omega) = S^2 \tau_c/ (1+\omega^2\tau_c^2) + (1- S^2)\tau/(1+\omega^2\tau^2) \qquad [6.20]$$

Here it can be seen that, if τ_c is known, there are only two unknowns, S^2 and τ_e, and the problem of under-determination is therefore made tractable. The two unknowns can be determined by just two measurements of relaxation parameters, such as T_1 and T_2, while the third parameter can be used as a check on the validity of the model.

It is useful to interpret the S^2 determined from the Lipari-Szabo formalism using the 'diffusion in a cone' model in which the NH bond vector is free to diffuse through polar angle θ', such that $0°<\theta'<\theta$, while the azimuthal angle ϕ is unrestricted, i.e. $0°<\phi<360°$ (Fig. 6.4A). The S^2 can be related to the angle θ by

$$S^2 = [0.5(\cos\theta)(1+\cos\theta)]^2 \qquad [6.21]$$

It can be seen that S^2 will vary from 1 to zero as θ varies from 0° (absence of motion of the NH bond vector, i.e. complete rigidity) to 180° (isotropic motion). A lower value of S^2 therefore indicates increasing mobility. In general, for residues in well-ordered secondary structure, the S^2 value is ~0.85. As can be seen from Fig. 6.4B, S^2 is particularly sensitive to changes for $0° < \theta < 90°$.

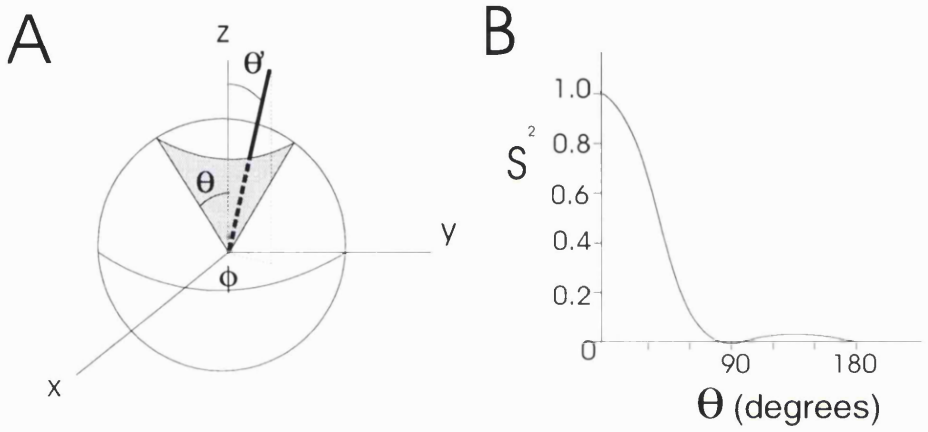


Fig. 6.4 A. Schematic representation of the diffusion in a cone model of internal motion. B. The variation of the square of the generalised order parameter with angle θ of NH bond vector in motion as described by diffusion in a cone.

It was found that, however, for some residues in a number of proteins the three experimental parameters cannot be fitted to this simple analytical dependence of the spectral density function, and additional degrees of freedom was therefore introduced. The model-free formalism was therefore extended to take account of intermediate motion (Clore *et al.*, 1990a; Clore *et al.*, 1990b). In situations where the internal motion occurs on two very different time-scales and one of these lies outside the extreme narrowing range, this approach is applicable when the R_1 and R_2 values are close to the mean value while the measured NOE is anomalous.

$$J(\omega_i) = S_f^2 S_s^2 \tau_c/(1+\omega_i^2\tau_c^2) + S_f^2(1- S_s^2)\tau_s/(1+\omega_i^2\tau_s'^2) \qquad [6.22]$$

where S_f^2 and S_s^2 are order parameters associated with fast and slow internal motions respectively, $\tau_l' = \tau_i\tau_c\ (\tau_c + \tau_i)$ where i = f,s, and τ_s' is the effective correlation time for

slower internal motions on the nanosecond time scale. It should be noted that while further parameters could be introduced to provide perfect fit to the spectral density function, such arbitrary parameters are of little usefulness as they may bear little relation to physical reality (Powers *et al.*, 1992).

A number of other theoretical methods for dynamic analysis of the ^{15}N relaxation data have been proposed, these include the spectral density mapping method (Peng & Wagner, 1992a; Peng & Wagner, 1992b) and its variants such as the reduced spectral density mapping (Farrow *et al.*, 1995; Lefevre *et al.*, 1996) and quasi-spectral density function analysis (Ishima & Nagayama, 1995). In contrast to the Lipari-Szabo model free approach in which the relaxation rates are fitted to the parameters defining the spectral density function as a sum of two or three Lorentzians, no *a priori* assumption of the analytical form of the spectral density function needs to be made in these analyses. Many other approaches to relaxation analysis have since been further developed, however these are not relevant to the studies in this thesis and will be not discussed further.

Diffusion experiment

The self-association or aggregation states of proteins can also be investigated by studying its self-diffusion which is a function of its molecular size (Altieri *et al.*, 1995; Dingley *et al.*, 1995). The translational diffusion constants can be measured in NMR experiment using pulsed field gradient (PFG) techniques (Stejskal & Tanner, 1964), such as PFGLED (pulsed field gradient longitudinal encode-decode (Altieri *et al.*, 1995) or pulsed field gradient longitudinal eddy-current delay (Dingley *et al.*, 1995)). The PFGLED has the advantage that it is fast compared to relaxation analysis, and does not require spectral assignment for the estimation of molecular size. It is therefore useful for the preliminary characterisation of a protein, however, here it was used as a comparison with the results from the relaxation analysis.

The PFGLED diffusion experiment used is based on the water-suppressed LED sequence by (Altieri *et al.*, 1995). An initial 90° - PFG - 90° sequence is used to creates gradient encoded *Z*-magnetisation. The PFG module corresponds to a specific time delay of a few milliseconds during which the PFG is applied. After an

additional period of time delay, a second 90° - PFG sequence then create a gradient-decoded stimulated echo which may then be acquired. Sequences for the suppression of water signal are inserted into this pulse sequence in order to achieve a maximum dynamic range for the measurement of attenuated signal. The pulse sequence was designed to avoid T_2 relaxation effects by keeping the magnetisation on the *z*-axis during the time-delay period and is insensitive to the effect of residual eddy-current which can result in signal distortion.

The echo measured in this experiment and its amplitude are dependent on the extent to which the molecules have diffused during the time delay. The molecules that have diffused across the field gradient will acquire a different Larmor frequency and will not be refocused. The strength of the PFG is gradually increased in a series of experiment and the amplitude of the echo will correspondingly decrease as the molecules that have diffuse through the same distance will now experience greater change in field gradient. The rate of translation diffusion is related to the self-diffusion coefficient D_s in which

$$A(2\tau) = A(0) \exp[-(\gamma\delta G)^2(\Delta-\delta/3)D_s] \qquad [6.23]$$

where $A(2\tau)$ is the amplitude of the echo, γ = ^{1}H gyromagnetic ratio, δ = PFG duration in seconds, G = gradient strength in G/cm, Δ = time interval between PFG pulses and therefore the diffusion time. The intensities of the signal can be extracted from each spectrum and plotted with gradient field strength. The D_s may be obtained by fitting the data to the exponential decay function in Equation [6.23]. For accuracy, the signals should be attenuated ten-fold which can be achieved by adjusting the PFG strength.

Experimental methods

Sample preparation

The CD2d1 protein samples were prepared as previously described in Chapter 2. Unless stated otherwise, all the protein samples used are in the concentration range of 40 μM to 1.2 mM in 20 mM phosphate buffer. The solution pH was adjusted by addition of aliquots of NaOH or HCL to the samples. The pH values quoted are the averages of the pH meter readings before and after each experiment.

NMR spectroscopy

For the relaxation analysis, standard pulse sequences for ^{1}H R_1 and [^{1}H]-^{15}N heteronuclear NOE (Barbato *et al.*, 1992; Kay *et al.*, 1989) were used, while the ^{15}N R_2 was measured using a modified sequence to suppress cross-correlation effects of CSA and dipolar interactions (Kay *et al.*, 1992b). The delays used in the CPMG sequence of the R_2 were set at 0.9 ms. Spectra were acquired with 1024 complex points and a spectral width of 10000 Hz in F2 dimension and 128 complex points and a spectral width of 1600 Hz in F1 dimension. 16 transients were added per increment and phase sensitivity was achieved using States-TPPI (States *et al.*, 1982) phase cycling combined gradient coherence selection sensitivity enhancement (Kay *et al.*, 1992a; Palmer *et al.*, 1992). Series of 2D spectra were recorded with relaxation time delays of 10.05, 60.29, 130.62, 231.09, 341.62, 482.28, 743.52, 1065.04 and 1507.13 ms for T_1, and 0, 16.87, 33.73, 50.59, 67.45, 84.32, 118.05 and 151.78 ms for T_2. Recycle delays were 2s in the T_2 and T_1 experiments. To assess the reproducibility of the experiment, the measurement for a relaxation time at 482.28 ms was repeated in the T_1 series, and 33.73 ms in the T_2 series. The heteronuclear NOE was measured by recording ^{1}H-^{15}N HSQC experiments with and without proton saturation for a duration of 3 s in a total recycle delay of 3.6 s between scans. Each of these two experiments was recorded twice in order to estimate the error. Noise levels were estimated from the root mean square (rms) noise in regions of the spectra void of signals. All the spectra for the relaxation experiments were recorded at a ^{1}H

frequency of 500 MHz on a Varian UnityPlus spectrometer with three radio frequency channels and triple resonance Z-axis pulsed field gradient probe.

Relaxation data were obtained for CD2d1 at six different protein concentrations between 1.2 mM and 40 μM. Data were first recorded at a sample concentration of 1.2 mM, which is subsequently diluted with appropriate volumes of sample buffer for measurement at lower concentrations. Protein concentrations were checked at each data point by UV absorbance at 280 nm. The number of transients were increase from 16 at a CD2d1 concentration of 1.2 mM to 128 at a concentration of 40 μM. Control experiments were performed at pH 6 using protein samples at concentration of 1.2 mM in 10 mM Bis-Tris/HCl buffer, pH 6.0, as well as in higher salt concentration of 100mM and 300 mM NaCl in 20 mM phosphate buffer.

All spectra were processed with nmrPipe (Delaglio *et al.*, 1995). FIDs were multiplied with a Gaussian window function in t_2 with a line broadening of 2.0 Hz and a weighting factor of 0.01 while a square sine function shifted by 80° was used in t_1 . No zero filling was applied in t_2 prior to Fourier transformation while FIDs in t_1 were first extended by 32 complex points (from 128 measured) by linear prediction using 8 coefficients followed by zero filling to 512 points. A second order polynomial baseline correction was performed both in t_1 and t_2. Spectra were analysed with ANSIG v3.3 (Kraulis, 1989), AZARA (Boucher, 1993) and XEASY (Bartels *et al.*, 1995).

For the diffusion experiments, the pulse sequence used is essentially as previously described (Altieri *et al.*, 1995; Dingley *et al.*, 1995). The duration δ of the encode and decode PFG pulsed was 10 ms. The diffusion delay time Δ used is 60 ms. The gradient strength (G) used was incremented in the range of 0.91 - 27.55 G/cm such that the spectra recorded ranged from the maximum down to close to zero. 30 1D spectra were recorded per data point. The data were processed using Felix version 2.3 (BIOSYM Inc. San Diego, CA) and the analysis and curve-fitting done using Prophet (Bolt Beranek and Newman Inc.) and Kaleidograph (Synergy Software).

Data Analysis

Relaxation analysis

The ^{15}N relaxation data were analysed according to the Lipari-Szabo model-free formalism (Lipari & Szabo, 1982a; Lipari & Szabo, 1982b) with its extension to motions on an intermediate time scale (Clore *et al.*, 1990b) in a manner similar to that described by Palmer and co-worker (Mandel *et al.*, 1995). Molecular rotational correlation time τ_c was calculated from a subset of residues that do not exhibit large amplitude local motions from the mean value of the R_2/R_1 ratios. These residues are selected according to the following criteria: $|(|R_2 - \langle R_2\rangle| / \langle R_2\rangle)-(|R_1 - \langle R_1\rangle|/\langle R_1\rangle)|$ >1.5 SD and only heteronuclear NOE values >0.65 are selected. All numerical analysis was performed in Mathematica (Wolfram Research) using a Levenberg-Marquardt non-linear least-squares algorithm.

Monomer-dimer equilibria

For the analysis of the concentration dependent effect of rotational correlation times, chemical shift variations and exchange broadening of resonance line, a model of equilibrium between monomeric CD2d1 and a symmetric CD2d1 homodimer is used:

$$A_2 \underset{k_{on}}{\overset{k_{off}}{\rightleftharpoons}} 2\,A \qquad [6.24]$$

The dissociation constant K_D can be expressed as follow:

$$K_D = \frac{k_{off}}{k_{on}} = \frac{[A_2]}{[A]^2} \qquad [6.25]$$

For the concentration-dependent chemical shift changes of CD2d1, the observed chemical shifts can be treated as population weighted averages of the free monomer and intact dimer values within the range of CD2d1 concentration used. The difference in chemical shift $\Delta\delta = |\delta(\text{dimer})-\delta(\text{monomer})|$ can therefore be used to

obtain the value of K_D by simple non-linear least square fit of the data for each residue.

The variation in relaxation rates at different protein concentrations was used to estimate a limiting value for the rotational correlation time of the monomeric CD2d1 at infinite dilution and the putative intact CD2d1 homodimer at high concentration. For this estimation, the appropriate averaging of observable values need to be considered. The apparent τ_c, however, is not a linear function of R_1 and R_2 decay rates or the R_2/R_1 ratio, the apparent τ_c at a given concentration is therefore not a simple weighted average of the monomer and dimer values. The analysis of τ_c as a function of concentration is therefore performed using a non-linear fit of measured and predicted R_2/R_1 ratios, with τ_c (monomer) and τ_c (dimer) set as the variables to be optimised, while the population weighting based upon K_D was also set as a further variable to be optimised (Grzesiek *et al.*, 1997).

In the case of interconverting monomer and dimer species, the observed R_1 and R_2 decay rates are expected to be biexponential. Fitting of such biexponential decay to the single exponential usually employed in the analysis of ^{15}N relaxation would be expected to degrade the quality of the fit and lead to bias in favour of the slower relaxation rate. However, in the case of dimerisation of molecules below 50 kDa the difference in the relaxation rates is too small to produce any significant error. Furthermore, as was noted recently (Fushman *et al.*, 1997), the averaging of monomer and dimer relaxation often occurs on a time scale much faster than the duration of the NMR experiment for the relaxation rates to be measured. Consequently, nuclear relaxation experiments can produce monotonic exponential decays that is characterised by population weighted relaxation rates.

Exchange line broadening was estimated using the following equation:

$$R_{ex} = (\Delta\omega)^2 \cdot p_A p_{A2} \cdot k_{ex}^{-1} \cdot [1-(2/k_{ex}\tau_{CPMG}) \cdot \tanh(k_{ex}\tau_{CPMG}/2) \qquad [6.26]$$

where p_A and p_{A2} are the mole fractions of monomeric and dimeric CD2d1, $\Delta\omega$ is the difference in chemical shift for a particular resonance between monomeric and dimeric state, and τ_{CPMG} is the CPMG refocusing period used in the R_1 pulse sequence. The values for $p_A p_{A2}$ as a function of concentration were calculated using

the K_D values obtained from the fitting of the chemical shifts as described earlier. The values of R_{ex} at different concentrations can then be used to obtained the value of $\Delta\omega$, the rate of association k_{on}, and the rate of dissociation k_{off}, by using equation [6.26] and the following relations:

$$k_{ex}^{-1} = \tau_{ex} = (\tau_A \cdot \tau_{A2}/(\tau_A + \tau_{A2}))$$

where $\tau_A = k_{on}^{-1} \cdot [A]^{-1}$ and $\tau_{A2} = k_{off}^{-1}$ [6.27]

PFGLED data analysis

For the diffusion data, a cluster of peaks in a range of chemical shifts within which the solvent peaks are absent in the 1D spectrum is selected. The integrated intensities of these peaks are then extracted from the series of 1D spectra obtained in the PFGLED experiments. The intensities can then be plotted against the gradients strength levels, from which the values of the diffusion coefficient D_s are obtained by a least squares fit according to equation [6.23]. The D_s can then be used to compute the apparent molecular weight which is proportional to ($1/ D_s^3$) (Dingley *et al.*, 1995):

$$M_r = \left(\frac{kT}{6\pi\eta F D_s}\right)^3 \left(\frac{4\pi N_A}{3[\bar{v}_2 + \delta_1 \bar{v}_1]}\right) \qquad [6.28]$$

where k is the Boltzmann constant (JK^{-1}), T is the temperature (K), η is the viscosity of the solution (assumed to be 9.2 x 10^{-4} kg m^{-1} s^{-1}), D_s is the translational diffusion coefficient (m^2s^{-1}), N_A is the Avagadro's number (mol^{-1}), δ_1 is the fractional amount of water bound to the molecule (set to be 0.34 g H_2O per g diffusant), v_1 (at 1 x 10^{-6} m^3Kg^{-1}) and v_2 (at 0.73 x 10^{-4} m^3Kg^{-1}) the partial specific volume of solvent water and the molecule respectively. The dimensionless Perrin factor, F, which relates the shape of the molecule, is set to be equal to 1.

Results

Analysis of the Backbone Dynamics of CD2d1

The ^{15}N relaxation time measurements and molecular dynamic analyses were performed on CD2d1 at a range of protein concentration and pH values. As shall be demonstrated, the analysis of the relaxation data revealed a distinct concentration and pH dependence in the dynamic properties of CD2d1. An initial description of the results is therefore based on experiments performed at a protein concentration of 1.2 mM, pH 6 in 20 mM phosphate to which results from other experiments at varying conditions will be compared.

CD2d1 has 99 backbone amide NH sites; apart from the N-terminal residues, all the resonances are observable at most pH. Most of the resonances are well resolved in the ^{1}H-^{15}N HSQC with a few such as Val6, Lys64, Arg89, and Leu95 showing considerable overlap. These overlapped residues are therefore not used in the analysis. A few very weak peaks such as Leu89 that may introduce greater error in the analysis are also excluded. The results of the various residue-specific relaxation parameters plotted against residue numbers are shown in Fig. 6.5.

The analysis yielded fairly uniform relaxation parameters for most of the residues in CD2d1, with the exception of the four residues at the N-terminus and the C-terminus residue Glu99. The average values for R_1 is 1.92 ± 0.08 s^{-1}, for R_2 10.00 ± 1.38 s^{-1}, and for heteronuclear NOE 0.71 ± 0.04. From the average R_2/R_1, the apparent molecular correlation time τ_c is calculated to be 7.52 ± 0.05 ns. For most residues, the square of the order parameter S^2 is greater than 0.6, with a mean S^2 values of 0.81 which accords with values reported for residues in highly ordered regions of protein structure (Kay *et al.*, 1989). The S^2 values are lower (0.6 - 0.8) for residues situated at the loops connecting the β-strands such as the BC loop (residues Asn20-Asp25), CC' loop (Gly35-Thr37), C'C'' loop (Lys45-Lys47), and the hairpin-turn connecting D and E strands (Lue58-Asn60), indicating greater mobility for these segments of polypeptide chain with larger amplitude motion. CD2d1 therefore can be considered to have a fairly rigid overall structure with flexible interconnecting loops. The derived estimates of the internal motion correlation times are uniformly

small (<50ps), perhaps indicating fast thermal vibrations. For the first eight residues of CD2d1, intermediate-fast time-scale dynamics has to be assumed. Here the local correlation times have values around 0.8-1.5 ns and the order parameters for the fast time-scale motion $S_f^2 \approx 0.7$.

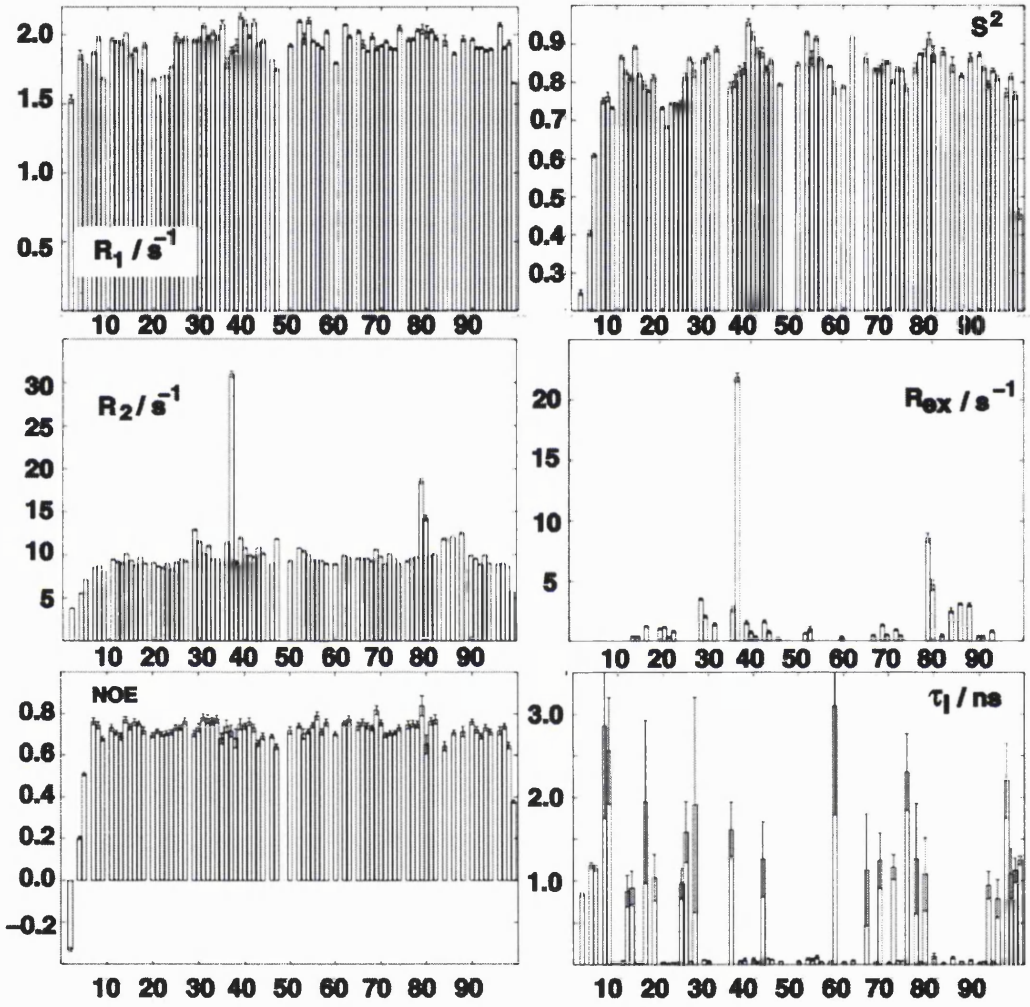


Fig. 6.5 Results of the Lipari-Szabo model free analysis for CD2d1 at 20 mM phosphate buffer pH6.0 at protein concentration of 1.2 mM at 25°C. On the left, from top to bottom: ^{15}N relaxation parameters R_1, R_2, and heteronuclear NOE. On the right are the major derived parameters of the Lipari-Szabo model-free analysis; from top to bottom: S^2, R_{ex} and local correlation time. The local correlation time is the τ_e in the ps range for a Lipari-Szabo model, but if an extended Lipari-Szabo model was employed, a very short value for τ_e was assumed and τ_i , which is in the ns range, is displayed.

