PREDICTION, DESIGN AND CHARACTERISATION OF METAL BINDING SITES IN ANTIBODIES

David St. John Gregory

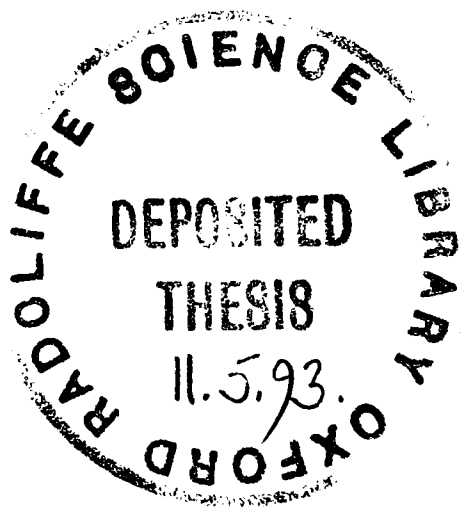
Thesis submitted for the Degree of

Doctor of Philosophy in the University of Oxford

Laboratory of Molecular Biophysics

and Merton College, Oxford

Trinity Term, 1992



Abstract

Prediction, Design and Characterisation of Metal Binding Sites in Antibodies

David St. John Gregory
Merton College

D. Phil. Thesis
Trinity Term, 1992

The design and creation of a functional metalloantibody is presented. A survey of metalloprotein structures was carried out to determine the common structural features of metal binding sites. Metal ligands were then introduced into hypervariable loop L1 of the anti-lysozyme antibody HyHEL-5 and the site comprehensively modelled. The final model was used as a basis for site-directed mutagenesis and mutant antibodies were produced using a *Xenopus laevis* oocyte expression system. Antigen binding studies showed the mutants to have an affinity for lysozyme equivalent to the parent antibody, while subsequent metal binding experiments have shown the new site to be capable of binding the transition metals cobalt, nickel, copper, zinc and cadmium.

A method for the rapid prediction and design of metal binding sites in proteins is described. A protein structure is screened for the location of putative metal sites by template matching. After the introduction of suitable liganding residues, the metal binding potential at each site is evaluated using hydrophobicity contrast. A point representing the metal ion within a new site may be optimised with-respect-to the ligands using multidimensional minimisation. The method has been tested on known metalloprotein structures. It was routinely able to identify the metal binding sites and position a metal point in the site to within 0.6Å of the actual metal ion.

Some of the work presented in this thesis has been published elsewhere:

Gregory, D. S., Staunton, D., Martin, A. C. R., Cheetham, J. C., Pedersen, J. and Rees, A. R. (1990). Antibody Combining Sites: Prediction and Design. *Biochem. Soc. Symp.*, **57**, 147–155.

Rees, A. R., Gregory, D. S., Pedersen, J., Jones, A., Hilyard, K., O'Neill, S. and Staunton, D. (1991). Antibody Combining Sites: Prediction and Design. Poster presented at the Protein Structure and Function Meeting in Honour of Sir David Phillips.

Gregory, D. S., Martin, A. C. R., Cheetham, J. C., and Rees, A. R. (1993). The Prediction and Characterisation of Metal Binding Sites in Proteins. *Protein Engineering.*, **6**(1), 29–35.

Acknowledgements

I wish to thank three people in particular for their help in the production of this thesis: my current supervisor, Professor Louise Johnson, for allowing me to continue to work in the laboratory and for her constant encouragement and support, including financial, without which completing this thesis would have been a much more difficult task; my previous supervisor, Dr Janet Cheetham for picking up the pieces of this project and this D. Phil student, post Rees, and turning the first into something worthwhile and believing in the second; and my surrogate supervisor, Dr Andrew Martin for his continued help, advice and support through thick and thin and for being a good friend.

Special thanks to my namesake, Dr Mary Gregoriou who managed to keep an eye me despite everything (and everyone!) else she had to deal with. I thank Dr David Staunton for his supervision of my foray into molecular biology, Dr Christina Redfield for performing the NMR experiments and Dr Werner Braun for supplying the coordinates of [Cd_7]-metallothionein-2. Thanks also to Professor R. J. P. Williams for useful discussions, James Robertson and Jan Pedersen for subroutines, Steven Lee for producing the photographs, and for his help and enthusiasm in making the video, Dr Richard Bryan for allowing me to have control of MACNAB as well as running ridiculously long batch jobs!

To the inhabitants of S1 both past and present: Robert Freeman, Sam Williams, Dr Jane Endicott, David Owen, John Timms, Daniel Crowther, Kate Smith, Irene Taylor, Simon Arthur, Ewald Schröder, other members of the lab and friends outside for making the last two years so enjoyable. Another special thanks to Kim Cameron, accommodation manager at Merton, for finding ways that allowed me to continue to live in college rooms and for her help in keeping the wolf from the door financially.

I thank Amersham International p.l.c. and latterly, the Access Fund for financial support. I especially thank my Aunt for her generous support in this respect which helped me to keep body and soul together, also the support of my Grandmother and the rest of my family. Finally, I'd like to thank my parents who have experienced the ups and downs of my D. Phil with me, but have always been there to help me in any way they could. It is to them that this thesis is dedicated.

Abbreviations

Amino Acids

A	Ala	Alanine
C	Cys	Cysteine
D	Asp	Aspartic Acid (Aspartate)
E	Glu	Glutamic Acid (Glutamate)
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
V	Val	Valine
W	Trp	Tryptophan
Y	Tyr	Tyrosine

Antibody Structure

ACS	Antibody Combining Site
CDR	Complementarity Determining Region (hypervariable loop = CDR)
L1, L2, L3	Light chain CDRs
H1, H2, H3	Heavy chain CDRs
Vκ	Kappa light chain variable domain
V_L	Light chain variable domain
V_H	Heavy chain variable domain
C_L	Light chain constant domain
C_H1	First heavy chain constant domain
C_H2	Second heavy chain constant domain
C_H3	Third heavy chain constant domain
C_H4	Fourth heavy chain constant domain (IgM only)
Fv	Variable fragment (V_L/V_Hdimer)
Fc	Constant (or crystallisable) fragment (C_H2,C_H3/C_H2,C_H3 dimer)
Fab	Antigen binding fragment (V_L,C_L/V_H,C_H1 dimer)

Chemicals

BSA	Bovine Serum Albumin
DMSO	Dimethyl sulphoxide
DTT	Dithiothreitol
EDTA	Ethylene diamine tetra-acetic acid
MBS	Modified Barth's Saline
MOPS	3-(N-morpholino) propane sulphonic acid
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate Buffered Saline
PMSF	Phenyl Methyl Sulphonyl Fluoride
SDS	Sodium Dodecyl Sulphate
Tris	Tris (hydroxymethyl) amino methane

General

Å	Ångström (10^{-10}m). 1Å = 0.1 nanometers
χ_n	Chi torsion angle about sidechain bond n
λ	Lambda (refers to molecular weight (Mw) marker)
ω	Omega torsion angle about C and N
ϕ	Phi torsion angle about N and Cα
$\Delta\phi$	Change in Phi
ψ	Psi torsion angle about Cα and C
σ	Standard deviation
θ	Theta (refers to torsion angle difference (in radians))
ASP	Atomic Solvation Parameter
CAMAL	Combined Algorithm for Modelling Antibody Loops
ccc	covalently closed circular
CPMG	Carr-Purcell-Meiboom-Gill (experimental technique in NMR)
CPU	Central Processing Unit (generally refers to computer time)
DNA	Deoxyribonucleic acid
cDNA	complementary DNA
ds	double-stranded
EM	Energy Minimisation
FID	Fourier Induction Decay
FPLC	Fast Prótein Liquid Chromatography
GAM	Goat anti-mouse (IgG polyclonal serum)
HEL	Hen Egg Lysozyme
HL	Human Lysozyme
Ig	Immunoglobulin
MOP	Maximum Overlap Procedure
Mw	Molecular weight
NIS	Non-immune serum
NMR	Nuclear Magnetic Resonance

PDB	Brookhaven Protein Databank coordinate or format file
PL	Polylinker
poly-A	poly-adenylated
ppm	parts per million
RF	Replicative form
r.m.s.	root mean square
RNA	Ribonucleic acid
mRNA	messenger RNA
SDM	Site-Directed Mutagenesis
SDR	Structurally-Determining Residue
ss	single-stranded
UT	Untranslated
UV	Ultraviolet
v/v	ratio of volume to volume
w/v	ratio of weight to volume
w/w	ratio of weight to weight

Contents

Abstract ii

Acknowledgements iv

Abbreviations v

1 Introduction 1

1.1 Antibody Structure and Function 2

1.1.1 Domain Structure 4

1.1.2 Conformation of Hypervariable Loops 6

1.1.3 Antigen Recognition 9

1.1.4 Applications of Antibody Specificity 13

1.2 Engineered Metal Binding Sites 14

1.3 Scope of This Thesis 17

2 A Survey of Metalloprotein Structures 19

2.1 Ligands and Metal-Ligand Distances 25

2.2 Secondary Structure 27

2.3 Metal Binding Sequences 28

2.4 Nearest Neighbour Residues 34
2.5 Discussion—Metals and Metalloproteins 36
2.5.1 Calcium and Magnesium 41
2.5.2 Manganese, Iron, Copper and Zinc 45
2.6 Conclusions 59

3 Metalloproteins in Metal Binding Site Design **60**
3.1 The Choice of Metal Ion 61
3.2 The Metallothionein Model 61
3.2.1 A Metal Binding Site in HyHEL-5 L1 62
3.2.2 Metal-Ligand Mutations in HyHEL-5 L1 64
3.3 Conformational Search of Hybrid L1 66
3.4 Energy Minimisation of Hybrid L1 69
3.5 Screening the Hybrid L1 Conformations 72
3.6 Conclusions 73
3.7 Rejection of the Hybrid L1 Model 75
3.7.1 The NMR Structure of [Cd_7]-metallothionein-2 76
3.7.2 Implications of the Metallothionein NMR Structure 78

4 Ab Initio Design of Metal Binding Sites **81**
4.1 Designing a Metal Binding Site in L1 82
4.1.1 Advantages of the New Metal Binding Site in L1 83
4.2 Conformational Search of Model1 L1 85
4.3 Energy Minimisation of the L1 Models in the Presence of Zinc . . . 96
4.3.1 Results of Energy Minimisation with Zinc 99

4.3.2 Discussion 104
4.4 Conclusions 107

5 Creation and Analysis of Metalloantibodies 108

5.1 Generation of Mutant cDNA 109
5.1.1 Oligonucleotide Site-Directed Mutagenesis 109
5.1.2 Synthetic Oligonucleotides for SDM 110
5.1.3 Site-Directed Mutagenesis 112
5.2 Sequencing the ssDNA Template 116
5.2.1 Preparation of M13G2H5-L1CEC DNA 118
5.3 Preparation of L1CEC Light Chain Insert 119
5.3.1 *Hpa* 1 Linearisation of M13G2H5-L1CEC DNA 119
5.3.2 *Nco* 1 Cleavage of Linearised M13G2H5-L1CEC DNA 121
5.3.3 Transformation of MC1061 Cells with pSP6G2H5-L1CEC . . 123
5.4 Preparation of L1CEC Light Chain DNA for Transcription 125
5.4.1 Small Scale Plasmid Preparations from Transformed Cells . 125
5.4.2 Sequencing of Three pSP6G2H5-L1CEC Clones 125
5.4.3 *Eco*R 1 Linearisation of pSP6G2H5-L1CEC 127
5.5 Synthesis of L1CEC Light Chain mRNA 127
5.5.1 Gel Analysis of RNA Transcripts 128
5.6 A *Xenopus laevis* Oocyte Expression System 130
5.6.1 Removal and Preparation of Oocytes 130
5.6.2 Microinjection of Oocytes 131
5.6.3 Labelling of Oocytes with ^{35}S-Methionine 132

5.7 Immunoprecipitation of Labelled Products 133
5.8 Characterisation of the Metalloantibodies 134
5.8.1 Binding Assay of the L1CEC Metalloantibody 134
5.8.2 NMR Studies of G2H5 and G2H5-HCEC Antibodies 137
5.9 Discussion 141

6 Metal Binding Site Prediction and Design **144**
6.1 Prediction and Design 145
6.1.1 Template Matching 146
6.1.2 Hydrophobicity Contrast 149
6.1.3 Multidimensional Minimisation 153
6.2 Analysis of Known Metalloproteins 157
6.3 Modelling Novel Metal Binding Sites 159
6.3.1 Modelling a Putative Zinc Binding Site in Gloop 2 159
6.3.2 Modelling a Ca^{2+} Binding Site in Human Lysozyme 160
6.3.3 Characterising the Zinc Binding Site in HyHEL-5 162
6.4 Conclusions 165

7 Modelling Hypervariable Loop SDR Mutations **166**
7.1 Altering CDR Conformation 167
7.1.1 Construction of L1 Val-29→Ile in J539 168
7.1.2 Construction of L3 Pro-94/Leu-95 Swap in J539 170
7.1.3 SDR Coupled Partner Mutations 174
7.2 A Mechanism to Explain Loss of Binding in Gloop 2 Mutants . . . 179
7.2.1 Mutations in Gloop 2 179

7.2.2 Construction of H2 Glu-50→Arg 181
7.2.3 Construction of L3 with H2 Glu-50→Arg 183
7.2.4 Construction of H2 Glu-50→Ser 184
7.2.5 Construction of L3 with H2 Glu-50→Ser 185
7.2.6 Construction of H1 with H2 Glu-50→Ser 186
7.2.7 Sidechain Construction around H2 Glu-50→Ser 187
7.3 Discussion 191
7.3.1 Implications for Protein Engineering 193

I A Metalloprotein Survey **195**

I.1 Calcium Metalloproteins 196
I.1.1 α-Lactalbumin 196
I.1.2 Thermitase 197
I.1.3 β-Trypsin 198
I.1.4 Galactose Binding Protein 198
I.1.5 Subtilisin Carlsberg 199
I.1.6 Elastase 200
I.1.7 Calmodulin 201
I.1.8 Calbindin 203
I.1.9 Parvalbumin B 204
I.1.10 Troponin C 205
I.2 Magnesium Metalloproteins 206
I.2.1 D-Xylose Isomerase 206
I.3 Manganese Metalloproteins 207

I.3.1 Glutamine Synthetase 207
I.3.2 Pea Lectin 208
I.3.3 Concanavalin A 209
I.4 Copper Metalloproteins 210
I.4.1 Azurin 210
I.4.2 Pseudoazurin 211
I.4.3 Azurin 212
I.4.4 Superoxide Dismutase 213
I.4.5 Plastocyanin 214
I.5 Iron Metalloproteins 215
I.5.1 Hemerythrin 215
I.5.2 Rubredoxin 216
I.6 Zinc Metalloproteins 217
I.6.1 Tonin 217
I.6.2 Carbonic Anhydrase B 218
I.6.3 Thermolysin 219
I.6.4 Carboxypeptidase A 221
I.6.5 Aspartate Carbamoyltransferase 222
I.6.6 Alcohol Dehydrogenase 223

II Standard Techniques in Molecular Biology 224

II.1 Materials 224
II.2 Biochemical Techniques 224
II.2.1 Agarose Gel Electrophoresis 224

II.2.2 Denaturing Polyacrylamide Gel Electrophoresis 225
II.2.3 Reducing SDS Polyacrylamide Gel Electrophoresis 225
II.2.4 Preparation of Modified Barth's Saline 226
II.3 Microbiological Techniques 226
II.3.1 Large Scale Preparation of M13 DNA 226
II.3.2 Gel Filtration of M13 DNA on Sephacryl-200 227
II.3.3 Preparation of Competent MC1061 Cells 227
II.3.4 Small Scale Preparation of Plasmid DNA 228
II.3.5 Very Large Scale Preparation of Plasmid DNA 228
II.3.6 CsCl Density Gradient Centrifugation 229

III Program CONRAD—Documentation **230**

List of Tables

1 Sequence alignments for the five protein families in the survey. . . . 21

2 Mean metal-ligand distances derived from Appendix 1. 26

3 Numbers of interactions used to derive each metal-ligand distance. . 26

4 Mean metal-ligand distances derived from small molecule structures. 26

5 Core secondary structures of the metalloproteins in Appendix 1. . . 29

6 Spacer lengths between ligands from the surveyed structures. 30

7 Numbers of each residue within the ligand spacer sequences. 33

8 Numbers of each residue as metal site nearest neighbours. 35

9 Specific physical properties of selected metal ions. 40

10 Biochemical classification of cations. 40

11 Basic requirements in the catalyses of selected transition metals. . . 47

12 Metallothionein substitutions in HyHEL-5 L1. 66

13 CONGEN constructions of the *Hybrid* L1. 69

14 Energy minimisation of the *Hybrid* L1 conformations with zinc. . . 70

15 Interaction energies for the minimised *Hybrid* L1 structures. 70

16 Solvent accessibilities of the *Hybrid* L1 lowest energy conformations. 74

17 R.m.s. deviations of the L1 models and HyHEL-5 L1. 94

18 Metal-ligand distances for the *Model1* L1 metal sites. 97

19 Comparison of metal site charges used in energy minimisation. . . . 100

20 Energies of the model structures after EM with zinc. 100

21 All atom r.m.s. deviations for the different zinc charges after EM. . 101

22 Zinc-ligand distances for the different zinc charges after EM. 101

23 MC1061 pSP6G2H5-L1CEC transformation results. 124

24 Zinc concentrations used during NMR of G2H5-HCEC Fv. 139

25 Cα and Cβ distances for template matching in CONRAD. 148

26 Mean metal-ligand distances derived from Appendix 1. 154

27 The results of CONRAD on 28 metalloprotein structures. 155

28 Chothia *et al.*'s canonical structures for CDR-L1. 169

29 Chothia *et al.*'s canonical structures for CDR-L3. 169

30 Shortest contact distances within 4.00Å of position 29 in J539 L1. . 175

31 R.m.s. deviations between J539 L1, L1 V29I and L1 V29II2A. . . . 177

32 R.m.s. deviations between Gloop 2 L3, L3E50R and L3E50S. 185

33 Sidechain constructions around H2 Glu-50→Ser in Gloop 2. 188

List of Figures

1 IgG structure. 3

2 CDR positions within the variable domains. 7

3 An EF-hand calcium binding site from Calmodulin. 43

4 The calcium, manganese binding site of Concanavalin A. 48

5 The iron binding sites of Hemerythrin. 51

6 The copper binding site of Plastocyanin. 54

7 The zinc binding site of Thermolysin. 58

8 HyHEL-5 L1 and Cd-Zn metallothionein residues 36–45 compared. . 63

9 Residues 36–45 from the crystal structure of Cd-Zn metallothionein. 65

10 NMR and crystal structures of metallothionein residues 36–45. . . . 77

11 Residues 36–45 from the NMR structure of Cd_7-metallothionein-2. . 79

12 The *Model1* L1 in the presence and absence of the N-terminus. . . . 93

13 The Eckstein Mutagenesis Method. 111

14 Amersham Mutagenesis Sample Analysis. 117

15 The SP64T RNA transcription vector light chain construct. 120

16 *Nco* 1 cleavage of linearised M13G2H5-L1CEC DNA. 122

17 Agarose gel of synthesized RNA transcripts. 129

18 Autoradiograph of ^{35}S-Met-labelled translation products. 135

19 Results of binding assay for the G2H5-L1CEC Antibody. 136

20 NMR spectra of the G2H5 and G2H5-HCEC Fv's. 140

21 Thermolysin coordinates coloured according to partial charge. . . . 150

22 Plot of C values calculated at the iron site in Rubredoxin. 152

23 The modelled zinc binding site in antibody Gloop 2. 161

24 The modelled calcium binding site in D86/D92 lysozyme. 163

25 The modelled zinc binding site in antibody HyHEL-5. 164

26 Comparison of J539 L1, J539 L1 V29I and REI L1. 171

27 Changes in ϕ for the J539 L1 V29I substitution. 171

28 Comparison of J539 L3, J539 L3 P94LL95P and REI L3. 173

29 Changes in ϕ for the J539 L3 SDR swap. 173

30 Comparison of J539 L1, J539 L1 V29I and J539 L1 V29II2A. 178

31 Changes in ϕ for J539 L1 V29II2A. 178

Chapter 1

Introduction

The high specificity of individual antibodies for their target antigens has found widespread application in science and medicine. The specificity is essentially confined to 6 hypervariable loops or 'complementarity-determining regions' (CDRs) which constitute the antibody combining site (ACS). The rapid advances in monoclonal antibody technology (Köhler and Milstein, 1975; Milstein, 1980) and CDR replacement strategies (Jones *et al.*, 1986; Verhoeyen *et al.*, 1988; Riechmann *et al.*, 1988) have allowed the development of antibodies with predetermined specificities, while advances in the development of bifunctional antibodies (Lewis and Hilvert, 1991) showed that the design of immunoglobulins capable of performing novel functions was feasible.

The recent interest in engineered metal binding sites (Higaki *et al.*, 1992) recognises the significance of the functions performed by metal ions. The incorporation of metals into antibodies could ultimately lead to the production of highly specific catalysts (Iverson and Lerner, 1989; Iverson *et al.*, 1990) and biosensors (Lowe, 1986; North, 1985), to almost any substrate (antigen). The ability to design metal binding sites in antibodies (and other proteins) will thus become increasingly important (Tainer *et al.*, 1991).

1.1 Antibody Structure and Function

[1]Antibodies or immunoglobulins serve two functions: (i) to bind to the epitope against which they have been raised, forming a specific high affinity complex; (ii) to stimulate effector functions such as complement and macrophage activation. The immunoglobulin structure combines both specific epitope recognition and common effector functions in a single multidomain protein. There are five classes of immunoglobulin: IgG, IgA, IgD, IgE and IgM. The general structure of an IgG is illustrated in Figure 1.

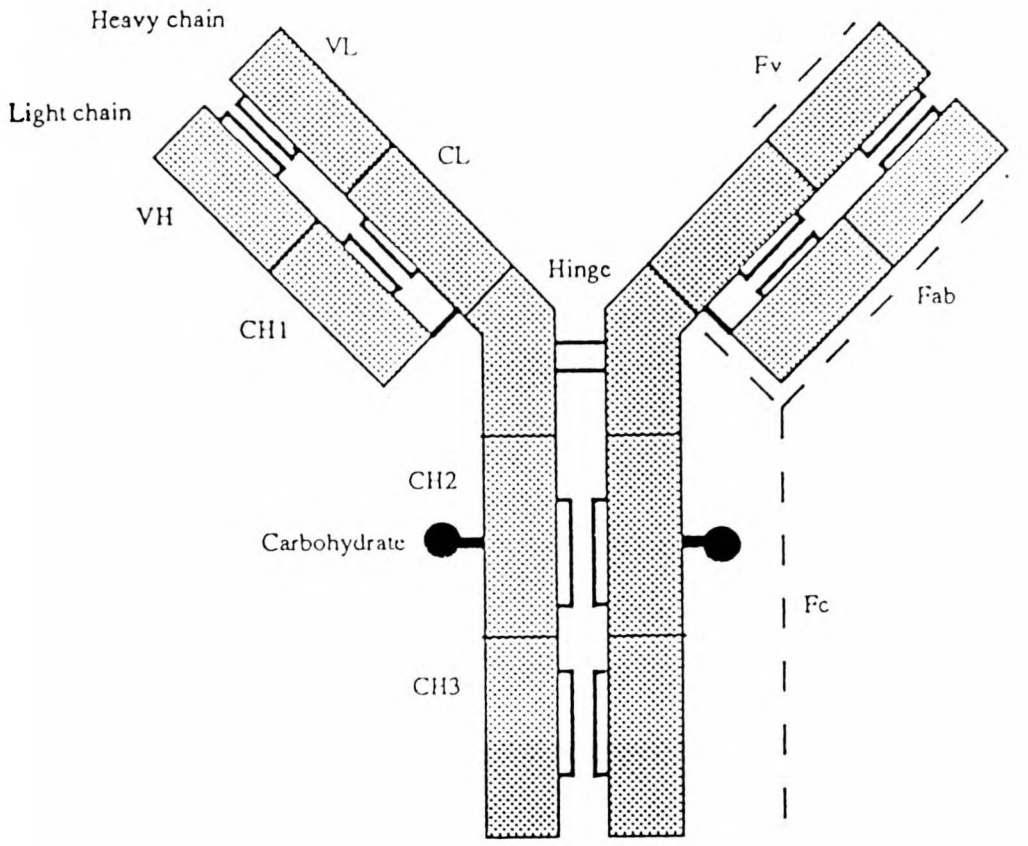
IgG consists of two identical heavy (H) and two identical light (L) chains linked by disulphide bonds. There are two antigen binding sites, one on each of the Fv arms. The site that activates effector functions is located in the Fc region. The variable domain supports the hypervariable sequence regions (CDRs) which determine specificity. Each chain contributes three CDRs that together compose the ACS. The four remaining classes contain essentially similar structures, but vary in the number and nature of their constant domains/identity in sub-domains, as follows:

IgA: The domain structure of IgA is identical to IgG but it can form a dimer linked by a 'J-chain' which causes association of the C-terminal ends of the heavy chains. IgA is the major antibody in secretions and thus forms the primary defence barrier of the acquired immune system against the entry of micro-organisms to tissues. It avoids digestion by means of a 'secretory piece', a polypeptide produced in epithelial cells and added to IgA dimers to enable transport across the epithelium.

---

[1]See Amzel and Poljak (1979), Davies and Metzger (1983), Mariuzza *et al.*, (1987), Alzari *et al.*, (1988), Davies *et al.*, (1988), (1990) and Chothia (1991) for reviews.

a)



b)

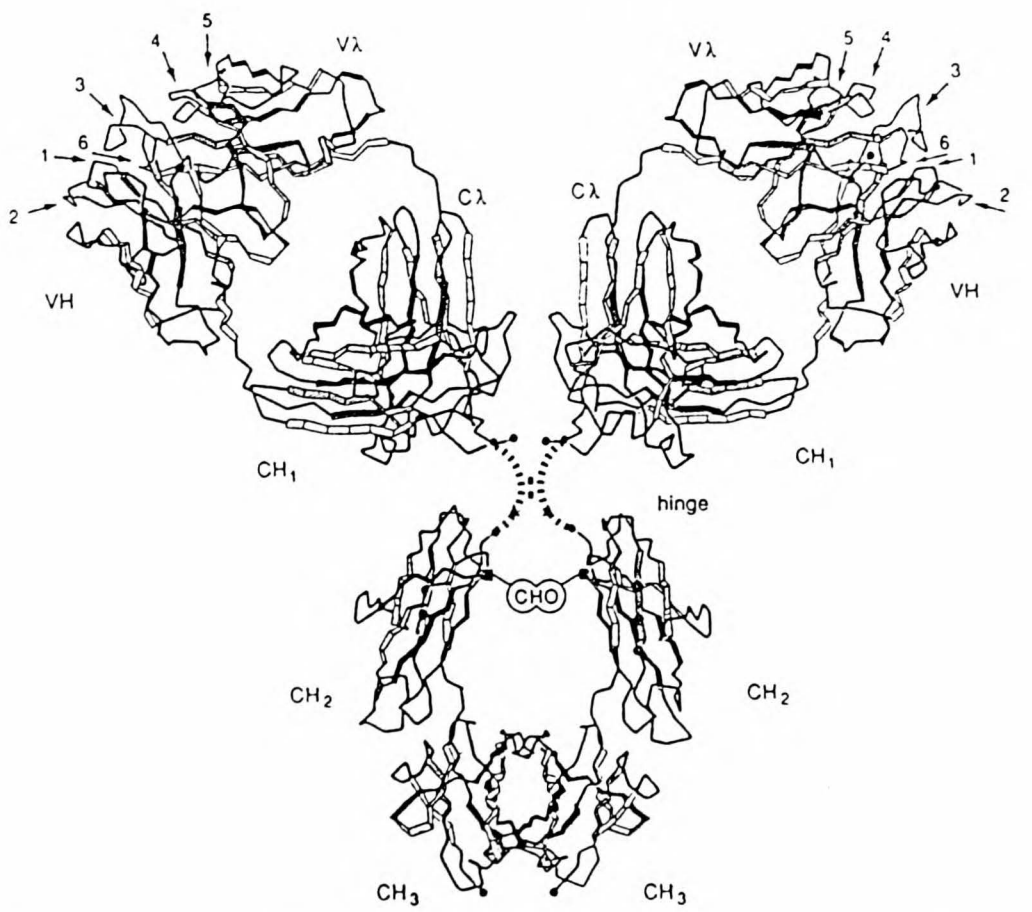


Figure 1: The structure of an IgG molecule. a) Basic structure; bold lines indicate the presence of intrachain and interchain disulphide bonds. b) Model of the three-dimensional structure based on the solved crystal structures of Fab and Fc domains of myeloma protein KOL (Matsushima *et al.*, 1978). The light chain is composed of the variable (V_L) domain and one constant (C_L) domain, and the heavy chain one V_H domain and three C_H domains (C_H1, C_H2, C_H3). The hypervariable regions are labelled 1–3(V_H) and 4–6(V_L). **CHO:** Carbohydrate. (Taken from Winter and Milstein (1991).)

IgD: In mouse, IgD has only two (rather than three) constant domains in the heavy chain while, in human, it has the same structure as IgG. The hinge region is extended and is particularly sensitive to proteolysis. The structure is heavily glycosylated. IgD functions on the B-cell surface and is thought to have some regulatory rôle causing differentiation of B-cells and may be involved in tolerance (Mosier *et al.*, 1977; Parkhouse and Copper, 1977; Calvert, 1986).

IgE: In IgE, the hinge region is replaced by an additional domain. The exact rôle of IgE is uncertain, but it is involved in the anti-parasitic response and is associated with hypersensitivity reactions such as asthma and hayfever. IgE binds to mast cells, promoting degranulation and histamine release.

IgM: A J-chain in IgM leads to pentamerisation (although monomers and other polymers are seen). The heavy chain has five domains: V_L, C_H1, C_H2, C_H3 and C_H4, and the hinge region is missing. The high concentration of binding sites means that IgM is very efficient at binding antigen and, on B-cells, it concentrates antigen for processing and presentation to T-cells. IgM is usually the first class of immunoglobulin to be synthesised and it is thought to be the earliest in evolutionary terms.

1.1.1 Domain Structure

X-ray crystallography has shown that both the variable and constant domains consist of two twisted anti-parallel β-sheets which form a β-sandwhich structure. The constant regions have three and four-stranded β-sheets arranged in a 'greek-key' like topology (Richardson, 1981), while variable regions have a further two short β

strands producing a five-stranded β-sheet. The two β-sheets are inclined at 30° to one another (Chothia and Janin, 1981), with a conserved disulphide bridge in each domain linking the two β-sheets. The relative orientation of the two β-sheets can vary by up to 18° in the variable domains and up to 10° in the constant domains (Lesk and Chothia, 1982).

The 'immunoglobulin fold' as this has been described, (Poljak *et al.*, 1973; Schiffer *et al.*, 1973; Chothia *et al.*, 1985) is a highly characteristic super-secondary structural motif and has been seen in a number of molecules of related and unrelated functions (Williams and Barclay, 1988); for example, Superoxide Dismutase (Richardson *et al.*, 1976) and PapD, a bacterial protein which mediates assembly of pili in *Escherichia Coli* (Holmgren and Bränden, 1989).

Interactions between V_L and V_H, C_L and C_H1, and C_H2 and C_H3 are extensive, with some 900Å^2 of surface area buried in each of the variable domains (Novotný *et al.*, 1983). Carbohydrate chains bound to the C_H2 domains form the interface between these domains. The V_L and V_H domains interact *via* the five-stranded β-sheet to form a nine-stranded β-barrel of about 8.4Å radius, with strands at the domain interface inclined at approximately 50° to one another (Novotný *et al.*, 1983). The loops linking the β-strands form the CDRs with the domain pairings bringing the loops from both the V_L and V_H domains into close proximity. The CDRs form 25% of the V_L/V_H domain interface (Colman, 1988) and are responsible for antigen binding and specificity (Figure 2). It has been shown (Jones *et al.*, 1986; Verhoeyen *et al.*, 1988; Riechmann *et al.*, 1988) that antigen binding specificity can be transferred from one antibody to another by CDR grafting. This serves to illustrate the rôle of the CDRs in determining specificity (Cheetham, 1989) (see

Section 1.1.4).

The C_L and C_H1 domains interact *via* the four-stranded β-sheets with the strands at the domain interface inclined at approximately 90° to one another. The angle between the pseudo-symmetry axes of V_L and V_H, and between C_L and C_H1 is described as the 'elbow bend' and varies between 133° in McPC603 (Satow *et al.*, 1986) and 178° in R19.9 (Lascombe *et al.*, 1989) which is almost fully extended. This bend has been shown to be flexible in JEL42 (Prasad *et al.*, 1988), HyHEL-5 (Sheriff *et al.*, 1987), D1.3 (Amit *et al.*, 1986) and KOL (Matsushima *et al.*, 1978). A 'ball and socket' model has been proposed for this region to describe the observed flexibility between the domains (Lesk and Chothia, 1988).

1.1.2 Conformation of Hypervariable Loops

Chothia *et al.*, (1989) examined the sequences of known immunoglobulin structures and found that many immunoglobulins had hypervariable loops which were the same size as those in other known antibody structures and contained the same or closely related residues at sites they proposed were responsible for the known conformations (Chothia and Lesk, 1987). From these observations, they reasoned that for five of the six hypervariable loops (excluding H3), there is only a small repertoire of canonical mainchain conformations. Kabat *et al.*, (1977) found conserved residues at sites within certain sets of hypervariable regions and suggested that they have a structural rôle. Padlan *et al.*, (1976) and de la Paz *et al.*, (1986) showed that some hypervariable loops in the immunoglobulins of known structure have the same mainchain conformation despite several differences in sequence.

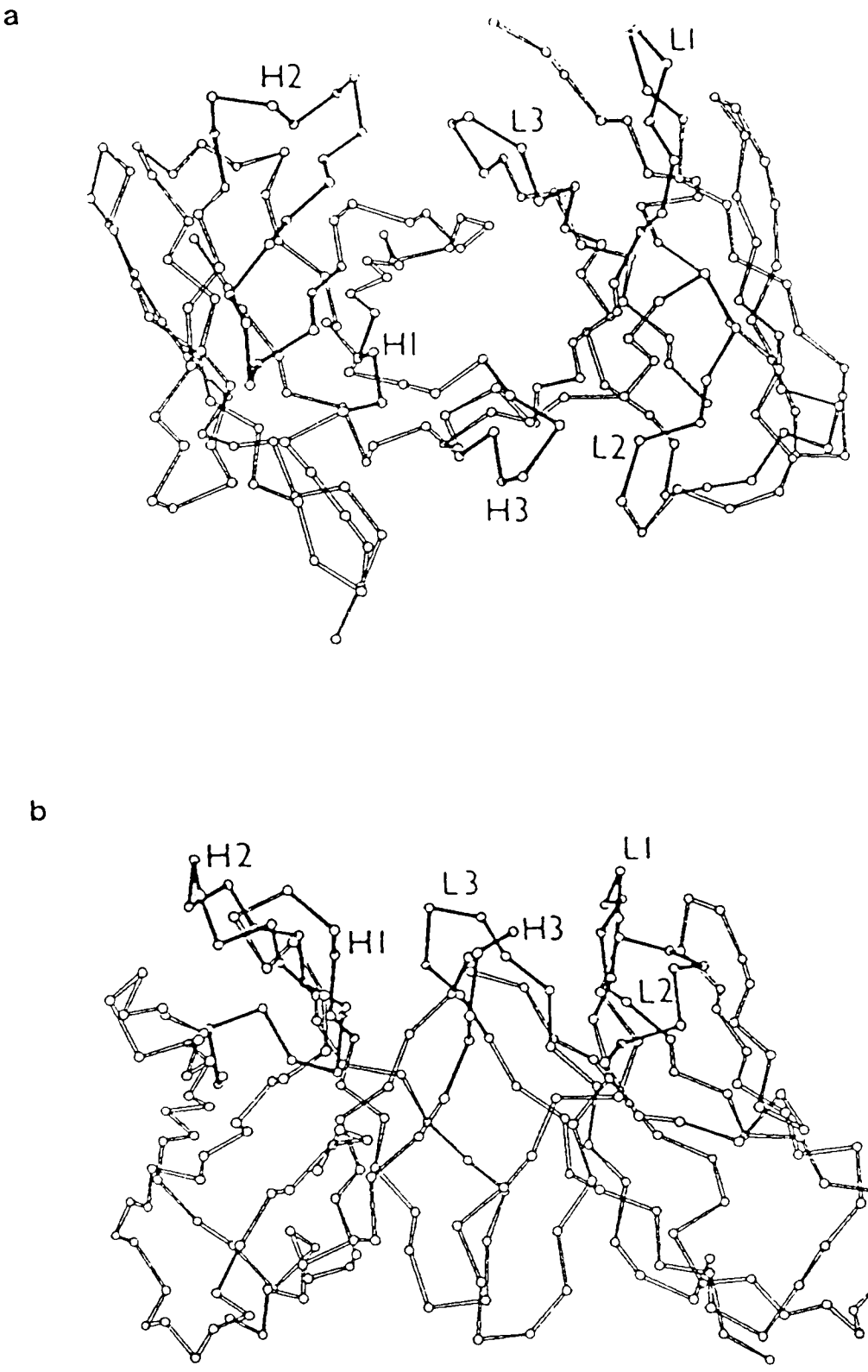


Figure 2: Cα plot of the HyHEL-5 Fv. a) View down onto the combining site. b) View across the combining site. Filled bonds: CDR residues; open bonds: framework residues. **L1, L2, L3:** Light chain CDRs; **H1, H2, H3:** Heavy chain CDRs. (Taken from Jeffrey (1989).)

Chothia and Lesk (1987) have identified the residues that are primarily responsible for CDR conformation in the structures of the Fab fragments NEW (Saul *et al.*, 1978), McPC603, KOL, J539 (Suh *et al.*, 1986) and the V_L domains of REI (Epp *et al.*, 1975) and RHE (Furey *et al.*, 1983). These 'structure-determining residues' (SDRs) affect loop packing (*e.g.* bulky residues such Trp, Tyr and Phe), can form hydrogen bonds or salt bridges (Ser, Thr, Asn, Gln, Asp, Glu, Arg, Lys) or are able to adopt unusual conformations (Gly and Pro). The conformations are determined by the interaction of a few residues at specific sites in the CDRs and framework.

Chothia *et al.*, (1989) used these findings to predict the conformations of the CDRs of four immunoglobulins: HyHEL-5, HyHEL-10 (Padlan *et al.*, 1989), NC41 (Colman *et al.*, 1987; Tulip *et al.*, 1989) and NQ10, and obtained good agreement (mean r.m.s. deviation 0.8Å) between the predicted structures and the observed mainchain conformations of the CDRs in these antibodies. Tramontano *et al.*, (1990) showed that the conformation and position of H2 residues 52–56 are determined by the packing against the V_H framework and by the identity of residue 71 in particular. They examined the environments of H2 and showed that in KOL and J539, Arg-71 packs between and hydrogen bonds to, H1 and H2. In HyHEL-5, residue 71 is an alanine. The cavity which would be created by this small sidechain is filled by the insertion of residue Pro-52a from H2. In KOL and J539, the sidechain at position 52a is on the surface. The relative movement of position 52a involves a change in the conformation of H2.

Sequence variations within the framework and hypervariable loops, especially at the SDRs, can shift the canonical structures relative to one another such that antigen binding is affected. Riechmann *et al.*, (1988) grafted the CDRs of a rat

antibody raised against the lymphocyte CAMPATH-1 antigen onto the human REI light chain and NEW heavy chain. However, in the first heavy chain framework region (HFR1) it was also necessary to make a single mutation, Ser-27H→Phe, to restore packing interactions within H1 in order to achieve binding. Mutagenesis studies by Roberts *et al.* (1987) on an anti-peptide antibody Gloop 2 (Jeffrey *et al.*, 1992) showed that the mutation Glu-50→Ser in H2, at a position close to the framework, resulted in the complete abolition of binding.

Chein *et al.* (1989) showed that alteration of CDR/framework contacts could abolish hapten binding. A single substitution, Asp-101 to Ala, abolished phosphocholine binding in antibody U10 even though this mutation is over 9Å from the phosphocholine binding pocket. They proposed that loss of binding is due to the elimination of a salt bridge between Asp-101 and Arg-94 which allows Arg-94, in H3, to disrupt binding by folding into the combining site. This study also underlines the importance of maintaining the conformation of H3 and shows that residues distant from the ACS can play a rôle in antibody specificity.

1.1.3 Antigen Recognition

The six hypervariable loops, which are supported on the β-barrel framework, form the ACS. The general arrangement of the loops is constant, forming a surface of approximately 700Å^2 (Colman, 1988). The topology of the ACS is determined by CDR sequence and length. Two of the CDRs, L1 and H3, have been proposed to be the most important in determining ACS topology (Padlan and Kabat, 1988) and are the most variable in length. Amongst the structures seen so far, anti-hapten (small molecule) antibodies form a flatter surface. H3 has the greatest

heterogeneity in both length and sequence and lies at the centre of the combining site, thus having the greatest potential for determining specificity. A relationship between CDR length and gross combining site topology has been proposed (de la Paz *et al.*, 1986; Padlan and Kabat, 1988). Flat combining sites with a shallow groove are characteristic of antibodies with a short L1 and H3, while cavities are characteristic of small sidechains in L3 and of a long L1 and H3.

1.1.3.1 General Features of Antigen Binding Sites

Details of the molecular interactions from the complexes of three anti-hapten antibodies (McPC603, NEW, 4-4-20 (Herron *et al.*, 1989)) and five anti-protein antibodies (HyHEL-5, HyHEL-10, D1.3, NC41, and NC10 (Colman *et al.*, 1989) have been reviewed.[2] In addition, the structure of an anti-peptide complex (Stanfield *et al.*, 1990) and an idiotope-anti-idiotope complex (Bentley *et al.*, 1990) has been solved.[3]

In anti-hapten complexes, the hapten binds in a pocket or groove in the centre of the combining site with approximately 90% of the ligand surface buried in an interaction area of 150–300Å (Davies *et al.*, 1990). The majority of interactions are with the heavy chain (55–60%) and four or more of the CDRs are involved. In contrast, the combining site surface of anti-protein antibodies tends to be flat and undulating, with depressions and protrusions complimenting the interacting surface of the antigen (the epitope).

---

[2](Amzel and Poljak, 1979; Davies and Metzger, 1983; Mariuzza *et al.*, 1987; Alzari *et al.*, 1988; Davies *et al.*, 1988; Davies *et al.*, 1990).

[3]See Chothia (1991) for a review of these two structures.

In the anti-lysozyme complexes (HyHEL-5, HyHEL-10 and D1.3), the interacting area is between 700–750Å^2 corresponding to approximately 14% of the lysozyme surface area. All six CDRs are involved in the interaction. There are no contacts made by L1 in the NC41 complex even though the area of interaction is larger for this antibody (880Å^2). Further refinement of the HyHEL-5 and D1.3 structures have shown that water molecules are present at the interface, and are involved in mediating contacts between antibody and antigen (Bhat *et al.*, 1990; Davies *et al.*, 1990). At least one framework residue is in contact with the antigen in all the anti-protein antibodies.

1.1.3.2 Antigen/Antibody Contacts

The binding energy and specificity of the antigen–antibody complex comes from ion-pair interactions, hydrogen bonds, van der Waals interactions and the removal of solvent around hydrophobic groups. The high complementarity between the interacting surfaces results in the virtual exclusion of solvent from the interface. Residues involved in hydrogen bonding and salt links are more likely to determine specificity rather than contribute to the binding affinity owing to their favourable interaction with solvent in the unbound state.

In all combining site sequences, there is a prevalence of tyrosine and tryptophan residues. It has been proposed that these residues could play an important rôle in ligand binding to the complex (de la Paz *et al.*, 1986; Padlan, 1990; Mian *et al.*, 1991). The large surface area of these sidechains has the potential to contribute greatly to the free energy of binding and there is a lower entropy change when they become immobilized at the interface (Padlan, 1990). Additionally, tyrosine and

tryptophan can participate in hydrogen bonding and thus can confer specificity as well as affinity to the complex.

The possibility of conformational changes upon antigen binding is still a matter of debate. In the anti-peptide antibody B13I2, the binding of peptide causes movement of three adjacent residues in H3 by 1.0–1.2Å and one sidechain rotates (Stanfield *et al.*, 1990). The structure of D1.3 Fv uncomplexed has been solved (Bhat *et al.*, 1990). A comparison of the unbound Fv structure with those of the D1.3 Fab complexed with lysozyme (Amit *et al.*, 1986) and with the anti-idiotypic Fab E225 (Bentley *et al.*, 1990), shows that the Fv combining site differs from those in the two complexes with the hypervariable regions of the V_H shifted relative to those of the V_L by about 0.7Å (Bhat *et al.*, 1990). In the idiotope-anti-idiotope complex, the antigen binding site differs from the other two structures by rotation of three sidechains (Bentley *et al.*, 1990).

Based on the information available from crystal structures to date, antigen recognition involves the use of the limited flexibility available to close-packed structures *i.e.* mainchain shifts of 1–2Å and rotation of surface sidechains. The recognition sites in the five anti-protein complexes known have very similar features: 33 ± 6 close-packed contact residues bury $1600 \pm 350 Å^2$ of surface and form between 10 and 13 hydrogen bonds (Bentley *et al.*, 1990; Amit *et al.*, 1986; Tulip *et al.*, 1989; Davies *et al.*, 1990). Approximately 20 residues of both antibody and antigen are in contact in the complex crystal structures with single mutations being sufficient to reduce binding drastically. For example, Arg-68→Lys in Turkey egg-white lysozyme reduces binding of HyHEL-5 by three orders of magnitude (Smith-Gill *et al.*, 1982) and Gln-121→His in lysozyme reduces binding of D1.3 by 450 times (Amit *et al.*,

1986). These features are identical to those found in protease-protein inhibitor complexes and other protein recognition systems (Janin and Chothia, 1990).

1.1.4 Applications of Antibody Specificity

Antibodies with predetermined specificities are widely used as analytic and diagnostic tools in biochemistry and medicine. The production of antibodies with particular specificities became practical with the invention of monoclonal antibodies produced by cell hybridisation techniques (Köhler and Milstein, 1975; Milstein, 1980). However, it may not always be possible to raise a monoclonal antibody against a particular molecule if its structure is not capable of illiciting an immune response. In addition, for antibodies to be useful in human therapy, they should ideally be of human origin in order to prevent an anti-globin response in the patient (Shawler *et al.*, 1985) which might interfere with therapy (Miller *et al.*, 1983) or cause immune complex hypersensitivity (Ratner, 1943). Unfortunately, the immortalisation of human antibody-producing cells has proven difficult (Carson and Freimark, 1986).

The antigen binding specificity of one antibody can be transferred to another antibody by grafting the CDRs conferring the specificity onto the framework regions of the second antibody. Murine CDRs of known specificity have been grafted onto human framework regions to produce a chimeric 'humanised' antibody. For example, humanised antibodies of the murine anti-4-hydroxy-3-nitrophenylacetyl caproic acid antibody have been produced (Jones *et al.*, 1986). The heavy chain CDRs of the murine monoclonal antibody (B1-8) were spliced into the heavy chain of a myeloma protein. This heavy chain was then combined with the B1-8 light chain to produce a chimeric antibody with the hapten binding affinity of native B1-8. Similarly mouse

anti-lysozyme activity from D1.3 has been transferred (Verhoeyen *et al.*, 1988) onto the human heavy chains from antibody NEW.

The ability to transfer antigen binding specificity between the framework of different antibodies illustrates the apparent flexibility of the CDRs to replacement and modification. This, coupled with the potentially vast range of specificities available, makes antibodies ideal candidates for protein engineering.[4]

1.2 Engineered Metal Binding Sites

Much interest has recently been focused on engineering metal binding sites in antibodies and other proteins (Arnold and Haymore, 1991). The production of bifunctional antibodies has allowed the inclusion of functions mediated by metals to be considered. Some functions, for example electron transport, are only possible by the incorporation of metal ions into proteins. A metal function which has received much attention is peptide hydrolysis.

The development of sequence-specific antibodies with protease activity has been achieved by Iverson and Lerner (1989) who introduced a catalytically active metal into the ACS by selecting antibodies which could bind a range of chelated metal complexes. The antibodies were produced by immunising with Co(III) triethylenetetramine (trien)-peptide hapten. One of these antibodies, 28F11, was identified as cleaving peptide bonds between Gly and Phe residues of two peptide substrates using a range of transition metal-trien complexes. It is interesting to note that peptide cleavage was most effective with the zinc complex (turnover number 6×10^{-4} s^{-1}),

[4]See Lewis and Hilvert (1991) for reviews.

since this metal is the primary cofactor in the reaction of natural metalloproteinases (see Chapter 2, Section 2.5.2.4).

The metal ion however, was not directly bound to the antibody. Therefore, the next logical step would be to produce antibodies having this, or any other function, available from a directly bound metal ion. This would require a metal binding site to be engineered into the ACS. The first steps to this end have been taken by Iverson *et al.*, (1990). They compared the backbone configurations around the catalytic metal binding sites from zinc metalloenzymes, with those of the CDRs in antibody light chains. A noticeable similarity was discovered between the anti-parallel β-sheet structure around the three His ligands of the zinc binding site from Carbonic Anhydrase B, and the β-sheet structure around residues 34, 89 and 91 in CDRs L1 and L3 of antibody light chains.

His residues were introduced at these positions in a single chain fluorescein binding molecule with a 212 linker derived from antibody 4-4-20 (Herron *et al.*, 1989), by site-directed mutagenesis using the method of Nakamaye and Eckstein (1986) (See Chapter 5). Metal binding was established by fluorescence quenching of Trp residues which surrounded the three-histidine site. The native protein gave a quenching constant of 1.2×10^4 M^{-1}, while in the presence of Cu(II), Trp fluorescence in the mutant protein was quenched until less than 20% residual fluorescence remained. They propose that quenching indicates binding of Cu(II) to the three-histidine site, with an approximate binding constant of 2×10^5 M^{-1}. The three-histidine site was also able to bind Zn(II) and Cd(II), with relative affinities in the order Cu(II) > Zn(II) > Cd(II). The molecule retained its ability to bind fluorescein with a binding constant of 3×10^5 M^{-1}, although this was between three and four orders

of magnitude lower than the native protein.

The engineering of metal binding sites in proteins is not confined to antibodies. Other work has concentrated on creating metal sites to enhance protein stability. Kuroki *et al.*, (1989) designed a calcium-binding site into human lysozyme based on the calcium site of α-lactalbumin which shows tertiary homology to lysozyme (Stuart *et al.*, 1986). They have shown that the enzyme containing the engineered site is more stable to heat and protease digestion than native human lysozyme. Pantoliano *et al.* (1988) altered the residues around a weak calcium binding site in subtilisin to increase its affinity for calcium and increase the stability of the enzyme.

Hellinga *et al.*, (1991) have inserted a copper binding site into *Escherichia coli* thioredoxin. Although this site resembles the copper binding sites found in the blue-copper proteins, it was designed using the program DEZYMER (Hellinga and Richards, 1991) and did not refer specifically to any particular natural copper protein structure. Variants of the mutant protein showing the original activity of thioredoxin could be isolated and although the metal binding site was specific for transition metals, Cu(II) did not bind as anticipated in the design, using only two of the four ligands originally introduced and two other native residues. To ensure metal binding at the engineered site it is necessary, as they point out, to consider and remove alternative modes of binding.

One final example, illustrating the further uses of engineered metal binding sites in proteins, is worthy of mention. Higaki *et al.* (1990; 1992), describe the regulation of serine protease activity by the binding of Cu(II) ions. Substitution of Arg-96 with His in rat trypsin places the imidiazole group near the essential active-site His-57. The relative position of these two histidines results in the formation of a

new metal binding site. Occupancy of this site by a metal ion prevents His-57 from participating as a general base in catalysis. The esterase and amidase activities of trypsin are thus inhibited. Enzyme activity can be fully restored by removing the bound copper ion with EDTA. The inhibition is similar to that of the serine proteinase Tonin (Fujinaga and James, 1987) in which His-57 is distorted owing to the formation of a zinc binding site with His-97 and His-99, such that the catalytic triad is no longer functional. However, in this case, three His sidechains are involved.

1.3 Scope of This Thesis

The first part of this introduction concentrated on the molecular architecture of the antibody molecule, and on the structure and function of the ACS in particular; the second part reviewed some of the engineered metal binding sites which have been created. An examination of antibody structure with metal site design reflects the underlying theme of this thesis: To use knowledge of antibody structure/function and metal/protein interactions to design novel metal binding sites into antibodies.

The first step in metal site design is to understand the structure of metal binding sites in natural metalloproteins. Accordingly, Chapter 2 examines the results of a survey of metalloprotein structures from the Brookhaven Protein Databank (Bernstein *et al.*, 1977) (individual entries in the survey are set out in Appendix 1) to determine the common structural features of metal binding sites. The observations are discussed in relation to the chemistry of the metal ion bound.

The modelling of a novel metal site into L1 of the anti-lysozyme antibody HyHEL-5 (Sheriff *et al.*, 1987) based on the structure of Cd-Zn metallothionein

is described in Chapter 3, while Chapter 4 describes the design of an L1 metal site *ab initio*. The molecular biology carried out to introduce the metal site designed in Chapter 4, into HyHEL-5 L1 is described in Chapter 5.

Chapter 6 describes a general method for the prediction, design and characterisation of novel metal binding sites in antibodies and other proteins, while in Chapter 7, the effects of mutations in CDRs with particular reference to alteration of the SDRs, are examined. The application of these findings to experimental results in other Chapters and the implications for protein engineering of antibody combining sites are discussed.

Chapter 2

A Survey of Metalloprotein Structures

The functions of metal ions are influenced by the proteins to which they are bound. The structure of these proteins and the nature of their metal binding sites are closely associated with the type of metal ion bound and its biochemical rôle. To introduce a metal binding site into an antibody, it is therefore necessary to understand how proteins bind metal ions. A survey of 27 metalloprotein structures from the Brookhaven Protein Databank (PDB) (Bernstein *et al.*, 1977) was performed to identify the common structural features of metal binding sites. A search of the protein databank was carried out using Group II *s*-block and *d*-block transition metals as search terms, to identify all examples of metalloprotein structures in the current release of the databank (1991).

The survey concentrated on proteins which bound transition metal ions since these are primarily responsible for the chemical functions (such as electron transport and catalysis) which could be usefully incorporated into an antibody. The main representatives from this group were manganese, iron, copper and zinc. Nevertheless, calcium, well represented in the PDB, and magnesium were also included as examples of metal ions other than transition metals, specifically bound by proteins, but involved in mechanical functions (*e.g.* protein activation). However, owing to the

limited number of metalloprotein structures available in the PDB, several examples of homologous metalloproteins were included in the survey to avoid basing assumptions about the binding of a particular metal ion on limited structural information. Nevertheless, the inclusion of homologous structures, with possibly similar metal binding sites, could affect the results.

Five protein families were represented in the survey: the serine proteinases; the EF-hand calcium binding proteins; the Lectins; the copper binding electron transport proteins and the zinc metalloproteinases. The extent of homology in each family was examined by sequence alignment using the program MULTALIGN (Barton, 1990; Barton and Sternberg, 1987). This implements the algorithm of Needleman and Wunsch (1970) and gives a significance score to indicate the quality of alignment. The sequences in each family were compared using a pairwise alignment and the results are shown in Table 1.

If the significance score obtained for the pairwise alignment of two sequences is greater than 5–6, then there is a high probability that the sequences will be correctly aligned over most of the regions that belong to core secondary structure. A score of less than 5.0 suggests that additional evidence is needed to support any relationship between the two sequences. Conversely, a score greater than 15.0 for the two sequences indicates that the alignment is likely to be correct over most of its length.

Examination of the significance scores given in Table 1 shows that six alignments have values greater then 5.0. Any common features of metal binding should be more apparent across a range of different metal sites rather than from sites of similar structure which might be expected in related metalloproteins. Homologous

Protein Family	Compared Sequences	Percentage Identity	Significance Score
Calcium binding serine proteinases	1TEC, 1TPO	15.25	1.40
	1TEC, 2SEC	48.13	24.25
	1TEC, 3EST	16.03	0.10
	1TPO, 2SEC	17.57	-0.24
	1TPO, 3EST	40.72	18.47
	2SEC, 3EST	14.17	-0.52
Calcium binding EF-hand proteins	3CLN, 3ICB	29.73	2.84
	3CLN, 5CPV	24.76	6.22
	3CLN, 5TNC	51.75	24.42
	3ICB, 5CPV	29.58	2.28
	3ICB, 5TNC	36.49	4.25
	5CPV, 5TNC	26.67	4.58
Ca/Mn Lectins	2LTN, 3CNA	18.33	-0.35
Copper binding electron transport proteins	1AZU, 1PAZ	9.17	-0.87
	1AZU, 2AZA	61.90	31.32
	1AZU, 5PCY	25.51	5.99
	1PAZ, 2AZA	16.18	-0.55
	1PAZ, 5PCY	20.41	3.57
	2AZA, 5PCY	24.49	3.70
Zinc proteinases	2CAB, 3TLN	15.69	0.78
	2CAB, 5CPA	15.35	-1.06
	3TLN, 5CPA	16.95	-0.34

Table 1: Sequence alignments for the five protein families represented in the survey. See text for description.

structures included in the survey contained differences in their metal binding sites as outlined below (see also Appendix 1):

- 1TEC contains two calcium binding sites, one of which differs from the single site in 2SEC.
- The calcium binding sites of 1TPO and 3EST contain different numbers of ligands.
- The similarity of the two EF-hand calcium binding sites in 5CPV with one of the EF hand pairs from 3CLN may be responsible for the observed homology between these two proteins. This is illustrated by the alignment of 3ICB with 3CLN which shows little homology owing to the presence of the long central helix. However, if the 3ICB sequence is aligned separately with each pair of EF-hand sequences from 3CLN, some homology between the two proteins is observed.
- Despite the high homology between 3CLN and 5TNC, 3CLN is involved in a much wider range of calcium-dependent functions than 5TNC which appears to be limited to skeletal muscle contraction. In addition, the four calcium binding sites in 5TNC have been characterised as two high-affinity and two low-affinity sites. This has not been described for 3CLN which may have sites of approximately equal affinity.
- 1AZU shows similarities to both 2AZA (with very high homology) and 5PCY (2AZA and 5PCY show little similarity). At the metal binding site however, it differs from both these proteins by the presence of a mainchain carbonyl group in place of the Met sidechain.

Chakrabarti (1989; 1990) also used examples of homologous metalloproteins in his investigations of the interactions of mainchain carbonyl groups and the sidechains of Cys and Met, with metal ions in protein structures. The introduction of metal

binding capability into an antibody is likely to be accomplished primarily by the mutation of amino acid sidechains. Hence, structures containing metal ions bound exclusively by mainchain carbonyl groups, in porphyrin rings or in inorganic clusters were excluded from the survey. To investigate the features of interest, the following questions were applied to each structure examined:

1. *What are the liganding residues?*

   Determine the sidechains primarily responsible for binding the metal ion.

2. *What are the metal-ligand distances?*

   These distances define the size of the cavity needed to accommodate the metal ion.

3. *What is the secondary structure of the protein and that which defines the metal binding site?*

   Examine any correlation between secondary structure and location of the metal binding site/association with particular metal ions.

4. *If the metal binding site is within a loop what is its sequence?*

   Determine the number and nature of the residues between the ligands.

5. *What are the nearest neighbour residues to the metal binding site?*

   Determine the nature of the residues immediately bordering the metal binding site by examining the sidechains within a sphere of radius of 7.5Å centered at the metal ion.

The metalloprotein structures included in the survey are shown below and the results for each one set out in Appendix 1. This chapter summarises the findings and attempts to identify any common features among them.

1. α-Lactalbumin (1ALC) (Acharya *et al.*, 1989)
2. Thermitase (1TEC) (Gros *et al.*, 1989)
3. β-Trypsin (1TPO) (Marquart *et al.*, 1983)
4. Galactose binding protein (2GBP) (Vyas *et al.*, 1988)
5. Subtilisin Carlsberg (2SEC) (McPhalen and James, 1988)
6. Elastase (3EST) (Meyer *et al.*, 1988)
7. Calmodulin (3CLN) (Babu *et al.*, 1988)
8. Calbindin (3ICB) (Szebenyi and Moffat, 1986)
9. Parvalbumin B (5CPV) (Swain *et al.*, 1989)
10. Troponin C (5TNC) (Herzberg and James, 1988)
11. D-Xylose Isomerase (5XIA) (Henrick *et al.*, 1989)
12. Glutamine Synthetase (2GLS) (Yamashita *et al.*, 1989)
13. Pea Lectin (2LTN) (Prasthofer *et al.*, 1989)
14. Concanavalin A (3CNA) (Hardman and Ainsworth, 1972)
15. Azurin (1AZU) (Adman and Jensen, 1981)
16. Pseudoazurin (1PAZ) (Petratos *et al.*, 1988)
17. Azurin (2AZA) (Baker, 1988)
18. Superoxide Dismutase (2SOD) (Tainer *et al.*, 1982)
19. Plastocyanin (5PCY) (Guss *et al.*, 1986)
20. Hemerythrin (1HMQ) (Stenkamp *et al.*, 1983)
21. Rubredoxin (5RXN) (Watenpaugh *et al.*, 1980)
22. Tonin (1TON) (Fujinaga and James, 1987)
23. Carbonic Anhydrase B (2CAB) (Kannan *et al.*, 1984)
24. Thermolysin (3TLN) (Holmes and Matthews, 1982)
25. Carboxypeptidase A (5CPA) (Rees *et al.*, 1983)
26. Aspartate Carbamoyltransferase (6AT1) (Stevens *et al.*, 1990)
27. Alcohol Dehydrogenase (7ADH) (Plapp *et al.*, 1983)

2.1 Ligands and Metal-Ligand Distances

Table 2 shows the mean metal-ligand distances from the 27 metalloprotein structures examined. (Table 3 shows a summary of the metal-ligand interactions used to derive these distances). They compare favourably with distances derived from small molecule structures (Table 4—Comparison with the distances from calcium proteins and their relationship to binding geometry is described in Section 2.5.1.1.) The mean difference between the data given in Tables 2 and 4 is 0.2Å. Generally, the protein-derived distances are longer. This may reflect the difference in resolution between protein and chemical crystallography. However, the possibility that protein distances could result from alterations in binding geometry to optimise metal ion function, should also be considered.

