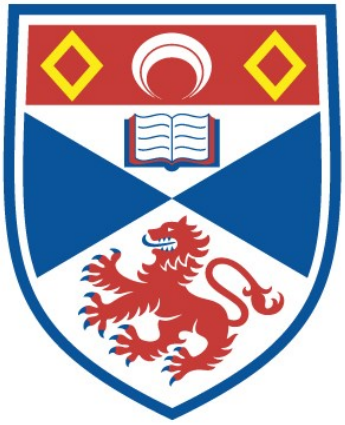
NMR INVESTIGATION OF MOLECULAR CONTACTS IN A HAPTEN-ANTIBODY COMPLEX

Douglas Graham Low

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews

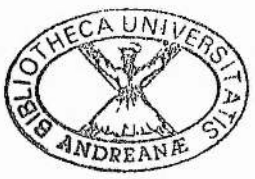


1997

NMR Investigation of Molecular Contacts in a Hapten-Antibody Complex.

Douglas Graham Low

A thesis submitted for the degree of Doctor of Philosophy.



Centre for Biomolecular Sciences,
University of St. Andrews,
August 1996.

Table of Contents

Acknowledgements

Firstly I would like to thank my two supervisors, Steve Homans (St. Andrews) and Paul Davis (Unilever Research), for all the help, encouragement and support throughout my Ph. D. project.

At Colworth I would like to thank John Windust, for teaching me all the molecular biology techniques needed, Karen Cromie and Paul van der Logt for lots of advice, Trevor Wattam for his help with the affinity chromatography, and Sandra Hemmington for her help with the protein expression.

At St. Andrews I would like to thank Mark Probert for providing the ^{13}C labelled compound, Trevor Rutherford, Charles Weller and Mark Milton for advice and assistance with the NMR experiments and molecular modelling, and Kothandaraman Seshadri his time and effort in doing the RCCAL calculations.

Finally I would like to thank all the people I have worked with at both St. Andrews and Colworth.

This work was funded by a BBSRC CASE studentship with Unilever Research (Colworth Laboratory).

Declarations.

I, Douglas Graham Low, hereby certify that this thesis, which is approximately 40 000 words in length, has been written by me, that it is a record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

28th August 1996 Douglas Graham Low

I was admitted as a research student in October 1992 and as a candidate for the degree of Doctor of Philosophy in October 1992; the higher study for which was carried out in the University of St. Andrews between October 1992 and August 1996.

28th August 1996 Douglas Graham Low

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Doctor of Philosophy in the University of St. Andrews and that the candidate is qualified to submit this thesis in application for that degree.

28th August 1996 Steve Homans

In submitting this thesis to the University of St. Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be supplied to any *bona fide* library or research worker.

28th August 1996 Douglas Graham Low

Abstract

An antibody Fv fragment was cloned from hybridoma cells of a monoclonal antibody against against the steroid derivative estrone-3-glucuronide (E3G). The detection of this molecule is important in a method for predicting fertility. The variable heavy and light chain genes were isolated and amplified and placed into a modified pUC 19 vector to allow expression of active protein. The immunochemistry of the fragment was determined by ELISA, demonstrating that the fragment did bind E3G, that this binding was specific, the antibody fragment competed against the parent monoclonal antibody, and that the fragment could inhibit binding of the monoclonal antibody to an E3G-alkaline phosphotase conjugate. The expression of the fragment in *E. coli* was optimised in various media with levels of 4 $mg.L^{-1}$ achieved.

The structure of E3G was probed in solution and in complex with the Fv fragment, utilising E3G with a fully ^{13}C labelled glucuronic acid moiety: estrone-[U-^{13}C]-glucuronide. Firstly in the in the solution structure studies by giving ^{13}C and ^{1}H assignments via a ^{13}C-^{13}C COSY and a ^{1}H-^{13}C HSQC, and secondly in the bound state by allowing the use of isotope editing techniques. In solution the glycan was found to exist in multiple conformations, with particularly large fluctuations about the glycosidic linkage ψ. The antibody selects a conformation from the free solution which does not correspond to either of the two minimum energy conformations of the free glycan. The glucuronic acid moiety undergoes a stacking interaction with an aromatic ring in the binding site, and both

NOEs and ring current shifts are in good agreement with the predicted bound state conformation. The predicted bound state conformation is also in good agreement with preliminary X-ray data.

A homology modelled protein structure of the Fv was built and compared with the X-ray crystal structure.

List of Figures.

Chapter 1.

Chapter 2.

Chapter 3.

List of Tables.

Chapter 2

Chapter 3.

Chapter 4.

Chapter 5.

Chapter 6

Chapter 7

List of Abbreviations.

2TY	bacto-tryptone, bacto-yeast media
BCIP	5-bromo-4-chloro-3-inodyl phosphate
BHI	brain heart infusion
BSA	bovine serum albumin
cDNA	complimentary DNA
CDR	complimentary determining region
COSY	correlated spectroscopy
ddNTP	dideoxynucleoside triphosphate solutions
DEPC	diethyl pyrocarbonate (or diethyl oxydiformate)
DNA	deoxyribonucleic acid
dNTP	deoxynucleoside triphosphate solutions
ds-scFv	disulphide scFv
DTT	dithiothreitol
E3G	estrone-3-glucuronide
E3G-AP	estrone-3-glucuronide-alkaline-phosphotase conjugate
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
ELISA	enzyme linked immunosorbent assay
ER	endoplasmic reticulum
Fab	antigen-binding antibody fragment
Fc	crystallisable antibody fragment
FR	framework region
Fv	variable region antibody fragment
g	gravity
HCCH-COSY	^{1}H-^{13}C-^{13}C-^{1}H correlation via $^{1}J_{CC}$ couplings
hnRNA	heterogeneous nuclear RNA

HSQC	heteronuclear single quantum correlation spectroscopy
IgG	immunoglobulin G
INEPT	insensitive nuclei enhanced by polarisation transfer
IPTG	isopropyl-β-D-thiogalactoside
LH	luetinising hormone
M9P	M9 minimal salts media
mAb	monoclonal antibody
mRNA	messenger RNA
NBT	nitro blue tetrazolium
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
NOESY	nuclear Overhauser effect spectroscopy
NOESY-HSQC	three-dimensional heteronuclear ^{1}H nuclear Overhauser ^{13}C-^{1}H HSQC
NP-40	nonylphenoxy polyethoxy ethanol
oligo-dT	oligo-thymidine RNA
P3G	pregnanediol-3-glucuronide
PCR	polymerase chain reaction
PEG	polyethylene glycol
PBS	phosphate buffered saline
PBSTA	phosphate buffered saline with tween and azide
poly(A)	poly adenyl
PNPP	*para*-nitrophenyl phosphate
RNA	ribonucleic acid
ROE	rotating frame Overhauser effect
ROESY	rotating frame Overhauser effect spectroscopy
rpm	revolutions per minute

RTB	reverse transcriptase buffer
scFv	single-chain Fv fragment
scRNA	small cytoplasmic RNA
SDS	sodium dodecyl sulphate
snRNA	small nuclear RNA
TE	Tris, EDTA buffer
Tris.Cl	tris(hydroxymethyl)aminoethane
tRNA	transfer RNA
TRNOE	transferred NOE
YE	yeast extract

Chapter 1.

Introduction.

1.1 Fertility Prediction.

The work in this thesis centres around the molecule estrone-3-glucuronide (E3G) and the Fv fragment of a monoclonal antibody (mAb) raised against it. E3G (figure 1.1) is a metabolite of the female hormone estrone, and when levels of E3G are measured in urine, along with the levels of luetinising hormone (LH) and pregnanediol-3-glucuronide (P3G), these give an indication of the probability of fertility. Antibodies or antibody fragments are suitable molecules for detecting these compounds, as they bind with high specificity, and the amount of antigen can be measured. A graph of the typical results are shown in figure 1.2, which shows how these measurements can used as a basis for fertility prediction.

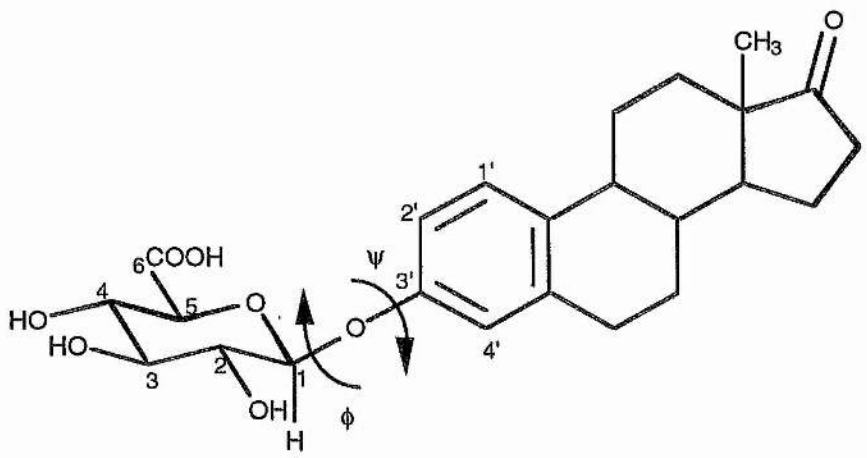


Figure 1.1 Structure of estrone-3-glucuronide showing glycosidic dihedral angles ϕ (H1-C1-O1-C3') and ψ (C1-O1-C3'-C4').

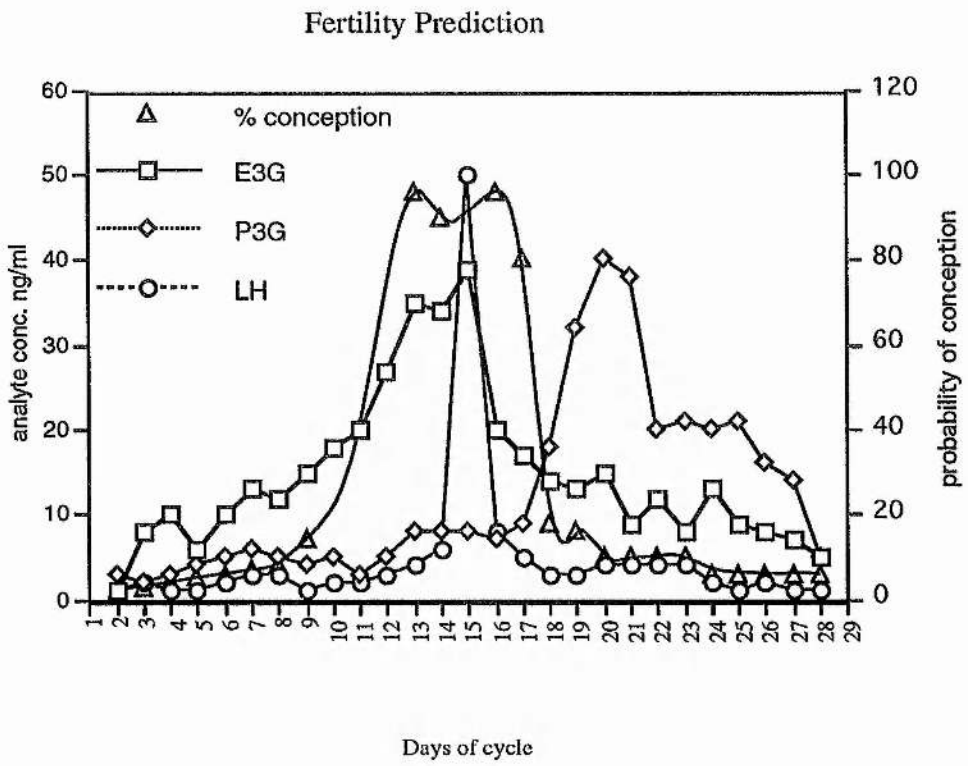


Figure 1.2 Graph showing level of hormones in urine as a basis for predicting fertility.

1.2 Antibodies

The recognition of a foreign antigen is the hallmark of the immune response. Two distinct types of molecules are involved in this process - the immunoglobulins and the T cell antigen receptor. Immunoglobulins, or antibodies, are a group of glycoproteins present in the serum and tissue fluids of all mammals. They are produced in large amounts by plasma cells which have developed from precursor B lymphocytes. Such lymphocytes carry membrane bound immunoglobulins of the same specificity of the plasma cell. Contact between these B lymphocytes and foreign antigen are required for the induction of antibody formation.

1.2.1 Immunoglobulin G

There are five distinct classes of immunoglobulins recognised in most mammals: IgG, IgA, IgM, IgD and IgE, these differ in size, charge, amino acid composition and carbohydrate content. Since the monoclonal antibody from which the Fv fragment was cloned is an IgG immunoglobulin, this will be the only class discussed in any detail. IgG is the major immunoglobulin in humans, it accounts for 70-75 % of total immunoglobulins in serum. IgG is a monomeric protein with a molecular weight of 146 kDa, although the IgG3 subclass is slightly larger. The IgG class is the main antibody of the immune response.

1.2.2 IgG Structure

The IgG structure is represented schematically in figure 1.3. The IgG consists of four polypeptide chains: two identical heavy chains with a molecular weight of 50-77 kDa; and two identical light chains of 25 kDa molecular weight. The polypeptide chains are held together with both covalent and non-covalent bonding. As shown in figure 1.3. There are disulphide bonds between each of the heavy chains and its adjacent light chain, and between the two heavy chains. The amino end of both heavy and light chains is characterised by sequence variability, these are referred to as V_H and V_L regions respectively. The constant (C) portion of the light chain is termed the C_L region. The constant portion of the heavy chain is further subdivided into three structurally discrete regions: C_H1, C_H2, C_H3. These subdivided regions which appear globular are called domains, and are further stabilised by intrachain disulphide bonds. The sites at which antibodies bind antigen are located in the variable region. The hinge region is a segment of the heavy chain located between the C_H1 and C_H2 domains. The flexibility this provides allows the two antigen binding sites

to operate independently. The C_H2 region is where the IgG is glycosylated.

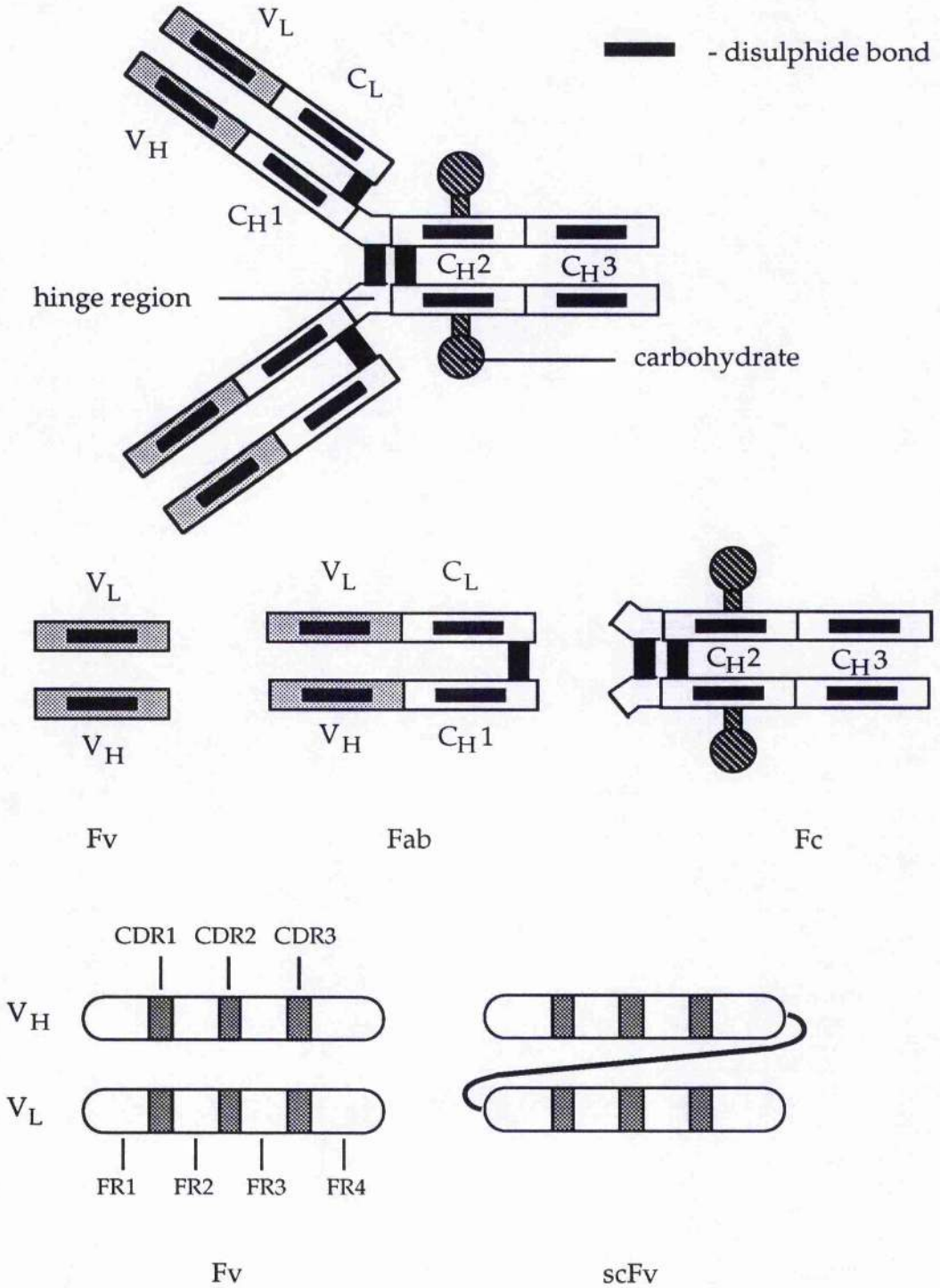


Figure 1.3 Schematic representation of Immunoglobulin G, and Fv, scFv and Fab and Fc antibody fragments, also showing CDR and FR regions.

1.2.3 Structure in Relation to Function.

Digestion of the IgG protein with the proteases papain and pepsin gave important information on structure/function relationships. Papain digestion releases two identical Fab fragments and an Fc fragment (figure 1.3). Fc fragments have several functions including binding to mononuclear cells, platelets, K cells and *staphylococcus aureus* protein A. The Fab fragment can be further reduced to fragments containing only the variable heavy and light chain domains (Fv fragment). These fragments are stable and bind to the antigen, as described in the next section all the antigen binding sites are in the variable regions of the antibody.

1.2.4 Structure in Relation to Antigen Binding.

When the primary amino acid structure of a large number of heavy and light chain were examined, it was found that the variability in these regions is not distributed evenly throughout these domains. Certain short segment of the variable domain show exceptional variability, these are known as hypervariable regions or complimentary determining regions (CDRs), as these are the areas directly involved in antigen binding. The intervening regions which show a much reduced degree of variability are known as the framework regions (FR). For both the light and heavy variable regions there are three complimentary determining regions (CDR1-CDR3) and four framework regions (FR1-FR4) (figure 1.3).

1.2.5 Monoclonal Antibodies

Antibodies produced from a single cell are homogeneous. However antibody molecules having a common specificity are normally heterogeneous because they are the product of many antibody-producing cells, that is to say, the response to the antigen is polyclonal. This

heterogeneity was a serious impediment to elucidating the molecular basis of antibody action. This hurdle was overcome by taking advantage of multiple myeloma, a malignant disorder of antibody-producing cells. In this cancer a single transformed lymphocyte or plasma cell divides uncontrollably, consequently a large number of cells of a single type are produced. They are a clone and have identical properties. Large amounts of immunoglobulin of a single type are excreted by these tumours. Myeloma immunoglobulins have a normal structure and are typical of normal immunoglobulins, but they are homogeneous. Myelomas also occur in mice, these tumours can be transplanted to other mice, where they proliferate. Furthermore, these antibody producing tumours synthesise the same kind of homogeneous antibody generation after generation.

Myeloma immunoglobulins are advantageous in being homogeneous but their corresponding antibodies are usually not known. In 1975, Köhler and Milstein (1975) discovered that large amounts of homogeneous antibody of nearly any desired specificity can be obtained by fusing an antibody producing cell with a myeloma cell. They are called hybridoma cells.

1.2.6 Enzyme Linked Immunosorbent Assay (ELISA)

ELISA is a very sensitive technique for detecting antigens and antibodies, it is very economical in the use of reagents and a large number of tests can be performed in a relatively short time.

1.3 Fv and Derived fragments.

As described previously Fv fragments are composed of the variable domains of both the heavy and light chains. The protein is stabilised by virtue of non-covalent interactions between the framework regions of two chains. With recombinant DNA technology, cloning of the genes for these variable domains have led to these fragments being expressed in bacteria. The relatively small size of the Fv fragment allied with the use of bacterial expression allowing stable isotope labelling has meant that elucidation of the three-dimensional solution structure of Fv fragments has been solvable, they are also easily crystallised so they are perfect molecules for studies of antibody-antigen interactions. Developments in the molecular biology techniques used in cloning and mutagenesis of wild type antibodies has led to several important developments in the scope and usefulness of antibody fragments.

1.3.1 Structure and Stability

Several methods have been used to stabilise the Fv fragment. The single-chain Fv (scFv) fragment is an Fv fragment which is connected by a polypeptide linker between the carboxy terminus of the heavy chain and the amino terminus of the light chain. This linker was first used in aiding expression of Fab fragments (Bird *et al.* 1988). This chain is usually $(Gly_4 Ser)_3$, the glycine residues allowing flexibility, and the serine residues to provide solubility. The advantage of this linker is two-fold: it can allow a greater range of expression systems to be used; and gives greater stability to the Fv fragment, preventing dissociation of the two chains, that may occur during some of the harsher conditions the fragment may experience during expression and purification. A recent publication

investigated the properties of an Fv fragment compared to six scFv counterparts with different polypeptide linkers (Glockshuber *et al.* 1990). In some cases the affinity for the antigen was improved compared to the affinity of the Fv fragment, but decreased the Fv fragments susceptibility to proteolysis.

By introducing an inter-domain disulphide bond into the Fv fragment the two chains can be covalently linked again increasing stability of the Fv or scFv fragment. In one example the disulphide-scFv (ds-scFv) was shown to have similar binding properties as the scFv and enhanced thermal stability (Young *et al.* 1995). The scFv has also been useful when selecting specific antibodies on phage.

The basic antibody structure has been known since 1973, (Poljak *et al.* 1973) when the first Fab structure was solved, but it was not until 1992 (Harris *et al.* 1992) that the first three-dimensional structure of a complete antibody was solved. The structure of Fv fragments have also been solved: in 1993 a scFv fragment three-dimensional structure was solved by X-ray crystallography (Essig *et al.* 1993); NMR would provide useful information in understanding structures and dynamics of antibodies and antibody-antigen binding, although hampered by poor solubility and short ^{1}H relaxation times, NMR techniques have been used to compare Fv and scFv structures (Freund *et al.* 1993) and an isolated V_L domain has been assigned (Constantine *et al.* 1992). Some of the existing predictive models for antigen binding have been used to improve antigen binding (Reichmann *et al.* 1992) and to design an antibody *de novo*. (Essen and Skerra 1994)

1.3.2. Expression of Antibody Fragments.

The expression of antibody fragments in *E. coli* brings the arsenal of gene technology techniques to antibodies. The fast growth of *E. coli* and its comparatively simple fermentation, makes large-scale fermentation of antibody fragments relatively convenient. There are several different procedures for expressing antibody fragments, the following methods have all been used to produce antibody fragments in *E. coli*. and have been reviewed by Plückthun (1990). It should be noted that only the simultaneous secretion of both chains (or use of a secreted single chain fragment) gives the advantage of folding to the native, functional state.

(i) Direct cytoplasmic expression. The most direct approach is the expression of the antibody chains without signal sequence (Chaudhary *et al.* 1989). Usually inclusion bodies have been obtained with this approach, therefore the resulting polypeptides will be denatured and will have to be renatured *in vitro.*

(ii) Expression as cytoplasmic fusion protein. Cleavable fusion proteins with N-terminal fragments of highly expressed cytoplasmic proteins can be constructed. This strategy can be useful for the construction of variants that are very unstable *in vivo.*

(iii) Secreted fusion proteins. Another strategy that can be used to express antibody domains is to fuse them to another secreted protein. Both N-terminus and C-terminus fusions to antibodies have been constructed (Tai *et al.* 1990; Plückthun and Skerra 1989; and Holland *et al.* 1990) The resulting protein is secreted to the periplasm, after partial lysis of the outer membrane to the medium. This method still requires development before being competitive as a general production method for antibody domains.

(iv) Functional expression by excretion. The simultaneous secretion

of both chains of the antibody gives rise to native and functional antibody fragments. This was first developed for Fv (Skerra and Plückthun 1988) and Fab fragments (Plückthun and Skerra, 1989) by Plückthun and co-workers (1988); and Better and co-workers (1988). The essence of this strategy is to reproduce in *E. coli* the normal normal folding and assembly pathway of antibodies within the eukaryotic cell. In antibody producing cells the two chains are expressed separately as precursors with N-terminal signal sequences and separately transported to the lumen of the endoplasmic reticulum (ER). There the signal sequences are cleaved by a membrane-bound signal peptidase. In the lumen of the ER, folding of the proteins, disulphide bond formation and assembly of the light and heavy chain to complete the antibody take place (Wall *et al.* 1983). The crucial hypothesis in the design of the secretory expression system for antibody fragments was that protein transport to the periplasm of *E. coli* is functionally equivalent to the transport of a protein to the lumen of the ER. A system was thus designed that directs both chains of the antigen binding fragment of the antibody to the periplasm of the same *E. coli*. The expression of antibody fragments was the first example, however, in which the assembly of a periplasmic heterodimer was required.

There are several crucial steps in the expression and secretion pathway that must occur correctly for the Fv or Fab fragment to assemble. (1) Approximately stoichiometric amounts of both chains must be synthesised, this was accomplished by having both genes under the control of the same artificial operon. (2) Both chains must be transported to the periplasm. (3) Both signal sequences must be cleaved at the correct position to yield the identical N-terminus, as in the original antibody molecule. (4) The folding to globular domains must then occur, (5) the

intramolecular disulphide bonds must form, and (6) the two chains must assemble to form the required heterodimer.

There are several advantages to this secretory expression system: firstly it leads directly to an assembled function product with correctly formed disulphide bonds without the need to refold the protein *in vitro;* secondly the problem of protease degradation is diminished as there are fewer proteases in the periplasm than the cytoplasm. Protection is also achieved by the folding to globular domains, accompanied by the oxidation of the S-S-bonds in the periplasm. Since the exposure to the oxidising milieu of the periplasm during the the folding process is necessary for the disulphide bonds to form, the secretion of both chains is an essential requirement. The expression rate must be commensurate with the rates for the transport, folding and assembly as protein will accumulate as insoluble material in either the periplasm or cytoplasm, therefore it is not necessarily useful to employ extremely strong promoters. The *lac* promoter/operator has been found to be useful.