The most prominent feature of the relaxation data is the very large deviations in the R_2 (Fig. 6.5) observed for a number of residue, in particular Thr37 and Thr79, while their values for R_1 and heteronuclear NOE is close to the average. In the

Lipari-Szabo analysis, these extra contribution to R_2 can be presented as the exchange term R_{ex} (Fig.6.6). The value of R_{ex} for Thr37 is 22 s^{-1}, for Thr79 it is 8 s^{-1}. A number of other residues have smaller but significant R_{ex} values ($R_{ex} > 1.5$ s^{-1}), these include Glu29, Val30, Ser36, Val39, Lys43, Val80, Asn84, Thr86 and Leu88. All these residues lie on the GFCC'C'' side which forms the binding surface in CD2d1.

Effect of protein concentration

The set of relaxation experiments was repeated for CD2d1 at a range of protein concentration from 40 μM to 1.2 mM. However, no NOE measurements were performed at 0.1 mM and 40 μM due to the low inherent sensitivity of the NOE experiments. NOE values were taken from measurement at 1.2 to 0.2 mM to perform the complete Lipari-Szabo analysis. In this analysis, no detectable trend in the variation of the NOE values over the entire concentration range can be discerned.

The majority of the residue-specific parameters are not significantly affected by the variation in protein concentration. For most residues, the value of S^2 is slightly greater at a higher concentration, and the maximal differences over the entire concentration range are around ±0.04, values that are within the order of magnitude of uncertainty of the derived values. Similarly, variations in values for the fast local motions as well as the time constants for motions on an intermediate time scale are also within the error limits.

The most significant change observed in this dilution series can be seen in Fig. 6.6 which compare the R_{ex} terms obtained at protein concentration of 1.2 mM to those obtained at 0.1 mM, the protein concentration at which the full set of relaxation parameters can still be measured. The R_{ex} terms for Thr37 and Thr79, as well as for some residues in region around Glu29-Trp32 and Thr79-Thr86, all exhibited pronounced effect from the changes in protein concentration. There is a systematic drop in the values for R_{ex} with decreasing concentration, and for some, the exchange contribution is absent at the lower concentration. At the lowest protein concentration (40 μM) for which R_2 is measured, the relaxation data for a number of residues (e.g. Val30, Trp32, Ser36, Val39, Phe42) which previously required an R_{ex} term in their analyses, can be analysed without recourse to this exchange contribution.

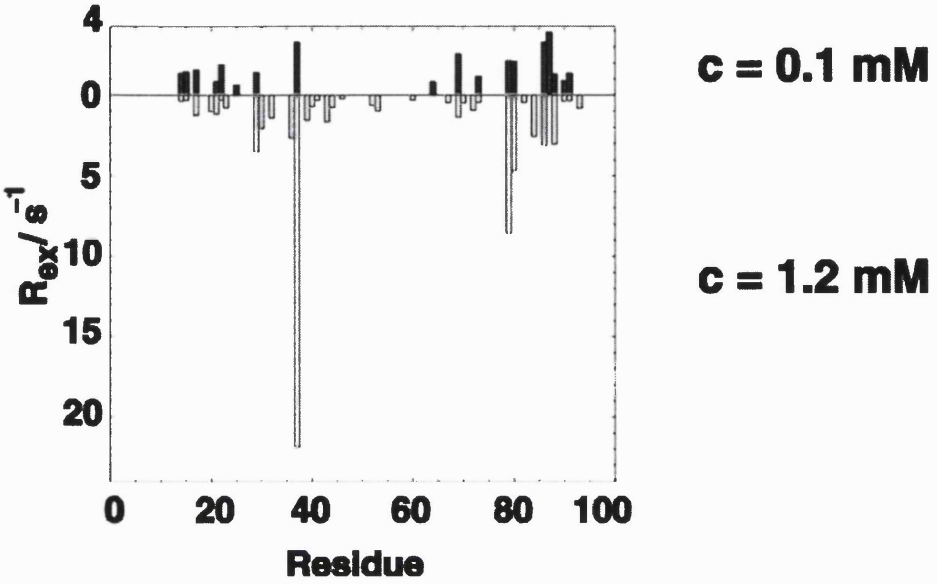


Fig. 6.6 Comparison of the exchange contribution R_{ex} for residues along the polypeptide chain of CD2d1 at 20 mM phosphate concentration, pH 6.0 at 25°C and proton resonance frequency of 500 MHz. The results obtained from proteins concentration at 0.1 mM is shown in black bars on top, while those from 1.2 mM concentration is shown at the bottom in grey bars.

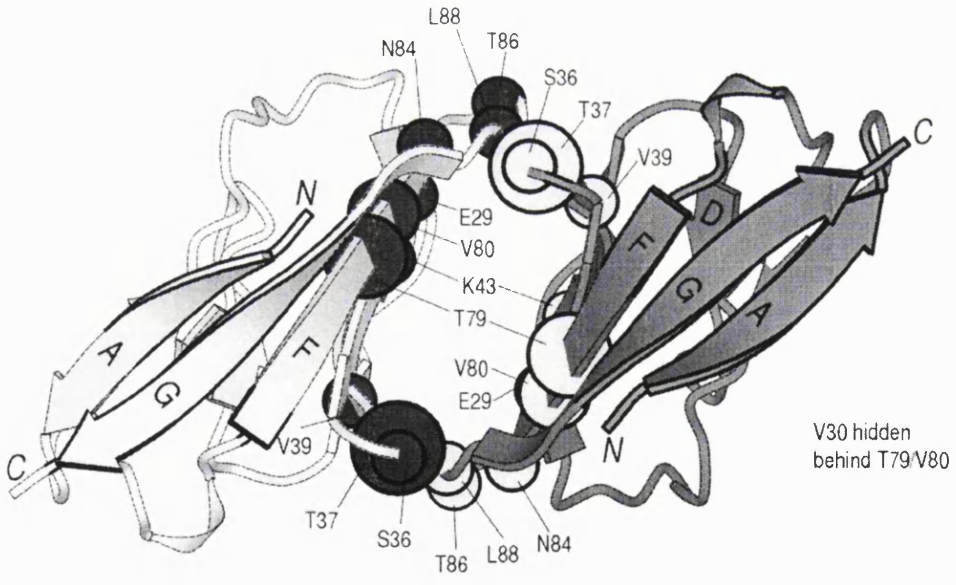


Fig. 6.7 Figure showing the location of residues with large R_{ex} on the CD2 crystal dimer. Residues with R_{ex} >1.5 s^{-1} are indicated by a sphere centred on its backbone nitrogen. The radius of the sphere is approximately proportional to the magnitude of R_{ex}.

The residues whose R_{ex} terms showed concentration dependent are indicated in the dimeric crystal structure of CD2d1 as shown in Fig. 6.7 (Jones *et al.*, 1992). It is clear from the figure that these residues are directly involve in the crystal

contact, and these concentration effects are therefore indicative of a dimerisation of CD2d1 in the specific manner seen in crystal structure. The large R_{ex} terms observed at higher protein concentration, which is commonly associated with conformational exchange, is actually the product of dimerisation. The fact that only a few residues are affected rules out the effect of bulk solvent properties, such as macroscopic sample viscosity as a cause of the observed exchange. A number of residues such as Thr69, Thr86, Leu88, however, have consistent values of ~2 s^{-1} throughout the dilution series. This would indicate that the R_{ex} values for these residue obtained are due to real conformational exchange rather than self-association.

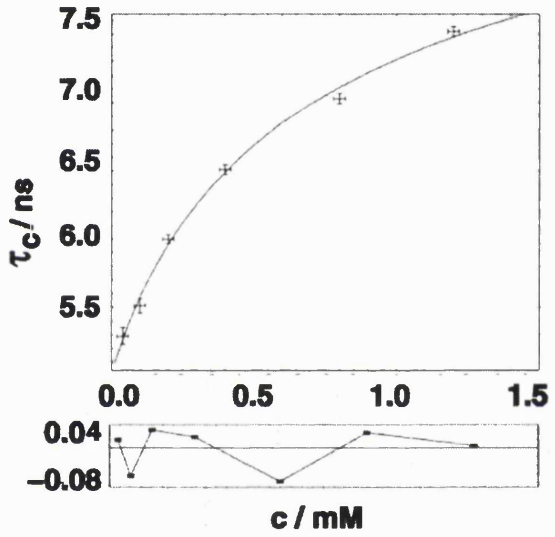


Fig. 6.8 Dependence on protein concentration of the apparent rotational correlation time τ_c of rat CD2d1 in 20 mM phosphate buffer, pH =6.0 and T = 25°C. The upper panel shows the fit of τ_c to a monomer-dimer equilibrium using equation [6.25]. The lower panel shows the residual from the fit.

This indication of the self-association is confirmed by the determination of the apparent rotational correlation time τ_c at various protein concentrations. The evaluation of the τ_c values is widely used to assess protein self-association (Fushman *et al.*, 1997; Gryk *et al.*, 1998; Grzesiek *et al.*, 1997), and assuming all else being equal, the rotational correlation time can be regarded as being roughly proportional to the relative molecular mass. For CD2d1, the τ_c is found to be strongly concentration dependent, and the variation of the τ_c at different concentrations is shown in Fig. 6.8. This variation in τ_c ranges over 2 ns for protein concentrations between 40 μM and 1.2 mM. The increase in τ_c at higher protein concentration therefore corresponds to

an apparent increase in relative molecular mass, which is a reflection of an increase in self-association of the protein.

A fit of the variations in τ_c values with concentration according to monomer-dimer equilibrium using Equation [6.25] yield and extrapolated values for CD2d1 monomer and dimer at τ_c of 5.02 ± 0.1 ns and 10.09 ± 0.5 ns respectively. These values agree very well with those calculated from the simple Stokes-Einstein relation (Equation 6.12]) which yield a theoretical values of 4.9 and 9.8 ns for monomer and dimer respectively. In this calculation of theoretical value, the fractional hydration of 0.32 was used, and the total molecular mass of the dimer, including the hydration layer, of the dimer is twice that of the monomer.

Apart from the changes in the R_{ex} exchange term and τ_c, the effect of the protein concentration can also be seen in the changes in chemical shifts of some of the peaks in the ^{1}H-^{15}N HSQC spectra. A number of residues showed significant ^{15}N chemical shift variation of up to 0.5 ppm over the concentration range of 40 μM to 1.2 mM. The majority of these residues were among those that exhibit considerable exchange broadening at the higher concentration of CD2d1. Residues with the largest values in chemical shift variation are sited on two regions of the polypeptide chain, around position 28-46 and 78-90. Particular large variations are seen for Thr37 and Thr79 (Fig. 6.9), which directly reflect the large exchange contribution observed for these residues.

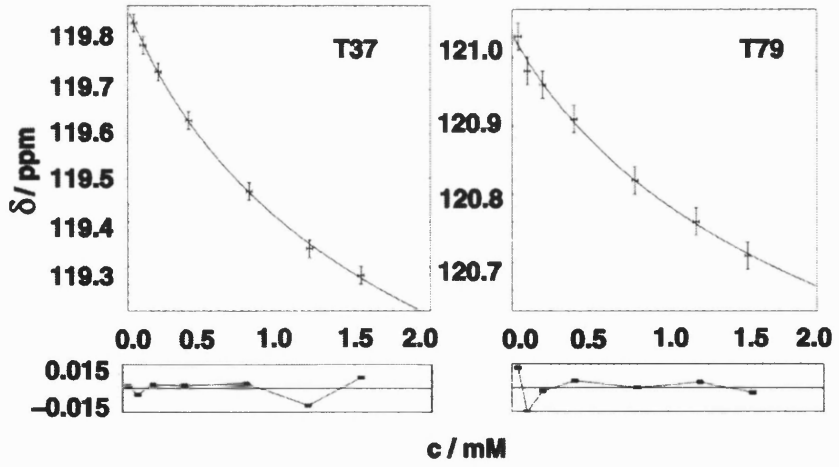


Fig. 6.9 Variation of the ^{15}N chemical shifts with CD2d1 concentration for Thr37 and Thr79 at 20 mM phosphate buffer, pH 6.0 and T = 25°C. The upper panels show the experimental data with a fit to a function representing the chemical shift as a population average of monomer and dimer values. The lower panels display the residuals from the fit.

The concentration dependent effect is also evident in the ^{19}F spectra of fluorine-labelled CD2d1 (Fig. 6.10). Exchange broadening as well as chemical shift changes of the peak from Phe49 can be observed. The peak height of Phe49 decreases by around a third from 1.2 to 0.2 mM protein concentration. In principle, it should be possible to use the exchange broadening of this residue for the analysis of the monomer-dimer equilibria. The fluorine spectra, however, suffer from severe rolling baseline at low protein concentrations and determination of peak height was therefore difficult a concentration lower than 0.2 mM. The data are therefore not used for further analysis.

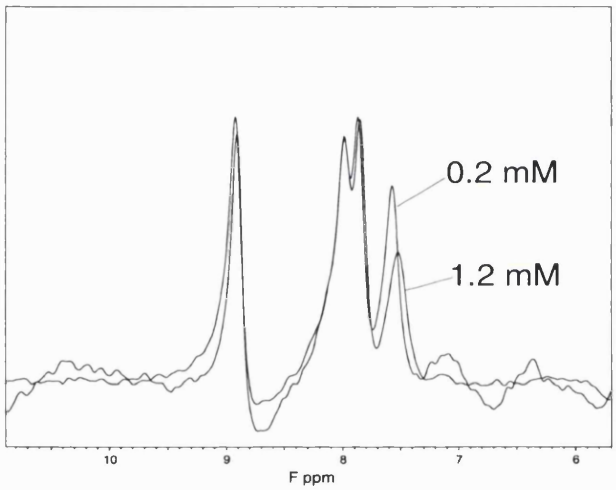


Fig. 6.10 The 1D spectrum of ^{19}F-labelled CD2d1 at protein concentration of 0.2 and 1.2 mM showing slight shift as well as broadening of the Phe49 resonance at higher protein concentration.

Determination of kinetic constants

The variations of the R_{ex} term, τ_c, and the chemical shift with the protein concentration can be used to obtain the kinetic parameters of the monomer-dimer equilibria of CD2d1. The exceptionally large R_{ex} terms and chemical shift changes from Thr37 and Thr79 are particularly ideal for such analysis. In Fig. 6.11, the R_{ex} terms of Thr37 and Thr79 are plotted against protein concentration.

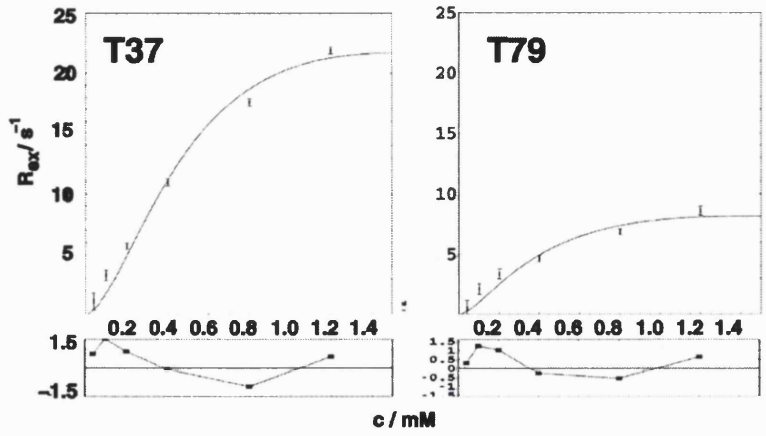


Fig 6.11 Variation of ^{15}N line-broadening as a function of protein concentration for Thr37 (left) and Thr79 (right) at 20 mM phosphate buffer, pH = 6.0 and T = 25°C. The upper panels show the experimental data and fit to the function defined by Equation [6.26]. The lower panels display the residuals from the fit.

In this analysis, the dissociation constant K_D is first obtained from the chemical shift changes of Thr37 and Thr79 (Fig.6.9). Caution was taken to maintain stable sample buffer conditions (concentration and pH) over the dilution range of 1.5 to 40 μM as the chemical shift can change significantly with a slight shift in pH. The dissociation constant K_D values estimated from the curve-fitting are 4.6±1.8 mM and 5.4±2.2 mM for Thr37 and Thr79 respectively. These estimated values of K_D were then used for curve-fitting of the R_{ex} data for the corresponding residue to a monomer-dimer equilibrium as described by Equation [6.26] in order to determine the fractional amounts of monomer and dimer. The estimated K_D is also used to relate k_{on} and k_{off} according to Equation [6.25]. This reduces the number of parameters that are optimised in the fitting procedure to $\Delta\omega$ and k_{on}, the latter in the form of τ_{ex} through Equation [6.27]. For Thr37, the curve-fitting yields the values for $\Delta\omega/2\pi = 130 \pm 5$ Hz (~2.6 ppm), $k_{on} = 4200 \pm 600$ $M^{-1}s^{-1}$ and $k_{off} = 9 \pm 2$ s^{-1}. The corresponding values for Thr79 are $\Delta\omega/2\pi = 79 \pm 6$ Hz (~1.6 ppm), $k_{on} = 4500 \pm 1200$ $M^{-1}s^{-1}$ and $k_{off} = 5 \pm 2$ s^{-1}.

For the chemical shift changes, curve-fitting to Equation [6.26] yields an extrapolated CD2d1 dimer-monomer chemical shift differences of $\Delta\delta = 1.97 \pm 0.11$ ppm for Thr37 and $\Delta\delta = 1.12 \pm 0.2$ ppm for Thr97. These values are in good

agreement with the values extracted from the fitting of the concentration dependence of R_{ex} terms which are $\Delta\delta = 2.6$ and 1.6 for Thr37 and Thr79 respectively.

The dissociation constant K_D can also be obtained from a fit using the variation of the rotational correlation time τ_c with concentration (Fig. 6.8). The value obtained from this fit is 2.6 ± 0.7 mM. The possible value for K_D therefore can range from ~2 to 7 mM. An analysis using ultracentrifugation was undertaken with a protein concentration of 0.1 mM in 20 mM phosphate at pH 6.0. As can be seen in Fig. 6.12, the protein is essentially monomeric at this concentration, with no dimer formation detected. This is further indication that K_D must lie in the mM range. Ideally the experiment should be repeated at a higher protein concentration so that the K_D may be determined by another biophysical method apart from NMR spectroscopy, however, due to insufficient time available for such study this was not pursued further.

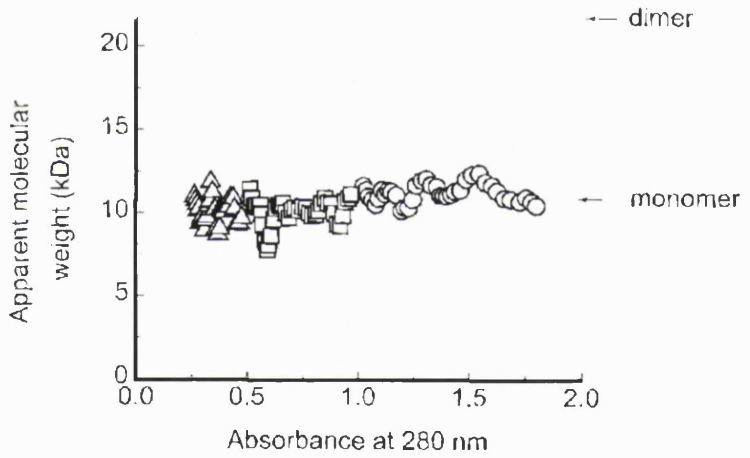


Fig. 6.12 Analytical ultracentrifugation data for a protein sample a concentration of 0.1 mM in 20 mM phosphate buffer, pH 6.0. The relative molecular mass of the monomer and dimer are indicated.

Effect of buffering condition

In order to exclude any spurious solvent or specific ion effects as a source for the observed dynamics phenomena, a number of control relaxation experiments were performed. The relaxation experiments were repeated at pH 6.0 using 1.2 mM protein concentration at 10 mM Bis-tris buffer, as well as in 20 mM phosphate buffer with 150 mM and 300 mM NaCl added.

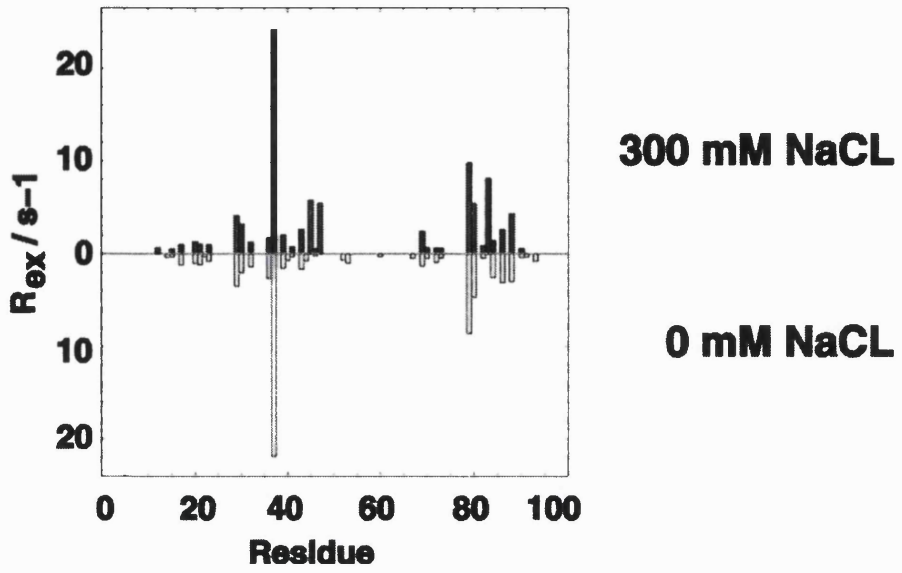


Fig. 6.13 The effect of salt on the exchange contribution on R_{ex} along polypeptide chain of rat CD2d1 determined at 20 mM phosphate buffer, pH = 6.0 and T = 25 °C and a proton resonance frequency of 500 MHz. Results obtained from sample with no added salt is shown in grey bars while those from sample with 300 mM added NaCl is shown in black bars.

On the residue-specific level, only small variations were seen in all the different buffering condition. The order parameter S^2 were largest in the sample without added salt and dropped only slightly (~0.04) above 150 mM NaCl (Fig.6.13). However, these changes were small and within the margin of experimental uncertainty for the majority of residues. Larger changes in the S^2 values (~0.06) were seen only for three highly mobile N-terminal residues. Internal motion correlation times were uniformly smaller than 80 ps for fast motions and between 1.0 and 4.0 ns for intermediate time scale motions. No consistent trend in the variation of the internal motion correlation value τ_e with salt concentration could be identified. Residues with substantial R_{ex} are essentially unaffected by the changes in ionic strength as shown in Figure 6.13. A slight increase in R_{ex} values is observed for some residues between Lys43 and Lys47 and Asn77-Ile88 at 300 mM NaCl with three residues (Lys45, Lys47, Thr83) exhibited more elevated values with $R_{ex} > 5\ s^{-1}$. This suggests that at these loop regions, electrostatic interaction may help to restrict the movement of the loops; and with increasing salt concentration, these electrostatic interactions are screened, resulting in greater movement of the loops. The residue-specific relaxation parameters measured in Bis-Tris/HCl buffer are identical within the error margin to those measured in phosphate buffer. For the τ_c, the values

measured at the different buffering concentrations are shown in Table 6.1. The variations of τ_c in all these different conditions are only marginally greater than the measurement error.

Table 6.1 *Rotation correlation time of CD2d1and 15N exchange line broadening of Thr37 in different buffering conditions at pH 6.0, protein concentration at 1.2 mM and T=25°C.*

Buffer condition	τ_c/ns	**Thr37** $R_{ex}/(s^{-1})$
20 mM phosphate	7.52 ± 0.04	21.9 ± 0.3
20 mM phosphate, 150 mM NaCl	7.67 ± 0.05	27.1 ± 0.8
20 mM phosphate, 300 mM NaCl	7.71 ± 0.06	24.1 ± 2.0
10 mM Bis-Tris/HCl	7.81 ± 0.05	23.2 ± 0.2

Effect of pH

With the confirmation of the low-affinity dimerisation of CD2d1, a further series of experiments was conducted in order to assess the effect of pH on the association states of CD2d1. Relaxation data were acquired at 7 different pH points - 2, 3, 4, 5, 6, 7 and 8. The measured relaxation rates and the results from the Lipari-Szabo analysis from experiment determined at pH 3.0 are shown in details in Fig 6. 14 and the variations in R_2 at varying pH are shown in Fig. 6.15.