There is no obvious correlation between mean liganding distance and ligand type. Overall, the size of the coordination sphere decreases in the order:

$$Ca > Cu > Mn > Mg > Fe > Zn$$

This sequence is slightly distorted given the unusually long Cu–Met bond length, but correlates with the change in ionic radius which decreases in the order:

$$Ca > Cu > Mn > Mg > Fe > Zn$$

Table 2 also shows how calcium, magnesium and manganese complex only with oxygen and nitrogen-donating groups, while zinc, copper and iron complex with oxygen, nitrogen and sulphur-donating groups. Complex formation can be regarded as a reaction between a Lewis acid (the metal ion) and a Lewis base (the ligand). Metals and ligands can be divided into two classes described as 'hard' or 'soft' according

Ligand	Ca	Fe	Cu	Zn	Mn	Mg
Carbonyl	2.3	—	2.5	—	—	—
Asp-Oδ	2.4	2.1	—	1.9	2.2	2.1
Glu-Oϵ	2.5	2.2	—	2.2	2.2	2.1
His-Nδ,Nϵ	—	2.2	2.2	2.1	2.1	—
Cys-Sγ	—	2.3	2.1	2.2	—	—
Met-Sδ	—	—	2.9	—	—	—
Asn-Oδ	2.4	—	—	—	—	—
Gln-Oϵ	2.4	—	—	—	—	—
Hydroxyl	2.5	—	—	—	—	—

Table 2: Mean metal-ligand distances derived from the 27 metalloprotein structures in Appendix 1.

	Ca		Fe		Cu		Zn		Mn		Mg	
Ligand	n	σ	n	σ	n	σ	n	σ	n	σ	n	σ
Carbonyl	35	0.13	—	—	1	N/A	—	—	—	—	—	—
Asp-Oδ	46	0.15	2	0.02	—	—	1	N/A	4	0.09	2	0.07
Glu-Oϵ	33	0.14	2	0.17	—	—	3	0.12	6	3.8	2	0.15
His-Nδ,Nϵ	—	—	5	0.03	12	0.11	14	0.08	3	0.09	—	—
Cys-Sγ	—	—	4	0.03	4	0.18	10	0.16	—	—	—	—
Met-Sδ	—	—	—	—	3	0.18	—	—	—	—	—	—
Asn-Oδ	12	0.16	—	—	—	—	—	—	—	—	—	—
Gln-Oϵ	3	0.10	—	—	—	—	—	—	—	—	—	—
Hydroxyl	2	0.08	—	—	—	—	—	—	—	—	—	—

Table 3: Total number of each metal-ligand interaction used to derive the distances given in Table 2. n: Number of examples observed; σ: Sample standard deviation.

Ligand	Fe	Cu	Zn	Mn	Mg
Carboxyl	2.0	—	1.9	2.2	2.3
Imidiazole	2.2	2.0	2.0	2.3	—
Thiol	2.3	2.3	2.3	—	—
Thioether	—	2.4	—	—	—

Table 4: Mean metal-ligand distances derived from small molecule structures in the Cambridge Structural Database (Orpen *et al.*, 1989). The structures used as reference were those that most closely resembled the major liganding residues His (imidiazole), Asp and Glu (carboxyl), Cys (thiol), and Met (thioether).

to their polarizability (as measured by the charge/radius ratio (charge density—see Section 2.5)). Non-polarizable ions such as Mg(II), Ca(II) and Mn(II) are hard as are ligands containing highly electronegative oxygen and nitrogen donor atoms, whereas ions like Fe(II), Cu(II) and Zn(II), and ligands containing sulphur donor atoms, are easily polarized and are considered borderline/soft. The most stable complexes are those formed between either hard metals and hard ligands (with largely electrostatic bonding) or soft metals with soft ligands (with predominantly covalent bonding) whilst mixed complexes have lower stability.

Complex stability is also affected by the number of binding sites on the ligand. If the ligand has two or more binding sites which may be used simultaneously (a polydentate ligand), then there will be an overall increase in the number of molecules according to the equation:

$$[M(H_2O)x]^{n+} + L \rightarrow [ML]^{n+} + H_2O \quad (1)$$

This may bring about an increase in the entropy of the system and with it a decrease in free energy, favoring the formation of $[ML]^{n+}$. This increase in stability is known as the *chelate effect* and it is significant that proteins are multidentate (Phipps, 1976).

2.2 Secondary Structure

Table 5 shows the secondary and super-secondary structure of the metalloproteins in Appendix 1, and the secondary structure of their metal binding sites. The general pattern which emerges from this Table is the association of calcium with proteins having a high helical content, copper with predominantly β-sheet proteins and zinc

with mixed α/β proteins. These associations have functional significance.

When calcium is acting as a mechanical trigger, the movement it initiates is transmitted through the helices. In contrast, the extensive β-sheet of the copper binding proteins is consistent with the rôle of this metal in electron transport. The rigid sheet structure permits very little movement, maintaining the copper in a geometry that facilitates the changes in oxidation state required for rapid electron transfer.

The combination of secondary structure types found in the zinc binding proteins may represent a compromise between the mechanical flexibility of α-helical systems and the rigidity of β-sheets, to optimise the function of the zinc ion as a Lewis acid catalyst.

2.3 Metal Binding Sequences

Table 6 shows the number of residues between the ligands of metal binding sequences from the metalloprotein structures examined. The length of the 'spacer' sequences is related to the mobility of the metal binding site and is associated with protein function (Vallee and Auld, 1989; Vallee and Auld, 1990). For each metal, the trends in the number of residues between ligands, as identified here, are tentative. This may be a consequence of the number of proteins which bound particular metals and that were included in the survey, were not sufficiently large for a pattern to emerge. The general spacer sequence preferred by each metal is described below. The numbers of examples observed are given in brackets.

PDB Code	Metal Ion	Secondary Structure	Structure of site
1ALC	Ca^{2+}	9h 1s	Helix-turn-helix
1TEC	$2\,Ca^{2+}$	7h 2s	Surface loops
1TPO	Ca^{2+}	3h	Strand turn
2GBP	Ca^{2+}	10h 2s	Random coil
2SEC	Ca^{2+}	9h 2s	Random coil
3EST	Ca^{2+}	2h 2b	Turn in barrel
3CLN	$4\,Ca^{2+}$	7h 2s	EF-hand
3ICB	$2\,Ca^{2+}$	5h	EF-hand
5CPV	$2\,Ca^{2+}$	6h 1s	EF-hand
5TNC	$2\,Ca^{2+}$	10h	EF-hand
5XIA	Mg^{2+}	15h 8b	β-barrel
2GLS	$2\,Mn^{2+}$	15h 3s	β-sheet
2LTN	Mn^{2+},Ca^{2+}	s	Strand loop
3CNA	Mn^{2+},Ca^{2+}	3s	β-sheet
1AZU	Cu^{2+}	s	β-sheet
1PAZ	Cu^{2+}	2h 1b	Type III turn in β-barrel
2AZA	Cu^{2+}	2h 2s	β-strand
2SOD	Cu^{2+},Zn^{2+}	b	β-barrel
5PCY	Cu^{2+}	2s	β-sheet
1HMQ	$2\,Fe^{2+}$	4h	α-helical
5RXN	Fe^{2+}	s	β-strand
1TON	Zn^{2+}	4h 4s	N/A
2CAB	Zn^{2+}	7h 1s	β-sheet
3TLN	Zn^{2+} $4\,Ca^{2+}$	12h 3s	α-helical Surface loops
5CPA	Zn^{2+}	8h 1s	α/β
6AT1	Zn^{2+}	3h 2s	Strand turns
7ADH	Catalytic Zn^{2+} Structural Zn^{2+}	10h 4s	α/β Surface loop

Table 5: Secondary and super-secondary structure of the metalloproteins in Appendix 1, and secondary structure of their metal binding sites. *Secondary Structure*: Gives the number and types of secondary structure in each protein. $\mathbf{h} = \alpha$-helices, $\mathbf{s} = \beta$-sheets and $\mathbf{b} = \beta$-barrels.

PDB Code	Metal Ion	Spacer Sequence	Metal site loop
1ALC	Ca^{2+}	79L-x-x-L-x-L-x-x-L-L88	√
1TEC	1 Ca^{2+}	5L-(41x)-L-(34x)-L-x-x-L-x-L-x-L89	√
	2 Ca^{2+}	57L-x-x-L-x-L-x-L-x-L66	√
1TPO	Ca^{2+}	70L-x-L-x-x-L-x-x-x-x-L80	√
2GBP	Ca^{2+}	134L-x-L-x-L-x-L-x-L-(62x)-L205	
2SEC	Ca^{2+}	2L-(38x)-L-(34x)-L-x-L-x-L-x-L81	
3EST	Ca^{2+}	70L-x-L-x-x-L-x-L-x-x-L80	√
3CLN	All Ca^{2+}	L-x-L-x-L-x-L-x-L-x-x-L	√
3ICB	1 Ca^{2+}	14L-x-x-L-x-L-x-x-L-x-x-x-x-L27	√
	2 Ca^{2+}	54L-x-L-x-L-x-L-x-x-x-x-L65	√
5CPV	1 Ca^{2+}	90L-x-L-x-L-x-L-x-x-x-x-L101	√
	2 Ca^{2+}	51L-x-L-x-L-x-L-x-L-x-x-L62	√
5TNC	All Ca^{2+}	L-x-L-x-L-x-L-x-x-x-x-L	√
5XIA	Mg^{2+}	180L-(35x)-L-(27x)-L-(47x)-L292	
2GLS	1 Mn^{2+}	131L-(80x)-L-(7x)-L220	
	2 Mn^{2+}	129L-(139x)-L-(87x)-L357	
2LTN	Mn^{2+}	119L-x-L-(7x)-L-(6x)-L136	√
	Ca^{2+}	121L-x-L-x-L-x-x-x-L129	√
3CNA	Ca^{2+}	10L-x-L-x-L-x-x-x-x-L19	√
	Mn^{2+}	8L-x-L-(8x)-L-x-x-x-x-L24	√
1AZU	Cu^{2+}	45L-L-(65x)-L-x-x-x-x-L117	
1PAZ	Cu^{2+}	40L-(38x)-L-x-x-L-x-x-x-x-L86	
2AZA	Cu^{2+}	46L-(66x)-L-x-x-x-x-L-x-x-x-L121	
2SOD	Cu^{2+}	44L-x-L-(13x)-L-(57x)-L118	
	Zn^{2+}	61L-(6x)-L-(7x)-L-x-x-L81	√
5PCY	Cu^{2+}	37L-(47x)-L-x-x-L-x-x-x-x-L92	
1HMQ	1 Fe^{2+}	58L-(14x)-L-x-x-x-L-(23x)-L-x-x-x-x-L106	
	2 Fe^{2+}	25L-(29x)-L-x-x-x-L-(48x)-L106	
5RXN	Fe^{2+}	6L-x-x-L-(29x)-L-x-x-L42	
1TON	Zn^{2+}	57L-(40x)-L-x-L99	
2CAB	Zn^{2+}	94L-x-L-(22x)-L119	
3TLN	Zn^{2+}	142L-x-x-x-L-(19x)-L166	
	1 Ca^{2+}	138L-(38x)-L-(7x)-L-x-L-x-x-L190	√
	2 Ca^{2+}	177L-(5x)-L-x-L-x-x-x-x-L190	√
	3 Ca^{2+}	57L-x-L-x-L61	√
	4 Ca^{2+}	193L-L-x-x-L-x-x-L200	√
5CPA	Zn^{2+}	69L-x-x-L-(123x)-L196	
6AT1	Zn^{2+}	109L-x-x-x-x-L-(23x)-L-x-x-L141	
7ADH	1 Zn^{2+}	46L-(20x)-L-(106x)-L174	
	2 Zn^{2+}	97L-x-x-L-x-x-L-(7x)-L111	√

Table 6: Number of residues between the ligands of the metal binding sequences from the metalloprotein structures examined. **L** = liganding residue; **x** = spacer residue. *Metal site loop*: A √ indicates that all the metal ligands forming the binding site are donated by a single loop of polypeptide. If no indication is given the ligands are donated by distinct regions of secondary structure (see also Table 5).

Calcium: Four of the proteins (Calmodulin (3CLN), Calbindin (3ICB), Parvalbumin (5CPV) and Troponin C (5TNC)) bind calcium using the EF-hand motif. The sequences of these motifs are all 12 residues in length. Positions 1, 3, 5, 7, 9 and 12 coordinate the metal ion (position 7 *via* the mainchain carbonyl oxygen). The preferred spacer sequences for the other calcium binding proteins are:

$$\text{L-x-L (34)} > \text{L-x-x-L (10)} > \text{L-x-x-x-x-L (5)}$$

Copper: The most common spacer sequences for the copper proteins are:

$$\text{L-x-x-x-x-L (3)} > \text{L-x-x-L (2)} > \text{L-x-x-x-L} = \text{L-x-L (1)}$$

Iron: No common spacer sequences were apparent for iron.

Magnesium: No common spacer sequences could be described for magnesium.

Manganese: The manganese binding sites in Glutamine Synthetase (2GLS) had similar spacer sequences:

$$\text{L-}(n\text{x})\text{-L-}(n\text{x})\text{-L}$$

The most common spacer sequences for the other proteins (Pea Lectin (2LTN) and Concanavalin A (3CNA)) are:

$$\text{L-x-L (6)} > \text{L-(4x)-L (2)} > \text{L-(3x)-L} = \text{L-(7x)-L} = \text{L-(6x)-L} = \text{L-(x)-L (=1)}$$

Note that L-x-L is also the most common spacer sequence for calcium.

Zinc: For zinc, the spacer sequences followed the order:

$$\text{L-x-x-L} > \text{L-x-x-x-L} = \text{L-x-x-x-x-L}.$$

This is broadly similar to the spacer sequences for the calcium binding sites in Thermolysin (3TLN):

$$\text{L-x-x-L (7)} > \text{L-x-L (4)} > \text{L-x-x-x-x-L (2)} > \text{L-x-x-x-L (1)}$$

A more complete analysis of the spacer sequences between the ligands in zinc metalloproteins is given in Vallee and Auld (1989).

Table 7 shows the number and type of residues within the spacer sequences between the ligands, from the metal binding sequences examined. The twenty amino acids have been divided into groups as described in Creighton (1984). Generally, the popularity of particular sidechains in spacer sequences follows the trend: small hydrophobic > hydroxyl > hydrogen (Gly) > aromatic > charged (negative $\approx$ positive) $\approx$ amide > pyrollidine (Pro) > sulphur-containing (Met) > imidiazole (His).

However, due to the larger number of examples, the proportions of each residue identified in the calcium proteins was greater than for the other metalloproteins. If the data for calcium are excluded the frequency changes to: small hydrophobic > hydroxyl = hydrogen (Gly) > aromatic = positively-charged = pyrollidine (Pro) > amide > negatively-charged > sulphur-containing (Met) > imidiazole (His). Comparison with the previous trend shows that the frequencies for the majority of residue types remain similar. Only relatively small changes in the distributions of

Residue	Ca	Cu	Fe	Mn	Zn	Total
Gly	25	9	0	2	5	41 (16)
Ala	11	3	1	4	2	21 (10)
Val	13	1	0	1	2	17 (4)
Leu	6	1	2	1	1	11 (5)
Ile	16	0	2	1	0	19 (3)
Pro	9	6	1	5	2	23 (14)
Ser	9	3	0	2	2	16 (7)
Thr	19	2	2	4	1	28 (9)
Asp	14	1	0	1	1	17 (3)
Glu	9	2	0	0	1	12 (3)
Asn	13	1	1	2	3	20 (7)
Gln	6	1	0	1	0	8 (2)
Lys	13	3	3	0	3	22 (9)
Arg	3	1	0	2	2	8 (5)
His	1	0	0	0	0	1 (0)
Phe	10	3	1	2	1	17 (7)
Tyr	7	1	0	3	1	12 (5)
Trp	2	1	0	1	0	4 (2)
Met	0	2	0	0	0	2 (2)
Cys	0	0	0	0	0	0 (0)

Table 7: Numbers of different sidechain types in the spacer sequences between the ligands of the metal binding sequences examined. The numbers in brackets are the totals for each residue when the data for calcium are excluded.

negatively-charged, amide and pyrollidine (Pro) sidechains is observed. Cys residues were not found in any of the spacer sequences.

Thus, despite the exclusion of data from the calcium proteins, the overall frequency of residue types is similar for proteins from all the metal ions examined. Note also that exclusion of the data for calcium removes the effects of homologous structures on the results, since the majority of these were for calcium (see Table 1). (Exclusion of the remaining homologous structures (for copper) does not affect either trend. This is also true for the frequencies of nearest neighbour residues given in the next Section.)

2.4 Nearest Neighbour Residues

Table 8 shows the number and type of nearest neighbour residues, excluding ligands, within a sphere of radius 7.5Å centered at the metal ion. This radius was chosen so that the residues closest to the metal-ligand complex, which may exert an influence on it, could be identified. Again, the twenty amino acids are divided into groups as described in Creighton (1984) and the small sample size for each metal must be considered. Since all the calcium binding sites examined were created within single surface loops, the residues between the ligands, as shown in Table 7, were also the nearest neighbours *i.e.* these sites had no nearest neighbour residues.

The frequency of particular sidechains as nearest neighbours follows the general trend: small hydrophobic > aromatic ≈ hydroxyl > negatively-charged > hydrogen (Gly) ≈ amide > positively-charged > pyrollidine (Pro) > imidiazole (His)

Residue	Ca	Cu	Fe	Mn	Zn	Total
Gly	25	3	2	0	1	31 (6)
Ala	11	2	0	0	1	14 (3)
Val	13	2	2	1	2	20 (7)
Leu	6	3	1	2	2	14 (8)
Ile	16	0	1	2	2	21 (5)
Pro	9	2	1	1	2	15 (6)
Ser	9	1	0	1	4	15 (6)
Thr	19	1	0	3	2	25 (6)
Asp	14	1	0	2	4	21 (7)
Glu	9	0	0	3	4	16 (7)
Asn	13	5	0	1	2	21 (8)
Gln	6	0	0	1	2	9 (3)
Lys	13	2	0	0	1	16 (3)
Arg	3	1	0	3	3	10 (7)
His	1	2	0	4	2	9 (8)
Phe	10	6	2	1	3	22 (12)
Tyr	7	1	2	3	1	14 (7)
Trp	2	0	1	1	1	5 (3)
Met	0	7	1	0	0	8 (8)
Cys	0	0	0	0	0	0 (0)

Table 8: Number and type of nearest neighbour residues, excluding ligands, around the metal binding sites examined. The numbers in brackets are the totals for each residue when the data for calcium are excluded.

≈ sulphur-containing (Met). However, if the data for calcium is excluded, this frequency changes to: small hydrophobic > aromatic > negatively-charged > hydroxyl ≈ amide > positively-charged > imidiazole (His) = sulphur-containing (Met) > Hydrogen (Gly) = pyrollidine (Pro). The first trend is distorted by the data for calcium owing to the prevalence of Gly and small hydrophobic residues, which maintain the conformational flexibility necessary for calcium binding, and the prevalent polar and charged residues, that may be involved in determining affinity (Pantoliano *et al.*, 1988).

It is interesting to note the lack of polar and charged residues in the iron proteins, the relative prevalence of aromatic residues, particularly Phe, in the copper proteins, and the greater distribution of polar and charged residues in the manganese and zinc proteins. This may be related to protein function and the rôle played by the metal ion.

2.5 Discussion—Metals and Metalloproteins

The selection of a metal ion for a particular function is based on its thermodynamic and kinetic parameters (Table 9). The energy change involved in the formation of a metal cation by removal of electrons from the neutral atom, the ionization energy, depends on the immediate environment of the atom. Ionization energy varies in a periodic fashion. Elements that have high values, implying electronic stability, are associated with atoms and ions having filled sets of orbitals.

The Group II *s*-block metal ions, Ca(II) and Mg(II), are particularly stable and generally do not undergo any further changes in electronic configuration, so that

the ionic charge is largely independent of the local environment. In contrast, the *d*-block transition metals (which include manganese, iron, copper and zinc) readily form several ions, each of different electronic configuration. The stability of these ions depends critically on their immediate environment. The larger atomic radii of the *d*-block metals compared to those of the *s*-block metals is reflected by an increase in ionization energies because of the greater nuclear (*i.e.* positive) charge.

An exception to this general trend is observed in the second ionization energy of copper which is anomalously high. This occurs because an electron is being removed from a filled d^{10} shell. The anomaly is also found when an electron is removed from a half-filled d^5 shell as in the formation of Mn(II) and Fe(III). This increased difficulty in removing an electron from such configurations is referred to as the stability of the half-filled or filled shell. It is useful in understanding the chemistry of manganese, iron, copper and zinc. Examples are: the stability of Mn(II), Fe(III), Cu(I) and Zn(II) ions; the anomalous ground state configuration of copper ($3d^{10}\,4s^1$); and the facts that the second ionization energy of zinc is lower than, and the third ionization energy of zinc greater than, those of copper.

The differences between the successive ionization energies of the *d*-block metals are relatively small because the *s* and *d*-electrons do not differ significantly in energy. This leads to the occurrence of various stable oxidation states. The lower energy *s*-electrons are lost in ion formation. In solution, the free energy of formation of the metal ion depends principally on the difference between two large factors; the ionization energy and the hydration energy. It is the balance between these two factors that ultimately determines how many electrons will be lost from a particular metal atom.

The size of a metal ion is given by its ionic radius which is defined by the outer electron configuration. The oxidation number is the most important factor affecting the ionic radius, since it represents the number of outer electrons lost. It can be seen from Table 9 that, with the exception of magnesium and zinc, for ions having the same oxidation state, the ionic radius decreases as the number of inner valence (*d*-shell) electrons increases. The extra attraction of the nucleus for the outer electrons (owing to an effective increase in positive charge) is not neutralized exactly by the extra valence electron in the outer shell, since electrons in the same shell do not screen each other well from the nucleus. Therefore, all the outer electrons are pulled in towards the nucleus and the radius decreases. The outer electrons in the Mg(II) ion are held more firmly owing to its small size so consequently, it has the smallest di-positive ionic radius. The radius of the Zn(II) ion is intermediate between those of Fe(II) and Cu(II). This could result from a balance between the greater nuclear attraction and the filled d^{10} shell.

The hydration energy of an ion is dependent on its charge, ionic radius (the charge/radius ratio defines the charge density) and the nature of the electronic core—the effective nuclear charge. The hydration energy decreases as the size of the ion increases. Thus, the small Mg(II) ion has a higher hydration energy than the larger Ca(II) ion. Ions of the same charge have similar hydration energies which are much greater than those of lesser charge. For the ions Mn(II), Fe(III), Cu(I) and Zn(II), the slight increase in the hydration energies correlates with the slight decrease in ionic radii.

The magnitude of the hydration energies are a measure of the strength of the bonds between the metal ion and the water molecules which form the first hydration

sphere. For highly charged ions such as Mg(II), the electrons on the oxygen atom are polarized by interaction with the metal ion to such an extent that the bonding is a combination of electrostatic and covalent components. This is also true for the *d*-block metal ions.

The hydrated metal ion is not an inert unit. Water molecules in the first hydration sphere exchange with the bulk solvent water. Solvent exchange rates for the hydrated metal ions are given in Table 9. Like the hydration energies, the rates are dependent on the effective nuclear charge. Hence, the small, highly charged Mg(II) ion has the slowest exchange rate, while the larger Ca(II), Mn(II) and Zn(II) ions have higher rates (Cu(II) has the highest rate owing to the rapid exchange of axial water ligands resulting from Jahn-Teller distortion (see Section 2.5.2.3)). This has biological significance.

The selection or rejection of, or the discrimination between metal ions may depend on the rates of their ligand exchange processes. The involvement of particular metal ions in certain biological functions can also be linked to their ligand exchange rates (Table 10). For example, the speed of ligand substitution around Ca(II) allows it to diffuse rapidly—a property that has been employed in signal transduction; the rapid exchange of substrates and products by Zn(II) is one of the factors that contributes to its catalytic function in zinc-dependent enzymes; and the slower exchange rate of Fe(II) could be involved in the oxygen transporting function of this ion.

Property	Mg	Ca	Mn	Fe	Cu	Zn
Outer electronic configuration	$3s^2$	$4s^2$	$3d^5\,4s^2$	$3d^6\,4s^2$	$3d^{10}\,4s^1$	$3d^{10}\,4s^2$
1st	738	590	717	762	745	906
2nd Ionization energies Kj/mol	1450	1146	1509	1561	1958	1732
3rd	7730	4941	3259	2958	3556	3828
Hydration energy M^{2+} Kj/mol	1900	1565	1815	1890	2075	2017
Principle oxidation states	+2	+2	+2 +3	+2 +3	+1 +2	+2
Ionic radius/Å	0.65	0.99	(+2) 0.80 (+3) 0.66	(+2) 0.76 (+3) 0.64	(+1) 0.96 (+2) 0.69	0.74
Ligand exchange rates/sec†	10^5	10^8	(+2) 10^7	(+2) 10^6	(+2) 10^9	10^7

Table 9: Specific physical properties of selected metal ions (Kneen *et al.*, 1984). †Approximate rate constants for the exchange of solvent water with the hydrated metal ion (Taken from Hay (1987)).

Classification	Mg^{2+} Ca^{2+}	Zn^{2+}	Mn^{2+} Fe^{2+} Cu^{2+}
Function	Structure formers and triggers	Super-acid catalysts	Redox catalysts
Mobility	Semi-mobile	Static	Static
Preferred ligands	Oxygen anion binding	Nitrogen/Sulphur donors	Nitrogen/Sulphur donors
Strength of complex	Moderately strong	Strong	Strong
Ligand exchange	Moderately fast exchange	No exchange	No exchange

Table 10: Biochemical classification of cations (Modified from Williams (1985)).

2.5.1 Calcium and Magnesium

2.5.1.1 Calcium

Calcium is a characteristically hard Lewis acid and its functions are due, at least in part, to its ionic radius ($r = 0.99$Å). This means that it has a much lower charge density so ligand substitution reactions are fast ($\approx 10^8$ sec^{-1}), allowing calcium to move quite rapidly through living tissue. Studies of the calcium complexes of small ligands have established that oxygen, a hard base, is the preferred donor atom (as observed in the survey). The coordination number is usually eight, but seven and six-coordinate are known. For eight-coordination, the geometry is best described as a square antiprism. Seven coordinate complexes are pentagonal bypyramidal and six-coordinate calcium is near octahedral. Calcium–oxygen bond distances can be correlated with coordination number (eight-coordinate 2.5Å, seven-coordinate 2.4Å and six-coordinate 2.3Å) (Hay, 1987) (*c.f.* Table 2). Proteins which employ calcium in their function select the metal on the basis of both its size and charge, thus excluding other hard metal ions (Snyder *et al.*, 1990).

The preference of calcium for oxygen ligands can be clearly seen from Table 2, with no nitrogen or sulphur donors represented. The mean metal-ligand distances indicate the type of coordination geometry involved, with values for Glu and hydroxyl (2.5Å) equal to eight-coordinate, Asp, Asn and Gln (2.4Å) equal to seven-coordinate and mainchain carbonyl groups (2.3Å) equal to six-coordinate ligation. The binding of calcium by carbonyl groups reflects the location of metal sites in loops or turns *i.e.* protein regions with no regular secondary structure (Chakrabarti, 1990). Water is also a common ligand and may function as a bridging group between

the metal ion and a distant protein ligand.

The secondary structure of the calcium binding proteins examined is consistent with the function of calcium in the maintenance of protein structure or as a trigger of mechanical action. Enzymes or sugar binding proteins use calcium to maintain structure. In these proteins the calcium ion plays a rôle analogous to that of the structural zinc ion in Alcohol Dehydrogenase (7ADH) for it is usually bound within a loop, but does not take part in protein function. However, when calcium is acting as a mechanical trigger, the proteins with which it interacts have evolved to convert the action of calcium binding into mechanical action (Williams, 1985). Consequently, these proteins are based on helices through which mechanical movement can be transmitted. This is best exemplified by the helical EF-hand proteins involved in contractile systems.

The EF-hand motifs also have characteristic spacer sequences. They are all 12 residues in length. Positions 1, 3, 5, 7, 9 and 12 coordinate the metal ion (position 7 *via* the mainchain carbonyl oxygen). Position 6 and 8 have conserved Gly and Ile residues (da Silva and Reinach, 1991) (Figure 3). The Gly at position 6 could be important for loop conformation, maintaining the mainchain carbonyl group at position 7 in the correct orientation to bind the metal ion; the Ile at position 8 may be involved in the small β-sheet which acts as a fulcrum for the movement of the helices upon calcium binding ((Williams, 1985) and references therein). The preference for small spacer sequences in the other calcium binding proteins, possibly demonstrates how the protein accommodates sufficient ligands in the correct orientation for the calcium binding geometry (six, seven or eight-coordinate) required. Within the spacer sequences, the general preference for Gly

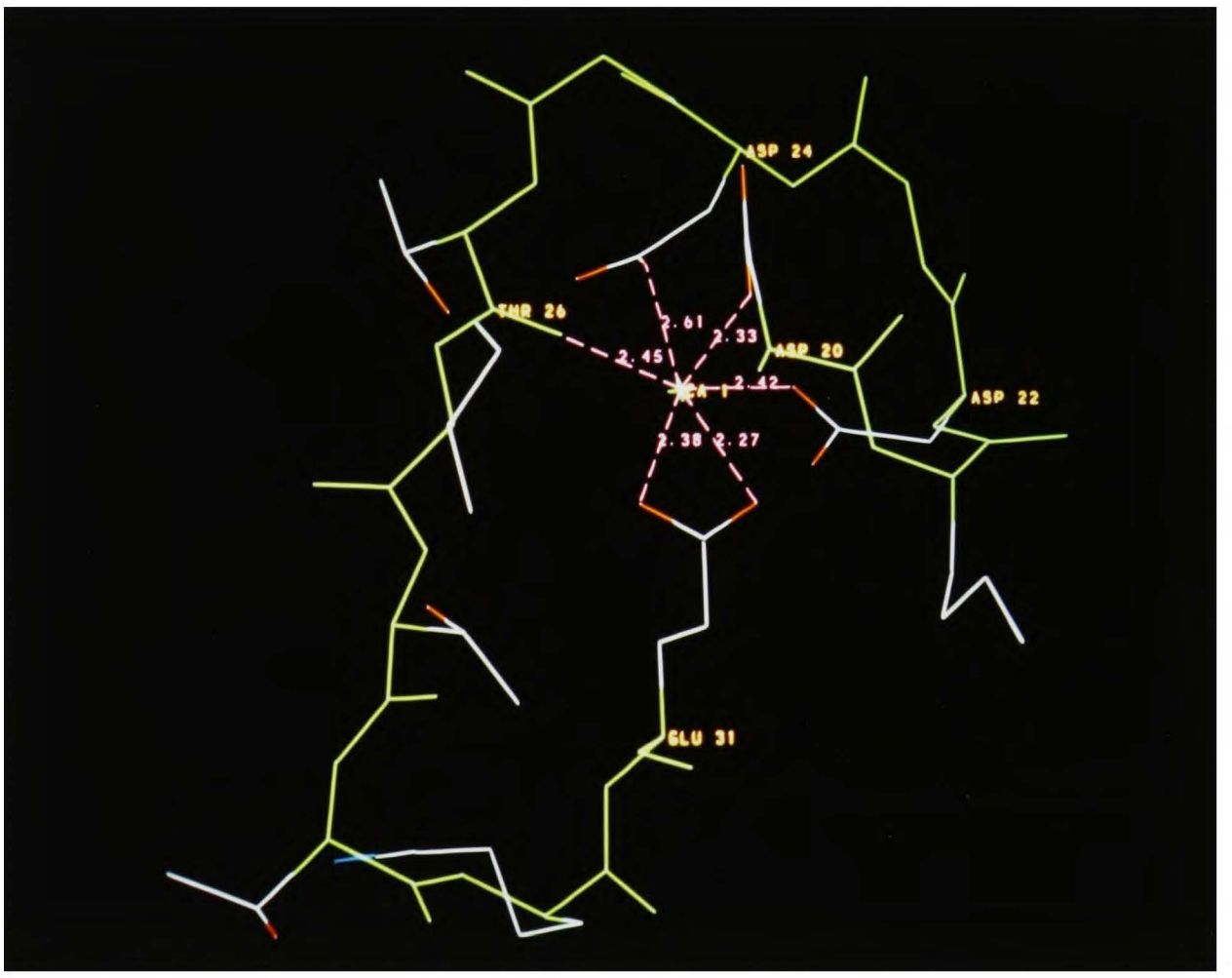


Figure 3: An EF-hand calcium binding site from Calmodulin (Babu *et al.*, 1988). Note the liganding residues Asp-20, Asp-22, Asp-24, Thr-26 carbonyl and Glu-31 which correspond to liganding positions 1, 3, 5, 7 carbonyl and 12, the conserved Gly-25 (not labelled) corresponding to position 6 and the conserved Ile-27 (not labelled) corresponding to position 8 in the EF-hand calcium binding sequence.

and residues with small sidechains may be indicative of the greater conformational flexibility required to enable calcium binding. The prevalence of polar and small charged sidechains within the spacer sequences, could be important in determining the binding affinity at the metal site (Pantoliano *et al.*, 1988).

2.5.1.2 Magnesium

Magnesium is another hard Lewis acid and, like calcium, oxygen is the preferred donor atom. Only one example of a protein which directly binds magnesium, Xylose Isomerase (5XIA), was available so no predictions as to the general nature of magnesium binding in proteins could be made. However, the metal-ligand distances in this protein are shorter than for calcium, demonstrating the smaller cavity required to accommodate the magnesium ion ($r = 0.65$Å).

Magnesium generally functions by reaction with the substrate, while in contrast, calcium exerts its function through interactions with proteins and usually has little contact with substrates[1]. The magnesium ion in 5XIA has direct contact with the substrate, but it is also bound within the protein and thus shows some similarity to calcium action in this case. By comparison, calcium can show similarities to magnesium action as seen from the calcium-substrate interaction observed in Staphyloccocal Nuclease, where the calcium is essential for substrate binding and

[1]One exception is the formation of trypsin from trypsinogen. The pro-enzyme contains two calcium binding sites; a low-affinity site and the high-affinity site also present in trypsin. The low affinity site is formed by four aspartate sidechains located in the N-terminal hexapeptide *1Val-$(Asp)_4$-Lys6*. Cleavage of this peptide between Lys-6 and Ile-7, produces active trypsin. In the absence of calcium, slow non-specific cleavage of trypsinogen occurs producing inert protein. The presence of calcium shields the negative charges of the aspartate carboxyl groups adjacent to Lys-6, directing and enhancing tryptic hydrolysis between Lys-6 and Ile-7. Removal of the hexapeptide promotes activation of the enzyme, although the mechanism is unclear. Calcium binding at the high affinity site induces a conformational change in trypsin which protects it against the formation of inert protein (Radhakrishnan *et al.*, 1969).

activation of the enzyme through a conformational change.

By far the largest group of enzymes which require magnesium in order to exhibit their full activity, are those that catalyse the hydrolysis or cleavage of various polyphosphates. This class includes enzymes such as alkaline phosphatase, the ATPase of the 'sodium pump', hexokinase, deoxyribonuclease (DNase 1) and the protein kinases. The magnesium, acting as a Lewis acid, binds to the phosphate groups to form the magnesium-polyphosphate complex found in the reactions of all these enzymes. This increases the rate of hydrolysis by modifying the shape, reducing the steric hindrance and polarizing the molecule thus increasing the likelihood of nucleophilic attack by water on the terminal phosphorus. (A similar rôle is played by zinc which can activate alkaline phosphatase in place of magnesium (see also Section 2.5.2.4)) (Phipps, 1976).

2.5.2 Manganese, Iron, Copper and Zinc

The ions of the *d*-block transition metals, unlike those of calcium and magnesium, are able to exist in a variety of oxidation states. In many cases, the interaction between the ligand(s) and the metal favours one particular oxidation state at the expense of others. The ligands tend to have oxygen, nitrogen and/or sulphur-donor atoms and, generally, either the +2 or +3 oxidation states are the most stable. However, those metals involved in electron transport or catalytic (enzymic) oxidations have the ability to move reversibly between different oxidation states.

The three common geometries for the transition metal complexes are octahedral, tetrahedral and square planar, but in proteins these geometries may be distorted due to the strict requirements of the metal binding site (Table 11). Such distortions

are important since they may result in considerable modification of the stability of any particular oxidation state (*e.g.* Cu(II) in the copper proteins (Section 2.5.2.3) (Phipps, 1976).

2.5.2.1 Manganese

Manganese is a hard cation, binding to both nitrogen and oxygen donors. Its enzymatic functions are primarily attributed to its Lewis acid nature. Mn(II) is d^5 with a high spin t_{2g}^3 e_g^2 configuration which does not provide any ligand field stabilisation energy. Complexes of Mn(II) are therefore thermodynamically less stable than the succeeding cations of the first row transition series (Fe(II) to Cu(II)). The ionic radius of Mn(II), 0.90Å, is greater than that for Mg(II), but less than that for Ca(II), allowing manganese to replace both magnesium and calcium *in vivo* (Phipps, 1976).

Of the three manganese binding proteins examined, only Glutamine Synthetase (2GLS) binds manganese exclusively. The others (Pea Lectin (2LTN) and Concanavalin A (3CNA)) bind the ion together with calcium. All the manganese binding proteins contain substantial regions of β-sheet and the ion binding sites are located within them.

In 2LTN and 3CNA, the number of ligands necessary to bind two metal ions in close proximity with the appropriate geometry, may generate the long sequences which contain virtually all of the liganding residues (Figure 4). The length of these sequences implies flexibility in metal coordination. This could be linked to their function as sugar binding proteins, which is dependent on metal binding (Sadhu and Magnuson, 1989).

Reaction	Requirements	Example
Electron transfer	1. Low rearrangement energy 2. Local mobility 3. Redox potential matching 4. Centre-to-centre distance of less than 20Å	Low-spin Fe (Rubredoxin), Cu (Copper proteins)
Acid catalysis (Lewis)	1. High electron affinity 2. Open coordination site 3. Easily replaceable ligand 4. Fast off reactions 5. No competing oxidations	Zn (Zinc enzymes)
Atom transfer (N-atom reduction)	1. Effective change of oxidation state by one, two or three units 2. Open coordination site or 3. Easily replaceable ligand 4. Easy exchange of atoms to transfer.	Oxo-state Mn (Glutamine Synthetase)

Table 11: Basic requirements in the catalyses of selected transition metals (Modified from Williams (1985)).

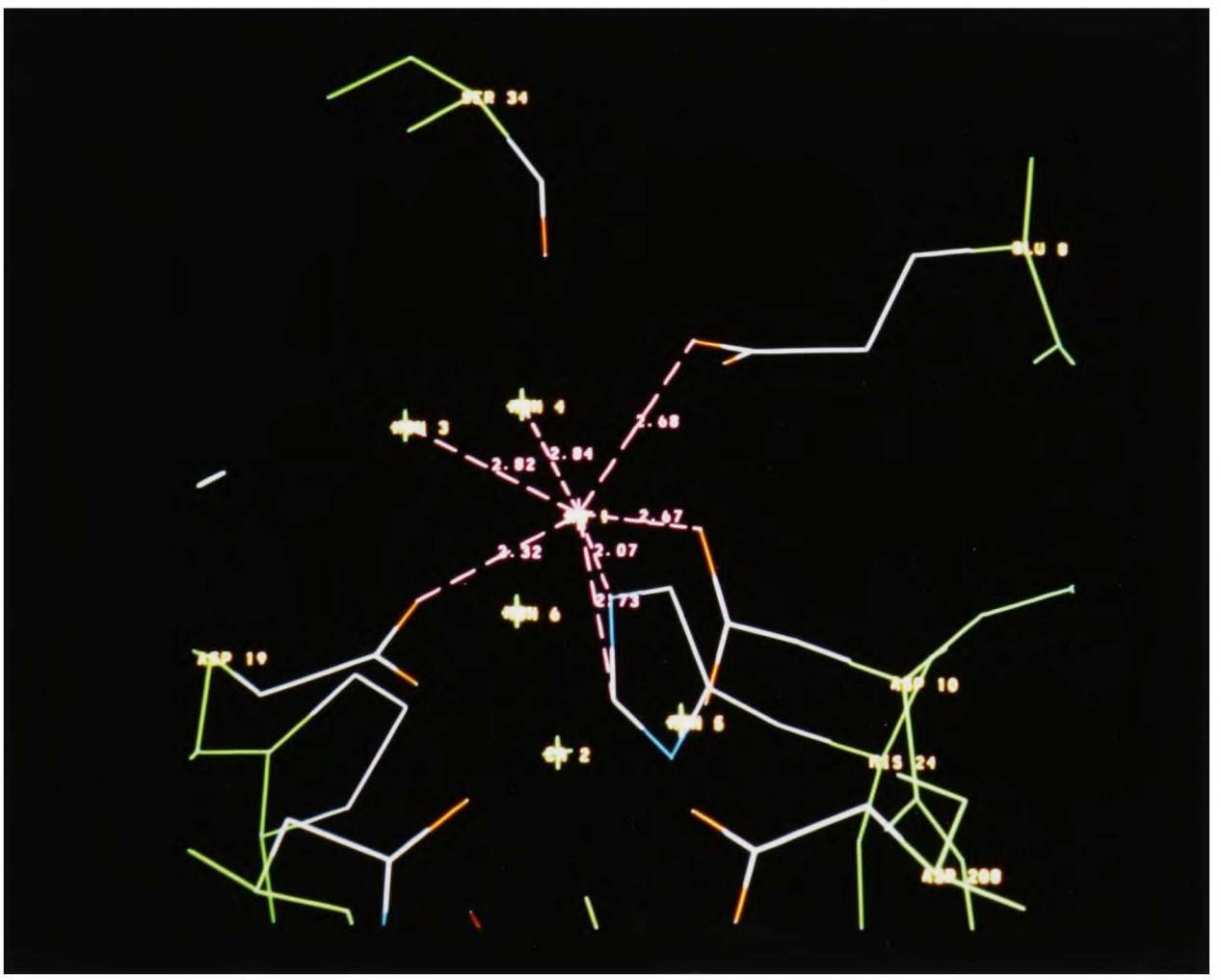


Figure 4: The calcium, manganese binding site of Concanavalin A (Hardman and Ainsworth, 1972). The metal binding sites are 4.5Å apart and all but two of the ligands coordinating the metal ions are contained within a 16 residue sequence. The coordination sphere of the manganese ion is highlighted. Ser-34 is coordinated *via* a bridging water molecule (HOH 4). A similar arrangement is found for Asp-208 in the calcium coordination sphere.

The spacer sequences and the nearest neighbour residues show a wide spread of residue types. The spacer sequences contain more small hydrophobic sidechains (which may be important for the structure of the metal binding sites), while the nearest neighbour residues contain more charged sidechains (which could influence metal binding affinity (Pantoliano *et al.*, 1988)).

2.5.2.2 Iron

Iron provided limited examples of directly bound ions compared to the number of examples of haem proteins available for study. This meant that few general features of iron binding proteins could be described. Iron is classified as a borderline cation. It binds to both hard (nitrogen and oxygen) and soft (sulphur) donors. However, there is a distinction between the function of iron bound by the two ligand types. Iron is involved in oxygen transport and redox reactions. The most characteristic iron proteins are those containing the haem group. These were not included in the survey, since the iron is bound by a porphyrin chelate.

Iron binding by the oxygen transport protein Hemerythrin (1HMQ) shows similarities with the oxygen-transporting haem iron. In haem, the iron is bound with hexa-coordinate octahedral geometry. Four nitrogen atoms are donated by the porphyrin ligand in a square-planar arrangement, while one fifth axial nitrogen comes from a His ligand. (The coordination sphere is completed by molecular oxygen which binds at the other free axial coordination position.) The similarity with 1HMQ results from the octahedral geometry around one of the two iron atoms. Four of the liganding atoms (two carboxyl oxygens and two histidine nitrogens) are in a square-planar arrangement and one axial position is occupied by a histidine

nitrogen. (The remaining axial position would be occupied by oxygen.) The other iron atom is bound with penta-coordinate ligation by equal numbers of both oxygen and nitrogen donors, in trigonal bypyramidal geometry (Figure 5). This contrasts with the other iron protein examined, Rubredoxin (5RXN), which is an electron transport iron-sulphur protein with a single iron atom bound tetrahedrally to four Cys sidechains.

The secondary structure of 1HMQ, a four helix bundle, shows additional similarities with the oxygen-transporting haem proteins which are also helical, while that of 5RXN shows similarities with the copper and electron transport haem proteins, consisting entirely of β-sheet. Electron transport is facilitated by the rigid structure that the β-sheet provides (Williams, 1985).

The structure/function relationship in 5RXN is clearly seen. The iron proteins show structural similarity to proteins with similar function, even though the mode of metal binding or the type of metal bound is different. This is further illustrated by the residues surrounding the iron binding site in 5RXN. The majority of spacer and nearest neighbour residues are hydrophobic, with approximately equal numbers of aliphatic and aromatic sidechains as nearest neighbours. This has functional importance, for solvent or hydrophilic sidechains close to the iron site could capture electrons, inhibiting the transfer process. In addition, the capture of electrons by either of these could result in the production of charged or radical species, such as the superoxide anion, that are highly reactive and toxic. It is interesting to note that the presence of aromatic side chains, particularly Trp, suggests a rôle for these residues in providing a conduction pathway for electrons onto the iron atom (Creighton, 1984).

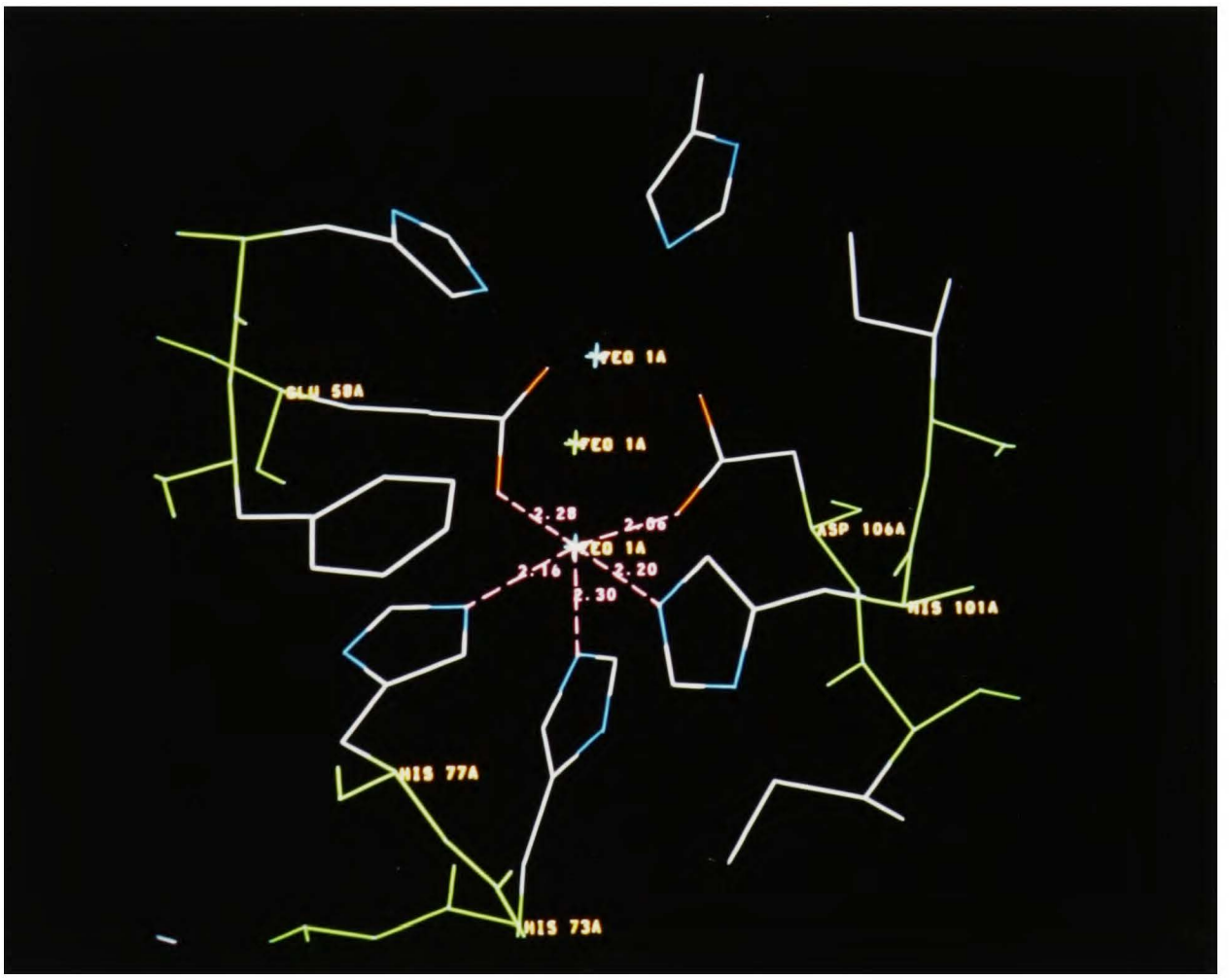


Figure 5: The iron binding sites of Hemerythrin (Stenkamp *et al.*, 1983). The iron binding sites are 3.2Å apart and are bridged by a mu-oxygen atom (yellow FEO 1A) in this structure. The hexa-coordinate iron atom is highlighted, showing the octahedral geometry of the coordination sphere.

2.5.2.3 Copper

Copper is classified as a borderline/soft cation. It is capable of forming strong complexes. Cu(I) and Cu(II) are the most usual oxidation states and reversible changes between them are not difficult. Cu(III) may also be important biochemically. The reversibility of oxidation states is reflected in the function of copper which forms an important part of many oxidases, redox and electron transport proteins. It is believed to act by cycling between oxidation states.

Cu(I) and Cu(II) have very different stereochemistries. Cu(I) is often linear, two-coordinate, but tetrahedral, four-coordinate stereochemistry is common. In contrast, Cu(II), with a d^9 configuration, adopts a tetragonally-distorted octahedral stereochemistry. This distortion results from Jahn-Teller effects that arise from the metal ion itself. It occurs when the electronic state of the ion is such as to give rise to more than one equivalent energy level—a degenerate state. Instead of six equivalent bonds, octahedral complexes generally have four short and two long bonds (Phipps, 1976).

An important principle which affects electron transfer reactions is the Franck-Condon principle. This states that there must be no movement of nuclei during the period of the electronic transition. As a result, the geometry of the two species after electron transfer must be identical to that existing before the transfer occurred (Hay, 1987). Vallee and Williams (1968) have proposed that the active site of a metalloenzyme is in a geometry approaching that of the transition state of the appropriate reaction and as such, is uniquely suited to catalytic action. They have used the term 'entatic state' to describe this situation. The geometry of other metals *e.g.* zinc, is adjusted by the protein for optimum function (see Section 2.5.2.4).

Four of the five copper proteins examined are electron transport proteins. In these proteins, the copper is bound by a combination of hard nitrogen and soft sulphur donors, which supports the borderline/soft characteristics of this metal and represents a compromise between the requirements of Cu(I) and Cu(II). The geometry of binding is square pyramidal. The sulphur atom of a Met sidechain lies at the apex of the pyramid. Of the 27 metalloprotein structures examined, Met ligation was observed only in the copper proteins. In Azurin (1AZU), a mainchain carbonyl group substitutes for Met. This was the only example of carbonyl ligation observed not involving calcium. These apical ligands may be responsible for the square pyramidal geometry which differs by up to 50° from tetrahedral values (Figure 6).

In Superoxide Dismutase (2SOD), the entatic state stereochemistry is generated and controlled, in part, by a zinc ion bound adjacent to the active site copper (Williams, 1984). The zinc ion also stabilises the structure at this point. The stereochemistry is intermediate between planar Cu(I) and tetrahedral Cu(II) geometries and may represent a compromise between the requirements of these two oxidation states.

The secondary structure of all the copper proteins studied, consists predominantly of β-sheet and the copper binding sites are all located within the sheet regions. This is consistent with a rigid binding site to maintain the entatic state. β-sheets are rigid, permitting very little movement of the sheet itself (Williams, 1985). This is in sharp contrast to the helical calcium proteins (see Section 2.5.1.1).

The distorted tetrahedral geometry and β-sheet structure of the copper proteins links the Franck-Condon principle with the entatic state of Vallee and Williams. The intermediate geometry facilitates the Cu(I)/Cu(II) cycling required for rapid

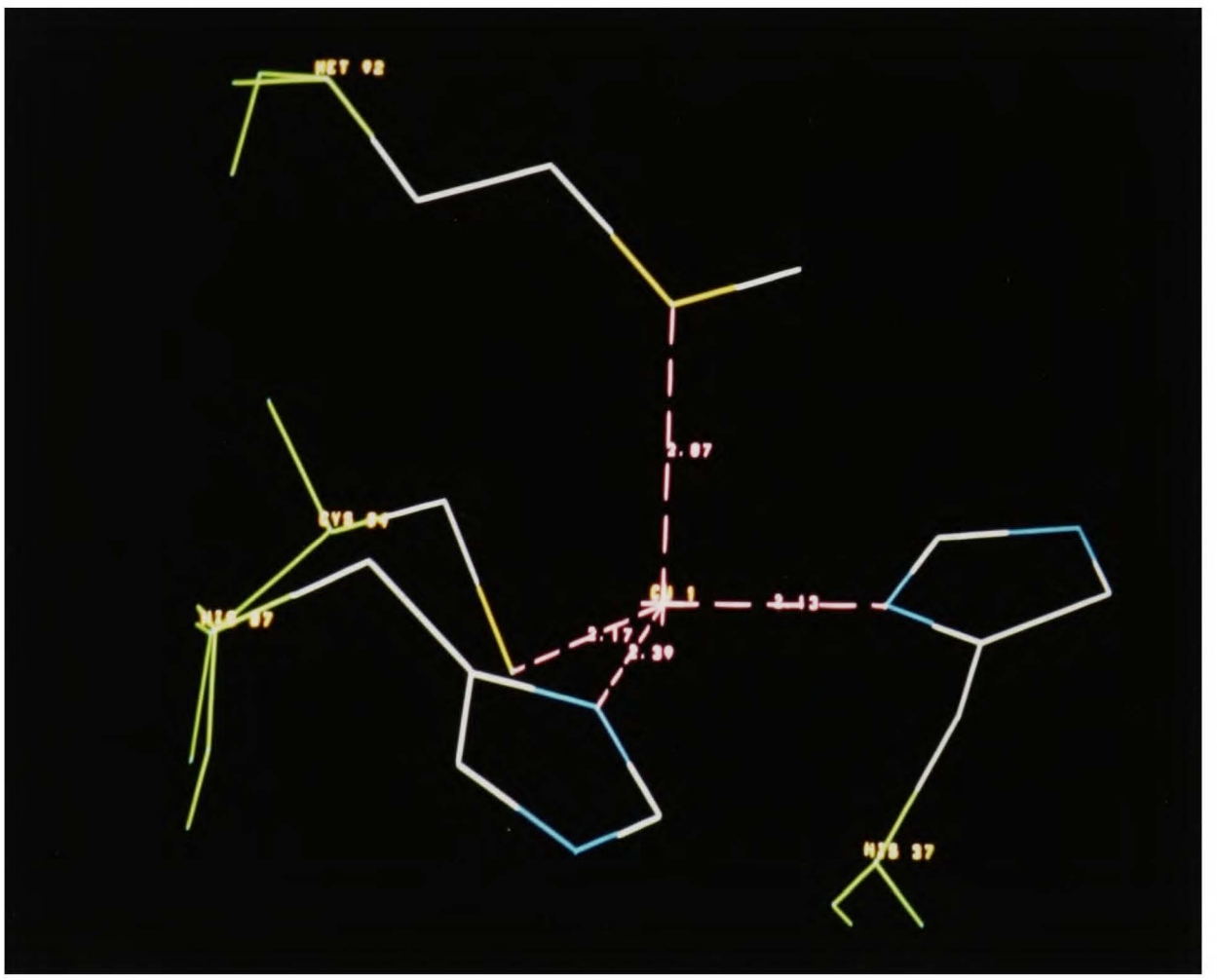


Figure 6: The copper binding site of Plastocyanin (Guss *et al.*, 1986). The square pyramidal geometry with Met-92 at the apex, allows the copper ion to cycle rapidly between the Cu(I) and Cu(II) oxidation states.

electron transfer (entatic state) without the adjustments required to flip the copper ion between Cu(I) and Cu(II) geometries (as stated by the Franck-Condon principle). Both these functional characteristics are maintained by the rigid β-sheet structure.

The preference for one short and one long spacer sequence between ligands, may suggest that flexibility is required to allow the distorted geometry necessary for optimal function. The large number of Gly residues found in these sequences may also be important in this respect as they allow greater conformational freedom. However, this freedom may be constrained by the preponderance of Pro residues. A conformational balance may thus exist between these two residues to create the distorted geometry required for optimum copper function.

For the copper proteins in general, the range of sidechain types found within spacer sequences or as nearest neighbours is small, with charged and polar side chains having the fewest examples. However, a broad preference for residues with hydrophobic sidechains can be discerned. The functional significance of this is the same as for the iron proteins (see previous Section).

2.5.2.4 Zinc

[2]Zn(II) is a symmetrical d^{10} ion and a borderline cation, forming strong complexes with nitrogen, oxygen and sulphur donors. It has no ligand field effects associated with it to determine a particular coordination number or geometry. Ligand exchange processes on Zn(II) are rapid ($\approx 10^7$ sec^{-1}) so that substrates and products can be readily introduced and removed. The enzymatic functions of zinc in small molecule

[2]A much more complete perspective of the function of zinc and the structure of zinc proteins can be found in Williams (1984; 1987) and Vallee and Auld (1990).

hydrolysis are primarily due to its Lewis acidity and in this respect it is similar to manganese. This activity can be 'tailored' by the liganding groups from the protein. Ligands are Lewis bases and as a result, the stronger the σ-electron donation from the ligand to the metal, the weaker the Lewis acidity of the metal centre. Ligands such as His which are both σ-donors and π-acceptors may tailor the metal ion's Lewis acid character by the appropriate degree of π-acceptance (Hay, 1987).

The secondary structure of the zinc proteins examined is a mixture of α-helical and β-sheet structural elements. The zinc binding site is usually located at the bottom of a deep central cleft which forms the substrate binding pocket. The site is found in regions of both α-helix and β-sheet, although β-sheet is favoured slightly.

With the exception of Alcohol Dehydrogenase (7ADH), the zinc enzymes examined showed a characteristic spacer sequence. Vallee and Auld (1989) have analyzed these spacer sequences. They propose that the short spacer sequence between two of the ligands ensures a nucleus for zinc binding and stabilises the protein conformation at that point, while the longer sequence between the second and third ligands not only secures the zinc coordination site, allowing flexibility in coordination geometry (the entatic state), but also brings into alignment the substrate binding groups critical for catalysis. Variation in spacer length could reflect differences in substrate specificity and catalytic mechanism.

It is interesting to note that the serine proteinase Tonin (1TON) which does not normally bind zinc, is inhibited by zinc binding to three His sidechains at the active site (one of which is involved in the catalytic triad). These three residues show the characteristic spacing of the ligands in the zinc enzymes, suggesting that a putative zinc binding site exists in this enzyme. Metal ion inhibition has been engineered

into other serine proteinases and other enzymes to control their activity (Higaki *et al.*, 1990; Higaki *et al.*, 1992).

The requirement for substrate binding and catalytic groups is reflected in the greater number of polar and charged sidechains close to the zinc binding site, particularly as nearest neighbours. The zinc is bound with three-coordinate ligation in approximately tetrahedral geometry. A fourth ligand is provided by a water molecule. This is displaced by an oxygen atom from the substrate which becomes activated by zinc during the catalytic cycle (Figure 7).

Aspartate Carbamoyltransferase (6AT1) and Alcohol Dehydrogenase show the use of zinc in a structural rôle. Both these enzymes contain zinc tetrahedrally bound to four Cys sidechains (*c.f.* Rubredoxin), cross-linking them. Cross-linking stabilises the structure in the region of the metal ion (*c.f.* calcium in Thermolysin). It can replace disulphide bridges and protects the thiolate groups because it does not catalyse their oxidation. Zinc, then, provides a centre free from one-electron reactions owing to the absence of redox properties shown by this metal (Williams, 1987). These reasons may also be important (Klug and Rhodes, 1987) for understanding the major structural rôle of zinc in zinc-fingers which are found in a large variety of proteins concerned with the regulation of gene expression[3].

Zinc-fingers are small peptide domains consisting of about 30 amino acids in which a zinc ion is tetrahedrally coordinated by various combinations of Cys and His sidechains. It has been proposed (Berg, 1990) that the use of zinc binding domains such as the zinc finger, to mediate interactions between different proteins and macromolecules involved in gene regulation, arose from the reduction in

[3]Reviews on the structure of zinc-fingers in DNA binding proteins can be found in Kaptein (1991) and references therein, and Vallee *et al.*, (1991).