The production of antibody fragments in *E. coli* shows a large variability of success, and may not be as straightforward as initially predicted. These problems have motivated many laboratories to develop production systems in alternative microorganisms including *Saccharomyces cervisiae,* (Wood *et al.,* 1985; Horwitz *et al.,* 1988) *Pichia pastoris,* (Ridder *et al.,* 1995) *Trichoderma reesei* (Nyyssönen *et al.* 1993) and insect cells. (zu Putlitz *et al.* 1990) Although production levels of several hundred milligrams per litre have been reported using some of these host systems, *E. coli* is still the system of choice in most laboratories, probably because of the overall simplicity of the cloning work using this host. Instead,

attempts are being made to improve *E. coli* production strains by, for example, co-expressing chaperone proteins (Dueñas *et al.* 1994) or by production as fusion proteins to facilitate recovery (Laukkanen *et al.*, 1994; Neri *et al.* ,1995).

Interesting data that explain the wide range of production levels in *E. coli* of different antibodies or antibody fragments are starting to emerge. The approach has been, where possible, to utilise protein engineering to convert an antibody that shows a low production level into a variant with high level production while only mutating regions of the molecule outside the CDRs.

1.3.3 Selection of Antibody Fragments Using Phage Display

The technology to display antibody fragments on the surface of filamentous phages (McCafferty *et al.*,1990) using the Gene III protein is an extension of the principle of peptide display (Smith, 1985). The method opened up the possibility of mimicking the B-cell selection system *in vitro* by specifically enhancing phage particles which display antibodies with a given desired specificity (Barbas *et al.*, 1991). Since the initial experiments this technique has been improved dramatically, and many of these improvements are described in these reviews: Hoogenboom and co-workers, (1992); Marks and co-workers, 1995 and Winter and co-workers, (1994).

One current trend in the field has been to apply antibody fragment combinatorial libraries to new applications, for example to raise specific antibodies against human self-antigens using human antibody fragments (Griffiths *et al.*, 1993), and to generate larger and larger libraries to select

for fragments that bind with a K_d in the nanomolar range from a single pot.

In nature the maturation of B-cells into plasma cells/memory cells involves somatic mutations of selected antibodies to improve further affinity from the primary response. This mechanism has been mimicked to some degree in the 'chain-shuffling' technology (Winter *et al.*, 1994). This method utilises a heavy or a light chain of a selected antibody which pairs with a library of heavy or light chains. From this second selection step significantly higher affinities can be isolated with comparison to the primary response.

An alternative to the chain-shuffling approach is to make larger and larger libraries and use these for selection with the hope that a single phage library could be used in obtaining high affinity antibodies against any antigen. It was demonstrated that recently that specific binders to as many as 18 different antigens can be obtained from a single phage-displayed scFv library (Nissim *et al.*, 1994).

1.3.4 Applications of Antibody Engineering.

Antibody engineering has many applications, an early example of this was humanisation, in which CDRs are grafted to from a murine antibody into the framework of a human antibody (Carter *et al.*, 1992). This has proven extremely useful in the generation of more biologically active antibodies for therapeutic agents. The use of the chain-shuffling method simplifies the cumbersome method of humanisation and the use of human antibody fragments on phages should eventually decrease the need for humanisation.

Another application for antibody engineering has been the generation of catalytic antibodies. Areas of this work have expanded the number of reactions catalysed by antibodies, phage display techniques have been used to select antibodies specifically for catalytic functions. The development of catalytic antibodies for use in biotechnological applications such as peptide synthesis and prodrug activation have been developed.

Antibody engineering plays an important role in therapy, of which humanisation is an example. Antibody fragments can also be engineered by making fusions to effector molecules that increase the potential for immunotherapy situations *in vivo*. This fusion approach is likely to expand the use of antibody fragments as targeting molecules in new clinical applications.

1.4 Molecular Biology

The advances described above in all the aspects of producing and the applications of antibody fragments, are only achievable by the important advances in molecular biology that have been made, that allow far easier manipulation of genetic material, thus making expression of the desired protein product possible.

In this section the molecular biological techniques that are required to clone the antibody fragment (Chapter 2) are explained in general terms and also in the application of cloning antibody fragments.

1.4.1 Messenger RNA (mRNA) and its Isolation.

Messenger RNA specifies the primary structure of proteins. Two other types of RNA are also involved in deciphering the genetic code, but

unlike mRNA do not specify the order of amino acid residues in the protein; these are transfer RNA (tRNA) and ribosomal RNA (rRNA). In both eukaryotes and prokaryotes the sequences of tRNAs and rRNAs are contained in larger precursor molecules (the primary transcript of the genes) known as pre-tRNA and pre-rRNA respectively.

In eukaryotes, cytoplasmic mRNAs are derived from much larger precursor molecules known collectively as heterogeneous nuclear RNA (hnRNA) which contain internal sequences known as small nuclear RNAs (snRNAs), and another set of RNAs distinct from tRNA called small cytoplasmic RNAs (scRNAs). At least one of the snRNAs is thought to be involved in the processing hnRNA into functional mRNA.

The structure of typical mRNA is shown in figure 1.4. The message itself is a linear sequence of nucleotide bases. Part of the sequence is made up of the coding region of the mRNA which contains the codons corresponding to the amino acid sequence of the protein, read in the 5′ to 3′ direction. The beginning of the sequence is represented by an initiator codon (usually AUG). The end of the codon sequence is a stop codon (UAA,UGA or UAG). The mRNA non-coding sequence (or untranslated region: UTR) at the 5′ end is the leader sequence, and parts of this sequence are involved in the binding of ribosomes to mRNA. All cytoplasmic eukaryotic mRNA molecules possess a cap structure at their 5′ end. The length of the untranslated sequence at the 3′ end are variable in length. Most eukaryotic cytoplasmic mRNA contain a sequence of 100-200 adenyl units at the 3′ end, which are added post-transcriptionally.

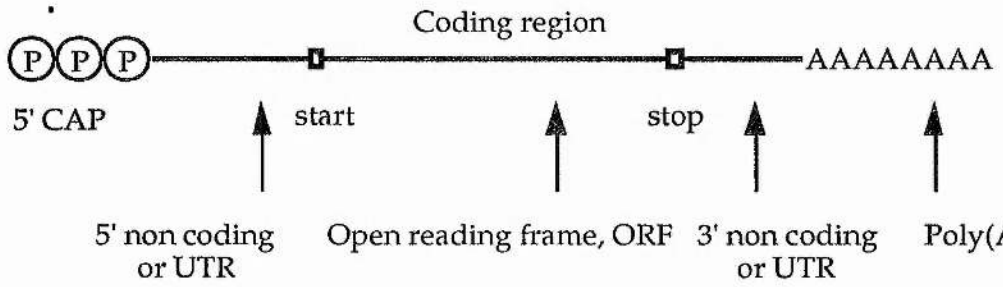


Figure 1.4 Schematic representation of eukaryotic mRNA showing the functional units.

The poly-adenyl or poly(A) tail allows a convenient method of purification (Edmonds *et al.*, 1971), by affinity chromatography with cellulose beads with poly-thymidyl or poly(T) nucleotides, which will base pair with the poly(A) tails of the mRNA at room temperature and high ionic strength. When working with any RNA care has to be taken from the presence of RNase contamination, so great care has to be taken in preparing glassware, plasticware and solutions used in the isolation experiment.

1.4.2 The Polymerase Chain Reaction.

Trying to isolate a small piece of DNA from a large eukaryotic genome represents a very difficult task, however the complimentary nature of DNA sequences allows small pieces of DNA to be used as probes. Cloning DNA by this method would be a laborious process: firstly the genome would have to be cut into smaller pieces by restriction endonucleases (see section 1.4.3); these pieces placed into bacterial plasmids, and each of these grown as 'clones' in bacterial colonies; and then a radiolabelled oligonucleotide to act as a probe to find which clone contained the piece of DNA required. This colony could be isolated and the required DNA amplified or cloned in this way. With the advent of PCR (Mullis *et al.*, 1986) the laborious task of finding the piece of the genome required by

simply using two oligonucleotide primers at each end of the piece of DNA that needs to be amplified, and are complimentary to the sequence of the ends.

The polymerase chain reaction works in the following way (figure 1.5). Two oligonucleotide primers are usually synthesised chemically, the DNA or cDNA to be amplified is heated to around 90 °C to separate the strands of DNA, the reaction is then cooled allow the annealing of the primers to the target strands of DNA. The temperature of this annealing step will decide the stringency with which the primers will bind to the target DNA. Annealing temperatures can be calculated by their base compostion of the primers. The annealing temperature can be varied in different reactions to assure the correct gene can be isolated and that other unwanted genes are not amplified. The final step of the reaction is incubate the sample in the presence of DNA polymerase and dNTPs (deoxynucleoside triphosphates - 2'-deoxyadenosine 5'-triphosphate (dATP), 2'-deoxycytidine 5'-triphosphate (dCTP), 2'-deoxyguanosine 5'-triphosphate (dGTP), thymidine 5'-triphosphate (TTP)) to allow synthesis of the complimentary strand. This would yield a new strand of DNA importantly with the sequence of the primers at each end of the amplified DNA. Each round of the reaction will double the isolated gene, so providing sufficient dNTPs, primers and active enzyme, any amount of DNA can be produced theoretically from one copy of the DNA. Use of the thermostable *Taq* DNA polymerase from the thermophilic bacterium *Thermus aquaticus*, avoids the tedious process of having to remove the inactive protein and add new DNA polymerase, which would be necessary if a non-heat stable enzyme was used. Another heat-stable polymerase, *Vent* DNA polymerase (New England Biolabs) can be used.

Vent DNA polymerase has a proof reading 3′ to 5′ exonuclease.

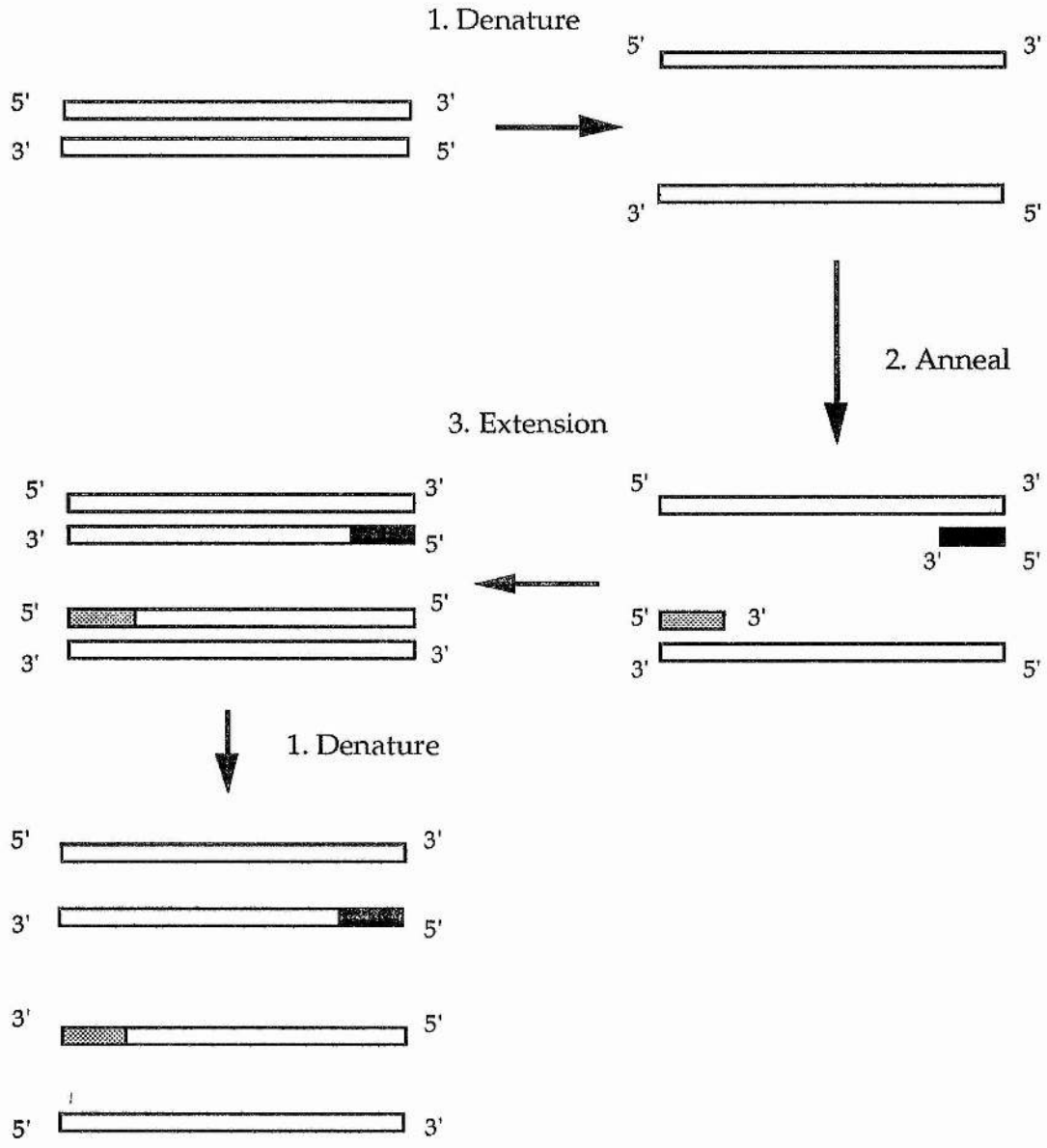


Figure 1.5 Diagram of the polymerase chain reaction (PCR).

PCR is of great value in isolating antibody light and heavy chain variable genes. Once cDNA has been synthesised with reverse transcriptase starting from an area in the C_H1 and C_L regions for heavy and light chains respectively, this cDNA product is suitable for PCR. The sequence of primers can be devised using DNA sequence data from Kabat (Kabat, 1991), and also from published PCR primers for isolating murine heavy and light chain variable genes.

1.4.3 Restriction Endonucleases.

Restriction endonucleases were discovered when it became apparent that bacterial DNA could detect 'foreign' DNA and degrade it, while DNA from the same strain was not degraded. The non-foreign DNA is modified such that it is no longer hydrolysed by the restriction endonuclease. A restriction endonuclease recognises a specific sequence in the DNA, and if these bases have not been modified, will hydrolyse the sugar-phosphate backbone.

Each bacterial species has one or more restriction endonuclease that recognises a different sequence. In the vast majority of cases the sequence recognised is a palindromic, the phosphate backbone is cleaved between the same two bases on each strand - it is possible to have a staggered cut, giving cohesive ends or a cut in the middle of the sequence leading to blunt ends (figure 1.6).

By the use of these endonucleases the whole idea of engineering genes is made simple. Two pieces of DNA from different sources, but cut with the same enzyme, can be joined by their cohesive ends due to the base pairing, and the broken phosphate backbone repaired by DNA ligase. By

using the different restriction endonucleases at each end of a particular gene and using the corresponding endonuclease sites in a plasmid, genes for expression can be added in the correct direction. Sequences recognised by particular restriction endonucleases can be placed into PCR primers, so that any gene amplified with this method can then be easily manipulated into suitable vectors.

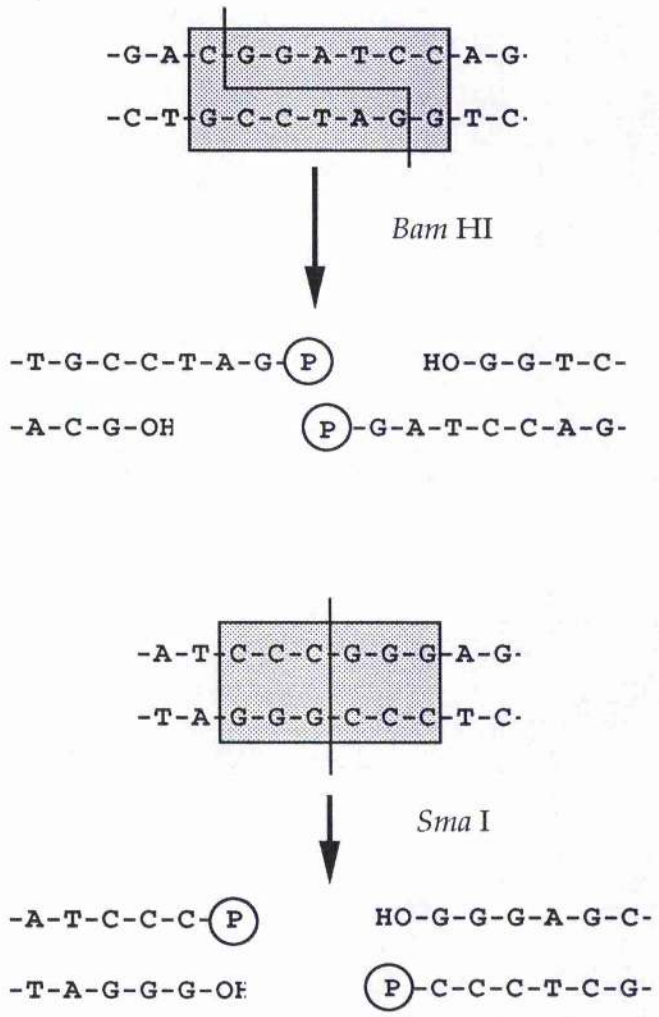


Figure 1.6 Diagram of the action of the restriction endonucleases *Bam* HI (cohesive ends) and *Sma* I (blunt ends).

In cloning the Fv variable genes, restriction endonuclease sites are added *via* the PCR primers and make the manipulation of these genes simple. One problem is some of the restriction sites often occur elsewhere in the gene, which therefore requires mutagenesis to change the sequence. Plasmids and PCR primers have been produced that use restriction endonuclease sites that are very uncommon in antibody variable genes.

1.4.4 Plasmids

Plasmids are self-sustaining pieces of DNA that are often found in bacteria and some lower eukaryotes. They are not essential for growth, but they confer some unusual properties to the cells that harbour them. Plasmids replicate independently of the chromosome and are generally circular DNA molecules, although a few linear plasmids have been found.

The properties conferred by plasmids are extremely varied. The first plasmid identified in *E. coli* had the ability to participate in simple conjugal exchange of genetic information. This plasmid is known as the F factor (fertility), it is 96 kb and it is transmissible: the ability to promote the act of conjugation and be transferred itself from one bacterial cell to another. The R factor plasmid confer resistance to antibiotics and are also transmissible.

One of the first accomplishments of genetic engineering was to develop plasmids into suitability for cloning and expressing pieces of DNA. The pUC plasmids (Messing 1983; Norrander *et al.*, 1983; Yanisch-Perron *et al.* 1983) have an ampicillin resistance gene and a *lac*Z gene that has a series of unique restriction sites into which cloned DNA can be inserted. The

ampicillin resistance allows only bacteria that contain that plasmid to be selected by growing in a medium that contains ampicillin. If these plasmids are transformed into a *lac*Z$^-$ strain of *E. coli* they will make the strain *lac*Z$^+$, and if grown on plates containing the artificial β-galactosidase substrate Xgal, then the colonies will appear blue. A plasmid which contains a DNA insert, thereby interrupting the lacZ site, thus unable to cleave Xgal, and will appear as white colonies - these white colonies growing on ampicillin medium will therefore be bacteria containing the transformed plasmid with a DNA insert. The plasmid also contains the *lac* operon and promoter that allows control of transcription, then translation to the protein coded by the cloned gene by the use of isopropylthiogalactoside (IPTG) (see section 1.4.6).

The plasmid used to express the Fv fragment is a modified pUC 19 vector and is described in detail in Chapter 2 (Better *et al.*, 1988; Ward *et al.*, 1989; Orlandi *et al.*, 1989).

1.4.5 Transformation of DNA into Bacteria.

Transformation has been described in many species of bacteria that have the ability to take up DNA naturally, only a small percentage of the cells have this ability and they are described as competent. In *E. coli,* competent cells do not occur naturally, but can be made so by treating with $CaCl_2$ or by electroporation. The electroporation technique is more efficient with up to about 10^9 transformed bacteria per microgram of DNA.

1.4.6 Expression of Proteins.

To express proteins coded for by DNA inserts in the pUC 19 vector, the

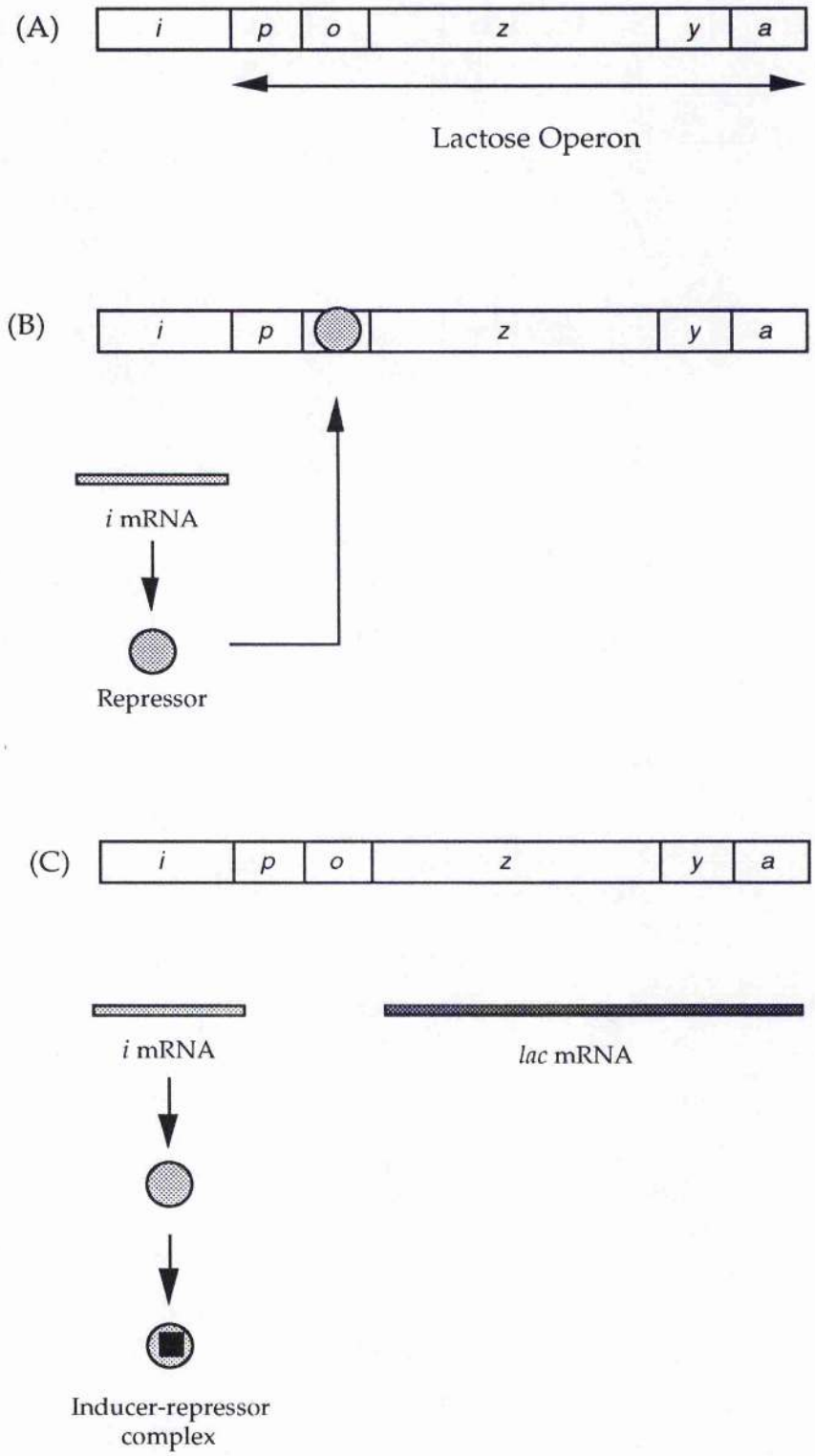


Figure 1.7 (A) Map of the lactose operon and its regulatory gene, (B) the lactose operon in the repressed state and (C) induced state.

lac operon gene, that controls the production of β-galactosidase in *E. coli.*, is used Experiments on the control of this gene led to the proposal of the operon model for the regulation of protein expression. The genetic elements of this model (figure 1.7) are known to work in the following way. The regulator gene produces a repressor protein which interacts with the operator. The operon also contains a promoter for the binding of RNA polymerase - a site for the initiation of transcription. An inducer, such as IPTG, binds to the repressor, preventing it from interacting with the repressor. The z,y and a genes can then be transcribed to a polygenic transcript, coding for all three molecules. A modified version of the *lac* operon is used in the pUC 19 vector. The different strategies for bacterial expression, especially when applied to antibody fragments is discussed in section 1.3.3.

1.5 Theoretical determination of conformation

Because nearly all of the relevant conformational information of E3G comes from the glycosidic linkage, using modelling procedures applicable to carbohydrates would be the most suitable method of theoretically determining the conformation of this glycan.

1.5.1 Modelling of Carbohydrates.

To determine the geometry of a oligosaccharide, it is necessary to apply theoretical analysis of computer-generated models. The atomic coordinates of an input structure are optimised to determine a minimum energy conformation. Methods for conformational searching and energy calculation procedures have been reviewed (French and Brady, 1989).

Full *ab initio* quantum mechanical computations of oligosaccharide conformations are unattainable, due to the computationally intensive calculations required. Hence computations of oligosaccharides conformation have relied almost entirely on molecular mechanical methods (Homans *et. al.*, 1987; Yan and Bush, 1990; Brady, 1986, 1987; Madsen *et. al.*, 1990). This approach treats the molecule concerned as a number of point masses and charges, and a series of springs for bonds between them. Empirical data from X-ray crystallography or *ab initio* studies of small organic molecules provides optimum values for bond-lengths and torsion angles, deviation from which increases the energy potential of the model. Classical mechanics can then be used to derive an energy function for the model, allowing identification of lower energy configurations by differentiation. Additional factors affect the application of these methods to carbohydrates, particularly the exo-anomeric effect (reviewed by Tvaroska and Bleha, 1989), the effect of which is to make the aglyconic carbon take up orientations of $\phi_H \sim +60°$ in β-D-glycosides, and $\phi_H \sim -60°$ in α-D-glycosides.

Early conformational calculations by Lemieux and co-workers utilised the hard sphere exo-anomeric (HSEA) approach (Lemieux *et. al.*, 1980), (Thogerson *et al.*, 1982). These calculations utilised a simple forcefield, including a Kityagordski term to describe nonbonded interactions, rigid ring geometries, neglect of partial atomic charges, and a torsional potential for the exo-anomeric effect. Results in good agreement with experimental and theoretical considerations were obtained for the blood-group oligosaccharides (Lemieux *et. al.*, 1980; Thogerson *et. al.*, 1982), sucrose (Bock and Lemieux, 1982), and structures related to the complex-

type N-linked glycans (Bock *et. al.*, 1982). These studies posit oligosaccharides as having essentially fixed orientations about the glycosidic linkages, with the exception of 1-6 linkages for which a number of orientations were predicted. Good agreement was seen between these studies, and NMR or X-ray crystallographic experimental data. However in extending this approach to structures larger than 4-5 residues, agreement is poorer, and it becomes apparent that a degree of conformational flexibility may be present.