There is a systematic decrease in the R_1 and R_2 measured when the pH is lowered from pH 6.0 for all the residues (Fig. 6.15). Variation is also seen in the R_{ex} exchange terms in manner similar to that seen in the dilution series, with those residues exhibiting large R_{ex} terms showing systematic decrease in values with decreasing pH. Again the exceptions to this decrease in R_{ex} terms are residues that are situated in and near the FG loop. Note that in Fig 6.14, the figure showing R_2 values shows a residue with an exceptionally large R_2 term. This residue Leu89, which is situated at the end of the FG loop, is not used in the analysis as it is weak and therefore may introduce error into the analysis. Nevertheless, it showed consistently exceptionally large R_2 value at all pH and protein concentration. This may indicate that Leu89 is one of the high flexibility points at the end of the loop

with the rest of the loop being relatively well-structure (Peng & Wagner, 1992b). It is also useful to note that while the R_{ex} value of T37 drops to normal value at pH 3.0, T79, whose R_{ex} values also decline with pH, still showed an elevated value at pH 3.0. Most of the residues that do not showed concentration-dependent variation in the R_{ex} terms also do not showed significant variations with pH, suggesting that the FG loop, together with residues close by, is exhibiting true conformational exchange.

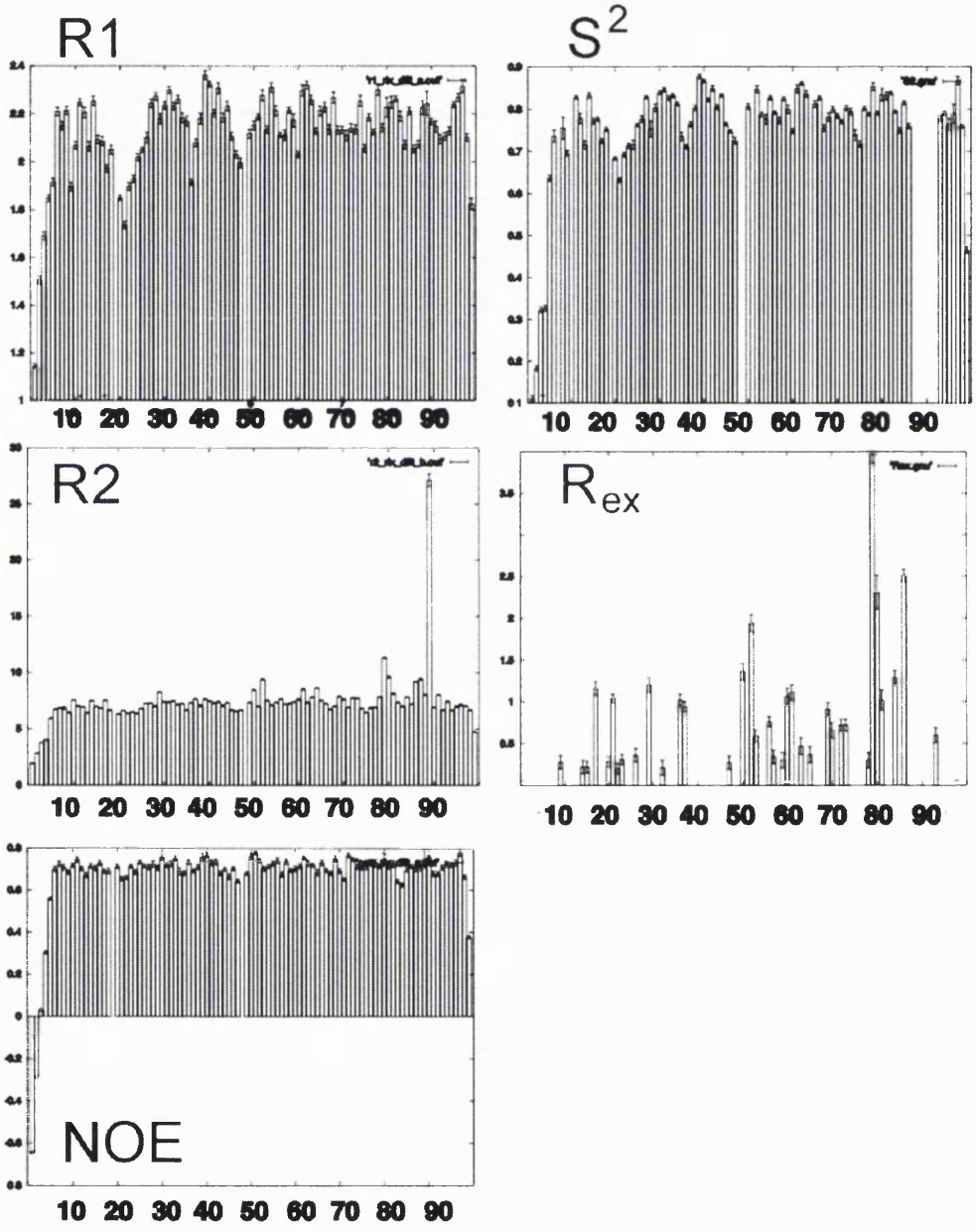


Fig. 6.14 Results of the Lipari-Szabo model free analysis for CD2d1 at pH 3.0 in 20 mM phosphate buffer with protein concentration of 1.2 mM at 25°C. On the left, from top to bottom: ^{15}N relaxation parameters R_1, R_2, and heteronuclear NOE. On the right are the major derived parameters of the Lipari-Szabo model-free analysis; from top to bottom: S2 and R_{ex}.

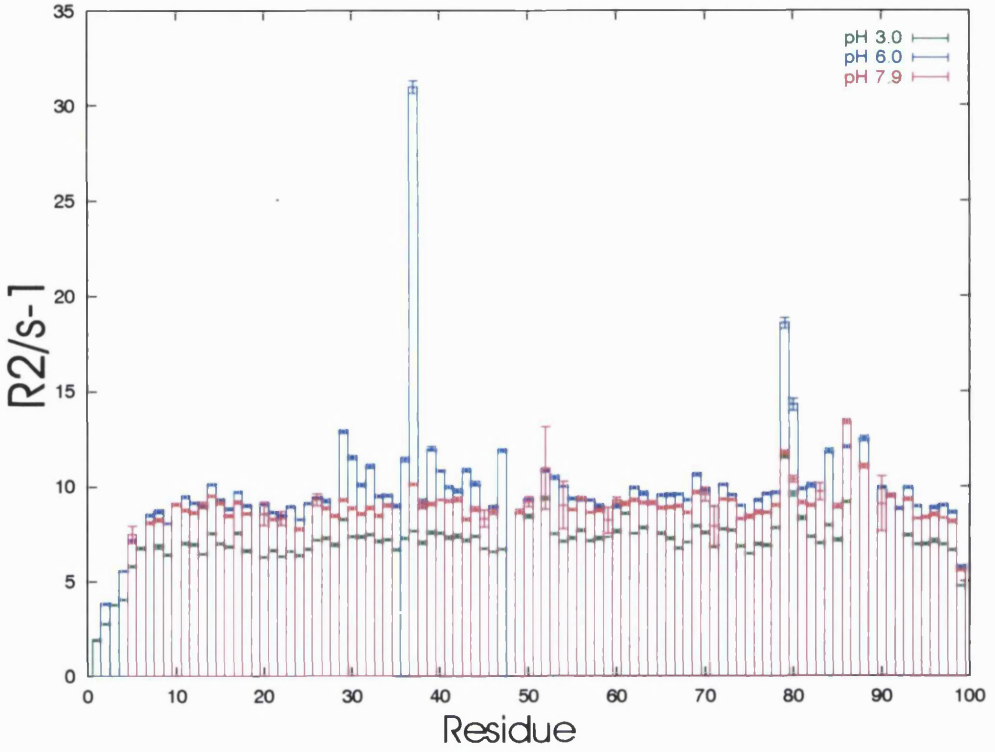


Fig. 6.15. The variation of R_2 values with pH. A total of 7 sets of data were acquired from pH 2 to 8.5, but only 3 are shown here for clarity. Maximal R_2 values are observed for most residues at pH 6 while the minimum is seen at pH 3.

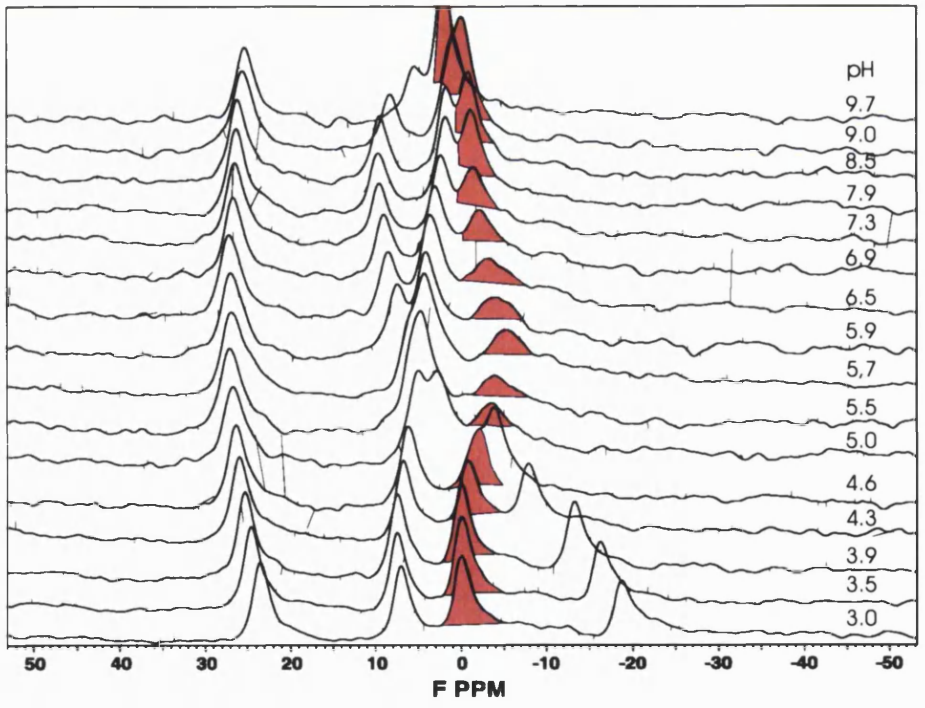


Fig. 6.16 pH dependent behaviour of the ^{19}F 1D spectra of CD2d1. The resonances marked in red are from Phe49. The experiments were performed at a protein concentration of 1.2 mM and T = 25 °C.

The pH dependent exchange phenomenon as seen in R_2 relaxation parameters can also be observed in the ^{19}F NMR spectrum of fluorine-labelled CD2d1 (Fig 6.16). For Phe49 which is directly situated on the binding surface and may be involved in the binding interaction, pronounced exchange broadening of its resonance can be seen. Such pH dependent exchange broadening is not seen for the other resonances including Phe42, suggesting that the exchange broadening effect is the direct result of interaction by Phe49 on CD2 dimerisation.

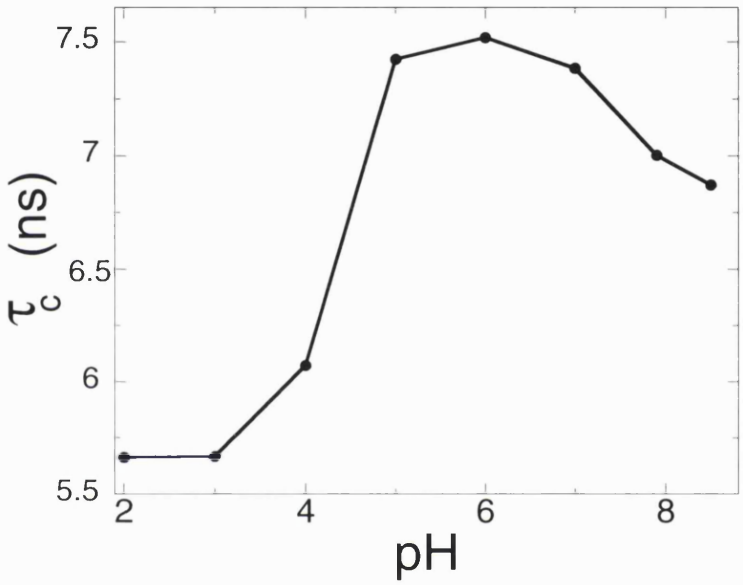


Fig. 6.17 Variation of the rotational correlation time τ_c with pH.

The correlation time τ_c at each pH point is determined from the relaxation parameters and the variations with pH in the τ_c calculated from R_2/R_1 ratio is shown in Fig 6.17. As can be seen, there is a distinct pH dependence for CD2d1 self-association, with a maximal value observed to be around pH 6.0. At pH 2.0 and pH 3.0, the values of τ_c is essentially unchanged ~5.7 ns, which is comparable to the extrapolated τ_c values for a monomer of 5.02 ± 0.1 ns, indicating that at the low pH, the CD2 can be regarded as essentially monomeric. At higher pH, the τ_c showed a gradual decrease, however, at pH 8.0, there is still some degree of self-association. When taken together with the result from the pH titration studies in which the pK_a of Glu41 is determined to be 6.7, this result strongly suggests that the maximal self-association for CD2 occurs when Glu41 is protonated. The presence of some degree of CD2d1 self-association at high pH, however, showed that the presence of the

charge on Glu41, while not ideal for CD2 self-association, may permit some degree of self-association.

The pH-dependent effect on CD2d1 self-association is also investigated using PFGLED experiment as self-diffusion can be used to assess the association states of protein (Nesmelova & Fedotov, 1998). The result of the diffusion experiment, as shown in Fig. 6.18, agrees well with that obtained from the relaxation in a general manner with similar pH dependent profile of CD2 self-association. The maximal increase in apparent M_r is at ~ pH 6 and the minimum value is obtained at pH values less than 4. Note that the relative molecular mass M_r obtained does not correlated with the expected value and cannot be regarded as reliable as its determination is prone to error. This is because some of the assumptions made in the calculations using Equation [6.28] may not be valid. For example, the water solvation δ_1 of the protein in not known *a priori*, and the Perrin (F) factor, if erroneous, can have a large effect on the calculated molecular weight as it has a inverse cubic relationship with M_r. The CD2d1 was assumed to be spherical while in reality the actual shape of CD2d1 is oblate. This method is therefore useful only as a comparative analysis and no significance could be attached to the calculated values unless the various parameters are also determined experimentally (Dingley *et al.*, 1995). Nevertheless, the greatly increased apparent relative molecular mass may reflect to a certain extent the occurrence of protein aggregation and precipitation which is observable at ~ pH 6.

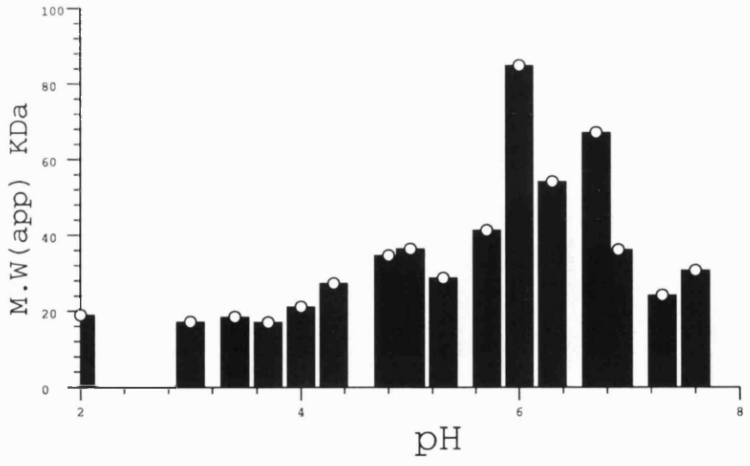


Fig 6.18 The pH dependence of the apparent relative molecular mass of CD2d1 as obtained from the PFGLED experiments performed with 1.2 mM CD2d1 in 20 mM phosphate at 25 °C.

Arginine Side-chain Mobility

From the heteronuclear [^{1}H]-^{15}N NOE experiments, the mobility of the arginine side-chain can also be examined. The heteronuclear NOE for some selected arginine side-chains NεH groups are displayed in Figure 6.19. A large negative NOE is indicative of high degree of flexibility of the side-chain while positive NOE suggest some degree of rigidity for the side chain. With the exception of Arg1 in the unstructured N-terminus, all the arginine Hε/Nε cross peaks are readily visible up to pH 6.0. The N$^{\varepsilon}$H cross peak of Arg31, however, did not show significant solvent exchange broadening until pH 8.5 and has the highest NOE values at all pH titration points studied. Together with the indication of transient hydrogen bonding with Glu29 and Glu41 as described in previous chapter, the data suggests that the Arg31 N$^{\varepsilon}$H group is protected from exchange by hydrogen bonding or by burial (Pascal *et al.*, 1995; Singer & Forman-Kay, 1997).

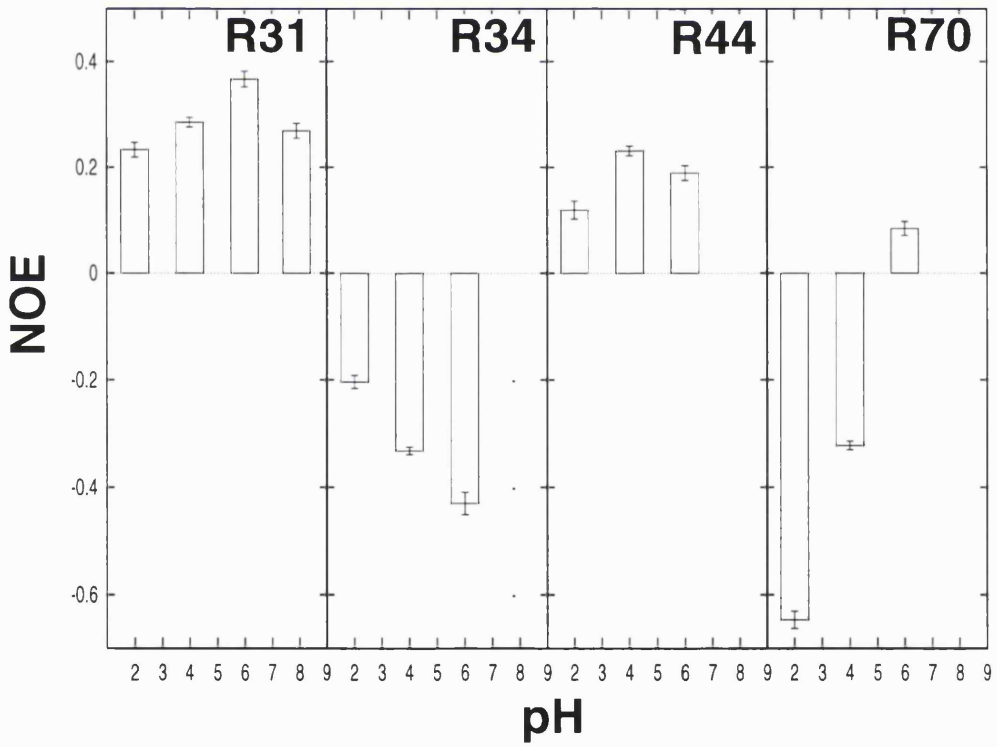


Fig. 6.19 pH dependent behaviour of the heteronuclear NOE of the arginine side-chains. Not shown are the residues R87 and R96 which have NOE profiles similar to that of R34.

The only other arginine with consistently positive N$^{\varepsilon}$H heteronuclear NOE values is Arg44 which could be a reflection of the hydrogen bonding to the backbone

carbonyl of Ile27, as well as transient hydrogen bonding to Asp25 and Asp28. The $N^{\varepsilon}H$ NOE values of Arg34, Arg87 and Arg96 are consistently negative which indicate that they are flexible at all pH values studied. The side chain of Arg70, while completely flexible at low pH, showed rising trends in NOE values with pH that indicates increasing rigidity which is most likely a result of the formation of a hydrogen bond with deprotonated Glu99.

Mutagenesis Studies

The relaxation analyses thus far showed that maximal self-association is achieved at pH value ~6.0. With the pK_a of Glu41 determined to be 6.7 at the protein concentration at which the experiments were done, the results strongly suggest that the maximal self-association requires Glu41 to be protonated while another residue, in all probability Glu29, should be deprotonated. The relaxation experiments to determined the relaxation parameters R_1 an R_2 were therefore repeated using the E41Q and E29Q mutants in order to assess the role played by Glu41 and Glu29 in CD2 self-association.

The results from these mutant studies strongly support the notion that dimerisation of CD2d1 requires a protonated Glu41 and a deprotonated Glu29 (Fig. 6.20). For the E29Q mutant, there is essentially no binding with only small variations of ~0.5 in the τ_c determined at all pH values in which the experiments were performed. For the E41Q mutant in which the charge on Glu41 is removed, there is little or no binding at low pH, while at pH 6 and above, the τ_c increased to ~7.2, indicating self-association of CD2d1 at higher pH. The τ_c value of the E41Q mutant however, is lower than that of the wild-type CD2 and does not showed variations above pH 6. This slightly lower binding affinity of E41Q mutant be due in part to possible rearrangement of the side-chain conformations in the wild-type CD2d1 at ~pH 6 when a protonated Glu41 can directly form a hydrogen bond with Glu29, while at higher pH the repulsive interaction of the two charged residue would move apart. The result from the mutants studies suggests that the lost of the dimerisation at pH lower than 6 is due to the protonation of Glu29, while the

decrease in binding affinity at pH higher than 6 is a result of the deprotonation of Glu41.

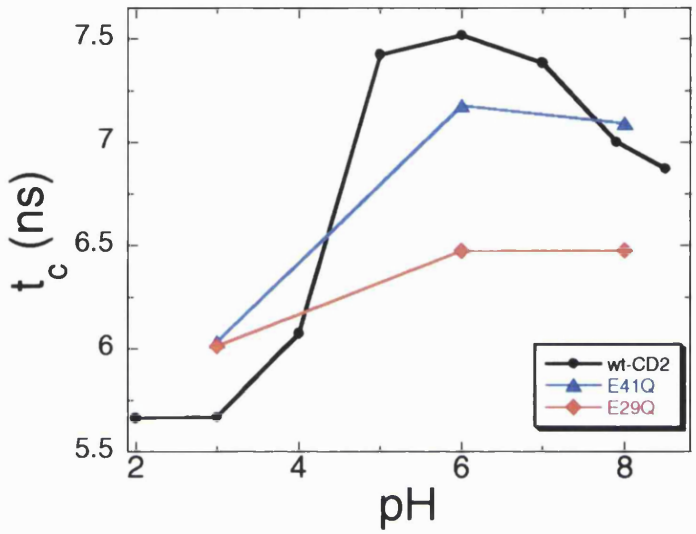


Fig. 6.20 The pH dependence of τ_c for the E29Q and E41Q mutants. The E29Q mutant is marked with red diamond while the E41Q is indicated by blue triangle, and the wild-type is in black.

Intertwined dimer

An unusual observation from the structural studies of CD2d1 and its mutants is their remarkable ability to adopt many different folds, either as a stable monomer, intertwined metastable dimers, or trimeric and tetrameric assembly (Hayes *et al.*, 1999; Murray *et al.*, 1998; Murray *et al.*, 1995). A small fraction of the wild-type protein (~15%) has been shown to form an intertwined IgSF domain. This unusual property of CD2d1 has been proposed to be the consequence of expressing the protein as a glutathione S-transferase (GST) fusion protein, in which the dimeric GST brings two polypeptide chains into close proximity during folding, and unfolding and refolding of the monomeric species can produce the dimeric folds. The unfolding of CD2d1 is permitted by the absence of the conserved V-set IgSF disulphide bond in

CD2d1. Such absence of disulphide bond is important in muscle protein such as titin and telokin in which partial unfolding is possible (Holden *et al.*, 1992). SDM of specific residues in CD2d1 can result in increased formation of the intertwined form while deletion of segments of the proteins can results in an alternative intertwined fold as well as a tetrameric form. In the analysis of the dimerisation of CD2d1, it is therefore necessary to examine such possibility in order to eliminate the misfolding aspect of CD2d1 as being responsible for the observed exchange phenomena.