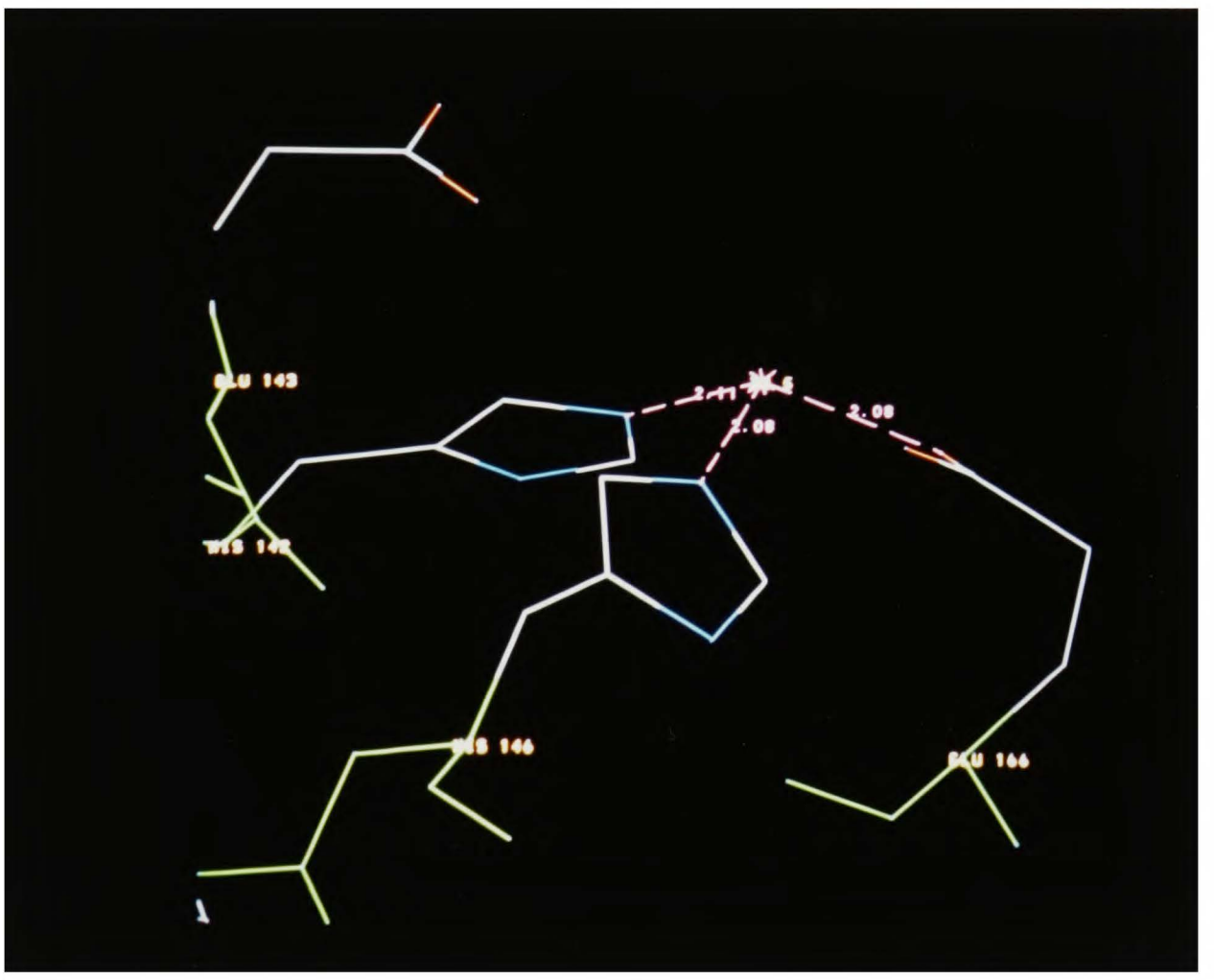


Figure 7: The zinc binding site of Thermolysin (Holmes and Matthews, 1982). The tetrahedral coordination sphere is completed by a water molecule (not shown) which is displaced by the substrate.

conformational entropy imposed by the cross-links that metal binding introduced. Metal-based domains are generally smaller than protein domains which are devoid of cross-links. A small domain (less than 50 amino acids) lacking cross-links may not be able to gain sufficient free energy from favourable interactions to overcome the conformational entropy of the unfolded form.

2.6 Conclusions

The results of this survey have shown that the common features concerning the coordination and function of metal ions, and the structure of metalloproteins, are present not only for a particular metal, but for metalloproteins of similar structure/function which bind different metals. It is these features that become important when a metal function is to be introduced into another protein by protein engineering. Transfer of function may be more challenging for particular metals which are dependent on specialised metal-protein interactions *e.g.* copper-mediated electron transport.

Generally however, metal binding could be achieved by the introduction of appropriate liganding residues in an approximately correct spacial arrangement. The problem is then reduced to selecting a possible position for the new metal binding site in the target protein, where its presence would be of greatest value. Such a task was undertaken and is described in the next Chapter.

Chapter 3

Metalloproteins in Metal Binding Site Design

The metalloprotein survey discussed in Chapter 2 showed the types of liganding arrangements which proteins use to capture and optimise the function of particular metal ions. To create a metal binding site, similar patterns of ligands must be introduced. The functional property required of the metal will dictate which metal ion is to be bound and this, in turn, dictates the nature of the ligands required. However, to ensure sufficient binding strength, it is essential that the ligands are inserted with the geometry and liganding distances preferred by the required metal ion.

While *ab initio* design might be used to create a metal binding site, a site based on a pre-existing metalloprotein structure from the Brookhaven Protein Databank (Bernstein *et al.*, 1977) has the advantage that the correct liganding geometry and distances are inherent in the metal binding pattern. The *ab initio* design of a metal binding site is discussed in Chapter 4. In this Chapter, the design of a metal binding site which mimics that from a pre-existing metalloprotein will be discussed.

3.1 The Choice of Metal Ion

Consideration of the various metal ion functions shown by different metalloproteins (Chapter 2, Section 2.5), led to the decision that the metal site should be modelled on those of zinc. Zinc plays a crucial rôle in many metalloproteinases, being directly involved in peptide bond cleavage. This reaction is probably the most interesting (and potentially the most useful) function to be transferred to an antibody. In addition, zinc does not support damaging redox reactions (Williams, 1987) which would hamper production of any metalloantibody. A second advantage is that zinc is also easily replaced by cobalt (Schneider-Bernlöhr *et al.*, 1988) which is more spectroscopically active, enabling the environment of the metal to be probed by physical techniques. A metal ion bound within the antibody combining site (ACS) may be perturbed by antigen binding. The disturbance may change the spectroscopic properties of the metal and detection of this change could signal the antigen recognition event. This is the first stage in the design of antibody-based biosensors (Lowe, 1986; North, 1985).

3.2 The Metallothionein Model

The ACS consists of six hypervariable loops or complementarity-determining regions (CDRs), which determine specificity and mediate antigen recognition. To be affected by antigen binding, the metal site must be formed within or between the CDRs. However, it must be designed such that antigen specificity and affinity are altered as little as possible.

Previous structural database searches (Martin, 1990) had shown that residues

36–45 from the crystal structure of the protein Cd-Zn metallothionein (Furey *et al.*, 1986) had a conformation similar to that of CDR-L1 in HyHEL-5 (Sheriff *et al.*, 1987) (R.m.s. deviation; 1.3Å on Cα and 1.3Å on backbone (N, Cα, C)[1]). A comparison of residues 36–45 from the metallothionein crystal structure, with the HyHEL-5 L1 crystal structure, is shown in Figure 8.

This chapter will go on to describe how residues 36–45 from the crystal structure of Cd-Zn metallothionein were used as a basis for the design of a metal binding site in CDR-L1 of HyHEL-5. It was later discovered that the metallothionein crystal structure was wrong, so the initial assumptions about conformational similarity and liganding pattern of this sequence (and the metal site design based on it) were now in doubt (see Section 3.7). It is included here to enable a comparison between the use of protein structure and ab initio *methods (Chapter 4) in metal site design. The final design used for creation of the actual metalloantibody (Chapter 5) was not based on the metallothionein structure.*

3.2.1 A Metal Binding Site in HyHEL-5 L1

The conformational similarity of the metallothionein sequence with L1 suggested a possible metal binding site within this CDR. However, before the design of such a site could be carried out, it was necessary to determine whether it could be accommodated by the CDR. The interactions of L1 with other loops in the combining site and with the antigen were examined by interactive molecular graphics using

[1]The r.m.s. deviations throughout this thesis were calculated using the program FIT (Martin, 1990).

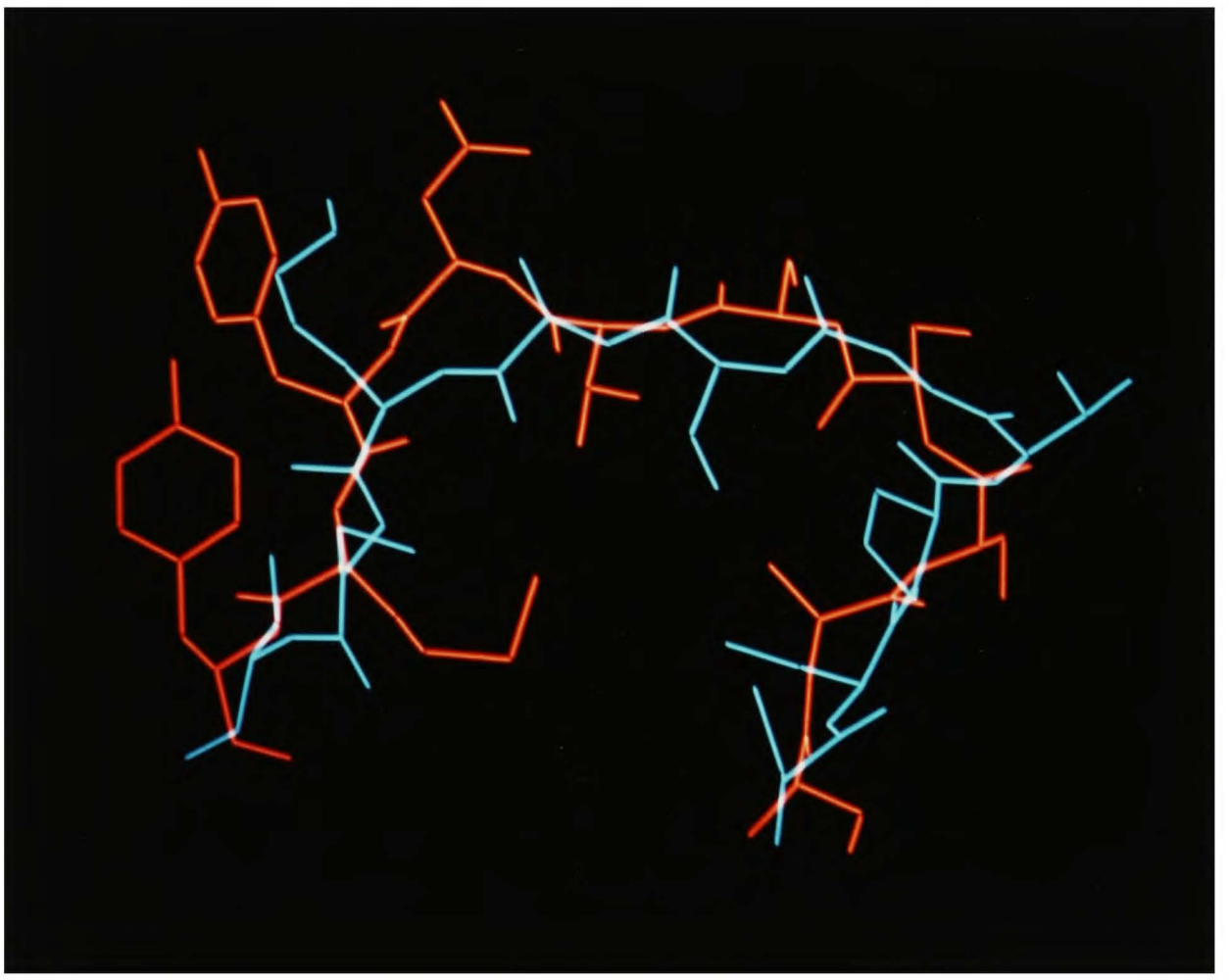


Figure 8: Comparison of the HyHEL-5 L1 crystal structure (red) with residues 36–45 from the Cd-Zn metallothionein crystal structure (blue).

the HyHEL-5/lysozyme complex (Sheriff *et al.*, 1987).

In the combining site, L1 interacts only with CDRs L2 and L3; there are no contacts with any of the heavy chain CDRs. These interactions only involve part of each loop: residues 30–33 of L1, residues 49 and 50 of L2, and residues 88 and 89 of L3. Seven residues from the light chain are in direct contact with the antigen, lysozyme. Of these, the only residues from L1 are Asn-30 and Tyr-31. Thus, all major interactions of L1 within the combining site and with the antigen, are confined to residues 30–33. Hence, residues 24–29 were available for mutation to create a metal binding site without disruption of the antigen contacting residues.

3.2.2 Metal-Ligand Mutations in HyHEL-5 L1

The ten residue sequence 36–45 from the crystal structure of metallothionein contains four cysteine residues. Cys-36, 37 and 41 form 75% of the the Cd-2 binding site in the crystal structure (Figure 9) (though note the comments in Section 3.2). When compared with HyHEL-5 L1, these three cysteines corresponded to Ser-24, Ala-25 and Val-29 (Note also that two of these residues, Ala-25 and Val-29 are structure-determining residues (SDRs). The significance of mutating SDRs is discussed in Chapter 7). Introducing cysteines at these positions in L1 would maintain the spatial relationship seen in the metallothionein crystal structure to give a tricoordinate metal site. This is the minimum number of ligands required to give a metal ion chealate stability (Williams, 1987).

The Cd-2 site of metallothionein (residues 36–41) also contains Proline and Glycine. The presence of these two residues in the sequence and their conformational significance, suggested that they may be important in the maintenance of the Cd-2

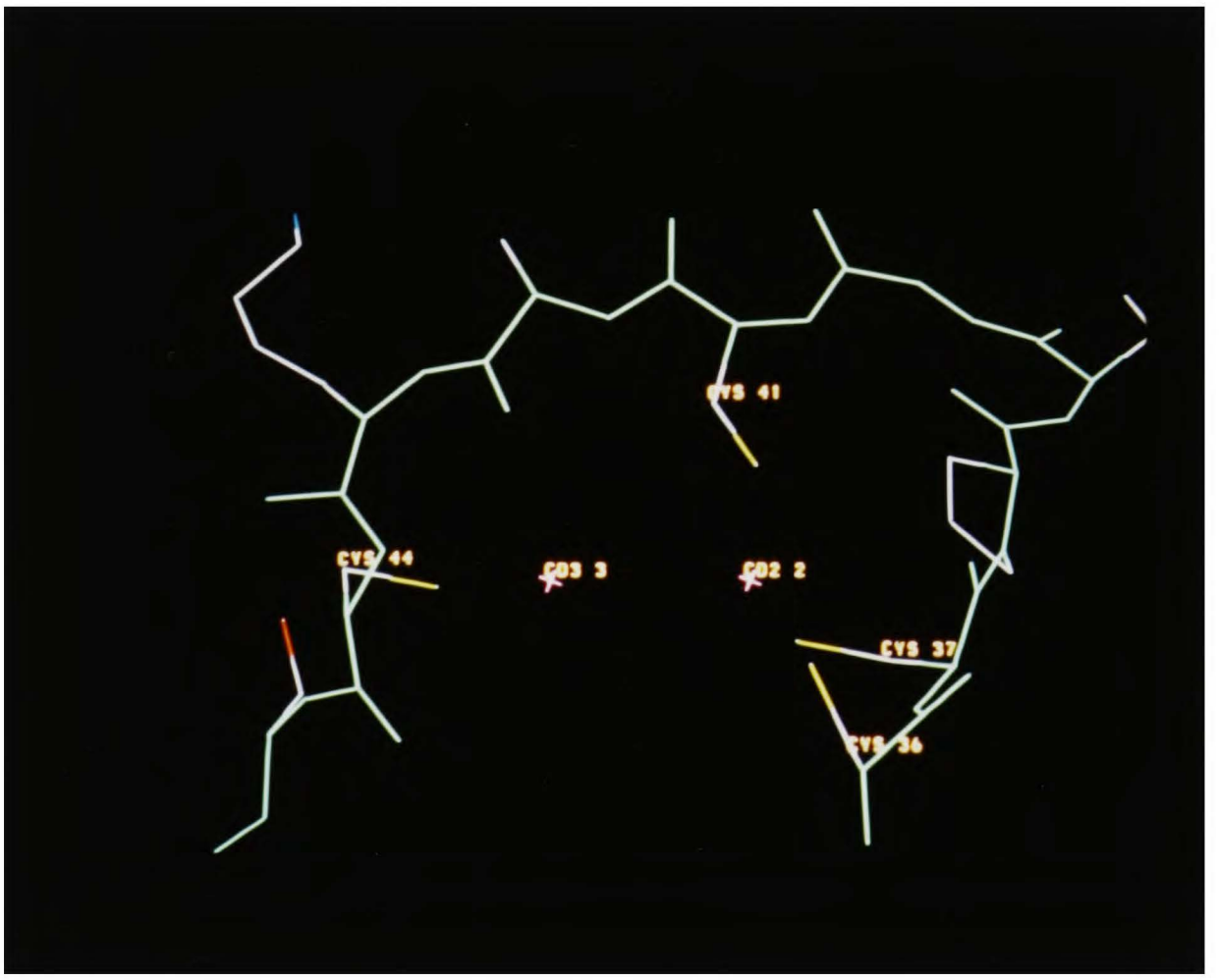


Figure 9: Residues 36–45 from the crystal structure of Cd-Zn metallothionein. Cysteines 36, 37 and 41 coordinate the Cd-2 ion, while Cysteine 44 coordinates the Cd-3 ion.

Structure	Loop sequence									
HyHEL-5 L1	*Ser*	*Ala*	*Ser*	*Ser*	*Ser*	*Val*	*Asn*	*Tyr*	*Met*	*Tyr*
2MT2 36–45	**Cys**	**Cys**	**Pro**	**Val**	**Gly**	**Cys**	**Ala**	**Lys**	**Cys**	**Ser**
*Hybrid*L1	**Cys**	**Cys**	**Pro**	*Ser*	**Gly**	**Cys**	*Asn*	*Tyr*	*Met*	*Tyr*

Table 12: The substitutions in HyHEL-5 L1 necessary to create the Cd-2 site from Cd-Zn metallothionein.

site. Table 12 shows the metallothionein (2MT2) and L1 sequences together with a third sequence, *Hybrid* L1, containing the minimum substitutions necessary to create the Cd-2 site in this CDR.

To determine the effects of the five mutations on the conformation of L1, particularly residues 30–33 which contact the antigen, the *Hybrid* L1 was modelled using both native CDR and metallothionein sequences as starting structures. These were modified to create two starting conformations for the *Hybrid* L1 sequence, using the REPLACE and REFI (Hermans and McQueen, 1974) procedures available within the molecular graphics program FRODO (Jones, 1985). The two conformations were then subjected to conformational search using the program CONGEN (Bruccoleri and Karplus, 1987).

3.3 Conformational Search of Hybrid L1

CONGEN performs a conformational search to saturate the conformational space available to a peptide fragment on a specified grid, by rotation about the backbone ϕ and ψ torsion angles and the sidechain χ angles. This is done by performing a depth first tree search with back-tracking. The ϕ and ψ angles for an amino acid are combined into a single degree of freedom and the amount of conformational

space explored is restricted by looking up ϕ/ψ combinations in a tabulated form of the Ramachandran plot (Ramachandran *et al.*, 1963) and examining only those combinations with energies below a specified cutoff. Three such Ramachandran maps are used: one for proline, one for glycine and one for alanine (which is taken as being representative of all other amino acids). Three residues of the section being constructed are generated using the analytical chain closure algorithm of Gō and Scheraga (1970) (as modified by Bruccoleri and Karplus (1985)). Thus, for a five residue section, two residues are constructed by full conformational search (and will be termed *backbone*), while three are constructed by chain closure (and will be termed *chain*).

In order to improve the efficiency of the algorithm, after each combination of *backbone* torsion angles has been generated, the distance between the end points is checked to ensure that it is possible to span the remaining gap with the number of residues left to construct (when these are in all-trans configurations with the bond angles stretched by 5° from the optimum). Only if this condition is met are further *backbone* residues constructed followed by the *chain* residues and, finally, the sidechains. In addition, conformations are rejected with van der Waals energy above a specified cutoff (typically 20 Kcal/mol for backbone, 100 Kcal/mol for chain and 20 Kcal/mol for sidechains).

Of the five sidechain construction protocols implemented in CONGEN, the ITERATIVE method was employed. This method starts with an energetically acceptable position for all the sidechains. All possible conformations for the first sidechain are then regenerated and the conformation with the lowest energy is selected. The process is repeated for each sidechain in turn and when all sidechains have been

searched, the process starts again from the first sidechain until the energy cannot be further reduced or an iteration limit is reached.

This procedure generates only one energetically feasible sidechain conformation per backbone conformation. If the number of iterations required is large, then the initial generation of a combination of sidechain conformations will bias the iterative procedure and the algorithm may miss the global minimum. However, this procedure has been shown to be the most effective (Bruccoleri and Karplus, 1985) in generating an accurate conformation in a minimum amount of computer time and has been used for all the sidechain constructions (with one exception discussed in Chapter 7, Section 7.2.7) presented in this thesis. In the protocol described by Martin *et al.*, (1989), a five residue fragment in the middle of the CDR is constructed using a 30° search of the torsion angles of the outer 2 *backbone* residues and the Gō and Scheraga chain closure on the middle three *chain* residues. If this fails to find more than 100 conformations, the torsional search is reduced to 15° or, if necessary, 5°. If the search still fails, additional *backbone* residues are searched until conformations are generated.

Four constructions were performed on the *Hybrid* L1. These are shown in Table 13. The whole loop was constructed in five residue sections covering the antigen/CDR contacting residues (CNYMY), the junction between these and the metal site substitutions (SGCNY), and all but one of the residues forming the metal site (CCPSG). In addition, a six residue construction which included all the substituents (CCPSGC), thus covering the entire metal site, was also carried out.

For all the constructions presented in this thesis, the CONGEN constructed

Construction	Total Confs	E_{low}	Energy Kcal/mol	CPU Time
CCPSG[C(NYM)Y]	137	28	30.4237	11
CCP[S(GCN)Y]MY	20	16	37.9705	1
[C(CPS)G]CNYMY	69	10	-26.9531	8
[C(CPS)GC]NYMY	588	154	-27.0275	20

Table 13: CONGEN constructions performed on the *Hybrid* L1. *Total Confs*: Total number of conformations generated; E_{low}: Lowest energy conformation; *CPU Time*: Computer time in hours.

sequence is shown in square brackets, while the *chain* residues are shown in parentheses. All CPU times quoted are for a VAXstation 3200.

3.4 Energy Minimisation of Hybrid L1

An initial investigation was made into the effect zinc might have on the new metal site. Accordingly, a zinc ion was introduced into the site using interactive molecular graphics, approximately equidistant from the Cys-25 and Cys-29 sidechains. The coordinates of the zinc ion were included with those of the lowest energy conformation from each construction generated by CONGEN (Table 13).

In addition, each conformation was minimised both in the presence and absence of zinc. By comparing the final energies of each conformation with and without zinc, the contribution of the metal to the stability of the loop could be estimated. All conformations were minimised in the absence of lysozyme, using the program GROMOS (Åqvist *et al.*, 1985); the results are summarised in Table 14.

The SHAKE algorithm (Ryckaert *et al.*, 1977; van Gunsteren and Berendsen, 1977) could not be used for the minimisation of conformation 154, both with and

Construction	E_{low}	Zinc	SHAKE	Cycles	CPU Time	Energy Kcal/mol	Difference Kcal/mol
CCPSG[C(NYM)Y]	28	No	Yes	330	152	-2163.4	
		Yes	Yes	144	68	-1999.9	163.5
CCP[S(GCN)Y]MY	16	No	Yes	651	295	-2484.9	
		Yes	Yes	1 000	423	-2446.7	38.2
[C(CPS)G]CNYMY	10	No	Yes	538	236	-2448.0	
		Yes	Yes	217	100	-2459.9	-11.9
[C(CPS)GC]NYMY	154	No	No	803	354	-2323.0	
		Yes	No	508	226	-2249.2	73.8

Table 14: Energy minimisation with and without zinc on the lowest energy conformations from the four constructions of *Hybrid* L1. *Construction*: CONGEN constructed residues; E_{low}: CONGEN lowest energy conformation; *Zinc*: Minimisation in the presence or absence of zinc; *SHAKE*: Minimisation with or without the use of SHAKE in GROMOS; *Cycles*: Number of cycles of minimisation completed; *CPU Time*: Computer time in minutes; *Difference*: Energy difference between the minimised conformation in the presence and absence of zinc.

Construction	E_{low}	Zinc	Energy (Kcal/mol)						
			E_{Total}	E_B	E_A	E_I	E_D	E_E	E_{vdw}
CCPSG[C(NYM)Y]	28	No	-2163.4	0.498	571	221	254	-1566	-1644
		Yes	-1999.9	0.494	582	227	279	-1498	-1592
CCP[S(GCN)Y]MY	16	No	-2484.9	12.5	320	76.0	262	1521	-1634
		Yes	-2446.7	12.3	323	77.9	268	-1507	-1622
[C(CPS)G]CNYMY	10	No	-2448.0	12.3	322	77.4	269	-1508	-1620
		Yes	-2459.9	11.8	271	80.2	261	-1490	-1593
[C(CPS)GC]NYMY	154	No	-2323.0	71.4	537	214	233	-1652	-1726
		Yes	-2249.2	71.6	541	217	245	-1617	-1707

Table 15: Contribution of interaction energies to the total energy after minimisation of the lowest energy conformations from the four constructions of *Hybrid* L1. *Construction*: CONGEN constructed residues; E_{low}: CONGEN lowest energy conformation; *Zinc*: Minimisation in the presence or absence of zinc; E_{Total}: Total energy of the structure after minimisation; E_B: Bond energy; E_A: Angle energy; E_I: Improper dihedrals energy; E_D: Dihedrals energy; E_E: Electrostatics energy; E_{vdw}: van der Waals energy.

without zinc. Absence of the bond length constraints imposed by SHAKE, can affect the extent of minimisation and the final energy of the structure. Therefore, a direct comparison of the energies for this conformation with the others cannot be made. Generally however, the reduction in energy is greatest where the construction included the metal site substitutions. Table 15 shows the contribution of various interaction energies to the total energy after GROMOS minimisation.

The electrostatics and van der Waals energies make the greatest contribution to the final energy of each structure, both with and without zinc. The optimisation of electrostatic metal-ligand and packing interactions results in a slightly higher total energy in the presence of zinc, although the difference in energy with and without zinc is small. The favourable electrostatic and van der Waals interactions between the loop and zinc indicates that an attractive electrostatic cavity of sufficient size to accommodate the metal ion exists.

No attempt was made to orientate the coordinating sidechains around the metal ion into a more 'typical' liganding configuration. This was not possible for the Cys-24 sidechain which was placed at the exterior of the loop. The negative partial charge of this sidechain would not have been opposed by the metal ion, reducing the overall metal-ligand electrostatic attraction at the site. Thus, the precise effect of the metal ion on loop stability could not be determined. A more rigorous investigation of metal site electrostatics in energy minimisation was carried out and is described in Chapter 4, Section 4.3.

3.5 Screening the Hybrid L1 Conformations

CONGEN is built around the CHARMM (Brooks *et al.*, 1983) energy minimisation and dynamics package, and 10 cycles of conjugate gradient CHARMM energy minimisation were applied to each conformation as it was generated. This allows the conformations to relax from the selected 30° torsional search grid. In his use of CONGEN, Bruccoleri chooses a conformation from those generated by selecting the lowest energy structure as evaluated using the CHARMM potential *in vacuo*. If however, the next lowest energy conformation is within 2 Kcal, then the conformation with the smaller solvent accessible surface (Lee and Richards, 1971) is chosen (Bruccoleri *et al.*, 1988). However, Martin *et al.*, (1989) have shown that the *in vacuo* CHARMM potential does not rank conformations particularly well. Conformations of low r.m.s. deviation compared with the crystal structure frequently do not fall amongst the lowest energy structures. The low energy conformations result from optimisation of van der Waals packing. In solvent, this attractive effect would be counter-balanced by an attraction of the loop towards solvent (Martin, 1990).

All the conformations from each construction were screened using a modified version of the GROMOS potential (Åqvist *et al.*, 1985) implemented in the program GROMSCAN (Martin, 1990). In order to prevent the optimisation of van der Waals packing which occurs *in vacuo*, the attractive ($-B/r_i^6$) term of the 12-6 Lennard-Jones potential was removed. In addition, the presence of solvent around the loops will lead to a high dielectric constant which reduces the importance of electrostatics. Thus, the electrostatic potential was also removed. The rankings obtained by this 'solvent-modified' potential are substantially better than those

from the *in vacuo* CHARMM potential, with low r.m.s. structures falling among the low energy conformations (Martin, 1990).

In the protocol described by Martin *et al.*, (1989), the final conformation is chosen by calculating the solvent accessible surface area of hydrophobic atoms, for each conformation of the five lowest energy structures evaluated by GROMSCAN and selecting the conformation with the lowest solvent accessibility. For this purpose all sidechain carbons are regarded as hydrophobic except those in charged delocalised groups or adjacent to hydrophilic groups. The solvent accessibility was calculated using the Lee and Richards algorithm (Lee and Richards, 1971) as implemented within CONGEN (Bruccoleri *et al.*, 1988). The accessibilities obtained are shown in Table 16.

Conformation 154 from the [C(CPS)GC]NYMY construction was selected as the final conformation for *Hybrid* L1, on accessibility criteria (Table 16). This construction included all the residues introduced to create the metal site, thus giving the most realistic representation of its overall structure. The r.m.s. deviation between this structure and the L1 crystal structure is 0.5Å on Cα and 0.4Å on backbone. Despite the presence of the conformationally significant residues Pro-26 and Gly-28, and cysteine substitutions in structure-determining positions, the chosen conformation for the *Hybrid* L1 was very similar to the crystal structure.

3.6 Conclusions

The close agreement between these two conformations supports the use of pre-existing metalloprotein structures in the design of novel metal sites (where the

Construction	Conformation	Accessibility Å^2
CCPSG[C(NYM)Y]	6	345.40
	10*	333.53
	26	343.52
	28	346.17
	103	366.16
CCP[S(GCN)Y]MY	9	372.44
	11	369.71
	12	372.62
	15	351.69
	16*	351.23
[C(CPS)G]CNYMY	9	404.93
	10	422.27
	12	436.36
	45	432.31
	46*	383.87
[C(CPS)GC]NYMY	59	441.54
	154*	407.03
	391	433.37
	409	416.40
	426	418.62

Table 16: Solvent accessibilities for the five lowest energy conformations, as calculated by CONGEN, from each construction of *Hybrid* L1. The conformation with the lowest solvent accessibility is indicated by an asterisk.

metalloprotein structure is known to be correct). The maintenance of native structure in the target antibody is more certain, and metal binding at an engineered site more likely, if a close structural similarity can be identified between the antibody and metalloprotein. Roberts *et al.*, (1990) have developed this approach into a method for designing metalloantibodies. They characterised structural patterns in the binding sites of antibodies and zinc enzymes, and identified structurally conserved sites in both the light and heavy chain variable domains which matched the stereochemistry of the zinc liganding pattern in Carbonic Anhydrase B. Iverson *et al.*, (1990) have used these finding to design a metal binding site between CDRs L1 and L3 of the anti-fluorescein antibody 4-4-20 (Herron *et al.*, 1989). The created metalloantibody was capable of binding copper with an affinity constant of about 10^6 M^{-1}.

3.7 Rejection of the Hybrid L1 Model

The success or failure of introducing any novel function into the ACS is dependent on the extent to which the mutations necessary to create the function, disturb the protein structure. When mutations are selected, based on knowledge gained from a crystal structure, the accuracy of that structure is important. Any doubt about its validity would call into question the usefulness of the proposed mutations.

After modelling of the metal site in L1 had been completed, the crystal structure of Cd-Zn metallothionein (Furey *et al.*, 1986) was shown to be wrong. This resulted from the publication of a metallothionein structure determined by NMR (Schultze *et al.*, 1988). It had not been detected earlier since use of the structure had originated

from a random database search of CDR-L1 in HyHEL-5 (Section 3.2), while the survey of metalloprotein structures was carried out. Comparison of the NMR and crystal structures showed that significant differences existed between them.

3.7.1 The NMR Structure of [Cd_7]-metallothionein-2

The coordinates for the *alpha* and *beta* domains of rat liver [Cd_7]-metallothionein-2, as derived from NMR, were obtained from Dr. Werner Braun. R.m.s. deviations between the NMR and crystal structures give an indication of the structural differences. The deviations are 5.6Å on Cα and 5.4Å on backbone for the *alpha* domain, and 3.7Å on Cα and 3.5Å on backbone for the *beta* domain. The values for the Cα deviations are identical to those given by Schultze *et al.*, (1988) and show a greater divergence in the *alpha* domain, between the NMR and crystal structures, than in the *beta* domain.

The *alpha* domain contains residues 36–45 which formed the basis of the metal site in L1. R.m.s. deviations for these residues between the NMR and crystal structures are 4.4Å on Cα and 4.2Å on backbone. A comparison of residues 36–45 from the NMR and crystal structures is shown in Figure 10.

It can be seen from Figure 10 that the conformations of these two loops differ significantly. In addition, the arrangement of the four cysteine ligands in this loop is different (*c.f.* Figure 9). In both the NMR and the crystal structures, Cysteines 36 and 37 serve the same functions. Cys-36 is a singly bound ligand, while Cys-37 is a bridging ligand[2]. However, the functions of cysteines 41 and 44 are reversed between the two structures. In the crystal structure, Cys-41 is a bridging ligand

[2]A bridging ligand is involved in the binding of two adjacent metal ions.

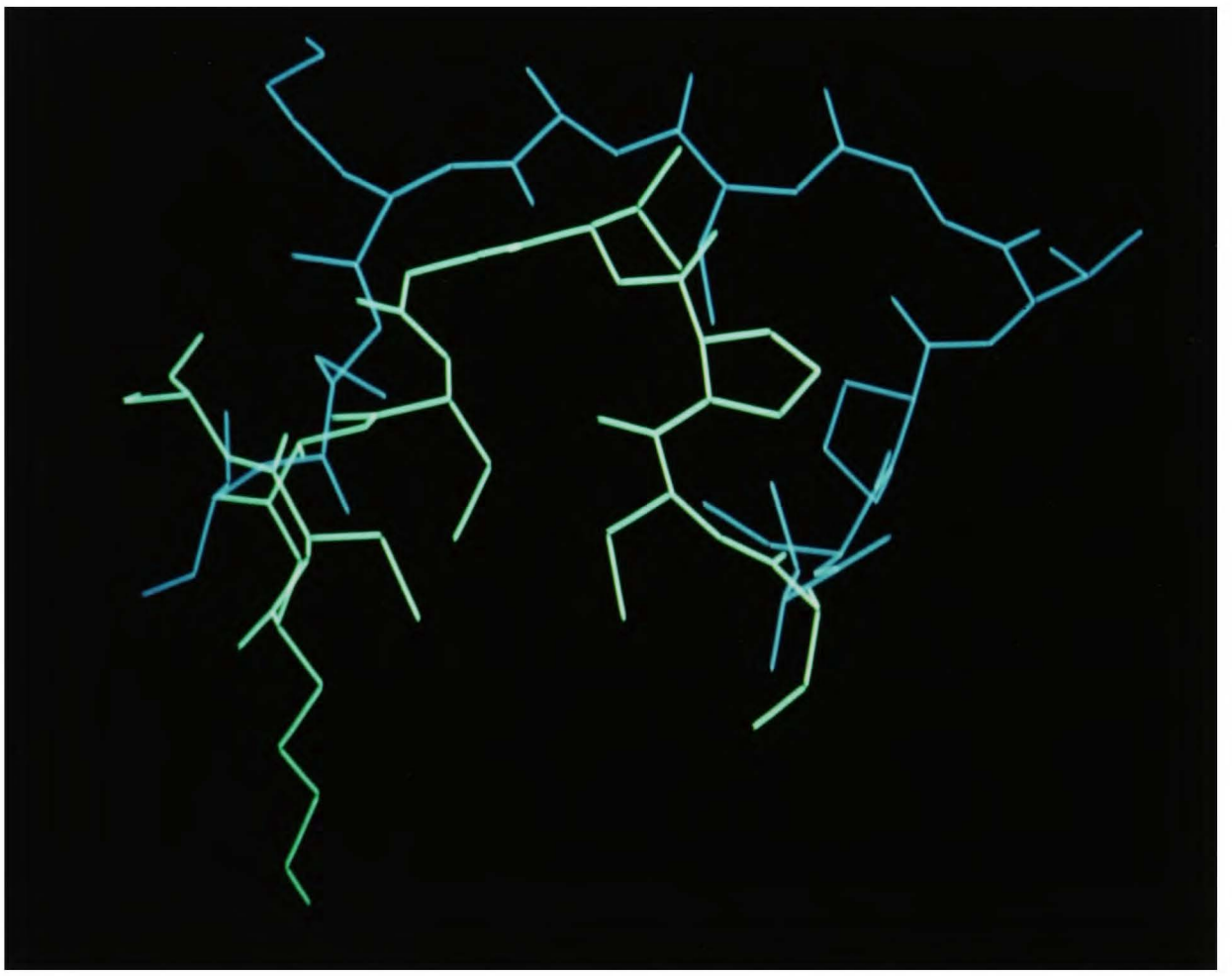


Figure 10: Comparison of the NMR (green) and crystal structure (blue), of residues 36–45 from metallothionein.

and Cys-44 is a singly bound ligand, while in the NMR structure Cys-41 is a singly bound ligand and Cys-44 is a bridging ligand (Schultze *et al.*, 1988).

3.7.2 Implications of the Metallothionein NMR Structure

In the NMR structure, residues 36–45 show no conformational similarity to HyHEL-5 L1 and the 4 liganding cysteines are now split into two pairs, with Cys-36 and 37, and Cys-41 and 44 coordinating separate metal ions. The difference in conformation of this loop suggests that the mutations necessary to introduce the metal binding site into L1 may not maintain the native CDR conformation. In addition, the division of the four cysteine ligands into pairs casts doubt upon the availability of Cys-41 as a potential third ligand (with Cys-36 and 37) in the L1 metal site. Two liganding residues would not be sufficient to give a metal ion chelate stability within the site (Williams, 1987) (although weak binding may be possible).

Lin *et al.*, (1990) have also shown that differences in the liganding configuration around a transition-metal site, exist between the solution and crystal structures of the protein Concanavalin A. The significance of these findings suggests caution is necessary in assuming that protein structure as determined from crystals, will always be applicable to physiological conditions. Another example is seen in the α-amylase inhibitor, tendamistat (Vértesy *et al.*, 1984). In the structure determined in solution (Kline *et al.*, 1988) by NMR techniques (Wüthrich, 1986), a surface arginine is invisible indicating high mobility. However, in the structure determined by X-ray crystallographic techniques (Pflugrath *et al.*, 1986), this arginine is immobile, as indicated by low temperature factors (Petsko and Ringer, 1984). The low mobility results from a chloride ion bridging to another arginine in the crystal lattice. A final

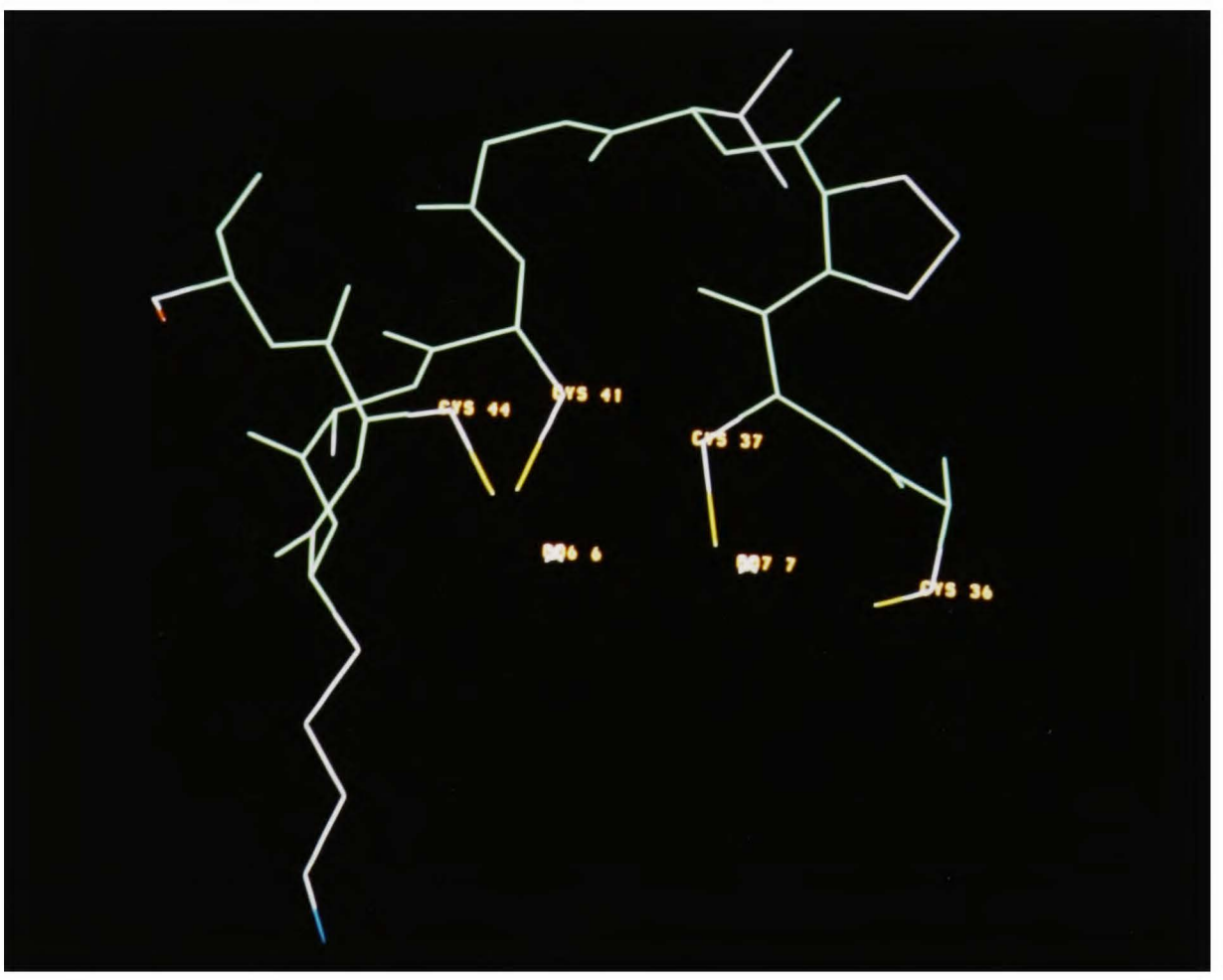


Figure 11: Residues 36–45 from the NMR structure of Cd_7-metallothionein-2. Cysteines 36 and 37 coordinate the Cd-7 ion, while cysteines 41 and 44 coordinate the Cd-6 ion. *c.f.* Figure 9.

example is provided by Baldwin *et al.*, (1991). They solved the crystal structure of interleukin 8 using the solution structure derived from NMR as a model in molecular replacement, and noted some potential differences between the structure in solution and in the crystalline states.

Therefore, on the basis of the differences between the two metallothionein structures and their implications for the L1 metal site, it was decided to abandon the metallothionein-based model. Nevertheless, the work on this model contributed to the development of a new metal site in L1. The site was comprehensively modelled using the combined algorithm method (Martin *et al.*, 1989) followed by more intuitive energy minimisation in the presence of zinc. The modelling protocols used, which supersede the methods described here, are described in the next chapter.

Chapter 4

Ab Initio Design of Metal Binding Sites

The design of a metalloantibody based on a pre-existing structure, carried out in Chapter 3, showed this method to be applicable where there are similarities between the antibody and metalloprotein structures. However, it was also apparent that the success of this approach was highly dependent on the accuracy of the parent and target crystal structures. If either of these are unproven, then a successful design is much less certain. It was for this reason that the metallothionein-based model (*Hybrid*) L1 was rejected.

Although other parent (and target) structures could have been found, it was felt that the work carried out on HyHEL-5 (Sheriff *et al.*, 1987) was still of value. Moreover, the *ab initio* design of a metalloantibody would, if successful, show that it is not necessary to copy a pre-existing structure in order to design a functional metal binding site, thus freeing the protein engineer from the constraints imposed by this approach.

4.1 Designing a Metal Binding Site in L1

The investigation of HyHEL-5 L1 as a target metalloantibody, described in Chapter 3, showed that it was an ideal candidate. The major interactions of this loop with the other light chain complimentarity-determining regions (CDRs) and the antigen lysozyme, involved only residues 30–33. This leaves residues 24–29 free to accommodate a metal binding site without disrupting these contacts and would place a metal ion in the antibody combining site (ACS). In this position it is more likely to be affected by antigen binding which is of functional importance for a biosensor or catalytic centre.

The liganding atoms must be within the appropriate spatial proximity, < 4.0Å, to bind a metal ion within L1. From the *Hybrid* L1, two of the liganding Cysteines corresponded to L1 residues Ala-25 and Val-29 . These two residues make contact with each other in L1 and were replaced with Cysteines, bringing the sulphur atoms of these sidechains well within the 4.0Å limit. This gave two of the minimum three ligands required to securely bind a metal ion. The most likely position for a third ligand was between residues 25 and 29 *i.e.* Serines 26, 27 and 28. However, in the crystal structure, these sidechains point into the solvent, away from the putative metal site. A third Cysteine could not be placed in any of these positions as its sidechain would not be long enough to reach the metal site without a rotation of the backbone, potentially disrupting the conformation of L1.

Nevertheless, a longer liganding sidechain such as Glutamate, might have sufficient flexibility to come into close proximity with the two Cysteines, creating a

tri-coordinate metal site. From an examination of Serines 26, 27 and 28 using interactive molecular graphics, it was decided to replace Ser-27 with Glutamate. This residue lies midway between the two Cysteines and places the Glu sidechain in an equidistant position between the two Cys sidechains.

The residue replacements were carried out using the REPLACE and REFI (Hermans and McQueen, 1974) procedures available within the molecular graphics program FRODO (Jones, 1985). The inclusion of a metal ion at an approximate position within the putative site, showed that it was possible, by manipulation of the Glu-27 sidechain torsion angles, to bring the carboxyl group within liganding distance of the metal ion (< 3.0Å).

Studying the putative metal site also revealed that Ile-2, at the N-terminus of the light chain, came within liganding distance (typically < 2.3Å for a *d*-block transition metal) of the metal ion *via* the Cδ atom. This suggested a possible fourth liganding position. A liganding sidechain of at least three atoms in length (Cβ, Cγ, Cδ) would be required at the Ile-2 position and it should be of the borderline/soft type to optimise for zinc (or other *d*-block metal) binding (see Chapter 2). Of the potential liganding residues available, only Histidine was of the correct length and type. When this replacement was made using FRODO, adjustment of the sidechain torsion angles allowed the Nδ1 of the His sidechain to come well within typical liganding distance.

4.1.1 Advantages of the New Metal Binding Site in L1

A four-coordinate metal binding site had now been created. Three ligands (Cys-25, Glu-27 and Cys-29) came from L1, while the fourth ligand (His-2) came from the

N-terminus of the light chain. This *ab initio* site (*Model1*) had several advantages over the *Hybrid* L1:

- Only 4 mutations (three in L1 and one at the N-terminus) would be required to create a four-coordinate site as opposed to 5 mutations (all in L1) to create a three-coordinate site in *Hybrid* L1.

- Of the five metallothionein substitutions in L1, two were conformationally significant Proline and Glycine residues. The effect of these on backbone conformation, and the differences between the crystal and NMR structures (Chapter 3), reduced the probability of maintaining the native CDR conformation. The *ab initio* design required fewer mutations in L1, but although each was introduced in order to maintain native CDR structure, three of them (His-2, Cys-25 and Cys-29) are defined by Chothia (1989) as structure-determining residues (SDRs) which may affect loop conformation (see Chapter 7).

- The spatial arrangement of the three ligands in *Hybrid* L1 required that Ser-24 be replaced by Cysteine. This had three disadvantages:

  1. The residue preceeding Ser-24, Cys-23, forms a disulphide bond with Cys-87 in the light chain. The orientation of Ser-24 places it in a *cis* configuration with Cys-23, outside the loop. A second Cysteine residue at position 24 might lead to competition with Cys-23 for Cys-87, causing the light chain to mis-fold through an incorrect disulphide pairing.

  2. Ser-24, although partially solvent accessible, is close enough to the framework to be influenced by it.

3. In order to function as a ligand, a Cysteine at this point would have to be in a *trans* configuration to Cys-23 *i.e.* pointing toward the inside of the loop. This would require rotation of the backbone at this point. The constructions which included Cys-24, carried out in Chapter 3, were unable to place this residue on the inside of the loop, indicating conformational restriction by the framework.

In the new design Ser-24 is retained, and acts as a spacer between Cys-23 and Cys-25. In addition, Cys-25 is in a *trans* configuration in relation to Cys-23, so the chance of an incorrect disulphide pairing is diminished.

4.2 Conformational Search of Model1 L1

The conformational effects of the metal site were investigated by modelling L1 in the presence of the liganding residues. At its simplest, this could be achieved using the maximum overlap procedure (MOP) (Snow and Amzel, 1986). Amino acid substitutions are made such that the new residue is constructed with as many atoms as possible in identical positions to the atoms in the parent structure; that is, wherever possible, the parent dihedrals are inherited by the mutated residue. The entire structure is then subjected to energy minimisation to obtain a prediction for the mutant protein. This procedure is easy to implement using interactive molecular graphics programs such as FRODO. This approach is successful particularly when the substitution is conservative and the residue makes few interactions with neighbouring residues. However, it is usually inadequate if a drastic change in the nature of the residue is made. In addition, it is possible for the structure to become trapped

in a local energy minimum resulting in an incorrect model (Martin, 1990).

Three of the mutations involved SDRs. Given the importance of these residues in CDR conformation (Chapter 7), the modelling of the metal site mutations in L1 required a more rigorous approach. Consequently, this CDR was constructed using the combined algorithm method (CAMAL) (Martin *et al.*, 1989). The method consists of four stages as outlined below.

The Database Search: The entire Brookhaven Protein Databank (Bernstein *et al.*, 1977) is searched for loops of similar conformation to those in the known antibody crystal structures, using distance constraints derived from an analysis of these structures. Cα distances within each of the 6 CDRs of the known antibody structures are calculated from the N- and C-terminal residues. A database of Cα distance constraints for all proteins in the Protein Databank is created and two searches are performed for each loop, applying distance constraints from the N- and C- terminal Cα's of the loop, respectively. The results from these two searches are then merged to produce a list of loops fitting both sets of constraints, and redundancies resulting from updated entries in the databank and multiple chains, removed.

Processing of the Loops: Each loop identified by the database search is extracted from the protein databank and positioned onto the framework take-off points of the loop being modelled. The sidechains are then replaced using standard conformations. Since these are repositioned by conformational searching, only the position of the Cβ which defines the overall orientation of the sidechain is important. Where a Cβ is present in the parent sidechain,

this position is used; when the parent is a Glycine, the position of the Cβ comes from the template conformation. 'Explicit hydrogens' (McCammon *et al.*, 1979) involved in hydrogen bonding are added using standard geometries. Having removed any loops which have bad angles at the framework junctions, or have bad clashes with the rest of the structure, the database loops are processed into the form of a CONGEN conformations file.

Conformational Searching: This is performed by the program CONGEN (Brucoleri and Karplus, 1987) and is described in Chapter 3. The protocol is to construct a 5 residue segment from the middle of the loop in the case of a whole loop replacement, or residues neighbouring a mutation, insertion or deletion. A variation in this procedure occurs for short loops (7 amino acid or fewer). For 6 or 7 amino acid loops, the database search appears to saturate the conformational space of the backbone satisfactorily so only sidechain positions are searched with CONGEN. For 5 amino acid loops (and shorter) the limited number of distance constraints means that an enormous number of conformations ($> 10\,000$) is selected from the database and processing this number of loops becomes impractical. Thus CONGEN is used alone in these cases.

Screening the Conformations: The conformations are screened using a modified version of the GROMOS (Åqvist *et al.*, 1985) potential implemented in the program GROMSCAN (Martin, 1990) as described in Chapter 3. For loops of 5 residues or shorter, where no database search has been performed, the final conformation is selected on the basis of solvent accessibility (Lee and

Richards, 1971) of hydrophobic atoms. For loops of 6 or 7 residues and longer, the final conformation can be selected from the five lowest energy structures using a structure determining-algorithm implemented in the program FILTER (Martin, 1990).

FILTER can only be applied to loops of 6 residues or longer as it requires information from the database search. For each residue of the loop in turn, the database loops are searched for the required residue type and the ϕ and ψ torsion angles are recorded. If the required residue type is not identified in any of the database loops, then the most similar residue types are identified from the Dayhoff mutation matrix and the confidence for the prediction at this residue is reduced by 5%. This procedure is repeated until ϕ/ψ angles for each residue are identified among the database loops.

The ϕ and ψ angles for each of the conformations being screened are then calculated and scored against the torsion angles observed in the database. The terminal residues start with a confidence of 50%, since one of the two torsion angles is undefined. Two scoring schemes are implemented. In the first scheme, torsion angles are scored as $1/\theta$, where θ is the difference between the model and the database torsion angles (in radians) and θ is less than a specified cutoff. This value is then multiplied by the confidence. The second scheme simply scores a 1 for a residue if it has ϕ/ψ angles within the cutoff of any ϕ/ψ angles observed in the database conformations and thus represents the distribution of scores across the loops.

Both scoring schemes are examined. If the first scoring scheme is unable to

distinguish between the model conformations, the distribution of the scores is also examined—conformations that have their scores distributed across the amino acids are chosen in preference to ones which score highly on only a few residues. In addition, if the conformation scoring highest in scheme 1 does not have a high distribution, this selection is rejected and a new angular cutoff is used.

The angular cutoff is initially chosen as 15°. If this is unable to select a conformation, the cutoff is increased to 30° and if this is still unable to distinguish between the models, the cutoff is further increased to 45°. This protocol is based on the ability of the different angular cutoffs to distinguish correctly between the low energy conformations. If FILTER still fails to select a unique conformation, this suggests that all the low energy conformations are very similar and the lowest energy conformation can be selected, or solvent accessibility can be used.

A total of four constructions were carried out on the *Model1* L1. For three of these constructions, the N-terminal five residues were removed so that each conformation was calculated in the absence of the N-terminus. Given that one of the metal ligands is located within this sequence, there should be sufficient flexibility in the light chain at this point for adjustment relative to any new conformation the *Model1* L1 might adopt. The first two constructions are shown in the Table below.

Construction	Total Conformations	CPU Time/h	
		CONGEN	GROMSCAN
S[C(SES)C]NYMY	1138	51	88
SCS[E(SCN)Y]MY	1039	47	26

The energies calculated by GROMSCAN revealed that the bottom 5 conformations from the [E(SCN)Y] construction were lower than those from the [C(SES)C] construction. The [C(SES)C] construction was repeated, but using the 5 lowest energy structures from the [E(SCN)Y] construction as starting conformations (see the Table below). Thus, construction of the metal site (CSESC) is influenced by the conformation of the junction between it and the main CDR/antigen interacting residues in L1.

Construction	Starting Conformation	Total Conformations	CPU Time/h	
			CONGEN	GROMSCAN
S[C(SES)C]NYMY	1	200	6.27	4.58
	2	142	4.49	3.25
	6	128	3.98	2.78
	15	114	3.58	12.50
	34	234	7.28	5.36

Starting conformation 34 produced the 5 lowest energy structures. These were screened using the FILTER algorithm with a 15° cutoff. The selected conformation is shown together with the native L1 in Figure 12 and the r.m.s. deviation between them is given in Table 17. There has been a significant change in the shape of L1 owing to the presence of the mutations. The generally wide and flat structure of the native CDR has been replaced by the taller, more compact structure of the *Model1* L1. This is illustrated by measuring Cα distances between residues across the long and short axes of the loop. In the crystal structure CDR, the distance between Ser-26 and Met-32 is 13.2Å, while that between Ser-24 and Ser-28 is 9.4Å. In *Model1* L1, the equivalent distances are: Ser-26 to Met-32, 11.8Å and Ser-24 to Ser-28 12.1Å.

The conformational differences between the two structures is a result of changes

in the ϕ/ψ torsion angles respectively, at the following residues: Ser-26 (91° and 204°), Glu-27 (214° and 123°) and Cys-29 (300° and 90°). The sidechain of Met-32 had rotated by 97° around χ_2 and by 154° around χ_3. This sidechain interacts with both Ala-25 and Val-29 within the native L1, and all three are SDRs (Chothia *et al.*, 1989). The rotation of this sidechain results from interactions with the Cys sidechains now at positions 25 and 29. While the change in overall conformation of the *Model1* L1 is most obviously a direct consequence of the removal of the N-terminus, the rôle played by the mutation of two SDRs must also be considered (see Chapter 7).

A fourth construction on the [C(SES)C] sequence was performed, but with the N-terminus left in place. In addition, Ile-2 was replaced by His (*Model1H2*). This was done to ascertain not only the effect of the N-terminus, but also the presence of the His ligand on the *Model1* L1.

Construction	Total conformations	CPU Time/h	
		CONGEN	GROMSCAN
S[C(SES)C]NYMY plus N-terminus and Ile-2→His	832	29	15

The selected conformation is shown together with the conformation obtained in the absence of the N-terminus and the crystal structure CDR, in Figure 12. The structures obtained in the presence and absence of the N-terminus differ significantly, giving an indication of the effect of the N-terminus on the conformation of L1. This serves to illustrate the importance of contacts with other CDRs or framework made by the CDR being modelled (Martin, 1990). The major conformational differences between the two structures results from different rotations around the ϕ/ψ torsion

angles respectively, of the three ligands Cys-25 (82° and 327°), Glu-27 (242° and 95°) and Cys-29 (300° and 100°). Again, the sidechain of Met-32 shows rotations of 138° around χ_2 and 195° around χ_3. These rotations are larger than those found in the absence of the N-terminus and may reflect not just interactions with the N-terminus, but also the presence of His-2 which itself is an SDR.

Table 17 shows the r.m.s. deviations between HyHEL-5 L1, *Hybrid* L1[1] and the model L1 in the presence (*Model1I2*[2] and *Model1H2*) and absence (*Model1*) of the N-terminus. The structures generated in the presence of the native (Ile-2) N-terminus (*Hybrid* and *Model1I2*) and with the N-terminal Ile-2→His substitution (*Model1H2*) tend to have relatively similar conformations to HyHEL-5 L1 and to each other. These conformations differ significantly from that generated in the absence of the N-terminus (*Model1*). This is to be expected given that the conformational constraints imposed by the N-terminus are similar in all cases where it is present. The conformations differ owing to changes in packing interactions resulting from the mutations, particularly at SDR positions.

However, the *Model1I2* L1 was an interesting exception. While this structure was modelled in the presence of the native N-terminus, its conformation differs markedly from those of HyHEL-5 L1, *Hybrid* L1 and *Model1H2* L1. This difference can be explained by the position of the Ile-2 sidechain in relation to L1. The loss of SDRs Ala-25 and Val-29 removes the major hydrophobic interactions of this residue. As a result, there is a large adjustment in the position of the loop to allow hydrophobic interaction with the sidechain carbons of Asn-30 and maximise

[1]The *Hybrid* L1 structure is included here for comparison, since it was calculated in the presence of the native (Ile-2) N-terminus.

[2]This structure was modelled to compare the effects of the N-terminal Ile-2→His substitution, itself an SDR mutation, on *Model1*.

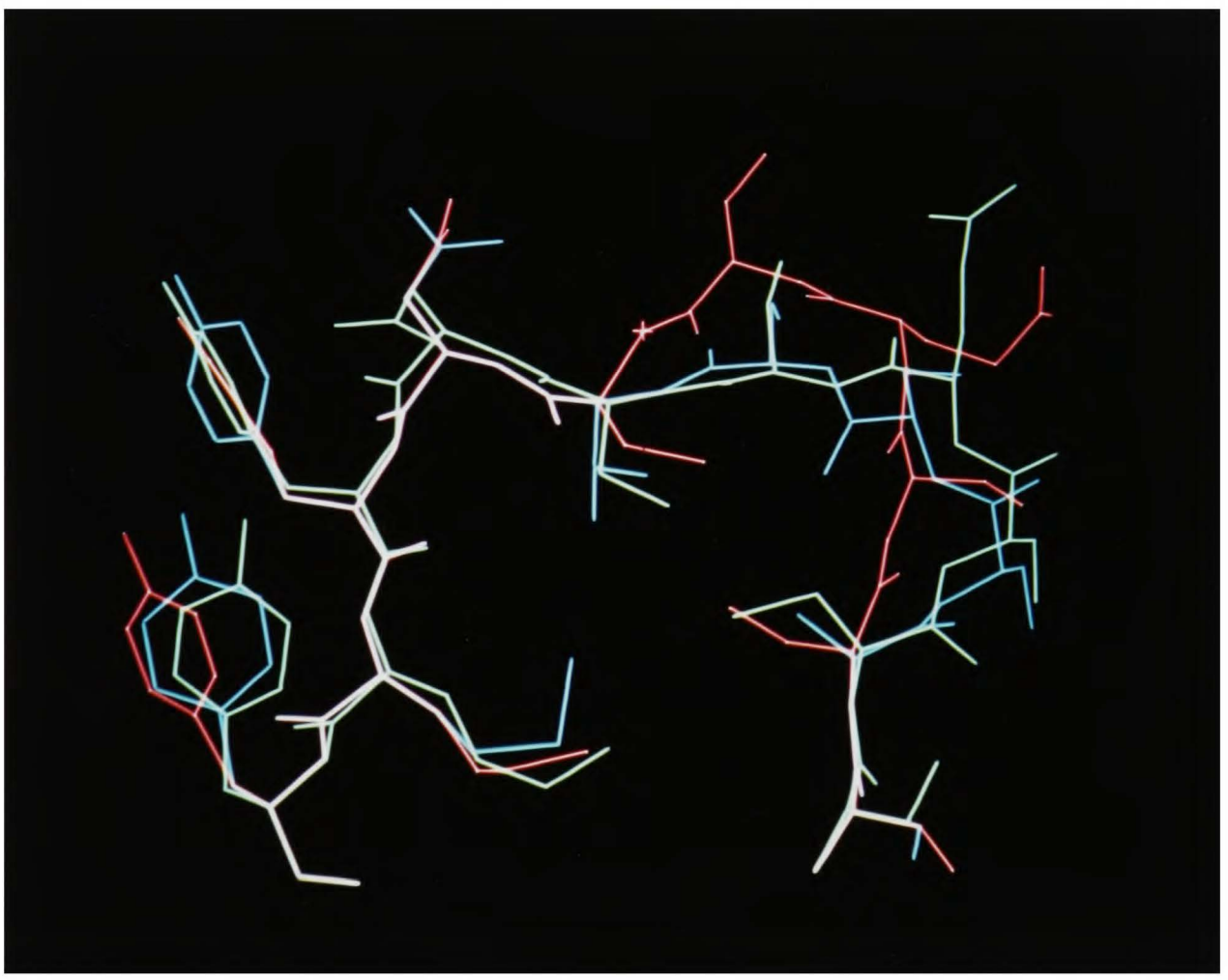


Figure 12: Comparison of the selected conformations for *Model1* L1 (red), *Model1H2* L1 (green) and HyHEL-5 L1 (blue). The r.m.s. deviations between these three structures are given in Table 17.