Inherent in the HSEA approximation is the exclusion of certain terms from the forcefield, and the result of a given simulation depends upon the terms included (Cumming and Carver, 1987). HSEA calculations give good results for static, minimum energy, structures. However, to investigate dynamic models, where the potential surface encompasses a wide range of ϕ and ψ angles, it now appears that a full molecular mechanical forcefield is more appropriate. The AMBER parametrisation (Weiner *et al.*, 1984, 1986), is a full forcefield, originally for nucleic acids and proteins, and has been extended to include carbohydrates by Homans (1990), based upon work by Ha and co-workers (1988a). A full molecular mechanical forcefield allows explicit inclusion of solvent water, in contrast to calculations done *in vacuo*.

HSEA calculations suggest deep potential wells, supporting the concept of rigid conformations. However, allowing all bonds, angles and torsion angles to vary during minimisation (flexible geometry) increases the area of conformational space accessible to a given glycosidic linkage (Scarsdale *et al.*, 1988; Homans *et al.*, 1987; French, 1989; Ha *et al.*, 1988b; Homans *et*

al., 1989; Yan and Bush, 1990; Imberty *et al.*, 1989). Energy contour map calculations (Ramachandran plots) in which ϕ_H and ψ_H are systematically varied in small steps (10° or 15°) and the geometry optimised for this value of the glycosidic angles to produce relaxed potential surfaces, allows exploration of the extent of accessible conformational space. The results may be presented as an iso-energy contour map, in which occupation is temperature dependent, governed by the Boltzmann law (French and Brady, 1989). Calculation of relaxed maps is preferable to the use of a rigid-residue approach, since bad contacts may be relieved by relatively small changes in the internal coordinates (French, 1988; Ha et. al., 1988b; Goebel et. al., 1970). Clearly this approach becomes impractical in larger molecules, since the number of degrees of freedom rapidly increases both the number of calculations that must be carried out, and consequent computation time. In addition, the problem of false or multiple minima is accentuated, since it cannot be ensured that at each ϕ/ψ point the molecule is at a minimum energy.

An alternative method for the determination of accessible conformational space is the calculation of molecular dynamics simulations (Homans et. al., 1987; Yan and Bush, 1990; Brady, 1987; Brady, 1986; Madsen et. al., 1990). Studies on the behaviour of model saccharides during molecular dynamics simulations has highlighted several important points. Firstly, whilst structures may be restricted to low-energy regions of conformational space, they are by no means rigid. A number of conformations may exist independently during the period of a simulation, each consistent with experimental data. Secondly, internal motions occur on a timescale of picoseconds, substantially faster than the

rate of overall tumbling for even moderately-sized molecules. In this case, calculation of theoretical NOEs (see following section) is complicated by the necessary inclusion of spherical harmonics terms, accounting for the time-dependent orientation of the vector between the NOE-coupled pair and the molecular axes (Tropp, 1980). Molecular dynamics simulations are presently limited by the length of simulated time which it is practicable to model. The timescale of the build-up of the experimental NOE is of the order of milliseconds, whilst current computing technology renders the simulation of periods greater then tens of nanoseconds impractical. Comparison of experimental and theoretical NOEs must be subject to this important caveat. Explicit inclusion of solvent water substantially increases the length of time required for a simulation, and hence it is often more practical to simulate the existence of solvent by using a dielectric constant ε=80.

Determination of the minimum energy structure may alternatively be reached by means of restrained simulated annealing (Clore et. al., 1985). Convergence upon similar final conformations, from a number of different starting conformations ensures that a larger proportion of conformational space has been sampled compared to minimisation by differentiation of the energy function.

1.5.2 Homology Modelling of Proteins

With the fast rate of newly published protein and nucleic acid sequences in comparison to the relatively slow rate of published X-ray and NMR structures, it would be advantageous to develop a method which could predict protein structure from its amino acid sequence. Since *a priori*

folding rules for proteins have not been developed, it is necessary to base any structural predictions on the conformation of reference proteins. Thus it is assumed that the structures of unknown and sequentially homologous reference proteins are homologous and that a model of the unknown protein can be built from the reference protein.

Early work dealing with building a protein by homology used only one known structure (Brown *et al.*, 1969; Shotton *et al.*, 1970). Amino acid similarities between the known and unknown proteins were used to determine where one protein would resemble the other. Sequence alignment was done and the coordinates of the reference protein were used to predict those of the unknown protein.

More structural approaches have been used by Greer, and Blundell and co-workers (Greer, 1980, 1981, 1985; Blundell *et al.*,1987, 1988). In their methods more than one reference protein is used and greater emphasis is placed on the conformational similarities between them. Less emphasis is placed on sequential alignment alone as a basis for a model. By determining the regions of the protein that do not vary from one member of a protein family to another, there is greater confidence that extrapolation of structure for that region to a new protein will be more accurate. To homology model a protein the following steps are needed:

1. Determine which proteins are related to the model protein.
2. Determine structurally conserved regions (SCRs).
3. Align the amino acid sequence of the unknown protein with those of the reference proteins within the SCR.
4. Assign coordinates in the conserved region.
5. Predict conformations for the rest of the peptide chain, including

loops between the SCRs and the N- and C-termini.

6. Search for the optimum side chain conformations for residues that differ from those in the reference proteins.

7. Use of energy minimisation and molecular dynamics to refine the molecular structure so that steric strain introduced during the model-building process can be relieved.

1.6 NMR spectroscopy.

The nuclear magnetic resonance phenomenon, first observed in 1946, has become a widespread tool for the non-destructive analysis of both organic and inorganic compounds. NMR techniques can provide information on structure, conformation, and internal mobility. Comprehensive introductions include Sanders and Hunter (1987), Derome (1987).

The fundamental basis of the NMR experiment is perturbation by a radiofrequency pulse, of equilibrium magnetisation from an axis parallel with the static magnetic field, B_0, into a vector perpendicular to this axis, precessing with a characteristic (Larmor) frequency. The value of the Larmor frequency is due to the extent to which the nucleus is shielded from the external field by nearby electron-inductive groups.

$$\omega = \gamma B0 \qquad \textit{Equation 1.1}$$

Resonance assignment, even of simple oligosaccharides, is complicated by the tendency for the majority of resonances in both ^{1}H and ^{13}C spectra to lie within an unresolved envelope, spanning only a few hundred Hertz. Anomeric proton or carbon resonances are distinct from this envelope,

being well resolved at low field, due to the electron-withdrawing ring oxygen. Problems of coincident resonances are partially resolved by separation into a second or third dimension, increasing spectral dispersion. The precise nature of the information contained in the spectrum depends upon the details of the pulse sequence and phase cycling used, however a number of key points are shared by all 2D experiments. A preparation pulse or sequence of pulses generates the desired coherences, which then evolve under a free precession Hamiltonian during an incremented delay (t_1). Transfer of magnetisation to other spins occurs, and the signal is detected during t_2. 2D spectra can then be processed by application of two orthogonal Fourier transformations. A 3-dimensional spectrum may be generated by insertion of a further evolution period, and a third orthogonal transformation is required.

Dispersion into 2 or more dimensions can be accomplished using either a homonuclear or heteronuclear methods. Homonuclear experiments include COSY, HOHAHA, and 3D HOHAHA-COSY. Heteronuclear methods allow observation of correlations between heterologous nuclei, such as ^1H and ^{13}C. Experiments can be 'tuned' to observe spin-couplings of particular size, by optimisation of delay periods, for instance $^1J_{CH}$~140 Hz, $^2J_{CH}$~0-4 Hz, or $^3J_{CH}$~0-7 Hz.

The sensitivity of C-H correlation spectra can be significantly improved by detection of the ^1H signal (inverse-detection), by sequences such as INEPT (Morris and Freeman, 1979) (Insensitive Nuclei Enhanced by Polarisation Transfer). This enhancement is due to two factors; firstly, the receiver coil

is more sensitive to signals of higher frequency (Cavanagh et. al., 1988), thus ^{1}H signals which resonate at a higher frequency than ^{13}C nuclei in the same magnetic field, will be detected with greater sensitivity. Secondly the intensity of the signal will be enhanced by γ_I/γ_S, where γ_S is the magnetogyric ratio of the less sensitive spin, due to the larger population difference for the sensitive spin. An additional benefit is that signal repetition can be on the order of T_1 for the sensitive nucleus, allowing faster acquisition.

1.7 Conformational analysis by NMR.

The following three parameters are particularly important in the conformational analysis of oligosaccharides:

Relaxation properties (NOE, T_1, T_2).

Spin-spin coupling constants (J).

Conformation dependent chemical shifts.

1.7.1 Relaxation properties of perturbed nuclei.

Detailed explanations of NMR relaxation phenomena are given by Noggle and Schirmer (1971), and Neuhaus and Williamson (1989).

After application of a pulse, or other perturbation, nuclei relax to their equilibrium states by one of two mechanisms. Longitudinal (T_1) relaxation causes the population difference between two spin states of a given nucleus to return exponentially to equilibrium, due to transfer of energy to the surroundings or 'lattice'. In terms of classical formalism, bulk magnetisation in the X-Y plane returns to the Z axis. The second

mechanism, spin-spin (T_2) relaxation causes a loss in phase coherence due to mutual exchange of spin energy, and resultant decay in the bulk intensity of transverse vectors. (Bloch model).

Both mechanisms of relaxation are due to time-dependent magnetic or electric fields, derived from random thermal motions within the sample. For both ^1H and ^{13}C nuclei, the major sources of such fields, the intensity of which is termed the spectral density $J(\omega)$, are the magnetic moments of neighbouring protons. Relaxation is most efficient for both T_1 and T_2 when the timescale of such interactions is at or near the Larmor frequency. T_2 is additionally sensitive to low-frequency modulations of spin energy levels. Thus the relaxation mechanisms are dominated by the effects of dipole-dipole interactions with adjacent ^1H nuclei. The strength of these dipolar interactions are primarily dependent, among other factors, upon the internuclear distance, and the mobility of the vector within the static B_0 field. Thus these relaxation phenomena can provide information about both geometric and dynamic behaviour.

Nuclear Overhauser Effect

Consider a system of two ^1H spins, I and S, that are not dipolar coupled, but are close in space.

We can define transition probabilities between states as shown in figure 1.8. The populations of each energy level are given by the Boltzmann distribution. Application of a radiofrequency pulse at the frequency of S saturates this spin causing the populations across the S spin transitions to be equalised. Relaxation will proceed by the various pathways, including

W_2 and W_0 double- and zero-quantum spin transitions. W_2 transitions are promoted by magnetic fields fluctuating at ~2ω, and W_0 by low frequency fluctuations. Hence in slowly tumbling molecules, W_0 processes dominate, causing a reduction in the intensity of the I spin, resulting in a -100% NOE in the limit. In a rapidly tumbling molecule, W_2 predominates, causing a positive enhancement. When the energy level populations are disturbed from equilibrium, the I spin also relaxes by transferring energy to its neighbouring protons. These indirect enhancements are also negative, and after prolonged irradiation of S the transfer of indirect NOE enhancements ('spin-diffusion') causes saturation of all spins in the molecule, and hence the distance proportionality of the NOE magnitude is lost. The rate of cross-relaxation, s, proportional to r^{-6}, is a more convenient parameter to use than the enhancement itself.

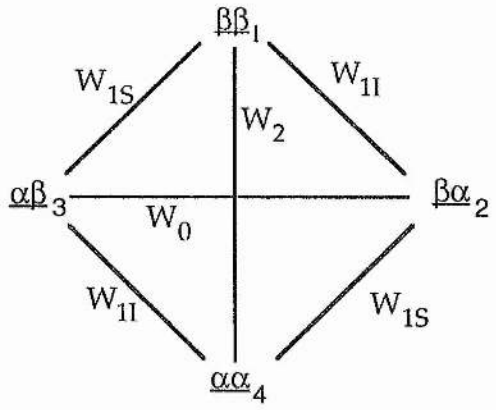


Figure 1.8 : Energy level diagram for two non-coupled spins, I and S, in close spatial proximity. W_1, W_2 and W_0 represent the transition probabilities for single double and zero quantum processes. Spin states are labelled α or β.

In a multispin system, the rate of intensity change for spin I is:

$$\frac{dI_Z}{dt} = -R\left(I_Z - I_Z^0\right) - \sigma_{IS}\left(S_Z - S_Z^0\right) - \Sigma\sigma_{IM}\left(M_Z - M_Z^0\right) \quad \textit{Equation 1.2}$$

where I_z and I_z^0 are I spin intensities along the Z-axis at time zero and time t respectively; R_I is the relaxation rate of I; and M represents all other spins in the system. The initial rate, when $I_z=I_z^0$, $M_z=M_z^0$, and $S_z=0$, is given by:

$$\frac{dI_Z}{dt} = \sigma_{IS}S_Z^0 \quad \textit{Equation 1.3}$$

For homonuclear ^1H-^1H interactions, s is given by:

$$\sigma_{IS} = \left(\frac{\mu_0}{4\pi}\right)^2 \frac{\gamma_I^2\gamma_S^2\hbar^2}{10r_{IS}^6}\left[6J(\omega_I + \omega_S) - J(\omega_I - \omega_S)\right] \quad \textit{Equation 1.4}$$

Where $\hbar$ is Planck's constant divided by 2π, μ_0 is the permeability of free space, r_{IS} is the internuclear distance, and γ is the magnetogyric ratio of a proton. J(ω), the spectral density function is defined as

$$J(\omega) = \frac{\tau_c}{1+\omega^2\tau_c^2} \quad \textit{Equation 1.5}$$

τ_c (the correlation time for molecular reorientation) is inversely related to rates of molecular motions, and is identical for each ^1H-^1H vector in a

rigid isotropically tumbling molecule.

Using the initial rate approximation, immediately after saturation of S, the NOE at I is affected by only one distance-dependent term, σ_{IS}. Therefore, in such situations, all enhancements behave as though they were in a two-spin system, and those measured during truncated driven experiments, when the saturation period is within the linear build-up region, are in proportion to the internuclear separation from S. With longer presaturation times, indirect enhancements develop, and distance specificity is lost.

When all constant terms in equation are gathered together as a single constant k, the equation simplifies to;

$$\sigma=kr^{-6} \qquad \textit{Equation 1.6}$$

Then, when an NOE enhancement is observed between two protons, (I and M), and the distance between I and S is known, the S-M distance can be estimated from a simple ratio calculation:

$$r_{SM}=r_{IS}(s_{IS}/s_{SM})^{1/6} \qquad \textit{Equation 1.7}$$

Due to the r^{-6} dependence of S small inaccuracies in measured NOE enhancements have negligible effects in calculated internuclear distances. However, when using the reference distance method, there are several points which must be appreciated:

1. The accuracy of the calculated S-M distance relies upon that of the reference distance.

2. Inaccuracies are introduced by non-instantaneous saturation.

3. The method is only valid for the initial rate approximation

4. Integration of spectra, particularly with overlapping or low intensity signals can be inaccurate.

5. Internuclear vectors connecting IS and SM must have the same effective correlation time (τ_e). This will not be the case if the molecule exhibits anisotropic tumbling or has flexibility in the S-M distance.

6. In the event of internal motions, enhancements are heavily weighted by the conformations with closest contact, due to the r^{-6} dependence. It is the internuclear distance, rather than the enhancement itself which is averaged.

Generally, in an isotropically-tumbling, rigid molecule, the S-M distance can be calculated to within ~10%. However, in flexible polysaccharides (5 and 6 above), inequalities of τ_c become relevant. Internal mobility adds uncertainties to the proportionality of σ and r^{-6} (Genest, 1989). For a flexible molecule in multisite conformational exchange that is slow on the τ_c timescale, the effective internuclear distance is simply a time-average of the separation at each individual conformation $\langle r^{-6}\rangle^{-1/6}$. However Tropp showed that during fast exchange (e.g methyl group rotations) s is a function of the $\langle r^{-3}\rangle^2$ average distance, and has also an angular dependence (Tropp, 1980). Use of Tropp averaging has important implications when computing simulated enhancements involving flexible molecules.

Lipari and Szabo (1982 a,b) have proposed a modification of the spectral density function to allow for the effects of fast internal motions. Their approach is independent of a motional model, but includes a

parameter (S^2 the generalised order parameter) for describing spatial restrictions of the internal motions. However S^2 describes movement of the internuclear vector, and is not applicable where the internuclear distance is varying. It is therefore particularly suitable for ^{13}C relaxation, since r_{CH} is constant. The Lipari-Szabo approach can be used to explain the experimental measurements, but does not assist in developing a dynamic model of the analyte (Dais and Perlin, 1987).

Calculated distances, with error limits, can thus be used to limit the available conformational space during molecular modelling calculations.

T_1 and T_2

The kinetics of relaxation are often expressed either as the rate at which net magnetisation is restored along the Z-axis (Spin-lattice relaxation, R_1), or as the rate at which signal is lost from the XY plane (spin-spin relaxation R_2). In macromolecules, the rates of these processes are not equal, in part because W_0 transitions increase the rate of spin-spin relaxation, but result in no longitudinal relaxation. Each process is described by a first-order rate expression, and therefore two rate constants must be defined: R1 and R2. Frequently these are expressed in reciprocal form as relaxation time constants $T_1=1/R_1$ and $T_2=1/R_2$.

$$M_0 - M_Z = C\exp\left(\frac{-t}{T_1}\right) \qquad \textit{Equation 1.8}$$

M_0 is the net magnetisation, M_z is the magnetisation component along

the Z-axis, C is a constant, and t is time. If relaxation proceeds by different pathways, the rate process is not described by a single exponential function.

The ^{1}H T_1 for a given spin is simply the sum of all cross-relaxation (σ) and auto-relaxation (ρ) rates. The dominant relaxation mechanism for ^{13}C nuclei is *via* its directly attached protons, and its rates can be calculated from:

$$\frac{1}{T_1} = \frac{\gamma_C^2 \gamma_H^2 \hbar^2}{4 r_{CH}^6} \left[J(\omega_H - \omega_C) + 3J(\omega_C) + 6J(\omega_H + \omega_C) \right]$$

Equation 1.9

and;

$$\frac{1}{T_1} = \frac{\gamma_C^2 \gamma_H^2 \hbar^2}{8 r_{CH}^6} \left[4J(0) + J(\omega_H - \omega_C) + 3J(\omega_C) + 6J(\omega_H) + 6J(\omega_H + \omega_C) \right]$$

Equation 1.10

The value of the ^{13}C-^{1}H NOE can be calculated from:

$$NOE = 1 + \frac{\gamma \left(6J(\omega_H + \omega_C) - J(\omega_H - \omega_C) \right)}{\gamma \left(J(\omega_H - \omega_C) + 3J(\omega_C) + 6J(\omega_H + \omega_C) \right)}$$

Equation 1.11

where r_{CH} is the ^{13}C-^{1}H internuclear distance, and τ_c in the spectral

density expression (equation 1.5), relates to the reorientation of the C-H vector. ^{13}C T_1 data are sensitive to the anomeric configuration of sugar residues (Hall and Preston, 1972), since the anomeric C-H bond lengths are shorter for a-sugars than for β sugars (Wolfe *et al*, 1990).

Equations 1.9, 1.10 and 1.11 indicate that like the ^1H-^1H NOE, T_1, T_2 and the $^{13}C\{^1H\}$ NOE are sensitive to τ_c, but since the only distance dependent term is r_{CH}, ^{13}C T_1, T_2 and NOE are influenced only by molecular motions, without complication by conformational effects, and may therefore be used for investigating molecular dynamics.

In large macromolecules, overall tumbling rates are relatively slow, but may exhibit rapid internal motions, which are then represented by an effective correlation time (τ_e), to account for both internal and external mobility. It can be shown that in such systems, T_1 is insensitive to slow molecular tumbling rates, being most strongly affected by internal mobility (Zhang et. al., 1990). However, W_0 transitions do not contribute to T_1 relaxation, but do affect the dependence of T_2 to slow motions, as the term J(0) in the spectral density expression. Hence T_1 can be considered a probe of internal mobility, and T_2 of slower motions, including molecular tumbling.

1.7.2. Spin-coupling constants.

The degree of atomic orbital overlap affects the magnitude of spin-coupling constants, which are therefore related to the dihedral angle (q) between vincinal coupled spins (Karplus, 1959; 1963). The generalised

Karplus relationship is applicable to both homonuclear and heteronuclear spin-coupling constants, and is of the form:

$$J = A\cos^2\theta + B\cos\theta + C$$

Where A, B and C are constants for which different values have been proposed, Haasnoot has included additional terms for substituent electronegativities (Haasnoot *et. al.*, 1980). The ^1H-^1H spin-coupling constants and chemical shifts of pro-R and pro-S hydroxymethyl resonances have been assigned unequivocally for several hexapyranosides (Nishida *et al.*, 1987). H-6 $_{proR}$ invariably has a greater chemical shift and larger H5-H6 coupling constant than H-6 $_{proS}$. The rotamer distribution around the C5-C6 bond can be found from analysis of the ^1H-^1H spin-coupling constants.

Of particular use for oligosaccharide analysis are the $^3J_{CH}$ values across glycosidic linkages, since these have been correlated with the glycosidic torsion angles (Mulloy *et.al.*, 1988; Tvaroska *et. al.*, 1989).

The dependence $^1J_{CH}$ on stereoelectronic effects has been termed the Perlin effect (Wolfe *et al.*, 1990). Tvaroska (1990) has recently proposed the following correlation between dihedral angle and $^1J_{CH}$ in small organic compounds;

$$^1J_{CH} = A\cos^2\theta + B\cos\theta + C\cos\theta + C\sin^2\theta + D\sin\theta + E$$

The variation of this function at torsion angles between ±100° is small in comparison with the error limits, but a single maximum coupling constant is observed close to 180°. The information may then be complementary to $^3J_{CH}$ in oligosaccharides, which shows two maxima at

0° and 180°

1.7.3 Conformation-dependent chemical shifts (not including ring current effects).

Conformation-dependent chemical shifts are observed when substituents on adjacent residues are in close proximity. For example when the distance between a proton Hx and a neighbouring substituent Y is short, a force exerted along the C-H bond axis affects the chemical shift of both Hx and its attached carbon (Grant and Cheney, 1967). In carbohydrates, the difference between the chemical shift of a resonance in a monosaccharide compared with the same signal when the residue is part of a larger structure is termed a glycosylation shift.

^{1}H-^{1}H proximity generally causes up-field glycosylation shifts on the resonances of both protons and attached carbons (Backman *et. al.*, 1988). Proton-oxygen proximity shifts the ^{1}H $^{13}\underline{C}$-H and $^{13}\underline{C}$-O resonances downfield (Lemieux and Bock, 1983; Baumann, 1988).

Glycosylation shifts are complicated by many factors including the orientation of oxygen lone-pairs (Baumann, 1988) and conformational averaging, hence there is no simple relationship between the magnitude of shifts and internuclear distances in carbohydrates. ^{13}C glycosylation shifts have been roughly correlated with O-H distances, (Baumann, 1988), torsion angles of skeletal bonds such as the glycosidic torsion angle ψ (Horii, 1984; Veregin *et. al.*, 198; Gidley and Bociek, 1988), and the deformation of bond angles (Saito and Ando, 1989).

Chemical shifts of carbon nuclei close to anomeric linkages are sensitive to the stereochemistry of the linkage (Hall and Johnson, 1969; Shashkov *et. al.*, 1988), including the absolute configuration of both residues, the configurations of anomeric and linkage carbon atoms and the position of the linkage on the aglycone ring. The magnitude of changes are small compared with the total range of chemical shifts and are rarely greater than 10 ppm for ^{13}C signal and 0.5 ppm for ^{1}H.

1.7.4 Ring Current Shifts.

Ring current effects are a major source of conformation dependent shifts observed in proton NMR spectroscopy. These spectroscopic results can be interpreted in terms of the local structure in terms of the aromatic rings in the local environment. These ring current effects provide a link between spectroscopy and structure both in static and in dynamic terms.

In proteins ring current effects originate from the aromatic rings of His, Phe, Tyr, and Trp residues. The heme in cytochromes, myoglobins and hemoglobins also give rise to large ring current effects.

Such effects arise from the circulation of the delocalised π electrons of aromatic rings around the periphery of the aromatic ring. (figure 1.9). This sets up a local magnetic field which opposes the external field set up by NMR spectrometer magnet. Chemical shift changes result for the resonances of protons which are close to the aromatic ring (<0.7 nm from the centre of the ring). The shifts are upfield above and below the ring plane, where the local field reinforces the external field, and downfield in the ring plane (figure 1.10). Therefore protons attached directly to the

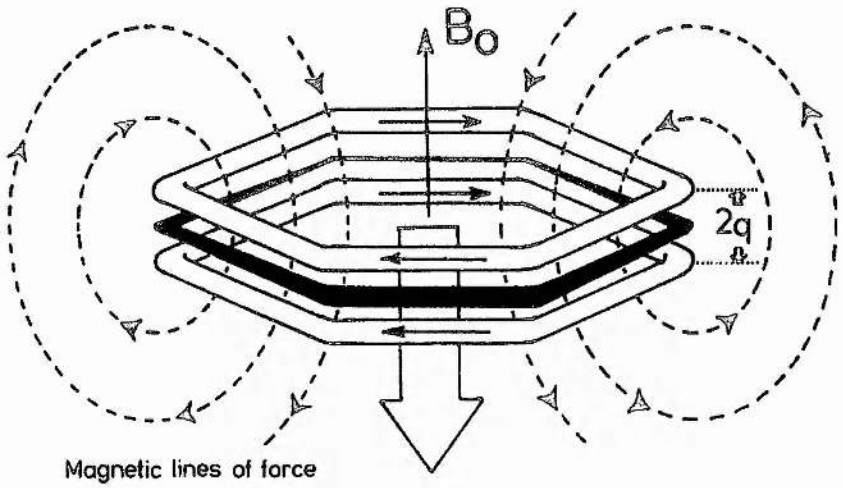


Figure 1.9 The ring current and the magnetic lines of force induced in a benzene ring by an external field B_0. The term q relates to the Johnson-Bovey equation. Taken from Perkins (1982)

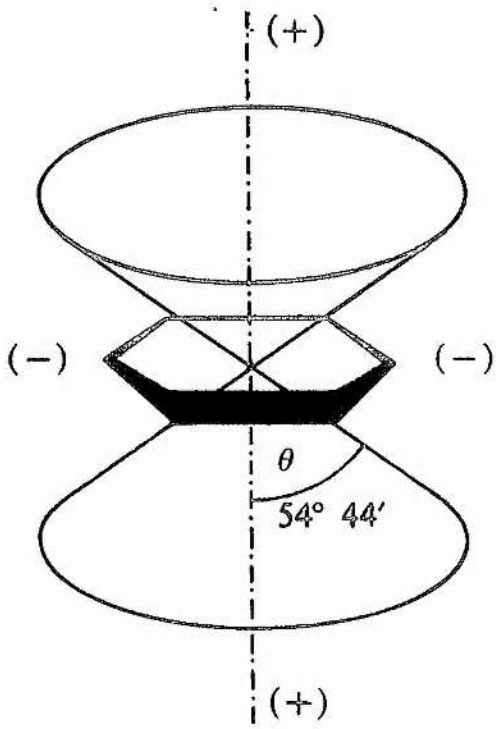


Figure 1.10 Schematic outline of the shielding (+) and deshielding (-) zones due to the ring current of a benzene ring. Taken from Perkins (1982)

aromatic ring, will lie in the plane of the ring and experience large 'internal' downfield shifts. Protons from other residues held close to the aromatic ring either by protein folding or a ligand being bound in a protein. In this case hydrophobic interactions normally place the proton(s) above the centre of the ring where the shielding is greatest. This would result in ring current shifts of up to 0.9 ppm upfield. The calculation of proton position from the aromatic ring (or conversely the position of the aromatic ring from the proton) are based on these 'external' ring current effects.