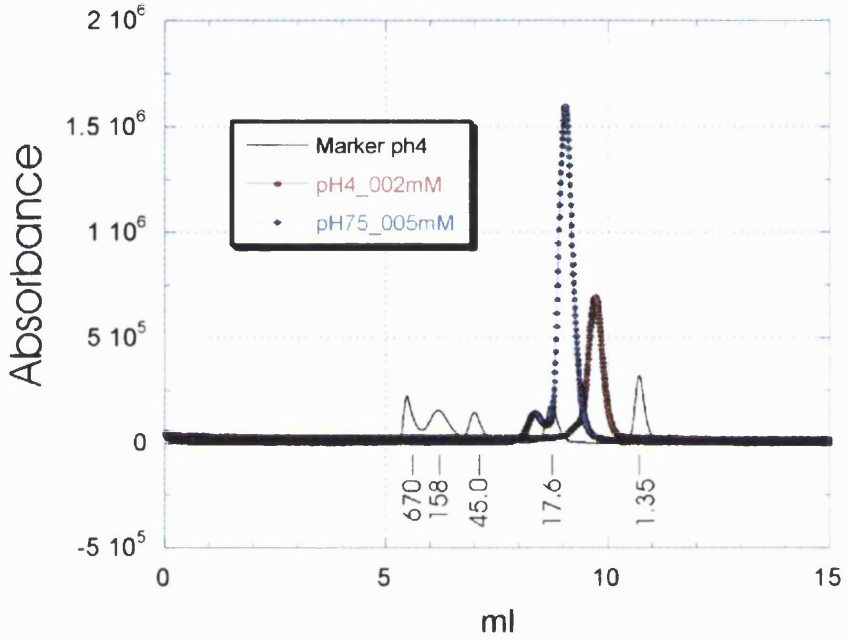


Fig. 6.21 The elution profile of CD2d1 at pH 7.5 (blue line) and pH 4.0 (red line). The molecular size of the marker in KDa is shown underneath. The concentration of CD2d1 used was 20 μM at pH 7.5 and 50 μM at pH 4.0.

The purified CD2d1, which was expressed as the native form (i.e. not expressed as a fusion protein), was concentrated and loaded onto an analytical gel filtration column. As can be seen from the elution profile of CD2d1 at pH 7.5 (Fig. 6.21), the CD2d1 can exist as two different species with apparent molecular sizes that suggest monomeric and dimeric forms. No change in the relative height of the two eluted peaks were observed when the experiments were repeated at varying protein concentrations (2 - 100 μM) and when the experiments were repeated at pH 5.7. Repeat of the experiment at pH 4, however, resulted in the protein being eluted as a single peak, suggesting that at lower pH the protein rapidly unfolds and forms the monomeric species. The CD2d1 monomer at pH 4.0 was also eluted with a lower

apparent molecular size when compared to the elution profile at pH 7.5. This may be a reflection of a change in its hydrodynamic properties at lower pH, or that the monomer-dimer equilibrium at higher pH resulted in the higher apparent molecular size.

The results suggests that the presence of the intertwined dimer need not be the product of an artefact from the use of GST fusion system, rather the formation of the intertwined dimer can arise naturally at a higher protein concentration. Such spontaneous dimer formation has previously been demonstrated for deletion mutant (Hayes *et al.*, 1999). This intertwining of wild-type CD2d1 appeared to be a slow process, and its presence is not apparent during the purification process when the protein is purified relatively rapidly.

The small amount of intertwined dimer species that is seen in the analytical gel filtration, however, is not reflected in the ^{15}N-^{1}H HSQC spectra. The formation of such intertwined form may be expected to result in heterogeneity of the spectra, however, such effect is not readily observable for wild-type CD2d1. This may be due to the much lower concentration of this dimeric form together with an increased in size of this intertwined dimer that produce broadened peaks.

Some mutants of CD2d1 have been previously demonstrated to show an increase the proportion of CD2d1 intertwined dimer formed (Hayes *et al.*, 1999; Murray *et al.*, 1998). For most of the mutants used for the studies of CD2d1, no sign of structural heterogeneity can be observed. A few, however, showed some degree of heterogeneity in the NMR spectra. For the F49A mutant, for example, the appearance of smaller companion peaks for many residues can be seen as an indication of the presence of an alternatively folded form (Fig. 6.22). However, these extra peaks may also be the products of cis-trans isomerisation as the heterogeneity appeared to be limited to a subset of resonances. Phe49 is situated next to a proline, Pro48, and removal of the aromatic groups may reduces the stability of the C'' and D strands. While the presence of bulky aromatic groups preceding a proline has been shown to substantially reduce cis-trans isomerisation (Reimer *et al.*, 1998), Phe49 is located after proline and any effect may reflect its importance in maintaining the structural integrity of the protein.

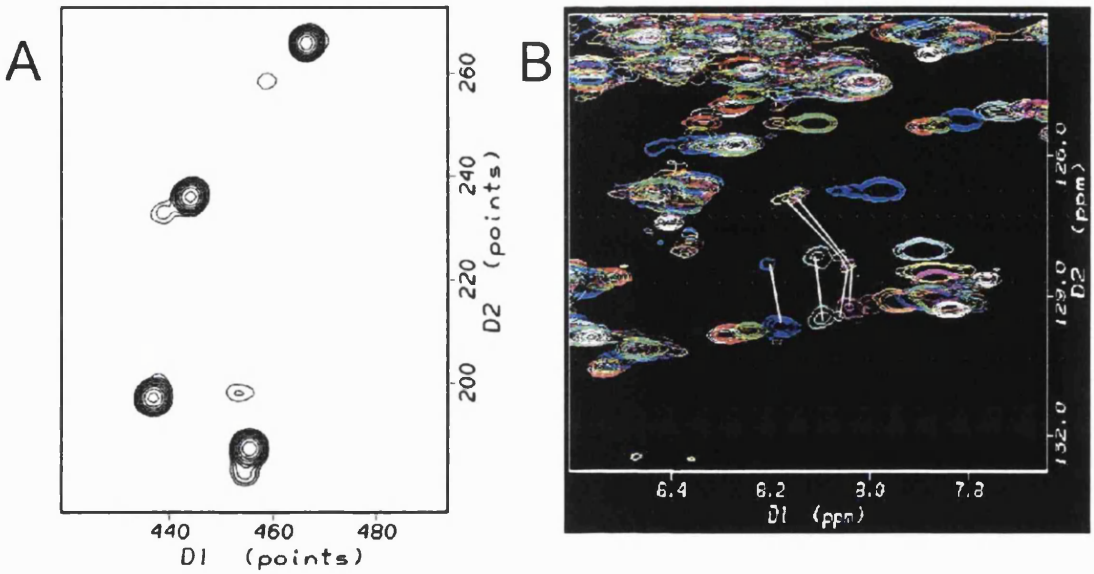


Fig. 6.22 Heterogeneity of the ^{1}H-^{15}N HSQC spectra of the F49A and K43A mutants. A - In the spectrum of the F49A mutant, some resonances showed the presence of a smaller companion resonance. B - The overlay spectra of the K43A mutant at different pH. Each colour represent a different pH titration point. For Lys43, the resonances showed multiple resonances between pH 3.5-5.5 and these resonances of Lys43 are joined by white line.

In the K43A mutant, the resonances of a large number of residues appeared to be exchange broadened with the maximum broadening seen at ~pH5. Closer inspection of the spectra, however, reveal the presence of multiple peaks which suggest a slow chemical exchange event rather than the intermediate chemical exchange suggested by exchange broadening (Fig. 6. 22). These may be indications of the formation of the intertwined dimer, however, it is noticeable that the effect is more pronounced for residues on the binding surface. This differential effect suggests that the appearance of these multiple resonances could therefore also be interpreted as the formation of a tighter 'head-to-head' dimer rather than that of an intertwined dimer.

It is therefore not possible to ascertain from such limited studies whether the intertwined form of CD2 is present in significant amount in these mutants. Within the time that was available, it had not been possible to assess the association states of these mutants properly. Further work on these mutants is therefore necessary in order to provide a more comprehensive analysis of these mutants.

Discussion

Dimerisation of CD2

The previous results as described in Chapter 5 have provide strong indication of CD2 self-association; while studies using other biophysical studies showed no observable dimer formation (Silkowski *et al.*, 1997). In this chapter, the CD2d1 is characterised by a variety of biophysical methods, but principally by the analysis of ^{15}N backbone dynamics. The result obtained here demonstrates unambiguously that the CD2 do dimerise, albeit with very low affinity, at higher protein concentration. This is indicated by the concentration dependence of a number of observed effects: the very fast ^{15}N transverse relaxation rate R_2 that results in significant line-broadening for some of the residues, the weak but detectable changes in chemical shifts and the variations in rotational correlation time.

All the residues that showed significant line-broadening as reflected by the large R_{ex} terms are found on the crystal dimer interface (Fig. 6.7). Some of the crosspeaks, such as Tyr37, Leu38, Tyr79, Tyr86 and Val80, showed significant chemical shift changes in the ^{15}N-^{1}H HSQC spectra in addition to the exchange broadening behaviour. Extreme broadening was also observed for Leu38 between pH 5 - 7. Leu38 is responsible for the β-bulge on the C' strand which is thought to be important for the V-set dimerisation (Chothia *et al.*, 1985). The CC' loop (residues Arg34-Leu38) in one CD2 molecule can form a reciprocal backbone-backbone hydrogen bonds with the FG loop (Thr83-Arg87) of another. Such an intermolecular hydrogen bond as observed in the crystal structure exists between the backbone amide of Leu38 and backbone carbonyl of Gly85. The other backbone amide (Arg87) involved in intermolecular hydrogen bonding cannot be used for the analysis as it overlaps with other resonances. The nearby Leu89, however, showed consistently broadened peak at all pH and concentrations. The concordance of residues exhibiting exchange broadening effects and chemical shift changes with residues on the crystal dimer interface strongly suggest that the CD2 dimerises via its binding surface in the specific manner seen in the crystal structure. This highly specific dimerisation, however, is extremely weak. All the observed concentration

dependent parameters - the exchange terms R_{ex}, the chemical shift changes and the rotational correlation time τ_c, can be fit reasonably well to a homodimerisation model with a K_D of the order of 3-6 mM.

The analysis of the relaxation data also produces other kinetic constants for the CD2d1 dimerisation. While the affinity/association constant can tell how energetically favourable an interaction is, it doesn't say anything about the stability of the interaction. The k_{on} and k_{off} may be more relevant to the analysis of the binding interaction of cell adhesion molecules such as CD2. For example, a fast k_{off} allows for the rapid detachment of bound ligand necessary for molecules involved in transient interaction. The exchange rate constant k_{ex} is determined to be 15000 s^{-1}, the k_{on} ~5000 M^{-1} s^{-1} and k_{off} ~ 7s^{-1}. The k_{off} determined for CD2 dimerisation obtained is very similar to the values previously obtained for rat CD2/CD48 interaction (>6 s^{-1} (van der Merwe *et al.*, 1994b) and ~7.6 s^{-1} (Pierres *et al.*, 1996)) as well as for human CD2-CD58 interaction (~7 s^{-1}) (Sun *et al.*, 1999). The k_{on}, however, is substantially lower than the 10^5 $M^{-1}s^{-1}$ determined from CD2/CD48 interaction (Davis *et al.*, 1998). It would therefore appear that the low affinity for CD2 dimerisation is primarily the result of a substantially reduced k_{on}.

The dimerisation of CD2d1 is also highly dependent on the pH. This pH-dependent dimerisation can be seen in the variation in τ_c from the relaxation analysis with pH as well as from diffusion studies. The pH effect on dimerisation is also evident in the site-specific exchange phenomenon as can be seen in the relaxation as well as ^{19}F NMR data. Other evidence comes from the pH titration experiments described in previous chapter. For proteins in which the associations with other protein is mediated by charged residues, the energetics of the protein association is strongly pH dependent. This pH-dependent self-association can be observed in a large number of proteins, for instance, in SH2 domain of Src homologous collagen-like (SHC) protein (Rety *et al.*, 1996) glycinamide ribonucleotide transformylase (Mullen & Jennings, 1996), HIV-1 protease (Xie *et al.*, 1999), thioredoxin (Andersen *et al.*, 1997), flagellar cap protein HAP-2 (Imada *et al.*, 1998), duck lens δ-crystallin (Chang *et al.*, 1997) and Bcl-2 family of proteins (Xie *et al.*, 1998). One or more titratable residues may be implicated in such pH-dependent protein-protein interactions. A glutamate and a histidine, for example, have been shown to mediate

the pH-dependent dissociation of protein the Fc fragments of mouse immunoglobulin-G from the B domain of staphylococcal protein A (Kato *et al.*, 1993); and a single glutamate or aspartate has been shown to be capable of mediating the self-association of protein and peptides (Andersen *et al.*, 1997; Dieckmann *et al.*, 1998; Mullen & Jennings, 1996; Whittaker & Whittaker, 1998).

From the relaxation analysis using the E29Q and E41Q mutants, the lost of binding capability of the E29Q mutant suggests that negative charge on Glu29 is crucially important for the binding interaction of CD2d1. The negative charge on Glu41, however, is dispensable for dimer formation and in fact maximal dimer formation is achieved when Glu41 is protonated while Glu29 is deprotonated. The association or dissociation of CD2 may therefore be mediated by ionisation events of Glu41 and Glu29 with the presence of the negative charge on Glu29 being of greater importance.

The significance of the observation of dimer formation is two fold. Firstly, CD2, CD48, CD58 form a subset of molecules within the IgSF and, by an examination of the sequences and chromosomal location of genes encoding CD2 and its homologous adhesion ligands CD48 and CD58, it has been suggested that adhesion molecules have evolved from the duplication of a primodial gene encoding a protein engaged in homophilic recognition (Williams & Barclay, 1988; Wong *et al.*, 1990). The model of interaction is postulated to be the 'head-to-head' interaction as seen in the crystal structure of human and rat CD2 (Jones *et al.*, 1992). Such mode of interaction has since been confirmed in the crystal structure of heterophilic complex of human CD2/CD58 (Wang *et al.*, 1999). There was, however, no previous evidence of CD2 dimerisation from biophysical studies performed in physiological conditions. This observation by NMR analysis of the dimer formation in solution, albeit with very low affinity, is therefore highly significant and lend credence to the evolutionary hypothesis.

Second, while the crystal structure of CD2/CD58 has been solved (Wang *et al.*, 1999), there is as yet no structure of the CD2/CD48 heterophilic adhesion complex. The evidence here showed that the self-association is mediated by the binding surface of CD2 which is consistent with other studies on CD2/CD48 interaction (McAlister *et al.*, 1996). This, together with the similarity between rat and human CD2 crystal dimer structures (Bodian *et al.*, 1994; Jones *et al.*, 1992), as

well as the observation of CD2/CD58 heterophilic complex that directly mirrors human CD2 dimer interaction (Wang *et al.*, 1999; Sun *et al.*, 1999), points to the validity of basing the CD2/CD48 interaction on the crystal dimer contact model. All the evidence therefore suggests that CD2 binds to CD48 in the same manner seen in CD2/CD58 interaction and rat CD2 crystal dimer.

CD2d1 dimer, intertwined dimer and higher order aggregates

The rat CD2d1 was observed to form intertwined dimer and has a tendency to aggregate at higher protein concentration. Such aggregation which is also observable in soluble human CD2 has been shown to affect other binding analyses (van der Merwe *et al.*, 1994a; van der Merwe *et al.*, 1993). It is therefore useful to consider if protein aggregation or indeed formation of intercalated dimer can produce the exchange phenomenon seen here. The Lipari-Szabo model-free formalism used in this analysis has been previously shown to be susceptible to the presence of dimer and aggregates (Schurr *et al.*, 1994). The analysis of the slow internal motion using the extend model-free approach in particular (Clore *et al.*, 1990a; Clore *et al.*, 1990b) can be distorted by the presence of a small amount of dimers and can introduce substantial amplitudes of non-existent slow internal motion. However, in this analysis, the observation of enhanced exchange contribution R_{ex} at higher protein concentration is highly site-specific as discussed earlier. The relaxation effect is also mirrored by the change in chemical shift of specific residues at different protein concentration and exchange broadening can be directly observed in the ^{15}N-^{1}H HSQC and ^{19}F-NMR spectra. The effect on the relaxation parameters seen is therefore unlikely to artefactual.

It is also suggested that the used of spectral density mapping methods may reduced the artefactual effect that can arise from the model-free approach (Fairbrother *et al.*, 1998). Relaxation analysis had been previously performed on human CD2 (Wyss *et al.*, 1997) and was interpreted using the reduced spectral-density mapping approach (Lefevre *et al.*, 1996). As shall be discussed later, a comparison of the two analyses showed the results of both analyses to be highly

similar and does not suggest any irregularity that may might be introduced by using the model free approach for this analysis.

As previously mentioned, little heterogeneity of the resonances that may have arisen from the presence of intertwined CD2d1 could be observed, suggesting that the small amount of intertwined CD2d1 that may be present does not have any significant observed effect. All the evidence suggest that the effects observed arise from the non-intertwined CD2 species involved in a monomer-dimer equilibria, for example, the chemical shifts changes can be directly fitted to an equation describing monomer-dimer equilibria and cannot be explained by the presence of stable intertwined dimer.

An interesting observation from the study of rat CD2 is the association between dimer formation and aggregation. Maximal CD2 dimer formation between pH 5-7 can be directly correlated with observed precipitation of the protein. Such association between dimer formation can also be seen in the mutants. For example, the E29Q mutant which showed little or no dimer formation also does not precipitate even at very high protein concentration, while other mutants of residues on the binding site that have strong influence on CD2 dimer interaction can precipitate readily at very low protein concentration. The tendency to precipitate appear to be a general property of all forms of CD2, and other CD2 homologues such as human and mouse CD48 also showed a tendency to aggregate (Davis *et al.*, 1993; van der Merwe *et al.*, 1993). The aggregation of rat CD2 is rapid, while the formation of the intertwined dimer appears to be a slow process, the formation of the intertwined dimer and aggregation of CD2d1 may therefore be two separate unrelated events.

The strong correlation between dimer formation and aggregation of CD2d may allow the aggregation effect to be delineated and dissected. The formation of CD2d1 dimer may produce one or more secondary low-affinity binding sites that permit further protein self-associations. Such a hypothesis can be easily tested by constructing mutants of residues away from the CD2 binding surface that can disrupt or enhance protein aggregation. The non-aggregating form of the various CD2 mutants may be useful for further biophysical studies, for example, in further delineating the relationship between the kinetics of protein self-association and the exchange process; while other mutants may also be useful for investigation into the effect of aggregation on the biological properties of the protein. Alternatively, the

aggregation may be the linked to the stability or unfolding of the protein (Kendrick *et al.*, 1998). This may be examined by protein folding and thermostability studies and the construction of a CD2d1 mutant that does not easily unfold by the insertion of disulphide bond (Gray *et al.*, 1993).

Backbone Dynamics of CD2d1

The analysis of the molecular dynamics permits a detailed description of the molecular motions in the picosecond to millisecond time scale for nearly all the NH groups. However, the low affinity dimerisation has been shown to have a significant impact on the relaxation analysis, and the significant line-broadening as reflected by the large R_{ex} terms for some residues can now be seen as the product of dimerisation. Any discussion on the internal dynamics of the protein should therefore exclude residues exhibiting strong concentration dependence in their relaxation parameters, and real internal motions of the molecules may be best observed at a lower protein concentration or pH at which protein dimerisation is at a minimal. From the analysis of the dynamics data, CD2d1 is shown to be a stable dynamic unit with some degrees internal motion. This result is not unexpected for a stably folded protein, where freedom of mobility is typically exhibited by termini and the loop regions of the polypeptide chain. The N-terminus residues showed particular high degree of mobility as shown by the negative NOE and low values for the order parameter S^2 of the first few residues. For most residues which form the hydrophobic core of the protein and may be important in maintaining the structural stability of V-set domain, the S^2 values are average or above average, suggesting restricted motions for these backbone amides.

A few residues, such as Thr69, Thr86, Leu88, have consistent R_{ex} values of ~2 s^{-1} throughout the dilution series and may reflect the true conformation exchange phenomenon in the ms to μs timescale. Thr69 lies on the EF loop while Thr86 and Leu88 are situated on the FG loop. The slower motional changes of the FG loop may have particular significance in the binding interaction of CD2. The Arg87 backbone amide forms a hydrogen bonds with the backbone carbonyl of Ser36 of its dimer partner, while its side-chain also participate in a intermolecular hydrogen-bonding with the carbonyl of Gly35 and salt bridge with the carboxyl group of Glu33. The

nearby Gly85 also forms backbone hydrogen bonding with Thr38. Movement of this loop may therefore disrupt these intermolecular interactions which can provide binding energy for CD2 adhesion, or allow a conformation optimal for binding to be adopted. It is therefore easy to envisage how movement of this region can have significant impact on the binding interaction of CD2.

It is useful to compare the relaxation analysis that has been performed on human CD2 with the present analysis. The main results from this study on human CD2d1 mirror that obtained from rat CD2d1 obtained at high protein concentration (Wyss *et al.*, 1997). The substantial line-broadening for a subset of resonances from many residues that form the CD58 binding surface is also observed. In particular, residues in the segment from Thr37 and Lys57 showed large exchange broadening; these residues are located at the C' strand, CC' and C'C'' loop in human CD2. The result was interpreted as multiple conformation fluctuations of the binding surface on the microseconds to millisecond time-scale (Wyss *et al.*, 1997). However, as the crystal structures of human and rat CD2 showed, despite significant differences in sequence, nearly identical homodimeric binding interaction (Bodian *et al.*, 1994; Jones *et al.*, 1992), and the results from the relaxation analyses of both molecules are very similar, the dimerisation that is seen for rat CD2d1 can therefore better explain the observed exchange phenomenon in human CD2d1. Correlation has also been made between the exchange phenomenon on the binding surface and the association rate constant k_{on} of the CD58 to CD2. However, as noted earlier, should any such correlation exist in CD2/CD48 interaction, it is likely to be limited to the motions seen on the FG loop instead of the conformational fluctuations of the entire binding surface as proposed by Wyss *et al.*, (1997).

Interpretation of exchange phenomenon

The comparison between the rat and human CD2d1 and the different interpretations of the exchange phenomenon bring into focus an important issue that needs to be addressed. The interpretation of the physical origin of the slower motion as reflected by the magnitude of the transverse relaxation rate R_2 is not straightforward. In the Lipari-Szabo formalism, the motions on the slower time-scale is analysed by using the exchange term R_{ex} which is a correction factor necessary for

the fitting of the relaxation parameters. This occurrence of the slower exchange phenomenon is often attributed to complex conformational rearrangements (Banci *et al.*, 1998; LeMaster & Kushlan, 1996; Volkman *et al.*, 1998). In a few specific cases, the observed exchange phenomena can indeed be satisfactorily explained by a detailed descriptions of well-defined conformational rearrangements, such as the isomerisation of disulphide bond (Otting *et al.*, 1993) and the backbone peptide bond of proline residues (Evans *et al.*, 1989). Such conformational changes in general result in resolution of distinct resonances in slow exchange. In many other cases, however, the explanation of the observed exchange phenomena involved generalised descriptions of the movement of surface loops. More unusual conformational changes, such as flexing and twisting of secondary structure elements or multiple conformational fluctuations have also been invoked to explain the observed exchange phenomena (Akke *et al.*, 1998; Banci *et al.*, 1998; Volkman *et al.*, 1998; Wong *et al.*, 1997; Wyss *et al.*, 1997). Such proposed exchange mechanisms, however, are difficult to verify. Given that flexibility and local motions of the proteins can have important implication for the functioning of the protein such as ligand-binding and protein-protein interaction, and the slower exchange phenomena are frequently interpreted in terms of conformation changes that are correlated with protein functions (Feher & Cavanagh, 1999; Kay *et al.*, 1998, Stock, 1999), it is of vital importance that the validity of such interpretations is examined closely.