Structure 1	Structure 2	R.m.s. Deviation/Å	
		Cα	Backbone
HyHEL-5 L1	*Hybrid* L1	0.29	0.49
	Model1 L1	1.29	1.48
	Model1H2 L1	0.73	0.95
	Model1I2 L1	1.88	2.20
Hybrid L1	*Model1* L1	1.23	1.55
	Model1H2 L1	0.59	0.88
	Model1I2 L1	1.89	2.23
Model1 L1	*Model1H2* L1	1.48	1.76
	Model1I2 L1	2.01	2.39
Model1H2 L1	*Model1I2* L1	1.95	2.31

Table 17: R.m.s. deviations between residues 24–33 of the different L1 metal site models and native HyHEL-5 L1. **Backbone** = N, Cα, C. Structure 2 was fitted onto Structure 1. *HyHEL-5 L1*: The HyHEL-5 crystal structure CDR; *Hybrid* L1: The metallothionein-based metal site L1 (see Chapter 3); *Model1* L1: The new metal site L1 calculated in the absence of the N-terminus; *Model1H2* L1: The new metal site L1 calculated with the N-terminus containing His-2; *Model1I2* L1: The new metal site L1 calculated with the native (Ile-2) N-terminus.

its contacts with the backbone between Ser-28 and Asn-30. The principal sidechain interactions of Ile-2 in *Hybrid* L1 are with Cys-25 Cβ, analogous to that of Ala-25 in the native L1. It is the additional contact with the sulphur atom at this position which contributes to the (smaller) conformational change.

For the *Model1H2* L1, the position of the His-2 sidechain is such that the imidiazole ring is face on to the backbone of L1, but packs onto the backbone of the N-terminus rather than L1. This provides greater room for the accommodation of the two Cys sidechains within L1 resulting in a smaller conformational change. However, the conformation of the His-2 sidechain was imposed by the REPLACE and REFI procedures in FRODO. Thus, the effects of His-2 in a position analogous to the native Ile-2 (or indeed, a 'liganding' configuration) could not be ascertained. Nevertheless, it is interesting to note that rotation of the imidiazole ring into a 'typical' liganding configuration (with the ring plane at right-angles to L1) did not cause undue disturbance to the conformation of L1 after energy minimisation (see next Section).

The similarity of the *Model1H2* L1 conformation to that of the native CDR suggests that contacts between the introduced liganding residues have conformational effects equivalent to those of the native SDRs. This also illustrates how the mutation of SDRs can be used in antibody engineering, since the maintenance of a native-like L1 conformation indicates the successful introduction of a novel metal binding site into HyHEL-5.

The liganding sidechains of the final conformations of *Model1* and *Model1H2* were manipulated using interactive molecular graphics, to see if they could be brought within liganding distance of a zinc ion (< 2.30Å) placed in the putative

metal site. An initial attempt was made to attain an approximately tetrahedral geometry around the metal ion. In both cases, slight adjustments of the two Cys sidechains were sufficient, while more extensive movement of the Glu-27 sidechain was required in the *Model1* L1, to orientate the carboxyl group over the metal ion.

For the His-2 sidechain, the lack of the N-terminus during conformational search of the *Model1* L1 meant that extensive adjustments were also required. However, for the *Model1H2* L1, the presence of the N-terminus necessitated only slight adjustments of His-2. In fact, it was possible to see how the His and Cys sidechains from this conformation could readily arrange themselves around a metal ion, giving a three-coordinate site sufficient to firmly bind a metal ion (Williams, 1987).

Hence, the Glu-27 sidechain for this structure was left in the position generated by CONGEN. To bring the carboxyl group within liganding distance would have required bending the sidechain over the metal ion and it is likely to prefer an extended conformation, with the carboxyl group fully solvent exposed[3].

4.3 Energy Minimisation of the L1 Models in the Presence of Zinc

In order to obtain some quantitative evidence for the effects of a bound metal ion in L1, both the *Model1* and *Model1H2* structures were subjected to energy minimisation (EM) in the presence of zinc. For each structure, the sidechains were

[3]It is probable that the metal ion in the actual metalloantibody is bound without the assistance of Glu-27. That this sidechain also does not appear to affect antigen affinity/specificity from antigen binding studies, lends support to the proposition that residues 25–29 in HyHEL-5 L1 do not contact the antigen. The characterisation of the functional metalloantibody is described in Chapter 5.

Structure	Ligand	Distance/Å		
		Actual	Mean	Difference
Model1	His-2 Nδ1	2.04	2.05	0.01
	Cys-25 Sγ	2.22	2.21	0.01
	Glu-27 Oϵ1	1.99	2.19	0.20
	Cys-29 Sγ	2.13	2.21	0.08
Model1H2	His-2 Nδ1	2.35	2.05	0.30
	Cys-25 Sγ	2.21	2.21	0.00
	Cys-29 Sγ	2.21	2.21	0.00

Table 18: Metal-ligand distances for the *Model1* L1 metal sites. The mean distances were obtained from the metalloprotein survey and are set out in Chapter 2, Table 2.

adjusted to within liganding distance (< 2.30Å) around the metal ion.

To optimise the conformations of the liganding residues, the metal and ligand positions were adjusted using several rounds of multidimensional (SIMPLEX) minimisation implemented within the program CONRAD (Chapter 6). The metal-ligand distances applied were derived from the metalloprotein survey (Chapter 2). A metal ion was placed within the putative metal site and its position refined with-respect-to the positions of the ligands, by SIMPLEX minimisation. From the resulting position of the metal ion, badly placed ligands, revealed by their close proximity to the metal, were identified and repositioned using interactive molecular graphics. The metal position was then subjected to a further round of minimisation with the new liganding positions, followed by further adjustment of the ligands if required. Final metal and ligand positions were taken when the position of the metal ion remained unchanged after minimisation. The metal-ligand distances obtained for both structures are shown in Table 18.

EM was performed using the program GROMOS (Åqvist *et al.*, 1985). Each structure was subjected to three rounds of EM initially employing the SHAKE

algorithm (Ryckaert *et al.*, 1977; van Gunsteren and Berendsen, 1977), then without SHAKE and finally with SHAKE again. For each structure, EM was carried out on both the *apo* and *metallo* forms. In addition, the *metallo* forms were minimised using three different charges on the zinc ion. All structures were minimised in the absence of lysozyme.

An examination of the charges used by GROMOS after EM of the *Model1* structure, revealed that no charge appeared to be applied to the zinc ion. This questioned the validity of the minimised structures with no charge applied to zinc, given the strong electrostatic interactions between a metal and its ligands. Hence, the minimisation was repeated including one of three charges on the zinc.

A value of 2.0, equivalent to a formal charge of +2 was first applied, since this is the common oxidation state for the zinc ion. However, a metal ion bound in a protein will be affected by its ligands as a result of electrostatic field effects (Kneen *et al.*, 1984). Quantum-mechanical calculations (Fischer-Hjalmars and Henriksson-Enflo, 1982) on a series of model metal complexes $[M(C_2H_2X_2)_2]^n$ (M = Be, Mg, Ni, Cu, Zn; X = NH, O, S; n = 2, 1, 0, 1, -2) indicated that the computed charges on the metal ions were generally much less than the formal charges.

To try to account for this in a metalloprotein complex, the value for the charge on the zinc ion within the metal site of the *Model1* structure, was calculated using the program GRID (Goodford, 1985). GRID determined the zinc charge in the site to be +0.22. It also indicated that the remainder of the formal +2 charge (+1.78) was distributed (delocalized) over the imidiazole ring of His-2 (Goodford—personal communication). However, since the exact nature of this distribution (*i.e.* its effects on the charges of each atom in the ring) was unknown, the extra charge on the His

ligand was not accounted for during EM.

Finally, an attempt was made to account not only for reduction in the formal charge of the metal ion, but also for the effects of the charge distribution on the ligands. The GROMOS potentials were modified to include the electrostatic partial charges calculated, *ab initio*, by Shen *et al.*, (1990) (*Shen*) (Table 19) for the metal-ligand complex of Cu-Zn Superoxide Dismutase (SOD) (Tainer *et al.*, 1982). The metal ions in this enzyme are bound by six His and one Asp ligands in total. Thus, the charges calculated for this metalloprotein complex do not include any values for Cys ligands. Therefore, only the charges of His-2 and Glu-27 could be altered. EM of the model structures used a mixture of GROMOS and *Shen* charges, so the results were not as accurate as if adjustment of the Cys charges had been possible. This is also true of the charge on the metal ion which would probably be modified in the presence of the Cys ligands.

4.3.1 Results of Energy Minimisation with Zinc

The results of EM on the two model structures using each of the zinc charges are shown in Tables 20, 21 and 22. Table 20 shows the effect of the difference between the two model structures on the stability of the metal site. Generally, the *Model1H2* L1 site, calculated in the presence of the N-terminus, had a lower energy, indicating a greater stability, than that of *Model1* L1, calculated in its absence. In addition, the *Model1H2* L1 site was much easier to build using interactive molecular graphics, suggesting a more plausible model.

Of particular significance (especially for the *Model1H2* structure) is that in both cases, except for *Model1* with Zn^{2+}, the *apo* form has a higher energy than any

Ligand	Atom	GROMOS	*Shen*
His-2	Nδ1	-0.58	-0.30
	Nϵ2	0	-0.30
	Cγ	0	0.10
	Cϵ	0.26	0.32
	Hϵ	0.19	0.30
Glu-27	Cδ	0.27	0.27
	Oϵ1	-0.635	-0.42
	Oϵ2	-0.638	-0.30

Table 19: Comparison of the charges used in energy minimisation of the metal site models with zinc. (Data reproduced and slightly modified from Shen*et al.*, (1990) Table III).

Structure	Zinc Charge	Energy Kcal/mol	R.m.s. Deviation/Å
Model1	*Apo*	-2223.9	0.60
	GROMOS	-2595.9	0.42
	Formal	-2105.7	0.51
	GRID	-2678.0	0.49
	Shen	-2446.5	0.36
Model1H2	*Apo*	-2187.0	0.44
	GROMOS	-2634.5	0.44
	Formal	-2714.2	0.48
	GRID	-2632.2	0.49
	Shen	-2627.4	0.54

Table 20: Energies of the two model structures after EM with zinc. **Zinc Charge:** The charge applied to the zinc ion during EM; *Apo*: EM of the open form from CONGEN without zinc; *GROMOS*: The GROMOS zinc charge; *Formal*: The formal +2 zinc charge; *GRID*: The zinc charge calculated by the program GRID; *Shen*: The metal site partial charges calculated by Shen *et al.* **R.m.s. Deviation:** The all atom deviation of the structure after EM, with the structure before EM.

Structure	Zinc Comparison		R.m.s.	Difference in
	Charge 1	Charge 2	Deviation/Å	Energy Kcal/mol
Model1	GROMOS	GRID	0.17	82.1
		Shen	0.11	149.4
		Formal	0.32	490.2
	GRID	*Shen*	0.11	231.5
		Formal	0.42	572.3
	Shen	Formal	0.38	340.8
Model1H2	GROMOS	GRID	0.18	2.3
		Shen	0.22	7.1
		Formal	0.10	79.7
	GRID	*Shen*	0.06	4.8
		Formal	0.17	82.0
	Shen	Formal	0.20	86.8

Table 21: All atom r.m.s. deviations for the *metallo* forms of the two model structures after EM. **Zinc Comparison:** The charge applied to the zinc ion during minimisation of the *metallo* forms—compared for the r.m.s. deviation; *GROMOS, GRID, Shen* and *Formal*: As for Table 20. **Difference in energy**: Energy difference between the two structures. Energies taken from Table 20.

Structure	Zinc-Ligand Distance/Å						
	Ligand	Pre-EM	GROMOS	GRID	*Shen*	Formal	Mean
Model1	His-2 Nδ1	2.0	2.7	2.6	2.8	2.1	2.6
	Cys-25 Sγ	2.2	3.1	2.9	2.8	2.7	2.9
	Glu-27 Oϵ1	2.0	3.2	2.9	2.7	2.8	2.9
	Glu-27 Oϵ2	—	2.9	2.7	2.4	2.1	2.5
	Cys-29 Sγ	2.1	2.8	2.8	2.6	2.3	2.6
Model1H2	His-2 Nδ1	2.4	3.0	2.6	2.8	2.1	2.7
	Cys-25 Sγ	2.2	3.6	3.2	3.1	2.7	3.2
	Cys-29 Sγ	2.2	3.1	3.1	2.8	2.5	2.9
	Ser-28 'O'	(2.6)	3.0	2.6	2.4	2.1	2.5

Table 22: Zinc-ligand distances for each zinc charge after EM. *Pre-EM*: Zinc-ligand distances of the structure before EM; *GROMOS, GRID, Shen* and *Formal*: As for Table 20; *Mean*: Average zinc-ligand distance from the combined distances of each zinc charge; shows the overall change in liganding distance when compared with the *Pre-EM* values.

of the *metallo* forms. This suggests co-operative metal binding (Williams, 1987). The unfavourable electrostatic repulsion between the ligands in the *apo* forms, is replaced by a strong metal-ligand electrostatic attraction plus additional van der Waals interactions with the bound metal. The table also gives an indication of the effect of structure and zinc charge on the extent of minimisation. The differences in energy and r.m.s. deviation between the *metallo* forms of the model structures is generally smaller for *Model1H2* than for *Model1*. This can be seen more clearly in Table 21 and reflects the greater inherent stability of the *Model1H2* structure compared to the *Model1* structure.

For both structures, except for the GROMOS/*Shen* comparison of *Model1H2*, the comparison r.m.s. deviations and energies using the formal zinc charge, are greater than those obtained from the other zinc charges. The calculated zinc charges (GROMOS, GRID and *Shen*) produced very similar structures suggesting that, despite adjustments in the zinc (and ligand) charges, the overall effect was the same (although this may also be partly owing to the use of identical (GROMOS) Cys charges in all cases). The larger differences in r.m.s. deviation or energy between the *metallo* forms of the same structure, reflects the differences in the zinc charges applied during EM. The success of each charge in maintaining the structure of the metal site, given the competing electrostatic repulsion between the ligands and any unfavourable van der Waals interactions, is indicated in Table 22.

This table shows the zinc-ligand distances for the bound form of each structure after EM. With one exception, all the zinc-ligand distances are longer after EM. Considering each model structure in turn, the zinc-ligand distance for the GROMOS, GRID and *Shen metallo* forms of *Model1* have all become much longer. In

addition, the carboxyl group of Glu-27 has rotated around the Cγ-Cδ bond, placing the Oϵ2 atom closer to the zinc ion (although either oxygen is capable of acting as a ligand). Also, the Cys-25 liganding distance is longer than the others by an average of 0.3Å. In contrast, the liganding distances for the *metallo* forms using the formal zinc charge have, except for Cys-25, maintained more typical metal-ligand distances ($<$ 2.30Å).

The imidiazole ring of His-2 has rotated in all forms after EM except that employing the formal zinc charge, placing it face-on to the metal site. In the presence of the formal zinc charge the ring has maintained an edge-on orientation. This ligand shows the greatest change in position after EM of all the *metallo* forms, as indicated by the mean r.m.s. deviation for the His sidechain of 0.3Å between the starting structure and the GROMOS, GRID and *Shen* zinc charges, 1.4Å for the formal zinc charge and 1.4Å between the formal zinc charge and the GROMOS, GRID and *Shen* zinc charges.

The zinc-ligand distances for the GROMOS, GRID and *Shen* forms of *Model1H2* have, like those of *Model1*, all got much longer and in particular, the Cys-25 distance again by an average of 0.3Å. Once more, the zinc-ligand distances of the *metallo* form employing the formal charge showed the smallest change after EM. Generally, the liganding distances for this structure are longer than those for the *Model1* structure.

The *Model1H2* structure did not employ Glu-27 as a ligand and was thus only a three-coordinate site. However, as shown in Table 22, EM brings the carbonyl group of Ser-28 into contact with the zinc ion. This group was not included in the starting structure and no examples of zinc-carbonyl ligation were observed in the

metalloprotein survey (see Chapter 2). Although the contribution of this group to metal binding in the actual metalloantibody is unknown, it should not be ruled out (Hellinga *et al.*, 1991).

4.3.2 Discussion

There is a clear distinction between the stability of the *Model1* and *Model1H2* structures resulting from the presence of the N-terminus and the Ile-2→His mutation in particular. When the selected conformation for the *Model1* L1 (calculated in the absence of the N-terminus) was examined with the N-terminus, containing this mutation, present, the position of the His ligand with-respect-to the other ligands in L1, was unfavourable. The sidechains of His-2 and Glu-27 in particular, required extensive manipulation using interactive molecular graphics, to create the metal site. In contrast, the selected conformation calculated in the presence of the N-terminus containing the His-2 mutation, required only slight adjustment of the liganding sidechains to create the metal site.

More significant however, is the effect of the different zinc (and ligand) charges on the metal site. Except for the energy of the *Model1* structure, all the data from Tables 20, 21 and 22 produced by the formal charge, differ markedly from those produced by the GROMOS, GRID and *Shen* charges. It appears that a zinc charge of +2, unshared over the ligands, is better at retaining a more 'typical' metal-ligand complex. This contradicts the findings of Shen *et al.*, (1990). Their *ab initio* calculated electrostatic partial charges account for a distribution of positive charge to each of the ligands. The EM of Cu-Zn SOD, for which they were derived, gave good agreement with the X-ray structure at the metal binding sites (mean

displacements for Cu and Zn are 0.32Å and 0.23Å respectively, as opposed to 0.67Å and 1.05Å after EM using formal charges (Shen *et al.*, 1989)). The deviations caused by the assignment of formal charges result, they propose, from the excessively strong electrostatic interaction between the metal ion and its ligands.

This was not shown by EM of the two model structures using a formal zinc charge. While maintaining more typical metal-ligand distances for the majority of the ligands, it exhibited no evidence of strong electrostatic interactions (*i.e.* particularly short metal-ligand distances). Adjustment of the zinc (and ligand) charge(s) to account for the distribution of positive charge, rather than balancing the magnitude of the opposing charges, results in electrostatic repulsion between the ligands owing to an insufficient attractive force on the zinc.

This is also in disagreement with the findings of Fischer-Hjalmars *et al.*, (1982) because any distribution of positive charge to the ligands should reduce the electrostatic repulsion between them. Any retention of more representative zinc-ligand distances as a result of this were not observed. The similarity of the GROMOS, GRID and *Shen* forms after EM, indicates a similar balance between the attractive and repulsive forces in all three, despite the (small) differences in zinc and ligand charges.

It should be remembered that it was not possible to adjust the charges of the Cys ligands to reflect the influence of the metal charge. This should be considered when comparing the results presented here with those presented by Shen *et al.* and Fischer-Hjalmars *et al.*

While electrostatic attraction may be the predominant factor influencing metal

binding, the importance of covalent interactions must not be forgotten. The electrostatics arise from the loss of the $4s^2$ electrons from the zinc atom to give the Zn^{2+} ion. This ion would be attracted to the electron-rich metal site created from the lone electron pair of the imidiazole ring nitrogen of His-2, the bonding electron pairs from the Cys-25 and Cys-29 sulphur atoms and possibly the Glu-27 carboxyl or Ser-28 mainchain carbonyl group.

In order to accommodate these electron pairs when bound in the site, the vacant $4s$ and $4p$ orbitals undergo sp^3 hybridisation to give 4 equal bonding orbitals at the points of a regular tetrahedron (*c.f.* carbon) (Williams—personal communication) leading to the tetrahedral stereochemistry characteristic of zinc metalloproteins (see Chapter 2, Section 2.5.2.4). The sharing of the bond pair and lone pair electrons which occupy these orbitals is the covalent component of metal binding.

It is the covalent bonding that is partly responsible for the observed stereochemistry at the metal binding sites. The ligands themselves, through steric interactions, also influence metal site stereochemistry (*e.g.* Methionine in the copper proteins—see Chapter 2, Section 2.5.2.3). However, more significantly, the ligands can, to an extent, tailor metal ion function *via* orbital overlap (Hay, 1987). Therefore, metal binding could be regarded as a dual phenomenon involving both electrostatic (ionic) and covalent interactions (Chapter 2, Section 2.5), and any charges derived for the study of metal binding sites should account for both of these components.

4.4 Conclusions

A metal binding site had been modelled in CDR-L1 of HyHEL-5 without reference to any particular metalloprotein structure. Conformational search followed by EM in the presence of zinc, indicated that the site could be readily formed and was stable (particularly in the case of the *Model1H2* structure). The modelling of this site showed that it was possible to create a metal binding site *ab initio* and introduces the possibility of tailor-made metalloantibodies (or other metalloproteins). Creation of the actual metalloantibody could now be carried out and is described in Chapter 5.

Chapter 5

Creation and Analysis of Metalloantibodies

The modelling studies described in Chapter 4 indicated that, from a structural point of view, the creation of a functional metal binding site within L1 of HyHEL-5 (Sheriff *et al.*, 1987) was a realistic possibility. However, the cDNA clones for this antibody were not available, so the six complimentarity-determining regions (CDRs) were grafted onto the framework of the anti-lysozyme antibody Gloop 2 (Jeffrey *et al.*, 1992). The chimeric antibody (henceforth referred to as G2H5) was prepared by Dr. D. Staunton and used in all subsequent protein engineering experiments.

Two metalloantibodies were constructed from G2H5. One contained the three mutations in CDR-L1; Ala-25→Cys, Ser-27→Glu and Val-29→Cys (G2H5-L1CEC), while the other contained an additional light chain N-terminal Ile-2→His mutation (G2H5-HCEC). This chapter describes the construction, subcloning, oocyte expression and characterisation of the three-mutant antibody. These experiments were carried out in conjunction with Dr. Staunton. Similar experiments on the four-mutant antibody were carried out independently by Dr. Staunton.

5.1 Generation of Mutant cDNA

5.1.1 Oligonucleotide Site-Directed Mutagenesis

The technique of site-directed mutagenesis (SDM) was used to create the mutations in the G2H5 sequence. The basis of SDM is the use of a short synthetic oligonucleotide (e.g 20-mer for a single base change) which is complementary to the wild type DNA except at the site of mutation. The oligonucleotide is used as a primer for DNA polymerisation using single-stranded (ss) wild type DNA as a template (usually in a bacteriophage M13 vector (Norrander *et al.*, 1983)).

The hetroduplex is ligated to form a covalently closed circular (ccc) double-stranded (ds) molecule, the replicative form (RF) of M13, which is transformed into E. *coli* cells. This results in a low frequency of mutation, often less than 5%. The major limitation to the efficiency is that the mutant strand, the (−) strand, is unmethylated and is acted upon preferentially by the host cell repair system using the (+) strand as the template.

The mutation efficiency can be increased by positive selection of the mutant strand during the reaction. The method of strand selection developed by Eckstein and coworkers (Taylor *et al.*, 1985a; Nakamaye and Eckstein, 1986; Sayers *et al.*, 1988) was employed in the mutagenesis of G2H5 (Figure 13). This method is based on the observation that certain restriction endonucleases (for example *Ava*I, *Ava*II, *Ban*II, *Hind*II, *Nci*I and *Pvu*I) are incapable of hydrolysing phosphorothioate internucleotide linkages (Taylor *et al.*, 1985b).

A mismatch oligonucleotide primer is annealed to a single-stranded (+) strand phage DNA and extended by Klenow polymerase, in the presence of T4 DNA ligase,

with a deoxynucleoside 5′-O-(i-thiotriphosphate) (dNTPαS) in place of one of the corresponding dNTP's. The phosphorothioate groups are incorporated into the (−) strand of the newly synthesized dsDNA. Addition of one of the above restriction enzymes results in the generation of nicks in the (+) strand which present sites for exonuclease III. This digests all or part of the wild type (non-phosphorothioate) strand in the 3′–5′ direction. Polymerisation and ligation using the mutant strand as a template, generates closed circular mutant dsDNA which can be transformed into the host E. *coli*. The method improves the efficiency of mutation by up to 95% (Taylor *et al.*, 1985a; Nakamaye and Eckstein, 1986). Unless otherwise stated, mutagenesis was carried out using the oligonucleotide-directed *in vitro* mutagenesis system as supplied by Amersham International p.l.c., according to the instructions provided therein.

5.1.2 Synthetic Oligonucleotides for SDM

Two oligonucleotides, a 42-mer containing the CDR-L1 sequence with mutations Ala-25→Cys, Ser-27→Glu and Val-29→Cys, and a 20-mer containing the light chain N-terminal mutation Ile-2→His, were synthesized by Mrs V. Cooper (Oxford Centre for Molecular Sciences) on an Applied Biosystems DNA synthesizer using phophoramidite chemistry. Each oligonucleotide was purified by polyacrylamide/urea gel electrophoresis using 10% for the 42-mer and 20% for the 20-mer. Approximately 50μl (0.2μmoles) of each oligonucleotide was resuspended in gel loading buffer (80% w/v formamide; 50mM Tris-borate, pH 8.3; 1mM EDTA, pH 8.0; 0.1% w/v xylene cyanol FF; and 0.1% w/v bromophenol blue), heated to 105° C for 5 minutes and 13μl aliquots loaded into 4 wells. The gels were run at 35mA for 6

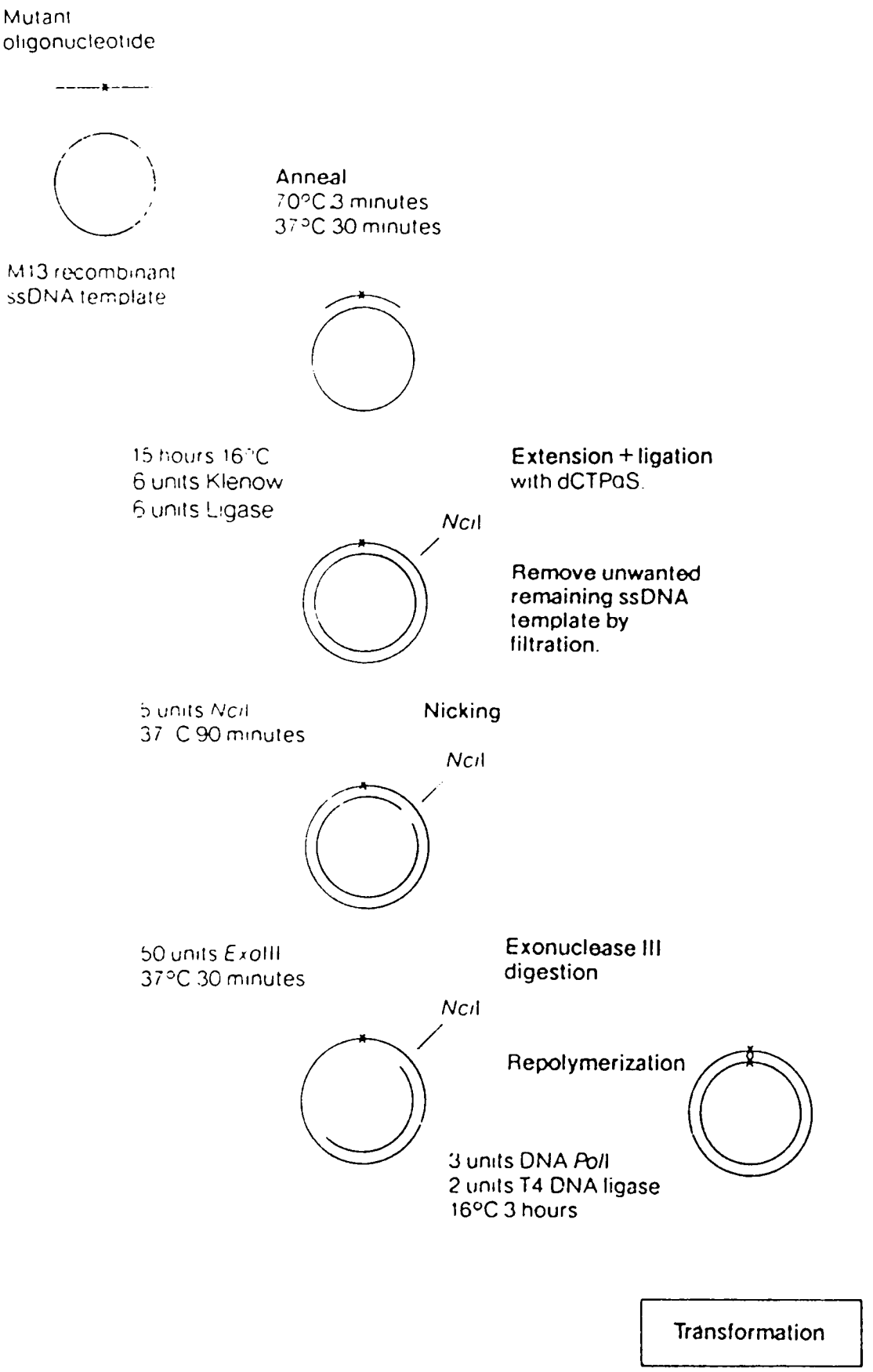


Figure 13: The Eckstein mutagenesis method as used in the Amersham oligonucleotide-directed *in vitro* mutagenesis system (Amersham International p.l.c.).

hours.

The gels were placed between two sheets of cling film and viewed under long wavelength ultraviolet (UV) light on a fluorescent TLC plate. The position of the slowest mobility band on each gel was marked, cut out and the gel slices dialysed in 1ml of Milli-Q water overnight to remove the urea. Each resuspended sample was filtered through a 0.2 micron filter to remove any remaining traces of polyacrylamide or other contaminants. Each sample was dried down in a speedvac and resuspended in a total volume of 50μl Milli-Q water before storage. The concentration of each oligonucleotide was determined by UV spectroscopy.

To protect the 5′ terminus of the mutagenic oligonucleotide from displacement by the newly growing DNA strand during mutagenesis and prevent 5′ exonucleolytic editing of mismatched nucleotides after transfection into E. *coli*, the 5′ terminus of the mutagenic oligonucleotide must be phosphorylated. A 50pmole sample of the 42-mer oligonucleotide was mixed with 3μl of kinase buffer (1M Tris pH8.0, 100mM $MgCl_2$, 70mM DTT and 10mM ATP), 22.5μl Milli-Q water and 2 units of T4 polynucleotide kinase in a total volume of 30.5μl. The reaction was performed at 37° C for 15 minutes, then the kinase was inactivated by heating at 70° C for 10 minutes. The 5′ phosphorylated oligonucleotide primer was stored at -20° C.

5.1.3 Site-Directed Mutagenesis

The Eckstein method outlined in the previous section was followed as detailed below. At various stages, samples were taken for subsequent analysis to check the efficiency of the mutagenesis.

Template Annealing: 1.6pmoles of phosphorylated mutagenic primer was mixed with 1μg of ssM13K19 DNA containing the CDR-replaced Gloop 2 light chain sequence (M13G2H5) in 120mM Tris-HCl, pH8.0; 120mM NaCl in a total volume of 17μl. The mixture was placed in a 70° C water bath for 3 minutes and then a 37° C water bath for 30 minutes.

Synthesis and Ligation: The annealing reaction was adjusted with 50mM Tris-HCl, pH8.0; 50mM NaCl; 10mM $MgCl_2$; 1mM ATP; 250μM of each dNTP, with dCTPαS replacing dCTP, in a total volume of 34μl. The reaction was started by the addition of 6 units of E. *coli* DNA polymerase Klenow fragment and 6 units of T4 DNA ligase, and incubated at 16° C overnight. A 1μl sample (sample 1) of the reaction mix removed for analysis, showed the presence of covalently closed circular (ccc) DNA with a smear of material above it.

Removal of ss DNA: To the reaction from the previous step, 30μl of 5M NaCl was added and the volume adjusted to 250μl with Milli-Q water. The ssDNA was removed by filtration through nitrocellulose. The filter was washed with 100μl of 500mM NaCl to collect any remaining RF DNA. To the filtrate, 28μl of 3M sodium acetate, pH5.2 and 700μl of cold ethanol (-20° C) was added, the solution left at -20° C for 20 minutes and then centrifuged for 15 minutes. The precipitated RF DNA was washed with 1ml of cold ethanol (-20° C) and the pellet air dried. The pellet was resuspended in 25μl of 10mM Tris-HCl, pH8.0; 20mM NaCl; 1mM EDTA.

***Nci*1 Nicking of Non-Mutant DNA:** 10μl of the filtered sample was taken (the remaining material was stored at -20° C) and 25mM Tris-HCl, pH8.0; 75mM NaCl; 50mM $MgCl_2$; 25mM DTT added in a total volume of 65μl. The non-mutant DNA strand was nicked by adding 5 units of *Nci*1 and incubating at 37° C for 90 minutes. A 10μl sample (sample 2) was removed from the reaction mix for analysis to determine the efficiency of the filtration and nicking with *Nci*1. Although a nicked circular band was evident, the presence of a band corresponding to closed circular DNA indicated that the efficiency of the nicking reaction was low.

Digestion of Non-Mutant DNA with Exonuclease III: To the reaction mix from the previous step 60mM Tris-HCl, pH8.0; 90mM NaCl; 6mM $MgCl_2$; 10mM DTT was added. The non-mutant DNA strand was digested by adding 50 units of Exonuclease III and incubating at 37° C for 30 minutes. The enzyme was inactivated by placing in a 70° C water bath for 15 minutes. A 15μl sample (sample 3) of the reaction mix was removed for analysis to monitor the efficiency of the Exonuclease III digest. The DNA band from this sample showed increased mobility relative to sample 2 and is slightly smeared due to exonuclease digestion from the nicked site.

Repolymerisation and Ligation: 800μM ATP, 200μM of each dNTP and 10mM $MgCl_2$ were added to the reaction mix from the previous step. The gapped DNA was repaired by adding 3 units of DNA polymerase I and 2 units of T4 DNA ligase, and incubating at 16° C for 3 hours. A 15μl sample (sample 4) of the reaction mix removed for analysis, showed the DNA after repolymerisation

and ligation. A band indicating the presence of closed circular mutant DNA was clearly visible.

Ammonium Acetate Precipitation: Unincorporated nucleotides were removed by precipitation of the mutant RF DNA in 4M ammonium acetate pH5.4 and ethanol (1:4 v/v) at room temperature for 10 minutes. The mutant RF DNA was collected by centrifugation for 10 minutes. The pellet was washed with 500μl of ethanol, air dried and resuspended in 100μl 10mM Tris-HCl pH8.0; 20mM NaCl.

Transformation Reaction: 300μl aliquots of competent TG1 cells (Biggin, unpublished) (prepared as described in the Amersham protocol) were transformed with 20μl of the previously prepared mutant DNA in pre-chilled, sterile polypropylene tubes. The cells were incubated on ice for 40 minutes then heat shocked at 42° C for 45 seconds. The cells were placed on ice for 5 minutes. 200μl of exponentially growing TG1 cells were then added to each tube. The cells were plated out in 3ml of molten (45° C) H-top agar and poured onto pre-warmed (37° C) H-agar plates. Once set the plates were incubated at 37° C overnight.

Examination of the plates after overnight incubation, showed that the transformation with 20μl of mutant DNA produced what was estimated to be over 2500 plaques resulting in an almost clear plate! If the reactions were efficient, the production of a large number of mutant RF DNA molecules may have resulted in the large number of plaques obtained. However, if the efficiency of one or more of the mutagenesis reactions were low, then the large number of plaques were due to transformation with

wild-type RF DNA. To establish whether the mutagenesis had worked correctly, the samples (1–4) taken previously, were analysed on an agarose gel (Figure 14).

The ambiguous result from sample 2 of the mutagenesis reactions suggested that, although mutant RF DNA had been produced, a large number of M13G2H5 RF DNA molecules may have been present. Two further transformations using 1μl and 5μl of DNA were carried out to allow picking of single plaques. The 1μl transformation yielded $\approx$ 125 plaques and the 5μl transformation yielded $\approx$ 25 plaques. 6 plaques from the 1μl transformation were picked for sequencing. Single-stranded template DNA was prepared according to the protocol described in the Amersham mutagenesis system, but completed at step 16, resuspending each DNA sample in 50μl Milli-Q water. An aliquot of each culture was stored at 4° C.

5.2 Sequencing the ssDNA Template

Each of the ssDNA samples were sequenced to determined which plaques contained the L1CEC mutations in the M13G2H5 template and to check that no spurious mutations were present. Sequencing was performed using the 'Sequenase kit version 2' (United States Biochemical Corporation, USA) which is based on the dideoxy chain termination method (Sanger *et al.*, 1977). This involves the synthesis of a DNA strand by DNA polymerase *in vitro* using the single-stranded DNA template. Synthesis is initiated at only the site where an oligonucleotide primer anneals to the template. The synthesis reaction is terminated by the incorporation of a nucleotide analogue which will not support continued DNA elongation. The analogues are 2′, 3′-dideoxynucleoside 5′-triphosphates (ddNTP's) which lack the 3′-OH group

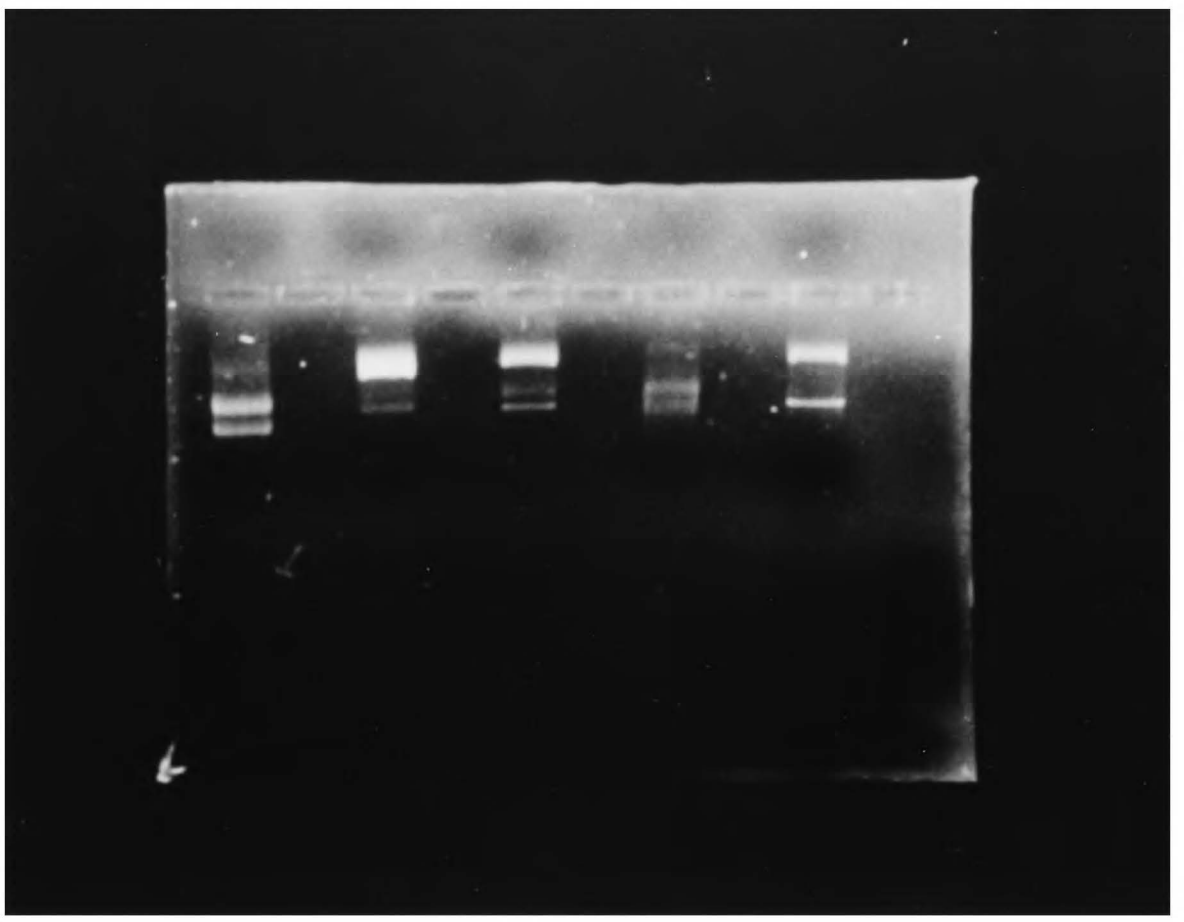


Figure 14: Samples 1–4 taken during mutagenesis according to the Amersham protocol (Amersham International p.l.c.) and analysed on a 0.8% w/v agarose gel. **M:** M13G2H5 single-stranded template marker; **S1:** After sysnthesis and ligation; **S2:** After nicking with *Nci*1; **S3:** After Exonuclease III digestion; **S4:** After repolymerisation and ligation. *ccc*: covalently closed circular

necessary for chain elongation.

When mixtures of dNTP's and one of the four ddNTP's are used, enzyme-catalysed polymerisation will be terminated in a fraction of the population of chains at each site where the ddNTP's can be incorporated. Four separate reactions, each with a different ddNTP give the complete sequence. A radioactively labelled nucleotide is also included in the synthesis, so the labelled chains of various lengths can be visualised by autoradiography after separation by high-resolution electrophoresis.

5.2.1 Preparation of M13G2H5-L1CEC DNA

Autoradiographic analysis of the six clones showed that 3 represented M13G2H5 template, 1 contained a sequence deletion in the region encoding the N-terminus, while 2 contained the L1CEC mutations successfully incorporated into M13G2H5 DNA. The DNA from these two clones was prepared for subcloning of the L1CEC light chain variable region into an RNA transcription vector.

Two separate 500μl volumes of 1$\times$ TY broth were inoculated with supernatant from phage stocks of the two M13G2H5-L1CEC clones. Each of the 1$\times$ TY cultures were streaked out separately onto three H-agar plates. 45ml of molten H-agar was inoculated with 100μl of overnight TG1 cells, poured onto the phage-streaked plates and allowed to set. The plates were incubated at 37° C overnight. A plaque was picked and placed in 10ml of 1$\times$ TY broth, shaken at 37° C for 5 hours, then transferred to 200ml of 2$\times$ TY broth and shaken at 37° C overnight. This ensured that the growth rates of the infected and non-infected cells remained in proportion. M13G2H5-L1CEC DNA was prepared from the overnight culture. The isolation of the DNA was carried out using the method of Birnboim and Doly (1979) with

modifications (Appendix 2, Section II.3.1), followed by Sephacryl-200 purification (Appendix 2, Section II.3.2).

5.3 Preparation of L1CEC Light Chain Insert

An L1CEC light chain variable region insert was prepared by restriction digests of the M13G2H5-L1CEC DNA and subcloned into the SP6 RNA transcription vector. This is based on the SP64T vector developed by Melton and co-workers (1984). Besides the ampicillin resistance gene, the plasmid contains the *Salmonella typhimurium* phage SP6 promoter, and the 5′ and 3′ flanking regions from the efficiently transcribed *Xenopus* β-globin mRNA. The cDNA for the Gloop 2 light and heavy chains had been cloned into the vector at the *Bgl*II and *Bgl*II-*Bam*H1 sites respectively (Roberts, 1987) (Figure 15). The chains were subsequently modified by Dr. Staunton to give the CDR-replaced Gloop 2 used in these experiments.

5.3.1 *Hpa*1 Linearisation of M13G2H5-L1CEC DNA

M13G2H5-L1CEC DNA was linearised by cleavage with the restriction endonuclease *Hpa*1. This enzyme recognises the palidromic hexonucleotide sequence GTTAAC, cleaving between the AT bases to produce 'blunt end' DNA.

20μl of M13G2H5-L1CEC DNA was mixed with 33mM Tris-acetate, pH7.9; 10mM magnesium acetate; 66mM potassium acetate; 0.5mM DTT; 1.5μl of *Hpa*1; 5.5μl of Milli-Q water in a total volume of 30μl and the reaction incubated at 37° C overnight. The 'completeness' of the digest was checked by running 1μl aliquots of both the uncut and linearised DNA on an 0.8% w/v agarose gel.

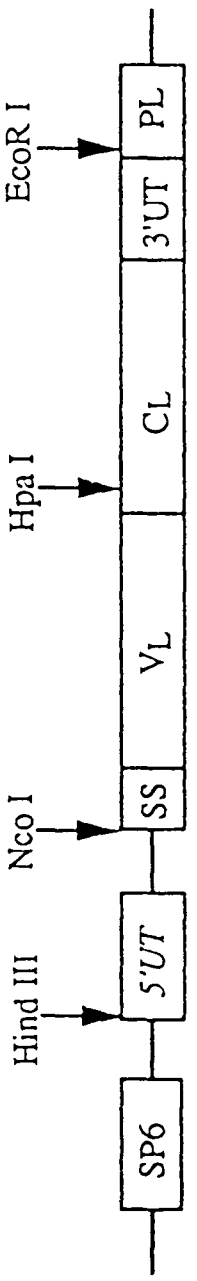


Figure 15: Detail of the SP64T RNA transcription vector light chain construct. **SS:** Immunoglobulin signal peptide; **V:** Variable region; **C:** Constant region; **UT:** Untranslated region of the Gloop 2 gene; *UT*: Untranslated region of the *Xenopus* β-globin gene; **PL:** Polylinker.

The linearised DNA was extracted with 0.5 v/v phenol:chloroform and the aqueous layer extracted with 0.5 v/v ether to remove any traces of chloroform. The DNA was precipitated with 5μl 3M sodium acetate, pH5.2 and 2.5 v/v ethanol in a dry-ice bath for 30–60 minutes. The DNA was collected by centrifugation for 20 minutes and the pellet resuspended in 50μl Milli-Q water.

5.3.2 *Nco* 1 Cleavage of Linearised M13G2H5-L1CEC DNA

The restriction endonuclease *Nco* 1 also recognises a palindromic hexonucleotide sequence, CCATGG, but cleaves between the CC bases producing 'sticky end' DNA. Cleavage of the linearised M13G2H5-L1CEC DNA with *Nco* 1 removes the L1CEC light chain variable region the from M13G2H5 DNA.

20μl of *Hpa* 1-cleaved M13G2H5-L1CEC DNA was mixed with 50mM Tris-HCl, pH7.5; 10mM $MgCl_2$; 100mM NaCl; 1mM DTE; 1.5μl *Nco* 1 and 5.5μl Milli-Q water in a total volume of 30μl. The reaction was incubated at 37° C overnight. 2μl aliquots of uncut M13G2H5-L1CEC DNA, *Hpa* 1-cleaved DNA, *Hpa* 1/*Nco* 1-cleaved DNA and a λ molecular weight (Mw) marker were then run on an 0.8% w/v agarose gel (Figure 16). A second gel containing 28μl *Hpa* 1/*Nco* 1-cleaved DNA and 4μl of λ Mw marker was run to separate the L1CEC light chain variable region from the M13G2H5 DNA.

The band containing the L1CEC light chain variable region DNA was cut out from both gels, the slices placed in a unidirectional electroelutor (International Biotechnologies, New Haven) in 20mM Tris-HCl; 0.2mM EDTA; 5mM NaCl and eluted at 100 volts for between 30–60 minutes into a 70μl cushion of 3M sodium acetate pH5.2, containing Bromophenol blue. Complete elution from the gel was

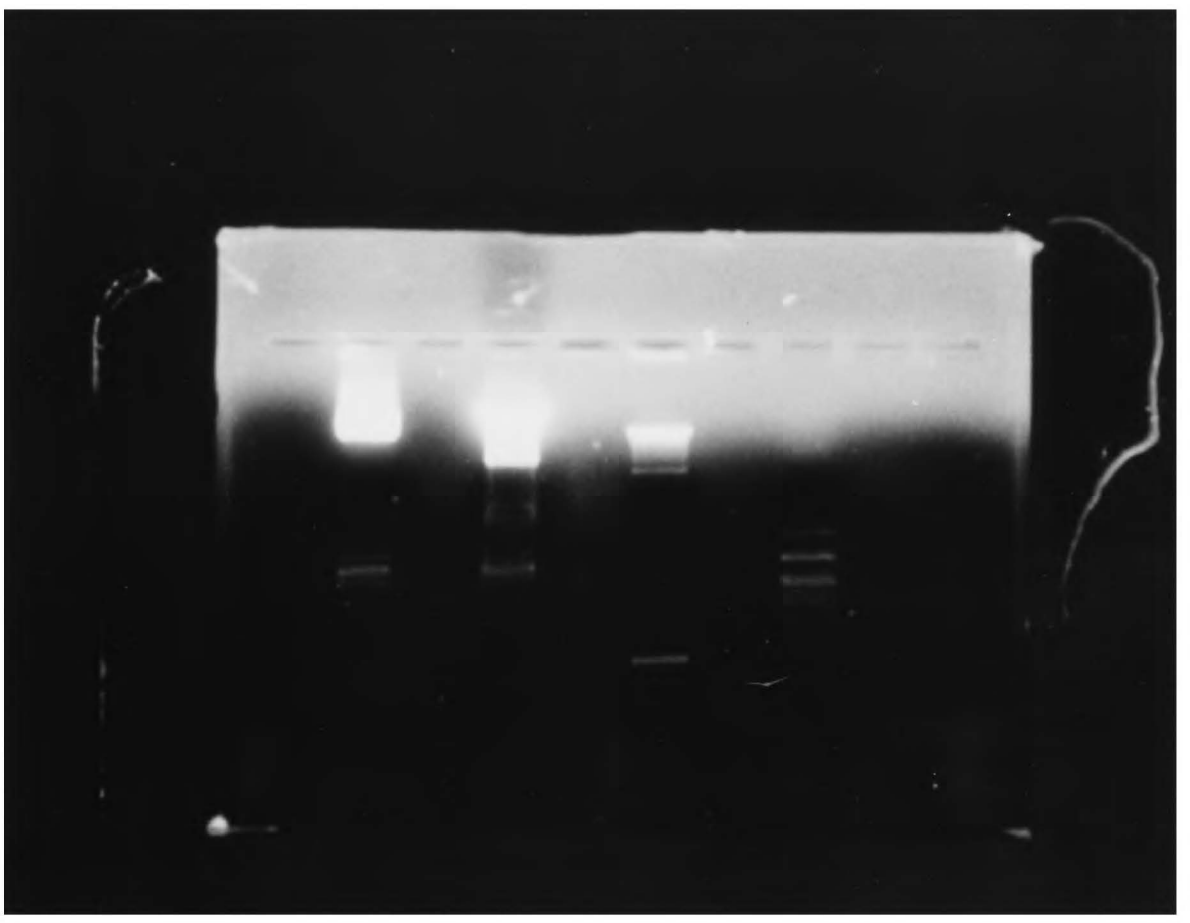


Figure 16: *Nco*1 cleavage of linearised M13G2H5-L1CEC DNA. Lane **1:** Uncut M13G2H5-L1CEC DNA; lane **2:** *Hpa*1-cleaved DNA; lane **3:** *Hpa*1/*Nco*1-cleaved L1CEC light chain DNA; lane **4:** λ Mw marker.

monitored by loss of UV fluorescence. The sodium acetate plus DNA was carefully transferred to a microfuge tube and the DNA precipitated by addition of 2.5 v/v ethanol then incubation at -20° C overnight.

5.3.3 Transformation of MC1061 Cells with pSP6G2H5-L1CEC

The L1CEC light chain variable region DNA was ligated into the SP64T vector containing the CDR-replaced Gloop 2 light chain constant and heavy chain regions (pSP6G2H5). Four ligation reactions were set up containing 1μl, 2μl, 4μl and 0μl (control) of L1CEC light chain DNA respectively. These were mixed with 10mM Tris-HCl, pH7.5; 10mM $MgCl_2$; 20mM DTT; 1mM rATP; 1μl of linearised pSP6G2H5 DNA, previously prepared by *Hpa*1/*Nco*1 digestion, and 0.5μl T4 DNA ligase. Milli-Q water was added to each tube to give a total volume of 20μl. The reactions were left at room temperature for 5 hours.

200μl of competent MC1061 cells (Casadaban and Cohen, 1980) (prepared as described in Appendix 2, Section II.3.3) were added to each tube and to a fifth microfuge tube containing 1μl of pBR used as a control to test cell competence. The cells were incubated at 4° C for 15–30 minutes. Heat shock was performed by placing the tubes in a 37° C water bath for 4 minutes and then on ice for 5 minutes. 800μl of LB media was added and the tubes placed in a 37° C water bath for 30 minutes to allow recovery from heat shock. Transformed cells from the 1μl, 2μl and 4μl ligations were selected by plating out 220μl aliquots onto L-agar containing 100μg/ml ampicillin. Cells from the 0μl control were pelleted by centrifugation for

Plate	Ligation				
	1μl	2μl	4μl	pSP6G2H5	pBR
1	13	28	18	11	66
2	15	19	16	—	—
3	35	35	37	—	—
4	30	54	30	—	—
5	5	20	13	—	—
Total	98	146	114	11	66

Table 23: Total number of colonies produced from the transformation of MC1061 cells with pSP6G2H5-L1CEC DNA for each ligation. *pSP6G2H5* and *pBR* are controls.

20 seconds and the supernatant poured off before plating out. A 100μl aliquot of cells from the pBR control were plated out.

Table 23 shows the number of transformed MC1061 colonies produced from each ligation, indicating that the overall efficiency of transformation was high. The pSP6G2H5 control measures the level of 'background' *i.e.* the level of pSP6G2H5 vector incorporated into the cells. The pBR control checks the quality of the competent cells. From these results, it can be seen that the 2μl ligation produced the largest number of transformed colonies. The difference in transformation totals is due to the ratio of L1CEC light chain:pSP6G2H5 DNA in each ligation. 12 colonies were picked from plate 4 of the 2μl ligation and grown up in 10ml culture tubes containing 1.5ml LB media with ampicillin (2μl/ml). The cells were then shaken at 37° C overnight.

5.4 Preparation of L1CEC Light Chain DNA for Transcription

5.4.1 Small Scale Plasmid Preparations from Transformed Cells

5ml plasmid preparations of the overnight cells were carried using the method of Birnboim and Doly (1979) with modifications, as described in Appendix 2 (Section II.3.4). Double *Hpa*1/*Nco*1 digests were set up for each of the 12 clones. In each case, 4μl of DNA was mixed with 10mM Tris-HCl, pH7.5; 10mM $MgCl_2$; 100mM NaCl; 1mM DTE; 0.5μl of *Hpa*1 and *Nco*1 in a total volume of 20μl. A control digest containing 1μl of pSP6G2H5 DNA was also set up. The reactions were incubated at 37° C overnight.

The completeness of the restriction digests was checked by electrophoresis on a 1% w/v agarose gel. 11μl from each clone and the control was loaded onto the gel and run at 60 volts. The gel showed that the L1CEC light chain variable region DNA had been cleaved from all but two of the 12 clones. This indicated that the ligations and restriction digests for 10 clones had been successful.

5.4.2 Sequencing of Three pSP6G2H5-L1CEC Clones

Three clones were sequenced to check that no spurious mutations had occurred during the ligation reaction. 2μl from each of the three small scale plasmid preparations was heated to 100° C for 3 minutes and immediately snap frozen on dry-ice. 200mM Tris-HCl, pH7.5; 100mM $MgCl_2$; 250mM NaCl and a 4 molar excess of

primer in a small volume were added and left at 37° C for 30 minutes, for annealing to take place. The standard labelling and termination reactions according the 'Sequenase' protocol were used, the products separated by denaturing polyacrylamide gel electrophoresis (described in Appendix 2, Section II.2.2) and visualised by autoradiography after 18 hours film exposure.

Of the 3 clones sequenced only one was visible on the autoradiograph. This sequence showed that the L1CEC light chain variable region DNA had been inserted into pSP6G2H5 in the correct orientation. The stock cells from this sample were grown up by innoculating 400ml of LB media containing ampicillin (100μg/ml), with 30μl of cells and shaking at 37° C overnight. 15μl of cells were streaked out onto ampicillin plates and incubated at 37° C overnight. Cells containing pSP6G2H5-L1CEC are ampicillin resistant and are selected preferentially. A very large scale plasmid preparation was carried out on the 400ml of cells using the method of Birnboim and Doly (1979) with modifications (Appendix 2, Section II.3.5), and the DNA purified by CsCl density centrifugation (Appendix 2, Section II.3.6).

The pSP6G2H5-L1CEC DNA obtained from the CsCl purification was sequenced to check the pSP6G2H5:L1CEC light chain junctions. These junctions are the restriction sites for *Nco*1 (5′-end) and *Hpa*1 (3′-end). Sequencing was carried out as before using pSP6 primer to sequence the 5′-end and a primer corresponding to the sequence of CDR-L3 to sequence the 3′-end. The autoradiograph showed the presence of *Nco*1 and *Hpa*1 sites, indicating that the orientation of the insert was correct and also the correct L1CEC light chain variable region sequence.

5.4.3 *Eco*R 1 Linearisation of pSP6G2H5-L1CEC

*Eco*R 1 cleaves pSP6G2H5-L1CEC at a unique site on the 3′ side of the L1CEC light chain sequence, creating a linear template in preparation for RNA transcription. RNA polymerase will synthesize mRNA consisting of a *Xenopus* Shine-Delgarno ribosome recognition sequence, followed by a signal peptide sequence then the Gloop 2 L1CEC light chain sequence. Linearisation prevents transcription of the whole plasmid. 25μl of pSP6G2H5-L1CEC DNA was mixed with 50mM Tris-HCl, pH7.5; 10mM $MgCl_2$; 100mM NaCl; 1mM DTE; 3μl *Eco*R 1 in a total volume of 40μl. The reaction was incubated at 37° C overnight.

The DNA was extracted with 0.5 v/v phenol mix and precipitated with 0.1 v/v 3M sodium acetate, 2.5 v/v ethanol at -70° C for 30–60 minutes. The DNA was spun down, washed with 70% ethanol and resuspended in 50μl of RNase free Milli-Q water. A small aliquot of the reaction was analysed by agarose gel electrophoresis to check the completeness of the reaction.

5.5 Synthesis of L1CEC Light Chain mRNA

mRNA was synthesized from the linearised pSP6G2H5-L1CEC DNA. All solutions and equipment involved in RNA production and manipulation were RNase free. The following were added to a microfuge tube at room temperature (to avoid precipitation of DNA by spermidine at 4° C): 10μl 5× SP6 buffer (200mM Tris-HCl, pH7.5; 30mM $MgCl_2$; 10mM spermidine); 1μl, 1M DTT; 100 units RNasein (Boehringer); 5μg/μl BSA-nuclease-free (BRL); 2.5mM each of ATP, CTP, GTP and UTP; 500μM

mono-methyl cap (Pharmacia), and 3μg of linearised DNA template, in a total reaction volume of 50μl. The synthesis was started by the addition of 15 units of SP6 RNA polymerase. After incubation at 40° C for 60 minutes, a further addition of 15 units of polymerase was made and the synthesis allowed to proceed for another 60 minutes at 40° C.

The DNA was digested with 1μg/μl DNase (RNase-free, Boehringer) at 37° C for 10 minutes and the reaction stopped by the addition of 1μl 0.5M EDTA. The reaction mixture was extracted once with 0.5 v/v phenol:chloroform and twice with chloroform/isoamyl alcohol (24:1 v/v). The RNA was precipitated in 4M ammonium acetate and 100% ethanol (Analar, BDH) at -70° C for at least 12 hours. The RNA was collected by spinning at room temperature for 15 minutes and resuspended in 20μl water (Analar, BDH).

5.5.1 Gel Analysis of RNA Transcripts

2μg (Abs_{260nm} 1.0 = 40μg/ml RNA) of sample RNA transcript and ribosomal RNA marker (Pharmacia) were dissolved in 70% v/v deionised formamide; 10× MOPS running buffer, pH7.0 (200mM 3-(N-morpholino)-propane-sulphonic acid (MOPS); 50mM sodium acetate; 10mM EDTA) and 2.2M formaldehyde. The sample was incubated at 60° C for 5 minutes and then placed on ice for 10 minutes. The RNA was separated by electrophoresis at 50 volts on a 1% w/v agarose gel containing 2.2M formaldehyde and 10× MOPS running buffer. After electrophoresis, the gel was stained with ethidium bromide (1μg/ml) for 15 minutes at room temperature, then destained in Milli-Q water at 4° C overnight and visualised by UV fluorescence (Figure 17).

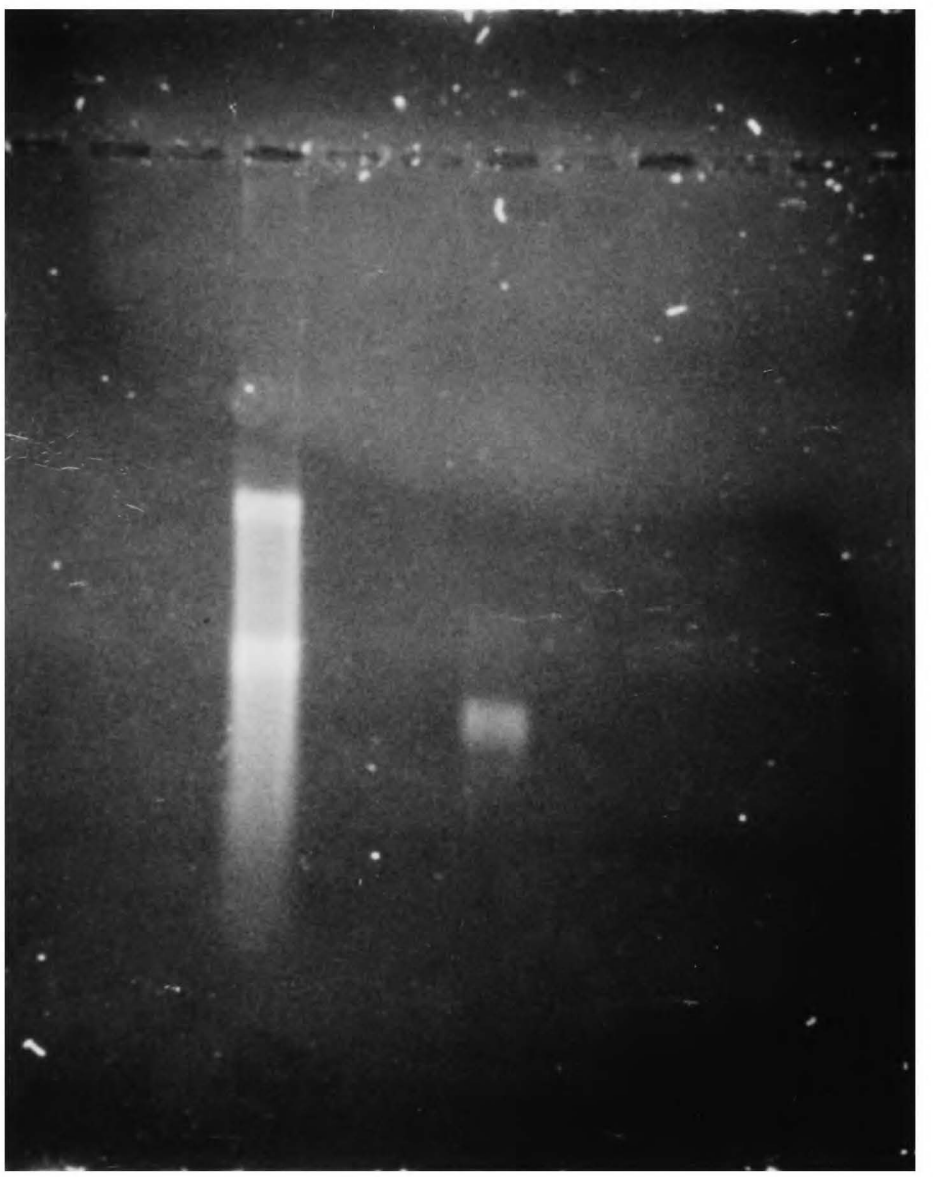


Figure 17: Agarose gel of synthesized RNA transcripts. Lane 1: 28S+18S ribosomal RNA marker; lane 2: G2H5-L1CEC light chain RNA transcripts. The position of the band in relation to the marker band indicates that the transcripts are of the correct size.