For a single aromatic ring, a ring current equation can be most usefully discussed in terms of a general expression, where the ring current shift in ppm is the product of these terms.

$$\delta_R \times 10^{-6} = iBG(\rho, z, \phi) \qquad \textit{Equation 1.12}$$

In the case of more than one ring δ_R is the summation of the individual δ_R calculated for each ring. The spatial term $G(\rho, z, \phi)$ relates the cylindrical coordinates ρ, z and ϕ of a proton relative to the ring plane and ring centre. It is of concern to establish the validity of the spatial term as a suitable geometrical description of the way in which ring current shifts vary in space. The general constant of proportionality B should be in principle related to the molecular susceptibility of the aromatic ring. The ring current intensity factor i takes into account the influence of the specific electronic structure of the aromatic ring n is ring current. For the aromatic amino acids, nucleic acid bases and the heme ring i thus takes different values.

Three main approaches have been used to relate the position of the proton relative to the aromatic ring for the calculation of a ring current shift. These are the classical dipole model of Pople (1956), the semi-classical current loop of Johnson and Bovey (1958) and the quantum mechanical method of Haigh and Mallion (1972). In this study only the Johnson-Bovey method was used.

The Johnson-Bovey equation evaluates the local magnetic field produced by two current loops parallel to the ring plane at a distance q above and below the ring (figure 1.12). These loops represent the six delocalised π electrons of the aromatic ring. The Johnson-Bovey equation is given in its 1958 form by:

$$\delta_R \times 10^{-6} = \frac{ne^2}{6\pi mc^2}\left\{\frac{1}{a\left[(1+\rho)^2+z^2\right]}\left[K+\frac{1-\rho^2-z^2}{(1-\rho)^2+z^2}E\right]\right\}$$

Equation 1.13

where n is the number of π electrons; a is the ring radius; e, m and c are the standard constants; ρ and z are cylindrical coordinates of the proton relative to the ring centre (figure 1.11); And K an E are the first And second elliptic integrals which are a function of ρ, z and q in units of a. The spatial term is enclosed within the large brackets of equation 1.12.

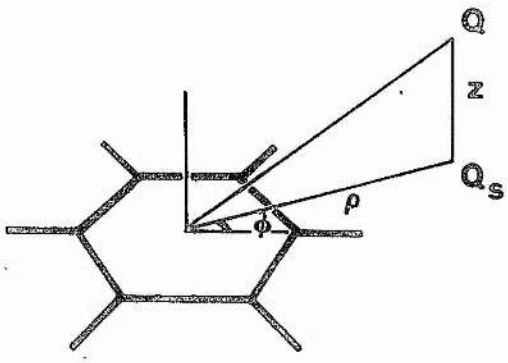


Figure 1.11 Geometrical input for the Johnson-Bovey equation in order to relate the position *Q* of the proton to the aromatic ring.

indicates the the position of *Q* when projected into the lane of the ring.

Chapter 2.

Cloning of Variable Heavy and Light Chain Anti-E3G Antibody Genes.

Abstract

A DNA plasmid which would allow the expression of an anti-E3G Fv fragment in *E. coli* was constructed. The Fv fragment is to be used in a diagnostic kit for fertility prediction. By isolating mRNA of the variable heavy and light chain genes from hybridoma cells, first strand cDNA was synthesised. The variable heavy and light chain genes (V_H and V_L respectively) were then isolated and amplified by PCR. This also introduces restriction enzyme sites that allow the DNA to be easily manipulated into an existing plasmid. The DNA and derived amino acid sequence data for the two genes has been given.

2.1 Introduction

Immunoglobulin Fv fragments, heterodimers containing only the variable domains V_H and V_L, are the smallest antibody fragments that have antigen binding activity. Their small size makes these molecules interesting targets for the development of immunodiagnostics and immunotherapeutic agents as well as for other applications. Having the the genetic material to express the Fv fragment in bacteria gives many opportunity to diversify the use of these fragments: the two chains of the fragment can be linked by a polypeptide chain (Bird *et al.*, 1988) (single chain Fv fragment, scFv) allowing a greater range of expression systems and adding stability to the protein; peptide tails can be added to aid in the purification or detection of the expressed fragment (Munro and Pelham, 1986); the antibody can be mutated to provide different levels of binding which in conjunction with structural studies can be used elucidate information of the nature of antibody-antigen interactions; bacterial expression also allows easy isotopic labelling for NMR studies. The small size of the Fv fragment makes the protein very amenable to structural

studies by either NMR or X-ray crystallography and several structures have been published. (Chitarra *et al.*, 1993; Eigenbrot *et al.*, 1882; Fan *et al.* 1992; Bundle *et al.*, 1994)

This chapter is concerned with obtaining the two variable chain genes from hybridoma cells and inserting these into a suitable plasmid (pUC 19) to allow expression of the active protein The hybridoma cells were from a mouse immunised to give a monoclonal response to E3G, and the fragment is envisaged to be used as a diagnostic agent in a commercially available kit to be used for fertility prediction. The main difficulty comes from the presence of 'rogue' sequence mRNA present in often very high levels in the hybridoma cells, if this is isolated its sequence appears as light chain DNA, but any expressed protein would be inactive. Fortunately, these sequences are well documented, so that any rogue sequence can be identified. (Strohal *et al.*, 1987) .

Expression of the recombinant protein in bacteria will also allow sufficient material to be generated for structural studies: firstly for my own use in an NMR study of the three dimensional structure of the bound antigen and binding site (Chapter 6), and also the plasmid was made available to collaborators for an X-ray crystallography study of the Fv complexed with its antigen and other closely related molecules. Once these genes are in a suitable plasmid, they can easily be moved into other plasmids which would allow the recombinant protein to be expressed with different polypeptide tails, and also as a single-chain protein.

2.2 Materials and Methods

Isolation and purification of mRNA.

The method used is a modification of the method described by (Favaloro *et al.*, 1980). Anti-E3G 4155 hybridoma cells (3×10^8) were washed twice in 500 mL of PBS, the cells were then pooled in 50 mL of PBS. To this 22 mL of cold lysis buffer with vanadyl ribonucleoside complexes was added. This was left on ice for 2 minutes, vortexed then centrifuged at 4000 g for 20 minutes at 4°C. The supernatant was transferred into 30 mg of protein kinase (predigested in 25 mL protein kinase buffer (200 mM Tris.Cl pH 7.4; 300 mM NaCl; 25 mM EDTA;2 % SDS; 1 $mg.mL^{-1}$ protein kinase K) for 15 minutes at 37°C), the kinase was inactivated by heating to 65°C for 3 minutes. To this 3.5 mL of 4M NaCl was added, then cooled to room temperature and added to the oligo dT cellulose solution. This was prepared as follows: 100 mg of oligo dT cellulose was washed 3 times with 10 mL of 0.1M NaOH, twice with DEPC treated H_2O, twice in elution buffer and once in binding buffer. For each wash the oligo dT cellulose was spun at 4000 g for 5 minutes and the supernatant decanted. The cells/oligo dT cellulose mixture was incubated for 30 minutes at room temperature on a rototorque. A column for purification was prepared using a 2 mL sterile syringe with sterile siliconised glass wool as a plug. The cells/oligo dT cellulose mixture was loaded on the column and washed 3 times with 1.5 mL of binding buffer and 3 times with 1.5 mL of washing buffer. The mRNA was eluted with three 1.5 mL washes with elution buffer (50 °C). The mRNA was precipitated with 500 µL of 3 M NaOAc and 12.5 mL of absolute ethanol, for 16 hours (overnight) at -20°C. The mRNA was centrifuged, the supernatant removed then resuspended in 500µL absolute ethanol, centrifuged and the supernatant removed as

before. The mRNA was resuspended in 500μL of absolute ethanol and stored at -70°C. The mRNA was quantified by measuring the absorbance at 260 nm, where one absorbance unit at 260 nm is equivalent to 40 $\mu g.mL^{-1}$. The ratio of absorbance at 260 nm over the absorbance at 280 nm was measured to give an indication of protein contamination.

Buffers used:

PBS: 8g NaCl, 0.2g KCl, 1.44g Na_2HPO_4, 0.24g KH_2PO_4 in H_2O to 1L, pH 7.4.

lysis buffer: 0.1 M NaCl, 15 mM $MgCl_2$, 10 mM Tris.Cl (pH 8.6) , 0.5 % NP-40.

binding buffer: 0.5 M NaCl, 10 mM Tris.Cl pH 8.0, 5 mM EDTA, 0.2% SDS.

washing buffer: 0.1 M NaCl, 10 mM Tris.Cl pH 8.0, 5 mM EDTA, 0.2 % SDS.

elution buffer: 10 mM Tris.Cl pH 8.0, 5 mM EDTA, 0.2 % SDS.

First Strand cDNA Synthesis.

Method modified from Sambrook *et al.* (1989): 2 μL of mRNA (~2 μg) and 3.2 μL of MG08 (heavy chain); or MK09 or PCR174 (light chain) primer (2.5 ng) were heated at 72 °C for 2 minutes, cooled on ice then 1μL 10x RTB (reverse transcriptase buffer, 50 mM Tris.Cl pH 7.6, 60 mM KCl, 1 mM dNTPs, 1 mM DTT, 1 $unit.\mu L^{-1}$ RNase inhibitor, 50μg.mL actinomycin D), 1 mL BSA (bovine serum albumin 100 $\mu g.mL^{-1}$, RNase free) 1μL of dNTP (5 mM), 0.4 μL of RNAguard (Pharmacia Biotech.), and 0.4 μL of reverse transcriptase (Anglian Biotech.) were added and the reaction incubated at 42 °C for 70 minutes. This reaction was stored at -70°C and can be used directly for the PCR.

Isolation and Amplification of V_H and V_L genes by the PCR.

Both reactions were variations on the method described by Ward and co-workers. (1989).

V_H. The following were added to a sterile 1.5 mL eppendorf tube: 3μL (10 pmol.mL^{-1}) of both 5′ primer (PCR 51) and 3′ primer (PCR 89), 5 μL 10x RTB, 5 μL dNTPs (2 mM), 32.5 μL sterile milli-Q H_2O, 0.5 μL *Taq* polymerase. 1μL of target DNA (10^3 dilution of first strand cDNA synthesis). The reaction mixture was covered with 50 μL of mineral oil, and the reaction tubes were placed in a thermocycler with the following program: 20 cycles of 93 °C for 1 minute, 50 °C for 1.5 minutes and 72 °C for 2 minutes, followed by another 2 minutes at 72 °C. 1 μL of these reactions were added to 1 μL of 10x running buffer, and 8μL of H_2O.

V_L: 10μL of *Vent* buffer (10x) (New England Biolabs), 2 μL dNTPs (20 mM), 2 μL 5′ primer, 2 μL 3′ primer (both 30 pmol.mL^{-1}), 83 μL sterile milli-Q H_2O.and 1 μL *Vent* polymerase (New England Biolabs). To this 1 μL of target DNA (10^3 dilution of cDNA reaction), and overlayed with 100 μL of mineral oil. The samples were placed in the thermocycler and were run with the following program: 94°C for 4 minutes, then 20 cycles of 94 °C for 1 minute, 55 °C for 1 minute. and 72 °C for 1 minute with an additional 2 minutes at 72 °C after the 20 cycles were completed. The results were viewed by running 1 μL of the sample on a 1% agarose gel.

Purification of PCR Products.

The DNA fragments were purified from a 1.5% agarose gel using the QIAEX DNA gel extraction kit. The DNA was excised from the gel with a clean scalpel blade and then the excess gel removed. The purification was

carried out using the protocol and chemicals supplied.

Sequencing of PCR Products.

V_H DNA in a modified pMM sequencing vector (M. Metcalfe, unpublished) was sequenced using the Sequenase 2.0 DNA sequencing kit. 7μL of dsDNA (1-2 μg), 2 μL primer (1-2 pmol) and 1 μL NaOH (0.8M) were incubated at 68°C for 10 minutes. 2 μL of 1.5 x TDMN (320 mM TES, 180 mM HCl, 75 mM DTT, 120 mM $MgCl_2$, 300 mM NaCl) was incubated at room temperature for 5 minutes. To this 1 μL of the labelling mix (2:5 with milli-Q H_2O), 0.5 μL ^{35}S-αATP, 2μL of Sequenase (1:8 ratio with milli-Q H_20), then mixed and left at room temperature for 5 minutes. 3.5 μL aliquots of this was added to each of the ddNTP solutions and incubated at 37°C for 5 minute. The sequencing reactions were stopped by adding 4 μL of the 'stop' reagent. The samples were incubated for 20 minutes at 80°C, before being loaded on a sequencing gel (4% polyacrylamide gel)

V_L DNA: Sequencing was done on an automated sequencer by Dean Sibthorpe

Preparation of DNA Plasmids.

A glycerol slant of pMM transfected in *E.coli.* JM109, was streaked across a 2TY amp plate and incubated at 37 °C for 16 hours (overnight). Colonies were picked and grown in 5 mL of 2TY amp. for 6 hours. DNA was prepared using the alkaline lysis method.

Restriction Enzyme Digests.

Pst I: 10 μL of DNA was digested by 10U of *Pst* I in Boehringer

Mannheim 'High-Salt' buffer: 50 mM Tris.Cl (pH 8.0), 10 mM $MgCl_2$, 100 mM NaCl, 1 mM 2-mercaptoethanol, at 37 °C for 2 hours

Xho I: 10 µL of DNA was digested by 10U of *Xho* I in Boehringer Mannheim 'High-Salt' buffer: 50 mM Tris.Cl (pH 8.0), 10 mM $MgCl_2$, 100 mM NaCl, 1 mM 2-mercaptoethanol. at 37 °C for 2 hours

Pst I-*Bst* EII: 10 µL of DNA was digested by 10U of *Pst* I in Boehringer Mannheim buffer 'B': 10 mM Tris.Cl (pH 8.0), 5 mM $MgCl_2$, 100 mM NaCl, 1 mM 2-mercaptoethanol, at 37 °C for 2 hours. 10 U (1µL) of *Bst* EII was added and this incubated at 60 °C for 2 hours.

Sac I-*Xho* I: 10µL of DNA was digested by 10U of *Sac* I in Boehringer Mannheim 'Low Salt' buffer: 10 mM Tris.Cl (pH 7.5), 10 mM $MgCl_2$, 1 mM dithiothreitol (DTT) for a final volume of 20µL, was incubated at 37 °C for 2 hours. To this 10U of *Xho* I, 2.5µL of NaCl(1M), and 1.5 µL of milli-Q H_2O (NaCl now 100 mM) and this was incubated at 37 °C for 2 hours.

Sac I-*Eco* RI This was done exactly as the *Sac* I - *Xho* I digest.
(All restriction enzymes were supplied by Boehringer Mannheim).

Ligation of V_H/V_L gene into vector. 5 µL of insert DNA, 1 µL of vector, 1 µL of 10x ligation buffer, 1µL DTT (0.1M), 1 µL rATP (10 mM), 1µL T4 DNA ligase. A negative reaction was also set up with no insert. These samples were incubated for 2 hours at room temperature.

Transformation of Ligated DNA into electrocompetent cells.
(A) Preparation of electrocompetent E. coli *JM109 cells.* 10 mL of 2TY media was inoculated with 10 µL of an *E. coli* JM109 culture (glycerol slant). This was incubated at 37 °C with shaking for 16 hours (overnight). 1 litre of 2TY media was inoculated with 5 mL of the overnight culture and incubated at 37 °C with shaking (250 rpm) until the absorbance at 600

nm was between 0.5 and 0.7. The cells were placed on ice for 30 minutes then centrifuged at 4 °C and 4000 g for 20 minutes. The supernatant was discarded and the cells resuspended in 1 litre of ice-cold HEPES (sterile, 1 mM) This suspension was centrifuged at 4000 g for 20 minutes at 4 °C. The supernatant was again discarded and the cells were resuspended in 500 mL of the HEPES, washed, then centrifuged in the same way. The washing and centrifuging steps were repeated this time with 20 mL of 1 mM HEPES, 10% glycerol. Finally the cells were resuspended in 2.5 mL of 1 mM HEPES, 10% glycerol. 50 μL aliquots were put into sterile eppendorf tubes. The cells were then used immediately, or frozen on dry-ice then stored at -70 °C for future use.

(B) Transformation. 1. Sample preparation: the ligation sample was extracted of protein impurities with a phenol/chloroform and chloroform extraction (as in DNA minipreps), and the DNA precipitated and washed, then resuspended in 20 μL of milli-Q H_2O.

2. Electroporation: The electrocompetent cells were thawed on ice then mixed with the DNA in a prechilled 0.2 cm electroporation cuvette. The cuvette was placed on ice for 1 minute then pulsed once at 200 Ω, 2.5 kV and 25 μF. Immediately 1 mL of 2TY was added and this transferred to a 6 mL falcon tube, the cuvette was rinsed with a further 1 mL of 2TY and this added to the falcon tube also, the 2 mL samples were incubated with shaking at 37 °C for 1 hour.

3. Plating out: For each transformed sample samples of 2 μL, 20 μL, and 200 μL samples were streaked onto 2TY amp plates. The remaining cells were spun down and resuspended in 200 μL of 2TY and plated out. The plates were incubated for 16 hours at 37 °C.

Preparation of DNA from Bacterial Colonies or Glycerol Slants: (alkaline lysis method or minpreps).

This method is taken from Sambrook *et al.* (1989), which is a modification of the method used by Birnboim and Doly, (1979); and Ish-Horowicz (1981)and Burke, (1981). 2 mL of 2TY amp media was inoculated with either a bacterial colony or ~10μL of a glycerol slant. and grown for 5 hours with shaking at 37 °C. 1.5 mL of the culture was transferred to a 1.5 mL eppendorf tube and centrifuged for 1 minute, the supernatant was discarded and the pellet shook dry (All centrifugation was done in a micro-centrifuge at 12,500 rpm) The pellet was resuspended in 100 μL of ice-cold solution I (50 mM glucose, 25 mM Tris.Cl pH 8, 10 mM EDTA), vortexed then left for 5 minutes on ice. 200 μL of solution II (0.2 M NaOH, 10% SDS - room temperature) was added, mixed by rapid inversion and this left at room temperature for 5 minutes. 150 μL of ice-cold solution III (3 M KaOAc, 2 M HOAc) was added, the mixture vortexed briefly, put on ice for 5 minutes and then centrifuged for 5 minutes. The supernatant was transferred to a fresh tube containing 450 μL of phenol/chloroform, the combined sample vortexed then centrifuged for 3 minutes. The aqueous layer was transferred to a fresh tube containing 450 μL of chloroform. The aqueous layer was extracted as before and transferred to a fresh tube containing 800 μL absolute ethanol; the DNA was precipitated at -20 °C for several hours or on dry-ice for 10 minutes. The DNA was centrifuged for 5 minutes, the supernatant discarded, the pellet washed with 70 % ethanol, centrifuged for 5 minutes, and the supernatant discarded. The pellet was allowed to dry for 15 minutes. The DNA was dissolved in 50 μL of TE buffer (1M Tric.Cl pH8, 0.5 M EDTA) containing 1 $\mu g.mL^{-1}$ of RNase A (DNase free) and the samples incubated at 37 °C for

30 minutes before being frozen away.

Precipitation of DNA.

1. Ethanol precipitation. To a DNA solution the following was added: 2.5 volumes absolute ethanol, 0.1 volumes of 3M NaOAc and 0.05 volumes of glycogen. This was precipitated at -20 °C for 16 hours, or on dry-ice for 10 minutes.

2. PEG Precipitation. 30 μL miniprep DNA and 50 μL 13% PEG, 1.5M NaCl, were added into a 1.5 mL eppendorf tube, and placed on iced water for at least 30 minutes, then centrifuged in a microcentrifuge at full speed for 20 minutes at 4 °C. The pellet was washed with 70% ethanol 3 times, the DNA pellet was air dried for 15 minutes and then resuspended in 15μL sterile milli-Q H_2O.

2.3 Results and Discussion.

mRNA Purification.

The absorbance measurements showed a total yield of 62 μg of mRNA, and the A_{260} /A_{280} ratio was 2.31. A ratio greater than 2 is required, which means that this mRNA is sufficiently free of protein contamination to suitable for use in the reverse transcriptase reaction.

First strand cDNA synthesis and PCR reactions.

Although ^{32}P labelled cDNA reactions could be run in parallel with 'cold' reactions as a check on the successful completion of the reaction, this procedure was found to be time consuming and largely unnecessary, as the first strand cDNA synthesis was reliable and in the time taken to run the labelled cDNA on gels and wait for the results the PCR reactions could

be carried out and analysed: figure 2.1 shows bands on the gel for both V_H reactions and the positive control reaction (anti-lysozyme V_H first strand cDNA), which indicate DNA of the correct size (340 bp). There is no band in the lane of the gel containing the reaction without V_H cDNA target, this control should ensure that the DNA amplified was the target cDNA. The V_H genes were successfully isolated and amplified, once the gene had been isolated it was subjected to a *Pst* I and *Bst* EII restriction enzyme digest, purified and ligated into a pMM cloning vector, previously treated to similar digestion and purification. The two restriction enzymes sites are introduced by the two PCR primers: PCR59 and PCR81, the sequences of these primers and the other PCR and cDNA primers used in this chapter are shown in table 2.1. The ligation reaction was transformed into *E. coli* JM109, several colonies were picked and grown in 2TY amp. Samples of DNA purified from these cultures were subsequently digested with *Pst* I and *Bst* EII and these digests run on an agarose gel (figure 2.2). Only those ligation reactions that were successful would have DNA with a V_H insert. This shows that for most of the cultures where DNA was isolated, there was a V_H insert. Lane number 2 showed the cleanest digest, and the DNA used for this digest was used for the sequencing reactions. Some of the culture remains were streaked onto a plate and several colonies picked and grown, to clone purify. The procedure of DNA isolation and restriction digest was repeated as just described, and the culture from which the cleanest DNA and digest derived was used to make a glycerol slant.

The gene was sequenced in this vector, and enough DNA sequence was elucidated to show it was a novel V_H gene. One noteworthy point was that the gene contained an internal *Xho* I site. No protein sequence data

was available from the monoclonal antibody, but since no similar 'rogue' sequence problem exists for variable heavy chain genes, the only problem in isolating the wrong gene would be contamination in the PCR reaction of some other heavy chain gene. Since the gene isolated is novel, and from the experience in the Colworth laboratory where no problems have been encountered with V_H gene isolation, it can be assumed with some confidence that the correct gene has been isolated.

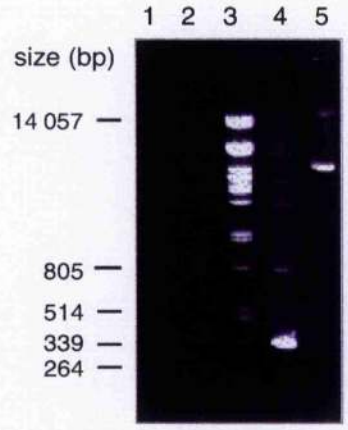


Figure 2.1 Heavy chain PCR reactions on an agarose gel.

Lanes:
1. V_H reaction
2. V_H reaction
3. λ phage DNA, *Pst* I digest
4. positive control (anti lys. heavy chain cDNA)
5. negative control (no cDNA)

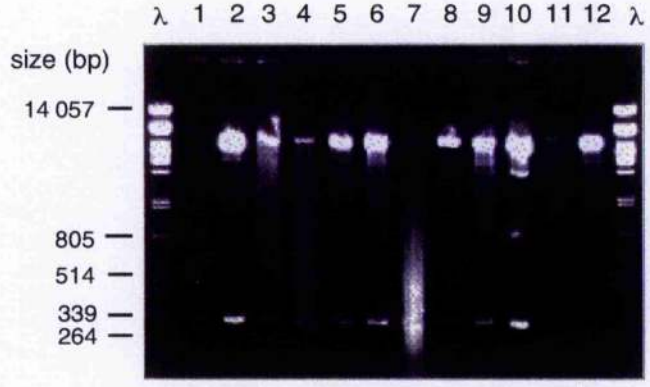


Figure 2.2 *Pst* I / *Bst* EII restriction enzyme digest of pMM vector with V_H DNA insert.

First strand cDNA synthesis primers.