The results obtained here demonstrated that protein self-association has a sizeable effect on the residue-specific line-broadening at the typical concentrations used in NMR experiments, and the effect cannot be explained by the neglect of anisotropic rotational diffusion in the analysis (Pfuhl *et al.*, 1999). The analysis showed that very low affinity protein self-association, perhaps beyond the limit that can be detected by other biophysical methods, can have a very pronounced effect on the relaxation data. Furthermore, the results of such low affinity self-association at high protein concentration need not be non-specific, rather, as can be seen for CD2d1, the effect on the exchange broadening is highly localised and residue-specific. It is also important to note that the because of the factor $p_A p_{A2}$ in Equation [6.26], the maximal effect of dimerisation exchange broadening can be achieved at concentrations well below K_D (the exchange broadening coefficients $p_A p_{A2}$ has a maximum of 0.25 for $p_A = p_{A2} = 0.5$). This is significant as NMR studies is

normally conducted using protein concentration in the mM range and large effect can be produced even when the association is of extremely low affinity. For CD2, assuming a K_D of ~5 mM, at the 1.2 mM protein concentration used in many of the experiments, ~30% of the CD2d1 may be in a dimeric state.

This observation of the pronounced effect from the low affinity binding is therefore of general significance in the interpretation of exchange phenomena. As mentioned earlier, many studies have attributed conformational changes, some more elaborate than others, to explain the observed exchange phenomena (Akke *et al.*, 1998; Banci *et al.*, 1998; Volkman *et al.*, 1998; Wong *et al.*, 1997; Wyss *et al.*, 1997). For example, barstar has been suggested to undergo rigid body movement of helix and twisting motion of β-sheet (Wong *et al.*, 1997), similarly the F-helix of haemoglobin-CO was also suggested to move as a hydrogen-bonded unit pivoting around a particular residue showing extreme line-broadening (Volkman *et al.*, 1998), while the fibronectin type III domain of human tenascin-C and the human CD2d1 were said to exhibit pervasive conformation fluctuations (Akke *et al.*, 1998; Wyss *et al.*, 1997). It is, however, not possible to discriminate true intramolecular conformational changes from protein self-association from the single set of relaxation data determined in a single specific experimental condition often used in such analyses. It can also be noted that in some of these studies, the rotational correlation time is significantly higher than what might be expected from theoretical calculations, (Banci *et al.*, 1998; Volkman *et al.*, 1998) and some have been shown to self-associate with low affinity or form aggregates (Akke *et al.*, 1998; Wyss *et al.*, 1997). While it is not possible to say with certainty that the suggestions of complex conformation changes for some of these protein are erroneous without conducting further experiments, self-association, by the principle of Occam's razor, may be a more reasonable explanation for the observed exchange phenomena in some of these proteins. Indeed, for human CD2d1, as discussed earlier, the exchange phenomenon observed is so similar to rat CD2d1 as to suggest that the interpretation of pervasive conformational fluctuations is likely to be wrong, while for fibronectin type III domain of human tenascin-C, the absence of the two C-terminal residues has been shown to be responsible for the observed chemical exchange events in the ms-μs timescale (Meekhof *et al.*, 1998). Exotic interpretations of relaxation data that suggest complex conformation changes should therefore be regarded with scepticism

if based upon only a single set of relaxation data and/or without careful consideration of the effect of low-affinity self-association.

Relaxation analysis should therefore require at the least two sets of experiments to be performed at two different protein concentrations. While protein self-association is known to occur frequently at the protein concentration in which NMR experiments are carried out (Schurr *et al.*, 1994; Zhang *et al.*, 1998), with the exception of a few previously reported cases (Fairbrother *et al.*, 1998; Fushman *et al.*, 1997; Grzesiek *et al.*, 1997), most relaxation analyses do not give explicit considerations to the effects of self-association on nuclear relaxation rates, and the deliberate measurements of relaxation rates over a range of protein concentration are rarely done. In principle, any self-association that is present can be detected by the careful monitoring of chemical shifts as a function of protein concentration (Fairbrother *et al.*, 1998). However, great care need to be taken to maintain identical sample condition over the dilution range as the chemical shifts may be strongly affected by other influences as was observed for CD2d1 in the pH 5.0-7.0 range. It is also noteworthy that the effects observed for CD2d1 were not significantly affected by the addition of salt and the use of a different buffer, which are typical measures taken to remedy problems associated with sample conditions. A dilution series of relaxation experiments, although time-consuming, would therefore provide better insight into the internal dynamics of the proteins.

Conclusion

The result obtained showed that the CD2d1 dimerised with a very low affinity of ~3-6 mM. This dimerisation is highly pH-dependent and may be mediated by the ionisation states of a few acidic residues. In particular, maximal association may be seen when Glu41 is protonated while Glu29 is deprotonated.

There relaxation analysis of CD2d1 also showed that the relaxation parameters can be strongly affected by this very low affinity binding interaction, and that such pronounced effect can occurs at protein concentration normally used for NMR analysis. Such low affinity binding is nevertheless highly site-specific with

severely enhanced transverse relaxation rates observed for many residues. The results suggest that low affinity protein self-association should be considered before the attribution of more elaborate conformational change mechanism. In order to avoid erroneous interpretation of the relaxation data, a full characterisation of protein dynamics will therefore require a more complete analysis of the relaxation rate over a range of protein concentration.

References

Abragam, A. (1961). *Principles of Nuclear Magnetism*, Oxfod University Press.

Akke, M., Liu, J., Cavanagh, J., Erickson, H. P. & Palmer, A. G. (1998). Pervasive conformational fluctuations on microsecond time scales in a fibronectin type III domain. *Nature Structural Biology* **5**(1), 55-59.

Altieri, A. S., Hinton, D. P. & Byrd, R. A. (1995). Association of Biomolecular Systems Via Pulsed-Field Gradient NMR Self-Diffusion Measurements. *Journal of the American Chemical Society* **117**(28), 7566-7567.

Andersen, J. F., Sanders, D. A. R., Gasdaska, J. R., Weichsel, A., Powis, G. & Montfort, W. R. (1997). Human thioredoxin homodimers: Regulation by pH, role of aspartate 60, and crystal structure of the aspartate 60->asparagine mutant. *Biochemistry* **36**(46), 13979-13988.

Banci, L., Felli, I. C. & Koulougliotis, D. (1998). Identification of slow motions in the reduced recombinant high- potential iron sulfur protein I (HiPIP I) from Ectothiorhodospira halophila via N-15 rotating-frame NMR relaxation measurements. *Journal of Biomolecular NMR* **12**(2), 307-318.

Barbato, G., Ikura, M., Kay, L. E., Pastor, R. W. & Bax, A. (1992). Backbone Dynamics of Calmodulin Studied By N-15 Relaxation Using Inverse Detected 2-Dimensional NMR-Spectroscopy - the Central Helix Is Flexible. *Biochemistry* **31**(23), 5269-5278.

Bartels, C., Xia, T. H., Billeter, M., Guntert, P. & Wuthrich, K. (1995). The Program XEASY For Computer-Supported NMR Spectral-Analysis of Biological Macromolecules. *Journal of Biomolecular NMR* **6**(1), 1-10.

Bloch, F. (1946). *Physical Review* **70**, 460-474.

Bodenhausen, G. & Ruben, D. J. (1980). Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Lett.* **69**(1), 185-189.

Bodian, D. L., Jones, E. Y., Harlos, K., Stuart, D. I. & Davis, S. J. (1994). Crystal-Structure of the Extracellular Region of the Human Cell- Adhesion Molecule CD2 At 2.5-Angstrom Resolution. *Structure* **2**(8), 755-766.

Boucher, W. (1993). AZARA v1.0, Department of Biochemistry, University of Cambridge, UK.

Cavanagh, J., Fairbrother, W. J., Palmer III, A. G. & Skelton, N. J. (1996). *Protein NMR spectroscopy - Principles and Practice*, Academic Press, London.

Chang, G. G., Lee, H. J. & Chow, R. H. (1997). pH-induced reversible dissociation of tetrameric duck lens delta- crystallin. *Experimental Eye Research* **65**(5), 653-659.

Chothia, C., Novotny, J., Bruccoleri, R. & Karplus, M. (1985). Domain Association in Immunoglobulin Molecules - the Packing of Variable Domains. *Journal of Molecular Biology* **186**(3), 651-663.

Clore, G. M., Driscoll, P. C., Wingfield, P. T. & Gronenborn, A. M. (1990a). Analysis of the Backbone Dynamics of Interleukin-1-Beta Using 2- Dimensional Inverse Detected Heteronuclear N-15-H-1 NMR-Spectroscopy. *Biochemistry* **29**(32), 7387-7401.

Clore, G. M., Szabo, A., Bax, A., Kay, L. E., Driscoll, P. C. & Gronenborn, A. M. (1990b). Deviations From the Simple 2-Parameter Model-Free Approach to the Interpretation of N-15 Nuclear Magnetic-Relaxation of Proteins. *Journal of the American Chemical Society* **112**(12), 4989-4991.

Crawford, D. A. (1994). Structure and Dynamics of a Cell Adhesion Protein, Hertford College, University of Oxford.

Davis, S. J., Davies, E. A., Tucknott, M. G., Jones, E. Y. & van der Merwe, P. A. (1998). The role of charged residues mediating low affinity protein-protein recognition at the cell surface by CD2. *Proceedings of the National Academy of Sciences of the United States of America* **95**(10), 5490-5494.

Davis, S. J., Jones, E. Y., Bodian, D. L., Barclay, A. N. & Vandermerwe, P. A. (1993). Analysis of the Structure and Interactions of Cd2. *Biochemical Society Transactions* **21**(4), 952-958.

Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRpipe - a Multidimensional Spectral Processing System Based On Unix Pipes. *Journal of Biomolecular NMR* **6**(3), 277-293.

Dieckmann, G. R., McRorie, D. K., Lear, J. D., Sharp, K. A., DeGrado, W. F. & Pecoraro, V. L. (1998). The role of protonation and metal chelation preferences in defining the properties of mercury-binding coiled coils. *Journal of Molecular Biology* **280**(5), 897-912.

Dingley, A. J., Mackay, J. P., Chapman, B. E., Morris, M. B., Kuchel, P. W., Hambly, B. D. & King, G. F. (1995). Measuring Protein Self-Association Using Pulsed-Field-Gradient NMR-Spectroscopy - Application to Myosin Light-Chain-2. *Journal of Biomolecular NMR* **6**(3), 321-328.

Evans, P. A., Kautz, R. A., Fox, R. O. & Dobson, C. M. (1989). A Magnetization-Transfer Nuclear Magnetic-Resonance Study of the Folding of Staphylococcal Nuclease. *Biochemistry* **28**(1), 362-370.

Fairbrother, W. J., Liu, J., Pisacane, P. I., Sliwkowski, M. X. & Palmer, A. G. (1998). Backbone dynamics of the EGF-like domain of heregulin-alpha. *Journal of Molecular Biology* **279**(5), 1149-1161.

Farrow, N. A., Zhang, O. W., Szabo, A., Torchia, D. A. & Kay, L. E. (1995). Spectral Density-Function Mapping Using N-15 Relaxation Data Exclusively. *Journal of Biomolecular NMR* **6**(2), 153-162.

Feher, V. A. & Cavanagh, J. (1999). Millisecond-timescale motions contribute to the function of the bacterial response regulator protein Spo0F. *Nature* **400**(6741), 289-293.

Folkers, P. J. M., Clore, G. M., Driscoll, P. C., Dodt, J., Kohler, S. & Gronenborn, A. M. (1989). Solution Structure of Recombinant Hirudin and the Lys-47-]Glu Mutant - a Nuclear Magnetic-Resonance and Hybrid Distance Geometry Dynamical Simulated Annealing Study. *Biochemistry* **28**(6), 2601-2617.

Fushman, D., Cahill, S. & Cowburn, D. (1997). The main-chain dynamics of the dynamin pleckstrin homology (PH) domain in solution: Analysis of N-15 relaxation with monomer/dimer equilibration. *Journal of Molecular Biology* **266**(1), 173-194.

Gray, F., Cyster, J. G., Willis, A. C., Barclay, A. N. & Williams, A. F. (1993). Structural-Analysis of the CD2 T-Lymphocyte Antigen By Site-Directed Mutagenesis to Introduce a Disulfide Bond Into Domain-1. *Protein Engineering* **6**(8), 965-970.

Gryk, M. R., Abseher, R., Simon, B., Nilges, M. & Oschkinat, H. (1998). Heteronuclear relaxation study of the PH domain of beta-spectrin: Restriction of loop motions upon binding inositol trisphosphate. *Journal of Molecular Biology* **280**(5), 879-896.

Grzesiek, S., Bax, A., Hu, J. S., Kaufman, J., Palmer, I., Stahl, S. J., Tjandra, N. & Wingfield, P. T. (1997). Refined solution structure and backbone dynamics of HIV-1 Nef. *Protein Science* **6**(6), 1248-1263.

Hayes, M. V., Sessions, R. B., Brady, R. L. & Clarke, A. R. (1999). Engineered assembly of intertwined oligomers of an immunoglobulin chain. *Journal of Molecular Biology* **285**(4), 1857-1867.

Holden, H. M., Ito, M., Hartshorne, D. J. & Rayment, I. (1992). X-Ray Structure Determination of Telokin, the C-Terminal Domain of Myosin Light Chain Kinase, At 2.8 Angstrom Resolution. *Journal of Molecular Biology* **227**(3), 840-851.

Imada, K., Vonderviszt, F., Furukawa, Y., Oosawa, K. & Namba, K. (1998). Assembly characteristics of flagellar cap protein HAP2 of Salmonella: Decamer and pentamer in the pH-sensitive equilibrium. *Journal of Molecular Biology* **277**(4), 883-891.

Ishima, R. & Nagayama, K. (1995). Quasi-Spectral-Density Function-Analysis For N-15 Nuclei in Proteins. *Journal of Magnetic Resonance Series B* **108**(1), 73-76.

Jardetzky, O. (1996). Protein dynamics and conformational transitions in allosteric proteins. *Progress in Biophysics & Molecular Biology* **65**(3), 171-218.

Jones, E. Y., Davis, S. J., Williams, A. F., Harlos, K. & Stuart, D. I. (1992). Crystal-Structure At 2.8-Angstrom Resolution of a Soluble Form of the Cell-Adhesion Molecule CD2. *Nature* **360**(6401), 232-239.

Karplus, M. & McCammon, J. A. (1983). Dynamics of Proteins - Elements and Function. *Annual Review of Biochemistry* **52**, 263-300.

Kato, K., Gouda, H., Takaha, W., Yoshino, A., Matsunaga, C. & Arata, Y. (1993). C-13 NMR-Study of the Mode of Interaction in Solution of the B Fragment of Staphylococcal Protein-A and the Fc Fragments of Mouse Immunoglobulin-G. *Febs Letters* **328**(1-2), 49-54.

Kay, L. E., Keifer, P. & Saarinen, T. (1992a). Pure Absorption Gradient Enhanced Heteronuclear Single Quantum Correlation Spectroscopy With Improved Sensitivity. *Journal of the American Chemical Society* **114**(26), 10663-10665.

Kay, L. E., Muhandiram, D. R., Wolf, G., Shoelson, S. E. & Forman-Kay, J. D. (1998). Correlation Between Binding and Dynamics at SH2 Domain Interfaces. *Nature Structure Biology* **5**(2), 156-163.

Kay, L. E., Nicholson, L. K., Delaglio, F., Bax, A. & Torchia, D. A. (1992b). Pulse Sequences For Removal of the Effects of Cross-Correlation Between Dipolar and Chemical-Shift Anisotropy Relaxation Mechanism On the Measurement of Heteronuclear T1 and T2 Values in Proteins. *Journal of Magnetic Resonance* **97**(2), 359-375.

Kay, L. E., Torchia, D. A. & Bax, A. (1989). Backbone Dynamics of Proteins As Studied By N-15 Inverse Detected Heteronuclear NMR-Spectroscopy - Application to Staphylococcal Nuclease. *Biochemistry* **28**(23), 8972-8979.

Kendrick, B. S., Carpenter, J. F., Cleland, J. L., & Randolph, T. W. (1998). A Transient Expansion of the Native State Precedes Aggregation of Recombinant Human Interferon-γ. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 14142-14146.

Kraulis, P. J. (1989). Ansig - a Program For the Assignment of Protein H-1 2d-NMR Spectra By Interactive Computer-Graphics. *Journal of Magnetic Resonance* **84**(3), 627-633.

Lefevre, J. F., Dayie, K. T., Peng, J. W. & Wagner, G. (1996). Internal mobility in the partially folded DNA binding and dimerization domains of GAL4: NMR analysis of the N-H spectral density functions. *Biochemistry* **35**(8), 2674-2686.

LeMaster, D. M. & Kushlan, D. M. (1996). Dynamical mapping of E-coli thioredoxin via C-13 NMR relaxation analysis. *Journal of the American Chemical Society* **118**(39), 9255-9264.

Lipari, G. & Szabo, A. (1982a). Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules. 1. Theory and Range of Validity. *Journal of the American Chemical Society* **104**(17), 4546-4559.

Lipari, G. & Szabo, A. (1982b). Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules. 2. Analysis of Experimental Results. *Journal of the American Chemical Society* **104**(17), 4559-4570.

Mandel, A. M., Akke, M. & Palmer, A. G. (1995). Backbone Dynamics of Escherichia-Coli Ribonuclease Hi - Correlations With Structure and Function in an Active Enzyme. *Journal of Molecular Biology* **246**(1), 144-163.

McAlister, M. S. B., Mott, H. R., van der Merwe, P. A., Campbell, I. D., Davis, S. J. & Driscoll, P. C. (1996). NMR Analysis of Interacting Soluble Forms of the Cell-Cell Recognition Molecules CD2 and CD48. *Biochemistry* **35**(19), 5982-5991.

Meekhof, A. E., Hamill, S. J., Arcus, V. l., Clarke, J. & Freund, M. V. (1998). The Dependence of Chemical Exchange on Boundary Selection in a Fibronectin Type III Domain from Human Tenascin. *Journal of Molecular Biology* **282** (1), 181-194.

Mullen, C. A. & Jennings, P. A. (1996). Glycinamide ribonucleotide transformylase undergoes pH-dependent dimerization. *Journal of Molecular Biology* **262**(5), 746-755.

Murray, A. J., Head, J. G., Barker, J. J. & Brady, R. L. (1998). Engineering an intertwined form of CD2 for stability and assembly. *Nature Structural Biology* **5**(9), 778-782.

Murray, A. J., Lewis, S. J., Barclay, A. N. & Brady, R. L. (1995). One Sequence, 2 Folds - a Metastable Structure of CD2. *Proceedings of the National Academy of Sciences of the United States of America* **92**(16), 7337-7341.

Nesmelova, I. V. & Fedotov, V. D. (1998). Self-diffusion and self-association of lysozyme molecules in solution. *Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology* **1383**(2), 311-316.

Otting, G., Liepinsh, E. & Wuthrich, K. (1993). Disulfide Bond Isomerization in BPTI and BPTI(G36s) - an NMR-Study of Correlated Mobility in Proteins. *Biochemistry* **32**(14), 3571-3582.

Palmer, A. G., Cavanagh, J., Byrd, R. A. & Rance, M. (1992). Sensitivity Improvement in 3-Dimensional Heteronuclear Correlation NMR-Spectroscopy. *Journal of Magnetic Resonance* **96**(2), 416-424.

Pascal, S. M., Yamazaki, T., Singer, A. U., Kay, L. E. & Forman-Kay, J. D. (1995). Structural and Dynamic Characterization of the Phosphotyrosine Binding Region of a Src Homology-2 Domain-Phosphopeptide Complex By NMR Relaxation, Proton-Exchange, and Chemical-Shift Approaches. *Biochemistry* **34**(36), 11353-11362.

Peng, J. W. & Wagner, G. (1992a). Mapping of Spectral Density-Functions Using Heteronuclear NMR Relaxation Measurements. *Journal of Magnetic Resonance* **98**(2), 308-332.

Peng, J. W. & Wagner, G. (1992b). Mapping of the Spectral Densities of N-H Bond Motions in Eglin-C Using Heteronuclear Relaxation Experiments. *Biochemistry* **31**(36), 8571-8586.

Pfuhl, M., Chen, H. A., Kristensen, S. M. & Driscoll, P. C. (1999). NMR exchange broadening arising from specific low affinity protein self-association: Analysis of nitrogen-15 nuclear relaxation for rat CD2 domain 1. *Journal of Biomolecular NMR* **14**(4), 307-320.

Pierres, A., Benoliel, A. M., Bongrand, P. & van der Merwe, P. A. (1996). Determination of the lifetime and force dependence of interactions of single bonds between surface-attached CD2

and CD48 adhesion molecules. *Proceedings of the National Academy of Sciences of the United States of America* **93**(26), 15114-15118.

Powers, R., Clore, G. M., Stahl, S. J., Wingfield, P. T. & Gronenborn, A. (1992). Analysis of the Backbone Dynamics of the Ribonuclease-H Domain of the Human-Immunodeficiency-Virus Reverse-Transcriptase Using N-15- Relaxation Measurements. *Biochemistry* **31**(38), 9150-9157.

Reimer, U., Scherer, G., Drewello, M., Kruber, S., Schutkowski, M. & Fischer, G. (1998). Side-chain effects on peptidyl-prolyl cis/trans isomerisation. *Journal of Molecular Biology* **279**(2), 449-460.

Rety, S., Futterer, K., Grucza, R. A., Munoz, C. M., Frazier, W. A. & Waksman, G. (1996). pH-dependent self-association of the Src homology 2 (SH2) domain of the Src homologous and collagen-like (SHC) protein. *Protein Science* **5**(3), 405-413.

Schurr, J. M., Babcock, H. P. & Fujimoto, B. S. (1994). A Test of the Model-Free Formulas - Effects of Anisotropic Rotational Diffusion and Dimerization. *Journal of Magnetic Resonance Series B* **105**(3), 211-224.

Silkowski, H., Davis, S. J., Barclay, A. N., Rowe, A. J., Harding, S. E. & Byron, O. (1997). Characterisation of the low affinity interaction between rat cell adhesion molecules CD2 and CD48 by analytical ultracentrifugation. *European Biophysics Journal With Biophysics Letters* **25**(5-6), 455-462.

Singer, A. U. & Forman-Kay, J. D. (1997). pH titration studies of an SH2 domain-phosphopeptide complex: Unusual histidine and phosphate pK(a) values. *Protein Science* **6**(9), 1910-1919.

States, D. J., Haberkorn, R. A. & Ruben, D. J. (1982). A Two-Dimensional Nuclear Overhauser Experiment With Pure Absorption Phase in 4 Quadrants. *Journal of Magnetic Resonance* **48**(2), 286-292.

Stejskal, E. O. & Tanner, J. E. (1964). Spin Diffusion Measurements: Spin Echoes in the Presence of a Time-dependent Field Gradient. *Journal of Chemical Physics* **42**, 288-292.

Stock, A. (1999). Biophysics - Relating dynamics to function. *Nature* **400**(6741), 221-222.