5.6 A *Xenopus laevis* Oocyte Expression System

The first demonstration that functional tetrameric IgG chains are secreted from oocytes after microinjection of natural immunoglobulin mRNA was by Valle and co-workers (1981). It has also been shown that RNA produced *in vitro* is processed and translated by *Xenopus* oocytes (Kreig and Melton, 1984; Drummond *et al.*, 1985). The RNA requires a 5′-methyl cap (included in the RNA synthesis) for efficient translation, although the requirement for a poly-A tail in uncertain (Drummond *et al.*, 1985). Roberts and Rees (1986) demonstrated the secretion of functional IgG after microinjection of *in vitro* synthesized light and heavy chain mRNA into the oocyte cytoplasm. The antigen binding properties of the oocyte secreted antibody were found to be identical to the native hybridoma antibody.

The expression system was chosen for its relative ease and rapidity of antibody production for antigen binding studies. This allowed any effects of the mutations on antigen affinity/specificity to be determined prior to large-scale bacterial expression.

5.6.1 Removal and Preparation of Oocytes

Laboratory-bred *Xenopus laevis* females were kept in water tanks as described by Colman (1984). Prior to the removal of a complete ovary, the frog was first immersed in iced water until sluggish and then injected subcutaneously with 150μl of the concentrated barbiturate, Euthetal (May and Baker Ltd, Dagenham). The frog was then placed in shallow water until fully anaesthetised which took 1–2 hours.

Once fully anaesthetised, the frog was placed on its back and an incision made through the loose skin and the body wall. Care was taken not to sever major

abdominal blood vessels. The ovary was removed from the body cavity using forceps and sharp scissors and the large clumps of oocytes placed in a petri-dish of Modified Barth's Saline (MBS—preparation described in Appendix 2, Section II.2.4).

Single oocytes were obtained by gripping a clump of oocytes with forceps and 'stripping' the oocytes off the clumps with another pair of forceps. The single oocytes were transferred into fresh MBS and sorted. Only those oocytes of large size and uniform unmarked dark pigment (animal pole) were chosen. Before microinjection the 'good' oocytes were incubated at 22° C overnight.

Any non-viable oocytes generally become visible during this pre-incubation period and could be removed. Old oocytes (which would subsequently lyse) are discoloured especially at the animal pole and often are very taught so cannot be depressed. After overnight incubation, the oocytes were sorted a second time, under a microscope, to remove any that had become non-viable. The remaining oocytes were kept in MBS at 22° C.

5.6.2 Microinjection of Oocytes

The microinjection equipment consisted of a stereomicroscope, a micromanipulator, a microsyringe and a fiberoptic lamp. Micropipettes were made from 10cm hard glass capillary tubes (BDH) which had been washed in chromic acid and autoclaved to ensure that they were RNase free. A micropipette was connected to the microsyringe and mounted in the micromanipulator. The barrel of the pipette was graduated using the scale in the microscope eyepiece.

1μl of L1CEC light chain, and heavy chain (prepared by Dr. Staunton), RNA transcripts were mixed, and 40nl of the RNA solution injected; approximately 10ng

of each chain per oocyte. The micropipette was calibrated with 1μl of RNase-free water (Analar, BDH). In order to load the pipette, 1μl of solution (water or RNA) was placed on a piece of nescofilm. This was sufficient for approximately 25 injections.

Up to five oocytes were transferred from MBS to a microscope slide onto which a small piece of plastic with a serrated edge had been glued. This immobilised each oocyte in the plastic teeth and facilitated injection. Without allowing the oocyte to dry out, the RNA was injected into the vegetable pole (light pigment) or the equatorial region, to avoid the animal pole which usually contains the nucleus. The injected oocytes were returned to MBS and kept at 22° C until the injections were completed. This allowed any damaged oocytes to lyse.

5.6.3 Labelling of Oocytes with ^{35}S-Methionine

The injected oocytes were incubated in a 96 well tissue culture plate (Falcon). Secreted proteins were prevented from absorbing onto the plastic by coating the plate with a solution of BSA (0.5μg/ml) in MBS at 37° C for 20 minutes. The plate was then washed twice with MBS. Five oocytes were placed in each well and any excess MBS aspirated off. 60μl of suspension fluid (10ml MBS containing 0.5ml 10mM cold L-methionine (to supplement the methionine pool in the oocyte) and 0.5ml foetal calf serum dialysed against MBS (to prevent reduction of the Ig chains)) were added to each well. For synthesis of ^{35}S-Met-labelled translation products, 60μl of ^{35}S suspension fluid (1020μl MBS; 60μl foetal calf serum; 19.2μl 50μM cold L-methionine and 100μl L-[^{35}S]-methionine (40–50 μCi: specific activity 1500Ci/mmol, Amersham International p.l.c.) was also added to each well. The

oocytes were incubated at 22° C for 22–24 hours.

A total of 71 injected oocytes were incubated and of these, 61 had lysed. Thus, only the supernatant from 10 oocytes could be used. This represents a survival rate of only 14.1%. The supernatant from the lysed and unlysed oocytes were pooled into separate sterile microfuge tubes. The debris from both supernatants was collected by centrifugation for 10 minutes at 4° C. The supernatant from the intact oocytes was divided into 60μl aliquots and stored at -20° C.

5.7 Immunoprecipitation of Labelled Products

The procedure for immunoprecipitation is an adaptation of the method described by Colman (1984) (Roberts (1987)). A 20μl aliquot of ^{35}S-labelled oocyte supernatant was mixed with 400μl of RIPA immunoprecipitation buffer (10mM Tris-HCl, pH8.5; 0.15M NaCl; 5mM EDTA; 1% v/v Triton X-100; 1% w/v sodium deoxycholate; 0.1% w/v SDS; 0.1mM phenylmethylsulphonyl fluoride (PMSF)). The mixture was divided into two 200μl aliquots and either 2.5μl of goat-anti-mouse (GAM) IgG polyclonal serum or 2.5μl of a non-immune serum was added. The aliquots were then incubated at 4° C overnight.

The immunocomplexes were precipitated by 50μl of a 50% suspension of protein-A-sepharose (Bioprocessing) in RIPA buffer, with gentle agitation at room temperature for 60 minutes. The protein-A-sepharose immunocomplexes were washed six times with 1ml RIPA buffer, 20μl of dissociation mix (10% β mercaptoethanol; 10% SDS; 25% Glycerol; 10mM Tris-HCl, pH6.9; 0.02% Bromophenol Blue) added and heated at 100° C for 10 minutes. The immunocomplexes were separated by

SDS-PAGE (Laemmli, 1970) as described in Appendix 2 (Section II.2.3) and visualised by autoradiography at -70° C with intensifying screens. The result of the immunoprecipitation is shown in Figure 18.

5.8 Characterisation of the Metalloantibodies

5.8.1 Binding Assay of the L1CEC Metalloantibody

The binding assay was done in triplicate.

HEL Binding: 50μl of a solution containing 25μg/ml hen egg-white lysozyme (HEL) in phosphate-buffered saline (PBS)[1] with 0.05% azide, was added to 24 wells of a PVC V-bottom microtitre plate (Dynatech).

Non-specific Binding: To a further 24 wells, 50μl of PBS containing 0.05% v/v Tween:20 (PBS/Tween)[2] was added.

All wells were checked for air bubbles, the plate covered with cling film and incubated at 4° C overnight. The wells were washed and blocked for 30 minutes, and after an additional wash, 50μl of the pooled ^{35}S oocyte supernatant from the intact oocytes, dissolved in 300μl of PBS/Tween was serially diluted into each well. The plate was covered with cling film and incubated at 4° C overnight.

The supernatant was removed by aspiration and the plate washed three times with PBS/Tween. 100μl of 0.1M NaOH was added to each well and incubated at 37° C for 20 minutes to strip the immune complexes from the plate. The contents of

[1]PBS was prepared by dissolving 80g NaCl; 2g KH_2PO_4; 29g Na_2HPO_4:10H_2O and 2g KCl in 1 litre of Milli-Q water.

[2]PBS/Tween was prepared by dissolving 100ml PBS; 0.5ml polyoxyethylene-sorbitan-monolaurate and 0.5g sodium azide in 1 litre of Milli-Q water.

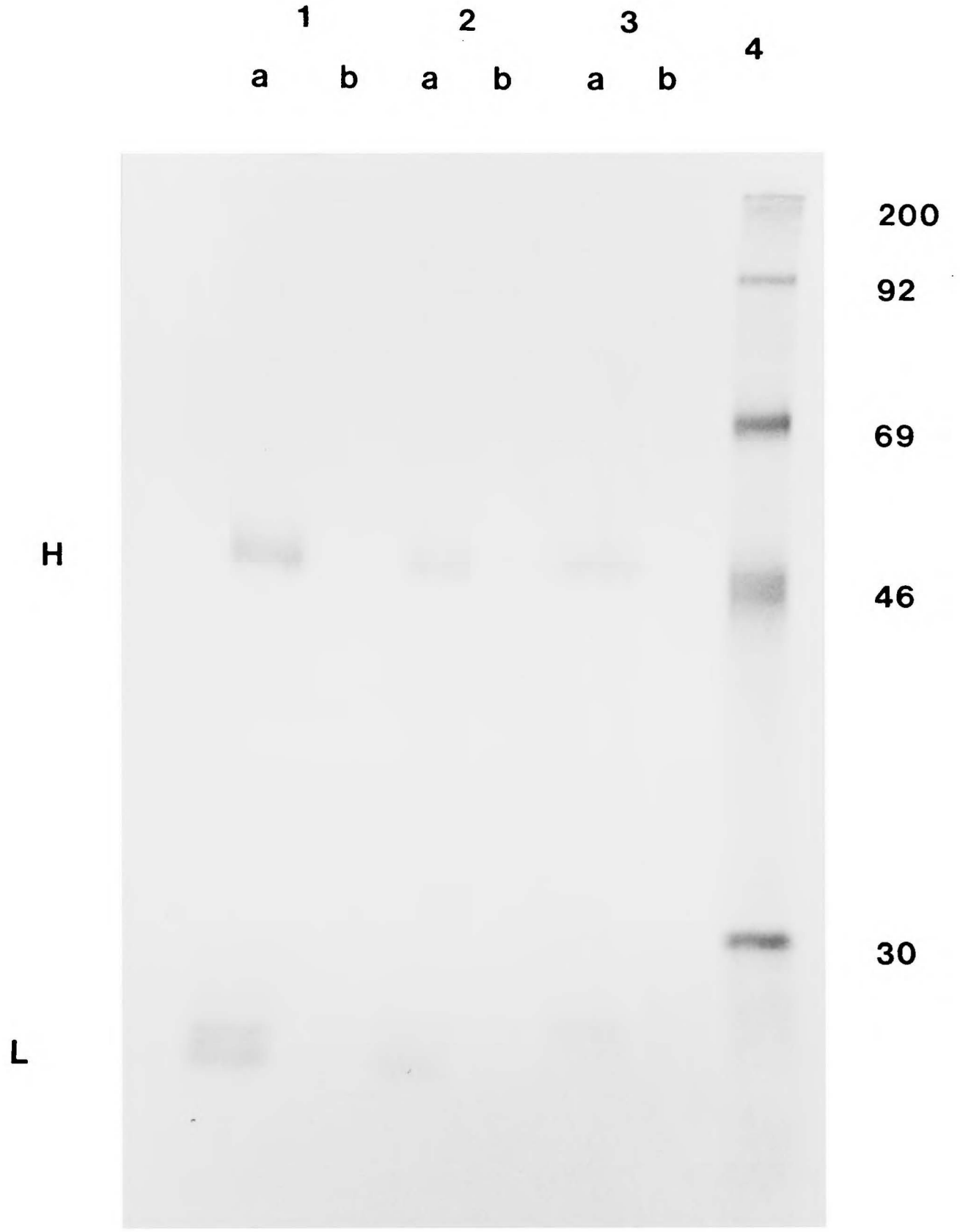


Figure 18: Autoradiograph of ^{35}S-Met-labelled translation products separated by reducing SDS-PAGE. The autoradiograph shows bands for **1:** Antibody G2H5-L1CEC; **2:** Antibody G2H5-HCEC (prepared by Dr. Staunton); **3:** Antibody HyHEL-5 (kindly provided by Dr Smith-Gill); **4:** Mw markers. **a:** Goat-anti-mouse serum; **b:** Non-immune serum; **H:** Heavy chain; **L:** Light chain.

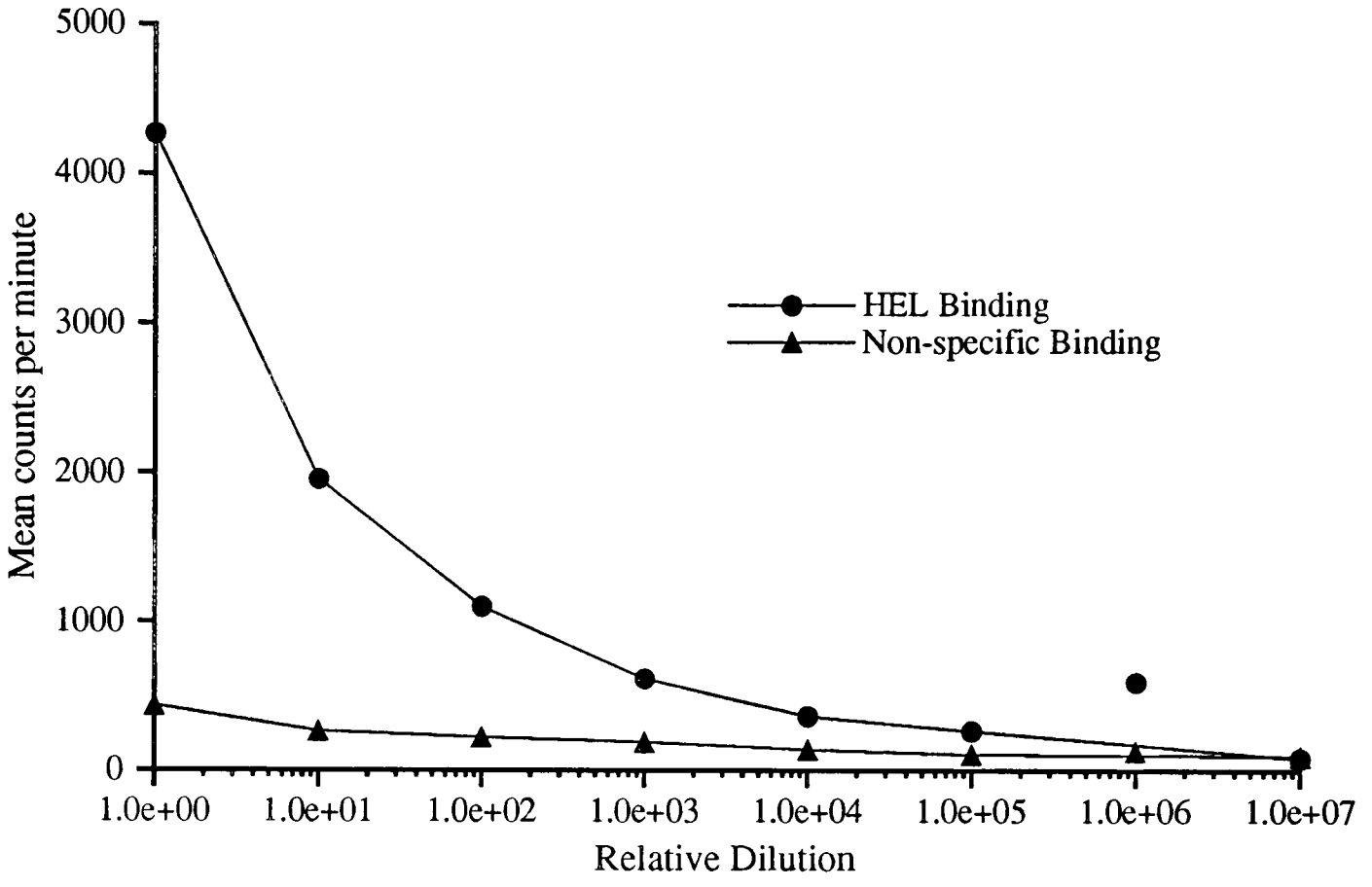


Figure 19: Plot of HEL and non-specific binding for the G2H5-L1CEC antibody. The counts per minute are mean values derived from the highest set of readings for each sample.

each well were then transferred to scintillation vials. The wells were washed twice with NaOH and the washings combined in the corresponding vials.

5ml of scintillant (ATOMLIGHT, Du Pont) was added to each vial, mixed well and counted for 4 minutes in a liquid scintillation counter (LKB 1211 minibeta). The results of the binding assay for the G2H5-L1CEC antibody, with both HEL and non-specific binding, are shown in Figure 19.

5.8.2 NMR Studies of G2H5 and G2H5-HCEC Antibodies

Nuclear magnetic resonance (NMR) provides a powerful tool with which to probe the interaction of protein molecules with ligands such as metal ions and small inhibitors (Campbell and Dwek, 1984). The positions (chemical shift values) of individual resonances in the 1H NMR spectrum of a molecule are determined by the environment of the proton within the protein structure, both local and three-dimensional (Wüthrich, 1986). For a given proton, if this environment is changed by a local interaction, such as with a ligand molecule or metal cation, or by a conformational change which affects the protein's tertiary structure, then the chemical shift position of its resonances will also change.

An antibody Fv molecule (Mw $\approx$ 22 000) is large by proton NMR standards, it's 1H spectrum being characterised by a large number of relatively broad ($\approx$ 15–20Hz), overlapping resonances. Any attempt to obtain sequence specific assignments for the residues in a protein of this size would require heteronuclear labelling. However, histidine residues (thought to be one of the ligands for zinc in the G2H5-HCEC antibody) together with aromatic residues, present a special case since the resonances of their ring protons (2H and 4H in the case of histidine) are often observed to be narrow even in proteins of high molecular weight. As a consequence of this difference in linewidth, the sharper resonances of the sidechain ring protons can be selectively observed in 1H NMR spectra by filtering out the broad components to the spectrum, which will derive from the majority of other residues in the protein by conventional homonuclear proton NMR.

5.8.2.1 Methods

Samples of G2H5 and G2H5-HCEC Fv were prepared for NMR by Dr. Staunton. For G2H5-HCEC, 10mg of the light and heavy chain Fv dimer was purified by FPLC and placed in a concentrator with a 10K cutoff. EDTA was added to a final concentration of 10mM, pH7.5. The EDTA was removed by spinning down to a final volume of 50μl. 1ml of Tris-saline buffer, pH6.8 and 5mM DTT (to prevent disulphide bond oxidation) was added. The buffer was then replaced with deuterated buffer by repeated washing.

The NMR experiments were performed on a home-built Nicolet 500MHz spectrometer (Oxford Center for Molecular Sciences) by Dr. C. R. Redfield. All data were collected at 300K (25° C). For each antibody Fv, a one-dimensional 1H spectrum was collected, employing a relaxation delay of approximately 1s and a sweep width of 10ppm (5 550Hz). 128 scans were collected per Fourier induction decay (FID). The spectra showed two areas of peaks, one upfield of the water resonances corresponding to aliphatic resonances (4.7–0.0ppm), and one downfield containing the aromatic and amide NH resonances (4.7–10.0ppm)—data not shown. The appearance of the amide NH resonances suggested that the protein was folded under the conditions used for the NMR studies.

To probe the interaction of zinc with each antibody Fv, a titration experiment was carried out using a modified 1-D pulse sequence which removes the broad resonance components to the FID. The Carr-Purcell-Meiboom-Gill (CPMG) experiment employs a delay (tau) between the 90 1H pulse and the start of acquisition (90-tau-180-tau), allowing the signals with short T2 relaxation times to decay prior to detection. Thus, only the sharper resonances persist in the 1-D spectrum. For

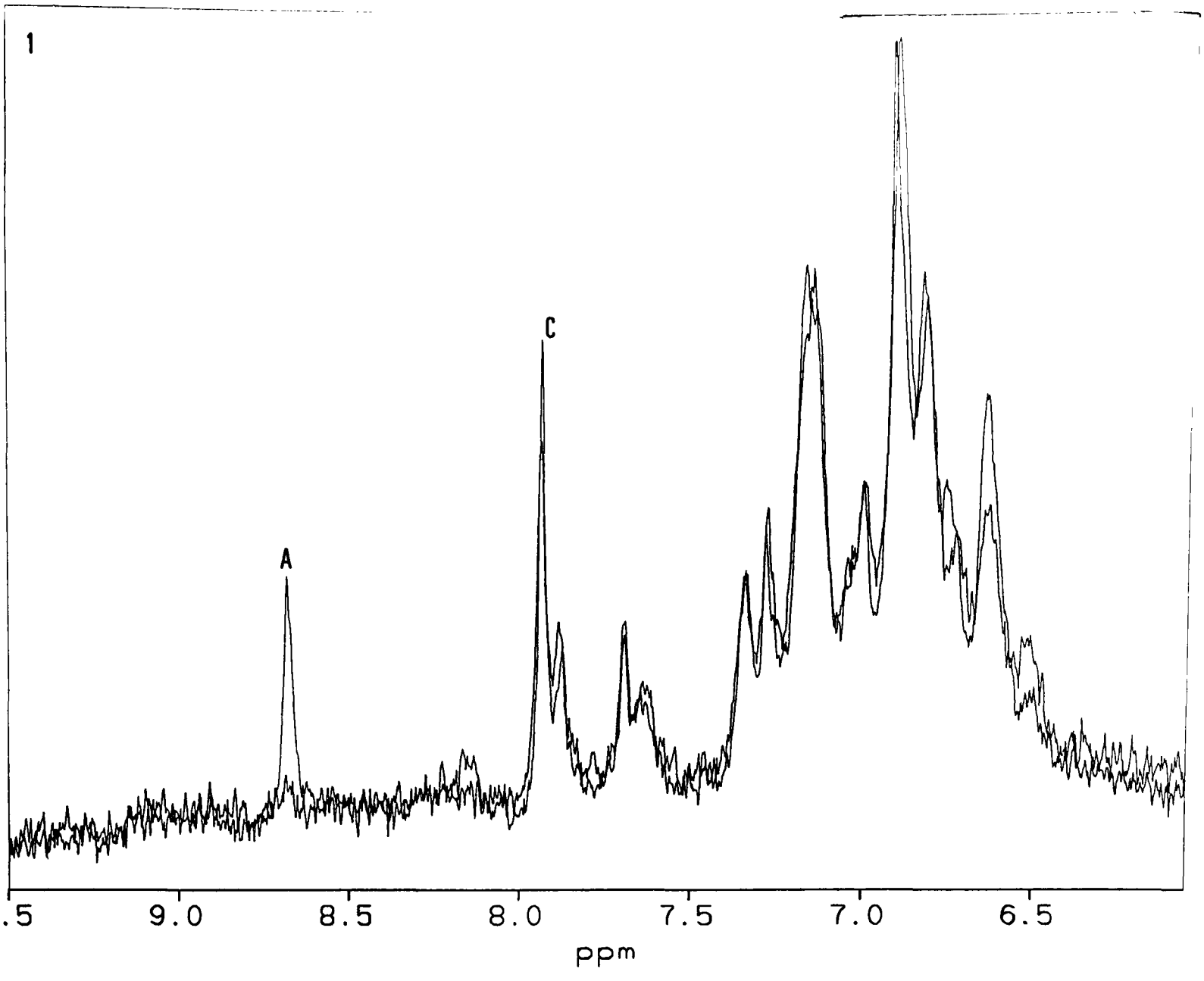
CPMG spectra	Volume $ZnCl_2$ added/μl	Zinc sample	
		Concentration in protein solution/M^{-1}	Zinc molar ratio/%
1	0	Zero	—
2	2	10^{-5}	—
3	2	10^{-4}	25
4	2	10^{-3}	50
5	2	10^{-2}	75
6	4	Saturated	125
7	10	Excess	—

Table 24: Zinc concentrations used for the CPMG spectra of the G2H5-HCEC Fv.

each Fv, the first CPMG spectrum was recorded in the absence of zinc, then Zn^{2+} was titrated directly into the NMR tube by the addition of aliquots of 25mM $ZnCl_2$ to excess (Table 24). 512 scans were collected per FID, with a delay (tau) of 40ms. The data were processed with Gaussian resolution enhancement and the spectra normalised to the $[Zn] = 0$ time point.

5.8.2.2 Results

The CPMG spectra recorded for the G2H5 and G2H5-HCEC antibody Fv's are shown in Figure 20. Resonances from the two histidine ring protons (2H and 4H) have random coil chemical shift values of 8.14 and 7.14ppm respectively. The latter therefore overlaps with resonances that will derive from the ring protons of aromatic residues such as tryptophan and phenylalanine, but the 2H proton which resonates further downfield can be distinguished easily at 8.65ppm in both proteins (peak A in Figure 20)—the value differs from random coil because of the influence of neighbouring residues in the protein structure. G2H5 Fv contains two histidine residues at positions 61 and 98 in the heavy chain, while the G2H5-HCEC Fv


1
A
C
9.5
9.0
8.5
8.0
7.5
7.0
6.5
ppm

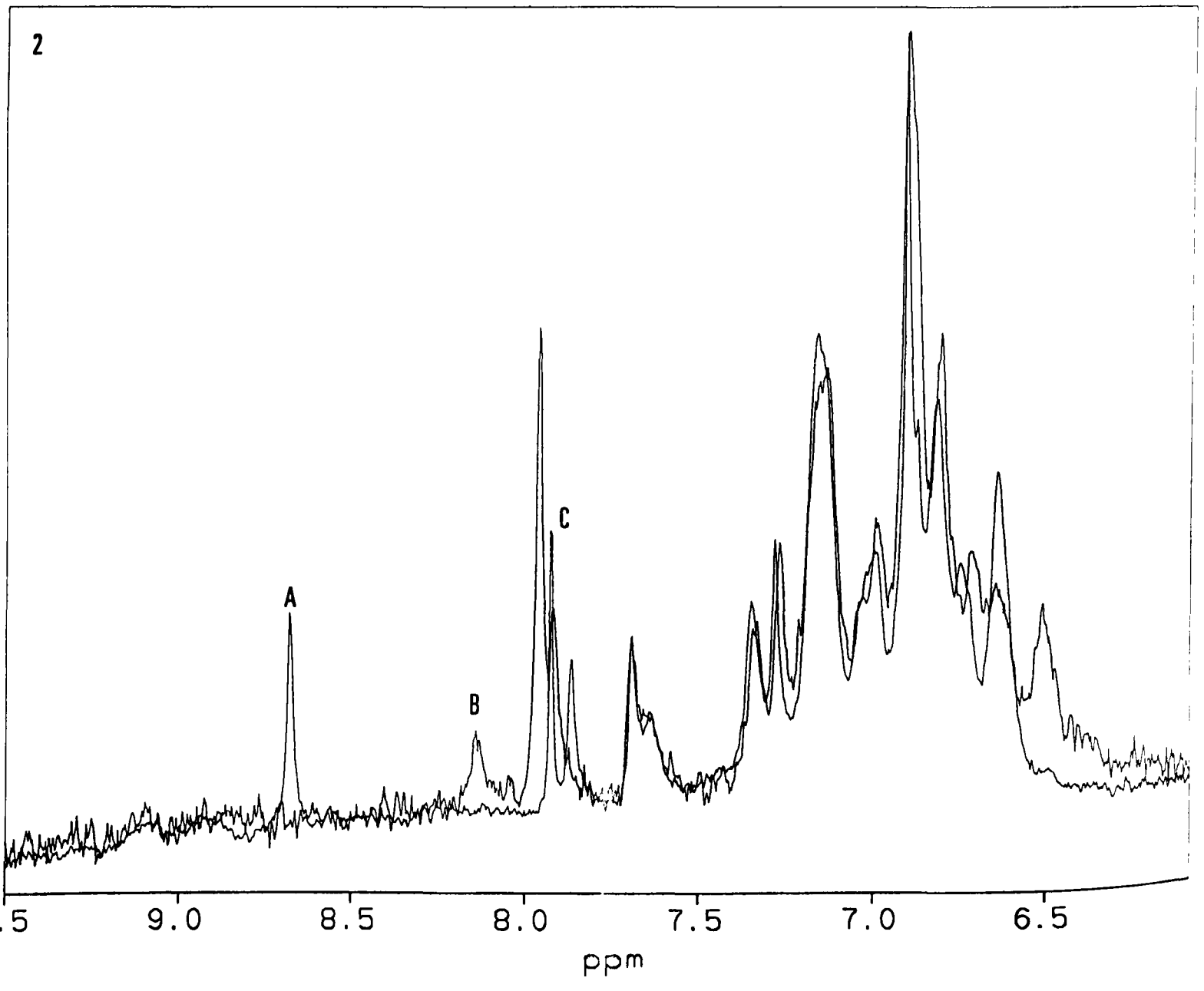

2
A
B
C
9.5
9.0
8.5
8.0
7.5
7.0
6.5
ppm

Figure 20: (Opposite page.) NMR spectra for 1) G2H5 Fv and 2) G2H5-HCEC Fv, with zinc concentrations at zero and excess.

contains three, the additional histidine being the one engineered into the light chain framework at position 2. We can therefore conclude from these spectra, that the 2H proton resonance of His-2 overlaps with those from His-61 and His-98.

Examining the series of spectra from the titration of each Fv with zinc, it can be seen that in each case, as the concentration of Zn^{2+} increases, the resonance at 8.65ppm (peak A) broadens and eventually disappears, and the relative intensity of the 'doublet' at 7.9ppm (peak C) is changed. However, for the G2H5-HCEC Fv, a new peak appears in the spectrum at $\approx$ 8.1ppm in the presence of excess zinc. This peak is not seen in the corresponding spectrum of the G2H5 Fv.

5.8.2.3 Conclusions

From the spectra shown in Figure 20, we can see that the 1H NMR spectra of both Fv's show changes on the addition of metal to the samples, indicating that in each case there is some interaction between the protein and zinc cations. The observation that peak A is broadened, rather than shifted on the addition of zinc, suggests that the metal-bound and metal-free protein forms are in intermediate exchange on the NMR timescale (Cheetham, 1986).

The appearance of a new peak at $\approx$ 8.1ppm for the G2H5-HCEC Fv is very interesting. Although the data presented here are only preliminary, they would suggest that in the case of this Fv, there is a second binding site for zinc in which the metal is in fast/slow exchange, the site being occupied only after saturation of the first site. Furthermore, they also suggest that one of the liganding groups at the

second site is a histidine ring, which would be consistent with the structure of the metal binding site engineered into G2H5-HCEC. Further studies are now needed to characterise this 'second' site and to obtain sequence specific assignments for the individual histidine residues of the native and engineered metal binding sites in G2H5-HCEC.

5.9 Discussion

The results of the binding assay and immunoprecipitation indicated that the G2H5-L1CEC antibody had an affinity for HEL which was similar to that of native HyHEL-5. This suggested that not only had the binding specificity of HyHEL-5 been successfully transferred onto the Gloop 2 framework, but also that the mutations in L1 (and the N-terminus) had not significantly altered the conformation of this CDR or disrupted its interactions with the rest of the combining site. Three of the mutations, Ala-25→Cys and Val-29→Cys in L1, and light chain N-terminal Ile-2→His, involve structure-determining residues (SDRs). The fact that the alteration of two L1 SDRs in G2H5-L1CEC does not disrupt HEL binding, suggests that gross (HyHEL-5) combining site topology was maintained.

Expression of G2H5-L1CEC and G2H5-HCEC Fv's has been carried out in E. *coli* by Dr. Staunton, and further binding studies confirmed these results. In addition, metal binding studies, also by Dr. Staunton, have been performed. G2H5-L1CEC showed no metal binding potential. The specific interaction of light chain N-terminal Ile-2 with L1 may have precluded metal binding by sterically blocking

the site. Alternatively, metal binding may not have taken place due to weak liganding interactions. Although sufficient ligands are present, Glu-27 may not have been able to ligate the metal ion either directly or *via* a bridging water molecule. This would leave only two effective ligands (Cys-25 and Cys-29) which may not have been sufficiently strong to bind the metal ion.

Evidence for both these possibilities comes from the metal binding studies of G2H5-HCEC. This has been shown to bind the transition metals cobalt, nickel, copper, zinc and cadmium (Staunton, 1992), in addition to HEL. His-2 may interact with L1 in a similar way to the native Ile-2, but it can also act as a ligand and therefore does not preclude metal binding. In addition, if Glu-27 is unable to act as a ligand, its loss is compensated by His-2. Metal and antigen binding may also be mutually dependent, the metal ion helping to create the correct L1 conformation and the antigen helping the metal binding site to form. NMR studies have allowed metal binding by the G2H5-HCEC Fv to be observed, but they were unable to prove conclusively that the L1 metal site was involved. Additional studies either by NMR or crystallography will be required to fully detail the structural basis of metal binding.

Finally, it is interesting to note that the characterisation of metalloantibody function carried out so far supports the predictions of the modelling studies, particularly for the site including the His-2 ligand. The main predictions were as follows:

1. Residues 24–29 in HyHEL-5 L1 are not involved in antigen binding or interactions with other CDRs and are therefore available for mutation (Chapter 3, Section 3.2.1).

2. The conformation of L1 in the presence of the N-terminus is similar to that seen in the HyHEL-5 crystal structure after conformational search, despite the presence of four mutations (three of which involved SDRs) (Chapter 4, Section 4.2). This and the previous prediction are supported by the retention of antigen affinity/specificity equivalent to the parent antibody.

3. The L1 site could be readily created with the His-2 ligand present and the stability of the site increased when a metal ion was introduced, according to energy calculations (Chapter 4, Section 4.3.1). This suggested co-operative metal binding; a finding reinforced by the observation that G2H5-HCEC has the ability to bind a range of transition metals.

These general observations, and the production of the functional metalloantibody described in this chapter, are evidence for the success of the modelling strategy which forms the basis of this thesis. Although specifically developed for antibodies, this strategy could allow metal binding sites to be readily designed in other proteins, enabling inclusion of the bioinorganic properties of metal ions into a wider range of protein functions.

Chapter 6

Metal Binding Site Prediction and Design

The structure and function of a protein is dependent on its amino acid sequence and, in some instances, on any interacting cofactors. Many proteins use metal ions as cofactors and the properties of protein and cofactor are interdependent. Williams (1985) proposes that the environment of a metal ion in a protein is frequently different from that found in an aqueous medium. This altered environment can confer unusual properties on the metal which, as a consequence, may be better fitted for a structural or catalytic rôle in the protein. The site of interaction within a protein for its metal cofactor is therefore of significant interest in understanding the function of metalloproteins and the mechanisms by which particular metal ions are selected (Snyder *et al.*, 1990) (see Chapter 2).

The interactions of proteins with coordinated metal ions have been characterised by a number of approaches, from the study of the primary sequence distribution of liganding residues (Vallee and Auld, 1989; Vallee and Auld, 1990), to the three-dimensional structural analysis of metal binding sites by X-ray crystallography (Chakrabarti, 1989; Chakrabarti, 1990; Yamashita *et al.*, 1990). From such studies, together with the analysis of 27 metalloprotein structures from the

Brookhaven Protein Databank (Bernstein *et al.*, 1977) (described in Chapter 2), two common features of metal-protein interactions were observed.

1. In general, Ca(II) and Mg(II), usually classified as hard metal cations, are bound by oxygen-containing groups; the predominant liganding sidechains for these metals are Asp and Glu; the sidechains of Asn, Gln, Ser and Thr are also found, but to a lesser extent.

2. Mn(II), Fe(II), Cu(II) and Zn(II), usually classified as borderline/soft metal cations, are generally bound by nitrogen and sulphur-containing groups where the predominant liganding sidechains are His and Cys, although the sidechains of Asp, Glu and Met are also represented.

Considering only Asp, Glu, Cys and His (Higaki *et al.*, 1992), a variety of metal binding sites with different selectivities and specificities can potentially be designed in proteins.

Based on these observations, the program CONRAD (written in FORTRAN) has been developed, capable of locating metal binding sites in known metalloprotein structures and engineered metal binding proteins (program documentation is given in Appendix 3). It has been used to design novel metal binding sites in two antibody structures.

6.1 Prediction and Design

The design of a novel metal binding site for the chosen cation, involves four separate stages:

1. The protein structure is searched with a geometric template for the chosen metal ion, using a pattern matching algorithm to find putative metal sites;

2. Ligands required for the chosen metal ion are built into the protein structure manually and given suitable coordinating conformations;

3. The modelled structure is then searched using a hydrophobicity contrast function similar to that of Yamashita *et al.*, (1990), to confirm that the modelled site satisfies the criteria for metal binding. Maximum function values are generally centred at or near a metal binding site and are shown as a cluster of points;

4. The mid-point of the cluster is selected and its position is refined using multi-dimensional minimisation (Press *et al.*, 1986). Interactive molecular graphics is used to refine the positions of individual ligands.

6.1.1 Template Matching

This procedure searches a protein structure to identify groups of 3 Cα and 3 Cβ atoms whose relative spatial arrangements, defined by their inter-atomic distances, match those of template sets of atoms, as evaluated by the equation:

$$R = \sum_{i=1}^{3} \mid d_i - d_{t_i} \mid \qquad (2)$$

where d_i is the distance between a pair of atoms in the protein and d_{t_i} is the distance between the equivalent pair in the template.

While a bidentate chelating site can provide sufficient energy to bind a metal ion, this process is reversible and no such sites were observed in any of the metalloprotein structures examined (Appendix 1). To reduce the complexity of designing

tetradentate chelating sites, templates of three atoms were chosen since this is the minimum number of ligands required for tight binding of *d*-block transition metal ions (Williams, 1987). The template positions identified could also be used to create bidentate chelating sites. Reversible metal binding at these sites has be used as a method of regulating enzyme activity (Higaki *et al.*, 1990; Higaki *et al.*, 1992).

Templates derived from metalloprotein structures avoid the errors which can be introduced by the use of small molecule parameters applied to proteins. Although the templates derived from copper binding sites (Table 25) represent a highly unusual distorted trigonal planar geometry, by its very nature it is likely to be highly specific for the redox functions of this cation. The structures from which the templates were derived are shown in Table 25.

For a set of three template atoms A, B, and C, each pair of distances A→B, A→C and B→C, are calculated. Corresponding distances between the equivalent atoms in the protein structure are calculated and atom numbers of pairs which match the template distances within a user-defined tolerance are stored. The tolerance defines the sensitivity of the search and therefore the number of atom pairs found.

The stored atoms are then sorted to remove identical pairs *e.g.* if atoms 2 and 10 have a separation equal to the A→B distance, then so will the identical pair 10 and 2. Finally, the remaining atom pairs are examined and groups equivalent to the search template identified.

Gly, Pro and disulphide cysteines may be ignored in the procedure, since these residues are often important in defining mainchain conformation and are thus often unsuitable candidates for mutation (Chothia *et al.*, 1989; Ramachandran *et al.*, 1963). However, the user may include these residues if so desired.

Metal Ion	Structures	Cα Templates				Cβ Templates		
		1–2	1–3	2–3		1–2	1–3	2–3
Calcium†	3CLN-1	4.29	6.50	7.00		4.10	6.44	5.49
	3CLN-2	4.47	6.56	7.20		4.14	6.47	5.71
	3CLN-3	4.58	6.62	6.97		4.33	6.47	5.51
	3CLN-4	7.09	6.32	11.04		6.99	6.08	9.40
	3ICB-1	2.96	7.10	6.87		2.96	5.60	5.39
	3ICB-2	4.33	6.80	7.07		4.04	6.65	5.56
	5CPV-1	4.34	6.69	7.06		4.04	6.48	5.60
	5CPV-2	4.33	6.54	7.11		3.94	6.36	5.65
	5TNC-1	4.27	6.91	6.85		4.04	6.62	5.37
	5TNC-2	4.47	6.47	6.77		3.97	6.10	5.34
Copper	1AZU	5.48	8.48	6.11		4.83	7.34	4.92
	1PAZ	5.87	7.60	5.39		5.11	6.87	4.27
	2AZA	5.92	8.23	5.69		4.82	6.76	4.02
	5PCY	6.11	7.76	5.44		5.26	7.10	4.37
Zinc	1TON	10.49	8.76	6.97		9.68	8.08	7.97
	2CAB	7.04	5.47	7.98		7.50	5.58	7.08
	3TLN	6.62	8.26	8.03		7.37	8.52	8.87
	5CPA	5.94	7.21	6.48		4.69	6.28	5.75
	7ADH	7.91	7.35	7.93		6.60	4.69	6.51

Table 25: Cα and Cβ templates used to search proteins, to identify groups of residues in the correct spatial arrangement to form potential metal ligands. Also shown are the structures from which the templates were derived. The columns headed 1–2, 1–3 and 2–3 represent the distances in Å between atom pairs in the templates. †Derived from EF-hand sequences (da Silva and Reinach, 1991).

6.1.2 Hydrophobicity Contrast

The concept of hydrophobicity contrast in a metal-ligand complex, was first described by Yamashita *et al.* (1990). From an examination of ionophores and metal-host molecules from the Cambridge Structural Database (Allen *et al.*, 1979) and metal binding sites in proteins, it was found that metal ions are bound at the centre of a shell of hydrophilic atoms (the liganding atoms) surrounded by an outer shell of hydrophobic carbon atoms (the sidechain carbons). This feature can be defined quantitatively by the expression:

$$C = \sum_{i=1}^{N_R} (\Delta\sigma_i \times r_i^2) - N_R < \Delta\sigma >< r^2 > \qquad (3)$$

where N_R is the number of atoms within a sphere of radius R, r_i is the distance of the i^{th} atom from the centre of a sphere and $\Delta\sigma$ is the Atomic Solvation Parameter (ASP)—essentially an atomic-level hydrophobicity term (Eisenberg and McLachlan, 1986).

In the implementation described here, the ASP was replaced with values derived from partial charges (Weiner *et al.*, 1984). It was considered that these gave a more intuitive representation of atomic-level hydrophobicity for the atoms in each amino acid, than the parameters described by Eisenberg and McLachlan. The central requirement of Yamashita *et al's* characterisation of metal binding sites (the concentric shells of hydrophilic and hydrophobic atoms) is reproduced using partial charges, as shown in Figure 21.

The square of the partial charge is used to represent the difference in energy between an atom in aqueous and non-aqueous environments. Thus, the greater the charge on an atom, the less hydrophobic it is. An increase in hydrophobicity with

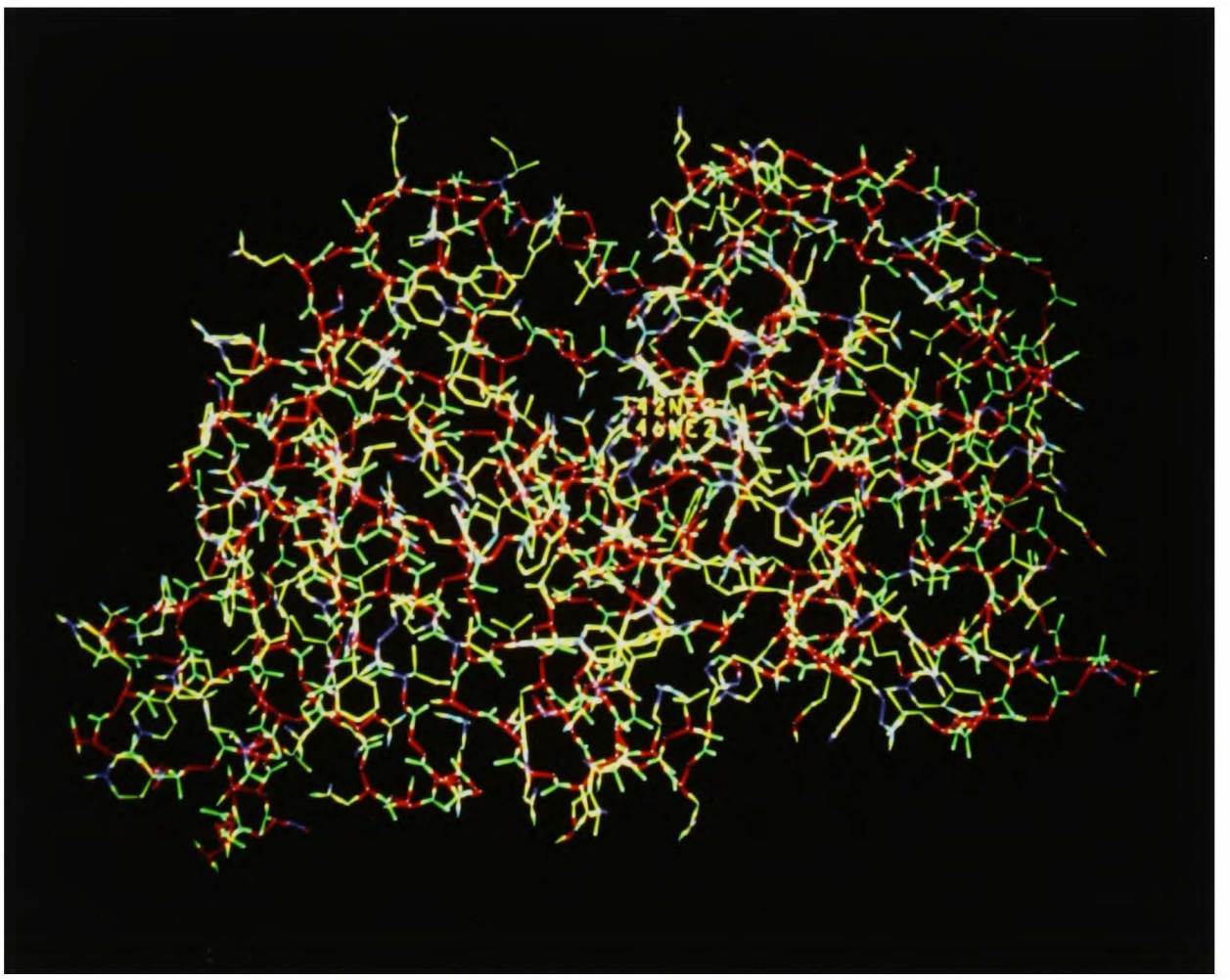


Figure 21: Thermolysin coordinates coloured according to partial charge. Blue = positive; red = negative; yellow = hydrophobic. 142 Nε2 and 146 Nε2 show the position of the metal site His ligands.

distance from a metal centre is shown by a decrease in the positive value of C (as calculated from Equation 3). Values for C are calculated at 23 concentric sphere radii between 0.5Å and 11.5Å. By analysing known metal binding sites in this way, curves such as that shown in Figure 22, are obtained. A sharp fall in the value of C is seen at a sphere radius around 6.0Å.

The hydrophobicity contrast function C_H, is defined here as the difference in slope between one pair of points with radius 3.5Å and 5.5Å and a second pair of points with radius 7.5Å and 9.5Å. Fourier transform digital filtering (Bevington, 1969) is performed on the 23 values to remove noise before the contrast is calculated. Noise results largely from the peptide unit which has a high intrinsic hydrophilicity (Furet *et al.*, 1988) compared with the sidechain carbons. Since C_H is evaluated from four discrete points of Equation 3, it is important that these are not affected by noise.

Initial studies using this function to locate known metal binding sites, showed that it was too sensitive to regions of high contrast which were not associated with metal binding. The major causes of this effect were charged surface residues, buried charged residues and disulphide bonds. Clearly, charged sidechains, hydrogen bonded hydrophilic sidechains and cysteines involved in disulphide bridges, while still hydrophilic in nature, would not be available to ligand a metal. A pre-screening stage was therefore introduced to eliminate these regions, so that hydrophobicity contrast was calculated only at putative metal binding sites. This is achieved as follows:

1. An orthonormal solvent grid is generated around the protein. The size of the grid is determined by the dimensions of the protein envelope plus a user

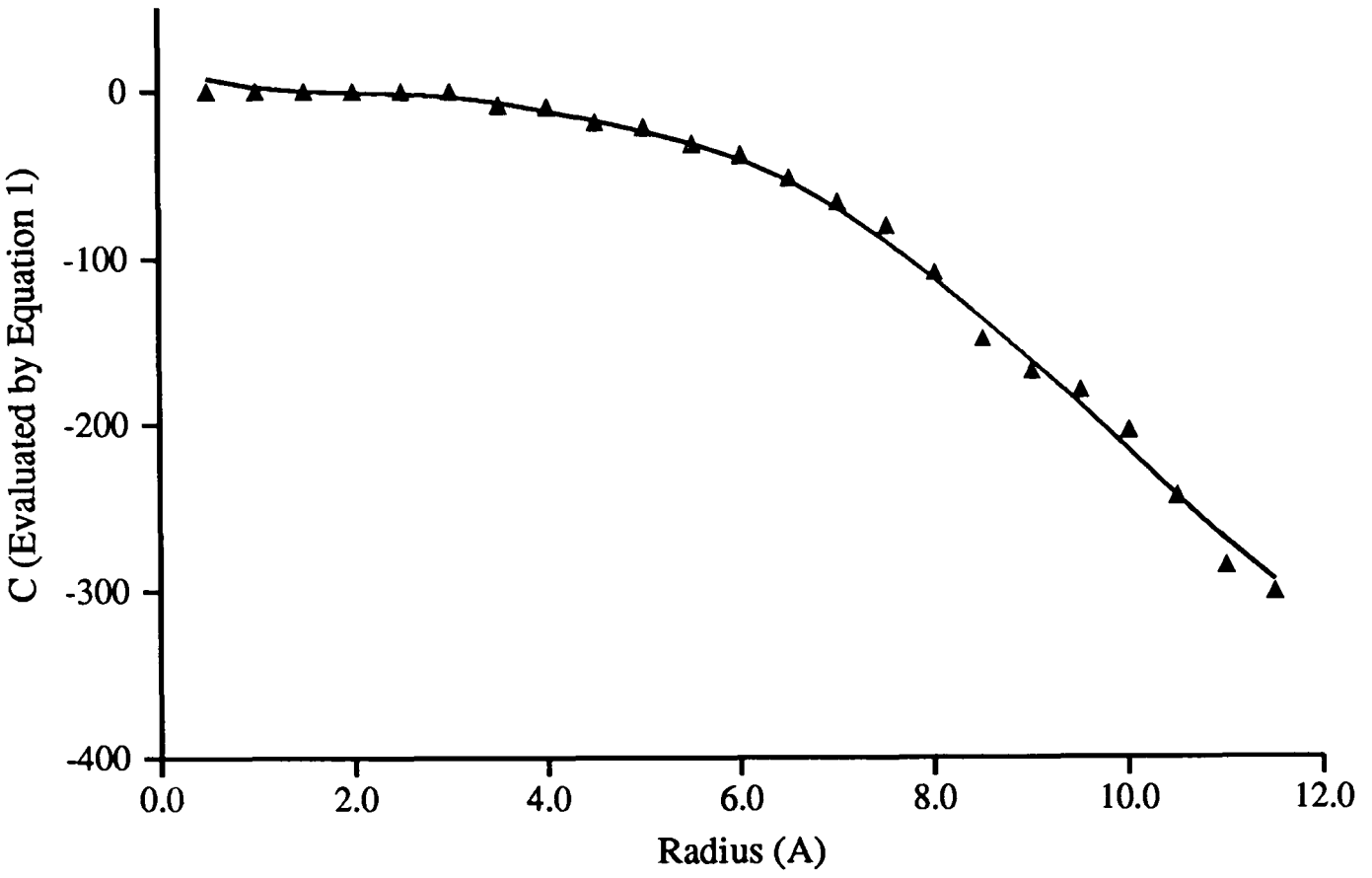


Figure 22: Values of C as calculated by Equation 3, at the Iron site in Rubredoxin, are shown by filled triangles. The data after Fourier smoothing are shown by the curve. The units of C are Coulomb^{2}Å^2.

definable solvent grid size. Since C is evaluated at radii up to 11.5Å, a grid of this size should be used. However, this creates a solvent shell which is large compared with the size of the protein, so grids of decreasing size were tested. A 1Å grid was found to give good results without incurring excessive computing time penalties. Solvent points are defined by searching from the limits of the grid along the X-, Y- and Z-axes in both directions until the protein envelope is encountered. Each solvent point has a charge applied to it, thus reducing the hydrophobicity contrast at surface charges. The optimal value for the charge was found to be -0.3 Coulombs.

2. A cavity search is performed at every point on a separate 1Å grid spanning

the protein structure; a grid point is determined to lie in a cavity if there are no atom or solvent point neighbours within 1.4Å. Having defined a cavity point, the shell of atoms within 4.0Å of the point is then analysed. If liganding atoms from Asp, Glu, His or Cys are present, the coordinates of the grid point are stored, but if atoms from Arg, Lys or disulphide Cysteines are found, the point is ignored.

3. For each stored cavity point, the hydrophobicity contrast function C_{H} is evaluated. When a histidine sidechain, which can exist in three protonation states, is encountered, the correctly protonated form must be selected. The closer ring nitrogen to the grid point is determined and charge values are assigned to the histidine from the species in which that nitrogen is unprotonated (thus having a lone pair available for liganding).

4. The values of C_{H} are sorted (Knuth, 1968) and their coordinates written to a file in Protein Databank format. Using interactive molecular graphics, these data can be superimposed upon the protein coordinates either as points or contour maps.

6.1.3 Multidimensional Minimisation

Within the modelled site, an initial position for the metal ion is chosen by calculating the mid-point of the cluster within the metal site. This is then refined with-respect-to ligand geometry using a multidimensional minimisation algorithm which implements the downhill simplex method of Nelder and Mead (1965). A set of mean metal–ligand distances (d_{0_i}) for each ligand type were derived from an

Ligand	Ca	Fe	Cu	Zn	Mn	Mg
Carbonyl	2.3	—	2.5	—	—	—
Asp-Oδ	2.4	2.1	—	1.9	2.2	2.1
Glu-Oϵ	2.5	2.2	—	2.2	2.2	2.1
His-Nδ,Nϵ	—	2.2	2.2	2.1	2.1	—
Cys-Sγ	—	2.3	2.1	2.2	—	—
Met-Sδ	—	—	2.9	—	—	—
Asn-Oδ	2.4	—	—	—	—	—
Gln-Oϵ	2.4	—	—	—	—	—
Hydroxyl	2.5	—	—	—	—	—

Table 26: Mean metal-ligand distances derived from an analysis of the 27 metalloprotein structures in Appendix 1, as described in Chapter 2, Section 2.1.

analysis of the metalloprotein structures in Appendix 1 (Table 26) (see Chapter 2). Although data from small molecules may give more accurate distances for metal-ligand interactions, the use of distances derived from protein structures will give results that are more typical of real metalloproteins. The program minimises the expression:

$$S = \sum_{i=1}^{N_{\mathrm{L}}} \mid d_i - d_{0_i} \mid \qquad (4)$$

where N_{L} is the number of ligands and $\mid d_i - d_{0_i} \mid$ is the absolute difference between the distance of the liganding atom i and its mean. Multidimensional minimisation is sensitive to local minima, so a multi-start minimisation is used in an attempt to locate the global minimum in Equation 4. For 25 of the metalloprotein structures shown in Table 27, the algorithm typically is able to move the mid-point of the cluster to within 0.6Å of the actual metal position.

PDB	Metal ion	Points examined	Cavity points	Points in site	Mid-point HCF	Deviation		CPU time
						CONRAD	SIMPLEX	
1ALC	Ca^{2+}	87 200	1	1	-0.0113	3.12	0.13	22
1AZU	Cu^{2+}	62 154	34	29	-0.2312	0.44	0.30	15
1HMQ	$1Fe^{2+}$	64 551	191	145			0.16	
	$2Fe^{2+}$			107	0.1161	0.24	0.13	22
1PAZ	Cu^{2+}	59 024	31	31	-0.2546	0.13	0.18	13
1TEC	$1Ca^{2+}$	109 021	46	0				
	$2Ca^{2+}$			14	-0.0132	2.58	0.36	51
1TON	Zn^{2+}	126 036	55	34	0.2198	1.71	0.01	56
1TPO	Ca^{2+}	99 310	21	21	-0.0500	1.73	0.44	39
2AZA	Cu^{2+}	58 787	38	25	-0.5243	0.25	0.29	14
2CAB	Zn^{2+}	109 475	134	67	-0.1656	1.01	1.71	55
2GBP	Ca^{2+}	152 695	15	14	-0.0346	0.63	0.12	83
2GLS	$1Mn^{2+}$	380 642	129	27			0.39	
	$2Mn^{2+}$			32	0.1618	1.27	0.60	389
2LTN	$1Mn^{2+}$	123 129	96	82			0.33	
	$2Ca^{2+}$			28	-0.0668	1.26	0.15	83
2SOD	Cu^{2+}	87 438	160	49			0.28	
	Zn^{2+}			82	-0.6502	1.03	0.13	32
3CLN	$1Ca^{2+}$	154 857	59	17	-0.0090	1.58	0.19	
	$2Ca^{2+}$			15	-0.0173	2.07	0.24	
	$3Ca^{2+}$			6	-0.0144	1.04	0.34	
	$4Ca^{2+}$			20	-0.0114	0.45	0.42	56
3CNA	Mn^{2+}	126 233	68	46			0.35	
	Ca^{2+}			35	-0.0483	0.56	0.41	58
3ICB	$1Ca^{2+}$	39 048	8	0				
	$2Ca^{2+}$			8	-0.0133	0.63	0.13	6
3TLN	$1Ca^{2+}$	176 853	140	60			0.13	
	$2Ca^{2+}$			35	-0.0170	1.31	0.07	
	$3Ca^{2+}$			0				
	$4Ca^{2+}$			0				
	Zn^{2+}			49	-0.2158	1.31	0.11	119
5CPA	Zn^{2+}	157 032	37	36	-0.5385	1.09	0.25	91
5CPV	$1Ca^{2+}$	44 116	74	33	-0.0113	0.68	0.18	
	$2Ca^{2+}$			33	-0.0224	1.20	0.17	10
5PCY	Cu^{2+}	42 451	26	26	-0.2027	0.28	0.23	8
5RXN	Fe^{2+}	24 857	52	49	-0.0252	0.43	0.05	3
5TNC	$1Ca^{2+}$	151 371	31	12	-0.0221	1.05	0.18	
	$2Ca^{2+}$			1	-0.0123	2.73	0.11	57
5XIA	Mg^{2+}	293 588	147	70	-0.0228	0.98	0.14	245
6AT1	Zn^{2+}	107 218	66	63	-0.0521	0.68	0.08	37
7ADH	$1Zn^{2+}$	242 863	115	67	-0.2622	0.98	0.22	
	$2Zn^{2+}$			40	-0.0264	1.46	0.31	193
G2Zn	M^{2+}	196 709	27	12	-0.0237			95
HCEC	M^{2+}	176 912	51	48	-1.4906			86
8692	Ca^{2+}	68 685	40	40	-0.0610			18

Table 27: See legend opposite for description.

Results from testing CONRAD on 25 metalloproteins from the Brookhaven Protein Databank (PDB) and 3 modelled metal binding proteins (G2Zn, HCEC and 8692).

PDB:-
1ALC:α-lactalbumin,
1AZU:Azurin,
1HMQ:Hemerythrin,
1PAZ:Pseudoazurin,
1TEC:Thermitase,
1TON:Tonin,
1TPO:β-Trypsin,
2AZA:Azurin,
2CAB:Carbonic Anhydrase B,
2GBP:Galactose Binding Protein,
2GLS:Glutamine Synthetase,
2LTN:Pea Lectin,
2SOD:Cu-Zn Superoxide Dismutase,
3CLN:Calmodulin,
3CNA:Concanavalin A,
3ICB:Calbindin,
3TLN:Thermolysin,
5CPA:Carboxypetidase A,
5CPV:Parvalbumin,
5PCY:Plastocyanin,
5RXN:Rubredoxin,
5TNC:Troponin C,
5XIA:Xylose Isomerase,
6AT1:Aspartate Carbamoyltransferase,
7ADH:Alcohol Dehydrogenase.

Points examined: Total number of grid points spanning the protein and solvent.
Cavity points: Number of points identified as being in cavities and at which the HCF was calculated.
Points in site: Number of points within 4.0Å of the actual metal position.
Mid-point HCF: Hydrophobicity contrast of the cluster mid-point within the site, evaluated as described in the text.
Deviation: Distance in Ångströms of the mid-point from the actual metal position; *CONRAD*: After calculation of the HCF; *SIMPLEX*: After simplex minimisation.
CPU time: Computer time in minutes on a VAXstation 3200.

6.2 Analysis of Known Metalloproteins

Table 27 presents the results of applying the method as described, to 25 metalloprotein and 3 engineered metal binding protein structures (Section 6.3). The known metalloproteins contained 42 metal binding sites of which the program found 38, representing a success rate of 90%. Although some points of high hydrophobicity contrast clustered outside the metal binding sites, the major cluster was always around a known metal site. The examples in which sites were not found, were from proteins containing multiple sites (1TEC, 3ICB and 3TLN). Of the twelve multi-site metalloproteins, the program successfully located all the sites in nine and at least one of the sites in the remaining three. The initial average deviation between the mid-point of a cluster within a site and the actual metal position was 1.1Å, but after simplex minimisation this improved to 0.2Å.