Heavy chain:

MG08	GAT AGA CAG ATG GGG GTG TCG TTT

Light chain:

MK09	AGA TGG ATA CAG TTG GTG CAG CAT
PCR174	TCA CTG GAT GGT GGG AAG ATG GAT ACA

PCR primers:

Heavy chain:

5'-PCR59	AGG T(C/G)(A/C) A(A/G)C AG(C/G) AGT A(A/T)G G
3'-PCR81	TGA GGA GAC GGT GAC CGT GGT CCC TTG GCC CC

Light chain:

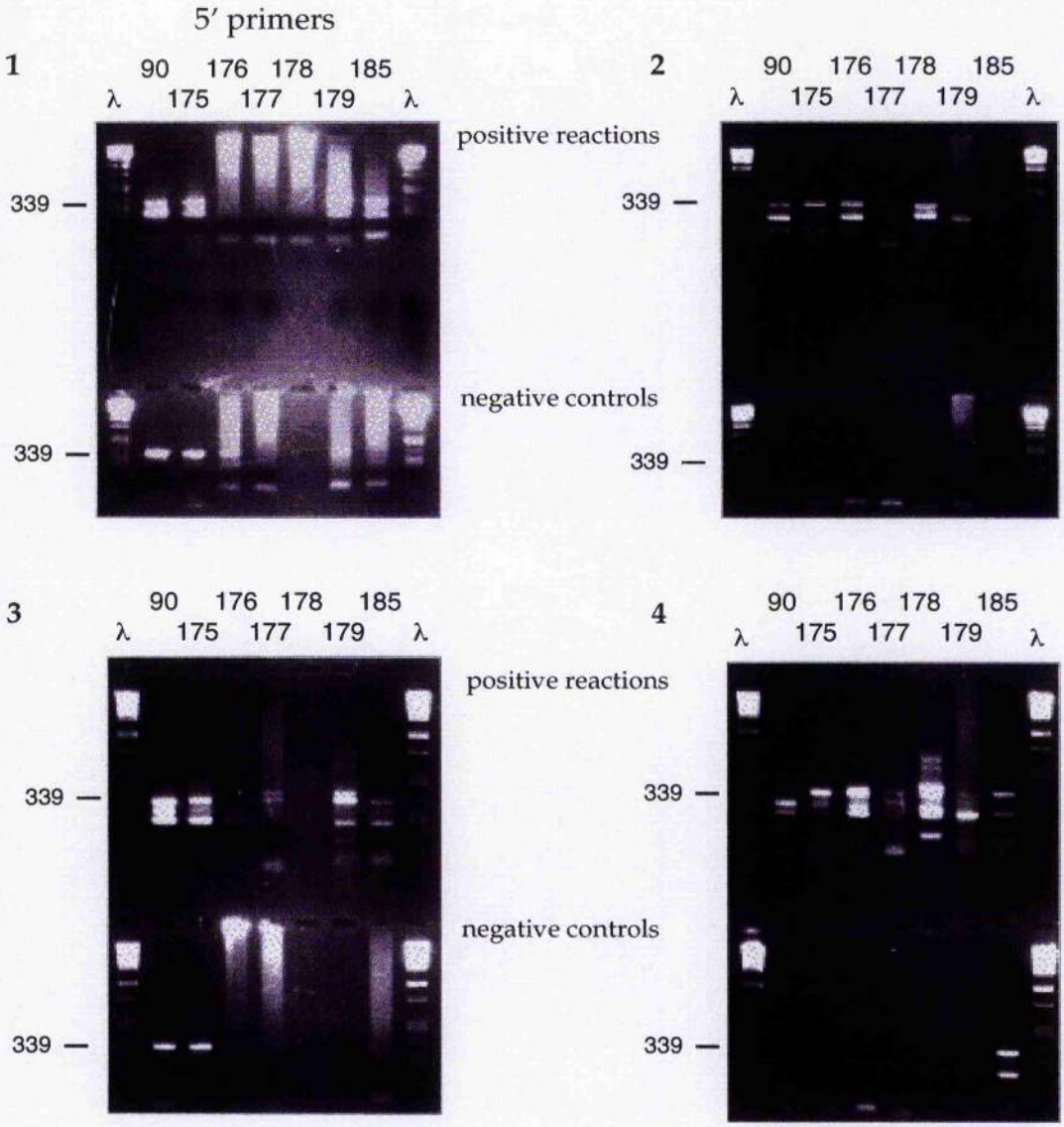
5'-PCR90	GAC ATT GAG CTC ACC CAG TCT CCA *
PCR175	GAA ATT GT(G/T) CTC AC(C/A) CA(G/A) TCT CC
PCR176	GAC ATC CAG ATG AC(C/A) CAG (T/A)CT (C/A)C
PCR177	GAT ATT GTG ATG AC(C/A) CAG G(C/A)T
PCR178	GAT GTT GTG ATG ACC CAA ACT CC
PCR179	A(A/G)(T/C) ATT GTG ATG ACC CAG (A/T)CT C
PCR185	GAC GTG GTG ATG ACC CAG (T/A C/G T/C) CC
3'-PCR116	GTT TGA TCT CGA GCT TGG TCC C *
PCR154	CCG TTT TAT TTC CAA CTT TGT CCC *

Table 2.1 Primers used for first strand cDNA synthesis and PCR reactions.

The PCR isolation and amplification of the variable light chain gene proved to be a much more difficult problem. The original primer set (PCR90 and PCR116, sequences in table 2.1), proved inadequate as no products were obtained unless very low annealing temperatures were used; which resulted in low stringency and several unwanted gene products. A gene product of the correct size (325 bp) was obtained but the band appeared very faint on the gel (data not shown). After ligating the correct size gene product into a sequencing vector, subsequent sequence analysis showed to this to be the variable light chain gene from the anti-lysozyme Fv, as this fragment had previously been cloned in this laboratory, a contamination of this DNA was in the sample and been subsequently isolated by PCR. When this process was repeated and the subsequent gene resulting from the primers PCR90 and PCR154, was subjected to DNA sequence analysis, the gene proved to be a rogue V_L sequence. From the problems in obtaining any gene products, and the subsequent sequence results it became obvious that the gene could not be cloned from the V_L primers previously used in the lab. To clone the V_L gene, a non-PCR method may have been used, or if the PCR route is to be developed a new set of primers for the isolation of the light chain DNA would be needed, it would also be helpful to be able to process the correct size DNA fragments sequence more quickly. Because the primer for the first strand cDNA synthesis may preferentially isolate the rogue light chain mRNA, it would be wise to try different first strand cDNA primers. At first a non-PCR approach was attempted which was unsuccessful. The availability of an automated DNA sequencer, which meant that a number of PCR products could be analysed directly without the need for cloning into a vector, together with a new number of 5′ V_L primers (Leung *et al.*, 1993) designed to negate the problem of isolating rogue sequence DNA,

gave hope of isolating the V_L gene by PCR. A primer was also designed from the light chain amino acid sequence (PCR185), the sequence of this and the other primers used are shown in table 2.1. A large scale PCR reaction was run using four different V_L cDNA reactions; using two different cDNA primers; and positive and negative reactions (without mRNA target) for each primer. For each of these a separate PCR reaction was run with every combination of the seven 5′ primers with each of the two 3′ primers. Results of the PCR reactions can be seen in figure 2.3. This shows several reactions which have bands of the correct size, where there are no corresponding bands in the negative control reactions. Primer combinations which give bands in the negative control reactions that correspond to the bands in the positive reactions are likely to have amplified a contamination.

The remaining volume of the promising reactions were run on an agarose gel and bands of the correct size purified direct from the gel. These were then sequenced directly using an automatic sequencer - not having to place all the genes into a suitable vector and test for positive ligations allows the sequence results of this PCR reaction to be processed much more quickly. These sequence results showed a number of the genes were found to be rogue sequence. One gene using the primer PCR174 in the first strand cDNA synthesis and primers PCR116 (3′) and PCR179 (5′) in the V_L PCR proved to match the small amount of V_L protein sequence data available (20 amino acids) and therefore almost certainly the correct gene product. The sequence also revealed an internal *Pst* I site that would be relevant in creating the final construct. The primers used in this experiment did not have any restriction enzyme sites, the cloned light chain gene was reacted in a further PCR reaction using the primers PCR90



Primers used:

1	cDNA:	MK09	2	cDNA	MK09
	3′:	PCR116		3′:	PCR154
	5:′	all		5′:	all
3	cDNA:	PCR174	4	cDNA	PCR174
	3′:	PCR116		3′:	PCR154
	5′:	all		5′:	all

Figure 2.3 Agarose gels of V_L PCR reactions.

(5′) and PCR116 (3′) to add the *Sac* I and *Xho* I restriction enzyme sites required for subsequent ligation.

Construction of final Fv E3G myc pUC 19 plasmid.

A modified pUC 19 vector was used as an expression plasmid for the variable genes of the anti-E3G antibody. This vector has been modified (Better *et al.*, 1988; Ward *et al.*, 1989; Orlandi *et al.*, 1989) to have the pel B leader sequence up-stream from each variable gene: these direct the expressed protein to the periplasm where the two chains associate. A map of the construct containing the two genes can be seen in figure 2.4.

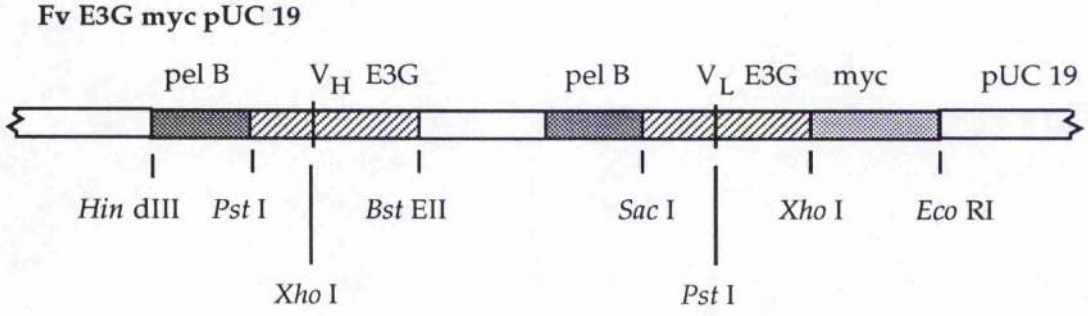


Figure 2.4 Representation of the final Fv E3G myc pUC 19 vector.

Because of internal restriction enzyme sites found in both V_L (*Xho* I) and V_H (*Pst* I) genes, it would not be possible to ligate both genes into the one vector, as for example, after ligating the V_H into the vector the subsequent *Sac* I/*Xho* I restriction digest to remove anti-lysozyme V_L DNA, would cut the newly added anti-E3G V_H DNA at the *Xho* I site. Therefore the following strategy was employed: Staring from the plasmid for expression of the anti lysozyme Fv: Fv lys myc pUC 19, and the pMM vector

containing the V_H gene, a *Pst* I and *Bst* EII restriction digest was carried out on both plasmids. Both reactions were run on agarose gels and the V_H E3G gene and the Fv lys. myc pUC 19 vector without the anti-lysozyme V_H gene were purified from the gel. The V_H gene was ligated into this vector, transformed in *E. coli* JM109, as previously described. From the transformation plates, 24 colonies were picked, grown in 2TY amp / glucose, the DNA purified and a *Xho* I restriction digest performed on each DNA sample and the reactions run on an agarose gel. Because of the internal *Xho* I site in the V_H E3G gene, the gel will show a band of about 560 bp where there was a successful ligation, as there are now two *Xho* I sites (figure 2.5). Sample number 14 was the cleanest, and this DNA was used in subsequent reactions.

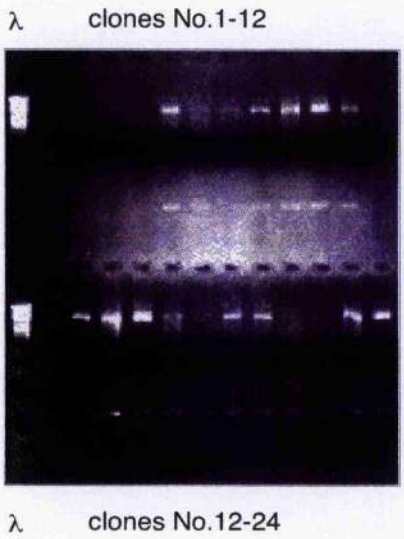


Figure 2.5 *Pst* I/*Bst* EII restriction digest of V_H E3G V_L lys myc pUC 19 showing clones with a positive ligation of heavy chain DNA.

The V_L PCR product with the *Sac* I and *Xho* I restriction enzyme sites and some more of the Fv lys myc pUC 19 vector were digested with these enzymes. The anti-E3G V_L DNA and the Fv lys myc vector, without the anti-lysozyme V_L were purified from the gel and ligated. The ligated vector was transformed in *E. coli* JM109 and colonies were picked and grown as for the V_H gene ligation. Because of the internal *Pst* I site a digest with this enzyme would show successful ligations, as only vectors with the anti-E3G V_L would give a band of around 600 bp (figure 2.6). This shows that most of the vectors contain anti-E3G V_L DNA insert. Number 5 was chosen as the cleanest digest and used in the making the final construct. To avoid the problem of the the *Xho* I and the *Pst* I internal restriction sites the final construct was prepared by removing both V_L myc genes by a *Sac* I and *Eco* RI digest, and the appropriate parts purified from the gel: V_H E3G V_L lys myc pUC 19 (-V_L lys myc) and V_L E3G myc. These were ligated and the ligated vector transformed. Again a *Pst* I restriction enzyme digest would show a successful ligation. (figure 2.7).

As each of the three different pUC 19 vectors were made, the colony which gave the best results was regrown and clone purified then made into glycerol slants. This allows more of any of these plasmids to be grown and purified when necessary. The final construct was purified in large amounts and a large part of this was further purified using PEG precipitation. This purified DNA was used in several sequence reactions using an automated DNA sequencer - again carried out by Dean Sibthorpe. The sequence for the two cloned variable genes are shown in table 2.2 and 2.3.

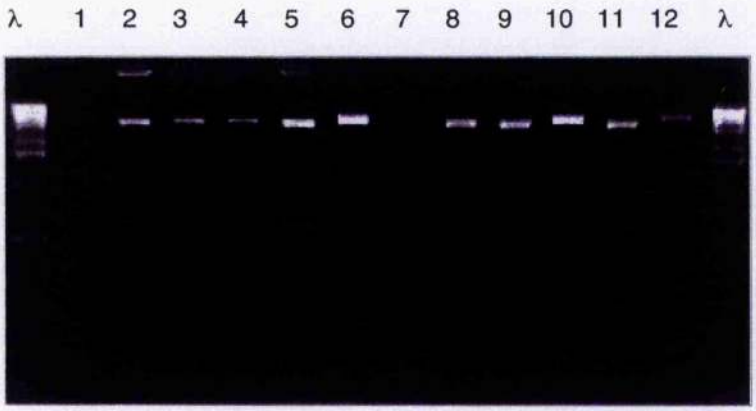


Figure 2.6 *Sac*I/*Xho*I restriction digest of V_H lys V_L E3G myc pUC 19 showing clones with a positive ligation of light chain DNA.

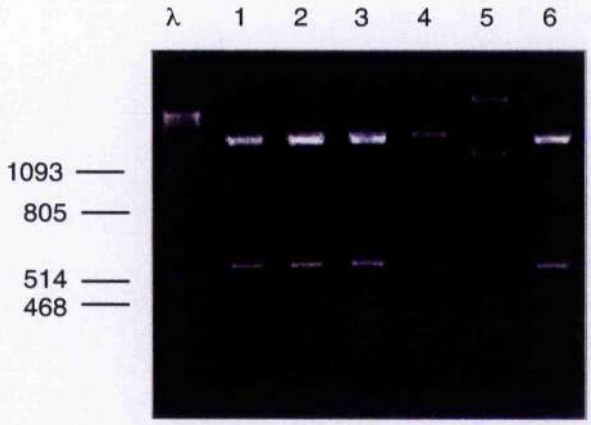


Figure 2.7 *Sac*I / *Xho*I restriction digest of Fv E3G myc pUC 19 showing clones with a positive ligation of light chain DNA.

										base no.	a.a. no.
GAC	CTC	GAG	CTC	ACC	CAA	ACT	CCA	CCC	TCC	- 30	
D	I	E	L	T	Q	T	P	P	S		- 10
CTG	CCT	GTC	AGT	CTT	GGA	GAT	CAG	GTT	TCC	- 60	
L	P	V	S	L	G	D	Q	V	S		- 20
ATC	TCT	TGC	AGA	TCT	AGT	CAG	AGC	CTT	GTG	- 90	
I	S	C	R	S	S	Q	S	L	V		- 30
TCC	AAT	AAT	AGA	AGG	AAC	TAT	TTA	CAT	TGG	-120	
S	N	N	R	R	N	Y	L	H	W		- 40
TAC	CTG	CAG	AAG	CCA	GGC	CAG	TCT	CCA	AAG	-150	
Y	L	Q	K	P	G	Q	S	P	K		- 50
CTC	GTG	ATC	TAC	AAA	GTT	TCC	AAC	CGA	TTT	-180	
L	V	I	Y	K	V	S	N	R	F		- 60
TCT	GGG	GTC	CCA	GAC	AGG	TTC	AGT	GGC	AGT	-210	
S	G	V	P	D	R	F	S	G	S		- 70
GGA	TCA	GGG	ACA	GAT	TTC	ACA	CTC	AAG	ATC	-240	
G	S	G	T	D	F	T	L	K	I		- 80
AGC	AGA	GTG	GCG	GCT	GAG	GAT	CTG	GGA	CTT	-270	
S	R	V	A	A	E	D	L	G	L		- 90
TAT	TTC	TGC	TCT	CAA	AGT	TCA	CAT	GTT	CCG	-300	
Y	F	C	S	Q	S	S	H	V	P		-100
CTC	ACG	TTC	GGT	TCT	GGG	ACC	AAG	CTC	GAG	-330	
L	T	F	G	S	G	T	K	L	E		-110
ATC	*GAA*	*CAA*	*AAA*	*CTC*	*ATC*	*TCA*	*GAA*	*GAG*	*GAT*	*-360*	
I	*E*	*Q*	*K*	*L*	*I*	*S*	*E*	*E*	*D*		*-120*
CTG	*AAT*	*TAA*									
L	*N*	•									

Table 2.2 DNA and derived protein sequence of anti-E3G antibody light chain variable gene with myc tail. The myc tail DNA and amino acid sequences are shown in italics.

										base no.	a.a. no.
CAG	GTG	CAG	CTG	CAG	GAG	TCT	GGG	GGT	GGC		- 30
Q	V	Q	L	Q	E	S	G	G	G		- 10
TTG	GTG	AAC	CTT	GGA	GGG	TCT	ATG	ACT	CTC	- 60	
L	V	N	L	G	G	S	M	T	L		- 20
TCC	TGT	GTA	GCC	TCT	GGA	TTC	ACT	TTC	AAT	- 90	
S	C	V	A	S	G	F	T	F	N		- 30
ACC	TAT	TAC	ATG	TCT	TGG	GTT	CGC	CAG	ACT	-120	
T	Y	Y	M	S	W	V	R	Q	T		- 40
CCA	GAG	AAG	ACG	CTG	GAG	TTG	GTC	GCA	GCC	-150	
P	E	K	T	L	E	L	V	A	A		- 50
ATT	AAT	AGT	GAT	GGT	GAA	CCT	ATC	TAT	TAT	-180	
I	N	S	D	G	E	P	I	Y	Y		- 60
CCA	AGA	CAC	TTG	AAG	GGC	CGA	GTC	ACC	ATC	-210	
P	D	T	L	K	G	R	V	T	I		- 70
TCT	CGA	GAC	AAT	GCC	AAG	AAG	ACC	CTA	TAC	-240	
S	R	D	N	A	K	K	T	L	Y		- 80
CTG	CAA	ATG	AGC	AGT	CTG	AAC	TTT	GAG	GAC	-270	
L	Q	M	S	S	L	N	F	E	D		- 90
ACA	GCC	TTA	TAT	TAC	TGT	GCA	AGA	CTT	AAT	-300	
T	A	L	Y	Y	C	A	R	L	N		-100
TAC	GCC	GTG	TAT	GGT	ATG	GAC	TAT	TGG	GGC	-330	
Y	A	V	Y	G	M	D	Y	W	G		-110
CAA	GGG	ACC	ACG	GTC	ACC	GTC	TCC	TCA	TAA		
Q	G	T	T	V	T	V	S	S	•		

Table 2.3 DNA and derived protein sequence of anti-E3G antibody heavy chain variable gene.

2.4 Conclusions.

Although the heavy chain variable gene was cloned without difficulty, the presence of rogue light chain variable genes in hybridoma cells, and an initial poor selection of 5′ primers made cloning of the anti E3G V_L gene difficult. With the likelihood of pulling out rogue sequences, the use of an automated sequencer along with the new primers was a much more

efficient method of cloning light chain variable genes: firstly by multiplying the probability of obtaining the right gene; and secondly, by allowing rapid screening of these products by negating the need to place all these genes into vectors. The use of a large variety of primers also gives every chance possible of by-passing the 'rogue' sequence problem.

Internal restriction endonuclease sites in both the V_H and V_L genes required a slightly convoluted process of constructing the final plasmid, but was extremely useful in providing a rapid method of differentiating DNA which had the correct insert present. This was especially important where there was a significant count on transformation plates for the negative control ligation. This meant once DNA from each culture was checked by the restriction enzyme digest, that this DNA and the culture itself could be used in subsequent reactions with confidence.

Once the genes are isolated there is a large scope of use for these genes to be placed in other plasmids, for example: several plasmids exist with different polypeptide tails, which can be of use in ELISA detection and also in purification. The variable genes could also be placed into a vector which contains the DNA which codes for a polypeptide linker between the C-terminus of the V_H and the N-terminus of the V_L. This would create a scFv which can provide a more stable fragment that is possible to use in different expression systems.

Chapter 3.

Detection, Specificity, Competition and Inhibition of Fv E3G myc by Enzyme Linked Immunosorbent Assay (ELISA)

Abstract

The expressed Fv fragment was subjected to a number of immunological tests using ELISAs and blots to determine several important properties of the recombinant protein. These assays showed the Fv E3G myc antibody fragment was firstly being expressed, the fragment was active in binding E3G and was specific to this antigen. It was also demonstrated that the Fv fragment competed against the parent antibody and in sufficient concentrations could inhibit binding of the parent antibody to an E3G-alkaline phosphotase conjugate (E3G-AP). An assay for detection of the Fv fragment was developed that does not rely on binding of the myc tail.

3.1 Introduction

The completed plasmid from Chapter 2 is to be used to express the anti-E3G Fv fragment and this protein used in immunological tests to determine several characteristics of the protein. By use of ELISA tests which rely on detection of activity by a series of recognition reactions by antibodies, we can detect properties such as activity and specificity of the Fv fragment, also competition with and inhibition against the parent mAb. Because of the sensitivity and selectivity of antibody binding these tests are very reliable as an indicator of the immunological properties of the expressed antibody fragment. It can often be the case that once the corresponding Fv fragment of a mAb can exhibit very weak binding in comparison to the parent mAb, therefore it is vital that the antibody fragment produced does bind the antigen in the first instance, and secondly that if good binding is detected that this effect is specific to the antigen. Once these properties have been demonstrated, information can be gained through developing assays that compare the properties of the antibody fragment in relation to its parent mAb. Once it has been

demonstrated that the fragment and monoclonal behave in a similar fashion there can be confidence that the fragment can be used and deliver the same results in its intended application in place of the more expensive mAb, produced from tissue culture. The ELISA used in the detection of initial activity depended on binding to a polypeptide tail. This may not be satisfactory for certain applications, therefore work in this chapter was also directed into developing an assay for activity that does not rely on the presence of this polypeptide tail. With the successful development of these assays and the required properties demonstrated, the antibody fragment could then be confidently used in the required applications, and with strong enough binding a meaningful NMR study of antibody-antigen interactions can be implemented.

3.2 Materials and Methods

Expression and Detection of Fv E3G myc.

Colonies from the final transformation plate were picked and grown in 3 mL of 2TY amp/glucose and grown until log phase, each sample was split into two parts: 2.5 mL was centrifuged at 5000 g for 5 minutes and the cells were resuspended in 5 mL of 2TY amp, 1 mM isopropylthiogalactoside (IPTG), and incubated at 25 °C for 16 hours (overnight) with shaking. Each culture was centrifuged at 5000 g for 10 minutes and the supernatant was kept, and lysates prepared from the cell pellet. The remaining 0.5 mL was added to 5 mL of 2TY amp/glucose and incubated with shaking for 16 hours at 37 °C, and the DNA isolated by the alkaline lysis method.

Anti-VH blots.

For each supernatant and lysate sample from the IPTG induced cultures,

5μL was spotted onto a nitrocellulose filter, then the filter was dried for 5 minutes at 37 °C. The filter was placed in a solution of anti-VH mAb, diluted 1:5000 in PBSTA (~2.5 $\mu g.mL^{-1}$), on a rocking platform for 30 minutes. The filter was washed three times in PBSTA then placed in a solution of anti mouse mAb-alkaline phosphotase conjugate, diluted 1:1000 in PBSTA (~2.5 $\mu g.mL^{-1}$), on a rocking platform for 30 minutes and washed as before. The filter was then placed into a solution of NBT and BCIP (150 $\mu g.mL^{-1}$ BCIP, 300$\mu g.mL^{-1}$ NBT, 1 M diethanolamine, 1 mM $MgCl_2$ pH 9.8) again on the rocking platform until the colour developed.

ELISA for detection of Fv 4155 myc.

1. Conjugation: 100μL of an E3G-ovalbumin complex (2$\mu g.mL^{-1}$) in conjugation buffer (0.2M Na_2CO_3, 0.2M $NaHCO_3$) was placed into the appropriate wells of an ELISA plate and left at 4 °C for 16 hours (overnight).

2. Assay: For each step of the ELISA 100μL samples were added to the appropriate wells and reacted for 1 hour at room temperature, the plates were then washed in a plate washer using three washes with PBSTA. Before adding the *para*-nitrophenyl phosphate (PNPP) substrate the plate was washed five times, the reaction was left until colour could be seen and then the absorbance at 410 nm was read in a plate reader. The steps were as follows:

Step 1: Supernatant and lysate samples from the induced cultures were used neat or diluted two or 4 fold in PBSTA.

Step 2: Anti-myc antibody at 1:1000 dilution (2.68$\mu g.mL^{-1}$) in PBSTA.

Step 3: Anti-mouse antibody/alkaline phosphotase conjugate at 1:1000

dilution in PBSTA (~2.5 $\mu g.mL^{-1}$).

Step 4: 1 $mg.mL^{-1}$ PNPP (1 $mg.mL^{-1}$) in 1M diethanolamine, 1 mM $MgCl_2$, pH 9.8

ELISA for Specificity of Fv E3G myc.

1. Conjugation: Solutions of E3G-ovalbumin complex (2$\mu g.mL^{-1}$), Haemoglobin, BSA and lysozyme (all 10 $\mu g.mL^{-1}$) in conjugation buffer were made up. A 96 well ELISA plate was marked into rows for different antigens and the appropriate antigens were placed into the appropriate wells in 100µL samples and left at 4 °C for 16 hours (overnight).

2. Assay: Each step was carried out in the same way as described in the Fv E3G myc detection assay.

Step 1: Supernatant and lysate samples from the induced cultures were diluted 1:2, 1:4, 1:8, through to 1:64 in blocking buffer (2% BSA, 0.1% tween)

Step 2 to Step 4: Reagents the same as the Fv E3G myc detection assay.

Competition Assay of Fv E3G myc versus *Parent Antibody.*

1.Conjugation: The ELISA plate was conjugated with 100µL of E3G ovalbumin complex as before except the concentration was 80 $\mu g.mL^{-1}$.

2. Assay: For each step of the ELISA 100 µL samples were added to the appropriate wells and reacted for 1 hour at room temperature, the plates were washed in a plate washer using three washes with PBSTA at the end of each step. Before adding PNPP substrate the plate was washed five times. The reaction was left until colour could be seen and then the absorbance at 410 nm was measured in a plate reader. The reagents used

in each step were as follows:

Step 1: Samples of 4155 mAb and various concentrations of Fv fragment were reacted for 1 hour before being added to the plates

Step 2: Anti-mouse antibody/alkaline phosphotase conjugate at 1:1000 dilution in PBSTA (~2.5 $\mu g.mL^{-1}$).

Step 3: 1 $mg.mL^{-1}$ PNPP (1 $mg.mL^{-1}$) in 1M diethanolamine, 1 mM Mg Cl_2, pH 9.8

Inhibition Assay:

1. Conjugation: 100 µL of a 1:1000 solution of anti-mouse antibody (~2.5 $\mu g.mL^{-1}$) in conjugation buffer was placed into the appropriate wells of an ELISA plate and left at 4 °C for 16 hours (overnight).