Sun, Z. Y. J., Dotsch, V., Kim, M., Li, J., Reinherz, E. L. & Wagner, G. (1999). Functional glycan-free adhesion domain of human cell surface receptor CD58: design, production and NMR studies. *Embo Journal* **18**(11), 2941-2949.

van der Merwe, P. A., Barclay, A. N., Mason, D. W., Davies, E. A., Morgan, B. P., Tone, M., Krishnam, A. K. C., Ianelli, C. & Davis, S. J. (1994a). Human Cell-Adhesion Molecule CD2 Binds CD58 (LFA-3) With a Very-Low Affinity and an Extremely Fast Dissociation Rate But Does Not Bind CD48 or CD59. *Biochemistry* **33**(33), 10149-10160.

van der Merwe, P. A., Brown, M. H., Davis, S. J. & Barclay, A. N. (1993). Affinity and Kinetic-Analysis of the Interaction of the Cell-Adhesion Molecules Rat CD2 and CD48. *EMBO Journal* **12**(13), 4945-4954.

van der Merwe, P. A., Brown, M. H., Davis, S. J. & Barclay, A. N. (1994b). Affinity and Kinetic-Analysis of the Interaction of the Cell-Adhesion Molecule CD2 With Its Ligands CD48 and CD58. *Journal of Cellular Biochemistry*(S18C SIC), 257-257.

Volkman, B. F., Alam, S. L., Satterlee, J. D. & Markley, J. L. (1998). Solution structure and backbone dynamics of component IV Glycera dibranchiata monomeric Hemoglobin-CO. *Biochemistry* **37**(31), 10906-10919.

Wang, J., Smolyar, A., Tan, K. M., Liu, J., Kim, M. Y., Sun, Z. J., Wagner, G. & Reinherz, E. L. (1999). Structure of a heterophilic adhesion complex between the human CD2 and CD58 (LFA-3) counterreceptors. *Cell* **97**(6), 791-803.

Whittaker, M. M. & Whittaker, J. W. (1998). A glutamate bridge is essential for dimer stability and metal selectivity in manganese superoxide dismutase. *Journal of Biological Chemistry* **273**(35), 22188-22193.

Williams, A. F. & Barclay, A. N. (1988). The Immunoglobulin Superfamily - Domains For Cell-Surface Recognition. *Annual Review of Immunology* **6**, 381-405.

Withka, J. M., Wyss, D. F., Wagner, G., Arulanandam, A. R. N., Reinherz, E. L. & Recny, M. A. (1993). Structure of the Glycosylated Adhesion Domain of Human T-Lymphocyte Glycoprotein CD2. *Structure* **1**(1), 69-81.

Wong, K. B., Fersht, A. R. & Freund, S. M. V. (1997). NMR N-15 relaxation and structural studies reveal slow conformational exchange in Barstar C40/82A. *Journal of Molecular Biology* **268**(2), 494-511.

Wong, Y. W., Williams, A. F., Kingsmore, S. F. & Seldin, M. F. (1990). Structure, Expression, and Genetic-Linkage of the Mouse BCM1 (Ox45 or Blast-1) Antigen - Evidence For Genetic Duplication Giving Rise to the Bcm1 Region On Mouse Chromosome-1 and the CD2/Lfa3 Region On Mouse Chromosome-3. *Journal of Experimental Medicine* **171**(6), 2115-2130.

Wyss, D. F., Dayie, K. T. & Wagner, G. (1997). The counterreceptor binding site of human CD2 exhibits an extended surface patch with multiple conformations fluctuating with millisecond to microsecond motions. *Protein Science* **6**(3), 534-542.

Xie, D., Gulnik, S., Gustchina, E., Yu, B., Shao, W., Qoronfleh, W., Nathan, N. & Reed, J. C. (1999). Drug resistance mutations can affect dimer stability of HIV-1 protease at neutral pH. *Protein Science* **8**, 1702-1707.

Xie, Z. H., Schendel, S., Matsuyama, S. & Reed, J. C. (1998). Acidic pH promotes dimerization of Bcl-2 family proteins. *Biochemistry* **37**(18), 6410-6418.

Zhang, W. X., Smithgall, T. E. & Gmeiner, W. H. (1998). Self-association and backbone dynamics of the Hck SH2 domain in the free and phosphopeptide-complexed forms. *Biochemistry* **37**(20), 7119-7126.

Chapter 7

Discussion

In this thesis the investigation into the electrostatic and dynamic properties of rat CD2d1 is described. It also includes the preliminary work necessary for the analyses such as cloning and optimisation studies for protein expression, as well the complete assignment of ^{1}H, ^{15}N and ^{13}C resonances of CD2d1 based on a uniformly labelled protein sample. The ionisation constants of all the acidic residues of rat CD2d1 were determined by monitoring the pH dependent chemical shift of ^{1}H, ^{13}C and ^{15}N resonances, in particular the ^{13}C carbonyl carbon resonances. The manifestation of the effect of these titrating groups on the other residues were examined by a detailed analysis of the chemical shift changes of the backbone amide nitrogen and proton, as well as aliphatic carbon and proton, resonances. The dynamic properties of the protein were also examined by a detailed investigation into the relaxation parameters which produced evidence that has general importance in the analysis of relaxation data. The association states of CD2d1 were determined by the relaxation analysis as well as other NMR and biophysical techniques. In this chapter, a synthesis of some of the findings is attempted and the implications of the results are discussed.

Protein electrostatics and CD2d1 dimerisation

The nature of electrostatic interaction between proteins has long been a subject of study (Honig & Nicholls, 1995). There are two competing forces at work in charge-charge interaction during the binding of a protein to its ligand and other proteins, or self-association: the effect of the charge desolvation penalty that comes from the loss of favourable interaction between the protein and the solvent which acts against interaction between the two molecules, while the formation of favourable electrostatic interaction between the charges in the protein may promote association. The unfavourable desolvation penalty increases with the number of charges. It has therefore been suggested that net effect of electrostatics interaction is generally to destabilise protein binding due to large desolvation penalties. Charge complementarity on the contact surface of two interacting proteins, though important and necessary, primarily defines specificity without favourably contributing to the stability of the complex (Hendsch & Tidor, 1994; Novotny & Sharp, 1992). The role played by the charged groups is particularly well-characterised for immunoglobulin in the protein-protein interaction between antibodies and antigens (Gibas *et al.*, 1997; Novotny & Sharp, 1992; Wibbenmeyer *et al.*, 1999). In antibodies, only a small number of residues within the same structural framework may change to bind an immense variety of antigens. Such high selectivity may originate to a large extent from specific electrostatic interaction governed by polar and charged groups. The main driving force for protein-protein interaction, however, is proposed to be the hydrophobic effect which provide the major stabilisation factor of complex formation, while the Van der Waals interaction or shape comlementarity factor would penalise intermolecular contacts that induced strain (Novotny & Sharp, 1992).

For adhesion molecules, the binding interaction is significantly different from that of antibody to antigen. As previously described, adhesion molecules such as CD2 may have much lower binding affinity compared to antibody-antigen interaction with very rapid binding kinetics (van der Merwe & Barclay, 1994). The weak binding and very fast binding kinetics may be partially explained by the relatively small binding surface that lacks surface shape complementarity (Davis *et al.*, 1998a). In contrast to the antibody-antigen interaction which can show pronounced surface shape complementarity (Davies *et al.*, 1990), the binding surface of CD2d1 is flat

with little surface complementarity. Compared to other protein-protein binding sites, the CD2 binding surface area (~700 $Å^2$) is also relatively small (Jones & Thornton, 1995; LoConte *et al.*, 1999). Such smaller interface may result in short-lived and low stability complexes (LoConte *et al.*, 1999), for example as seen in the complex between cytochrome *c* and cytochrome peroxidase (Pelletier & Kraut, 1992).

Another factor that may determine the binding properties of CD2 is the highly-charged character of its binding surface. Charged residues account for 35% of the rat CD2 and 70% for human CD2 binding surface compared to the average of 15% of other protein-protein interaction sites (LoConte *et al.*, 1999). The presence of the charged residues may be especially important in determining the binding kinetics of adhesion molecules, and previous analysis have shown that transient protein complexes contained more hydrophilic residues on the interfaces than the permanent complexes (Jones & Thornton, 1996). Many molecules involve in cell-cell recognition have significantly polar or charged binding surface, for example, the ligand-binding sites of peripheral nerve myelin P0 (Shapiro *et al.*, 1996), N-cadherin (Shapiro *et al.*, 1995) and T-cell receptor (Garboczi *et al.*, 1996) are hydrophilic in character, with 83%, 71% and 68% of residues forming their respective ligand binding sites being polar or charged. The studies on CD2 may therefore have a more general application.

For rat CD2, the roles played by the charged and hydrophobic residues on the binding surface accord with the canonical views as exemplified by the antibody-antigen interaction. The charged residues are postulated to be primarily responsible for the specificity of CD2 ligand-binding without a concomitant increase in binding strength, thereby maintaining the low affinity necessary for the transient and reversible interaction of T-cell with other cells (Davis *et al.*, 1998a), while the binding energy is largely derived from hydrophobic contact. A different interpretation, however, is proposed for the interaction between human CD2 and CD58. From the crystal heterodimeric structure, the binding surface is shown to be dominated by interdigitating charge-charge interactions, and the binding energy is proposed to be largely derived from electrostatic interaction with hydrophobic interaction playing only a minor role (Wang *et al.*, 1999). The interdigitation of the charged residues are suggested to contribute to the binding energy by neutralising unfavourable interaction between like-charges in each binding surface. Interactions

among charged and polar groups at the protein interface can also result in large energetic contributions to protein association (Buckle *et al.*, 1994). For example, some intermolecular salt bridges have been shown to contribute significantly to the energetics of the antibody/antigen interaction (Wibbenmeyer *et al.*, 1999). While both rat and human CD2 have highly hydrophilic binding surface, human CD2 is considerably more hydrophilic than rat CD2. When the surface is very hydrophilic, the driving force for protein association may be ionic in character and both hydrophobic and hydrophilic interactions may work cooperatively with each other.

In the work presented in this thesis, the binding interaction of CD2 is largely examined in the context of dimer formation and focused on the charged residues. Many protein-protein interactions have been shown to be driven by a small subset of residues on the binding surface (Clackson & Wells, 1995; Dall'Acqua *et al.*, 1998). Previous studies have shown that the 'structural epitopes' of rat CD2 consist of eight charged or polar residues (Asp28, Glu29, Arg31, Glu41, Lys43, Tyr86 and Arg87), two aromatic residues (Tyr81, Phe49) and a single aliphatic residue, Leu38 (Davis *et al.*, 1998b). Of these, Asp28, Glu29, Arg31, Leu38, Phe49 and Tyr81 are particularly important with 20-fold or greater decreases in ligand binding affinity for alanine mutants of these residues. Asp28, Glu29 and Arg31 were proposed to be important primarily for conferring specificity while the hydrophobic residues, Tyr81, Leu38 and Phe49, provide the binding energy. Analysis of the dimer interface of CD2 crystal structure, however, suggests that the binding interaction in CD2 dimer may be different in crucial aspects from CD2/CD48 interaction.

From Fig. 8.1, it can be seen the dimer interface is largely populated by charge-charge interaction with only limited hydrophobic contact. There are 12 hydrogen bonds and salt bridges, of particular importance are the four intermolecular salt bridges formed between Glu29 and Arg31, and between Glu33 (a residue not significant in CD2/CD48 interaction) and Arg87. Also observed is the intermolecular hydrogen-bonding between the backbone amide of Gly85 and carbonyl of Leu38 that may be particularly significant for V-set dimerisation (Chothia *et al.*, 1985). There are four regions of hydrophobic contacts, formed by the $C^{\alpha}H_2$ of Gly85 of one monomer to the aromatic ring of Phe49 of the other, and the $C^{\varepsilon}H_3$ of Met46 to the $C^{\varepsilon}H_2$ of Lys43. It can therefore be seen that, of the three

hydrophobic residues suggested to provide binding energy in CD2/CD48 interaction, only Phe49 is involved in hydrophobic contact in the dimer.

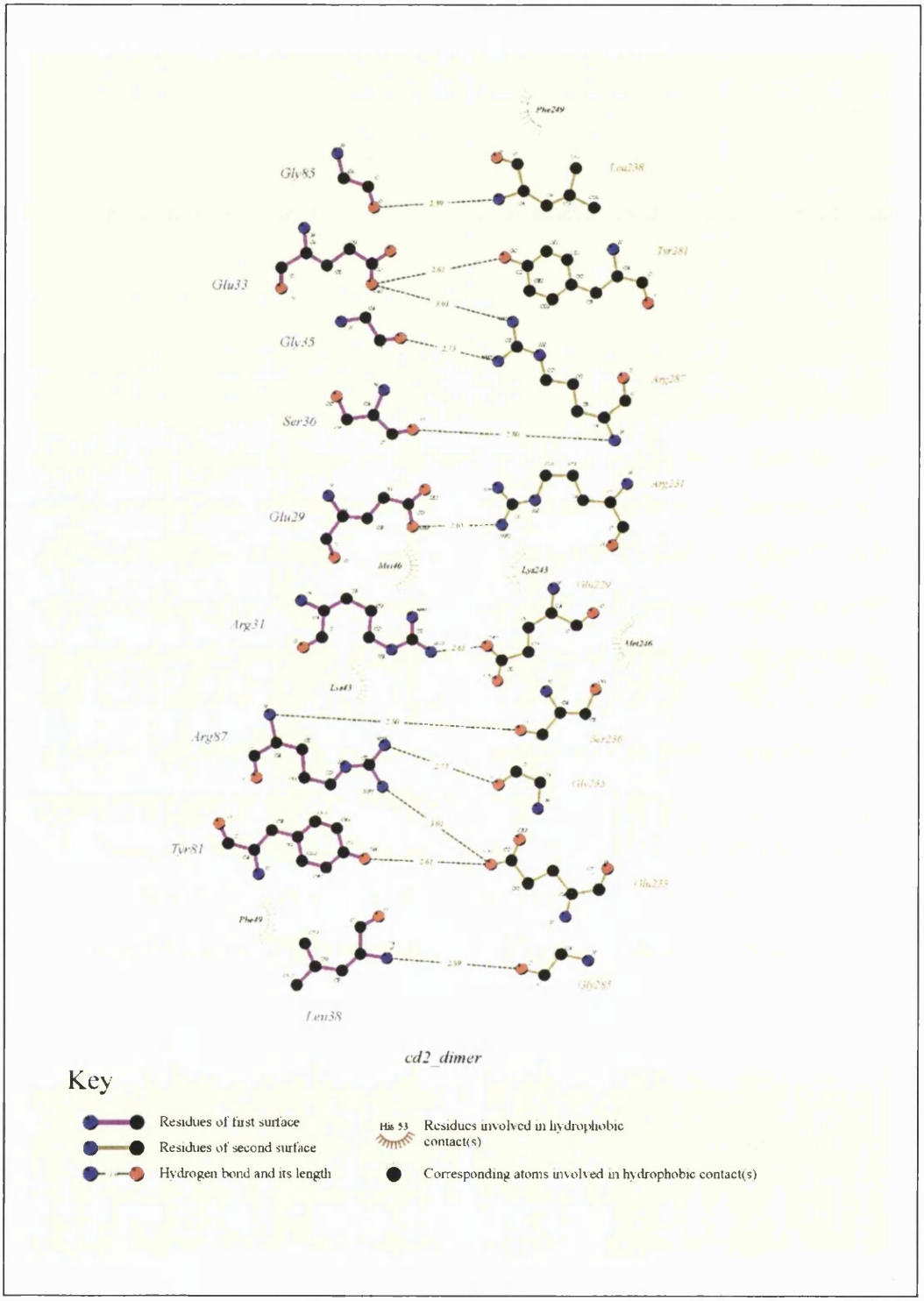


Fig. 8.1 Residues on the CD2 dimer interface analysed using Ligplot (Wallace et al., 1995) based on the rat CD2 crystal structure (Jones et al., 1992). The residues of one CD2 are number from 1-199, while the other is numbered from 201-399. The hydrogen-bonds and salt bridges and their bond lengths are indicated. Hydrophobic contact is indicated by spoked arc.

The few hydrophobic contacts that exist lie at the periphery of the contact interface of CD2 dimer, with the charge-charge interactions populating the centre of the binding surface. This paucity and the peripheral nature of hydrophobic contacts suggest that charge-charge interaction plays the dominant role in mediating homodimeric interaction. For proteins in which the associations with other proteins are mediated by charged residues, the energetics are strongly pH-dependent and assessing the ionisation states of titratable residues at the ligand-binding surface can therefore reveal much about the binding interactions.

The analysis of the ionisation states of acidic residues showed that Glu41 has an anomalously high pK_a of 6.7. Furthermore, the protonation state of this residue as well as that of Glu29 affects the self-association of CD2d1, and maximal dimeric interaction is achieved when Glu41 is protonated and Glu29 deprotonated. Changes in protonation states can influence the binding interaction of protein, for example in the binding of protein antigens to antibodies (Gibas *et al.*, 1997) and the binding of cofactors to enzymes (Cannon *et al.*, 1997). This CD2d1 self association can be rationalised, in part, to the intra-molecular interaction between Glu29 and Glu41 which reduces the desolvation penalties associated with intermolecular interaction as desolvation cost of the charged group on the binding sites can have strong effect on the binding interaction (Yang & Honig, 1993; Zeng *et al.*, 1999). Intramolecular electrostatic interactions can therefore enhance molecular association by reducing the effective dielectric constant on binding (Chong *et al.*, 1998).

However, the results obtained showed that CD2d1 dimerises with very low affinity, with a dissociation constant of the order of ~3-6 mM which is ~30 -100 times weaker than CD2/CD48 interaction. The results also indicated that the greatly reduced affinity of CD2 homodimerisation is due to a much lower k_{on}, indicating that very few of the random collisions between CD2d1 resulted in productive binding. The limited number of significant hydrophobic contacts may partly contribute to the low binding affinity, however, apart from the flatness of the binding surface, there is no obvious adverse shape complementary factor that can penalise dimeric contact. There are also no obvious unfavourable charge-charge interactions as the charges are in fairly complementary positions, and strict charge-charge complementarity is in any case not necessary (McCoy *et al.*, 1997). There are four acidic residues: Asp28, Glu29, Glu33, and Glu41; and three basic residues: Lys43 and two arginines, Arg31

and Arg87, which form intermolecular salt bridges. Arginine occur relatively frequently in interface, and may reflect its comparatively large number of hydrogen-bonding sites (McCammon, 1998). Asp28 lies on the periphery and not directly engaged on the binding surface but form intramolecular interaction with Lys43; it may, however, play a significant role in the binding as charged residues not directly on the binding site can affect the binding affinity (Hawkins *et al.*, 1993). Glu29 and Glu33 form salt bridges with Arg31 and Arg87 respectively, which leave Glu41 as the carboxylate not directly forming intermolecular salt-bridge, but the nearby Lys43 may be a potential partner. Glu41, however, forms hydrogen bond with Glu29 when protonated and is likely to be uncharged on the dimer interface.

The increase in pK_a of Glu41 upon dimer formation suggests that the energetic contribution from Glu41 is likely to be negative as the protonated form is favoured in the dimer interface. The pK_a shift may indicate energetic contribution of a particular charged residue on the stability of the binding interaction. The charge on Glu41 is therefore unfavourable to homodimeric interaction which can also partly explain the low binding affinity in CD2 dimer formation. However, even at optimal binding condition, the binding affinity is still very low. It is possible that the arrangement of the side chains of residues such as Glu41 and Glu29 on the binding surface that is necessary for dimeric interaction is energetically unfavourable. The crystal structure may be taken as the conformation adopted by CD2d1 as it dimerises, however, the disposition of the side chains as well as backbone in the monomer may show some differences. For human CD2, the binding surface of the monomer is suggested to fluctuate rapidly between two different conformation, one of which is favourable for the heterophilic interaction with CD58 (Kitao & Wagner, 2000). It is not known if such rapid fluctuation is present for rat CD2, however, slower motion of the EF and FG loops are observed that may have implications for the binding interaction of CD2. For CD2 dimerisation, further subtle changes may be necessary for dimer formation and such conformation changes may be unfavourable. Minor structural changes have been shown to significantly affect the binding affinity in other complexes (Chacko *et al.*, 1995; Tulip *et al.*, 1994).

Structural rearrangement of the surface residues may also be observed at varying pH which can contribute to the change in binding affinity. The side chains of the charged residues on the binding surface lacks mobility as indicated by the

heteronuclear NOE of the arginine side chains as well as the non-degeneracy of the γ-CH_2 protons of Glu41 and Glu29. The side chains therefore may be held by a complex network of hydrogen bonds and salt bridges as such network of salt bridges is stabilising (Kumar & Nussinov, 1999). Changes in the protonation state of the charged residues therefore requires the rearrangement of this network of salt bridges and hydrogen bonding resulting in changes in the disposition of the side chains. Some perturbation of the resonances due to the anomalous pK_a of Glu41 may be observed part of D strand, perhaps indicating that subtle conformation changes on dimer formation is promulgated across from the binding surface to ABED sheet. An interesting observation is the degeneracy of resonances $C^{\beta}H$ of Asp28 at the pH at which Glu41 is protonated and Glu29 deprotonated (i.e. when there is maximal dimer formation), which suggests free rotation of the side chain. Whereas at higher and lower pH the two protons resonances are distinct implying restricted motion. In the crystal structure, however, Asp28 forms a salt bridge with Lys43 which should restrict somewhat the movement of Asp28, whether this indicates there may be subtle differences between the crystal structure and dimer in solution remains to be resolved.

Apart from Glu41, the pK_as of most acidic residues, such as that of Glu29, are little changed on dimer formation, suggesting that the energetic contribution of these residues are largely neutral. The intermolecular charge-charge interaction may compensate for intramolecular charge-charge interaction and desolvation cost. The role of Glu29 is therefore largely for specificity but whose charge is highly important for CD2 binding interaction as protonation or mutation of this residue abolishes CD2d1 dimerisation. Glu33 showed a slight drop of 0.15, suggesting its intermolecular salt bridge interaction is somewhat favourable, however, the significance of the salt bridge from Glu33 to Arg87 has not been assessed. This absence of significant effect of these residues can also be observed in the insensitivity of the pK_a as well as the exchange phenomenon to changes in ionic strength and type of buffer used, and this parallel the result from the study of CD2/CD48 interaction which showed the heterophilic interaction to be insensitive to salt concentration (Davis *et al.*, 1998a). If the charge groups contribute significantly to the binding energy between CD2d1 monomer, the electrostatic screening effect at high salt concentration should affect the dimer formation. The results could therefore be

taken as indications that, overall, the contribution of the charged groups are energetically neutral. In the binding interaction of adhesion molecules involving large numbers of electrostatic interactions, the energetics may be finely balanced with favourable interactions counterbalanced by other unfavourable interactions.

CD2/CD48 binding interaction

The protonation of Glu41 upon dimerisation raises an interesting question about its role in CD2 binding interaction. From a comparison of the human and rat CD2 ligand binding sites, the species and ligand specific binding may be determined by just three residues, namely Thr37, Leu38 and Glu41 (McAlister *et al.*, 1996). The involvement of Glu41 in the binding interaction has been tested in a number of mutagenesis studies. Charge reversal mutation of Glu41 to arginine abrogates CD48 binding; while complementary charge reversal mutagenesis of Arg31 residue in CD48 restored the binding capability (van der Merwe *et al.*, 1995). An alanine mutant of Glu41, however, can tolerate drastic substitution at Arg31 position of CD48 and is highly promiscuous in being able to bind to many CD48 mutants (Davis *et al.*, 1998a), similarly the K43A mutant of CD2d1 can bind to a variety CD48 mutants with residue mutated at the Glu44 position (Davis *et al.*, 1998a). These results are taken as indication that the charge on Glu41 has the primary role of conferring specificity to the interaction of CD2 to CD48. However, the anomalously high pK_a of Glu41 suggests that it cannot be taken as granted that Glu41 should be charged. It is useful to note at this point that for E41, the corresponding residue in human CD2 is replaced by a glutamine (Gln46). The Gln46 side chain sits opposite Leu27 on CD58, representing an island of nonpolar contacts surrounded by polar and ionic salt bridge interactions. If the structure of rat CD2/CD48 is highly homologous to human CD2/CD58, the Glu41 of rat CD2 could be sited opposite Thr33 of CD48, and only obliquely juxtaposed with the proposed salt bridge partner Arg31 on CD48.