For one of the proteins, 2CAB, simplex minimisation fails to place the cluster mid-point within 0.6Å of the actual metal position; the deviation in this case is 1.7Å. In 2CAB, the metal is bound by 3 protein ligands. Examination of the poorly positioned mid-point, using interactive molecular graphics, showed that it formed the fourth vertex of a tetrahedron which was the mirror image of the tetrahedron formed by the metal and the three ligands in the crystal structure. Since the minimisation protocol simply optimises a set of distances from the metal to the ligands, it is unable to distinguish between the two enantiomeric cases where the residual calculated by Equation 4 will be equal.

In cases where only three ligands are provided by the protein and the metal is not coordinated in a planar arrangement, a fourth ligand is generally provided by

a water molecule. The crystal structure of 2CAB (Kannan *et al.*, 1984) does not include any waters bound to the metal ion; however, placing a water at a location similar to that found at the copper ion in 2SOD (Tainer *et al.*, 1982) and re-running, the simplex minimisation correctly positioned the mid-point to within 0.3Å of the actual metal coordinates. This was the only example where an additional ligand was required to decide between two equivalent positions. The 36 other examples given in Table 27, show that simplex minimisation is normally able to reposition the mid-point accurately using only the ligands given.

The program was unable to locate four calcium binding sites; one in 1TEC, one in 3ICB and two in 3TLN. An analysis of these sites has suggested an explanation for this discrepancy. The contribution of carbonyl groups to hydrophobicity contrast is small compared with that of carboxyl sidechains, the other main ligand in these sites. The intrinsic hydrophilicity of the peptide unit (including the carbonyl carbon) as mentioned earlier, reduces the magnitude of the contrast at these sites.

In addition, hydrophobicity contrast is further reduced by the accessibility of these sites to solvent as generated by the solvent grid. This can be seen when C_H is calculated in its absence. In 3ICB (Szebenyi and Moffat, 1986), 12 points are now found, one of which is situated in the calcium-1 site, while in 3TLN (Holmes and Matthews, 1982), five of the now 200 points found are located in the calcium-3 site (compare with the results in Table 27).

The only metal binding sites of the 42 examined which could not be located by the program, either with or without a solvent grid, was the calcium-1 site in 1TEC and the calcium-4 site in 3TLN. The coordinating groups at these sites included few of the principle liganding sidechains for calcium (Asp and Glu—see Chapter 2,

Section 2.1). Thus, as the method has been parameterised to consider only the most frequently used metal ligands (since they are the obvious starting point in any metalloprotein design experiment), these particular sites were not identified.

6.3 Modelling Novel Metal Binding Sites

Table 27 shows that the program was able to identify and optimise the metal positions in a range of known metalloproteins. However, to illustrate how it can be used to assist in the design of novel metal binding sites, the program was used to evaluate the metal binding potential of three model metalloprotein structures.

6.3.1 Modelling a Putative Zinc Binding Site in Gloop 2

A putative zinc binding site was modelled into the light chain of the anti-lysozyme antibody Gloop 2 (Jeffrey *et al.*, 1992). The template search was performed using all the zinc templates (Table 25) on the coordinates of the Gloop 2 P_1 crystal structure containing both light and heavy chains. A total of 346 putative metal site positions were identified. The position chosen, which was selected from the 346 by its location in the antibody being close to the combining site, was a $C\beta$ pattern composed of residues Ser-34, Ile-89 and Tyr-91 in the light chain. The spatial agreement of these residues with the zinc site $C\beta$ template from carboxypeptidase A (5CPA) was 0.39Å. These residues place the metal site at the base of the complementarity-determining regions (CDRs) L1 and L3, within the antibody combining site (ACS).

Considering the nature of the residues to be replaced, their interactions within the combining site and whether they were structurally determining (Chothia *et al.*,

1989), zinc ligands were modelled at these positions by making the substitutions Ser-34→His, Ile-89→Glu and Tyr-91→His using the REPLACE and REFI (Hermans and McQueen, 1974) procedures available within the molecular graphics program FRODO (Jones, 1985). The hydrophobicity contrast was evaluated for both native and modelled Gloop 2 (Table 27: G2Zn).

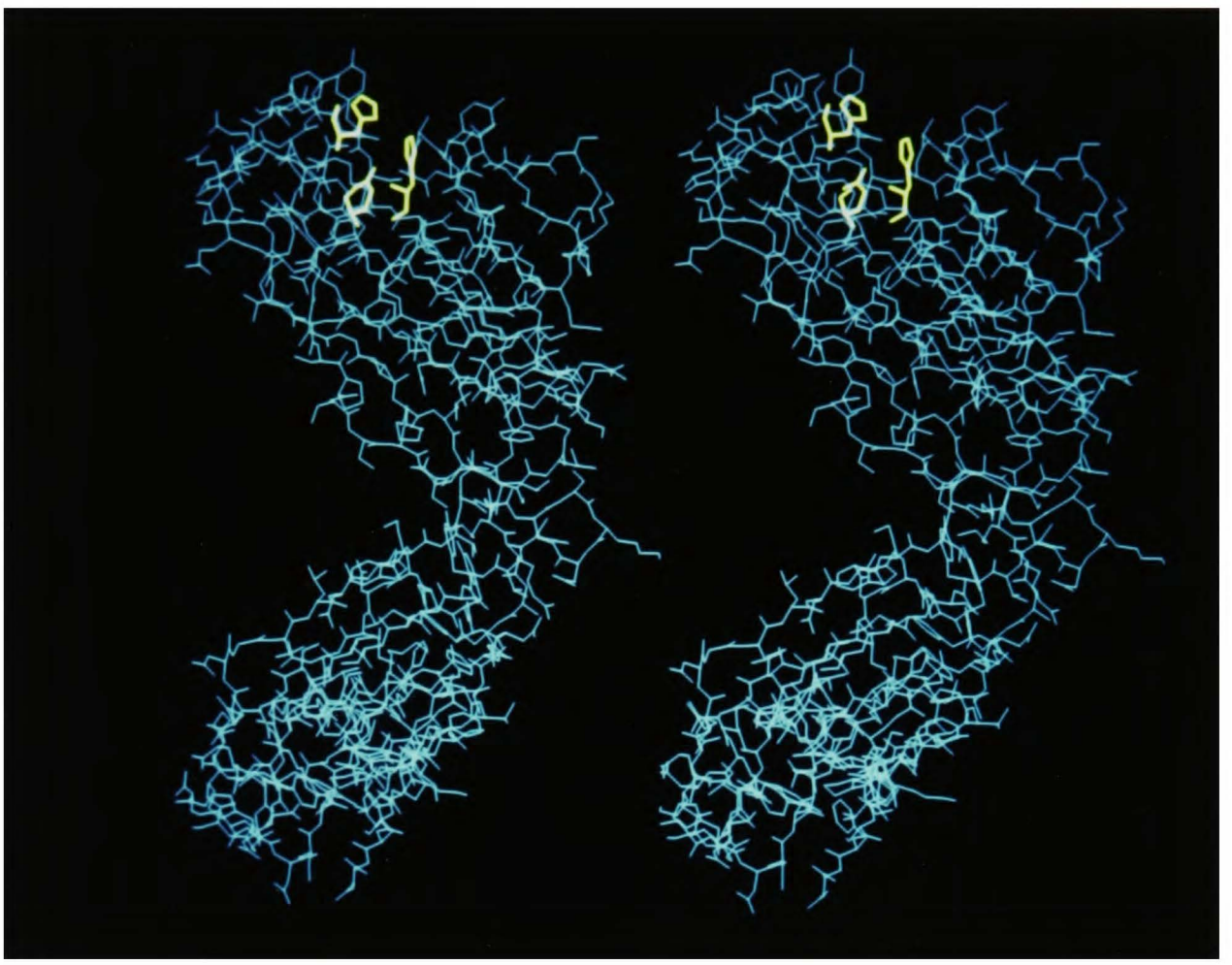
Native Gloop 2 showed 15 points of high contrast owing to the presence in the light chain, of a putative metal binding site composed of Asp-151, Glu-154 and His-189 located in the constant domain at the C-terminus of the Fab fragment. This site may be of significance owing to the presence of both Asp and His. It has been proposed (Higaki *et al.*, 1992) that Asp–His couples might naturally play a rôle in protein folding.

Nevertheless, an additional 12 points of high hydrophobicity contrast were found within the modelled metal site in the antibody (Figure 23). Although this site has not been tested for metal binding, its position is equivalent to the metal site created in the anti-fluorescein antibody 4-4-20 (Herron *et al.*, 1989) which has been shown to bind copper (Roberts *et al.*, 1990; Iverson *et al.*, 1990).

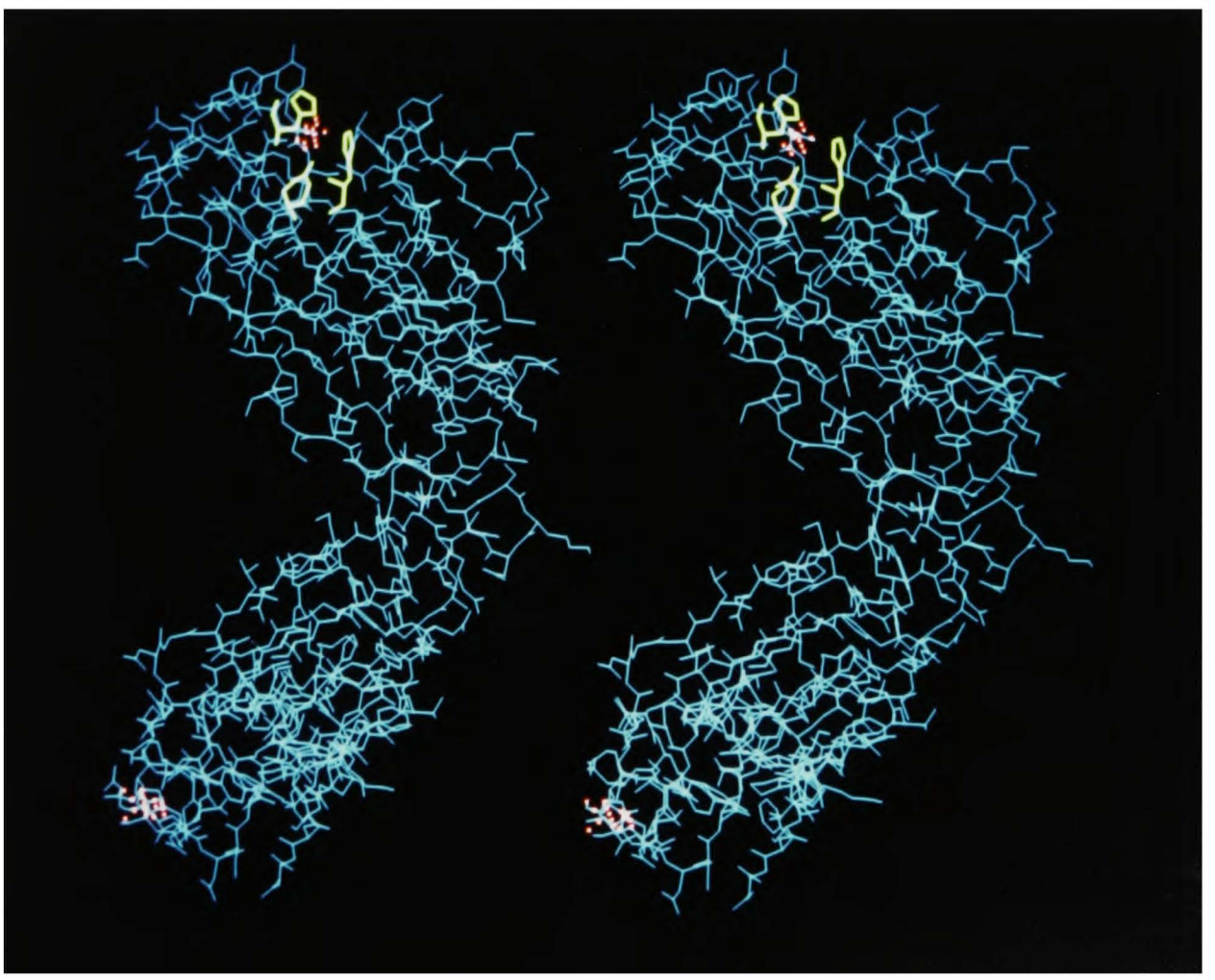
6.3.2 Modelling a Ca^{2+} Binding Site in Human Lysozyme

Human lysozyme (HL) is not a metal binding protein, but shows high structural homology with the protein α-lactalbumin, which binds calcium with an affinity of 4.0×10^7 M^{-1}. Kuroki *et al.*, (1989) showed that by replacing Gln-86 and Ala-92 in HL with aspartate residues to mimic the pattern of ligands in the calcium binding site of α-lactalbumin, a mutant HL with a calcium binding affinity of 5.0×10^6 M^{-1} could be generated. Since the native protein does not bind metal, while the mutant

Figure 23: Stereo-view of the modelled zinc binding site in antibody Gloop 2. a) Positions of the introduced zinc ligands.



b) Points of high hydrophobicity contrast. Note that in addition to the cluster within the new metal site, a second cluster of points was found indicating the presence of a putative metal site in the native protein.



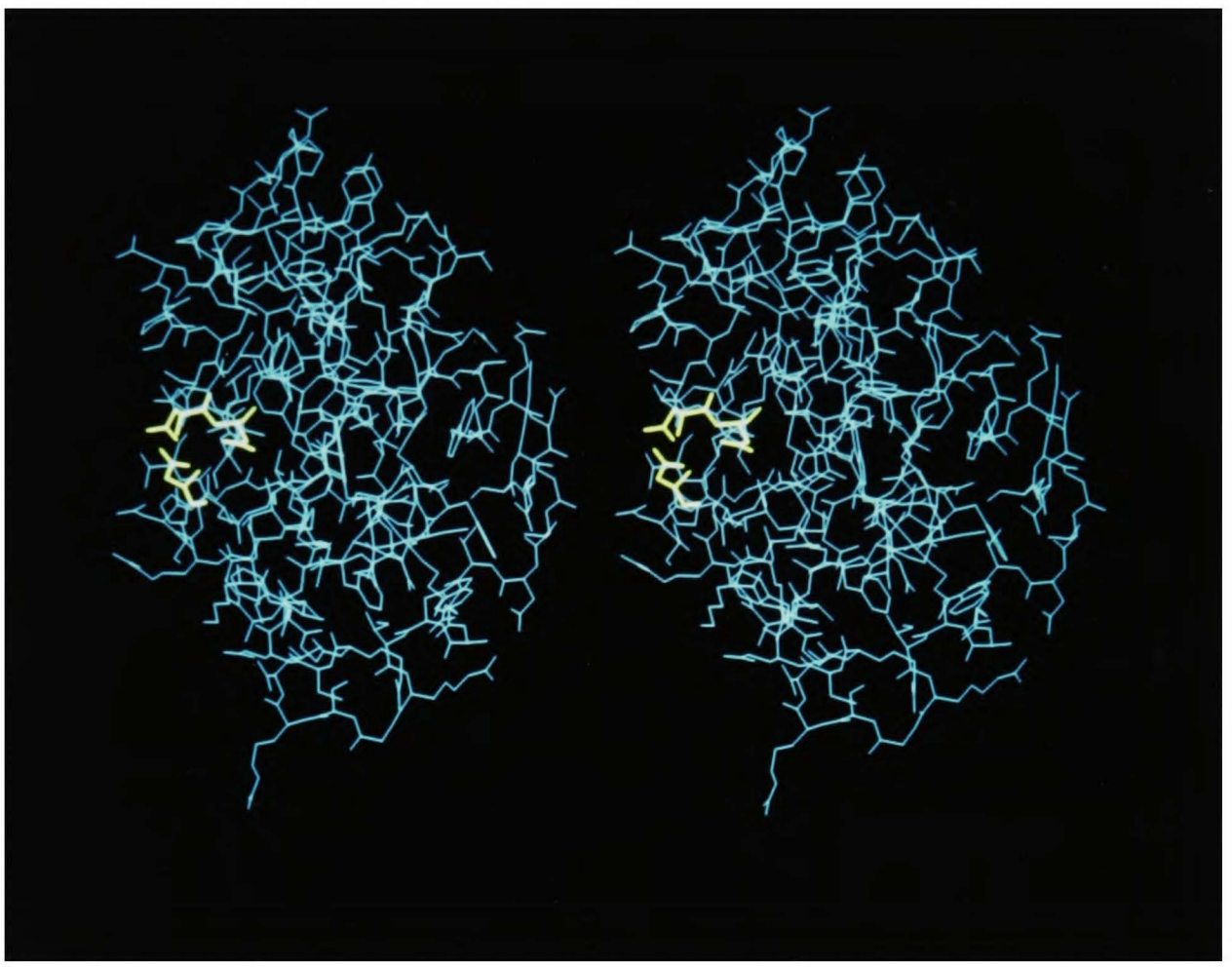
D86/D92 HL does, the system offered a good test of the predictive power and accuracy of the method as a tool in the *ab initio* design of metal binding sites.

From the refined 1.5Å crystal structure of human lysozyme (Artymuik and Blake, 1981), a model for the D86/D92 HL was generated by replacing the sidechains of Gln-86 and Ala-92 with Asp, using the REPLACE and REFI procedures in FRODO. Runs were then performed on both the native and mutant enzyme structures. In native HL, no regions of high hydrophobicity contrast were found, in good agreement with the experimental observation that the native protein does not bind calcium. For the D86/D92 lysozyme, 40 points of high contrast, localised within a 4.0Å sphere (Table 27: 8692) were found. The sphere coincided with the engineered metal binding site where Kuroki *et al.* had introduced their mutations (Figure 24). The method was thus able to discriminate correctly between the metal binding potential of the native and engineered protein.

6.3.3 Characterising the Zinc Binding Site in HyHEL-5

The program has been used to characterise the metal binding potential of a zinc binding site previously modelled (Chapter 4) into CDR-L1 of antibody HyHEL-5 (Sheriff *et al.*, 1987). Hydrophobicity contrast was evaluated for both native HyHEL-5 and the modelled metalloantibody. No points of high contrast were found for the native antibody, but 48 points of high hydrophobicity contrast, localised within a 4.0Å sphere around the metal binding site, were found in the modelled metalloantibody (Figure 25). The antibody has been created (Chapter 5) and expressed in E. *coli*. It has been shown to exhibit binding of its protein antigen, lysozyme, and the transition metals cobalt, nickel, copper, zinc and cadmium (Staunton, 1992).

Figure 24: Stereo-view of the modelled calcium binding site in D86/D92 lysozyme. a) Position of the created calcium binding site.



b) Points of high hydrophobicity contrast. All the points found cluster at the calcium binding site.

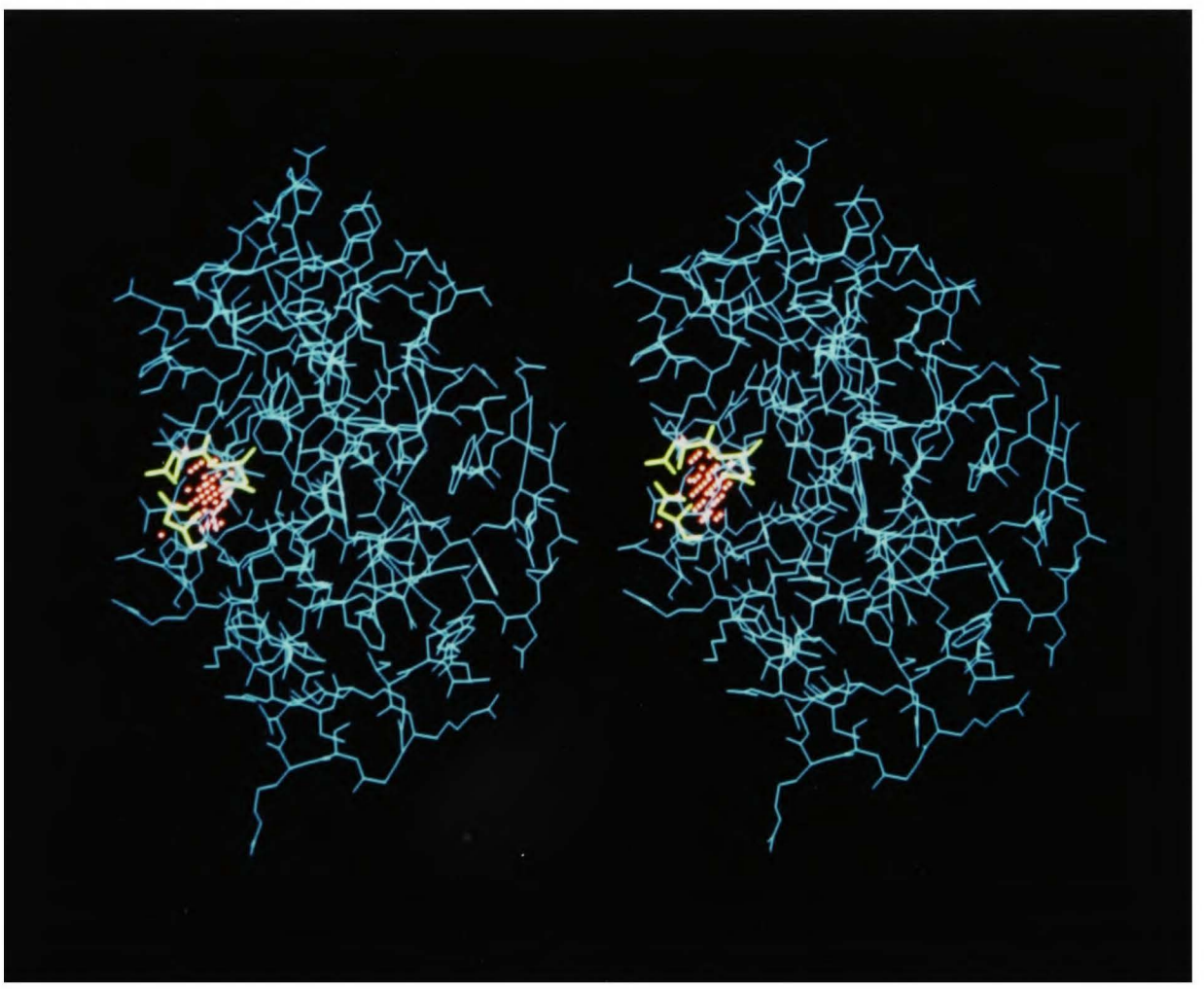
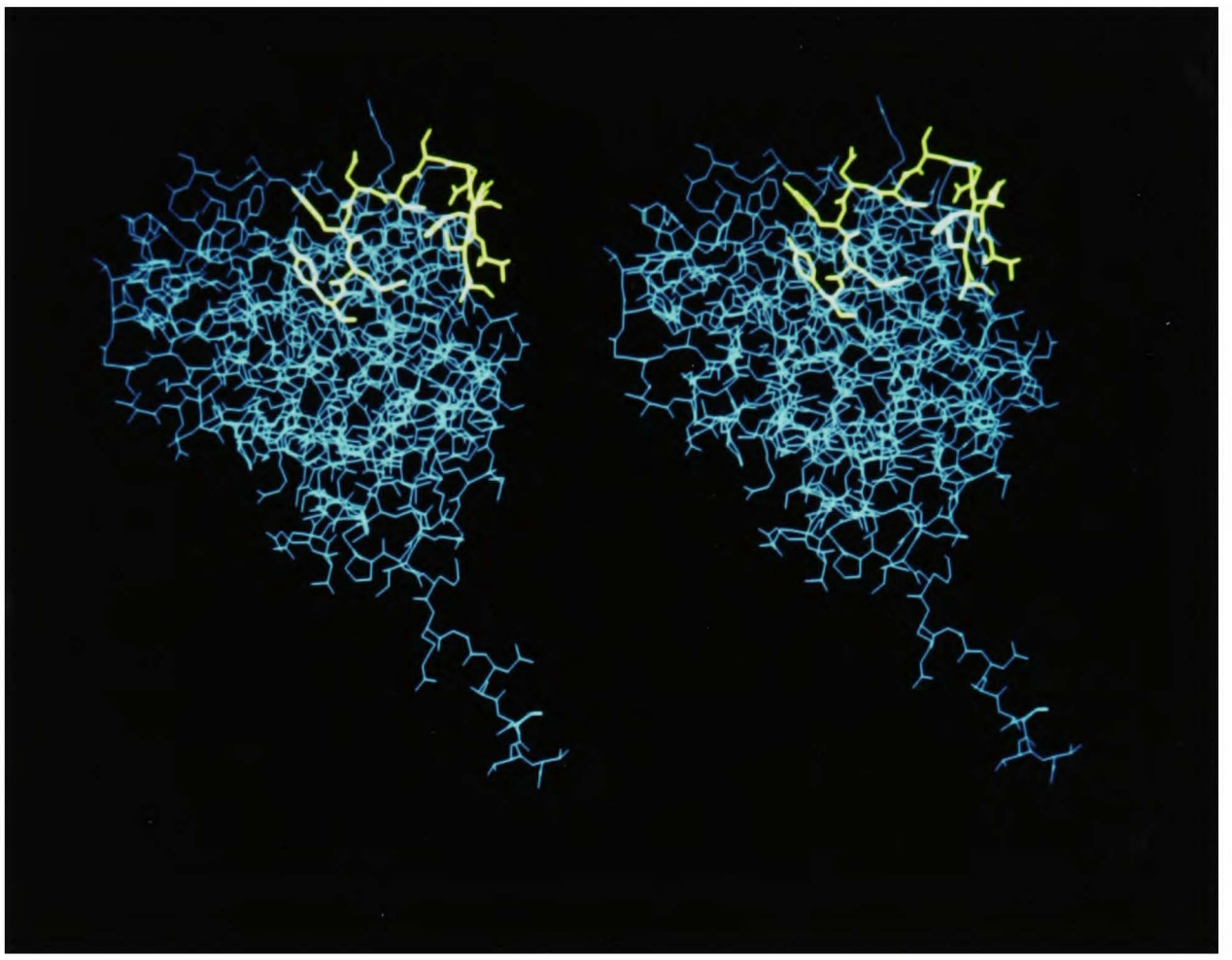
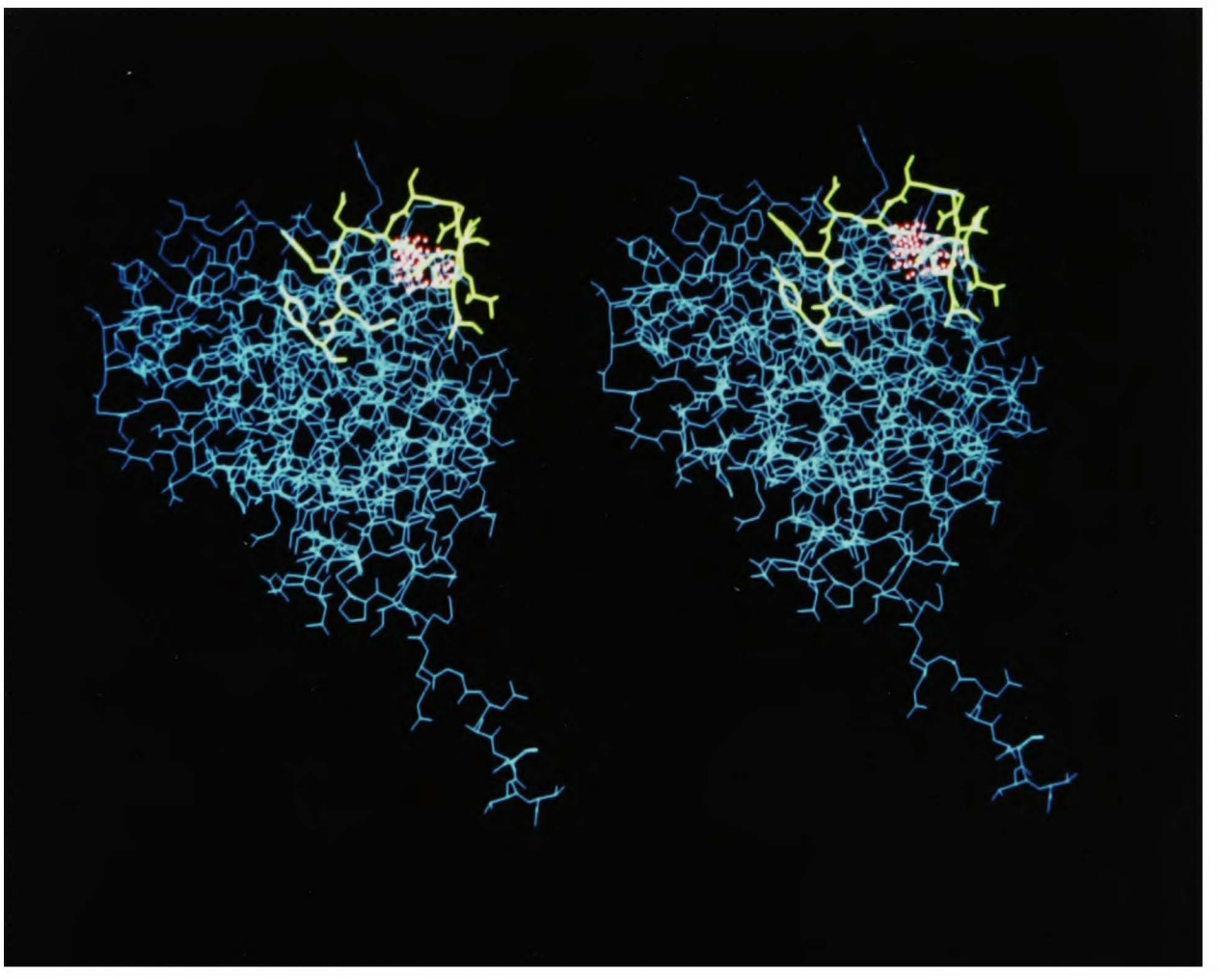


Figure 25: Stereo-view of the modelled zinc binding site in antibody HyHEL-5. a) The metal binding site created in L1. (The predicted position of the zinc atom is shown in magenta).



b) Points of high hydrophobicity contrast. 48 of the 51 points found, cluster within 4.0Å of the predicted zinc position.



6.4 Conclusions

Interest in the design of metal binding sites in proteins has grown rapidly in recent years, as increasing numbers of potential applications for such engineered metalloproteins have been recognized (Tainer *et al.*, 1991). The modelling procedure presented in this Chapter allows a rapid, automated search of the target protein structure for groups of residues, whose relative geometric arrangement makes them good candidates for mutation to form a binding site for metal ions. An initial model of the putative metalloprotein can then be evaluated by using the hydrophobicity contrast function to asses its relative binding potential.

In addition, it may be possible to use hydrophobicity contrast to obtain a qualitative measure of the strength of metal binding in the modelled protein; good metal binding sites in natural metalloproteins are characterised by tightly clustered regions of contrast. If so, one could assess the effects of using different combinations of ligands on the specificity and/or metal binding affinity of the engineered protein, prior to any mutagenesis.

Chapter 7

Modelling Hypervariable Loop SDR Mutations

The antibody combining site (ACS) is constructed from six hypervariable loops or complimentarity-determining regions (CDRs), of differing lengths and sequence from one antibody to the next. Chothia *et al.* observed, from an extensive analysis of the crystal structures of antibody Fabs and light chain dimers, that the conformations of each of these loops appeared to be critically dependent on the nature of the amino acids at 'key' positions along the loop sequence (Chothia and Lesk, 1987). On the basis of these observations, the different antibodies could be grouped into canonical sets of structures, each representing a particular loop conformation. It was then shown that the structures of unknown antibody combining sites could be successfully predicted on the basis of the classification of their CDRs into the known canonical sets according to their 'key' residue composition (Chothia *et al.*, 1989). These residues are termed structure-determining residues (SDRs).

Protein engineering experiments which modify the residues within the antibody combining site, clearly also have the potential to alter its three-dimensional structure and hence, antigen binding. Ideally one would avoid mutating the SDRs, since they are known to play a critical rôle in determining the structure of the combining

site. However, one can envisage situations where it may indeed be desirable to mutate one of the 'key' amino acids to introduce or remove particular functional properties at that position in the combining site (*e.g.* add a positive charge, reduce sidechain bulk). In practical terms, it would be highly advantageous if the molecular biologist could evaluate the consequences of these mutations prior to doing the wet experiments and even construct a list of acceptable amino acid substitution choices. Extending this further, if the desired mutation of a key residue appears to disrupt the combining site structure, can the original conformation be restored by making a complementary mutation elsewhere, such as is seen to occur in the natural evolutionary path of many homologous proteins (Creighton, 1984).

To address these questions, it was necessary to determine whether the combined algorithm methods developed by Martin *et al.*, (1989), were sensitive enough to accurately model the structural consequences of individual mutations of the SDRs. The work presented in this Chapter focuses primarily on reproducing the structural rôle of SDRs suggested by Chothia *et al.*, using molecular modelling. An explanation for the loss of antigen binding in a CDR-H2 mutation of the anti-lysozyme antibody Gloop 2 (Jeffrey *et al.*, 1992; Roberts *et al.*, 1987) is then suggested on the basis of similar analyses.

7.1 Altering CDR Conformation

According to Chothia *et al.*, CDRs with the same canonical structures have similar conformations owing to the presence of the same or similar residues at key positions in the antibody sequence. Chothia *et al.* have so far defined four canonical structures

for L1 (Table 28) and three for L3 (Table 29) from the available antibody structures.

Even though canonical structures 1 and 2 are of different length, the replacement of an SDR in an L1 from canonical Group 1 with the corresponding residue from an L1 in canonical Group 2, should cause the mutated L1 to adopt a conformation similar to that observed in canonical Group 2.

7.1.1 Construction of L1 Val-29→Ile in J539

From an examination of the L1 sequences shown in Table 28, it can be seen that the only clear difference at an SDR between canonical Groups 1 and 2 is at position 29 in the structures J539 (Suh *et al.*, 1986) and REI (Epp *et al.*, 1975). In J539, the SDR is a Valine, while in REI it is an Isoleucine. To determine the affect of SDR 29 in J539 L1, this CDR, containing a Val-29→Ile substitution (V29I), was modelled using the combined algorithm method (CAMAL—see Chapter 4). This suite of programs was also used to model the native J539 L1 as a control. If Isoleucine is responsible for the conformation of L1 in REI, then this substitution should cause the J539 L1 to exhibit conformational similarities with the CDRs in canonical Group 2.

CDR	Construction	Total Conformations	CPU Time/h	
			CONGEN	GROMSCAN
Native J539 L1	SAS[S(SVS)S]LH	529	46	130
J539 L1 V29I	SAS[S(SIS)S]LH	530	46	30

The control construction selected the conformation found in the J539 crystal structure, as would be expected for the native L1 sequence. The conformation selected for the J539 L1 V29I substitution is shown together with native J539 L1 and native REI L1 in Figure 26. Figure 27 shows a bar chart comparing the ϕ angles for native

CS	Protein	26	27	28	* 29	30	31	a	b	c	d	e	f	32	* 2	* 25	* 33	* 71
1	J539	S	S	S	V	S	—	—	—	—	—	—	—	S	I	A	L	Y
	HyHEL-5	S	S	S	V	N	—	—	—	—	—	—	—	Y	I	A	M	Y
	NQ10	S	S	S	V	R	—	—	—	—	—	—	—	Y	I	A	M	Y
2	REI	S	Q	D	I	I	K	—	—	—	—	—	—	V	I	A	L	Y
	D1.3	S	G	N	I	H	N	—	—	—	—	—	—	Y	I	A	L	Y
	HyHEL-10	S	Q	S	I	G	N	—	—	—	—	—	—	N	I	A	L	F
	NC41	S	Q	D	V	S	T	—	—	—	—	—	—	A	I	A	L	Y
3	McPC603	S	E	S	L	L	N	S	G	N	E	K	N	F	I	S	L	F
4	4-4-20	S	Q	S	L	V	H	S	—	N	G	N	T	Y	V	S	L	F

Table 28: Chothia *et al.*'s canonical structures for CDR-L1.

CS	Protein	91	92	93	94	95	96	90
						*		*
1	REI	Y	Q	S	L	P	Y	Q
	McPC603	D	H	S	Y	P	L	N
	D1.3	F	W	S	T	P	R	H
	HyHEL-10	S	N	S	W	P	Y	Q
	NC41	H	Y	S	P	P	W	Q
	4-4-20	S	T	H	V	P	W	Q
	NQ10	W	S	S	N	P	L	Q
					*			*
2	J539	W	T	Y	P	L	I	Q
						*		*
3	HyHEL-5	W	G	R	N	P	—	Q

Table 29: Chothia *et al.*'s canonical structures for CDR-L3.

Sequences of Chothia *et al.*'s structural definitions for the Vκ and V_H hypervariable regions of known structure. The residues listed here, using the standard numbering of Kabat and Wu (Kabat *et al.*, 1987), are those that form the CDRs and those in the framework regions which are important for the observed CDR conformations (Chothia and Lesk, 1987). The sequences are grouped so that those having the same backbone conformation or canonical structure (**CS** in the Tables) are adjacent. The residues in the hypervariable and framework regions which are primarily responsible for these conformations are indicated by an asterisk. (Data reproduced and slightly modified from Chothia *et al.*, (1989) Table 1.)

J539 L1, J539 L1 V29I and native REI L1. The r.m.s. deviation between the native J539 L1 and the L1 V29I construction is 0.87Å on Cα and 0.77Å on backbone (N, Cα, C). The deviation chiefly results from a rotation around ϕ of 220° at Ser-30 and of 61° at Ser-31. The ϕ angle is now positive for residue 30, as in REI.

This result implies that substituting an SDR from one canonical structure with that from another, alters CDR conformation in a manner which conforms to the parent structure of the substituent amino acid. This supports the findings of Chothia *et al.*

Table 28 shows that canonical groups always contain hypervariable regions of identical length *i.e.* length is the primary determinant of canonical structure. While it is clear that variation in loop length between J539 and REI introduces considerable conformational differences, this experiment has shown that similar conformational changes can be induced simply by substitution of an SDR. As a further test of the importance of SDRs, 2 canonical groups containing loops of identical length were selected (Groups 1 and 2 in CDR-L3).

7.1.2 Construction of L3 Pro-94/Leu-95 Swap in J539

In CDR-L3, Proline at position 95 is defined to be an SDR in all sequences of all canonical groups except Group 2, where it is found at position 94. J539 in Group 2 has a Leucine at position 95. Comparing this sequence with REI in Group 1, a swap of Leu and Pro at positions 94 and 95 is observed (Table 29). Thus, to investigate the structure-determining effects of the Proline at these positions, L3 of J539 was reconstructed, with Pro and Leu swapped to the positions observed in REI *i.e.* Pro-94→Leu and Leu-95→Pro (P94LL95P). If SDRs are important in defining the loop's

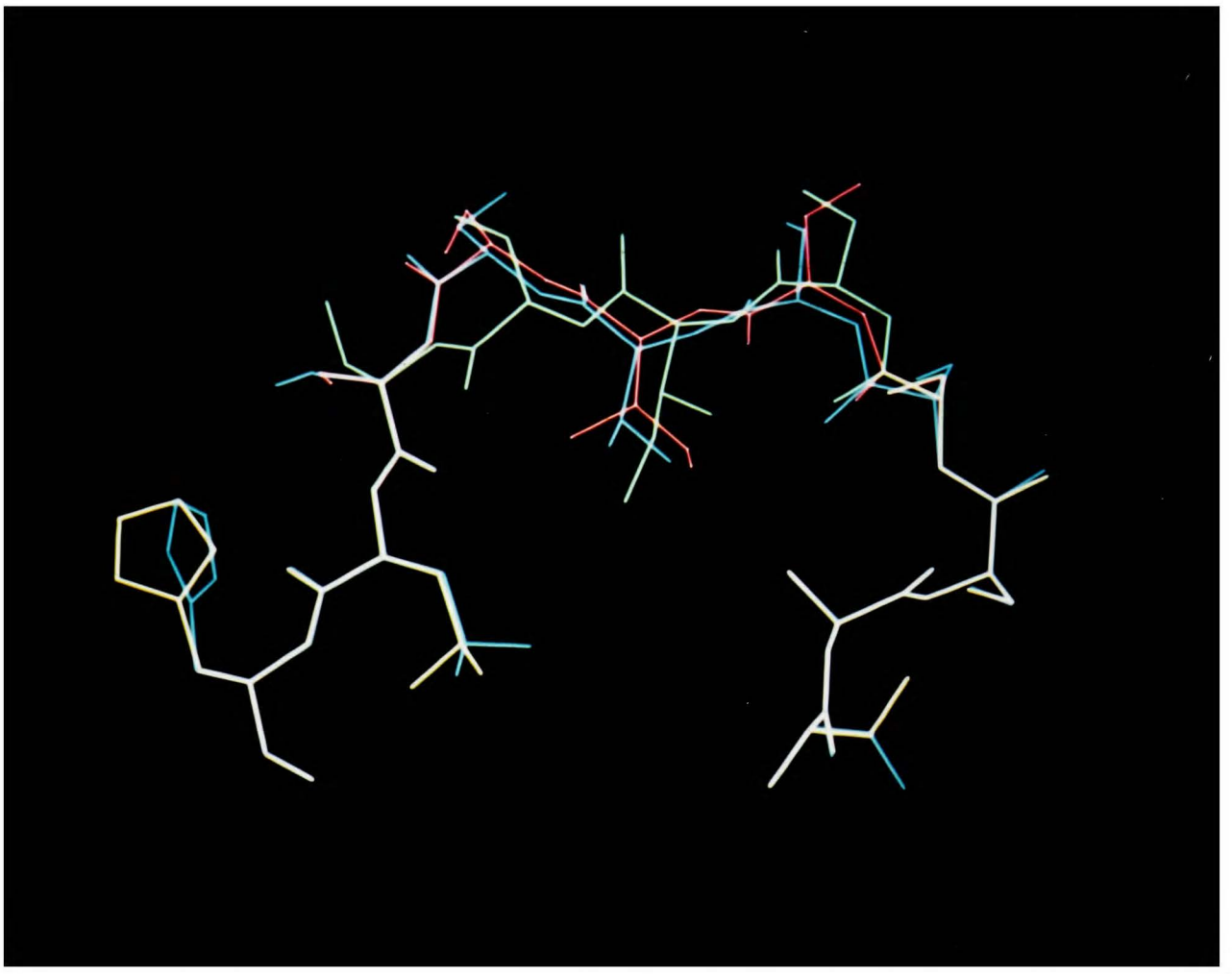


Figure 26: Comparison of native J539 L1 (blue), J539 L1 V29I (green) and native REI L1 (red).

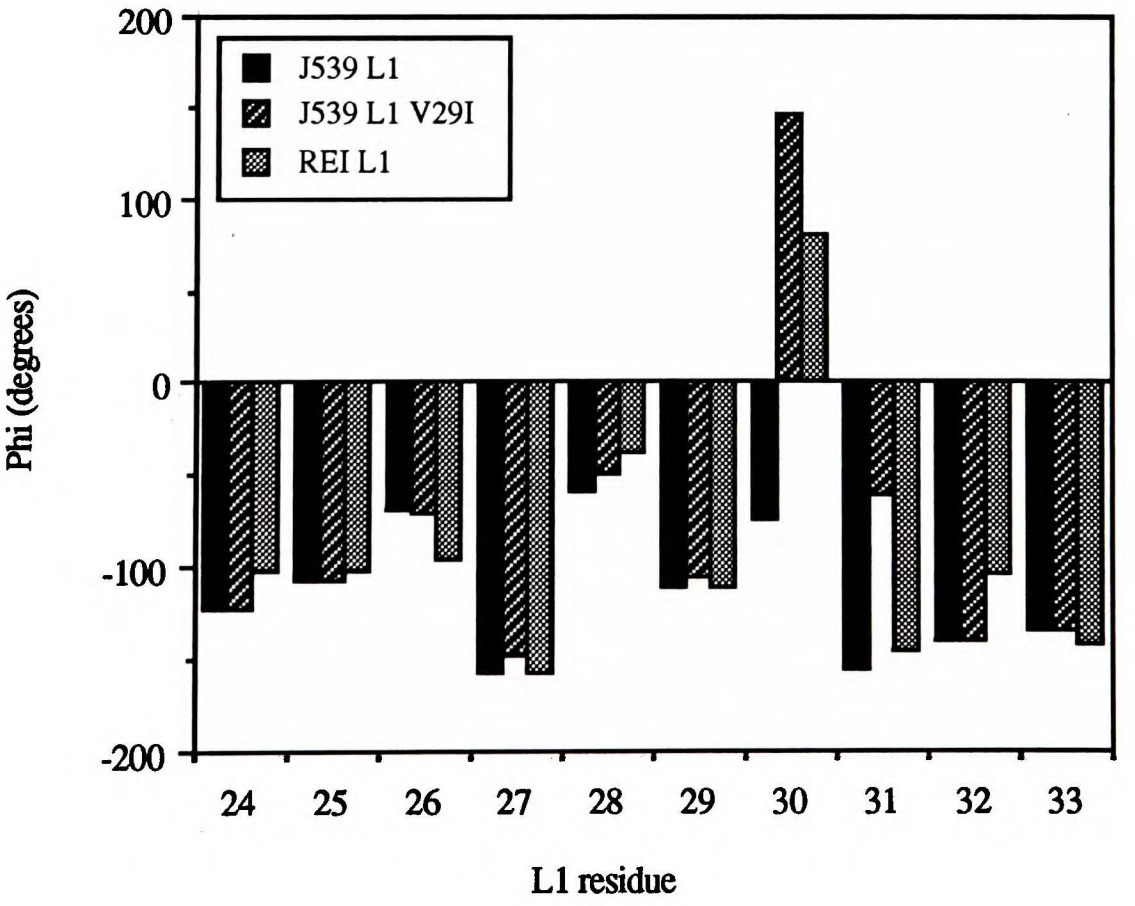


Figure 27: Bar chart showing a comparison of the ϕ angles for native J539 L1, J539 L1 V29I and native REI L1. Note the positive angle for residue 30 of J539 L1 V29I.

conformation, a conformation similar to canonical Group 1 would be expected. The P94LL95P construction was performed using CAMAL and native J539 L3 was built as a control.

CDR	Construction	Total Conformations	CPU Time/h	
			CONGEN	GROMSCAN
Native J539 L3	QQ[W(TYP)L]IT	12480	1320	490
J539 L3 P94LL95P	QQ[W(TYL)P]IT	1216	129	48

Again, the control construction selected the conformation found in the crystal structure. The conformation selected for the J539 L3 P94LL95P construction is shown together with native J539 L3 and native REI L3 in Figure 28. Figure 29 shows a bar chart comparing the ϕ angles for native J539 L3, J539 L3 P94LL95P and native REI L3. The r.m.s. deviation between the native J539 L3 and the P94LL95P L3 is 0.59Å on Cα and 0.74Å on backbone. The deviation is mainly the result of rotations around ϕ of 224°, 54°, 34° and 62° at residues Thr-92, Tyr-93, Pro-94 and Leu-95 respectively. This experiment has shown that altering the position of the Proline affects the conformation of J539 L3, lending further support to the rôle of this residue as an SDR. However, as Figure 29 shows, the selected conformation has little similarity with REI L3. This can be seen from the inter-Cα distance between residue 92 and residue 95 for the native and mutant loops.

CDR	Distance/Å
Native J539 L3	5.37
J539 L3 P94LL95P	6.59
Native REI L3	8.67

The P94LL95P L3 has expanded in size, but occupies a conformation between native J539 L3 and REI L3. An examination of these three loops using interactive molecular graphics, revealed a possible explanation for this result.

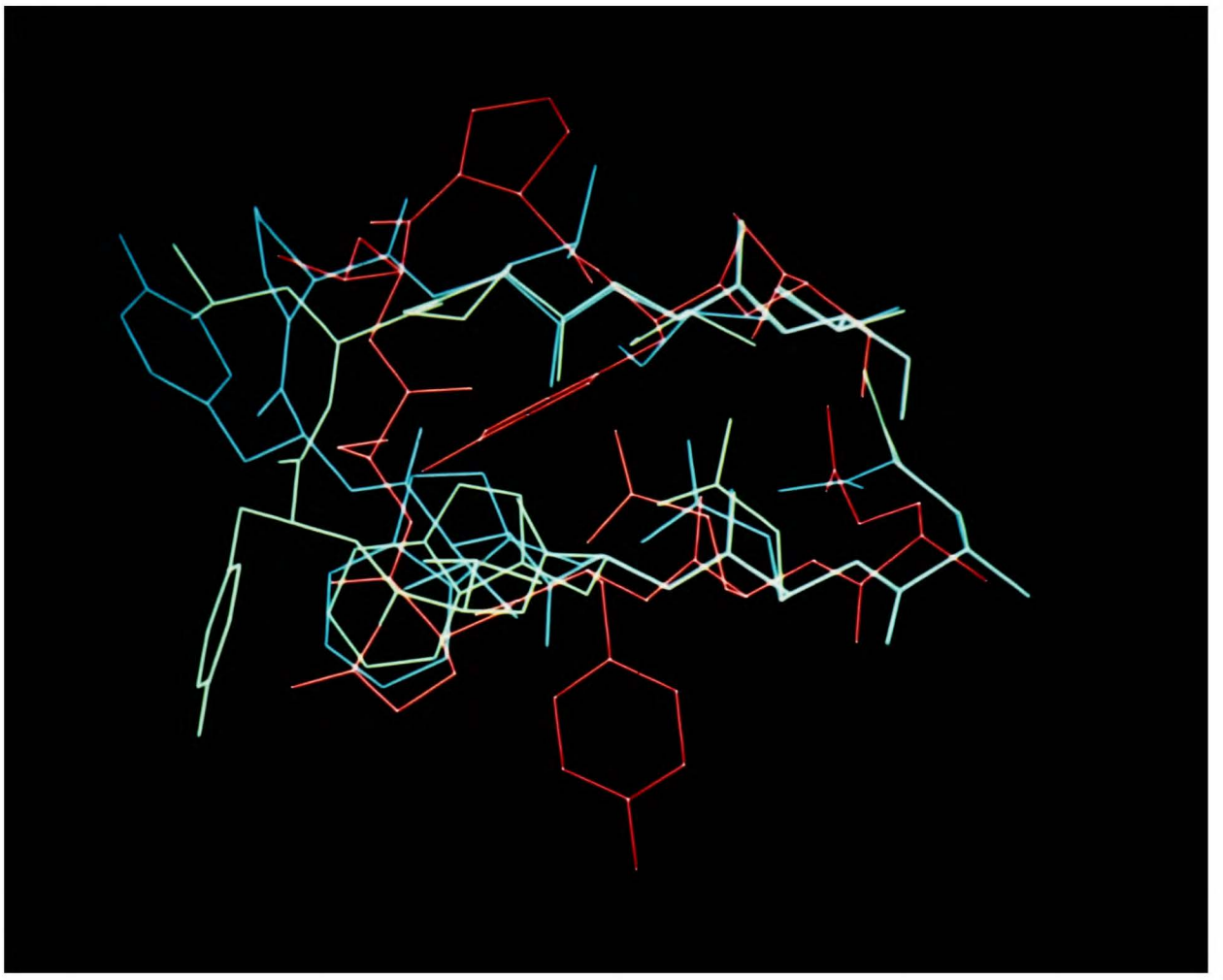


Figure 28: Comparison of native J539 L3 (blue), J539 L3 P94LL95P (green) and native REI L3 (red).

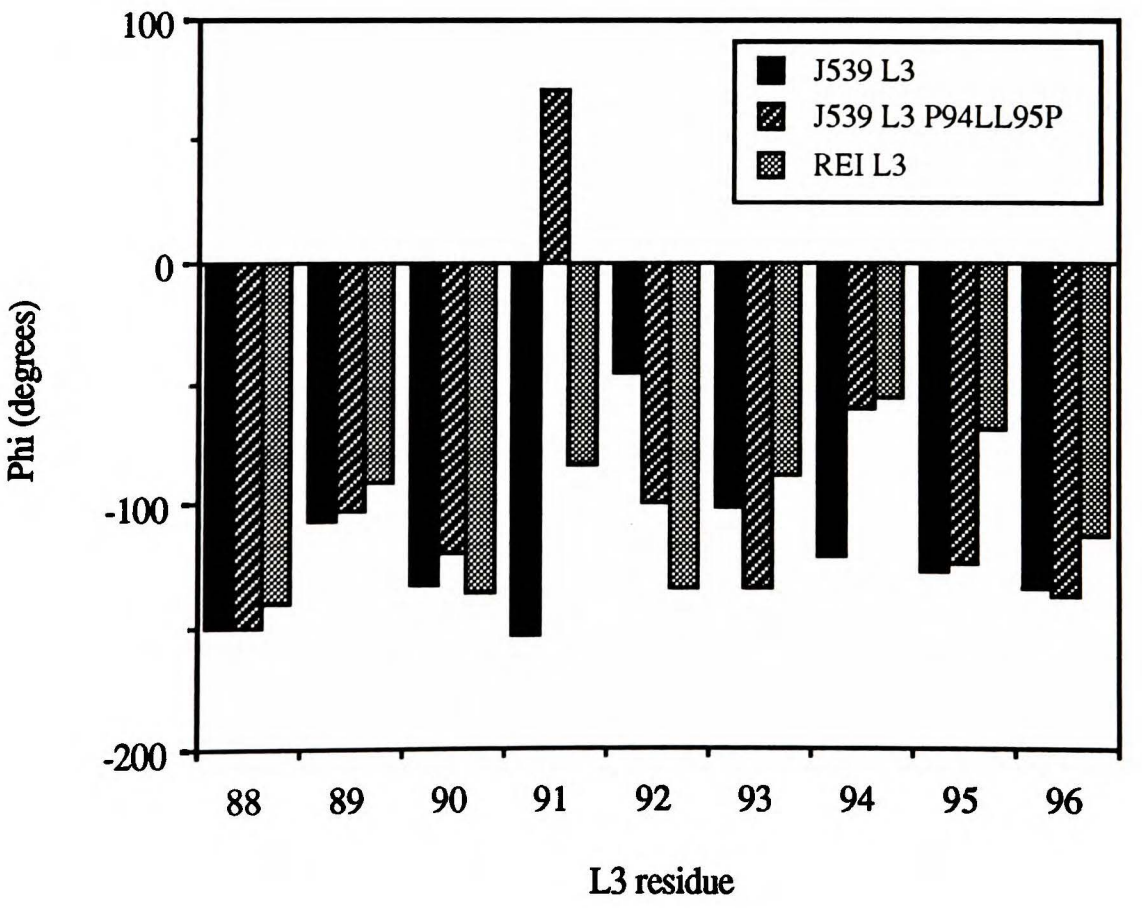


Figure 29: Bar chart showing a comparison of the ϕ angles for native J539 L3, J539 L3 P94LL95P and native REI L3. Note the positive angle for residue 91 of the P94LL95P L3.

In J539, native L3 has no adjacent CDR or framework residues with which it interacts. The conformation is thus defined only by intra-loop residue contacts. The only major interaction involves the sidechain of Tyr-93 which packs face to face onto the pyrollidine ring of Pro-94. This interaction may cause some restriction in the conformations that can be adopted. Altering the position of the Proline as in P94LL95P, removes this interaction. In addition, possibly as a result of steric hindrance by Leu-94, the sidechain of Tyr-93 flips over and now interacts with Thr-92. This interaction is much less substantial and may be less conformationally restricting.

7.1.3 SDR Coupled Partner Mutations

The results of the above experiments give further support to the hypothesis of Chothia *et al.*, that certain residues at particular positions within hypervariable (and framework) regions, determine CDR conformation. In addition, it was shown that CDR conformation could be changed to a different canonical structure, simply by substituting a single SDR.

An SDR exerts its effects through interactions with other residues or SDRs, as described by Chothia *et al.* Conformational changes are, therefore, a result of changes in residue interactions. By maintaining packing interactions through alteration of the residues contacting the SDR, it might be possible to retain the native conformation. Alteration of the SDR-contacting residues, represents 'coupled partner' substitutions. An examination of the J539 L1 V29I result provided a test for this proposition.

The substitution of Val-29 by Ile produces an increase in size owing to the

J539 L1 Structure	Ile-2	(L1) Ala-25	Tyr-71	(L3) Gln-90	(L3) Thr-92
Native J539 L1	3.79	3.81	3.37	3.49	3.92
Native J539 L1 backbone; CONGEN Ile-29	3.78	3.82	3.23	3.53	3.88
J539 L1 V29I	3.63	3.79	3.93	3.32	3.54

Table 30: Shortest interaction distances for residues within a 4.00Å radius of position 29 in three J539 L1 structures. *Native J539 L1 backbone; CONGEN Ile-29*: CONGEN sidechain search of Ile-29 on the native J539 L1 backbone.

presence of the Ile Cδ1 atom. Since Val and Ile are identical except for this atom, the change in conformation must be a result of the larger Ile sidechain at position 29 in L1. The native J539 L1 conformation was examined, with Val-29 replaced by Ile using the REPLACE and REFI (Hermans and McQueen, 1974) procedures available within the molecular graphics program FRODO (Jones, 1985). However, it was noticed that the orientation of the Ile sidechain calculated within FRODO, was significantly different from that calculated by the CAMAL for the V29I L1.

The preferred position of this sidechain is important because the location of the Ile Cδ1 atom determines the number and distances of any residue contacts. Therefore, to determine which orientation was correct, a sidechain only construction of Ile-29 was carried out using CONGEN (Bruccoleri and Karplus, 1987) and a total of 5 conformations were produced. The r.m.s. deviation between Ile-29 from the lowest energy conformation and the position of this residue in the J539 V29I L1, was 0.49Å. In contrast, the deviation with the Ile-29 position generated by FRODO was 2.13Å. This confirmed that the position calculated by CAMAL was preferred.

Table 30 shows the shortest interaction distances between a residue at position

29 in three J539 L1 structures and residues within a 4.00Å radius of this. Three of the proximal residues, Ile-2, Ala-25 and Tyr-71 are SDRs. The other two, Gln-90 and Thr-92, are the first and third residues of CDR-L3. An examination of these distances suggested which residue should be investigated as a possible coupled partner for Ile-29.

The Cβ of Gln-90 interacts with Ile-29 in all three structures, indicating that even substituting with the smallest sidechain, Ala, the Ile-29 sidechain may not be accommodated without a conformational change in L1. Thr-92 as part of L3 should not be changed even if it is a coupled partner of Ile-29. Mutating this residue could alter the conformation of L3. Hence, any effect on antigen binding may result from disruption of this CDR, rather than of L1.

This leaves the three SDRs Ile-2, Ala-25 and Tyr-71. The contact distances between Ala-25 and Ile-29 within L1 remain similar across the structures where Ile-29 is present, implying that the Ala-25 interaction was not significantly affected by the presence of this sidechain. The contact distances of Tyr-71 with a sidechain at position 29 decreases by 0.14Å from the native L1 structure to the L1 CONGEN Ile-29 structure, but then increases by 0.70Å in the J539 L1 V29I structure. It is the accommodation of the Ile-Cδ1 atom through movement of the L1 backbone in the V29I structure, which causes the increase in the Tyr-71 distance, thus reducing the effective interaction with this residue.

Finally, the contact distance for Ile-2 with Ile-29 is shorter in L1 V29I. The presence of the Ile-29 Cδ1 atom reduces the contact distance with Ile-2 in the structure. This interaction could be the main determinant in the conformational change of L1. Therefore, it was decided to reconstruct the J539 L1 V29I substitution

Structure 1	Structure 2	R.m.s. Deviation/Å	
		Cα	Backbone
Native J539 L1	J539 L1 V29I	0.87	0.77
Native J539 L1	J539 L1 V29II2A	0.35	0.31
J539 L1 V29I	J539 L1 V29II2A	0.76	0.69

Table 31: R.m.s. deviations between native J539 L1, J539 L1 V29I and J539 L1 V29II2A. Structure 2 was fitted onto Structure 1.

exactly as before, but with Ile-2 replaced by Ala (V29II2A).

CDR	Construction	Total Conformations	CPU Time/h	
			CONGEN	GROMSCAN
Native J539 L1	SAS[S(SVS)S]LH	529	46	130
J539 L1 V29I	SAS[S(SIS)S]LH	530	46	30
J539 L1 V29II2A	SAS[S(SIS)S]LH	561	49	32

The selected conformation for J539 L1 V29II2A is shown together with native J539 L1 and J539 L1 V29I, in Figure 30. Figure 31 shows a bar chart comparing the ϕ angles for native J539 L1, J539 L1 V29I and J539 L1 V29II2A.

Table 31 shows the r.m.s. deviations between native J539 L1, J539 L1 V29I and J539 L1 V29II2A. The deviation between L1 V29I and L1 V29II2A chiefly results from a rotation around ϕ of 221° at Ser-30 and of 89° at Ser-31, and contrasts sharply with rotations around ϕ of 1° at Ser-30 and of 7° at Ser-31 between L1 V29II2A and the native L1. This shows that the conformation selected for the V29II2A L1 is very similar to the native L1.