2. Assay: For each step of the ELISA 100µL samples were added to the appropriate wells and reacted for 1 hour at room temperature, the plates were then washed in a plate washer using three washes with PBSTA at the end of each step. Before adding PNPP substrate the plate was washed five times and the reaction was left until colour could be seen, the absorbance at 410 nm was read in a plate reader. The solutions for each step are as follows:

Step 1: A 1:1000 solution of 4155 MAb in PBSTA (~2.5 $\mu g.mL^{-1}$).

Step 2: Samples with varying concentrations of E3G-AP and Fv E3G myc were reacted together for an hour before being used in the ELISA.

Step 3: 1 $mg.mL^{-1}$ PNPP (1 $mg.mL^{-1}$) in 1M diethanolamine, 1 mM $MgCl_2$, pH 9.8

Total Fv Assay.

Conjugation: 100 µL of a 1:1000 solution of 4155 mAb (~2.5 $\mu g.mL^{-1}$) in conjugation buffer was placed into the appropriate wells of an ELISA plate and left at 4 °C for 16 hours (overnight).

2. Assay: For each step of the ELISA 100µL samples were added to the appropriate wells and reacted for 1 hour at room temperature, the plates were then washed in a plate washer using three washes with PBSTA at the end of each step. Before adding PNPP substrate the plate was washed five times and the reaction was left until colour could be seen and then the absorbance at 410 nm was read in a plate reader. The solutions for each step are as follows:

Step 1: E3G-AP at a final dilution of 1:500, and and the samples of Fv E3G myc at various dilutions were reacted for one hour at room temperature, before being added to the ELISA plate.

Step 2: 1 $mg.mL^{-1}$ PNPP (1 $mg.mL^{-1}$) in 1M diethanolamine, 1 mM $MgCl_2$, pH 9.8

Preparation of supernatants and periplasmic Lysates.

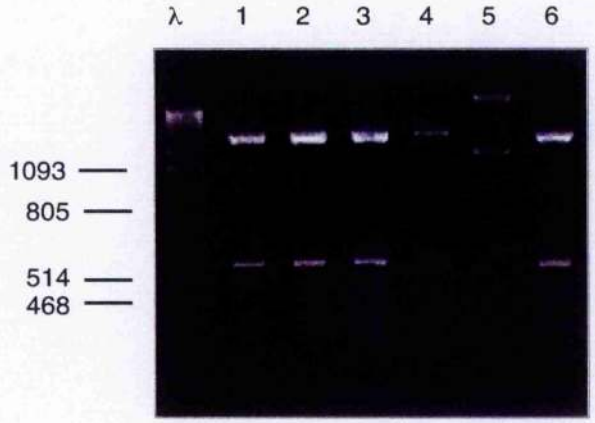
The culture was spun at 5000 g for 20 minutes, and supernatants decanted. The cell pellet was resuspended in 0.1 volume (0.5 mL) of lysis buffer(20 % sucrose, 200 mM Tris.Cl pH 7.5; 500 $\mu g.mL^{-1}$ lysozyme), and incubated at room temperature for 10 minutes. A 0.1 volume of sterile milli-Q H_2O was added and the sample incubated for another 10 minutes. This was placed into a sterile eppendorf tube, and centrifuged in a micro-centrifuge at full speed for 5 minutes and the supernatants moved into a fresh tube - periplasmic lysate.

3.3 Results and Discussion

The detection and expression experiments were done in three parts, the results from these can be seen in figure 3.1 and table 3.1. The first part the *Pst I* restriction digest shows that from five of the colonies picked DNA was isolated, and for all of these the anti-E3G V_L myc DNA, is present in the plasmid (this result is previously shown in Chapter 2). The anti-V_H blots show that again for all the clones except for number 5 (which did not grow) protein has been expressed. It also indicates that most of the recombinant protein is in the supernatant as the spots are much weaker for the corresponding periplasmic lysates. This serves a useful purpose, because it will give proof of expression, even if the antibody fragment produced does not bind to the antigen. Because the anti-V_H mAb was raised against the peptide that the primer PCR51 expresses, the antibody will bind to any Fv fragment that uses this primer when cloned. From the detection ELISA to show binding of E3G (of which a schematic representation is given in figure 3.2), these results indicate all the clones, apart from number 5, show anti-E3G activity, and because none of the negative controls gave positive results, this activity must be due to the expressed recombinant protein being active in binding E3G. The negative controls are deficient in one of the necessary reagents to give a positive ELISA result, therefore if none of these appear as positive we can be sure there is no background binding by any of the reagents, that is no binding of any of the reagents to the ELISA itself.

Now the expression and binding of E3G of the Fv fragment has been demonstrated it is important to show that the Fv fragment binding to E3G is specific. To do this the detection assay is repeated the with various other antigens. The results from this assay can be seen in table 3.2.

Pst I digest.



Anti-V_H blot

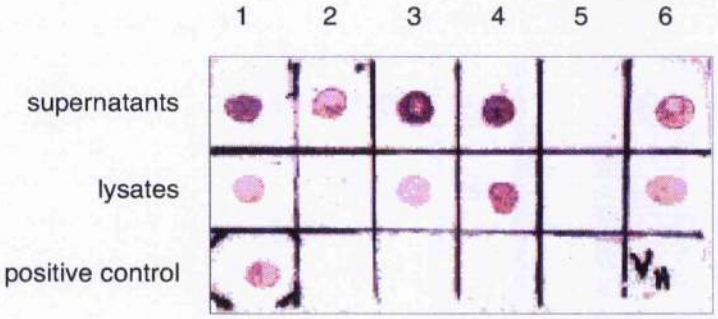


Figure 3.1 Results from expression and detection experiment: results of anti-V_H blots and agarose gel of *Pst* I digest of final construct - Fv E3G myc pUC 19.

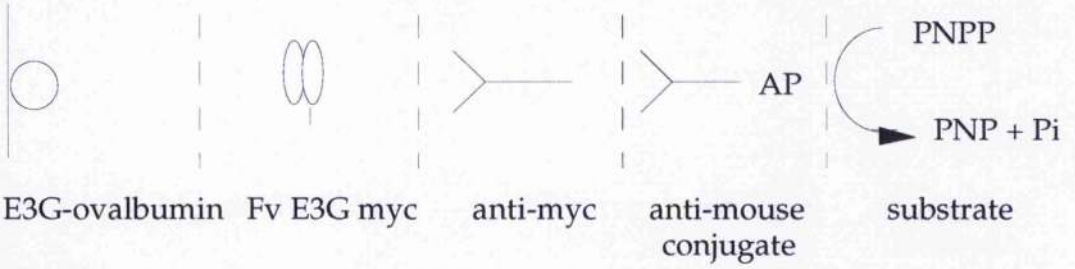


Figure 3.2 Representation of the ELISA used for detection of E3G binding by the Fv fragment. (The specificity assay uses exactly the same scheme, but other antigens are used as well as E3G-ovalbumin).

supernatant dilution		Sample number					
		1	2	3	4	5	6
	1:1	1.200	1.289	1.029	1.227	0.006	1.203
		1.133	1.234	1.318	1.315	0.009	1.269
	1:2	1.257	1.225	1.192	1.245	0.013	1.098
		1.162	1.073	1.215	1.156	0.007	1.016
	1:4	0.985	0.953	1.152	1.143	0.008	0.933
		1.147	0.959	1.145	1.163	0.005	0.923

Lysate dilution		Sample number					
		1	2	3	4	5	6
	1:1	0.956	0.472	0.796	1.168	0.007	0.343
		1.001	0.458	0.564	1.117	0.007	0.671
	1:2	0.862	0.314	0.520	1.079	0.009	0.656
		0.860	0.339	0.551	0.908	0.006	0.764
	1:4	0.614	0.132	0.427	0.694	0.006	0.661
		0.572	0.140	0.423	0.730	0.004	0.451

	E3G	Fv	anti-myc	conj.	Absorbance
1	+	+	+	+	0.942
2	+	+	+	-	0.010
3	+	+	-	+	0.008
4	+	-	+	+	0.006
5	+	-	-	+	0.008
6	+	-	+	-	0.009
7	-	+	+	+	0.003
8	-	+	+	-	0.001
9	-	+	-	+	0.001
10	-	-	+	+	0.003
11	-	-	-	+	0.002
12	-	-	+	-	0.005

Table 3.1 ELISA results for detection of Fv E3G myc expression in *E. coli* culture supernatants and periplasmic lysates.

E3G-ovalbumin concentration (μg/ml)

Fv supernatant dilution	3.2	1.6	0.8	0.4	0.2	0.1	
0	0.001	0.006	0.001	0.053	0.002	0.001	No anti-myc
1.2	0.002	0.003	0.002	0.006	0.001	0.004	
1:2	0.828	0.547	0.211	0.156	0.014	0.016	anti-myc
1:4	0.561	0.401	0.165	0.121	0.011	0.013	
1:8	0.322	0.215	0.088	0.090	0.008	0.008	
1:16	0.093	0.060	0.037	0.044	0.003	0.002	
1:32	0.006	0.007	0.022	0.129	0.005	0.004	
0	0.004	0.016	0.040	0.119	0.058	0.039	

Various antigens (4.0 μg/ml)

Fv supernatant dilution	Haem.	BSA	Lys	
0	0.000	0.001	0.083	No anti-myc
1.2	0.001	0.001	0.020	
1:2	0.002	0.001	0.024	anti-myc
1:4	0.000	0.000	0.039	
1:8	0.003	0.000	0.052	
1:16	0.020	0.002	0.0380	
1:32	0.001	0.001	0.031	
0	0.005	0.003	0.074	

Table 3.2 Results of specificity ELISA for Fv E3G myc towards other antigens.

These data demonstrate reduced binding when either the amount of immobilised E3G-ovalbumin or the concentration of the supernatant containing the expressed Fv fragment is reduced as would be expected. More importantly the results show that there is no binding to either of the other antigens used: haemoglobin, bovine serum albumin (BSA) and lysozyme. The results for the lysozyme do show absorbance readings above background, but these also occur where there is no Fv present, and also where there is no anti-myc mAb present, so cannot be due to non-specific binding of the anti-E3G Fv to lysozyme. The list of antigens used in this experiment is in no way exhaustive, but the activity of the antibody fragment to its own antigen and to the other antigens shown in this assay, it would be fair to assume that there is no significant activity to other antigens and that the binding is specific.

Competition Assay.

To further characterise the behaviour of this recombinant protein, an experiment is required to assess the Fv fragment in competition with its parent mAb. For the Fv fragment to bind E3G it must compete in the presence of the parent mAb, and therefore give an indication of the relative affinities of the two. It is evident that if there was to be an excess of E3G-ovalbumin bound to the plate, there would be no competition between the Fv fragment and its parent mAb, therefore the concentration of the E3G-ovalbumin must be limiting. The results from the specificity assay show that at a concentration of 0.8 $\mu g.mL^{-1}$ of E3G-ovalbumin there is indeed a limiting concentration as there is increased binding with each increase in Fv concentration. In the described competition assay (figure 3.3) any binding of the Fv fragment to the immobilised E3G ovalbumin in

preference to the 4155 mAb would result in a reduced absorbance reading in the assay. The concentrations and results of this assay are shown in table 3.3. At the lower concentration of Fv there is no significant effect, but as the concentration is increased to around 1:8 of the original supernatant concentration the Fv starts to compete out some of the parent mAb binding to E3G-ovalbumin, which is shown by the reduced absorbance readings. The controls in this ELISA experiment only gave a positive result when all the reagents were present except the Fv fragment, and where any reagent required in the chain of binding reactions is missing, there was no positive result. Because no protein has been purified from the supernatant there is no absolute concentration for the antibody fragment - as there is for the mAb, therefore a direct affinity comparison is not possible. Despite this lack of quantitative data, the assay tells us qualitatively that the Fv fragment is capable of competing out the parent antibody, which would not be possible if the fragment bound very weakly to the antigen.

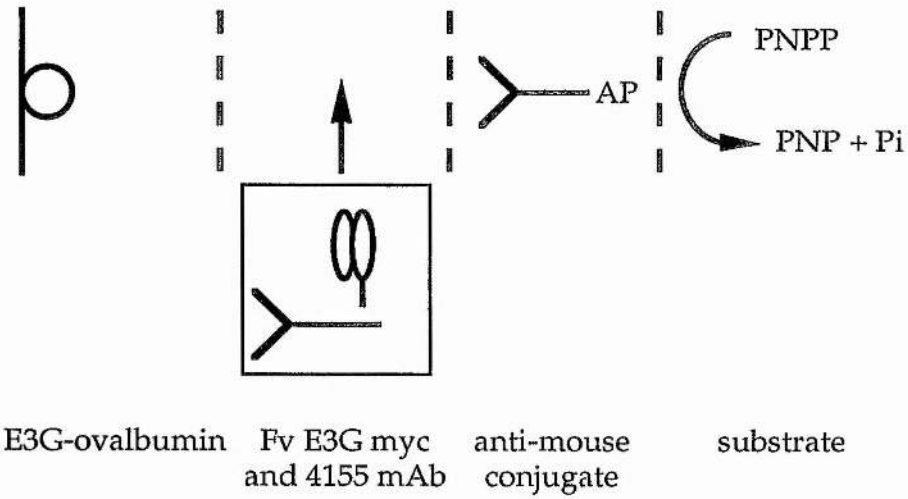


Figure 3.3 Representation of the competition ELISA, used in demonstrating the Fv fragment competes with its parent mAb (4155).

4155 mAb conc. (µg/ml)	Fv supernatant concentration							
	0	1:128	1:64	1:32	1:16	1:8	1:4	1:2
1.0	1.060	0.825	0.901	0.941	0.788	0.705	0.613	0.526
	1.049	1.023	0.987	0.930	0.861	0.779	0.683	0.583
0.5	0.789	0.778	0.703	0.764	0.668	0.481	0.393	0.373
	0.767	0.831	0.795	0.748	0.687	0.533	0.440	0.396
0.25	0.602	0.617	0.584	0.507	0.514	0.440	0.295	0.238
	0.660	0.661	0.611	0.579	0.484	0.445	0.372	0.231

Experiment Controls

Number	E3G	4155 / Fv		anti-mouse	Absorbance
1	+	+	-	+	0.721
2	+	-	+	+	0.016
3	+	+	-	-	0.010
4	+	-	+	-	0.011
5	-	+	-	+	0.009
6	-	-	+	+	0.010
7	-	+	-	-	0.011
8	-	-	+	-	0.010

Table 3.3 Results of ELISA to demonstrate that Fv E3G myc competes with its parent mAb.

Inhibition assay.

This assay (figure 3.4) relies on inhibition of 4155 mAb binding an estrone-3-glucuronide-alkaline phosphotase conjugate (E3G-AP), by adding Fv E3G myc to inhibit this binding. Again if there was poor binding of the of the recombinant antibody fragment to the E3G-AP conjugate the fragment would be easily displaced by the mAb. Before this assay can be run it is important that there is not an excess of E3G-AP conjugate, if this was the case, if inhibition did occur it would not be detected as the excess E3G-AP would still bind the 4155 mAb therefore not affecting the final absorbance reading. By using varying concentrations of

4155 mAb and E3G-AP conjugate the limiting concentration of E3G-AP was determined. The results of this test assay was to show that any of the dilutions of the E3G-AP conjugate used (from 1:250 - 1:1000) would be suitable. A concentration of 1:500 was chosen as the maximum for the inhibition assay and also concentrations of 1:1000 and 1:2000 were used. The results obtained from this assay are displayed table 3.4. These demonstrate, that even at low concentrations of Fv, the E3G-AP binding of the Fv fragment inhibits the binding of the conjugate to the parent monoclonal. Therefore even at large excess of mAb, this does not displace Fv binding from the E3G-AP conjugate.

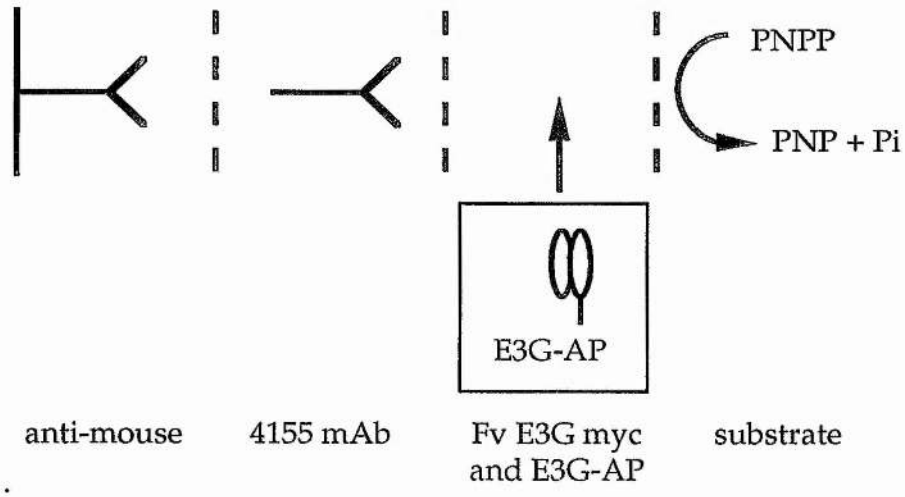


Figure 3.4 Representation of the ELISA used to demonstrate the inhibition of 4155 mAb (parent antibody) by Fv E3G myc.

E3G-AP dilution	Fv supernatant concentration							
	0	1:128	1:64	1:32	1:16	1:8	1:4	1:2
1:500	0.755	0.389	0.271	0.193	0.101	0.054	0.025	0.019
	0.788	0.403	0.279	0.198	0.093	0.061	0.036	0.019
1:1000	0.416	0.171	0.128	0.086	0.042	0.029	0.012	0.009
	0.334	0.193	0.123	0.087	0.055	0.032	0.018	0.013
1:2000	0.230	0.101	0.075	0.050	0.031	0.015	0.012	0.011
	0.224	0.109	0.078	0.052	0.032	0.019	0.012	0.018

Controls

Number	anti-mouse	4155	E3G-AP	/ Fv	Absorbance
1	+	+	+	-	0.658
2	+	+	-	+	0.008
3	+	-	+	-	0.000
4	+	-	-	+	0.006
5	-	+	+	-	0.006
6	-	+	-	+	0.009
7	-	-	+	-	0.003
8	-	-	-	+	0.005

Table 3.4 Results of ELISA used to determine that Fv E3G myc can inhibit binding of the parent 4155 mAb to E3G-AP.

Total Fv assay.

When antibody fragments with myc peptide tails are expressed in *E. coli* cultures this tail can often be subject to proteolysis, and since the detection of the anti E3G antibody fragment depends on binding of this tail by the anti-myc mAb, loss of this tail would result in apparently low concentrations of active protein. This assay has also made it possible to quantify concentrations of expressed anti-E3G Fv fragment in supernatants and periplasmic lysates, when compared against standards of the purified recombinant protein at known concentration. This assay was shown to be of particular use in the optimisation of expression of the

Fv fragment (Chapter 4). It is clear from the plan of this assay (figure 3.5) that an excess of the E3G-AP conjugate would result in positive detection of the PNP product despite the presence of the anti-E3G Fv fragment, therefore the conjugate must be at limiting concentration. To find this concentration an assay must be carried out without any Fv to determine this limiting concentration, at all concentrations from 1:250 to 1:2000 the E3G-AP was found to be at limiting concentration. A dilution of 1:500 of the E3G-AP conjugate was decided as the suitable concentration for the total anti-E3G Fv assay. The assay was now suitable to be applied in quantitative assays for determining the concentration of active protein in cell cultures.

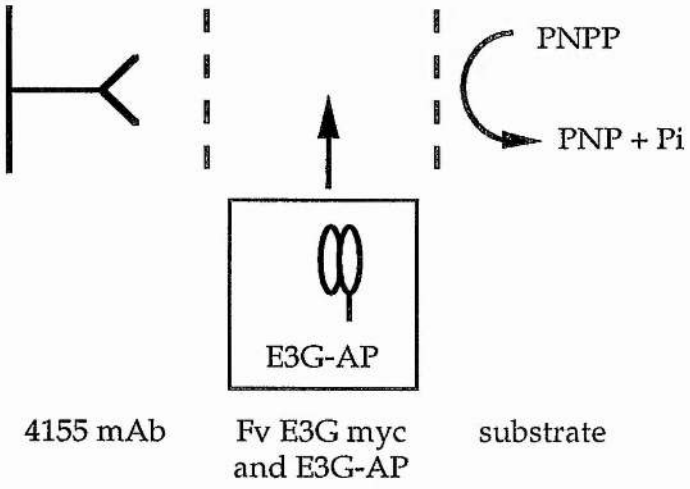


Figure 3.5 Representation of the ELISA developed for detection of Fv binding to E3G, that does not rely on the myc tail.

3.4 Conclusions

The successful cloning of the two variable light and heavy chain genes into the expression vector described in the previous chapter had provided the means to express the recombinant protein in a bacterial culture. The initial experiment proved the correct construct was indeed produced, that protein was being expressed in the anti-V_H blots, and from the initial ELISA experiment, that the expressed recombinant protein was active in binding E3G. Because the protein was to be used in a NMR study of the antibody-antigen interaction this was vital. The parent monoclonal antibody was developed to be used in a diagnostic test, because of the cost of producing intact antibodies from mammalian cell cultures it is vital that the Fv fragment - if it is to replace the use of the mAb - shows specificity to the antigen, and that the recombinant protein demonstrates the same properties as the parent mAb. The specificity ELISA demonstrates this selectivity, although if the fragment was to be used in preference to the mAb it would have to be tried against the large panel of antigens that the panel of mAb raised against E3G was in its selection, as the role the fragment would need to fulfil it would need to show specificity over many molecules of similar structure to the antigen. The fact that the Fv fragment shows good competition against its parent antibody, and also good inhibition of 4155 mAb-E3G binding, would suggest strongly that these properties are very similar. This may be due to the small nature of the antigen; the strong avidity effects seen in some mAbs, which lead to much improved binding over their Fv counterparts, are not important. From these reactions there can be confidence that the protein will be suitable for both the desired applications.

Chapter 4.

Expression and Purification of Fv E3G myc.

Abstract

To fully exploit the use of recombinant Fv fragments, it is important that significant quantities of the protein can be expressed and subsequently purified. This chapter describes the optimisation of expression levels of the recombinant protein in *E. coli*, using two different media, and the method of purification of the protein from the cell cultures. Initial expression levels were found to be low: 1 $mg.L^{-1}$. Problems of denaturation of the protein especially in the unlabelled celtone media (Martek Biosciences) were overcome by using this media buffered with phosphate buffer. Expression levels of up to 4 $mg.L^{-1}$ were then obtained. Purification of the Fv fragment was achieved by affinity chromatography, relying on the different affinities of the Fv fragment to antigen analogues.

4.1 Introduction

The ability to express Fv antibody fragments in bacterial or yeast expression systems represents a significant advantage over mAbs isolated from tissue culture. The reduced cost of producing proteins in bacterial culture open up the range of commercial applications to which these Fv fragments can be applied. Allied with the use of modern molecular biology techniques such by-passing immunisation (Marks *et al.*, 1991), which opens up exciting developments in their use as therapeutic agents, the potential of these molecules are massive.

Bacterial and yeast expression systems (Plückthun, 1990; Plückthun and Skerra, 1989; Better *et al.* 1988; Skerra and Plückthun, 1988; Horwitz *et al.* 1988; Wood *et al.* 1985; Ridder *et al.* 1995; Nyyssönen *et al.* 1993; zu Putlitz *et al.* 1990) allow the quantities required for structural studies by X-ray

crystallography and NMR to be obtained more easily, together with isotopic labelling and the size of the Fv fragment it is possible to elucidate the full solution structure of these proteins by NMR. When molecular biology mutagenesis techniques are allied to structural studies of these mutagens, a far better understanding of the specific nature of antigen-antibody binding can be obtained.

In this chapter work in optimising expression levels of the recombinant protein in different media is described. Because of denaturation of the protein during the culture experiments, a number of derivative media were made from the commercially available celtone media. The quantitative analysis of each culture for active Fv fragment was accomplished using the assay for Fv detection that does not rely on the myc tail (developed in Chapter 3).

When purifying mAbs and Fab fragments, the constant regions on these allow easy affinity purification, as the Fv fragments do not have any constant regions, affinity purification therefore, must rely on the specificity of the antigen. The mAb and Fab purification can be mimicked by using polypeptide tails at the end of the V_L and using antibodies raised against these in purification. In this chapter a purification method is described that relies on the different affinities of various antigen analogues to the anti-E3G Fv fragment.

4.2 Materials and Methods

Plasmid Stability.

From cultures of the plasmid in *E. coli* on 2TY plates different colonies were selected, and streaked onto 2TY amp plates. Two colony types were

found and for both colony types a single colony was picked and grown in 10 mL BHI, and incubated overnight at 37°C. For each colony type duplicate samples were set up at 25 °C, 30 °C and 37°C, and these samples were monitored at 7 and 11 hours. At 7 hours identical samples were set up which were allowed to grow overnight so these could be monitored at 17 hours. The cultures were set up as follows: 100 mL of M9P, 5 mL yeast extract (10%), 1 mL thiamine (0.01%), 200µL $MgSO_4$ (1M), and 100µL $CaCl_2$, was inoculated with 1 mL of the overnight culture in BHI (brain-heart infusion) in a 250 mL baffle flask. These were grown at the required temperature with shaking (200 rpm) for the required time. 1 mL samples were taken from each sample at the required times, and a series dilution of each sample was made in sterile saline from a range of 10^{-1} - 10^{-8}. For each sample 50 µL of the 10^{-3} through to the 10^{-8} dilutions were pipetted onto 2TY plates. (three 50µL samples per plate) and onto 2TY amp plates in an identical manner.

Production screen of Fv 4155 myc pUC 19 E. coli *JM109.*

For one particular colony type an overnight culture was set-up in BHI as before. Cultures were set up for each temperature (25 °C, 30 °C or 37°C) for each different length of induction time (7, 11 or 17 hours) and again in duplicate for each condition (18 flasks). The cultures were set up as follows: 100 mL of M9P and 5 mL yeast extract (10%), 1 mL thiamine (0.01%), 200µL $MgSO_4$ (1M), and 100µL $CaCl_2$, was inoculated with 1 mL of the overnight culture in BHI in a 250 mL baffle flask. These were grown at the required temperature with shaking (200 rpm) then induced with 100µL IPTG (0.5M).at the required time for each flask. For each flask samples were taken prior to induction (1 mL samples), after 24 hours

growth and after 48 hours growth (both 1 mL and 5 mL samples). The samples were analysed as follows.

Pre-induction samples: Viable counts were made as for the plasmid stability experiment.

24 hour and 48 hour 1 mL samples: Viable counts were made as for the plasmid stability experiment, then the absorbance of the 10^{-1} dilution was measured at 600 nm against a 10^{-1} dilution of the growth media in sterile saline as a reference.