A previous study, however, showed a marked reduction in the binding of CD2 to CD48 below pH ~ 6 (McAlister *et al.*, 1996). The loss of binding begins at ~ pH 6.5 and with complete loss of binding at pH below 4, thereby implicating the presence of one or more ionisable residues on the binding interface that mediate the

binding interaction. The anomalous pK_a of Glu41 suggests that it may be the candidate. The loss of binding when the pH is lowered may suggest that protonation of Glu41, in contrast to the case of CD2d1 dimerisation, is unfavourable for complex formation with CD48. However, as previously described, it is difficult to assess the protonation states of Glu41 on interaction with CD48 and the pK_as of this and other titratable residues upon CD48 binding is yet to be determined. Furthermore, two histidines (His36 and His90) are also present on the putative binding surface of CD48. Mutagenesis studies of the binding surface of CD48 (van der Merwe *et al.*, 1995) in which the histidine was substituted for an acidic residue showed no loss of binding for the H90D mutant, while the H36E mutant showed reduced binding, suggesting a possible role in the binding interaction for the His36. More work is required in order to assess the importance and relative significance of these titrating residues on the binding properties of these adhesion molecules.

For the relaxation analysis of the mutants, the charge on Glu29 is shown to be more important for CD2d1 dimerisation, this corresponds with studies on CD2-CD48 binding which showed loss of binding for the various Glu29 mutants while alanine mutant of Glu41 showed essentially unchanged binding affinity. There is as yet no structure of the rat CD48 or CD2/CD48 complex, it is therefore not known precisely how the charged groups might interact. As binding studies also suggest that CD2/CD48 binding interaction may closely mirrored that of the CD2 dimer interaction, it may be possible to examine how the charge groups interaction by comparison to the crystal dimer structure. However, it may not simple to directly model CD2/CD48 interaction using the CD2 dimer. For example, Glu29 form intermolecular interaction with Arg31 in CD2 dimeric structure, but from sequence alignment, the equivalent position of Arg31 in rat CD48 is replaced by a threonine. For Glu41 the nearest intermolecular basic residue is Lys43 from the crystal dimer. And from complememtary mutagenesis, Lys43 interacts with Glu44 of CD48 which is the equivalent of Glu41, however, Glu41 of CD2 does not interact with the equivalent residue of Lys43 in CD48 (which is replaced by a proline), but instead interact with Arg31 of CD48 (which is equivalent to Glu29 in CD2). Glu33 interacts with Arg87 and Tyr81 in CD2 dimer, but there is no positive charge in the equivalent position of Arg87 in CD48 and the tyrosine is replaced by leucine. The topology of interaction between CD2 and CD48 may varies, and indeed, the crystal structures of

CD2 dimer also suggest that the orientation of interaction may differ with subtly different modes of dimeric interaction possible (Jones *et al.*, 1992). The binding interaction in the homodimer is symmetrical but for the heterophilic interaction, it may be asymmetrical as observed in the CD2/CD58 structure (Wang *et al.*, 1999). Heterodimer model of the rat CD2-CD48 adhesion complex based on the human CD2-CD58 crystal structure has shown less favourable interactions compared CD2-CD58 complex (Wang *et al.*, 1999). Further work, however, need to be done in order to elucidate the heterophilic binding properties of rat CD2.

Biological Significance

The work has demonstrated that CD2d1 dimerised with very low affinity, but such low affinity interaction is nevertheless highly specific. The homodimeric interaction occurs via the GFCC'C'' face of the molecule in the same manner as observed in the crystal structure, and this dimerisation is pH-dependent with a maximal dimer formed at ~ pH 6. Many immunoglobulins are known to dimerise, most notably in antibodies in which tight complexes are formed between two immunoglobulin chains. ICAMs are also thought to exist as dimers (Casasnovas *et al.*, 1997; Casasnovas *et al.*, 1998), while the first two domains of neural CAM (NCAM) has also been shown to dimerised, however, the dimerisation occurs via interaction between the different domains (Atkins *et al.*, 1999; Jensen *et al.*, 1999) Other immunoglobulins that form homodimers includes human CD4 which dimerises with mM affinity (Wu *et al.*, 1997), hemolin which dimerises via both faces of the β sandwich (Su *et al.*, 1998). N-cadherin (Shapiro *et al.*, 1995) which showed structure similar to V-type Ig domain, as well as E-cadherin (Takeda *et al.*, 1999) and P_0 (Shapiro *et al.*, 1996) also interact via homodimerisation.

The immunoglobulins may dimerise in *cis* or *trans* mode, i.e. the homodimerisation may be between adhesion molecules on the same cell or between different cells. For ICAM-1, the interaction is proposed to occur via *cis* orientation on the BED surface, while NCAM can dimerised in both *cis* and *trans* orientation on the ABED face. For CD2, the *trans* mode of interaction may be possible between the two T-cells, however, the *cis* interaction is suggested to be inhibited by the glycosylations on the ectodomain which orientate the binding surface upward and

may also prevent non-specific aggregation. Nevertheless, the linker region between the two extracellular domain is flexible and *cis* interaction between two adjacent CD2 molecules is therefore possible.

It is difficult to ascribe a physiological role for the pH dependent self-association of CD2d1 as the binding affinity is in the millimolar range. However, at the adhesion contact region, there is a high local concentration of CD2, such low affinity interaction therefore may have some physiological relevance. The self-association may explain some observed phenomena that are currently unsatisfactorily explained. For example, the emergence of a neo-epitope termed the CD2R epitope during T-cell activation upon ligation by CD58 (Dustin *et al.*, 1987). This CD2R epitope is also expressed when the number of cell surface CD2 receptors reached a threshold, and it can be induced by cross-linking of adjacent receptors (Li *et al.*, 1996). This neo-epitope has been mapped to the interdomain linker region and a contiguous region in CD2 domain 2, and its emergence has been ascribed to conformational changes involving interdomain reorientation and close-packing of the receptors forming molecular lattice (Li *et al.*, 1996). Such close-packing of CD2, however, has yet to be observed *in vivo*, and the presence of the glycans may preclude such lattice formation of the receptors. Moreover, the formation of the immunological synapse requires the dynamic interaction of the TCR and peptide-MHC molecules as well as CD2 to CD48/CD58, a fact not compatible with the formation of a stable lattice. It is therefore interesting to speculate if the low affinity *cis* homodimeric interaction can result in the appearance of the interdomain neo-epitope by inducing the conformational changes necessary to expose the interdomain surface. While noting the possibility that glycans may prevent such *cis* interaction, the low affinity interaction can better explain the observed phenomenon, for example, the fact that CD2R epitope is observed at high concentration of cell surface CD2 receptors, and permit the CD2 to form clusters of interacting receptors. The rapid dissociation of CD2 dimer also means that the binding surface is not sequestered by such interaction and is readily available for heterophilic interaction with CD58 and its low affinity would not interfere with the higher affinity heterophilic interaction. The CD2 dimerisation also allows for possible *trans* mode of homophilic interactions between T-cells that only become significant when there is a high local concentration of CD2.

It is not clear whether the conformation changes of CD2 has any relevance to the functioning of CD2. TCR has been postulated to undergo conformational changes upon peptide/MHC ligation (Sakihama *et al.*, 1995) while conformational changes have also been implicated in other cell adhesion molecules such as integrins, however, the mechanisms of such changes is unclear. For CD2, the emergence of the CD2R epitope coincides with the changes in avidity of CD2 observed upon T-cell activation (Dustin *et al.*, 1987), however, it is not clear how the two events may be related. CD2R expression is also enhance in certain rheumatoid disease (Potocnik *et al.*, 1991), and it is also interesting to note that CD58 is also upregulated in inflammatory sites (Dengler *et al.*, 1992) suggesting that such changes may facilitate the entry of T-cell into inflamed tissues. The change in reactivity may also be produced by infection with oncogenic virus *H. saimiri* (Meinl & Hohlfeld, 2000) as the infected T-cells can be activated by a single crosslinked anti-CD2 mAb instead of the pairs of mAbs normally required for activation. However, there is as yet no satisfactorily explanation for the change in avidity. The change in avidity of human CD2 has been proposed to involve the formation of molecular lattice as described earlier (Li *et al.*, 1996). The increase in avidity has also been ascribed to tethering of CD2 to the cytoskeleton (Davis *et al.*, 1998b), however, tethering merely permits a higher affinity interaction between CD2 and its ligands, and does not determine its binding affinity. Other possibilities raised include the unfolding of CD2 and the formation of intercalated dimer at high protein concentration (Murray *et al.*, 1995), or the production of cytoplasmic factors. The pH dependence of the binding behaviour, however, suggests an intriguing possibility for the regulation of the binding property of adhesion molecules.

The activation of T-cell via CD2 is accompanied by a rapid increase in intracellular pH, an effect also seen in stimulation with alloantigens or lectins and only occasionally seen in activation via CD3 (Fischer *et al.*, 1988). The pH change is mediated by the Na^+/H^+ antiporter which increases the cytoplasmic pH by pumping out protons. The same effect may be observed to greater effect by activation via integrins which regulates initial events in focal adhesion and cytoskeletal reorganisation via the Na^+/H^+ antiporter NHE1 (Schwartz *et al.*, 1991, Tominaga & Barber, 1998). The adhesion interaction of integrins and CD2 may be increased upon T-cell activation via an 'inside-out' mechanism. The NHE1 is localised at sites of

the focal contact and it is therefore interesting to speculate if the action of NHE1 may result in transient acidification of the extracellular space, thereby regulating the action of the adhesion molecules by changes in pH. Slight acidification of blood serum has been observed upon activation of T-cells via CD2, however, such small changes is unlikely to affect in any way the binding affinity of CD2. Within a confined space, however, the change in pH may be significant. For example, within inflamed tissues, the pH may be lowered significantly. However, it has yet to be shown that significant local changes in pH can be achieved in the immunological synapse that may influence the binding affinity of the cell adhesion molecules. If local variations in pH may be controlled, the binding interaction of adhesion molecules can be modulated which may facilitate the attachment to as well as the detachment from the target cells by T-cell.

Scope for further study

CD2 is an extremely well-defined molecule with interesting properties that render it useful for biophysical studies. Its small size and good solution properties make it particularly well suited for analysis by NMR. The many mutants that are available also allow the various aspect of the molecule to be dissected and analysed in detail, for instance, in its binding behaviour and folding properties (Lorch *et al*, 2000; Murry *et al.*, 1998).

In this thesis, the effect of low affinity self-association which has significant implication for relaxation analysis was examined. Relaxation is an important part of NMR and it is therefore of vital importance that the factors that affect its analysis, such as protein self-association, should be carefully delineated. The various mutants of CD2 which shown varying propensity for self-association and these mutants can therefore provide further opportunity for detailed analysis of the relaxation parameters, as well providing useful information on the role played by the various residues on the binding interaction. However, thus far, only two of the mutants are examined in details. The properties that make CD2 such interesting molecule also complicates its analysis. For example, mutation of the surface molecules affects not only the binding characteristics and aggregational properties, but it can also have profound effect on the structural integrity and may induce structural perturbations,

making interpretation of the results difficult. A careful examination of the various elements in the structural and physical properties of CD2 would therefore be invaluable.

For a molecule such as CD2 in which electrostatic properties play a significant role in its function, it is necessary to carefully study the effect of the properties of the charge residues may impinge on the binding behaviour. However, the large number of charges present results in complex electrostatic interactions that are not easy to unravel. For the reasons described above, analysis of the mutants has thus far been limited. An alternative approach may be using electrostatic modelling which can provide further insight into the electrostatic of the protein, but analysis of CD2d1 thus far had yielded unsatisfactory results. Improved algorithms and methodologies for theoretical modelling of electrostatics, however, are available. These analytical methods allow for the inclusion of the energetic contribution from the solvent and hydration layer, optimisation of hydrogen-bonding network, and consideration of the mobility of side-chains and the electrostatic microenvironments of the charged groups (Beroza & Case, 1996; Havranek & Harbury, 1999; Mehler & Guarnieri, 1999; Nielsen *et al.*, 1999; vanVlijmen *et al.*, 1998). Calculations, however, should be performed on a structure which may better reflect the structure of a monomeric CD2d1, rather than relying on that of the dimeric crystal structure. A further refinement of the solution structure of CD2 would therefore contribute immeasurably to the understanding of its function.

The ultimate aim in the structural analysis of CD2 is to understand the biophysical basis for the adhesion properties of CD2 in its interaction with its physiological partner, CD48. The analysis suggests that subtle changes can significantly affect the binding. However, the studies are hampered considerable by the absence of a form CD48 that is amenable for NMR analysis as severe broadening of the resonance with the attendant difficulty in studying any observed effect render any detailed analysis impossible. It is therefore imperative to reclone the CD48 such that an unglycosylated and more tractable form of CD48 can be produced. It may be instructive in this case to examine the production of the soluble form of human CD2 and CD58 (Sun *et al.*, 1999). As CD48 is homologous to CD2 and CD58, it may be possible to model the structure which may then be used to create mutant that favours formation of soluble monomer. By judicious application of SDM of specific

residues, it may be possible to increase the solubility and/or stability of the protein (Sun *et al.*, 1999). There is also, as yet, no structure of CD48 or CD2/CD48 complex; a soluble and tractable form of CD48 would therefore permit the determination of its structure which would contributes significantly to the understanding of adhesion events on the T-lymphocytes.

References

Atkins, A. R., Osborne, M. J., Lashuel, H. A., Edelman, G. M., Wright, P. E., Cunningham, B. A. & Dyson, H. J. (1999). Association between the first two immunoglobulin-like domains of the neural cell adhesion molecule N-CAM. *Febs Letters* **451**(2), 162-168.

Beroza, P. & Case, D. A. (1996). Including side chain flexibility in continuum electrostatic calculations of protein titration. *Journal of Physical Chemistry* **100**(51), 20156-20163.

Brown, L. S., Kamikubo, H., Zimanyi, L., Kataoka, M., Tokunaga, F., Verdegem, P., Lugtenburg, J. & Lanyi, J. K. (1997). A local electrostatic change is the cause of the large-scale protein conformation shift in bacteriorhodopsin. *Proceedings of the National Academy of Sciences of the United States of America* **94**(10), 5040-5044.

Buckle, A. M., Schreiber, G. & Fersht, A. R. (1994). Protein-Protein Recognition - Crystal Structural-Analysis of a Barnase B arstar Complex At 2.0-Angstrom Resolution. *Biochemistry* **33**(30), 8878-8889.

Cannon, W. R., Garrison, B. J. & Benkovic, S. J. (1997). Electrostatic characterization of enzyme complexes: Evaluation of the mechanism of catalysis of dihydrofolate reductase. *Journal of the American Chemical Society* **119**(10), 2386-2395.

Casasnovas, J. M., Springer, T. A., Liu, J. H., Harrison, S. C. & Wang, J. H. (1997). Crystals structure of ICAM-2 reveals a distinctive integrin recognition surface. *Nature* **387**(6630), 312-315.

Casasnovas, J. M., Stehle, T., Liu, J. H., Wang, J. H. & Springer, T. A. (1998). A dimeric crystal structure for the N-terminal two domains of intercellular adhesion molecule-1. *Proceedings of the National Academy of Sciences of the United States of America* **95**(8), 4134-4139.

Chacko, S., Silverton, E., Kammorgan, L., Smithgill, S., Cohen, G. & Davies, D. (1995). Structure of an Antibody Lysozyme Complex Unexpected Effect of a Conservative Mutation. *Journal of Molecular Biology* **245**(3), 261-274.

Chong, L. T., Dempster, S. E., Hendsch, Z. S., Lee, L. P. & Tidor, B. (1998). Computation of electrostatic complements to proteins: A case of charge stabilized binding. *Protein Science* **7**(1), 206-210.

Chothia, C., Novotny, J., Bruccoleri, R. & Karplus, M. (1985). Domain Association in Immunoglobulin Molecules - the Packing of Variable Domains. *Journal of Molecular Biology* **186**(3), 651-663.

Clackson, T. & Wells, J. A. (1995). A Hot-Spot of Binding-Energy in a Hormone-Receptor Interface. *Science* **267**(5196), 383-386.

Dall'Acqua, W., Goldman, E. R., Lin, W. H., Teng, C., Tsuchiya, D., Li, H. M., Ysern, X., Braden, B. C., Li, Y. L., SmithGill, S. J. & Mariuzza, R. A. (1998). A mutational analysis of binding interactions in an antigen-antibody protein-protein complex. *Biochemistry* **37**(22), 7981-7991.

Davies, D. R., Padlan, E. A. & Sheriff, S. (1990). Antibody-Antigen Complexes. *Annual Review of Biochemistry* **59**, 439-473.

Davis, S. J., Davies, E. A., Tucknott, M. G., Jones, E. Y. & van der Merwe, P. A. (1998a). The role of charged residues mediating low affinity protein-protein recognition at the cell surface by CD2. *Proceedings of the National Academy of Sciences of the United States of America* **95**(10), 5490-5494.

Davis, S. J., Ikemizu, S., Wild, M. K. & van der Merwe, P. A. (1998b). CD2 and the nature of protein interactions mediating cell-cell recognition. *Immunological Reviews* **163**, 217-236.

Dengler, T. J., Hoffmann, J. C., Knolle, P., Albertwolf, M., Roux, M., Wallich, R. & Meuer, S. C. (1992). Structural and Functional Epitopes of the Human Adhesion Receptor- CD58 (LFA-3). *European Journal of Immunology* **22**(11), 2809-2817.

Dustin, M. L., Selvaraj, P., Mattaliano, R. J. & Springer, T. A. (1987). Anchoring Mechanisms For Lfa-3 Cell-Adhesion Glycoprotein At Membrane-Surface. *Nature* **329**(6142), 846-848.

Evans, E., Berk, D. & Leung, A. (1991). Detachment of Agglutinin-Bonded Red-Blood-Cells .1. Forces to Rupture Molecular-Point Attachments. *Biophysical Journal* **59**(4), 838-848.

Fischer, G. F., Holter, W., Majdic, O., Cragoe, E. J. & Knapp, W. (1988). T-Cell Stimulation Via Cd2 Molecules Is Regularly Accompanied By an Increase in Cytoplasmic Ph - Different Effects of Lectins and Cd3 Antibodies. *Journal of Immunology* **141**(2), 404-409.

Garboczi, D. N., Ghosh, P., Utz, U., Fan, Q. R., Biddison, W. E. & Wiley, D. C. (1996). Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. *Nature* **384**(6605), 134-141.

Gibas, C. J., Subramaniam, S., McCammon, J. A., Braden, B. C. & Poljak, R. J. (1997). pH dependence of antibody/lysozyme complexation. *Biochemistry* **36**(50), 15599-15614.

Havranek, J. J. & Harbury, P. B. (1999). Tanford-Kirkwood electrostatics for protein modeling. *Proceedings of the National Academy of Sciences of the United States of America* **96**(20), 11145-11150.

Hawkins, R. E., Russell, S. J., Baier, M. & Winter, G. (1993). The Contribution of Contact and Noncontact Residues of Antibody in the Affinity of Binding to Antigen - the Interaction of Mutant D1.3 Antibodies With Lysozyme. *Journal of Molecular Biology* **234**(4), 958-964.

Hendsch, Z. S. & Tidor, B. (1994). Do Salt Bridges Stabilize Proteins - a Continuum Electrostatic Analysis. *Protein Science* **3**(2), 211-226.

Honig, B. & Nicholls, A. (1995). Classical Electrostatics in Biology and Chemistry. *Science* **268**(5214), 1144-1149.

Jensen, P. H., Soroka, V., Thomsen, N. K., Ralets, I., Berezin, V., Bock, E. & Poulsen, F. M. (1999). Structure and interactions of NCAM modules 1 and 2, basic elements in neural cell adhesion. *Nature Structural Biology* **6**(5), 486-493.

Jones, E. Y., Davis, S. J., Williams, A. F., Harlos, K. & Stuart, D. I. (1992). Crystal-Structure At 2.8-Angstrom Resolution of a Soluble Form of the Cell-Adhesion Molecule CD2. *Nature* **360**(6401), 232-239.

Jones, S. & Thornton, J. M. (1995). Protein-Protein Interactions - a Review of Protein Dimer Structures. *Progress in Biophysics & Molecular Biology* **63**(1), 31 et seq.

Jones, S. & Thornton, J. M. (1996). Principles of protein-protein interactions. *Proceedings of the National Academy of Sciences of the United States of America* **93**(1), 13-20.

Kitao, A. & Wagner, C. (2000). A space-time structure determination of human CD2 reveals the CD58- binding mode. *Proceedings of the National Academy of Sciences of the United States of America* **97**(5), 2064-2068.

Kumar, S. & Nussinov, R. (1999). Salt bridge stability in monomeric proteins. *Journal of Molecular Biology* **293**(5), 1241-1255.

Laskowski, R. A. (1995). Surfnet - a Program For Visualizing Molecular-Surfaces, Cavities, and Intermolecular Interactions. *Journal of Molecular Graphics* **13**(5), 323 et seq.

Li, J., Smolyar, A., Sunderplassmann, R. & Reinherz, E. L. (1996). Ligand-Induced Conformational Change Within the CD2 Ectodomain Accompanies Receptor Clustering - Implication For Molecular Lattice Formation. *Journal of Molecular Biology* **263**(2), 209-226.

LoConte, L., Chothia, C. & Janin, J. (1999). The atomic structure of protein-protein recognition sites. *Journal of Molecular Biology* **285**(5), 2177-2198.

Lorch, M., Mason, J. M., Sessions, R. B., Clarke, A. R. (2000). Effects of mutations on the thermodynamics of a protein folding reaction: implications for the mechanism of formation of the intermediate and transition states. Biochemistry **39**(12), 3480-3485.

McAlister, M. S. B., Mott, H. R., van der Merwe, P. A., Campbell, I. D., Davis, S. J. & Driscoll, P. C. (1996). NMR Analysis of Interacting Soluble Forms of the Cell-Cell Recognition Molecules CD2 and CD48. *Biochemistry* **35**(19), 5982-5991.

McCammon, J. A. (1998). Theory of biomolecular recognition. *Current Opinion in Structural Biology* **8**(2), 245-249.

McCoy, A. J., Epa, V. C. & Colman, P. M. (1997). Electrostatic complementarity at protein/protein interfaces. *Journal of Molecular Biology* **268**(2), 570-584.

Mehler, E. L. & Guarnieri, F. (1999). A self-consistent, microenvironment modulated screened Coulomb potential approximation to calculate pH-dependent electrostatic effects in proteins. *Biophysical Journal* **77**(1), 3-22.

Meinl, E. & Hohlfeld, R. (2000). T cell transformation with Herpesvirus saimiri: a tool for neuroimmunological research. *Journal of Neuroimmunology* **103**(1), 1-7.

Murray, A. J., Head, J. G., Barker, J. J., Brady, R. L., (1998). Engineering an intertwined form of CD2 for stability and assembly. *Nature Structure Biology* **5**(9), 778-782.

Murray, A. J., Lewis, S. J., Barclay, A. N. & Brady, R. L. (1995). One Sequence, 2 Folds - a Metastable Structure of CD2. *Proceedings of the National Academy of Sciences of the United States of America* **92**(16), 7337-7341.

Nielsen, J. E., Andersen, K. V., Honig, B., Hooft, R. W. W., Klebe, G., Vriend, G. & Wade, R. C. (1999). Improving macromolecular electrostatics calculations. *Protein Engineering* **12**(8), 657-662.