The result of this investigation shows that by adjusting the interactions between two (non-native) SDRs through residue substitution, a conformation very similar to that of the native CDR can be obtained. This suggests that mutations can be incorporated into the ACS without significant alteration in antigen affinity/specificity,

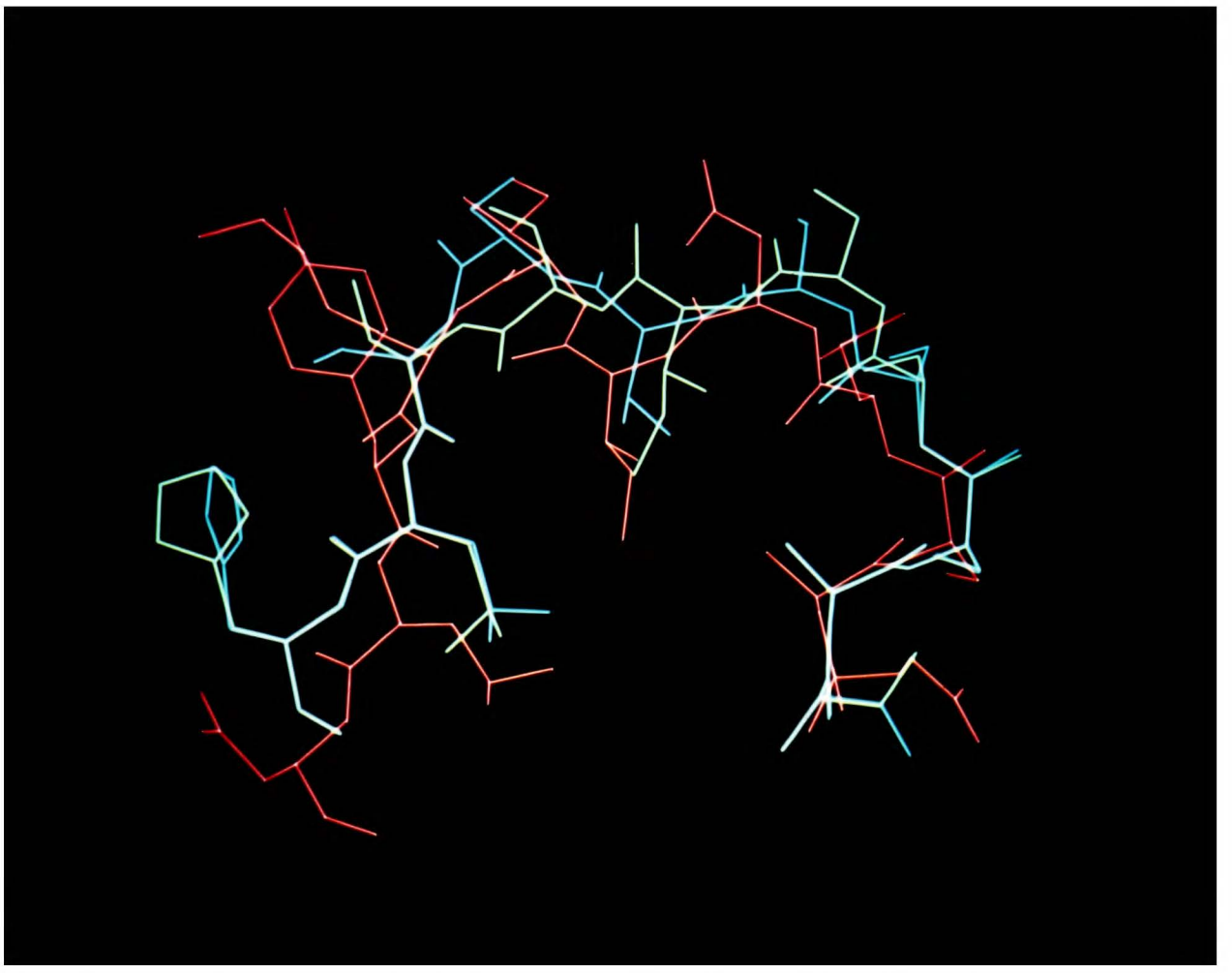


Figure 30: Comparison of native J539 L1 (blue), J539 L1 V29I (red) and J539 L1 V29II2A (green).

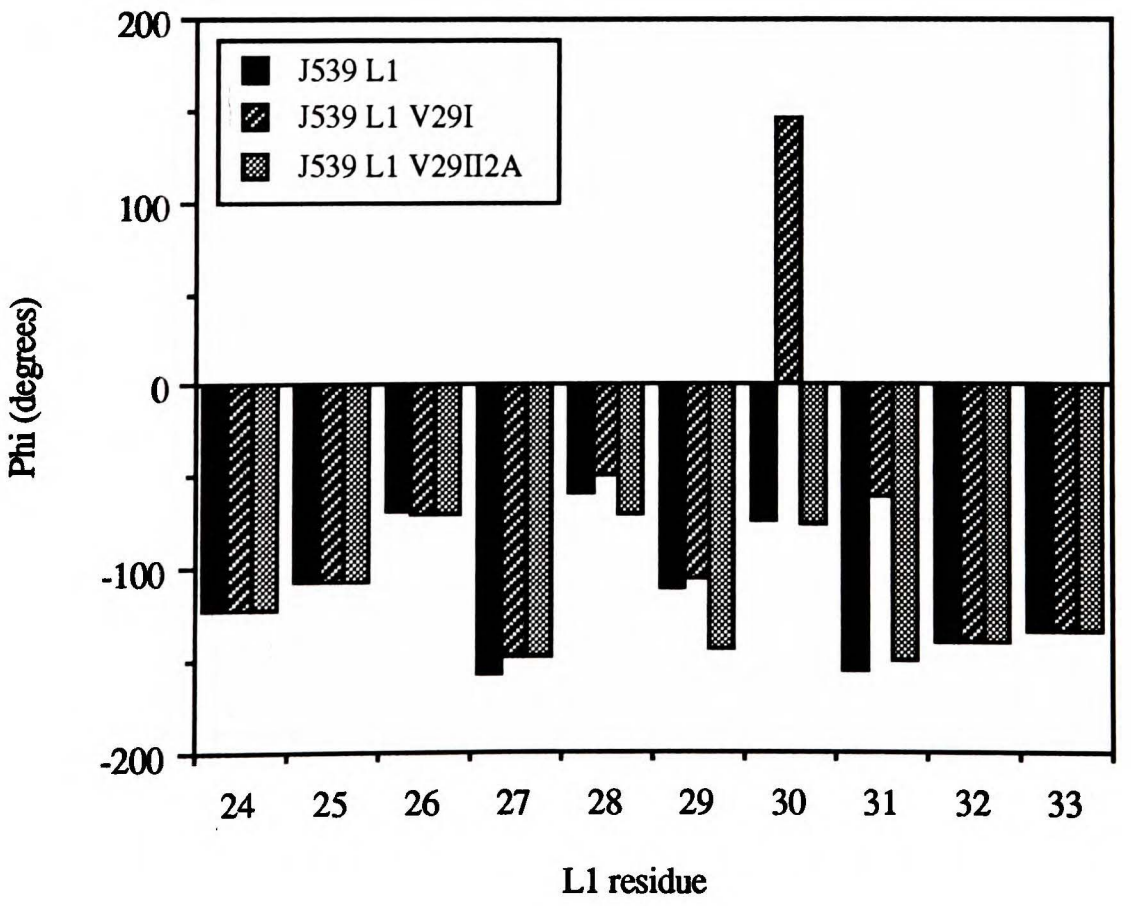


Figure 31: Bar chart showing a comparison of the ϕ angles for native J539 L1, J539 L1 V29I and J539 L1 V29II2A. Note the similarity in angles for the native L1 and the V29II2A L1.

owing to the retention of a native-like CDR conformation.

7.2 A Mechanism to Explain Loss of Binding in Gloop 2 Mutants

The experiments carried out so far have shown the significant influence that SDRs have on CDR conformation. Chothia *et al.*, (1989) have used the presence of these residues within CDRs, to predict their structure. The previous section demonstrated that substitution of SDRs can be used to control CDR conformation, with the possibility of altering or retaining antigen affinity/specificity. These residues exert their effect through specific interactions with other residues. The nature of these interactions determines CDR conformation. In this section, the interactions of two mutations in Gloop 2, shown experimentally to abolish binding of the antigen lysozyme, are investigated in an attempt to explain why this loss of binding occurs.

7.2.1 Mutations in Gloop 2

From an analysis of a computer model of the Gloop 2-lysozyme complex, Roberts *et al.*, (1987) found residues partially buried within the combining site, inaccessible to antigen contact and variable between individual antibodies. One such residue was Glu-50, the first residue in heavy chain CDR-H2. A mutant of Gloop 2 was expressed in which Glu-50 was replaced by Ser (E50S). This mutation resulted in the complete abolition of binding to both ^{125}I-labelled Pep1 (a loop-containing peptide derived from lysozyme) and lysozyme. This residue shows high variability between heavy chains of the same V_H II subgroup (Kabat *et al.*, 1987). In the model

generated by de la Paz *et al.*, (1986) and later refined by Roberts *et al.*, (1987), this residue appeared to be both hydrogen bonded to Tyr-94 in neighbouring CDR-L3 and Thr-35 in CDR-H1.

It was predicted that substitution at this position would cause a significant change in the relative conformations of adjacent, interacting CDRs owing to the loss of these interactions and that such perturbations would more than account for the observed loss of binding. A second Gloop 2 mutant in which Glu-50 was replaced by Arg also showed a complete loss of binding to both Pep1 and lysozyme. This substitution could be predicted to disrupt the above interactions as a result of its large size and electrostatic repulsion with Arg-101 in CDR-H3.

An examination of Glu-50 in the Gloop 2 P_1 crystal structure (Jeffrey *et al.*, 1992) using interactive molecular graphics, confirmed the predicted interactions with Tyr-94 and Thr-35. Visual inspection of the Glu-50 environment indicated possible consequences of a Ser or Arg at this position. In the E50S mutant, a reduction in the size of this sidechain would allow Tyr-94 at the apex of L3, to move into the resulting cavity. This would perturb the conformation of L3 resulting in the loss of antigen binding. Residues from CDR-H2 are less likely to move into this cavity because the mutation is at the base of H2 adjacent to the framework. Similarly, Thr-35 is the last residue of H1 preceeding the framework, making it unlikely that any gross movement would occur.

The consequences of an Arg at this position are more obvious. There is a steric clash with Tyr-94 owing to the large bulk of the Arg sidechain. This may alter the conformation of L3 in order to allow the Tyr-94 sidechain to move out of the way. Apart from possibly losing a hydrogen bond, Thr-35 should be able to maintain

similar interactions with Arg-50; it interacts with all atoms in the native Glu-50 sidechain.

To test these predictions and to determine the extent of the conformational disruption owing to Ser and Arg, the CAMAL was used to reconstruct H2 and L3 in the presence of both mutations. Control constructions of native H2 and L3 were also carried out. For construction purposes the definition of H2 follows that of Martin (1990), running from the N-terminus of the Kabat and Wu definition (Kabat *et al.*, 1987) to its strand partner at the C-terminus. All the constructions were carried out using the Gloop 2 P_1 crystal structure (first molecule in the unit cell).

7.2.2 Construction of H2 Glu-50→Arg

CDR	Construction	Total Conformations	CPU Time/h	
			CONGEN	GROMSCAN
H2	[E(IFP)G]NSKTY	1576	140	120
H2E50R	[R(IFP)G]NSKTY	1578	140	123

The r.m.s. deviation between the selected conformation from the control construction and the crystal structure CDR is 1.22Å on all atoms and 0.54Å on backbone. This compares with an r.m.s. deviation of 2.85Å on all atoms and 1.36Å on backbone, for the selected conformation of the H2 construction performed by Martin (Martin, 1990). The selected conformation for the H2 construction shown here, is closer to that in the crystal structure by 1.63Å on all atoms and 0.82Å on backbone. Thus, a conformation for the native H2 similar to the crystal structure, was obtained.

This H2 conformation was calculated from a set of database loops which were

than the one used by Martin (1990). The greater amount of structural information available from the larger databank resulted in a considerable improvement in the r.m.s. deviation of the selected conformation.

The r.m.s. deviation between the selected conformation for the H2 Glu-50→Arg mutation (H2E50R) and the Gloop 2 crystal structure CDR was 1.15Å on Cα and 0.91Å on backbone. These values are not significantly higher than the r.m.s. deviations achieved for the native construction.

The H2E50R mutant model H2 produced two interesting results. The original interactions with L3 (a hydrogen bond with Tyr-94) and H3 (a salt-link with Arg-101) had been maintained, while the hydrogen bond with Thr-35 had been lost. The salt-link with H3 remained, but no longer involved Arg-101 owing to electrostatic repulsion. It now occurred between Arg-50 and Glu-102, reversing the polarity of the salt-link.

The sidechain positions had radically altered, as shown by two groups of interactions. In the crystal structure, there is a hydrophobic interaction between Ile-51 and Pro-53 in H2. The face of the pyrollidine ring packs onto the planar Ile sidechain. In H2E50R, this interaction is no longer present. Additionally, a small hydrophobic cluster has formed between the ring of Phe-52 (which packs face-to-face onto the ring of Tyr-59) and the sidechain carbons of Lys-57[1]. The Cβ of Asn-55 was also orientated towards this cluster. The backbone deviation resulting from the Arg-50 mutation caused the whole loop to move sideways relative to the crystal structure CDR. Loss of binding owing to the effect of the Arg-50 mutation may thus be caused by the gross changes in sidechain packing, removing antigen-contacting

[1]Martin (1990) has found evidence for hydrophobic interactions of Arg and Lys *via* the sidechain carbons, in the structures of numerous proteins.

residues from their correct positions.

7.2.3 Construction of L3 with H2 Glu-50→Arg

CDR	Construction	Total Conformations	CPU Time/h	
			CONGEN	GROMSCAN
L3	LQ[Y(LSY)P]LT	10743	1219	433
L3E50R	LQ[Y(LSY)P]LT	3857	251	243

The r.m.s. deviation between the selected conformation from the control construction and the crystal structure CDR is 2.34Å on all atoms and 1.34Å on backbone. This shows close agreement with the r.m.s. deviation of 2.83Å on all atoms and 1.27Å on backbone, for the selected conformation of the L3 construction performed by Martin (1990). Again both conformations compare favourably with the crystal structure CDR. The selected conformation for the L3E50R construction also shows close similarities with the crystal structure CDR. The r.m.s. deviation between these two structures is 2.16Å on all atoms and 0.71Å on backbone. The effect of the H2 Glu-50→Arg mutation on L3 is seen by comparison with the crystal structure CDR.

In the crystal structure, Tyr-94 is involved in a small hydrophobic cluster with framework residue Trp-47 and Tyr-59 from H2. In the presence of the Glu-50→Arg mutation, this cluster is destroyed. The sidechain of Tyr-94 has rotated by 143° around χ_1 away from the Arg (relieving the steric clash which would otherwise result) so that it now protrudes from the surface of the antibody. The movement of this sidechain causes a peptide flip between Ser-93 and Tyr-94 (98° and 172° around ϕ for Ser-93 and Tyr-94 respectively). The sidechain of Tyr-91 has rotated, by 231° around χ_1, towards the antibody surface as a result of the peptide flip. The position of Leu-93 has adjusted to accommodate the Arg-50 sidechain.

Apart from the peptide flip, the shape and relative position of the loop within the ACS had not changed (backbone r.m.s. deviation 0.71Å). The effect on antigen binding owing to the Glu-50→Arg mutation, could result from changes in the sidechain positions of the residues within L3, possibly causing unfavourable interactions with the antigen rather than a change in shape of the whole CDR (as for the previous H2E50R construction).

7.2.4 Construction of H2 Glu-50→Ser

CDR	Construction	Total Conformations	CPU Time/h	
			CONGEN	GROMSCAN
H2	[E(IFP)G]NSKTY	1576	140	120
H2E50R	[R(IFP)G]NSKTY	1578	140	123
H2E50S	[S(IFP)G]NSKTY	1575	138	258

The construction of the H2 Glu-50→Ser mutation (H2E50S) is shown together with the native H2 and H2E50R constructions performed previously, in the table above. The selected conformation for this construction was identical to the crystal structure H2, despite the presence of the Glu-50→Ser mutation. This suggested that a reduction in size of the residue at position 50 does not, in itself, affect the conformation of H2. In addition, the loss of the hydrogen bonds with Tyr-94 and Thr-35 and the salt-link with Arg-101, indicate that these interactions were not important in the maintenance of the H2 conformation.

Structure 1	Structure 2	R.m.s. Deviation/Å	
		Cα	Backbone
L3	L3E50R	2.16	0.71
L3	L3E50S	2.07	0.65
L3E50R	L3E50S	0.92	0.45

Table 32: R.m.s. deviations between native Gloop 2 L3, Gloop 2 L3E50R and Gloop 2 L3E50S. Structure 2 was fitted onto Structure 1.

7.2.5 Construction of L3 with H2 Glu-50→Ser

CDR	Construction	Total Conformations	CPU Time/h	
			CONGEN	GROMSCAN
L3	LQ[Y(LSY)P]LT	10743	1219	433
L3E50R	LQ[Y(LSY)P]LT	3857	251	243
L3E50S	LQ[Y(LSY)P]LT	4674	319	241

The construction of L3 in the presence of the H2 Glu-50→Ser mutation (L3E50S) is shown together with the native L3 and L3E50R constructions performed previously, in the Table above. The r.m.s. deviations between the conformations from these three constructions are shown in Table 32.

It was predicted that the Tyr-94 sidechain of L3 would occupy the cavity created by the Glu-50→Ser mutation. This was not observed in the model. As shown by the r.m.s. deviation, the selected conformation is very similar to that chosen for the L3E50R construction. Visual inspection of these conformations showed the only difference to be a peptide flip at Leu-92 ($\Delta\phi = 236°$) in L3E50S. Thus, changes in the residue contacts are similar to those described for L3E50R (see Section 7.2.3). In addition, comparison of these two conformations with the native L3 construction, showed all three conformations to be almost identical. This suggested that a residue at position 50 in H2 has little effect on the conformation of L3 given that, when

replaced by two distinctly different sidechains (a large, positive Arg or a small, polar Ser), no apparent change in the conformation of L3 was observed.

The constructions carried out so far on the Glu-50 mutants have provided an explanation only for the possible effects of the Arg substitution. The model of the Ser mutation created a conformation for H2 identical to the crystal structure, indicating that the CDR is apparently unaffected by this substitution. Contrary to the initial predictions, the neighbouring CDR-L3 constructions were also, ostensibly, unaffected by either the Arg or Ser mutations.

Given the interaction of Thr-35 from H1 with Glu-50 in the Gloop 2 crystal structure, it was decided to construct this CDR to establish whether it was affected by a mutation at position 50 in H2. Thr-35 interacts with the whole of the Glu sidechain in the native antibody including a hydrogen bond with the carboxyl group. Much of the sidechain interaction would be maintained with the longer Arg mutation (except, possibly, for the hydrogen bond), but $\approx$ 50% of this interaction would be lost with the Ser mutation. Hence, the H1 construction was performed only in the presence of this mutation.

7.2.6 Construction of H1 with H2 Glu-50→Ser

CDR	Construction	Total Conformations	CPU Time/h	
			CONGEN	GROMSCAN
H1	[T(FGI)T]	511	44	27

The selected conformation for this construction was identical to the crystal structure CDR, despite the presence of the Glu-50→Ser mutation. Again, this suggested that a reduction in size of the residue at H2 position 50 does not affect the conformation

of H1. The smaller number of sidechain contacts, including the loss of the hydrogen bond between Thr-35 and Ser-50, indicate that these interactions are not involved in maintaining the conformation of H1 (*c.f.* H2E50S construction (Section 7.2.4)).

7.2.7 Sidechain Construction around H2 Glu-50→Ser

The loss of binding by the Arg mutation could be attributed to a change in the sidechain positions of H2, which could disrupt the topology of the ACS so that it no longer recognises the antigenic epitope. However, the loss of binding caused by the Ser mutation was still unexplained. The L3 and H1 constructions suggested that both these CDRs were unaffected by the mutation, contradicting the predictions made by Roberts *et al.*, (1987) that it would alter their conformations.

It has been shown for the H2E50R construction and for the L3 constructions, that the positions of the sidechains show the greatest change. This suggests that loss of binding may result from subtle sidechain rearrangements rather than gross changes in CDR backbone conformation. Novotný *et al.*, (1990) have shown that changes in the hydrogen bonding network owing to the presence of a spontaneous Ser→Arg mutation in the anti-digoxin/digitoxin antibody 40-150 A2.4, are sufficient to alter antigen specificity, despite this mutation being distant from the ACS.

It was decided to carry out a detailed sidechain construction of residues around the Ser mutation. An identical construction around the native Glu-50 was also carried out as a control. Accordingly, residues within 8.00Å of Glu-50 were identified, to include all sidechains which might possibly be affected by the Ser substitution. These residues fell into four groups: Tyr-94, Pro-95 and Leu-96 (L3); Phe-30, Ile-32 and Thr-35 (H1); Trp-45 (heavy chain framework) and Glu-97 (H3).

Permutation	Residue Construction Order	Energy Kcal/mol	
		E50 Control	E50S Mutant
1	94,95,96,30,32,35,45,97	11.2981	39.8830
2	97,45,35,32,30,96,95,94	22.2136	43.7749
3	96,30,97,35,94,32,45,95	22.3144	43.7887
4	94,45,32,96,35,97,30,95	22.3144	43.7887
5	30,32,35,97,45,94,95,96	18.3196	33.4433
6	96,95,94,97,45,35,32,30	22.3144	43.7887
7	30,32,35,94,95,96,45,97	11.2981	39.8830
8	97,45,96,95,94,35,32,30	22.3144	43.7887

Table 33: Permutations of sidechain constructions performed around the Glu-50→Ser mutation in H2. 30 = Phe, 32 = Ile, 35 = Thr, 45 = Trp, 94 = Tyr, 95 = Pro, 96 = Leu and 97 = Glu.

The sidechain constructions were performed using CONGEN. Each conformation was generated using the ITER construction protocol (see Chapter 3) and given 10 steps of CHARMM conjugate gradient energy minimisation. To avoid biasing the iterative process towards the initial sidechain arrangement selected, the residues were constructed in a total of 8 different permutations and each set of sidechain conformations took approximately 3 minutes of CPU time to generate.

Table 33 shows that for both the native and E50S constructions, 4 of the 8 conformations generated have the same energy and interactive molecular graphics showed them to have the same conformation. Nevertheless, 4 different conformations had been generated. Examination of each one confirmed that neither the sidechains of L3 (YPL) or H1 (FIT) are affected by the Ser mutation. None of these sidechains has significantly altered its position except for minor adjustments relative to one-another, in each CDR.

In the control construction, permutation 1 had the lowest energy and was thus

selected as the final model. This conformation accurately reproduced the sidechain positions seen in the crystal structure, except for Trp-45 (all atom r.m.s. deviation 0.21Å). The remaining control permutations (except for permutation 7 which had the same conformation as permutation 1) and *all* the E50S mutant constructions showed the poor orientation of Trp-45, but also displayed a different orientation of Glu-97.

In the crystal structure, Glu-97, from H3, interacts with H1, contacting Phe-30 and making a hydrogen bond with the backbone nitrogen of Gly-31. It is located at the surface of the ACS and is fully exposed. (Although H3 is a neighbouring CDR to H2, Glu-97 has no contacts with Glu-50 so H3 was not considered for construction.) In the presence of the Glu-50→Ser mutation, the Glu-97 sidechain rotates by 31° around χ_1, 234° around χ_2 and 133° around χ_3, away from Phe-30 and down into the ACS, while maintaining the hydrogen bond with Gly-31. The r.m.s. deviation between the two sidechain positions is 0.90Å.

The interactions of framework residue Trp-45 are *via* the five-membered ring of the indole system. It contacts residues Tyr-94, Pro-95 and Leu-96 in L3, as well as Thr-35 (H1) and Glu-50 (H2) in the crystal structure. The Trp-45 sidechain also rotates, by 346° around χ_1 and 144° around χ_2, in the presence of the Glu-50→Ser mutation. The r.m.s. deviation between the two sidechain positions is 1.46Å. The six-membered ring of the indole system now has close contacts with Thr-35 (H1) and Leu-96 (L3), but the interaction with Tyr-94 (L3) is lost.

The movement of the Trp-45 sidechain in all the conformations generated, can be explained by a 'surface tension' effect. Since the constructions were carried out using the standard CHARMM potential, *in vacuo*, surface residues tend to pack

down to optimise their van der Waals contacts. By rotating into the ACS, the Trp-45 sidechain can interact more closely with Ile-96 (L3), Thr-35 (H1) and the sidechain of the residue at position 50 in H2, thus making the van der Waals energy more negative.

Table 33 also shows that the energies of the control constructions were approximately half the value of those generated for the mutant. While this can be explained partly as a result of better packing between the Trp-45 sidechain and the longer Glu-50 sidechain, it also suggests that other factors (like the loss of the hydrogen bond between Glu-50 and Thr-35, for example) not seen in this investigation, affect the Glu-50→Ser substitution and contribute to the loss of binding shown by this mutation. It also suggests that more extensive sidechain rearrangements may take place.

One possible mechanism for the loss of binding, resulting from sidechain movements, can be described. The location of Glu-97 at the surface of the ACS and its accessibility suggests that it may interact with the antigen. It is maintained in this position by a repulsive electrostatic interaction with Glu-50. When Glu-50 is replaced by Ser, the repulsive interaction is lost, allowing the Glu-97 sidechain to rotate into the ACS.

Abolition of binding could thus be attributed to loss of the H3 Glu-97 interaction with the antigen and the presence of this sidechain in the ACS. This is in addition to any other interactions which have been lost owing to the Glu-50→Ser mutation. While Glu-50 cannot be regarded as an SDR by the Chothia *et al.* definition, since it does not appear to determine backbone conformation, it is a critical residue which appears to exert influence on antigen binding by determining the conformations of

a network of sidechains. It may, therefore, be necessary to define a new class of key residues—sidechain conformation-determining residues (SCDRs). This investigation shows that sidechain as well as backbone conformations are important in protein-protein interactions.

7.3 Discussion

The investigation of SDRs carried out in this Chapter, support the findings of Chothia *et al.* concerning the rôle that these residues play in CDR conformation. In each case, the mutation of an SDR at a particular position in the antibody sequence, alters the conformation of the CDR with which it is associated. SDRs exert their influence through specific contacts that they make with other residues in the CDR and framework. These contacts take the form of packing interactions between the residues. A change in CDR conformation owing to an SDR substitution, is a result of changes in these interactions caused by differences in the shape and size of the substituent.

The mechanisms by which SDRs exert their effect fall into 2 categories. Tables 28 and 29 show that SDRs tend to occur at the beginning and end or around the mid-point of the CDR sequence; SDRs occurring in the framework are in contact with these hypervariable SDRs (coupled partner SDRs). SDRs which occur at the beginning or end of CDR sequences can dictate the backbone conformation for the rest of the loop, while those at the mid-point, although capable of distorting the loop conformation, may determine how tightly the loop is packed against the surrounding structure. Framework SDRs could influence both situations.

The presence of similar SDRs found in the same position within a CDR canonical class, suggests the conservation of particular residue contacts in order to generate the same conformations. Nevertheless, other combinations of residues at the same positions may also be able generate similar CDR conformations which fall into the same structural canonical class. This is likely to be observed in only a very small percentage of antibodies, since it requires two specific mutations to take place with no selective advantage. By its very nature, such a coupled mutation will not result in increased ACS diversity.

By examining the interactions that SDRs make, it may become possible to predict the precise effects of mutating these residues on the conformation of their associated CDR. These simulations have shown that both mutations which increase the size of a sidechain (J539 L1 V29I) and mutations that alter the packing between residues (J539 L3 P94LL95P), can alter CDR conformation. Native CDR conformation can be retained, despite an increase in the size of one SDR sidechain, if it is accommodated by a decrease in the size of a contacting SDR (or other) sidechain (J539 L1 V29II2A).

However, a change in the size of a sidechain must be considered in conjunction with the change in shape that this can represent. For example, the residues Leucine and Isoleucine have the same number of sidechain carbon atoms (4), but because of the position of the methyl substituent (attached to Cβ in Ile and Cγ in Leu), their shape is quite different. A mutation which may accommodate the asymmetric Ile sidechain may not do so for the more symmetric Leu sidechain.

A second example is provided by the incorporation of a metal binding site into L1 of HyHEL-5, as described in Chapter 4. The mutation of Ala-25→Cys increases

the size of the sidechain at this position by one atom. This is partially offset by the mutation of Val-29→Cys, a reduction in the size of the sidechain by one carbon atom. However, sulphur is larger than carbon (covalent radii: sulphur 1.02Å; carbon 0.68Å (Allen *et al.*, 1979)). Therefore, the sidechain at position 25 has increased in size by more than one carbon atom, while the sidechain at position 29 has decreased in size by less than one carbon atom. Thus, there is an overall increase in bulk within the loop which may contribute to the conformational change observed in the metal binding L1 (although this does not affect antigen binding—see Chapter 4).

7.3.1 Implications for Protein Engineering

The implications of these findings for the protein engineering of antibodies (and potentially, other proteins) are clear. The investigations of L1 and L3 SDRs show that it may become possible to alter the specificity of an antibody to that of another by substitution of corresponding SDRs, at appropriate positions within the antibody sequence. This would reduce the number of mutations necessary compared with CDR grafting (Jones *et al.*, 1986; Verhoeyen *et al.*, 1988; Riechmann *et al.*, 1988).

The results of the coupled partner SDR investigation together with the design of the metal binding site discussed in Chapter 4, indicate that, through careful manipulation of CDR and framework SDRs in the ACS, novel functions may be introduced into this region, while still retaining native antibody specificity and affinity.

The current interest in the production of antibodies with protease, metal binding or other functional properties which are also able to bind to particular (predetermined) antigens, requires a method for the rational design of bi-functional

immunoglobulins. The work on the manipulation of SDRs presented in this chapter and the method for the design of a metalloantibody which forms the basis of this thesis, represent a step closer towards that goal.

Appendix I

A Metalloprotein Survey

The survey was performed by applying the following set of questions to 27 metalloprotein structures from the Brookhaven Protein Databank (PDB) (Bernstein *et al.*, 1977).

1. What are the liganding residues?
2. What are the metal-ligand distances?
3. What is the secondary structure of the protein and that which defines the metal binding site?
4. If the metal binding site is within a loop what is its sequence?
5. What are the nearest neighbour residues to the metal binding site?

The results for each structure are set out below and summarised in Chapter 2. The residue numbers of some of the metal ions were changed to aid manipulation of the metal binding sites with interactive molecular graphics. Liganding mainchain carbonyl oxygens are highlighted in single inverted commas.

I.1 Calcium Metalloproteins

I.1.1 α-Lactalbumin

PDB code: 1ALC (Acharya *et al.*, 1989)
Resolution: 1.70Å (R-factor: 0.220)
Length: 123 residues.
Function: Involved in the synthesis of lactose.

Ligand	Atom	Distance /Å
Lys-79	'O'	2.22
Asp-82	Oδ1	2.36
Asp-84	'O'	2.27
Asp-87	Oδ1	2.32
Asp-88	Oδ1	2.29

The calcium binding site of α-Lactalbumin. The coordination shell is completed by the addition of two water molecules bound to the calcium at distances of 2.40Å and 2.50Å. The overall geometry is a slightly distorted pentagonal bipyramid with carbonyl oxygen atoms at the vertices, while the carboxylate groups and waters are arranged with the waters approximately opposed (Stuart *et al.*, 1986).

Secondary structure: 9 α-helices and 1 β-sheet of 2 strands. The 5 protein ligands come from a tight turn akin to a β-turn with an extra residue, Asp-82, acting as a spacer between two helices (Stuart *et al.*, 1986). The sequence of this turn (which also includes the nearest neighbour residues) is: *79Lys-Phe-Leu-Asp-Asp-Asp-Ile-Thr-Asp-Asp88.*

I.1.2 Thermitase

PDB code: 1TEC (Gros *et al.*, 1989)
Resolution: 2.20Å (R-factor: 0.179)
Length: 279 residues.
Function: Serine proteinase.

Ligand	Atom	Distance /Å
Asp-5[1]	Oδ2	2.37
Asp-47[1]	Oδ1	2.60
	Oδ2	2.48
Val-82[1]	'O'	2.19
Asn-85[1]	Nδ2	2.47
Thr-87[1]	'O'	2.36
Ile-89[1]	'O'	2.42
Asp-57[2]	Oδ2	2.68
Asp-60[2]	Oδ1	2.84
Asp-62[2]	Oδ1	2.56
	Oδ2	2.56
Thr-64[2]	'O'	2.50
Gln-66[2]	Nϵ2	2.56

The calcium binding sites of Thermitase. [1]The residues involved in the coordination of calcium 343. [2]The residues involved in the coordination of calcium 344. The protein also binds a sodium ion, but since it is ligated by 3 carbonyl groups it has not been included in this entry.

Secondary structure: 7 α-helices and 2 β-sheets. The calcium binding sites lie in surface loops at either end of a 10 residue helix formed between His-71 and Ala-81. The sequences of these loops (which also include the nearest neighbour residues) are: **Ca 343:** *82Val-Thr-Asn-Asn-Ser-Thr-Gly-Ile89*; **Ca 344:** *57Asp-Phe-Val-Asp-Asn-Asp-Ser-Thr-Pro-Gln66.*

I.1.3 β-Trypsin

PDB code: 1TPO (Marquart *et al.*, 1983)
Resolution: 1.70Å (R-factor: 0.180)
Length: 223 residues.
Function: Serine proteinase.

Ligand	Atom	Distance /Å
Glu-70	Oϵ1	2.39
Asn-72	'O'	2.26
Val-75	'O'	2.29
Glu-80	Oϵ2	2.28

The calcium binding site of β-Trypsin.

Secondary structure: 3 α-helices. The calcium binding site is located in a turn of an anti-parallel β-strand. The sequence of this turn (which also includes the nearest neighbour residues) is: *70Glu-Asp-Asn-Ile-Asn-Val-Val-Glu-Gly-Asn-Glu80.*

I.1.4 Galactose Binding Protein

PDB code: 2GBP (Vyas *et al.*, 1988)
Resolution: 1.90Å (R-factor: 0.146)
Length: 309 residues.
Function: Periplasmic carbohydrate binding protein.

Ligand	Atom	Distance /Å
Asp-134	Oδ1	2.42
Asn-136	Oδ1	2.37
Asp-138	Oδ2	2.29
Gln-140	'O'	2.28
Gln-142	Oϵ1	2.37
Glu-205	Oϵ1	2.51
	Oϵ2	2.59

The calcium binding site of Galactose binding protein.

Secondary structure: 10 α-helices and 2 six-stranded β-sheets. The calcium binding site is located between helix 5 and strand 2 of sheet 2. The residue sequence which contains the liganding residues (and also the nearest neighbour residues) is: *134Asp-Leu-Asn-Lys-Asp-Gly-Gln-Ile-Gln142.*

I.1.5 Subtilisin Carlsberg

PDB code: 2SEC (McPhalen and James, 1988)
Resolution: 1.80Å (R-factor: 0.136)
Length: 274 residues.
Function: Serine proteinase.

Ligand	Atom	Distance /Å
Gln-2	Oϵ1	2.39
Asp-41	Oδ1	2.35
	Oδ2	2.55
Leu-75	'O'	2.23
Asn-77	Nδ2	2.36
Thr-79	'O'	2.41
Val-81	'O'	2.28

The calcium 1 binding site of Subtilisin Carlsberg. The protein binds a further two calcium ions, but since each is ligated by three carbonyl groups they have not been included in this entry. Both of these ions have two water molecules bound to them.

Secondary structure: 9 α-helices and 2 β-sheets of seven and two strands respectively. The calcium 1 binding site is located in the polypeptide chain immediately following helix C. The Gln-2 ligand is donated by the N-terminus while the Asp-41 ligand comes from a type 1 turn. The sequence of the chain which contains the other four ligands (and also the nearest neighbour residues) is: *75Leu-Asp-Asn-Thr-Thr-Gly-Val81.*

I.1.6 Elastase

PDB code: 3EST (Meyer *et al.*, 1988)
Resolution: 1.65Å (R-factor: 0.169)
Length: 240 residues.
Function: Serine proteinase.

Ligand	Atom	Distance /Å
Glu-70	Oϵ2	2.46
Asn-72	'O'	2.42
Gln-75	'O'	2.41
Asn-77	Oδ1	2.73
Glu-80	Oϵ2	2.39

The calcium binding site of Elastase.

Secondary structure: 2 α-helices and 2 six-stranded β-barrels. The calcium binding site is located in the turn between strands 5 and 6 of the N-terminal β-barrel. The sequence of this turn (which also includes contains the nearest neighbour residues) is: *70Glu-His-Asn-Leu-Asn-Gln-Asn-Asn-Gly-Thr-Glu80.*

I.1.7 Calmodulin

PDB code: 3CLN (Babu *et al.*, 1988)
Resolution: 2.20Å (R-factor: 0.175)
Length: 148 residues.
Function: Calcium binding regulatory protein.

Ligand	Atom	Distance /Å
Asp-20[1]	Oδ2	2.33
Asp-22[1]	Oδ1	2.42
Asp-24[1]	Oδ2	2.61
Thr-26[1]	'O'	2.45
Glu-31[1]	Oϵ1	2.27
	Oϵ2	2.38
Asp-56[2]	Oδ2	2.21
Asp-58[2]	Oδ2	2.48
Asn-60[2]	Oδ1	2.46
Thr-62[2]	'O'	2.17
Glu-67[2]	Oϵ1	2.49
	Oϵ2	2.31
Asp-93[3]	Oδ2	2.14
Asp-95[3]	Oδ2	2.22
Asn-97[3]	Oδ1	2.39
Tyr-99[3]	'O'	2.05
Glu-104[3]	Oϵ1	2.32
	Oϵ2	2.76
Asn-129[4]	Oδ1	2.42
Asp-131[4]	Oδ2	2.37
Asp-133[4]	Oδ2	2.43
Gln-135[4]	'O'	2.38
Glu-140[4]	Oϵ1	2.57
	Oϵ2	2.32

The calcium binding sites of Calmodulin. [1–4]Residues involved in coordinating calcium ions 1–4 respectively. All four metal ions are bound with pentagonal bipyramidal geometry.

Secondary structure: 7 α-helices and 2 β-sheets of 2 strands. The calcium binding sites are located in turns between the helices. This is characteristic of the 'helix-loop-helix' or 'EF-hand' motif. A long central helix separates the two pairs of EF-hands (*c.f.* Troponin C). The sequences of these metal binding turns (which also include all the nearest neighbour residues) are:

Ion	Residue range	Turn sequence											
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Ca 1	20–31	Asp	Lys	Asp	Gly	Asp	Gly	Thr	Ile	Thr	Thr	Lys	Glu
Ca 2	56–67	Asp	Ala	Asp	Gly	Asn	Gly	Thr	Ile	Asp	Phe	Pro	Glu
Ca 3	93–104	Asp	Lys	Asp	Gly	Asn	Gly	Tyr	Ile	Ser	Ala	Ala	Glu
Ca 4	129–140	Asn	Ile	Asp	Gly	Asp	Gly	Gln	Val	Asn	Tyr	Glu	Glu

These sequences are all 12 residues in length. Positions 1, 3, 5, 7, 9 and 12 coordinate the metal ion (position 7 *via* the mainchain carbonyl oxygen) (da Silva and Reinach, 1991). Note also positions 4, 6 and 8.

I.1.8 Calbindin

PDB code: 3ICB (Szebenyi and Moffat, 1986)
Resolution: 2.30Å (R-factor: 0.178)
Length: 75 residues.
Function: Vitamin-D-dependent calcium binding protein.

Ligand	Atom	Distance /Å
Ala-14[1]	'O'	2.16
Glu-17[1]	'O'	2.39
Asp-19[1]	'O'	2.38
Gln-22[1]	'O'	2.21
Glu-27[1]	Oϵ1	2.51
	Oϵ2	2.27
Asp-54[2]	Oδ1	2.38
Asn-56[2]	Oδ1	2.45
Asp-58[2]	Oδ1	2.29
Glu-60[2]	'O'	2.26
Glu-65[2]	Oϵ1	2.29
	Oϵ2	2.61

The calcium binding sites of Calbindin. [1]The residues involved in the coordination of calcium 76. [2]The residues involved in the coordination of calcium 77. Both calcium ions are bound with pentagonal bipyramidal geometry.

Secondary structure: 5 α-helices. The calcium ions are bound in the turns which connect helices 1 and 2 (Ca 76), and helices 3 and 4 (Ca 77). This is characteristic of the 'helix-loop-helix' or 'EF-hand' motif. The sequences of these metal binding loops (which also include all the nearest neighbour residues) are: **Ca 76:** *14Ala-Ala-Lys-Glu-Gly-Asp-Pro-Asn-Gln-Leu-Ser-Lys-Glu-Glu27*; **Ca 77:** *54Asp-Lys-Asn-Gly-Asp-Gly-Glu-Val-Ser-Phe-Glu-Glu65.*

I.1.9 Parvalbumin B

PDB code: 5CPV (Swain *et al.*, 1989)
Resolution: 1.60Å (R-factor: 0.187)
Length: 109 residues.
Function: Chordate white muscle calcium binding protein.

Ligand	Atom	Distance /Å
Asp-90[1]	Oδ2	2.25
Asp-92[1]	Oδ1	2.41
Asp-94[1]	Oδ1	2.44
Lys-96[1]	'O'	2.29
Glu-101[1]	Oϵ1	2.50
	Oϵ2	2.49
Asp-51[2]	Oδ1	2.23
Asp-53[2]	Oδ1	2.36
Ser-55[2]	Oγ	2.60
Phe-57[2]	'O'	2.27
Glu-59[2]	Oϵ1	2.48
Glu-62[2]	Oϵ1	2.55
	Oϵ2	2.43

The calcium binding sites of Parvalbumin B. [1]The residues involved in the coordination of calcium 109. A water molecule is also usually bound to this metal ion. [2]The residues involved in the coordination of calcium 110. Both calcium ions are bound with pentagonal bipyramidal geometry.

Secondary structure: 6 α-helices and 1 β-sheet of 2 strands. The calcium binding sites are located in the turns between helices C and D (Ca 110), and helices E and F (Ca 109). This is characteristic of the 'helix-loop-helix' or 'EF-hand' motif. The sequences of these metal binding loops (which also include all the nearest neighbour residues) are:

Ion	Residue range	Turn sequence											
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Ca 109	90–101	Asp	Ser	Asp	Gly	Asp	Gly	Lys	Ile	Gly	Val	Asp	Glu
Ca 110	51–62	Asp	Gln	Asp	Lys	Ser	Gly	Phe	Ile	Glu	Glu	Asp	Glu

These sequences are both 12 residues in length. Positions 1, 3, 5, 7, 9 and 12 coordinate the metal ion (position 7 *via* the mainchain carbonyl oxygen) (da Silva and Reinach, 1991). Note also positions 6, 8 and 11.

I.1.10 Troponin C

PDB code: 5TNC (Herzberg and James, 1988)
Resolution: 2.00Å (R-factor: 0.155)
Length: 162 residues.
Function: Calcium regulation of muscle contraction.

Ligand	Atom	Distance /Å
Asp-106[1]	Oδ2	2.39
Asn-108[1]	Oδ1	2.44
Asp-110[1]	Oδ1	2.26
Phe-112[1]	'O'	2.13
Glu-117[1]	Oϵ1	2.24
	Oϵ2	2.57
Asp-142[2]	Oδ2	2.38
Asn-144[2]	Oδ1	2.28
Asp-146[2]	Oδ1	2.50
Arg-148[2]	'O'	2.45
Glu-153[2]	Oϵ1	2.46
	Oϵ2	2.37

Only two of the four calcium binding sites in this structure are occupied and are shown in this Table. [1]The residues involved in the coordination of calcium 163. [2]The residues involved in the coordination of calcium 164. Both calcium ions are bound with pentagonal bipyramidal geometry.

Secondary structure: 10 α-helices. The calcium binding sites are located in the turns between helices E and F (Ca 163), and helices G and H (Ca 164). This is characteristic of the 'helix-loop-helix' or 'EF-hand' motif. A long central helix separates the two pairs of EF-hands (*c.f.* Calmodulin). The sequences of these metal binding loops (which also include all the nearest neighbour residues) are:

Ion	Residue range	Turn sequence											
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Ca 163	106–117	Asp	Lys	Asn	Ala	Asp	Gly	Phe	Ile	Asp	Ile	Glu	Glu
Ca 164	142–153	Asp	Lys	Asn	Asn	Asp	Gly	Arg	Ile	Asp	Phe	Asp	Glu

These sequences are both 12 residues in length. Positions 1, 3, 5, 7, 9 and 12 coordinate the metal ion (position 7 *via* the mainchain carbonyl oxygen) (da Silva and Reinach, 1991). Note also positions 6, 8 and 11.

I.2 Magnesium Metalloproteins

I.2.1 D-Xylose Isomerase

PDB code: 5XIA (Henrick *et al.*, 1989)
Resolution: 2.50Å (R-factor: 0.135)
Length: Subunit A; 393 residues.
Function: Conversion of D-Xylose to L-Xylose.

Ligand	Atom	Distance /Å
Glu-180	Oϵ2	2.22
Glu-216	Oϵ1	2.01
Asp-244	Oδ2	2.07
Asp-292	Oδ1	2.17

The magnesium binding site of Xylose Isomerase. Two further ligands are donated by the substrate giving overall hexacoordinate geometry to the metal ion.

Secondary structure: Each subunit consists of 15 α-helices and an 8-stranded β-barrel. The magnesium binding site lies at the C-terminal end of strands 5, 6, 7, and 8 in the β-barrel.

Nearest neighbours: Trp-15 and 136, Asn-214, His-219, Asn-246, Asp-256 and His-290.

I.3 Manganese Metalloproteins

I.3.1 Glutamine Synthetase

PDB code: 2GLS (Yamashita *et al.*, 1989)
Resolution: 3.50Å (R-factor: 0.258)
Length: Subunit A; 469 residues.
Function: ATP-driven amidation of glutamate with ammonium ions.

Ligand	Atom	Distance /Å
Glu-131[1]	Oϵ2	2.32
Glu-212[1]	Oϵ2	1.93
Glu-220[1]	Oϵ1	2.18
Glu-129[2]	Oϵ2	1.55
His-269[2]	Nδ1	2.11
Glu-357[2]	Oϵ2	1.93

The manganese binding site of Glutamine Synthetase. [1]The residues involved in the coordination of manganese ion 470. A water molecule is also bound to this ion at a distance of 1.92Å giving overall four-coordinate ligation. [2]The residues involved in the coordination of manganese ion 471.

Secondary structure: Glutamine Synthetase is a dodecamer consisting of 12 subunits. Each subunit consists of 15 α-helices and 3 β-sheets, two of 2 strands, and an inter-subunit sheet of 8 strands. The manganese binding sites are 5.83Å apart and lie in the inter-subunit sheet. Residues Glu-129 and Glu-131 are separated by Pro-130. Residues Glu-212 and Glu-220 define the turn between strands 3 and 4 in the inter-subunit sheet and approximately half of the length of strand 4. This has the sequence: *212Glu-Val-Ala-Thr-Ala-Gly-Gln-Asn-Glu220.*

Nearest neighbours: *Mn 470*: Glu-129, Tyr-179, His-210, Gln-218 and His-269; *Mn 471*: Glu-220, His-271, Arg-344 and Arg-359.

I.3.2 Pea Lectin

PDB code: 2LTN (Prasthofer *et al.*, 1989)
Resolution: 1.70Å (R-factor: 0.177)
Additional metal ion: Calcium.
Length: Subunit A; 181 residues.
Function: Carbohydrate binding protein.

Ligand	Atom	Distance /Å
Glu-119[1]	Oϵ1	2.22
Asp-121[1]	Oδ2	2.07
Asp-129[1]	Oδ1	2.19
His-136[1]	Nϵ2	2.24
Asp-121[2]	Oδ1	2.47
	Oδ2	2.66
Phe-123[2]	'O'	2.30
Asn-125[2]	Oδ1	2.23
Asp-129[2]	Oδ2	2.41

The manganese and calcium binding sites of Pea Lectin. [1]The residues involved in coordinating the manganese ion. [2]The residues involved in coordinating the calcium ion.

Secondary structure: The Pea Lectin molecule normally exists as a dimer. Each monomer consists of two separate polypeptide chains, *alpha* and *beta*. The *alpha* chain, containing the metal binding sites, consists of 181 residues and the *beta* chain consists of 52 residues. The *alpha* chain secondary structure consists exclusively of β-sheet. The metal binding sites are 4.24Å apart and lie in a large loop which forms the turn of an anti-parallel β-strand on the surface of the structure. The sequence of this loop is: *119Glu-Phe-Asp-Thr-Phe-Tyr-Asn-Ala-Ala-Trp-Asp-Pro-Ser-Asn-Arg-Asp-Arg-His136* (*c.f.* Concanavalin A).

Nearest neighbours: *Manganese*: Leu-101, Asn-125, Trp-128, Pro-130, Asp-140, Ile-144, Ser-146 and Thr-149; *Calcium*: Asp-81, Leu-101, Glu-119, Thr-122, Phe-123, Tyr-124, Trp-128, His-136 and Ile-144.

I.3.3 Concanavalin A

PDB code: 3CNA (Hardman and Ainsworth, 1972)
Resolution: 2.40Å (R-factor: None given.)
Additional metal ion: Calcium.
Length: 237 residues.
Function: Carbohydrate binding protein.

Ligand	Atom	Distance /Å
Asp-10[1]	Oδ2	2.09
Tyr-12[1]	'O'	1.96
Asn-14[1]	Oδ1	2.04
Asp-19[1]	Oδ2	2.26
Asp-208[1]	Oδ2	4.47
Glu-8[2]	Oϵ2	2.68
Asp-10[2]	Oδ1	2.67
Asp-19[2]	Oδ1	2.32
His-24[2]	Nϵ2	2.07
Ser-34[2]	Oγ	3.96

The calcium and manganese binding sites of Concanavalin A. [1]The residues involved in coordinating the calcium ion. Asp-208 coordinates the metal ion *via* a bridging water molecule. [2]The residues involved in coordinating the manganese ion. Ser-34 coordinates the metal ion *via* a bridging water molecule. Two additional water molecules are bound to each of the metal ions. Other residues may also indirectly coordinate the metal ions *via* the water molecules.

Secondary structure: Concanavalin A is exclusively β-sheet in structure consisting of 3 sheets I, II, and III of 6, 7 and 5 strands respectively. The metal binding sites are 4.46Å apart and are located in sheet II in a cavity near the surface of the molecule. All the liganding residues are donated by this sheet and, with the exception of Ser-34 and Asp-208 which also coordinate through bridging water molecules, the remaining ligands are contained within the sequence: *8Glu-Leu-Asp-Thr-Tyr-Pro-Asn-Thr-Asp-Ile-Gly-Asp-Pro-Ser-Tyr-Pro-His24* (*c.f.* Pea Lectin).

Nearest neighbours: *Calcium*: Tyr-12, Ile-17, Arg-228 and Leu-230; *Manganese*: Ile-17, Val-32, Thr-37 and Arg-228.

I.4 Copper Metalloproteins

I.4.1 Azurin

PDB code: 1AZU (Adman and Jensen, 1981)
Resolution: 2.70Å (R-factor: 0.350)
Length: 128 residues.
Function: Electron transport.

Ligand	Atom	Distance /Å
Gly-45	'O'	2.47
His-46	Nδ1	2.15
Cys-112	Sγ	1.79
His-117	Nδ1	2.42

The copper binding site of Azurin. The ligands are arranged in a highly distorted tetrahedron with the sidechains of His-46, Cys-112 and His-117 almost planar. The carbonyl group of Gly-45 is at the apex. (*c.f.* Azurin (2AZA)).

Secondary structure: The predominant secondary structure is β-sheet. The copper binding site lies in a cavity at the C-terminal end of the molecule. The ligands Cys-112 and His-117 define a six residue surface loop which has the sequence: *112Cys-Thr-Phe-Pro-Gly-His117.*

Nearest neighbours: Asn-3 and 47, His-35, Leu-86, Met-13,44 and 121, and Phe-15 and 114.

I.4.2 Pseudoazurin

PDB code: 1PAZ (Petratos *et al.*, 1988)
Resolution: 1.55Å (R-factor: 0.180)
Length: 123 residues.
Function: Electron transport.

Ligand	Atom	Distance /Å
His-40	Nδ1	2.16
Cys-78	Sγ	2.16
His-81	Nδ1	2.12
Met-86	Sδ	2.76

The copper binding site of oxidized Pseudoazurin. The ligands are arranged in a highly distorted tetrahedron with the sidechains of His-40, Cys-78 and His-81 almost planar. The Sδ atom of Met-86 is at the apex.

Secondary structure: 2 α-helices and an eight-stranded β-barrel. The copper binding site lies within the β-barrel. Cys-78 and His-81 define a type III reverse turn. The sequence between Cys-78 and Met-86 which includes this turn is: *78Cys-Thr-Pro-His-Tyr-Ala-Met-Gly-Met86.*

Nearest neighbours: Met-7 and 16, Phe-18, Gly-39 and Asn-41.

I.4.3 Azurin

PDB code: 2AZA (Baker, 1988)
Resolution: 1.80Å (R-factor: 0.157)
Length: 129 residues.
Function: Electron transport.

Ligand	Atom	Distance /Å
His-46	Nδ1	2.08
Cys-112	Sγ	2.12
His-117	Nδ1	2.01
Met-121	Sδ	3.12

The copper binding site of oxidized Azurin. The ligands are arranged in a highly distorted tetrahedron with the sidechains of His-46, Cys-112 and His-117 almost planar. The Sδ atom of Met-121 is at the apex. The carbonyl group of Gly-45 at a distance of 3.16Å may also interact with the copper ion (*c.f.* Azurin (1AZU)).

Secondary structure: 2 α-helices and 2 β-sheets, of 4 and 5 strands respectively. The copper binding site is located in a cavity at the C-terminal end of the molecule. The strand containing Cys-112, His-117 and Met-121 has the sequence: *112Cys-Ser-Phe-Pro-Gly-His-Trp-Ala-Met-Met121.*

Nearest neighbours: Ser-9, Met-13, Tyr-15, His-35, Met-44, Gly-45, Asn-47, Val-86, Phe-114, Pro-115 and Gly-116.

I.4.4 Superoxide Dismutase

PDB code: 2SOD (Tainer *et al.*, 1982)
Resolution: 2.00Å (R-factor: 0.256)
Additional metal ion: Zinc.
Length: Subunit O; 152 residues.
Function: The dismutation of superoxide into hydrogen peroxide and water.

Ligand	Atom	Distance /Å
His-44[1]	Nδ1	2.01
His-46[1]	Nϵ2	2.11
His-61[1]	Nϵ2	2.21
His-118[1]	Nϵ2	2.10
His-61[2]	Nδ1	2.09
His-69[2]	Nδ1	2.14
His-78[2]	Nδ1	2.04
Asp-81[2]	Oδ1	1.91

The copper and zinc binding sites of Superoxide Dismutase. The enzyme can be regarded as a heterodinuclear coordination compound. The two metal ions share a common bridging imidiazole residue from the sidechain of His-61. [1]The coordination around the copper ion is a distorted square pyramid, with a bound water molecule 3.20Å at the apex. [2]The coordination around the zinc ion is pseudotetrahedral.

Secondary structure: Superoxide Dismutase contains four subunits designated O, Y, B and G. Each subunit consists of an eight-stranded anti-parallel β-barrel, with 'Greek-key' topology similar to an immunoglobulin domain. The metal binding sites are 6.26Å apart and are situated at the N-terminal end of the subunit, close to the surface. Ligands His-44 and His-46 are separated by Val-45. The other ligands, except for His-118, come from a strand of polypeptide which has the sequence: *61His-Phe-Asn-Pro-Leu-Ser-Lys-Lys-His-Gly-Gly-Pro-Lys-Asp-Glu-Glu-Arg-His-Val-Gly-Asp81.*

Nearest neighbours: *Copper*: Val-116, Asp-122, Thr-135 and Ala-138; *Zinc*: Lys-68, Arg-77, Asp-122, Lys-134 and Thr-135.

I.4.5 Plastocyanin

PDB code: 5PCY (Guss *et al.*, 1986)
Resolution: 1.80Å (R-factor: 0.160)
Length: 99 residues.
Function: Electron transport.

Ligand	Atom	Distance /Å
His-37	Nδ1	2.13
Cys-84	Sγ	2.17
His-87	Nδ1	2.39
Met-92	Sδ	2.87

The copper binding site of reduced Plastocyanin. The ligands are arranged in a highly distorted tetrahedron with bond angles differing by up to 50°. The side-chains of His-37, Cys-84 and His-87 are almost planar. Met-92 Sδ is at the apex.

Secondary structure: 1 helical turn and 2 β-sheets both of four strands. The copper binding site is located in sheet II. His-37 comes from the N-terminal end of strand 4; Cys-84 and His-87 define a turn at the end of strand 3; and Met-92 comes from the N-terminal end of strand 2. Ligands 84, 87 and 92 are contained in the sequence: *84Cys-Ser-Pro-His-Gln-Gly-Ala-Gly-Met92.*

Nearest neighbours: Leu-5 and 12, Ala-13, Phe-14 and 35, Pro-36 and Asn-38.

I.5 Iron Metalloproteins

I.5.1 Hemerythrin

PDB code: 1HMQ (Stenkamp *et al.*, 1983)
Resolution: 2.00Å (R-factor: 0.173)
Length: Subunit A; 113 residues.
Function: Oxygen transport.

Ligand	Atom	Distance /Å
Glu-58[1]	Oϵ1	2.28
His-73[1]	Nϵ2	2.23
His-77[1]	Nϵ2	2.16
His-101[1]	Nϵ2	2.20
Asp-106[1]	Oδ1	2.06
His-25[2]	Nϵ2	2.20
His-54[2]	Nϵ2	2.19
Glu-58[2]	Oϵ2	2.04
Asp-106[2]	Oδ2	2.09

The iron binding site of Hemerythrin. [1]Iron 1 has hexacoordinate ligation. [2]Iron 2 has pentacoordinate ligation.

Secondary structure: Hemerythrin consists of 4 subunits (A–D) of equal size. Each subunit consists of a four helix bundle containing two iron atoms. The iron binding sites are 3.23Å apart and are created by the donation of ligands from each helix into the central cavity. The sequences *54His-Phe-Leu-Asn-Glu58*, *73His-Lys-Lys-Ala-His77* and *101His-Ile-Lys-Thr-Ile-Asp106* represent one turn of helices 2, 3 and 4 respectively.

Nearest neighbours: Phe-55, Met-62, Trp-97, Ile-102 and Tyr-109.

I.5.2 Rubredoxin

PDB code: 5RXN (Watenpaugh *et al.*, 1980)
Resolution: 1.20Å (R-factor: 0.096)
Length: 54 residues.
Function: Electron transport.

Ligand	Atom	Distance /Å
Cys-6	Sγ	2.33
Cys-9	Sγ	2.29
Cys-39	Sγ	2.31
Cys-42	Sγ	2.25

The tetrahedral iron-sulphur site of oxidized Rubredoxin.

Secondary structure: Random coil. The iron-sulphur site is located at one end of the molecule. Pairs of ligands are donated by the N and C-terminal regions which lie adjacent to each other in the structure. Each ligand pair defines a 4 residue loop with sequences *6Cys-Thr-Val-Cys9* and *39Cys-Pro-Leu-Cys42*.

Nearest neighbours: Val-8, Gly-10, Tyr-11, Pro-40, Leu-41, Gly-43, Val-44 and Phe-49. Virtually all of the nearest neighbour residues around the iron-sulphur site are hydrophobic. Those with charged or polar groups have the hydrophobic (carbon) portion of their sidechains orientated towards the site. The metal binding site is thus predominantly hydrophobic.

I.6 Zinc Metalloproteins

I.6.1 Tonin

PDB code: 1TON (Fujinaga and James, 1987)
Resolution: 1.80Å (R-factor: 0.196)
Length: 235 residues.
Function: Zinc-inhibited serine proteinase.

Ligand	Atom	Distance /Å
His-57	Nϵ2	2.04
His-97	Nϵ2	2.05
His-99	Nϵ2	2.05

The zinc binding site of Tonin. The ligands are arranged tetrahedrally around the zinc ion.

Secondary structure: 4 α-helices and 4 β-sheets. The sheets are divided into pairs, each pair forming a β-sandwich. Ligands 97 and 99 are separated by Asp-98. The liganding residues include the active-site histidine. Coordination of a zinc ion causes inhibition of the enzyme which does not have a native zinc binding site.

Nearest neighbours: Tyr-94, Pro-95, Asp-98, Ser-214, Gly-215 and Ala-217.

I.6.2 Carbonic Anhydrase B

PDB code: 2CAB (Kannan *et al.*, 1984)
Resolution: 2.00Å (R-factor: 0.193)
Length: 261 residues.
Function: The reversible hydration of carbon dioxide.

Ligand	Atom	Distance /Å
His-94	Nϵ2	1.93
His-96	Nϵ2	1.94
His-119	Nδ1	1.90

The zinc binding site of Carbonic Anhydrase B. The ligands are arranged in a distorted tetrahedron around the zinc ion. The free fourth coordination position is usually occupied by a water molecule.

Secondary structure: 7 α-helices and 1 β-sheet of 10 strands. The molecule is roughly ellipsoidal ($40 \times 45 \times 55$Å), with the zinc binding site lying near the bottom of a deep cleft near the centre of the molecule. The ligands His-94 and His-96 are separated by Phe-95.

Nearest neighbours: Thr-199 (involved in the hydration-dehydration mechanism), His-67, Gln-92, Glu-117, Val-143 and Trp-209.

I.6.3 Thermolysin

PDB code: 3TLN (Holmes and Matthews, 1982)
Resolution: 1.60Å (R-factor: 0.213)
Additional metal ion: Calcium.
Length: 316 residues.
Function: Metalloproteinase.

Ligand	Atom	Distance /Å
His-142	Nϵ2	2.11
His-146	Nϵ2	2.08
Glu-166	Oϵ1	2.08
Asp-138[1]	Oδ1	2.46
Glu-177[1]	Oϵ1	2.53
	Oϵ2	2.87
Asp-185[1]	Oδ2	2.52
Glu-187[1]	'O'	2.29
Glu-190[1]	Oϵ1	2.40
	Oϵ2	2.51
Glu-177[2]	Oϵ2	2.42
Asn-183[2]	'O'	2.37
Asp-185[2]	Oδ1	2.43
Glu-190[2]	Oϵ2	2.40
Asp-57[3]	Oδ1	2.23
	Oδ2	2.71
Asp-59[3]	Oδ1	2.33
Gln-61[3]	'O'	2.18
Tyr-193[4]	'O'	2.43
Thr-194[4]	Oγ1	2.48
	'O'	2.54
Ile-197[4]	'O'	2.31
Asp-200[4]	Oδ1	2.24

The metal binding sites of Thermolysin. Residues His-142, His-146 and Glu-166 coordinate the zinc ion. [1–4]Residues involved in coordinating calcium ions 1–4 respectively. These metal ions are responsible for maintaining the quaternary structure of the enzyme.