24 hour and 48 hour 5 mL samples: Supernatants and periplasmic lysates were prepared and these samples were analysed with the specific E3G assay using 0.5, 0.25 and 0.125 dilutions.

Purification of Fv 4155 myc.

1. Preparation of Estrone linked sepharose: 4 g of CMB epoxy sepharose (Pharmacia) was allowed to swell in 250 mL of milli-Q H_2O for 45 minutes. This was washed by filtration with 1 litre of milli-Q, then 200 mL of 85% DMF (pH 13). 50 mg of estrone was dissolved in 8 mL of DMF and the sepharose added to this and placed in a hybridisation oven at 45°C with rotation overnight (16 hours). The sepharose was washed by filtration with 200 mL of of 85% DMF, then with 1 L milli-Q H_2O and packed into a P16 column (Pharmacia). The column was washed through with milli-Q H_2O.

2. Column purification: 5M NaCl was added to the cell supernatant or periplasmic lysate to a final salt concentration of 1 M. This was loaded on the column at 3 $mL.min.^{-1}$. After loading the column was washed with PBSA. The protein was eluted with 10 mL of estradiol glucuronide (0.2 $mg.mL^{-1}$), and the eluted material put through a gel filtration column

to remove excess estradiol glucuronide. The column was then washed with 50 mM citric acid, followed by PBSA to clean it prior to the next usage.

4.3 Results and Discussion

Plasmid stability.

The results of the plasmid stability experiments are shown in table 4.1. The stabilities after 7 hours show that in general the type II colonies show excellent stability for each temperature, apart from what appears an anomalous result for culture number 11. The observations for the type I colony are less encouraging, in general stability is poorer for the higher temperatures as would be expected. The 11 hour results suggest that that both colony types are stable. in some cases more so than the 7 hour reactions. Because the 11 hour results match much better between duplicate samples, these would be more reliable. Again after 17 hours plasmid stability is excellent for both colony types apart from at 37 °C, where there is a definite reduction in stability. Since the duplicates give similar results these results would seem valid. The results show that both colony types have plasmids which remain stable for the length of time needed to acquire biomass before induction with IPTG, so that production screens will have to be carried out for both colony types.

			Plasmid Stability (%)		
No.	Colony	Temp.	7 hours	11 hours	17 hours
1	Type I	37	20	100	-
2	Type I	37	67	100	-
3	Type I	30	21	65	-
4	Type I	30	34	62	-
5	Type I	25	40	100	-
6	Type I	25	40	100	-
7	Type II	37	100	100	-
8	Type II	37	100	81	-
9	Type II	30	100	100	-
10	Type II	30	100	100	-
11	Type II	25	56	100	-
12	Type II	25	100	52	-
13	Type I	37	-	-	72
14	Type I	37	-	-	77
15	Type I	30	-	-	61
16	Type I	30	-	-	100
17	Type I	25	-	-	100
18	Type I	25	-	-	100
19	Type II	37	-	-	69
20	Type II	37	-	-	58
21	Type II	30	-	-	100
22	Type II	30	-	-	92
23	Type II	25	-	-	100
24	Type II	25	-	-	100

Table 4.1 Results of plasmid stability experiments of type I and type II colonies of Fv E3G myc pUC 19 E. coli JM109 in M9P and yeast extract.

Type II production screen in M9P and YE (yeast extract).

Results from this production screen show very little production of Fv, compared to the samples grown previously for use in ELISA experiments. It was concluded there may have been a problem with the batch of IPTG that was used. A test experiment showed this not to be the case, therefore this colony type is probably no use for expressing the recombinant protein.

Type I production screen in M9P and YE.

Again cultures were grown in duplicate at the required temperature and induction time. Results from the samples taken are shown in table 4.2. The plasmid stability data shows that the stability is generally very good prior to induction and this generally remains the case except for two exceptions: all the 37 °C samples loose most of the plasmid at 24 hours, although, again there is poor comparison between some of the duplicates. The pre-induction absorbance measurements as expected show higher levels of growth for the samples at higher temperatures and longer induction times, but all continued growth after induction. The ELISA results indicate that the best conditions for the type I colony type in M9P+ YE are 25 °C, 7 hours induction time and a growth of 24 hours. One other point is their is very little activity in the periplasmic lysate samples, so nearly all of the protein was in the cell supernatant. When these samples are analysed against a series dilution of known concentrations of anti-E3G Fv the best expression levels are at best 1 $mg.mL^{-1}$. This assumes that the purified Fv used as a standard is 100 % pure, that there is no loss of activity on purification and all the protein can be recovered from the media. The mild purification conditions will not affect activity but it is certain that the protein will not be 100 % pure and that all the protein will not be recoverable, therefore the estimated value will be a best guess, and the amount recovered would be lower than this. This value is still too low for producing significant amounts of labelled material.

Culture Conditions			Absorbance			Plasmid Stability		
No.	Temp.	Induction	P.I.	24 hr.	48 hr.	P.I.	24 hr.	48 hr.
1	37	7 hr.	2.063	1.703	2.182	100	2	11
2	37	7 hr.	2.075	2.066	2.257	100	2	53
3	30	7 hr.	1.813	1.754	2.350	47	17	100
4	30	7 hr.	1.787	1.830	2.384	59	11	100
5	25	7 hr.	1.164	1.148	2.080	11	91	35
6	25	7 hr.	1.219	0.837	2.210	63	100	87
7	37	11 hr.	0.923	1.756	2.212	63	3	1
8	37	11 hr.	1.165	1.976	2.453	57	3	72
9	30	11 hr.	1.462	2.053	2.261	100	100	96
10	30	11 hr.	1.503	2.114	2.290	81	95	100
11	25	11 hr.	0.802	1.328	1.592	60	100	43
12	25	11 hr.	0.812	1.320	2.128	82	4	74
13	37	17 hr.	1.802	1.774	2.051	2	0	0
14	37	17 hr.	1.980	1.732	2.454	4	0	0
15	30	17 hr.	1.130	2.015	2.156	67	100	100
16	30	17 hr.	1.381	2.174	2.304	78	100	61
17	25	17 hr.	1.407	1.943	2.307	95	65	100
18	25	17 hr.	1.362	2.034	2.325	100	65	100

Table 4.2 (A) Plasmid stability and cell culture absorbance results from a type I production screen in M9P and yeast extract.

Supernatants:

Culture Conditions			Absorbance after 24 hours			Absorbance after 48hours		
No.	Temp.	Induction	1:2	1:4	1:8	1:2	1:4	1:8
5	25	7 hr.	0.318	0.167	0.082	0.010	0.010	0.011
6	25	7 hr.	0.646	0.422	0.188	0.009	0.012	0.010
11	25	11 hr.	0.377	0.255	0.157	0.066	0.058	0.030
12	25	11 hr.	0.422	0.302	0.246	0.072	0.065	0.036
17	25	17 hr.	0.019	0.019	0.017	0.006	0.009	0.007
18	25	17 hr.	0.018	0.019	0.015	0.008	0.004	0.007

Lysates:

Culture Conditions			Absorbance after 24 hours			Absorbance after 48hours		
No.	Temp.	Induction	1:2	1:4	1:8	1:2	1:4	1:8
5	25	7 hr.	0.047	0.021	0.018	0.009	0.010	0.010
6	25	7 hr.	0.046	0.019	0.017	0.012	0.010	0.011
11	25	11 hr.	0.523	0.033	0.020	0.813	0.125	0.009
12	25	11 hr.	0.631	0.329	0.025	0.864	0.104	0.008
17	25	17 hr.	0.276	0.056	0.021	0.011	0.012	0.010
18	25	17 hr.	0.281	0.062	0.020	0.012	0.012	0.012

Table 4.2 (B) ELISA results from cell supernatants and periplasmic lysates of cell cultures from a type I production screen in M9P and yeast extract for detection of Fv E3G myc

Interestingly ELISA samples of the same reaction after 48 hours has elapsed actually indicate a lower detection of Fv in the anti-myc based assay. This means that some of the Fv must be denatured in some way after 48 hours. There are two possibilities: the protein could be denatured enough to loose function and then would be inactive in the E3G binding part of the assay; another possibility is the loss of the exposed myc tail, which has been shown previously to be subject to proteolysis, and protein affected in this manner would not be detected by the anti-myc based assay. Therefore it is possible that more Fv is being produced but that this protein is undetectable in the assay. This can be resolved by using the 'total Fv' assay which does not require detection of the myc peptide tail and therefore would give an accurate analysis of active protein present. Again using known concentration of fully active anti-E3G in the assay would give a standard curve from which the concentration of active protein in the culture samples could be estimated. When the samples taken after 48 hours have lower levels of active protein than their corresponding 24 hour samples when measured with the total-Fv assay,

therefore the loss in levels of active protein in this time can not only be due to loss of the myc tail, but in fact there must be denaturation of the protein to an extent it looses activity.

Type I production screen in celtone.

This experiment was carried out in a similar fashion to the previous experiments, but this time an additional induction time of 4 hours was added to the experiment for each of the three temperatures. The various details of each culture, along with their results for plasmid stability absorbances and ELISA experiments are shown in table 4.3.

Culture Conditions			Absorbance			Plasmid Stability		
No.	Temp.	Induction	P.I.	24 hr.	48 hr.	P.I.	24 hr.	48 hr.
1	37	4 hr.	0.047	0.952	1.012	61		100
2	37	4 hr.	0.054	1.021	1.125	100		100
3	30	4 hr.	0.020	1.041	1.867	100	45	0
4	30	4 hr.	0.019	1.124	1.750	93	33	0
5	25	4 hr.	0.053	1.028	1.625	70	95	0
6	25	4 hr.	0.063	1.102	1.439	68	56	0
7	37	7 hr.	1.128	1.071	1.527	45	100	0
8	37	7 hr.	0.936	1.085	1.562	62	6	0
9	30	7 hr.	0.062	1.152		100	100	0
10	30	7 hr.	0.053	1.085		100	100	0
11	25	7 hr.	0.823	1.159		100	1	0
12	25	7 hr.	0.815	1.324		100	6	0
13	37	11 hr.	0.926	1.191		2	1	0
14	37	11 hr.	0.824	1.315		0	53	0
15	30	11 hr.	1.229	1.410		100	76	0
16	30	11 hr.	1.116	1.325		100	46	0
17	25	11 hr.	1.211	1.426		100	1	0
18	25	11 hr.	1.287	1.325		65	0	0
19	37	17 hr.	1.054	1.210		63	3	1
20	37	17 hr.	1.071	1.252		12	2	0
21	30	17 hr.	1.119	1.316		100	21	4
22	30	17 hr.	1.196	1.324		100	100	1
23	25	17 hr.	1.086	1.205		100	43	1
24	25	17 hr.	1.126	1.327		100	100	3

Table 4.3 (A) Plasmid stability and cell culture absorbance results from a type I production screen in celtone media.

Supernatants:

Culture Conditions			Absorbance after 24 hours			Absorbance after 48 hours		
No.	Temp.	Induction	1:2	1:4	1:8	1:2	1:4	1:8
5	25	4 hr.	0.134	0.049	0.016	0.024	0.012	0.006
6	25	4 hr.	0.107	0.032	0.01	0.016	0.008	0.007
9	30	7 hr.	0.441	0.292	0.086	0.067	0.032	0.014
10	30	7 hr.	0.464	0.271	0.093	0.067	0.036	0.013
21	30	17 hr.	0.357	0.168	0.056	0.091	0.039	0.014
22	30	17 hr.	0.266	0.118	0.040	0.046	0.028	0.011
23	25	17 hr.	0.090	0.029	0.011	0.035	0.014	0.007
24	25	17 hr.	0.169	0.053	0.040	0.051	0.020	0.011

Lysates:

Culture Conditions			Absorbance after 24 hours			Absorbance after 48 hours		
No.	Temp.	Induction	1:2	1:4	1:8	1:2	1:4	1:8
5	25	4 hr.	0.453	0.343	0.169	0.016	0.014	0.013
6	25	4 hr.	0.412	0.243	0.098	0.014	0.014	0.012
9	30	7 hr.	0.621	0.498	0.280	0.012	0.013	0.012
10	30	7 hr.	0.683	0.511	0.316	0.013	0.016	0.011
21	30	17 hr.	0.937	0.898	0.453	0.020	0.010	0.005
22	30	17 hr.	0.769	0.718	0.444	0.012	0.008	0.006
23	25	17 hr.	0.444	0.354	0.233	0.009	0.006	0.005
24	25	17 hr.	0.482	0.232	0.157	0.005	0.004	0.003

Table 4.3 (B) ELISA results from cell supernatants and periplasmic lysates of cell cultures from a type I production screen in celtone for detection of Fv E3G myc.

Results from the plasmid stability experiments again proved to be quite poor in their duplication, but did show a lack of stability again at higher temperatures. The absorbance experiments illustrated the much slower growth of the cultures in this medium compared to the M9P + YE, especially at 25 °C. Although all the cultures do grow throughout the experiments, and the growth at the higher temperatures is good. ELISA results are fairly conclusive in pointing to the best conditions for growth in this media. Sample numbers 9, 10 and 21, 22 give the highest levels of detection in the ELISA experiment. These refer to 30 °C samples after 24 hours for 7 hours and 11 hours induction times respectively, these ELISA data also show significant levels of activity from the periplasmic lysates, so that in a large scale culture both supernatants and periplasmic lysates would have to be prepared to maximise the yield of Fv. Again lower reactivities in the ELISA were found for the corresponding cultures after 48 hours.

A quantitative anti-myc based assay was carried out on the best producing samples after 24 hours. Because there were significant levels of expression detected in both the supernatants and periplasmic lysates, both types of sample were used in this assay. This assay predicted expression levels of up to 1.5 $mg.L^{-1}$ of active protein, which although slightly improved on the experiment with the M9P media does not reach the required levels. The conditions which gave the best results had the samples taken at 48 hours analysed with the total-Fv assay, again gave insight into the problem of denaturation of the protein. These were again lower than the corresponding estimates for the samples taken after 24 hours (using the anti-myc based assay), so therefore suggesting a denaturation of active

protein over this time period.

Because the two previous production screens have showed a definite time dependency on the best levels of detected active protein, and also that the celtone media has shown slightly the best results, it was decided to repeat the experiment in celtone for those reaction which gave the best results, but to monitor these reactions over several more sampling points. From the previous screen the conditions which gave the best expression levels were those grown at 30 °C, and induced after either 11 hours or 17 hours. The 11 hour induced reactions were tested prior to induction then at 24 hours, 26 hours, 28 hours and 31 hours. The sample induced at 11 hours had samples tested prior to induction and at 19 hours 21 hours and 24 hours. All samples had the absorbance at 600 nm recorded and those taken after induction had supernatants and lysates prepared in the usual way and those tested by ELISA. These results are displayed in table 4.4. The absorbance results for the duplicate 11 hour induced samples show good growth prior to induction, which increases after induction but drops at 48 hours. The 17 hour induction time samples shows a higher level of growth prior to induction as would be expected, but this level stays fairly consistent. The anti-myc based assay results demonstrate for the the 11 hour induction time the best results are obtained after 31 hours whilst for the 17 hour induction time the best results are achieved for the 21 hour sample. Because of these improvements these samples were tested against known standards in a quantitative assay using the total-Fv assay. The best producing samples showed very similar results to the previous screen, with optimum levels $\sim 1.5\ mg.L^{-1}$.

Absorbance at 600 nm for cell cultures:

time		11 hours	26 hours	28 hours	31 hours
culture no.	1	0.881	1.116	1.102	0.843
	2	0.904	1.116	0.945	0.501
		17 hours	19 hours	21 hours	24 hours
	3	1.295	1.268	1.225	1.112
	4	1.201	1.275	1.226	1.063

ELISA results (O.D. at 410 nm).

		24 hours			time	26 hours		
culture no.		1/2	1/20	1/100	dilution	1/2	1/20	1/100
supernatant	1	0.389	0.009	0.008		0.581	0.012	0.007
	2	0.322	0.008	0.008		0.571	0.011	0.009
lysate	1	0.802	0.029	0.008		0.852	0.263	0.008
	2	0.769	0.015	0.006		0.979	0.229	0.008

		28 hours			31 hours		
culture no.		1/2	1/20	1/100	1/2	1/20	1/100
supernatant	1	0.660	0.029	0.006	0.876	0.059	0.009
	2	0.722	0.056	0.006	0.696	0.023	0.006
lysate	1	0.938	0.116	0.008	0.875	0.044	0.004
	2	0.913	0.160	0.006	0.906	0.097	0.003

		17 hours			19 hours			21 hours		
culture no.		1/2	1/20	1/100	1/2	1/20	1/100	1/2	1/20	1/100
supernatant	3	0.755	0.036	0.005	0.953	0.051	0.051	0.854	0.058	0.004
	4	0.603	0.015	0.005	0.964	0.043	0.043	0.865	0.022	0.004
lysate	3	0.921	0.195	0.006	1.043	0.137	0.137	1.041	0.087	0.007
	4	0.920	0.218	0.006	0.984	0.154	0.154	1.063	0.151	0.007

Table 4.4 Results of cell growth and ELISA production for the celtone production screen monitored over more time points.

From all the previous production screens, the levels of production remain fairly low, at best 1.5 mg.L^{-1}. To produce enough labelled material for a full NMR analysis of the protein, much higher levels of expression would be required. Due to the high cost of the ^{13}C, ^{15}N dual labelled celtone media (US$ 4500 L^{-1}, Martek Biosciences), it would be required to get the levels of production to about 15 mg in 1 litre of celtone media - these levels are quite a distance away. There may be two reasons for the relatively poor production of Fv: firstly, it is often the case that there is a pH imbalance in such media - as amino acids are used in the production of protein the pH of the reaction will rise which may cause problems. In the previous screen using the celtone media many of the cultures had a pH of 9 or 10. Secondly cultures have been started from plates with different colony types which have been started from previously grown plates, but cultures normally show better growth when inoculated from freshly transformed cells. Therefore a method was needed to maintain the pH balance of the cultures to find if this helped in increasing the expression levels of the recombinant protein in any way. This could be done by buffering the celtone with phosphate buffer, or by spiking M9P media with a certain amount of celtone.

Type I production screen in celtone derived media.

In table 4.5 the reactions are referred to as either P, C or M, these refer to the media used. P is the 1:1 ratio of celtone to 0.2 M phosphate buffer, C is the celtone at the conditions which gave the best results in the previous production screen. Samples M15 to M22 are M9P with a further 10 % of celtone added, samples M23 to M30 are M9P with a further 20 % celtone.

Culture Conditions			Absorbance			pH		
No.	Temp.	Induction	P.I.	24 hr.	48 hr.	P.I.	24 hr.	48 hr.
P1	37	7 hr.	0.185	0.245	0.666	7-8	7-8	8
P2	37	7 hr.	0.086	0.070	1.245	7-8	7-8	8
P3	30	7 hr.	0.600	0.673	0.737	7-8	7-8	8
P4	30	7 hr.	0.268	0.274	1.087	7-8	7-8	8
P5	37	11 hr.	0.098	0.238	1.370	7-8	7-8	8
P6	37	11 hr.	0.188	0.216	1.112	7-8	7-8	8
P7	37	11 hr.	0.419	0.326	0.791	7-8	7-8	8
P8	37	11 hr.	0.463	0.345	0.983	7-8	7-8	8
P9	30	17 hr.	0.022	0.248	0.668	7-8	8	8
P10	30	17 hr.	0.017	0.135	0.490	7-8	8	8
P11	37	17 hr.	0.352	0.230	0.634	7-8	8	8
P12	37	17 hr.	0.393	0.247	0.723	7-8	8	8
C13	30	11 hr.	1.209	1.692	1.463	6-7	7	9
C14	30	11 hr.	1.150	1.308	1.809	6-7	7	9
M15	30	7 hr.	0.461	0.180	0.365	6	6	6
M16	30	7 hr.	0.459	0.156	0.421	6	6	6
M17	25	7 hr.	0.351	0.124	0.326	6-7	6	6
M18	25	7 hr.	0.259	0.115	0.418	6-7	6	6
M19	30	17 hr.	0.311	0.278	0.525	6-7	6	6
M20	30	17 hr.	0.347	0.520	0.441	6-7	6	6
M21	25	17 hr.	0.479	0.471	0.580	6-7	6	6
M22	25	17 hr.	0.502	0.473	0.462	6-7	6	6
M23	30	7 hr.	0.814	0.311	0.606	6	6	6
M24	30	7 hr.	0.850	0.298	0.904	6	6	6
M25	25	7 hr.	0.408	0.173	0.508	6-7	6	6
M26	25	7 hr.	0.610	0.180	0.520	6-7	6	6
M27	30	17 hr.	0.525	0.485	0.654	6-7	6	6
M28	30	17 hr.	0.518	0.577	0.655	6-7	6	6
M29	25	17 hr.	0.750	1.040	0.788	6-7	6	6
M30	25	17 hr.	0.750	0.759	0.738	6-7	6	6

Table 4.5 Plasmid stability, pH and cell culture absorbance results from a type I production screen in celtone derived media.

This table shows the temperatures and induction times corresponding to the reaction numbers along with the absorbance measurements at 600 nm, pH measurements and the ELISA results. There are no plasmid stability experiments for this production screen. The reasons for this are firstly: previously this experiment has shown poor repeatability between duplicate cultures but also in general there are no stability problems with this colony type, secondly, the process of dilutions, marking out plates, plating out and counting the large number of plates and the quality of these results did not justify the amount of effort required, especially in this experiment where 30 separate cultures are grown. This generates an enormous amount of samples which that would not be possible to process in the time-scale of the reaction.

From the absorbance results it is quite clear for some of the buffered celtone samples there is quite poor growth, although in these cases growth does improve throughout the experiment. The worst case is P9 and P10, corresponding to a culture at 30 °C with a 11 hour incubation time. The same is true for some of the celtone 'spiked' M9P cultures, and in fact none of them attain high levels of biomass. This was in some way anticipated as much less media is used in these reactions, and is why the lowest temperature (25 °C) was not used in this experiment. What is much more encouraging is the pH stability of the experiment. These results were obtained using pH paper which detects over a range of 1-14, although fairly crude, it is accurate enough to detect significant deviation from neutral pH, and that any damage to the anti-E3G Fv in this experiment is not caused by extreme acidity or alkalinity. The assay results displayed in table 4.6, give valuable information in the relative performance of the different media tried in this experiment.

Supernatants:

Culture Conditions			Absorbance after 24 hours			Absorbance after 48hours		
No.	Temp.	Induction	1:2	1:4	1:8	1:2	1:4	1:8
P3	30	7 hr.	0.809	0.727	0.699	0.729	0.466	0.202
P4	30	7 hr.	0.782	0.663	0.316	0.453	0.107	0.027
P7	30	11 hr.	0.826	0.676	0.401	0.457	0.151	0.065
P8	30	11 hr.	0.738	0.678	0.354	0.500	0.199	0.064
C13	30	11 hr.	0.929	0.833	0.631	0.907	0.449	0.095
C14	30	11 hr.	1.054	0.864	0.554	0.930	0.598	0.207
M17	25	7 hr.	0.265	0.165	0.072	0.036	0.019	0.008
M18	25	7 hr.	0.175	0.132	0.057	0.013	0.012	0.007
M25	25	7 hr.	0.448	0.214	0.076	0.017	0.008	0.005
M26	25	7 hr.	0.224	0.105	0.031	0.042	0.014	0.008
M29	25	17 hr.	0.103	0.031	0.017	0.495	0.062	0.014
M30	25	17 hr.	0.362	0.078	0.028	0.626	0.081	0.019

Lysates:

Culture Conditions			Absorbance after 24 hours			Absorbance after 48hours		
No.	Temp.	Induction	1:2	1:4	1:8	1:2	1:4	1:8
P3	30	7 hr.	0.352	0.257	0.138	0.015	0.009	0.008
P4	30	7 hr.	0.025	0.015	0.009	0.007	0.006	0.005
P7	30	11 hr.	0.086	0.028	0.011	0.020	0.010	0.006
P8	30	11 hr.	0.042	0.022	0.011	0.077	0.012	0.008
C13	30	11 hr.	0.234	0.163	0.044	0.003	0.005	0.004
C14	30	11 hr.	0.007	0.010	0.009	0.005	0.004	0.004
M17	25	7 hr.	0.025	0.012	0.009	0.020	0.006	0.006
M18	25	7 hr.	0.018	0.010	0.010	0.007	0.007	0.009
M25	25	7 hr.	0.030	0.012	0.010	0.007	0.005	0.006
M26	25	7 hr.	0.039	0.020	0.011	0.006	0.006	0.006
M29	25	17 hr.	0.017	0.008	0.007	0.045	0.006	0.004
M30	25	17 hr.	0.013	0.011	0.007	0.055	0.014	0.004

Table 4.6 ELISA results from cell supernatants and periplasmic lysates of cell cultures from a type I production screen in celtone derived media for detection of Fv E3G myc.

When comparing the results it is important to remember the relative amounts of celtone used. In general the performance of the spiked M9P media was poor, the buffered celtone reactions particularly samples numbers P7 and P8 (30 °C, 11 hours induction) and P11 and P12 (30 °C, 17 hours induction) seem to give very similar results to the best conditions for the pure celtone media. The low levels of expression for most of the trial media can be explained by the low levels of growth in these samples prior to induction.

These results may not seem particularly encouraging, but as the point of these media was to decrease the inactivation of these proteins, these were to have the property of reducing the inactivation of the proteins due to high temperatures at extremes of pH. If this was a problem in the expression of the recombinant protein, and the perceived lack of detected protein was now due to proteolysis of the myc peptide tail, then the total anti-E3G Fv activity assay would show higher levels of protein than the cultures in the pure celtone media. The assay indicated levels of about 4 $mg.mL^{-1}$ of Fv in the best samples of the phosphate buffered celtone media. If this experiment was to be done with a litre of labelled celtone a yield of 8 mg of labelled protein would be the maximum achievable, which may be enough for a full structural analysis by NMR.

Purification of anti-E3G antibody fragment.

The sepharose column with the epoxy linked estrone was used in purifying Fv expressed in the fermentation experiment. Protein is detected by an absorbance measurement at 280 nm in a flow cell, after the sample has passed through the column. When the supernatant is loaded

in the presence of 1 M NaCl a high absorbance reading is observed due to the large amount of unbound protein flowing through the cell. On washing with PBSA the non-binding protein is removed from the column and the absorbance reduces to background, after further washes with 3M NaCl to remove proteins bound non-specifically, and again with PBSA the protein is ready for elution. The protein is eluted with estradiol glucuronide and the eluant was diverted through a column with ion-exchange resin, to separate the excess steroid compound from the protein. The protein can be quantified from absorbance measurements at 280 nm. Only two amino acids absorb radiation at this wavelength: tryptophan and tyrosine. The molar extinction coefficients for these residues at 280 nm are 5550 and 1340 respectively. For this protein their are three tryptophan residues and 14 tyrosine residues therefore the molar extinction coefficient (ε) can be calculated thus:

$$\varepsilon = \frac{(\text{no. of W} \times 5550) + (\text{no. of Y} \times 1340)}{\text{molecular weight of Fv}}$$

This gives a value of 1.316 for the Fv fragment.