Novotny, J. & Sharp, K. (1992). Electrostatic Fields in Antibodies and Antibody Antigen Complexes. *Progress in Biophysics & Molecular Biology* **58**(3), 203-224.

Pelletier, H. & Kraut, J. (1992). Crystal-Structure of a Complex Between Electron-Transfer Partners, Cytochrome-C Peroxidase and Cytochrome-C. *Science* **258**(5089), 1748-1755.

Potocnik, A. J., Menninger, H., Yang, S. Y., Pirner, K., Krause, A., Burmester, G. R., Broker, B. M., Hept, P., Weseloh, G., Michels, H., Truckenbrodt, H., Emmrich, F. & Kroczek, R. A. (1991). Expression of the Cd2 Activation Epitope T11-3 (Cd2r) On T-Cells in Rheumatoid-Arthritis, Juvenile Rheumatoid-Arthritis, Systemic Lupus- Erythematosus, Ankylosing-Spondylitis, and Lyme-Disease - Phenotypic and Functional-Analysis. *Scandinavian Journal of Immunology* **34**(3), 351-358.

Sakihama, T., Smolyar, A. & Reinherz, E. L. (1995). Molecular Recognition of Antigen Involves Lattice Formation Between Cd4, Mhc Class-Ii and Tcr Molecules. *Immunology Today* **16**(12), 581-587.

Shapiro, L., Doyle, J. P., Hensley, P., Colman, D. R. & Hendrickson, W. A. (1996). Crystal structure of the extracellular domain from P-0, the major structural protein of peripheral nerve myelin. *Neuron* **17**(3), 435-449.

Shapiro, L., Fannon, A. M., Kwong, P. D., Thompson, A., Lehmann, M. S., Grubel, G., Legrand, J. F., Alsnielsen, J., Colman, D. R. & Hendrickson, W. A. (1995). Structural Basis of Cell-Cell Adhesion By Cadherins. *Nature* **374**(6520), 327-337.

Su, X. D., Gastinel, L. N., Vaughn, D. E., Faye, I., Poon, P. & Bjorkman, P. J. (1998). Crystal structure of hemolin: A horseshoe shape with implications for hemophilic adhesion. *Science* **281**(5379), 991-995.

Sun, Z. Y. J., Dotsch, V., Kim, M., Li, J., Reinherz, E. L. & Wagner, G. (1999). Functional glycan-free adhesion domain of human cell surface receptor CD58: design, production and NMR studies. *Embo Journal* **18**(11), 2941-2949.

Schwartz, M.A., Ingber, D.E., Lawrence, M., Springer, T. A., Lechene, C. (1991). Multiple integrins share the ability to induce elevation of intracellular pH. *Exp Cell Res.*, **195**(2), 533-5.

Takeda, H., Shimoyama, Y., Nagafuchi, A. & Hirohashi, S. (1999). E-cadherin functions as a cis-dimer at the cell-cell adhesive interface in vivo. *Nature Structural Biology* **6**(4), 310-312.

Tominaga, T. & Barber, D. L. (1998). Na-H exchange acts downstream of RhoA to regulate integrin-induced cell adhesion and spreading. *Molecular Biology of the Cell* **9**(8), 2287-2303.

Tulip, W. R., Harley, V. R., Webster, R. G. & Novotny, J. (1994). N9 Neuraminidase Complexes With Antibodies Nc41 and Nc10 - Empirical Free-Energy Calculations Capture Specificity Trends Observed With Mutant Binding Data. *Biochemistry* **33**(26), 7986-7997.

van der Merwe, P. A. & Barclay, A. N. (1994). Transient intermolecular adhesion: the importance of weak protein-protein interactions. *Trends in Biochemical Science* **19**, 354-358.

van der Merwe, P. A., McNamee, P. N., Davies, E. A., Barclay, A. N. & Davis, S. J. (1995). Topology of the CD2-CD48 Cell-Adhesion Molecule Complex - Implications For Antigen Recognition By T-Cells. *Current Biology* **5**(1), 74-84.

vanVlijmen, H. W. T., Schaefer, M. & Karplus, M. (1998). Improving the accuracy of protein pK(a) calculations: Conformational averaging versus the average structure. *Proteins-Structure Function and Genetics* **33**(2), 145-158.

Wallace, A. C., Laskowski, R. A., Thornton, J. M. (1995). LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Engineering* **8**(2),127-134.

Wang, J., Smolyar, A., Tan, K. M., Liu, J., Kim, M. Y., Sun, Z. J., Wagner, G. & Reinherz, E. L. (1999). Structure of a heterophilic adhesion complex between the human CD2 and CD58 (LFA-3) counterreceptors. *Cell* **97**(6), 791-803.

Wibbenmeyer, J. A., Schuck, P., SmithGill, S. J. & Willson, R. C. (1999). Salt links dominate affinity of antibody HyHEL-5 for lysozyme through enthalpic contributions. *Journal of Biological Chemistry* **274**(38), 26838-26842.

Wu, H., Kwong, P. D. & Hendrickson, W. A. (1997). Dimeric association and segmental variability in the structure of human CD4. *Nature* **387**(6632), 527-530.

Xu, D., Lin, S. L. & Nussinov, R. (1997). Protein binding versus protein folding: The role of hydrophilic bridges in protein associations. *Journal of Molecular Biology* **265**(1), 68-84.

Yang, A. S. & Honig, B. (1993). On the pH-Dependence of Protein Stability. *Journal of Molecular Biology* **231**(2), 459-474.

Zeng, J., Fridman, M., Maruta, H., Treutlein, H. R. & Simonson, T. (1999). Protein-protein recognition: An experimental and computational study of the R89K mutation in Raf and its effect on Ras binding. *Protein Science* **8**(1), 50-64

Appendix A

DNA and protein sequence of CD2d1 from *Rattus norvegicus* (rat).

I) CD2d1 DNA sequence

1 atgagagaca gtgggaccgt ctggggtgcc ctgggtcatg gcatcaacct
51 gaacatccct aactttcaaa tgactgatga tattgatgag gtgcgatggg
101 agagggggag caccctggtt gccgagttta aaaggaagat gaagcctttt
151 ttgaaatcgg gagcatttga gatcttagca aatggagact tgaagataaa
201 gaatctgaca agagatgaca gtggcaccta taatgtaacg gtatacagca
251 caaatgggac acgtatcctg aacaaggcac tggacttgag gattctagag

II) CD2d1 protein sequence

1 MRDSGTVWGA LGHGINLNIP NFQMTDDIDE VRWERGSTLV
41 AEFKRKMKPF LKSGAFEILA NGDLKIKNLT RDDSGTYNVT
81 VYSTNGTRIL NKALDLRILE

Appendix B

II) Assignment list for rat CD2d1

The following is the list of assignments for the first 99 residues of CD2 (CD2d1). The protein is expressed in E.Coli and contained an uncleaved N-terminal methionine (marked M0 in the list).

Residue	13C'	15N (1HN)	13CA (1HA)	13CB (1HB)	Others
Met0	172.4		55.1 (4.19)	33.0 (2.22)	CG 30.91(2.65)
Arg1	175.7	124.8 (8.81)	56.5 (4.43)	31.1 (1.90,1.85)	CG 27.18(1.69,1.65)
					CD 43.40(3.23)
					NE 85.58(7.21)
					CZ 171.72
Asp2	176.2	123.2 (8.63)	54.6 (4.72)	41.3 (2.85,2.76)	CG 180.15
Ser3	175.2	117.6 (8.46)	58.6 (4.67)	64.2 (3.96)	
Gly4	173.4	111.7 (8.56)	45.6 (4.27,4.05)		
Thr5	173.4	118.8 (8.03)	62.6 (4.91)	70.7 (3.81)	CG2 21.31(0.48)
Val6	173.1	128.9 (9.01)	61.0 (4.32)	34.6 (1.84)	CG1 21.03(0.94)
					CG2 21.68(0.91)

Residue	13C'	15N (1HN)	13CA (1HA)	13CB (1HB)	Others
+++++++	++++	+++ +++	++++ +++	++++ +++	++++++
Trp7	176.6	126.5 (8.52)	56.1 (5.30)	31.1 (2.99)	HD1 7.12
					NE1 130.24
					HE1 10.56
Gly8	170.9	109.5 (8.87)	44.0 (4.65,3.26)		
Ala9	177.0	124.8 (8.30)	50.3 (4.89)	20.5 (1.15)	
Leu10	179.4	123.0 (8.69)	56.2 (3.81)	42.0 (1.65,1.56)	CG 26.62(1.76)
					CD1 25.22(0.97)
					CD2 24.38(0.82)
Gly11	174.0	112.4 (9.37)	45.6 (4.09,3.91)		
His12	174.4	117.8 (7.90)	53.9 (4.96)	31.3 (3.43,3.32)	HD1 7.21
Gly13	174.3	110.1 (8.64)	45.1 (4.69,3.58)		
Ile14	171.1	120.8 (8.66)	59.2 (4.54)	41.8 (1.84)	CG2 18.04(0.74)
					CG1 29.04(1.43,1.04)
					CD1 14.78(0.88)
Asn15	174.2	124.0 (7.86)	51.2 (5.32)	39.1 (2.59,2.23)	CG 175.96
					ND2 111.04(7.34,6.49)
Leu16	173.9	125.8 (9.37)	54.0 (4.30)	41.3 (1.24,1.05)	CG 27.27(1.16)
					CD1 24.10(0.48)
					CD2 25.41(0.11)
Asn17	174.9	120.5 (8.07)	51.9 (4.84)	40.7 (2.71,2.55)	CG 176.89
					ND2 112.63(7.97,6.79)
Ile18	174.7	124.3 (8.23)	60.5 (3.96)	37.3 (1.97)	CG2 18.14(0.39)
					CG1 28.95(2.13,0.69)

					CD1 12.92(0.48)
Pro19	*176.6*		*63.3 (4.36)*	*32.2 (2.30,1.86)*	*CG 27.55(2.02)*
					CD 50.39(4.23,3.25)
Asn20	*173.5*	*117.6 (8.86)*	*54.2 (4.29)*	*38.0 (3.00,2.72)*	*CG 178.76*
					ND2 113.43(7.60,6.93)
Phe21	*173.6*	*118.7 (7.55)*	*57.6 (4.76)*	*43.4 (2.79,2.44)*	*HD1 7.20*
Gln22	*172.8*	*126.9 (7.02)*	*53.7 (4.22)*	*31.4 (1.72,1.64)*	*CG 33.42(2.12)*
					CD 180.68
					NE2 113.02(7.41,6.74)
Met23	*175.6*	*124.3 (8.42)*	*55.5 (4.08)*	*31.3 (1.98,1.91)*	*CG 31.93(2.74,2.36)*
Thr24	*175.5*	*117.2 (6.83)*	*60.2 (4.71)*	*72.5 (4.52)*	*CG2 21.40(1.07)*
Asp25	*175.7*	*119.3 (8.73)*	*56.4 (4.53)*	*40.6 (2.79,2.69)*	*CG 180.15*
Asp26	*175.1*	*116.5 (8.13)*	*54.4 (4.74)*	*40.8 (2.73,2.60)*	*CG 180.43*
Ile27	*175.2*	*120.2 (7.52)*	*59.2 (4.18)*	*35.57(2.68)*	*CG2 18.60(1.01)*
					CG1 27.09(2.06,1.72)
					CD1 9.65(0.86)
Asp28	*174.7*	*127.8 (8.95)*	*55.1 (4.96)*	*45.9 (2.31)*	*CG 177.82*
Glu29	*174.8*	*116.2 (7.77)*	*54.4 (5.71)*	*34.5 (2.10,1.95)*	*CG 34.82(2.61,2.10)*
					CD 182.58
Val30	*174.5*	*126.1 (9.11)*	*62.2 (4.75)*	*35.4 (2.17)*	*CG1 21.03(0.94)*
					CG2 21.77(0.84)
Arg31	*174.2*	*127.9 (9.35)*	*55.3 (5.13)*	*33.9 (1.89)*	*CG 28.39(1.66,1.49)*
					CD 43.77(3.23,3.07)
					NE 84.67(7.23)
					CZ 171.45

Residue	13C'	15N (1HN)	13CA (1HA)	13CB (1HB)	Others
+++++++	++++	+++ +++	++++ +++	++++ +++	++++++
Trp32	175.9	125.1 (9.34)	56.5 (5.45)	32.2 (3.22,3.00)	HD1 6.80
					NE1 130.01(10.16)
Glu33	174.8	123.8 (9.59)	54.4 (5.27)	34.7 (1.94)	CG 36.97(2.13,2.08)
					CD 182.58
Arg34	176.9	124.1 (8.75)	55.0 (4.51)	31.3 (1.55,1.09)	CG 26.62(1.07,0.98)
					CD 43.40(2.87,2.72)
					NE 85.47(6.88)
					CZ 171.88
Gly35	174.6	119.9 (8.98)	47.4 (4.03,3.65)		
Ser36	174.1	122.0 (8.82)	58.2 (4.53)	63.3 (4.06,3.90)	
Thr37	173.3	120.2 (8.19)	62.1 (4.42)	70.4 (4.26)	CG2 21.49(1.17)
Leu38	175.8	130.9 (8.77)	56.2 (4.36)	42.4 (1.82,1.44)	CG 26.99(1.38)
					CD1 23.64(0.95)
					CD2 26.06(0.46)
Val39	174.9	124.5 (8.97)	62.5 (4.29)	32.6 (1.40)	CG1 20.56(0.68)
					CG2 21.77(0.43)
Ala40	175.5	119.0 (7.88)	51.8 (5.43)	23.8 (1.37)	
Glu41	174.5	122.3 (9.71)	55.4 (5.64)	34.2 (2.25)	CG 34.08(2.44,2.15)
					CD 182.58
Phe42	173.5	129.2 (9.36)	56.1 (4.97)	41.8 (3.24,3.01)	HD1 7.09
					HE1 7.22
Lys43	174.7	130.9 (8.52)	54.0 (4.90)	35.8 (1.65,1.49)	CG 25.22(1.30)
					CD 29.04(1.52)

					CE 42.28(2.93,2.87)
Arg44	176.7	126.2 (8.18)	58.5 (3.74)	30.0 (2.10,1.94)	CG 27.74(1.99,1.83)
					CD 43.03(3.42,3.38)
					NE 84.46(7.54)
					CZ 171.70
Lys45	175.3	116.6 (8.73)	58.4 (3.88)	30.3 (2.15,2.07)	CG 25.97(1.44)
					CD 29.14(1.75,1.70)
					CE 42.19(3.02)
Met46	176.1	119.1 (7.99)	53.7 (4.83)	35.6 (2.09,1.81)	CG 32.12(2.48)
Lys47	175.0	125.2 (8.45)	55.0 (4.57)	31.9 (1.86)	CG 24.85(1.61)
					CD 29.42(1.78)
					CE 42.19(3.09)
Pro48	177.1		63.5 (4.81)	32.8 (2.42,1.83)	CG 27.36(2.23,2.12)
					CD 50.58(3.96,3.76)
Phe49	174.1	122.6 (8.67)	57.5 (4.76)	41.0 (3.25,2.85)	
Leu50	177.0	128.3 (7.63)	52.8 (4.64)	44.9 (1.50)	CG 27.18(1.62)
					CD1 24.94(1.07)
					CD2 23.54(0.97)
Lys51	175.7	123.3 (8.48)	58.7 (3.76)	33.6 (2.31,1.83)	CG 24.94(1.66,1.63)
					CD 29.42(1.91,1.83)
					CE 42.47(3.12)
Ser52	173.7	108.6 (7.64)	56.4 (4.66)	65.8 (4.18,4.00)	
Gly53	173.8	106.4 (8.68)	45.8 (4.33,3.81)		
Ala54	176.5	122.2 (7.96)	53.2 (4.05)	19.0 (0.85)	

Residue	13C'	15N (1HN)	13CA (1HA)	13CB (1HB)	Others
+++++++	++++	+++ +++	++++ +++	++++ +++	++++++
Phe55	174.2	115.3 (7.20)	55.7 (5.44)	43.7 (2.99,2.77)	HD1 7.08
					HE1 7.21
Glu56	173.8	119.1 (8.85)	55.6 (4.49)	34.5 (1.89,1.79)	CG 36.31(2.15,2.12)
					CD 182.58
Ile57	175.3	123.7 (8.57)	58.8 (5.06)	40.7 (1.59)	CG2 16.65(1.11)
					CG1 28.76(1.68,1.22)
					CD1 14.97(1.07)
Leu58	178.9	129.7 (8.91)	54.4 (4.50)	41.9 (1.99,1.70)	CG 27.64(1.57)
					CD1 25.22(0.84)
					CD2 22.80(0.77)
Ala59	177.7	123.0 (8.48)	54.8 (4.02)	18.1 (1.45)	
Asn60	175.6	112.5 (7.53)	52.4 (4.46)	37.7 (3.35,2.78)	CG 176.89
					ND2 109.00(7.26,6.42)
Gly61	172.0	108.2 (8.18)	44.8 (4.29,3.24)		
Asp62	173.7	119.9 (7.82)	55.1 (4.74)	40.9 (2.77,2.47)	CG 179.78
Leu63	174.1	121.4 (7.25)	53.1 (4.38)	42.5 (0.27,0.00)	CG 26.15(0.90)
					CD1 23.26(0.45)
					CD2 25.87(0.13)
Lys64	175.4	128.2 (9.01)	54.5 (5.04)	35.2 (1.60,1.45)	CG 24.66(1.27,0.83)
					CD 29.51(1.55)
					CE 42.19(2.83,2.79)
Ile65	176.3	127.6 (8.61)	60.7 (4.28)	38.6 (1.62)	CG2 17.30(0.56)
					CG1 27.27(1.30,0.84)

					CD1 13.76(0.52)
Lys66	*176.4*	*125.7 (7.93)*	*59.7 (3.74)*	*33.1 (1.73,1.47)*	*CG 25.97(1.24,1.13)*
					CD 29.32(1.62)
					CE 42.28(2.84,2.71)
Asn67	*172.9*	*117.1 (8.40)*	*53.1 (3.97)*	*39.5 (2.56,2.30)*	*CG 176.67*
					ND2 111.57(7.21,6.55)
Leu68	*177.2*	*124.0 (8.07)*	*56.9 (4.08)*	*42.8 (1.76,1.68)*	*CG 26.81(1.60)*
					CD1 25.87(0.82)
					CD2 25.59(0.73)
Thr69	*175.7*	*118.0 (8.88)*	*59.9 (4.76)*	*72.6 (4.48)*	*CG2 21.59(1.25)*
Arg70	*178.5*	*120.9 (9.00)*	*59.9 (3.87)*	*29.6 (1.88,1.84)*	*CG 28.76(1.85,1.69)*
					CD 43.12(3.29,3.23)
					NE 85.00(7.49)
					CZ 171.78
Asp71	*176.2*	*117.0 (7.97)*	*55.6 (4.55)*	*40.6 (2.75,2.50)*	*CG 180.06*
Asp72	*177.0*	*116.2 (7.63)*	*55.1 (4.81)*	*42.1 (2.92)*	*CG 178.38*
Ser73	*173.1*	*115.6 (7.46)*	*59.1 (4.22)*	*64.1 (3.92)*	
Gly74	*172.4*	*111.3 (8.78)*	*44.4 (4.50,3.94)*		
Thr75	*174.8*	*117.4 (8.35)*	*63.2 (5.15)*	*69.9 (3.89)*	*CG2 22.43(1.17)*
Tyr76	*174.4*	*130.0 (9.91)*	*57.1 (5.18)*	*41.0 (3.16,3.01)*	
Asn77	*173.8*	*121.8 (9.60)*	*52.7 (5.60)*	*42.7 (2.60,2.54)*	*CG 176.80*
					ND2 113.72(7.24,6.85)
Val78	*173.2*	*126.5 (9.16)*	*57.4 (5.07)*	*32.4 (-0.32)*	*CG1 18.60(0.28)*
					CG2 21.49(-0.38)
Thr79	*172.8*	*120.7 (8.38)*	*59.9 (4.91)*	*71.3 (3.65)*	*CG2 22.70(1.01)*

Residue	13C'	15N (1HN)	13CA (1HA)	13CB (1HB)	Others
+++++++	++++	+++ +++	++++ +++	++++ +++	++++++
Val80	174.7	123.8 (7.74)	60.3 (5.01)	35.4 (1.60)	CG1 23.36(1.11)
					CG2 21.21(0.40)
Tyr81	177	125.3 (9.17)	56.3 (5.63)	42.5 (2.99,2.86)	HD1 6.79
					HE1 6.60
Ser82	176.8	119.3 (9.30)	57.1 (5.37)	65.5 (4.50,3.95)	
Thr83	175.2	116.2 (8.70)	65.4 (3.96)	69.1 (0.00,3.95)	HB 4.36
					CG2 22.14(1.31)
Asn84	176.3	118.6 (8.21)	52.6 (4.85)	37.7 (3.23,2.89)	CG 177.64
					ND2 112.26(7.65,6.93)
Gly85	174.5	108.6 (8.23)	45.3 (4.36,3.75)		
Thr86	173.7	117.7 (7.72)	63.5 (4.19)	69.4 (4.18)	CG2 21.49(1.15)
Arg87	175.8	128.7 (8.98)	56.1 (4.13)	30.0 (1.89,1.57)	CG 26.99(1.03,0.92)
					CD 43.96(3.06)
					NE 84.30(7.07)
					CZ 171.97
Ile88	175.8	127.1 (8.62)	61.7 (4.19)	39.0 (1.82)	CG1 26.81(1.27,1.15)
					CG2 (0.97)
					CD1 12.82(0.88)
Leu89	174.4	119.0 (7.18)	55.4 (4.47)	45.4 (1.50,1.41)	CG 26.71(1.38)
					CD1 26.15(0.72)
					CD2 26.81(0.57)
Asn90	173.5	122.6 (8.42)	53.0 (5.38)	40.7 (2.60,2.42)	CG 176.33
					ND2 112.64(7.20,6.78)

Lys91	*173.3*	*124.7 (8.67)*	*54.1 (4.75)*	*37.0 (1.69,1.39)*	*CG 25.22(1.30)*
					CD 28.95(1.84,1.67)
					CE 42.28(3.08,2.97)
Ala92	*176.9*	*125.3 (8.29)*	*50.7 (5.47)*	*21.6 (1.29)*	
Leu93	*174.7*	*124.4 (9.43)*	*53.9 (4.93)*	*45.8 (1.79,1.50)*	*CG 27.93(1.72)*
					CD1 25.87(1.05)
					CD2 25.13(0.99)
Asp94	*173.5*	*123.8 (8.73)*	*53.3 (5.09)*	*41.8 (2.57)*	*CG 179.41*
Leu95	*175.8*	*128.3 (8.99)*	*54.4 (4.99)*	*43.4 (2.37,1.16)*	*CG 28.48(1.27)*
					CD1 26.06(0.71)
					CD2 24.57(0.71)
Arg96	*174.1*	*129.3 (8.89)*	*53.7 (4.50)*	*32.4 (1.15,0.16)*	*CG 26.81(1.12,0.95)*
					CD 42.93(2.72,2.67)
					NE 86.34(7.16)
					CZ 171.59
Ile97	*176.8*	*121.1 (8.39)*	*58.8 (4.70)*	*38.1 (1.95)*	*CG1 27.93(1.40,1.29)*
					CG2 18.14(0.88)
					CD1 12.26(0.70)
Leu98	*175.7*	*130.5 (8.85)*	*55.2 (4.51)*	*43.1 (1.67,1.60)*	*CG 27.19(1.69)*
					CD1 24.75(0.76)
					CD2 24.20(0.77)
					CD 24.75
Glu99	*180.7*	*129.0 (8.12)*	*57.7 (4.25)*	*31.0 (2.12,1.95)*	*CG 35.85(2.31)*
					CD 184.07