Secondary structure: 12 α-helices and 3 β-sheets of 5 stands each. The residues forming the zinc binding site are donated by helices 3 and 4. His-142 and His-146 define a helical turn which has the sequence: *142His-Glu-Leu-Thr-His146.* The 4 calcium binding sites are located in loops close to the surface of the molecule, between α-helices (Ca 1, 2 and 4) or β-strands (Ca 3 in N-terminal sheet 3). The sequences of these turns are: **Ca 1 and 2:** *177Glu-Phe-Tyr-Ala-Asn-Lys-Asn-Pro-Asp-Trp-Glu-Ile-Gly-Glu-Asp-Val192.* Glu-177 to Ala-180 form the last turn of helix 4 and Glu-190 to Val-192 form the single turn of helix 5. These sites are 3.83Å apart; **Ca 3:** *57Asp-Ala-Asp-Asn-Gln61*; **Ca 4:** *193Tyr-Thr-Pro-Gly-Ile-Ser-Gly-Asp200.*

Nearest neighbours (Zinc site): Glu-143 and His-231 (involved in catalysis), Tyr-157, Asn-165, Ser-169, Asp-170 and Arg-203.

I.6.4 Carboxypeptidase A

PDB code: 5CPA (Rees *et al.*, 1983)
Resolution: 1.54Å (R-factor: 0.190)
Length: 307 residues.
Function: Metalloproteinase.

Ligand	Atom	Distance /Å
His-69	Nδ1	2.13
Glu-72	Oϵ1	2.18
	Oϵ2	2.31
His-196	Nδ1	2.07

The zinc binding site of Carboxypeptidase A. The zinc is coordinated with a distorted tetrahedral stereochemistry. A water molecule or hydroxide ion may also be bound to the zinc.

Secondary structure: 8 α-helices and 1 β-sheet of 8 strands. The molecule is roughly elipsoidal (50 × 40 × 38Å). His-69 and Glu-72 come from the end of a 4 residue turn at the beginning of helix 2. The sequence of this turn is *69His-Ser-Arg-Glu72*. His-196 comes from strand 5 of the β-sheet. The zinc binding site lies at one end of a large central cleft.

Nearest neighbours: Glu-270 (activates the attacking water molecule during catalysis), Ile-68, Arg-127, Asp-142, Asn-144, Ser-197 and Phe-279.

I.6.5 Aspartate Carbamoyltransferase

PDB code: 6AT1 (Stevens *et al.*, 1990)
Resolution: 2.60Å (R-factor: 0.160)
Length: Regulatory subunit B; 153 residues.
Function: Involved in the formation of carbamoyl aspartate.

Ligand	Atom	Distance /Å
Cys-109	Sγ	2.32
Cys-114	Sγ	2.34
Cys-138	Sγ	2.35
Cys-141	Sγ	2.32

The tetrahedral zinc site of Aspartate Carbamoyltransferase regulatory subunit B.

Secondary structure: 3 α-helices and 2 β-sheets of 5 and 4 strands respectively. The zinc binding site is located in the 4-stranded sheet between the turns of strands 1 and 2 (Cys-109 to Cys-114), and strands 3 and 4 (Cys-138 to Cys-141). The sequences of these turns are: *109Cys-Pro-Asn-Ser-Asn-Cys114*; and *138Cys-Lys-Tyr-Cys141.*

Nearest neighbours: Ile-115, Ser-116, His-117, Phe-125, Lys-143 and Phe-145.

I.6.6 Alcohol Dehydrogenase

PDB code: 7ADH (Plapp *et al.*, 1983)
Resolution: 3.20Å (R-factor: 0.290)
Length: 374 residues.
Function: NAD-dependent reversible oxidation of alcohols to aldehydes.

Ligand	Atom	Distance /Å
Cys-46[1]	Sγ	2.16
His-67[1]	Nϵ2	2.11
Cys-174[1]	Sγ	2.07
Cys-97[2]	Sγ	2.42
Cys-100[2]	Sγ	2.04
Cys-103[2]	Sγ	2.11
Cys-111[2]	Sγ	1.95

The zinc binding sites of Alcohol Dehydrogenase. [1]The residues involved in the coordination of the active site zinc ion. A water molecule usually completes the coordination shell to give a distorted tetrahedral stereochemistry. [2]The residues involved in the coordination of the structural zinc ion.

Secondary structure: 10 α-helices, 4 β-sheets of 6(S1), 4(S2), 6(S3) and 6(S4) strands. The enzyme consists of two identical subunits of Mw $\approx$ 40 000. The active site zinc is situated at the bottom of a hydrophobic pocket in a deep cleft about 25Å from the protein surface. The structural zinc is close to the surface of the enzyme, but not exposed to solvent. The binding site for this ion is reminiscent of the catalytic centre of FeS_4 proteins (*c.f.* Rubredoxin). The ligands which form this site are contained within the sequence: *97Cys-Gly-Lys-Cys-Arg-Val-Cys-Lys-His-Pro-Glu-Gly-Asn-Phe-Cys111.*

Nearest neighbours: *Active site zinc*: Ser-48 (accepts a proton from the zinc-bound water molecule), Arg-47, Asp-49, Glu-68, Phe-93 and 140, and Thr-178; *Structural zinc*: Pro-95, Gln-96, Lys-99, Val-102, Leu-112, Lys-113 and Ile-155.

Appendix II

Standard Techniques in Molecular Biology

II.1 Materials

- All restriction and DNA modifying enzymes were purchased from Boehringer Mannheim.
- All radiochemicals were purchased from Amersham International p.l.c.
- Chemical reagents (analytical grade) were obtained from Sigma, Fisons and BDH.
- Agar no. 1, Tryptone and Yeast extract were purchased from Difco. Antibiotics were purchased from Sigma.
- All media and antibiotic stock solutions were prepared as described in Sambrook *et al.*, (1989).

II.2 Biochemical Techniques

II.2.1 Agarose Gel Electrophoresis

50ml gels were prepared by dissolving 0.8–1% w/v agarose (Type II) in electrophoresis buffer (1× TBE: 89mM Tris-borate; 89mM boric acid and 2mM EDTA) by boiling, then allowed to cool (60–70° C). Ethidium bromide was added to a concentration of 0.5μg/ml and the mixture poured into a horizontal gel tank. Once set, the comb was removed and the gel submerged in 1× TBE. DNA samples were taken up in gel loading buffer (3.5% w/v Ficoll; 0.1% w/v bromophenol blue; 0.1% w/v xylene cyanol and 4mM EDTA) and electrophoresed at 50 volts for 30–60 minutes. The DNA was visualised by fluorescence with short-wavelength ultraviolet (UV) light.

II.2.2 Denaturing Polyacrylamide Gel Electrophoresis

7.5ml of a 40% solution of acrylamide (38% w/v acrylamide and 2% w/v bis-acrylamide (N,N′-methylene bis-acrylamide) filtered through a 0.45 micron membrane and stored in the dark at 4° C (40% stock acrylamide)) was diluted in a solution of 25g urea dissolved in 5ml 1× TBE and 17.5ml Milli-Q water. Polymerisation was initiated by addition of 300μl 10% ammonium persulphate and 30μl TEMED (N,N,N;N′-tetramethyl-ethylene-diamine) to each 50ml of solution. After mixing, the solution was poured between sealed plates of dimensions 400×200mm and the comb clamped into position. The gel was allowed to set in a horizontal position for 30–60 minutes.

Before sample electrophoresis, the gel was pre-electrophoresed in 1× TBE buffer for at least 5 minutes at 30 watts. The DNA samples were taken up in 80% v/v deionised formamide; 1× TBE; 0.1% w/v bromophenol blue and 0.1% w/v xylene cyanol, incubated at 75° C for 2 minutes then loaded immediately into pre-washed wells. A thermostating metal plate was clamped onto the gel plates to equilibrate the heat generated during electrophoresis, which was performed at 25–30 watts (1200–1500 volts). After electrophoresis, the plates were separated and the gel fixed with 500ml 10% acetic acid; 10% methanol before drying under vacuum on Whatman No. 1 filter paper at 80° C for 30–60 minutes. The autoradiograph was exposed for 24–48 hours.

II.2.3 Reducing SDS Polyacrylamide Gel Electrophoresis

SDS-PAGE was performed essentially as described by Laemmli (1970). The running gel was prepared by diluting acrylamide stock (40% w/v acrylamide; 1% w/v bis-acrylamide) and SDS buffer (0.75M Tris-HCl, pH 8.8; 2.5% w/v SDS) with water. Polymerisation was initiated by the addition of 2.5ml 1.5% w/v ammonium persulphate and 500μl 5% v/v TEMED, the gel poured to within 4cm of the top of the gel plates and allowed to set. A stacking gel was prepared in 0.15M Tris-HCl, pH6.8; 2.5% w/v SDS and polymerised with 1ml 1.5% w/v ammonium persulphate and 200μl 5% v/v TEMED. The stacking gel was poured on top, the comb inserted and the gel allowed to set.

Samples were taken up in 20μl of dissociation mix (10mM Tris-HCl, pH6.9; 25% v/v glycerol; 0.02% w/v bromophenol blue; 10% w/v SDS and 10% v/v β-mercaptoethanol. Electrophoresis was performed in 1× running buffer[1] at 50 volts overnight. For autoradiography, the gel was dried under vacuum and the film presensitised before exposure which was carried out at -70° C.

[1]The running buffer was prepared by dissolving 21.6g glycine; 4.5g Tris-HCl, pH8.3 and 1.5g SDS in 1.5 litres of water.

II.2.4 Preparation of Modified Barth's Saline

Modified Barth's Saline (MBS) is made from the following solutions:

Solution A: 128g NaCl; 2g KCl; 5g $NaHCO_3$; 45g Tris base. Made up in water, pH adjusted to 7.6 with concentrated HCl, final volume 1 litre. Stored as 40ml aliquots at -20° C.

Solution B: 1.9g $CaNO_3{:}4H_2O$; 1.54g $CaCl_2{:}6H_2O$. Dissolved in 1 litre of water and stored as 40ml aliquots at -20° C.

Solution C: 5g $MgSO_4{:}7H_2O$. Dissolved in 1 litre of water and stored as 40ml aliquots at -20° C.

Solution D: Sodium Penicillin and Streptomycin Sulphate each at 10mg/ml. 8ml stored at -20° C.

MBS was prepared by mixing 40ml of solutions A, B and C, and 1ml of solution D in a total volume of 1 litre. The final composition of MBS is:-
88.0mM NaCl,
1.0mM KCl,
2.4mM $NaHCO_3$,
15.0mM Tris-HCl, pH7.6,
0.3mM $CaNO_3{:}4H_2O$,
0.41mM $CaCl_2{:}6H_2O$,
0.82mM $MgSO_4{:}7H_2O$,
10μg/ml Sodium Penicillin,
10μg/ml Streptomycin Sulphate.

II.3 Microbiological Techniques

Stocks of pSP6G2H5-L1CEC were prepared for long term storage at -70° C by mixing 0.9ml of a fresh overnight culture with 0.6ml of sterile 80% glycerol in cryo-vials.

II.3.1 Large Scale Preparation of M13 DNA

A 200ml culture was prepared by innoculating media with 200μl of an overnight culture and incubating at 37° C, 250rpm overnight. The cells were pelleted by centrifugation using a GSA rotor in a Sorval RC-5 centrifuge, at 5 000rpm for 20 minutes. The cells were resuspended in 3ml of *Solution 1* (50mM Glucose; 25mM Tris-HCl, pH8.0; 10mM EDTA), transferred to a 30ml Corex tube containing 15mg lysozyme and incubated at 37° C for 5–10 minutes. The cells were completely lysed by the addition of 6ml fresh *Solution 2* (1ml 2M NaOH; 1ml 10% w/v SDS; 8ml

Milli-Q water), mixed gently and incubated on ice for 5 minutes. 5ml of *Solution 3*[2] was added, mixed and the solution incubated on ice for a further 5 minutes.

After centrifugation in an SS34 rotor at 9 000rpm for 10 minutes, the supernatant, (containing the nucleic acids) was poured off and extracted once with 0.5 v/v of phenol:chloroform.[3] Following a second centrifugation at 9 000rpm for 5 minutes, the nucleic acids were precipitated from the aqueous phase by the addition of 0.6 v/v isopropanol and incubating at room temperature for 15 minutes. The nucleic acids were pelleted by centrifugation at 9 000rpm for 10 minutes, washed in 100% ethanol (to remove all traces of isopropanol) and left to dry. The pellet was resuspended in 3ml TE buffer (0.1M Tris-HCl, pH8.0; 0.01M EDTA), 12μl RNase A (10mg/ml) added and the solution incubated at room temperature for 30 minutes.

The solution was extracted once with 0.5 v/v phenol and the DNA precipitated with 0.1 v/v 3M sodium acetate; 2.5 v/v 70% ethanol at -70° C for 30 minutes. The DNA was pelleted, resuspended in 300μl of 0.3M sodium acetate, pH5.2 and transferred to a microfuge tube. 800μl of 70% ethanol was added and precipitation carried out at -20° C overnight. The DNA was purified by gel filtration.

II.3.2 Gel Filtration of M13 DNA on Sephacryl-200

The free ribonucleotides generated by RNase treatment during preparation of the DNA, were removed by filtration down a 10ml column of Sephacryl-200 (Pharmacia) pre-equilibrated with sephacryl running buffer (0.2M NaCl; 10mM Tris-HCl, pH8.0; 0.5mM EDTA). The precipitated DNA was collected by centrifugation, the pellet resuspended in 200μl of running buffer and loaded onto the column. 30 fractions were collected and 3μl from each fraction was spotted onto an 0.8% w/v agarose mini-gel containing ethidium bromide (0.5μg/ml).

After drying at room temperature, the fractions containing DNA were identified by UV fluorescence. These fractions were pooled and the DNA precipitated with 0.1 v/v 3M sodium acetate, pH5.2 and 2.5 v/v 70% ethanol at -70° C for 30 minutes. Following centrifugation, the pellet was washed with 100% ethanol and resuspended in 100μl Milli-Q water.

II.3.3 Preparation of Competent MC1061 Cells

A 10ml overnight culture of the MC1061 strain was prepared and 400μl used to inoculate 40ml LB media. The culture was grown at 37° C, 250rpm, until an OD_{600nm} $\approx$ 0.24 units was reached. The cells were cooled on ice for 10 minutes, then pelleted

---

[2] *Solution 3* was prepared as follows: To 60ml of 5M potassium acetate, 11.5ml of glacial acetic acid and 28.5ml of (Milli-Q) water were added. The solution was mixed by inversion and left on ice for 10 minutes.

[3] Phenol:chloroform was prepared by dissolving 800g of phenol and 5g of hydroxyquinone in 100ml chloroform. The solution was saturated with a low salt buffer (10mM Tris-HCl, pH7.0; 0.1M NaCl) and stirred for 12 hours until the layers separated.

at 2 000rpm in an MSE centrifuge at 4° C for 10 minutes. The cells were resuspended gently in 10ml of sterile, ice-cold 50mM $CaCl_2$ and incubated on ice for 20 minutes. The cells were pelleted, resuspended in 1.5ml 50mM $CaCl_2$ and incubated at 4° C overnight prior to transformation.

II.3.4 Small Scale Preparation of Plasmid DNA

An isolated colony was picked and used to inoculate 1.5ml of liquid media containing antibiotic, in 10ml culture tubes. After overnight growth at 37° C 250rpm, the cultures were transferred to microfuge tubes and centrifuged for 5 minutes. The infected cells remaining in the culture tubes were stored at 4° C. The pelleted cells were resuspended in 200μl of *Solution 1*, 50μl lysozyme (5mg/ml) added and incubated at room temperature for 5 minutes. The cells were completely lysed by the addition of 500μl of fresh *Solution 2*, mixed gently and incubated on ice for 5 minutes. 375μl of ice-cold *Solution 3* was added, mixed and the solution incubated on ice for a further 5 minutes.

After centrifugation, the supernatant (containing the nucleic acids) was transferred to a fresh microfuge tube and extracted once with 500μl phenol:chloroform. The nucleic acids were precipitated with 600μl isopropanol and incubated at room temperature for 25 minutes. Following centrifugation, the pellet was washed with 50μl 100% ethanol and resuspended in 200μl TE buffer. 2μl RNase A (10mg/ml) was added and the solution incubated at 37° C for 30 minutes.

The solution was extracted once with 0.5 v/v phenol:chloroform and once with 0.5 v/v ether (water equilibrated). The DNA was precipitated with 0.1 v/v sodium acetate, 2.5 v/v 70% ethanol and incubated on dry ice for 30 minutes. After centrifugation, the DNA pellet was washed with 50μl 100% ethanol and resuspended in 12μl Milli-Q water.

II.3.5 Very Large Scale Preparation of Plasmid DNA

The cells from a 400ml overnight culture were pelleted by centrifugation using a GSA rotor in a Sorval RC-5 centrifuge, at 5 000rpm, 4° C for 20 minutes. The cells were resuspended in 6ml of *Solution 1* containing 30mg lysozyme and incubated at 37° C for 5 minutes. The solution was transferred to Corex tubes and the cells completely lysed by the addition of 12ml of fresh *Solution 2*, mixed gently then incubated on ice for 5 minutes. 9ml of *Solution 3* was added, mixed and the solution incubated on ice for a further 5 minutes.

After centrifugation in an SS34 rotor at 9 000rpm for 10 minutes, the supernatant, containing the nucleic acids, was poured off and extracted once with 0.5 v/v phenol:chloroform. Following a second centrifugation at 9 000rpm for 5 minutes, the supernatant was transferred to a fresh Corex tube and the nucleic acids precipitated from the aqueous phase by the addition of 0.6 v/v isopropanol and incubating at room temperature for 15 minutes. The nucleic acids were pelleted by

centrifugation at 9 000rpm for 10 minutes, washed in 100% ethanol (to remove all traces of isopropanol) and dissolved in 6ml TE buffer. 24μl RNase A (10mg/ml) was added and the solution incubated at room temperature for 30 minutes.

The solution was extracted once with 0.5 v/v phenol:chloroform and the DNA precipitated with 0.1 v/v 4M NaCl; 2.5 v/v 70% ethanol at -70° C for 30 minutes. The DNA was pelleted, resuspended in 400μl 0.3M sodium acetate, pH5.2 and transferred to a microfuge tube. 1ml of 70% ethanol was added and precipitation carried out at -20° C for 48 hours. The DNA was pelleted, washed with 1ml 100% ethanol and dissolved in 2ml TE buffer. The DNA was purified by CsCl density gradient centrifugation.

II.3.6 CsCl Density Gradient Centrifugation

8ml of a salt cushion, made by dissolving 16g of CsCl in 16ml TE buffer and adding 400μl ethidium bromide (10mg/ml), was transferred into two Beckman Quick Seal Ultra-clear centrifuge tubes. 1ml of the DNA sample was layered on top of the salt cushion and the tubes balanced with any remaining salt cushion. The tubes were filled to the top and balanced with Parafin Oil, then heat sealed. The CsCl gradient was established by centrifugation in a Beckman L8-M Ultracentrifuge using a 70.1Ti rotor accelerated to 45 000rpm for 17 hours.

The DNA bands were identified by UV fluorescence (lower band), the solution drawn off using a 1ml syringe and placed in a 15ml Corex tube. The DNA solution was extracted with an equal volume of water-saturated butanol (to remove the ethidium bromide) until clear. 2 volumes of Milli-Q water were added to dilute the CsCl. The DNA was precipitated in 0.6 v/v isopropanol at room temperature for 60 minutes and collected by centrifugation at 10 000rpm for 10 minutes. The pellet was resuspended in 300μl 0.3M sodium acetate, pH5.2. The solution was transferred to a microfuge tube and the DNA reprecipitated with 900μl 70% ethanol. The pellet was washed with 100% ethanol, dried and resuspended in 100μl Milli-Q water.

Appendix III

Program CONRAD—Documentation

Program CONRAD incorporates the major features of the method for modelling metal binding sites (template matching, hydrophobicity contrast and simplex minimisation) described in Chapter 6. The program is controlled by a total of 41 separate commands and user communication is *via* a command parser. Despite this large number, many commands have default values or run optional utilities and are thus not normally required. The data for the 28 metalloprotein structures given in Chapter 6 (Table 27), were obtained using a maximum of 9 commands. These are shown in the following example command file (VAX/VMS commands begin with a $ sign):

```
$!
$RUN CONRAD
TITLE                              ! Specify a title for the run.
Hydrophobicity contrast in 2AZA.
SETPRN                             ! Set the run process name.
2AZA
PDB                                ! Specify an input PDB format file.
PDB2AZA.ENT
CHARGES                            ! Specify an input charge datafile.
KOLLUNI.DAT
HCF                                ! Specify an output HCF file.
2AZA.HCF
RANGE 1A 129A                      ! Specify the residues to include.
SSBOND 3A 26A                      ! Specify any disulphide bonds.
RANK                               ! Rank any points found in site.
CENTRE 1.400 11.630 12.143         ! Centre point for RANK.
END                                ! End RANK.
GO                                 ! Start program execution.
$EXIT
```

The commands may be specified in any order and in upper or lower case, but must be placed before the GO command. Certain commands are only available within utilities or subprograms. The program will ignore commands beginning with an exclamation mark (!). The functions of all the commands available are described below. <cr> indicates carriage return.

$

Class: Utility.
Default: N/A

```
CONRAD> $DIR<cr>
```

Uses the VAX specific routine LIB$SPAWN to allow access to DCL commands from within the program. The example shown would produce a listing of files in the current directory.

ATOMVDW

Class: Optional.
Default: 0.66

```
CONRAD> ATOMVDW 1.12<cr>
```

Real value representing a van der Waals radius applied to all atoms. (The 0.66 default is the mean of the covalent radii (Allen *et al.*, (1979)) for all atom types found in amino acids).

CAVITY

Class: Optional.
Default: 1.40

```
CONRAD> CAVITY 2.80<cr>
```

A point on the protein grid is defined as being in a cavity if no atoms or solvent points are within the cutoff radius set by this real value.

CENTRE

Class: Optional.
Default: N/A

```
                    (X)    (Y)    (Z)
CONRAD> CENTRE 1.400 11.630 12.143<cr>
_Write RAW data to file: 2AZA.RAW<cr>
_Write FTS data to file: 2AZA.FTS<cr>
```

Command to enter the XYZ coordinates of a centre point, in free format, at which values of C will be calculated for each radius of the range set by the `FIXED` command. The user is prompted for the names of the output RAW and FTS (Fourier Transform Smoothing) data files. It is analogous to the `RANK> CENTRE` command.

CHARGES

Class: Run-dependent.
Default: N/A
Additional commands: $, STOP

```
CONRAD> CHARGES<cr>
_Read charge file: KOLLUNI.DAT<cr>
```

Command to enter the name of a partial charge (or equivalent) datafile to be read; maximum name length, 40 characters. The datafile has the format:

```
GLY                  ! Residue name (A4).
   1                 ! Number of atoms in residue (I4).
CA         0.246000 ! Atom name and partial charge (A4,1X,F13.6).
```

Required for hydrophobicity contrast only. If the user fails to specify a file before entering the GO command, the program is halted and a file prompted for in interactive mode, but terminated with error status in batch mode.

COLOUR

Class: Optional.
Default: 100.0

```
CONRAD> COLOUR 10.0<cr>
```

Arbitrary real value by which values of C_H are multiplied before they are written to a 4th column (equivalent to the PDB occupancy column) in the output HCF file. Allows the points to be displayed in colour using interactive molecular graphics. The values from a previously written HCF file will be divided by this value when used in the **RANK** utility.

DISTANCE

Class: Utility.
Default: N/A
Additional commands: $, STOP

```
CONRAD> DISTANCE<cr>
                (X)     (Y)     (Z)
_Point:1> -20.628 2.496 -14.600<cr>
_Point:2> -17.865 1.322 -20.093<cr>
The distance between these two points = 6.26
_Point:1>
```

Enters a utility to calculate the distance between to points. The utility prompts for the XYZ coordinates of each point in free format. Pressing <cr> before entering coordinates at either of the prompts, exits the utility.

END

Class: Optional.
Default: N/A

```
RANK> END<cr>
CONRAD>
```

Indicates the end of data input to **RANK**.

FIXED

Class: Optional.
Default: Min:0.50, Max:11.50, Step:0.50

```
                (Min) (Max) (Step)
CONRAD> FIXED  1.50  7.00  0.25<cr>
```

Real values which set the minimum and maximum sphere radii, and stepsize, defining a radii range at which values of C are calculated for the centre point set by the `CENTRE` command.

FLOAT

Class: Optional.
Default: Min:0.50, Max:11.50, Step:0.50

```
                (Min) (Max) (Step)
CONRAD> FLOAT  1.00  10.00  2.00<cr>
```

Real values which set the minimum and maximum sphere radii, and stepsize, defining a radii range at which values of C are calculated for each cavity point.

GO

Class: Required.
Default: N/A

```
CONRAD> GO<cr>
```

In all cases, this command initiates execution of the subprogram or utility and must, therefore, be entered only after all parameters have been set, in both interactive and batch mode.

GRID

Class: Optional.
Default: Automatic.

```
                (Xmin)  (Xmax) (Ymin) (Ymax)  (Zmin) (Zmax)
CONRAD> GRID  -0.014  2.981  9.249  13.269  10.657 13.540<cr>
```

By default, the program determines the minimum and maximum X, Y and Z coordinates which will define a grid spanning the whole structure. A grid spanning only part of the structure can be defined by giving the minimum and maximum X, Y and Z coordinates from the region of interest (see also **SOLGRID**).

HCF

Class: Optional.
Default: 'CONRAD.HCF'
Additional commands: STOP

```
CONRAD> HCF<cr>
_Write HCF file: 2AZA.HCF<cr>
```

Command to enter the name of an output file for the points of hydrophobicity contrast, written in PDB format; maximum name length, 40 characters. If none is given and the `NOHCF` command has not been entered, points are written to the default file.

LIGANDS

Class: Optional.
Default: 3

```
CONRAD> LIGANDS 3<cr>
```

Integer value which determines the number of potential metal ligands falling with in the cutoff radius, defined by the `SITE` command, which are required for the current cavity point to be accepted as being in a putative metal site.

MATCH

Class: Run-dependent.
Default: N/A
Additional commands: $, GO, MATCH, NOSCREEN (Default:No), PDB, RANGE, RESET, SCREEN, TOL (Default:2.03), SSBOND, STOP

```
CONRAD> MATCH<cr>
MATCH> PDB<cr>
_Read coordinate file: PDB2CAB.ENT<cr>
MATCH> MATCH
_Read template file: ZN_SITES.DAT<cr>
MATCH> GO
```

The template match subprogram. This requires a metal site templates datafile, maximum name length 40 characters, in the following format:

```
5CPA.  Resolution. 1.54A.  Zinc-site C-BETA pattern.   ! 1
ATOM   554  CB  HIS    69      -1.567  23.787  -4.043 ! 2
ATOM   581  CB  GLU    72      -5.744  23.955  -1.908 ! 3
ATOM   1555 CB  HIS   196      -6.360  26.567  -6.998 ! 4
```

Line 1: Title line (not optional); maximum length, 80 characters.

Lines 2,3,4: Atoms and coordinates comprising the match template, in PDB format.

If the user fails to specify either file before entering the `GO` command, the program is halted and a file prompted for in interactive mode, but terminated with error status in batch mode. The `TOL` command controls the sensitivity of the search by altering the level of agreement between the template and matched atoms, before a match is accepted. By default, the search will ignore atoms from Gly and Pro residues. These can be included by entering the `NOSCREEN` command. The default action is restored using the `SCREEN` command. Disulphide cysteines are excluded from the search using the `SSBOND` command and the search can be performed over a particular region using the `RANGE` command. The `RESET` command returns all match parameters to their default values, while the `STOP` command terminates program execution in interactive mode.

NOHCF

Class: Optional.
Default: HCF file created.

```
CONRAD> NOHCF<cr>
** CONRAD ** No HCF file will be written.
```

By default, the program writes the contrast points to a PDB format HCF file. The output filename will be either that given previously by the user or the default 'CONRAD.HCF' (see **HCF**). This command suppresses the creation of an output HCF file.

NOSCREEN See **MATCH**

NOSELECT

Class: Optional.
Default: No

```
CONRAD> NOSELECT<cr>
```

By default, the program examines only cavity points which are close (*i.e* within the cutoff set by the **SITE** command) to potential metal ligands. Entering **NOSELECT** causes all cavity points found to be examined. The default action is restored by entering the **SELECT** command.

NOSOLVENT

Class: Optional.
Default: The solvent grid is calculated.

```
CONRAD> NOSOLVENT<cr>
** SOLVENT ** The solvent grid will not be calculated.
```

By default, a solvent grid is calculated and the points defined used in the evaluation of hydrophobicity contrast. This command suppresses generation of the solvent grid, so that solvent points are not considered. The default action is restored by entering the **SOLVENT** command.

PDB

Class: Required.
Default: N/A
Additional commands: $, STOP

```
CONRAD> PDB<cr>
_Read PDB file: PDB2AZA.ENT<cr>
```

Command to enter the name of a PDB format coordinate file to be read; maximum name length, 40 characters. Required for hydrophobicity contrast. By default, the program ignores coordinates that begin with 'HETATM'. All coordinates to be read must, therefore, begin with 'ATOM'. If the user fails to specify a file before entering the `GO` command, the program is halted and a file prompted for in interactive mode, but terminated with error status in batch mode.

PROSTEP

Class: Optional.
Default: 1

```
CONRAD> PROSTEP 2<cr>
```

Integer value which defines the stepsize through the protein grid (see **GRID**).

PTS

Class: Optional.
Default: 5.75

```
CONRAD> PTS 3.00<cr>
```

Number of points over which to perform Fourier Transform digital filtering. Related to the number of values of C calculated from the radii ranges set by the `FLOAT` and `FIXED` commands.

RANGE

Class: Optional
Default: All

```
CONRAD> RANGE 5L 105L<cr>
```

The program will read in the coordinates of the whole structure, unless a residue range is specified. The chain identifier must be included, if given, with the residue numbers, as shown in the example. (The command format is identical to SSBOND).

RANK

Class: Utility.
Default: N/A
Additional commands: $, CENTRE, END, GO, RESET, STOP, TITLE

```
CONRAD> RANK<cr>
RANK> CENTRE 1.400 11.630 12.143<cr>
RANK> RANK<cr>
_Read HCF file: 2AZA.HCF<cr>
RANK> GO<cr>
```

Enters a utility which calculates the rank and deviation of each contrast point from a given centre. The utility can be used to rank points from a previously calculated HCF file, as shown in the example (2AZA.HCF), or following calculation of hydrophobicity contrast itself. The CENTRE command is required in both cases, but when employed during the calculation either the END command or <cr> must be given to exit the utility, since only the centre coordinates are required (see the example command file). A maximum of 20 centre point coordinates may be entered at one time and run with the same HCF file. All points are cleared when the utility has executed. The GO, RESET and STOP commands are intended for use only when the utility is operated interactively. The TITLE command allows a title line, maximum length 80 characters, to be entered which will be output during execution. The format is identical to **TITLE**.

RESET

Class: Condition-dependent.
Default: N/A

```
CONRAD> RESET<cr>
** CONRAD ** Restoring default command parameters.
```

During interactive operation, if a number of tasks have been performed, erroneous operation may result due to the alteration of default parameters. This command restores all parameters to their default values. In **RANK**, it restores only parameters relevant to this utility, deletes all centre point coordinates and closes the current HCF file.

SELECT See NOSELECT

SETPRN

Class: Optional.
Default: 'CONRAD'
Additional commands: STOP

```
CONRAD> SETPRN<cr>
_Process name: 2AZA<cr>
```

Uses the VAX specific routine SYS$SETPRN to set the process name; maximum name length, 15 characters. If none is given, the default is used.

SIMPLX

Class: Optional.
Default: N/A
Additional commands: $, GO, STOP

```
CONRAD> SIMPLX<cr>
_Read control file: MODEL2BNT_ZN.DAT<cr>
```

Command to enter the name of a control file, maximum name length 40 characters, required by the Simplex multidimensional minimisation subprogram. Execution begins immediately once a filename has been entered (and its existence verified). The subprogram normally expects the data to be in a file. However, this can be overridden by entering the GO command at the file prompt. No file is read and the user enters the subprogram. Data may then be entered manually at the SIMPLX> prompt. The data must be in the format shown in the following example. The lines can appear in any order, but they must all precede the GO command.

```
* Optimization of the active-site Copper HCF in 2AZA.          ! 1
ATOM    356  ND1 HIS A  46       -.014  11.034  13.540  2.17 ! 2
ATOM    850  SG  CYS A 112       1.606  13.269  10.814  2.05 ! 3
ATOM    885  ND1 HIS A 117       2.981  10.530  12.712  2.17 ! 4
ATOM    922  SD  MET A 121        .040   9.249  10.657  2.91 ! 5
SIMPLX    1  IGN HCF     1       0.000  11.000  11.000       ! 6
CENTRE  989 CU    CU A 130       1.400  11.630  12.143       ! 7
WRITE 2AZA_SMM.PDB                                           ! 8
GO                                                           ! 9
```

Line 1: Optional title line; maximum length, 80 characters. Must begin with an asterisk.
Lines 2,3,4,5: The actual liganding atoms and coordinates, in PDB format. Each line must start with 'ATOM'. The 4th column (the column normally containing PDB occupancy values) must contain the metal-ligand distance.
Line 6: The coordinates of the point to be minimised, in PDB format. This line must begin with 'SIMPLX'.
Line 7: The coordinates of a comparison point. This is used to compare the final position of the minimised point with that of a reference position (the actual metal coordinates are shown in the example). This line must begin with 'CENTRE'.
Line 8: The coordinates of the minimised point will be written to the file given, in PDB format. If the file already exists, the coordinates are appended to it. If it does not, a new file with this name is created.
Line 9: While the Simplex subprogram will minimise only one point at a time, the control file can contain the coordinates of several metal sites to be minimised. The GO command indicates when all the data for a particular site has been read. The program executes only after this command has been encountered.

SITE

Class: Optional
Default: 4.00

```
CONRAD> SITE 4.00<cr>
```

Atoms from potential metal ligands may form a metal site if they fall within the cutoff radius set by this real value.

SOLCUT

Class: Optional.
Default: 1.40

```
CONRAD> SOLCUT 2.80<cr>
```

Points on the solvent grid are defined as solvent if no atoms are found within the cutoff radius set by this real value.

SOLGRID

Class: Optional.
Default: 1

```
CONRAD> SOLGRID 2<cr>
```

Integer value added to the minimum and maximum X, Y and Z coordinates to create a solvent grid around the protein (see also **GRID**).

SOLPAR

Class: Optional.
Default: -0.30

```
CONRAD> SOLPAR -0.50<cr>
```

Real value representing the charge applied to all solvent points.

SOLSTEP

Class: Optional.
Default: 2

```
CONRAD> SOLSTEP 1<cr>
```

Integer value which defines the stepsize through the solvent grid.

SOLVDW

Class: Optional.
Default: 1.40

```
CONRAD> SOLVDW 0.70<cr>
```

Real value representing a van der Waals radius applied to all solvent points. (The 1.40 default represents the van der Waals radius of a water molecule).

SOLVENT See **NOSOLVENT**

SORT

Class: Optional.
Default: All

```
CONRAD> SORT 20 50<cr>
```

Integer values specifying the range of contrast points to be written. By default, the program sorts all the cavity points found into descending order of C_H values, before they are written. Allows a selected range of sorted points to be written instead.

SPHERES

Class: Optional.
Default: 3.50, 5.50, 7.50, 9.50

```
CONRAD> SPHERES 2.00 4.00 6.00 8.00<cr>
```

Real values for the pairs of sphere radii which define the difference in slope used for calculation of the hydrophobicity contrast function C_H at each cavity point.

SSBOND

Class: Structure-dependent.
Default: None

```
CONRAD> SSBOND 23L 87L<cr>
```

Required when the protein contains disulphide bonds. A maximum of 20 disulphide bonds may be entered. The chain identifier must be specified, if given, with the residue numbers, as shown in the example. (The command format is identical to **RANGE**). If the disulphide bonds are ignored, the program will identify them as regions of high hydrophobicity contrast in preference to any metal sites present, which could lead to unrepresentative results.

STOP

Class: Optional.
Default: No

```
CONRAD> STOP<cr>
Exit CONRAD (Y/N)? [N]: Y<cr>
** CONRAD **
```

In interactive mode, terminates execution of any subprogram or utility at any stage *except* after the `GO` command. The user is requested to confirm that they wish to exit the program. If none is given, the user is returned to the subprogram or utility from where the command was issued. This command must not be used in batch mode.

TITLE

Class: Optional.
Default: None

```
CONRAD> TITLE<cr>
_Title: Hydrophobicity contrast in 2AZA.<cr>
```

Allows a title line, maximum length 80 characters, to be entered which will be output during execution. The format is identical to the **RANK** title command.

Bibliography

K. R. Acharya, D. I. Stuart, N. P. C. Walker, M. Lewis, and D. C. Phillips, (1989). *Journal of Molecular Biology*, **208**, 99–127.

E. T. Adman and L. H. Jensen, (1981). *Israel Journal of Chemistry*, **21**, 8–12.

F. H. Allen, S. Bellard, M. C. Brice, B. A. Cartwright, A. Doubleday, H. Higgs, T. Hummelink, B. G. Hummelink-Peters, O. Kennard, and W. D. S. Motherwell, (1979). *Acta Crystallographer*, **B35**, 2331–2339.

P. M. Alzari, M. B. Lascombe, and R. J. Poljak, (1988). *Annual Review of Immunology*, **6**, 555–580.

A. G. Amit, R. A. Mariuzza, S. E. V. Phillips, and R. J. Poljak, (1986). *Science*, **233**, 747–753.

L. M. Amzel and R. J. Poljak, (1979). *Annual Review of Biochemistry*, **48**, 961–967.

J. Åqvist, W. F. van Gunsteren, M. Leifonmark, and O. Tapia, (1985). *Journal of Molecular Biology*, **183**, 461–477.

F. H. Arnold and B. L. Haymore, (1991). *Science*, **252**, 1796–1797.

P. J. Artymuik and C. C. F. Blake, (1981). *Journal of Molecular Biology*, **152**, 737–762.

Y. S. Babu, C. E. Bugg, and W. J. Cook, (1988). *Journal of Molecular Biology*, **204**, 191–204.

E. N. Baker, (1988). *Journal of Molecular Biology*, **203**, 1071–1095.

E. T. Baldwin, I. T. Weber, R. St. Charles, J-C. Xuan, E. Appella, M. Yamada, K. Matsushima, B. F. P. Edwards, G. Marius Clore, A. M. Gronenborn, and A. Wlodawer, (1991). *Proceedings of the National Academy of Science*, **88**, 502–506.

G. J. Barton. *Methods in Enzymology.* Volume 183, Academic Press, New York, (1990).

G. J. Barton and M. J. E. Sternberg, (1987). *Journal of Molecular Biology*, **198**, 327–337.

G. A. Bentley, G. Boulot, M.-M. Riottot, and R. J. Poljak, (1990). *Nature*, **348**, 254–257.

J. M. Berg, (1990). *Journal of Biological Chemistry*, **265**, 6513–6516.

F. C. Bernstein, T. F. Koetzle, G. T. B. Williams, E. F. Meyer, M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi, (1977). *Journal of Molecular Biology*, **112**, 535–542.

P. R. Bevington. *Data Reduction and Error Analysis for the Physical Sciences.* New York: McGraw-Hill, (1969).

T. N. Bhat, G. A. Bentley, T. O. Fischmann, G. Boulot, and R. J. Poljak, (1990). *Nature*, **347**, 483–484.

H. C. Birnboim and J. Doly, (1979). *Nucleic Acids Research*, **7**, 1513–1523.

B. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, (1983). *Journal of Computational Chemistry*, **4**, 187–217.

R. E. Bruccoleri and M. Karplus, (1985). *Macromolecules*, **18**, 2767–2773.

R. E. Bruccoleri and M. Karplus, (1987). *Biopolymers*, **26**, 137–168.

R. E. Bruccoleri, E. Haber, and J. Novotný, (1988). *Nature*, **335**, 564–568.

J. E. Calvert, (1986). *Immunology Today*, **7**, 136–137.

I. D. Campbell and R. A. Dwek. *Biological Spectroscopy.* Benjamin/Cummings, California, (1984).

D. A. Carson and B. D. Freimark, (1986). *Advances in Immunology*, **38**, 275–311.

M. J. Casadaban and S. N. Cohen, (1980). *Journal of Molecular Biology*, **138**, 179–207.

P. Chakrabarti, (1989). *Biochemistry*, **28**, 6081–6085.

P. Chakrabarti, (1990). *Biochemistry*, **29**, 651–658.

J. C. Cheetham. *Studies of Enzyme-Inhibitor Interactions and Protein Dynamics.* PhD thesis, University of Oxford, (1986).

J. C. Cheetham, (1989). *Protein Engineering*, **1**, 170–172.

N. C. Chien, V. A. Roberts, A. M. Giusti, M. D. Scharff, and E. D. Getzoff, (1989). *Proceedings of the National Academy of Science*, **86**, 5532–5536.

C. Chothia, (1991). *Current Opinion in Structural Biology*, **1**, 53–59.

C. Chothia, A. M. Lesk, A. Tramontano, M. Levitt, S. J. Smith-Gill, G. Air, S. Sheriff, E. A. Padlan, D. R. Davies, W. R. Tulip, P. M. Colman, S. Spinelli, P. M. Alzari, and R. J. Poljak, (1989). *Nature*, **342**, 877–883.

C. Chothia and J. Janin, (1981). *Proceedings of the National Academy of Science*, **78**, 4146–4150.

C. Chothia and A. M. Lesk, (1987). *Journal of Molecular Biology*, **196**, 901–917.

C. Chothia, Novotný J., R. E. Bruccoleri, and Karplus M., (1985). *Journal of Molecular Biology*, **186**, 651–663.

A. Colman. *Transcription and Translation - A Practical approach.* IRL Press, Oxford, (1984).

P. M. Colman, (1988). *Advances in Immunology*, **43**, 99–132.

P. M. Colman, W. G. Laver, J. N. Varghese, A. T. Baker, P. A. Tullock, G. M. Air, and R. G. Webster, (1987). *Nature*, **326**, 358–363.

P. M. Colman, W. R. Tulip, J. N. Varghese, P. A. Tulloch, A. T. Baker, G. M. Air, and R. G. Webster, (1989). *Philosophical Transactions of the Royal Society of London*, **B323**, 511–518.

Thomas E. Creighton. *Proteins. Structure and Molecular Properties.* W. H. Freeman and Company, New York, (1984).

A. C. R. da Silva and F. C. Reinach, (1991). *Trends in Biochemical Science*, **16**, 53–57.

D. R. Davies and H. Metzger, (1983). *Annual Review of Immunology*, **1**, 87–117.

D. R. Davies, E. A. Padlan, and S. Sheriff, (1990). *Annual Review of Biochemistry*, **59**, 439–473.

D. R. Davies, S. Sheriff, and Padlan E. A., (1988). *Journal of Biological Chemistry*, **263**, 10541–10544.

P. de la Paz, B. J. Sutton, M. J. Darsley, and A. R. Rees, (1986). *EMBO Journal*, **5**, 415–425.

D. R. Drummond, J. Armstrong, and A. Colman, (1985). *Nucleic Acids Research*, **13**, 7375–7394.

D. Eisenberg and A. D. McLachlan, (1986). *Nature*, **319**, 199–203.

O. Epp, E. E. Lattman, M. Schiffer, R. Huber, and W. Palm, (1975). *Biochemistry*, **14**, 4943–4952.

I. Fischer-Hjalmars and A. Henriksson-Enflo. *Advances in Quantum Chemistry.* Volume 16, Academic Press, New York, (1982).

M. Fujinaga and M. N. G. James, (1987). *Journal of Molecular Biology*, **195**, 373–396.

P. Furet, A. Sele, and N. C. Cohen, (1988). *Journal of Molecular Graphics*, **6**, 182–189.

W. F. Furey, A. H. Robbins, L. L. Clancy, D. R. Winge, B. C. Wang, and C. D. Stout, (1986). *Science*, **231**, 704–710.

W. J. Furey, B. C. Wang, C. S. Yoo, and M. Sax, (1983). *Journal of Molecular Biology*, **167**, 661–692.

N. Gō and H. A. Scheraga, (1970). *Macromolecules*, **3**, 178–187.

P. J. Goodford, (1985). *Journal of Medicinal Chemistry*, **28**, 849–857.

P. Gros, M. Fujinaga, B. W. Dijkstra, K. H. Kalk, and W. G. J. Hol, (1989). *Acta Crystallographer*, **B45**, 488–499.

J. M. Guss, P. R. Harrowell, M. Murata, V. A. Norris, and H. C. Freeman, (1986). *Journal of Molecular Biology*, **192**, 361–387.

K. D. Hardman and C. F. Ainsworth, (1972). *Biochemistry*, **11**, 4910–4919.

Robert W. Hay. *Bio-Inorganic Chemistry.* Ellis Horwood, (1987).

H. W. Hellinga and F. M. Richards, (1991). *Journal of Molecular Biology*, **222**, 763–785.

H. W. Hellinga, J. P. Caradonna, and F. M. Richards, (1991). *Journal of Molecular Biology*, **222**, 787–803.

K. Henrick, C. A. Collyer, and D. M. Blow, (1989). *Journal of Molecular Biology*, **208**, 129–157.

J. Hermans and J. E. McQueen, (1974). *Acta Crystallographer*, **A30**, 730–739.

J. N. Herron, X. M. He, M. L. Mason, E. W. Voss, and A. B. Edmundson, (1989). *Proteins: Structure, Function and Genetics*, **5**, 271–280.

O. Herzberg and M. N. G. James, (1988). *Journal of Molecular Biology*, **203**, 761–779.

J. N. Higaki, R. J. Fletterick, and C. S. Craik, (1992). *Trends in Biochemical Science*, **17**, 100–104.

J. N. Higaki, B. L. Haymore, S. Chen, Fletterick R. J., and C. S. Craik, (1990). *Biochemistry*, **29**, 8582–8586.

M. A. Holmes and B. W. Matthews, (1982). *Journal of Molecular Biology*, **160**, 623–639.

A. Holmgren and C. I. Bränden, (1989). *Nature*, **342**, 248–251.

B. L. Iverson and R. A. Lerner, (1989). *Science*, **243**, 1184–1188.

B. L. Iverson, S. A. Iverson, V. A. Roberts, E. D. Getzoff, J. A. Tainer, S. J. Benkovic, and R. A. Lerner, (1990). *Science*, **249**, 659–662.

J. Janin and C. Chothia, (1990). *Journal of Biological Chemistry*, **265**, 16027–16030.

P. Jeffrey. PhD thesis, University of Oxford, (1989).

P. D. Jeffrey, R. E. Griest, G. L. Taylor, and A. R. Rees, (1992). Manuscript in preparation.

P. T. Jones, P. H. Dear, J. Foote, M. S. Neuberger, and Winter G., (1986). *Nature*, **321**, 522–525.

T. A. Jones. *Methods in Enzymology.* Volume 115, Academic Press, New York, (1985).

E. A. Kabat, T. T. Wu, and H. Bilofsky, (1977). *Journal of Biological Chemistry*, **252**, 6609–6616.

E. A. Kabat, T. T. Wu, M. Reid-Miller, H. M. Perry, and K. S. Gottesman. *Sequences of Proteins of Immunological Interest.* U.S. Department of Health and Human Services, Fourth edition, (1987).

K. K. Kannan, T. A. Jones, and M. Ramanadham, (1984). *Annals of the New York Academy of Sciences*, **429**, 49–60.

R. Kaptein, (1991). *Current Opinion in Structural Biology*, **1**, 63–70.

A. D. Kline, W. Braun, and K. Wüthrich, (1988). *Journal of Molecular Biology*, **204**, 675–724.

A. Klug and D. Rhodes, (1987). *Trends in Biochemical Science*, **12**, 464–469.

W. R. Kneen, M. J. W. Rogers, and P. Simpson. *Chemistry: Facts, Patterns and Principles.* Addison-Wesley, (1984).

D. E. Knuth. Fundamental Algorithms, *Vol. 1* of The Art of Computer Programming. Reading, Mass.: Addison-Wesley, (1968).

G. Köhler and C. Milstein, (1975). *Nature*, **256**, 495–497.

P. A. Kreig and D. A. Melton, (1984). *Nucleic Acids Research*, **12**, 7057–7070.

R. Kuroki, Y. Taniyama, C. Seko, H. Nakamura, M. Kikuchi, and Ikehara M., (1989). *Proceedings of the National Academy of Science*, **86**, 6903–6907.

U. K. Laemmli, (1970). *Nature*, **227**, 680–685.

M.-B. Lascombe, P. M. Alzari, G. Boulot, P. Saludjian, P. Tougard, C. Berek, S. Haba, E. M. Rosen, A. Nisonoff, and R. J. Poljak, (1989). *Proceedings of the National Academy of Science*, **86**, 607–611.

B. K. Lee and F. M. Richards, (1971). *Journal of Molecular Biology*, **55**, 379–400.

A. M. Lesk and C. Chothia, (1982). *Journal of Molecular Biology*, **160**, 325–342.

A. M. Lesk and C. Chothia, (1988). *Nature*, **335**, 188–190.

C. T. Lewis and D. Hilvert, (1991). *Current Opinion in Structural Biology*, **1**, 624–629.

S.-L. Lin, E. A. Stern, A. J. Kalb (Gilboa), and Y. Zhang, (1990). *Biochemistry*, **29**, 3599–3603.

C. R. Lowe, (1986). *Biosensors*, **1**, 3–16.

R. A. Mariuzza, S. E. V. Phillips, and R. J. Poljak, (1987). *Annual Review of Biophysics and Bioengineering*, , 139–159.

M. Marquart, W. Bode, J. Deisenhofer, R. Huber, and J. Walter, (1983). *Acta Crystallographer*, **B39**, 480–490.

A. C. R. Martin. *Molecular Modelling of Antibody Combining Sites.* PhD thesis, University of Oxford, (1990).

A. C. R. Martin, J. C. Cheetham, and A. R. Rees, (1989). *Proceedings of the National Academy of Science*, **86**, 9269–9272.

M. Matsushima, M. Marquart, T. A. Jones, P. M. Colman, K. Bartels, R. Huber, and W. Palm, (1978). *Journal of Molecular Biology*, **121**, 441–459.

C. A. McPhalen and M. N. G. James, (1988). *Biochemistry*, **27**, 6582–6598.

J. A. McCammon, P. G. Wolynes, and M. Karplus, (1979). *Biochemistry*, **18**, 927–942.

E. Meyer, R. Radhakrishnan, G. Cole, and O. Epp, (1988). *Acta Crystallographer*, **B44**, 26–38.

I. S. Mian, A. R. Bradwell, and A. J. Olson, (1991). *Journal of Molecular Biology*, **217**, 133–151.

R. A. Miller, P. T. Oseroff, A. R. Stratte, and R. Levy, (1983). *Blood*, **62**, 988–995.

C. Milstein, (1980). *Scientific American*, **243**, 56–74.

D. E. Mosier, I. M. Zitron, J. J. Mond, A. Ahmed, I. Scher, and Paul W. E., (1977). *Immunological Reviews*, **37**, 89–104.

K. Nakamaye and F. Eckstein, (1986). *Nucleic Acids Research*, **14**, 9679–9698.

S. B. Needleman and Wunsch C. D., (1970). *Journal of Molecular Biology*, **48**, 443–453.

J. A. Nelder and R. Mead, (1965). *Computer Journal*, **7**, 308.

J. Norrander, T. Kemple, and J. Messing, (1983). *Gene*, **26**, 101–106.

J. R. North, (1985). *Trends in Biotechnology*, **3**, 180–186.

J. Novotný, R. E. Bruccoleri, and E. Haber, (1990). *Proteins: Structure, Function and Genetics*, **7**, 93–98.

J. Novotný, R. E. Bruccoleri, J. Newell, D. Murphy, and M. Karplus, (1983). *Journal of Biological Chemistry*, **258**, 14433–14437.

A. G. Orpen, L. Brammer, F. H. Allen, O. Kennard, and D. G. Watson, (1989). *Journal of the Chemical Society — Dalton Transactions*, **Issue 12**, S1–S83.

E. Padlan, (1990). *Proteins: Structure, Function and Genetics*, **7**, 112–124.

E. A. Padlan and E. A. Kabat, (1988). *Proceedings of the National Academy of Science*, **85**, 6885–6889.

E. A. Padlan, D. R. Davies, I. Pecht, D. Givol, and C. Wright, (1976). *Cold Spring Harbor Symposia in Quantitative Biochemistry*, **41**, 627–637.

E. A. Padlan, E. W. Silverton, S. Sheriff, G. H. Cohen, S. J. Smith-Gill, and D. R. Davies, (1989). *Proceedings of the National Academy of Science*, **86**, 5938–5942.

M. W. Pantoliano, M. Whitlow, J. F. Wood, M. L. Rollence, B. C. Finzel, G. L. Gilliland, T. L. Poulos, and P. N. Bryan, (1988). *Biochemistry*, **27**, 8311–8317.

R. M. E. Parkhouse and M. D. Copper, (1977). *Immunological Reviews*, **37**, 105–126.

K. Petratos, Z. Dauter, and K. S. Wilson, (1988). *Acta Crystallographer*, **B44**, 628–636.

G. A. Petsko and D. Ringer, (1984). *Annual Review of Biophysics and Bioengineering*, **13**, 331–342.

J. W. Pflugrath, G. Wiegand, and R. Huber, (1986). *Journal of Molecular Biology*, **189**, 383--386.

D. A. Phipps. *Metals and metabolism.* Oxford University Press, (1976).

B. V. Plapp, H. Eklund, T. A. Jones, and C. I. Brändén, (1983). *Journal of Biological Chemistry*, **258**, 5537–5547.

R. J. Poljak, L. M. Amzel, H. P. Avey, B. L. Chem, R. P. Phizackerley, and F Saul, (1973). *Proceedings of the National Academy of Science*, **70**, 3305–3310.

L. Prasad, M. Vandonselaar, J. S. Lee, and L. T. J. Delbaere, (1988). *Journal of Biological Chemistry*, **263**, 2571–2574.

T. Prasthofer, S. R. Phillips, J. A. Engler, and F. L. Suddath, (1989). *Journal of Biological Chemistry*, **264**, 6793–6796.

W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes—The Art of Scientific Computing.* Cambridge University Press, (1986).

T. M. Radhakrishnan, K. A. Walsh, and H. Neurath, (1969). *Biochemistry*, **8**, 4020–4027.

G. N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan, (1963). *Journal of Molecular Biology*, **7**, 95–99.

B. Ratner. *Allergy Anaphylaxis and Immunotherapy.* Williams and Wilkins, Baltimore, (1943).

D. C. Rees, M. Lewis, and W. N. Lipscomb, (1983). *Journal of Molecular Biology*, **168**, 367–387.

J. S. Richardson, (1981). *Advances in Protein Chemistry*, **34**, 168–339.

J. S. Richardson, D. C. Richardson, and K. A. Thomas, (1976). *Journal of Molecular Biology*, **102**, 221–235.

L. Riechmann, M. Clark, H. Waldmann, and G. Winter, (1988). *Nature*, **332**, 323–327.

S. Roberts. PhD thesis, University of Oxford, (1987).

S. Roberts and A. R. Rees, (1986). *Protein Engineering*, **1**, 59–65.

S. Roberts, J. C. Cheetham, and A. R. Rees, (1987). *Nature*, **328**, 731–734.

V. A. Roberts, B. L. Iverson, S. A. Iverson, S. J. Benkovic, R. A. Lerner, E. D. Getzoff, and J. A. Tainer, (1990). *Proceedings of the National Academy of Science*, **87**, 6654–6658.

J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, (1977). *Journal of Computational Physics*, **23**, 327–341.

A. Sadhu and J. A. Magnuson, (1989). *Biochemistry*, **28**, 3197–3204.

J. Sambrook, E. F. Fritsch, and T. Maniatis. *Molecular Cloning. A Laboratory Manual. Second Edition.* Cold Spring Harbor Laboratory Press, (1989).

F. Sanger, S. Nicklen, and A. R. Coulson, (1977). *Proceedings of the National Academy of Science*, **74**, 5463–5467.

Y. Satow, G. H. Cohen, E. A. Padlan, and D. R. Davies, (1986). *Journal of Molecular Biology*, **190**, 593–604.

F. A. Saul, L. M. Amzel, and R. J. Poljak, (1978). *Journal of Biological Chemistry*, **253**, 585–597.

J. R. Sayers, W. Schmidt, and F. Eckstein, (1988). *Nucleic Acids Research*, **16**, 791–802.

M. Schiffer, R. L. Girling, K. R. Ely, and A. B Edmunds, (1973). *Biochemistry*, **12**, 4620–4631.

H. Schneider-Bernlöhr, G. Formicka-Kozłowska, R. Bühler, J-P von Wartburg, and M. Zeppezauer, (1988). *European Journal of Biochemistry*, **173**, 275–280.

P. Schultze, E. Wörgötter, W. Braun, G. Wagner, M. Vašák, J. H. R. Kägi, and K. Wüthrich, (1988). *Journal of Molecular Biology*, **203**, 251–268.

D. L. Shawler, R. M. Bartholomew, L. M. Smith, and R. O. Dillman, (1985). *The Journal of Immunology*, **135**, 1530–1535.

J. Shen, S. Subramaniam, C. F. Wong, and J. A. McCammon, (1989). *Biopolymers*, **28**, 2085–2096.

J. Shen, C. F. Wong, S. Subramaniam, T. A. Albright, and J. A. McCammon, (1990). *Journal of Computational Chemistry*, **11**, 346–350.

S. Sheriff, E. W. Silverton, E. A. Padlan, G. H. Cohen, S. J. Smith-Gill, B. C. Finzel, and D. R. Davies, (1987). *Proceedings of the National Academy of Science*, **84**, 8075–8079.

S. J. Smith-Gill, A. C. Wilson, M. Potter, E. M. Prager, R. J. Feldmann, and C. R. Mainhart, (1982). *The Journal of Immunology*, **128**, 314–322.

M. E. Snow and L. M. Amzel, (1986). *Proteins: Structure, Function and Genetics*, **1**, 267–279.

E. E. Snyder, B. W. Buoscio, and J. J. Falke, (1990). *Biochemistry*, **29**, 3937–3943.

R. L. Stanfield, T. M. Fieser, R. A. Lerner, and I. A. Wilson, (1990). *Science*, **248**, 712–719.

D. Staunton, (1992). Manuscript in preparation.

R. E. Stenkamp, L. H. Jensen, and L. C. Sieker, (1983). *Acta Crystallographer*, **B39**, 697–703.

R. C. Stevens, J. E. Gouaux, and W. N. Lipscomb, (1990). *Biochemistry*, **29**, 7691–7701.

D. I. Stuart, K. R. Acharya, N. P. C. Walker, S. G. Smith, M. Lewis, and Phillips D. C., (1986). *Nature*, **324**, 84–87.

S. W. Suh, T. N. Bhat, M. A. Navia, G. H. Cohen, D. N. Rao, S. Rudikoff, and D. R. Davies, (1986). *Proteins: Structure, Function and Genetics*, **1**, 74–80.

A. L. Swain, R. H. Kretsinger, and E. L Amma, (1989). *Journal of Biological Chemistry*, **264**, 16620–16628.

D. M. E. Szebenyi and K. Moffat, (1986). *Journal of Biological Chemistry*, **261**, 8761–8777.

J. A. Tainer, V. A. Roberts, and E. D. Getzoff, (1991). *Current Opinion in Biotechnology*, **2**, 582–591.

J. A. Tainer, E. D. Getzoff, K. M. Beem, J. S. Richardson, and D. C. Richardson, (1982). *Journal of Molecular Biology*, **160**, 181–217.

J. W. Taylor, J. Ott, and F. Eckstein, (1985). *Nucleic Acids Research*, **13**, 8765–8785.

J. W. Taylor, W. Schmidt, R. Cosstick, A. Okruszek, and F. Eckstein, (1985). *Nucleic Acids Research*, **13**, 8749–8764.

A. Tramontano, C. Chothia, and A. M. Lesk, (1990). *Journal of Molecular Biology*, **215**, 175–182.

W. R. Tulip, J. N. Varghese, R. G. Webster, G. M. Air, W. G. Laver, and P. M. Colman, (1989). *Cold Spring Harbor Symposia in Quantitative Biochemistry*, **54**, 257–263.

G. Valle, J. Besley, and A. Colman, (1981). *Nature*, **291**, 338–342.

B. L. Vallee and D. S. Auld, (1989). *FEBS Letters*, **257**, 138–140.

B. L. Vallee and D. S. Auld, (1990). *Biochemistry*, **29**, 5647–5659.

B. L. Vallee and R. J. P. Williams, (1968). *Proceedings of the National Academy of Science*, **59**, 498–505.

B. L. Vallee, J. E. Coleman, and D. S. Auld, (1991). *Proceedings of the National Academy of Science*, **88**, 999–1003.

W. F. van Gunsteren and H. J. C. Berendsen, (1977). *Molecular Physics*, **34**, 1311–1327.

M. Verhoeyen, C. Milstein, and G. Winter, (1988). *Science*, **239**, 1534–1536.

L. Vértesy, V. Oeding, R. Bender, K. Zepf, and G. Nesemann, (1984). *European Journal of Biochemistry*, **141**, 505–512.

N. K. Vyas, M. N. Vyas, and F. A. Quiocho, (1988). *Science*, **242**, 1290–1295.

K. D. Watenpaugh, L. C. Sieker, and L. H. Jensen, (1980). *Journal of Molecular Biology*, **138**, 615–633.

S. J. Weiner, P. A. Kollman, D. A. Case, U. Chandra-Singh, C. Ghio, G. Alagona, S. Profeta Jr, and P. Weiner, (1984). *Journal of the American Chemical Society*, **106**, 765–784.

A. F. Williams and A. N. Barclay, (1988). *Annual Review of Immunology*, **6**, 381–406.

R. J. P. Williams, (1984). *Endeavour, New Series*, **8**(2), 65–70.

R. J. P. Williams, (1985). *European Journal of Biochemistry*, **150**, 231–248.

R. J. P. Williams, (1987). *Polyhedron*, **6**(1), 61–69.

G. Winter and C. Milstein, (1991). *Nature*, **349**, 293–299.

K. Wüthrich. *NMR of Proteins and Nucleic Acids.* John Wiley & Sons, New York, (1986).

M. M. Yamashita, L. Wesson, G. Eisenman, and D. Eisenberg, (1990). *Proceedings of the National Academy of Science*, **87**, 5648–5652.

M. M. Yamashita, D. Eisenberg, D. Cascio, R. J. Almassy, and C. A. Janson, (1989). *Journal of Biological Chemistry*, **264**, 7681–7690.