Fermentation of Type II Colonies in M9P + Y.E.

At the Colworth laboratory it is usual procedure to take the optimised growing conditions in the M9P media into the fermentation stage. This work was carried out by Mark Griffiths. In the fermentation reaction there was a problem with obtaining culture growth. The different samples were not made available to myself so I could not perform a quantitative assay on them. Instead the protein was purified from which an estimated 1.5

mg was recovered.

Expression of ^{13}C, ^{15}N dual labelled protein.

This experiment was carried out by Sandra Hemmington, and around 5 mg of dual labelled active protein was purified from 1 L of dual labelled celtone media.

4.4 Conclusions

The plasmid created in Chapter 2, proved in the first instance to be present in two colony types both of which were found to be stable over the length of time required for the expression experiments to be carried out. However one of these colony types proved to be useless in the production of the Fv fragment. From the M9P media experiments fairly low levels of expression of the recombinant protein even on optimisation (~ 1.0 $mg.L^{-1}$). This approach would be of no use in producing large quantities of labelled protein that would be required for structural studies by NMR. Using the celtone media definite improvements were noticed in the optimised expression levels (~1.5 $mg.L^{-1}$). In all these reactions a noted decrease in detected, active protein levels after 48 hours was noted, which could only be explained by lost activity in the protein. An assay which gave an indication of the total amount of active protein which did not rely on the presence of the peptide myc tail would go a long way in explaining this problem. These detected levels were still low which suggested denaturation of the protein due to the extended time which it spent in the environment of the shake flask experiments. When the celtone media was buffered to try and negate the rises in pH observed in earlier experiments, the levels of active protein detected did rise after the

the initial test at 24 hours. There was a detected stability in pH which gave rise to increasing detection of active protein after 24 hours and therefore better expression levels per litre of celtone. Not enough protein was obtained, dual labelled to use as a sample for the full elucidation of the solution structure by NMR, but significant levels of unlabelled protein (~18 mg) was obtained that would prove adequate for a series of NMR experiments based on the finding the bound conformation of E3G.

Chapter 5.

Structure of Estrone-3-glucuronide in Free Solution.

Abstract

Using heteronuclear (^{13}C, ^{1}H) NMR methods, the three dimensional structure and dynamics of E3G has been probed in free solution. The glycan is found to exist in multiple conformations, with particularly large fluctuations about the glycosidic linkage ψ. The use of a ^{13}C labelled analogue, estrone-3-[U-^{13}C]-glucuronide was utilised to obtain the ^{13}C assignments and subsequent proton assignments for the glucuronic acid moiety *via* a ^{13}C-^{13}C COSY. These could not be obtained from the natural abundance ^{13}C compounds due to resonance overlap.

5.1 Introduction.

To help understand the interaction of a glycan with is protein ligand, knowledge of the behaviour of the molecule in solution is vital. A major difficulty in the conformational analysis of glycans such as E3G in solution is the small number of good quality distance constraints available across the glycosidic linkage. It is very common that only one or two NOE or ROE constraints are available, and because these are the primary source of conformational variance, the overall three-dimensional structure may be seriously underdefined by NOEs or ROEs alone. Additional constraints can be obtained by long range heteronuclear (^{13}C-^{1}H) coupling constants across the glycosidic linkage (Hamer *et al.*, 1978; Tvaroska *et al.*, 1989; Poppe and van Halbeek, 1991a,b), which are possible in this study due to ^{13}C labelling and the relatively large amount of material.

In solution several conformers exist in equilibrium, but the experimental

studies in solution give a time averaged conformation. Molecular dynamics simulations can be used in an attempt to describe the motion of the glycan in solution using the NMR restraints. Problems can arise, for example, where a molecule exists in two distinct conformations, and each of these conformations are characterised by a distinct set of NOEs. If the rate of conversion between these conformers is rapid with comparison to the NMR timescale, then the NOEs from each conformation will be observed simultaneously. In attempting to generate a single structure from this data, there may be no single conformation which satisfies this averaged data and the resultant structure may be highly strained and physically unrealistic. It is possible their could be many more than two solution conformations, in which case the calculated structure may be even further from the true situation. Using a method suggested by Torda and co-workers (1989) and demonstrated by the same group (1990), instead of the normal energy penalty imposed if the two nuclei, described by the constraint, move outwith the allowed range, this penalty is only applied when this distance, averaged over a period of time, is out with the constraint - time-averaged restraints. The length of time over which the distance is averaged, has to be an order of magnitude smaller than the length of the dynamics simulation.

This system has been applied in calculating the solution structure and dynamics of estrone-3-glucuronide.

5.2 Materials and Methods

NMR Measurements.

Homonuclear ^{13}C-^{13}C COSY data were acquired with spectral widths

of 4800 Hz in each dimension, with 2048 complex points and 256 complex points in the t2 and t1 dimensions respectively. Sixteen transients were acquired per t1 increment, resulting in a total acquisition time of ~2 hours.

Homonuclear ^{1}H-^{1}H COSY data were acquired with spectral widths of 4000 Hz in each dimension, with 4096 complex points and 256 complex points in the t2 and t1 dimensions respectively. Sixteen transients were acquired per t1 increment, resulting in a total acquisition time of ~2 hours.

Homonuclear ^{1}H-^{1}H ROESY data were acquired with spectral widths of 4000 Hz in each dimension, with 4096 complex points and 256 complex points in the t2 and t1 dimensions respectively. Sixteen transients were acquired per t1 increment, resulting in a total acquisition time of ~2 hours.

HSQC data were acquired with spectral widths of 3000 Hz and 10000 Hz and with 1024 complex points and 128 complex points in the proton and carbon dimensions respectively. Sixteen transients were acquired per t1 increment, and the data were adopised with cosine-bell weighting function, followed by zero-filling once in each dimension prior to Fourier transformation. In HSQC experiments on estrone-3-U-^{13}C-glucuronide in free solution, the total solute concentration was ~10 mM in 100 mM phosphate buffer (pH 7.0) prepared in D_2O (99.96%).

Estrone-3-glucuronide was purchased from the Sigma Chemical Company, the estrone-3-[U-^{13}C]-glucuronide was synthesised for me by Dr. Mark Probert.

Molecular Modelling and Dynamics.

For the unrestrained and restrained models the molecular modelling package DISCOVER was used (Biosym Technologies Inc.), with a molecular mechanical forcefield suitable for carbohydrates (Homans, 1990). Modifications were required to include several atom types in the steroid part of this molecule, these were taken directly from the AMBER forcefield.

Ten pseudo-random geometries were generated by dynamical quenching: The initial minimised structure was subjected to unrestrained dynamics at 750 K for 1000 ps, during which the torsional term ϕ was scaled by a factor of 7 to prevent distorted ring geometries, with a structure written after every 100 ps. Simulated annealing was carried out for each model in the following manner: The structure was minimised with a steepest decent algorithm until the maximum derivative was less than 1 kcal.Å^{-2}, with ϕ scaled by 5; then equilibrated with a thermal path by running dynamics successively for 1 ps at 500 K, 450 K, 400 K, 350 K, 300 K, and then successively for 1 ps in decreasing steps of 10 K. After 1 ps at 10 K a further 1 ps at 5 K was followed by minimisation using a steepest decent algorithm until the maximum derivative was less than 0.1 kcal Å^{-2}. The molecular Dynamics was run with a dielectric constant of 80, for 510,000 steps of 1 fs (510 ps) with the last 500 ps used for further analysis.

A gridsearch about ϕ and ψ was calculated by varying each angle in 15° steps and minimising for each of the 24^2 geometries created until the maximum derivative was less than 0.1 kcal.Å. The energy of each

geometry was displayed in an iso-energy ϕ,ψ plot.

XPLOR minimisations and dynamics were calculated using the the same modified AMBER forcefield with added terms for some of the steroid atoms as used in the DISCOVER calculations. Minimisations were calculated without restraints and with restraints taken from the ROESY spectrum as a comparison with DISCOVER. A series of XPLOR minimisations and dynamics were run with time averaged restraints as described by Torda and co-workers (1990). Typically these were run with a τ of 10 ps, and non-conservative energy, with flat-well constraints.

NOE Simulations.

All rotating frame Overhauser effects (ROE) simulations were computed with the program MDNOE (written by S. W. Homans.), which is a package for general full relaxation matrix simulations of NOEs, ROEs and transferred NOEs for an arbitrary homonuclear or heteronuclear spin system. Time averaging of internuclear distances is included using formalisms appropriate for motions which are either slow or fast with respect to overall molecular tumbling (Tropp, 1980)(Homans and Forster, 1992).

5.3 Results and Discussion

Modelling studies of E3G without NMR restraints.

A model of E3G was built in DISCOVER and was minimised as described. The minimised model gave values for the torsional angles ϕ (H1-C1-O1-C3') and ψ (C1-O1-C3'-C4') as 20° and 87° respectively. This structure was

subsequent simulated annealing. From the simulated annealed structures two families of structures were obtained, these can be seen in figure 5.1. The minimum energy structures of each of the families correspond to values of ϕ and ψ about the glycosidic linkage of +18 °, -84° and +28°, +92°, and each of these structures were used as input of independent 500 ps molecular dynamics simulations *in vacuo*. When these simulations were plotted (figure 5.2) the molecule spends most of its time in the two minima suggested by the simulated annealing, this matches the minimum energy wells given by the iso-energy plot also displayed in figure 5.2.

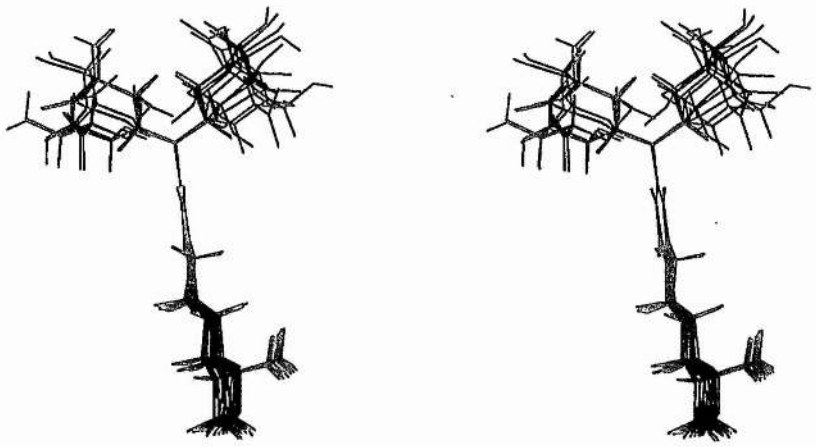


Figure 5.1 Results of unrestrained dynamical simulated annealing on ten pseudo-random structures of estrone-3-glucuronide showing the two families of structures.

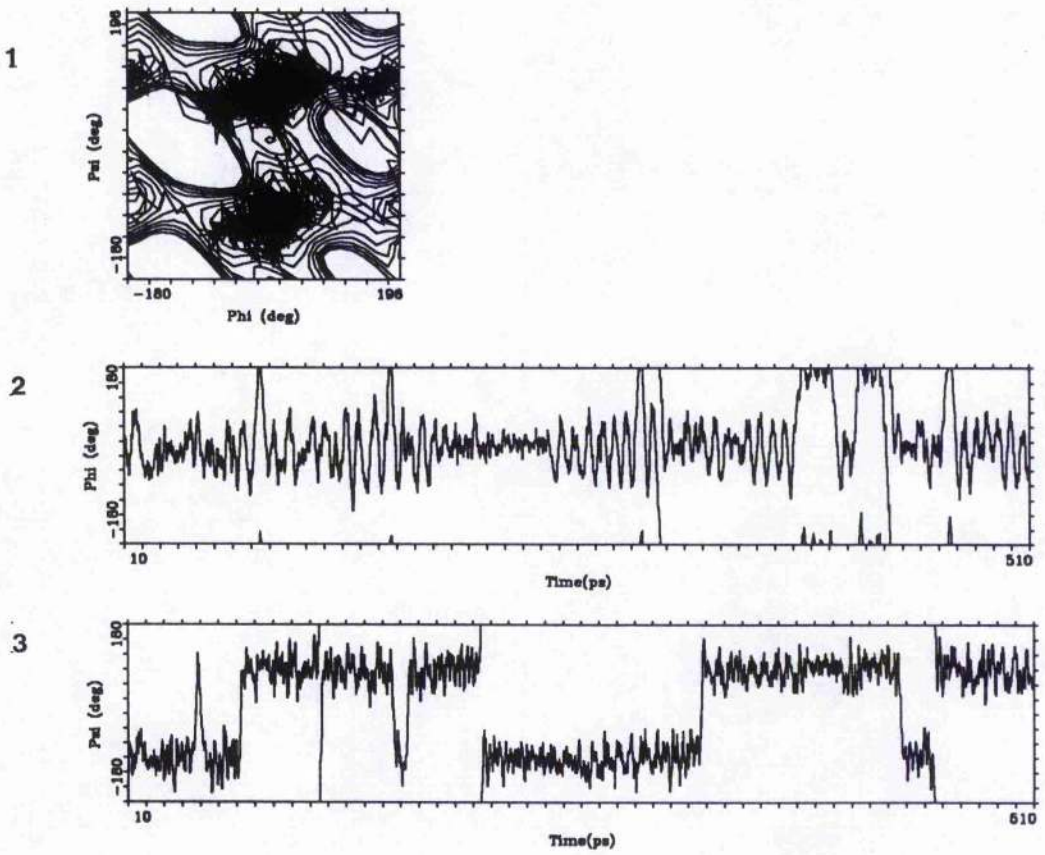


Figure 5.2 Minimum energy structure of E3G subjected to 510 ps of unrestrained dynamics.

1. Plot of ϕ and ψ during dynamics run, superimposed on potential energy surface for ϕ and ψ.

2. Plot of angle ϕ over the timecourse of the dynamics.

3. Plot of angle ψ over the timecourse of the dynamics.

Modelling studies of E3G with NMR restraints.

Initially, the solution structure and dynamics of E3G were to be determined using conventional rotating frame Overhauser effects measurements in combination with restrained dynamical simulated annealing and molecular dynamics simulations (Homans and Forster, 1992).

Proton resonance assignments for the glycan, a prerequisite for these studies, could not however be obtained from conventional ^{1}H-^{1}H correlation methods due to extreme overlap of resonances corresponding to the glucuronic acid moiety (figure 5.3). Inspection of the ^{13}C NMR spectrum of the glycan showed that all the of the carbon resonances were well resolved. The ^{13}C spectrum of the glucuronic acid moiety was therefore assigned by use of a ^{13}C-^{13}C COSY experiment on estrone-3-[U-^{13}C]-glucuronide, from which proton resonance assignments were derived by conventional ^{13}C-^{1}H HSQC methods (figures 5.4 and 5.5 respectively). These assignments were then utilised in the interpretation of a ^{1}H-^{1}H ROESY measurements on E3G (natural abundance).

As can be seen in figure 5.6 two inter-residue ROEs were observable, from the C-1 proton of the glucuronic acid moiety to the C-2′ and C-4′ protons of the estrone residue, together with several intra-residue ROEs. The inter-residue ROEs were quantified together with the intra residue ROE between the C-1 and C-5 protons of the glucuronic acid moiety by measurement of crosspeak volumes, and these are given in table 5.1.

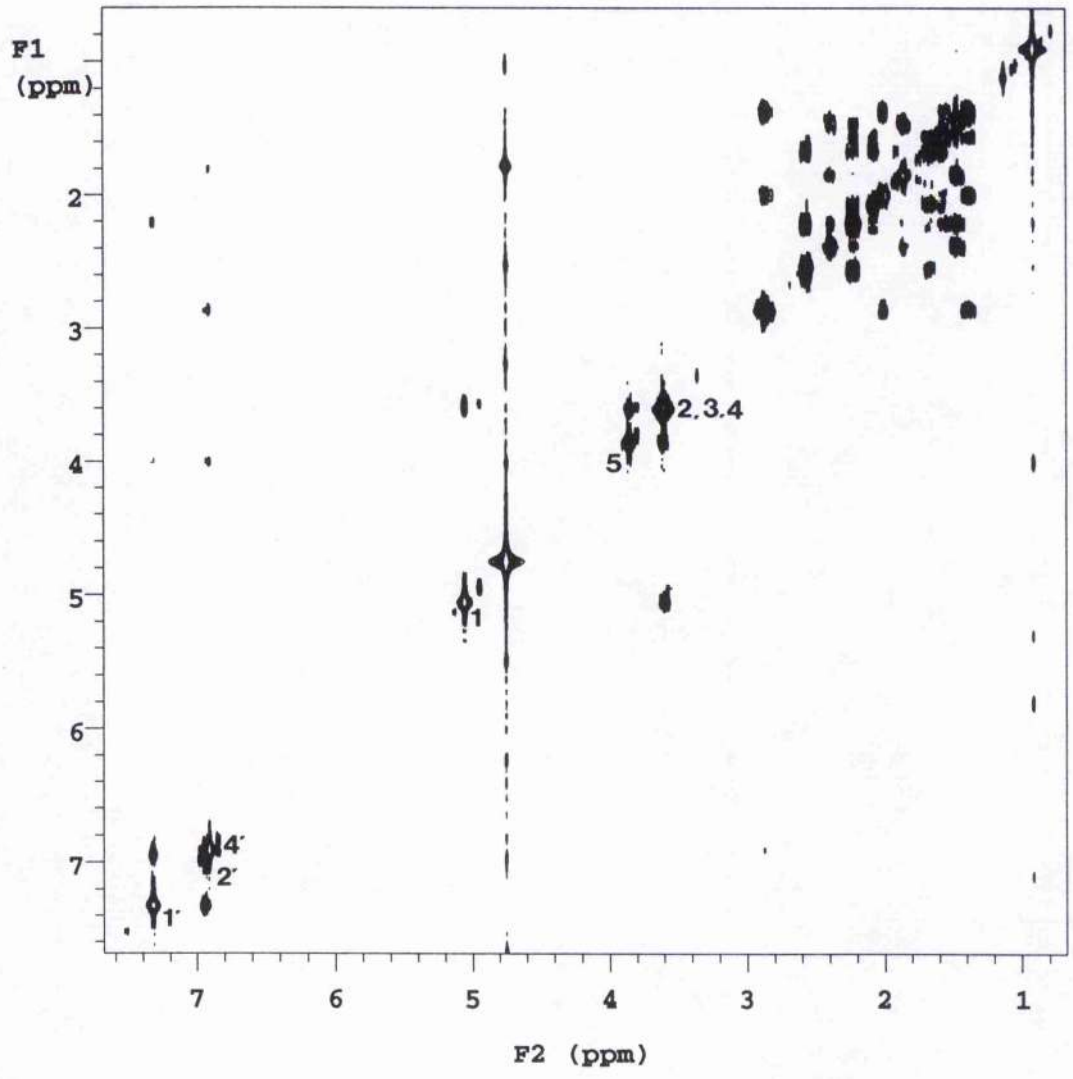


Figure 5.3 (A) ^{1}H-^{1}H COSY spectrum of E3G in D_2O.

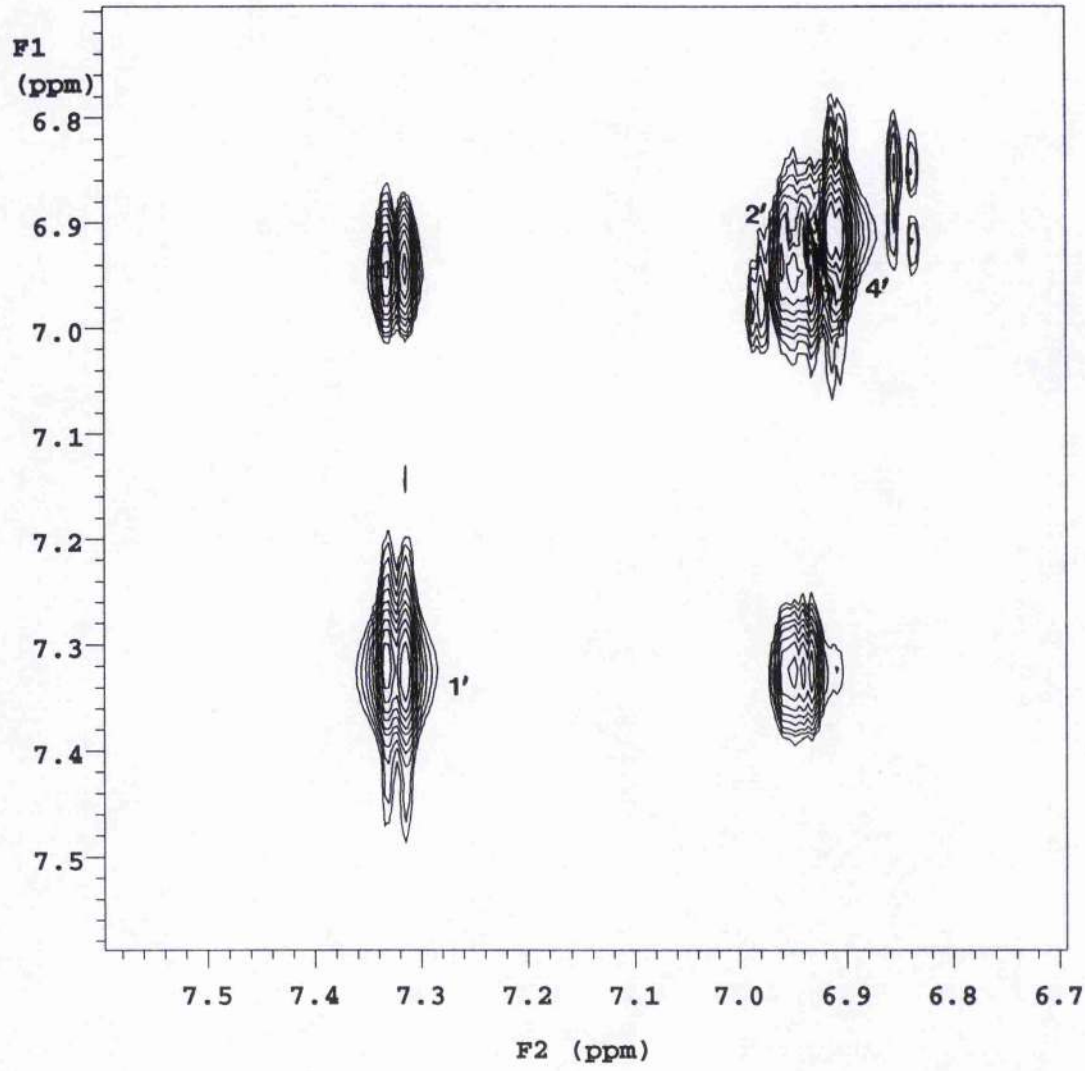


Figure 5.3 (B) ^{1}H-^{1}H COSY spectrum of E3G in D_2O. (aromatic region). Showing proton assignments for the aromatic region of the steroid.

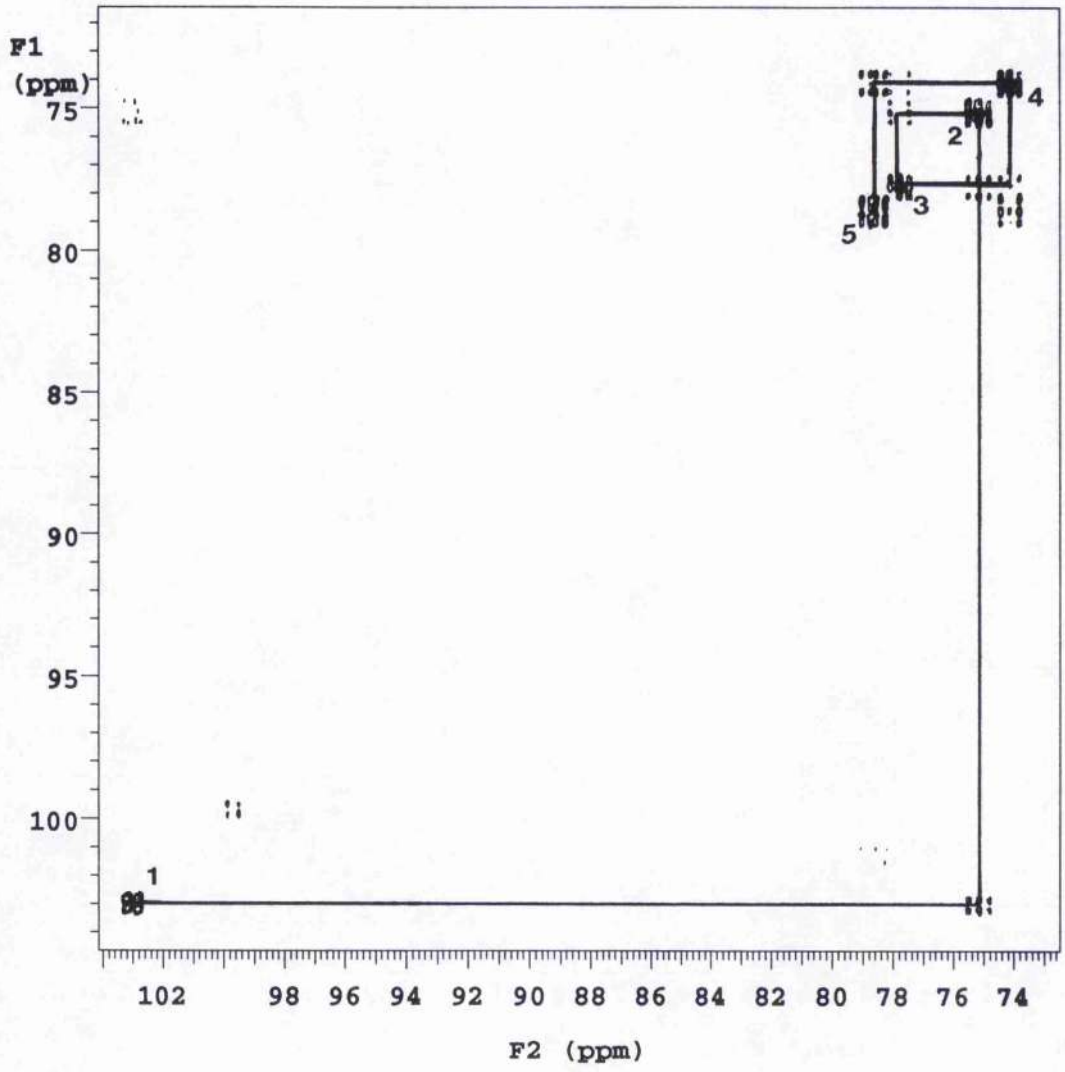


Figure 5.4 ^{13}C-^{13}C COSY of estrone-3-[^{13}C-U]-glucuronide in D_2O. Showing ^{13}C assignments for the glucuronic acid moiety.

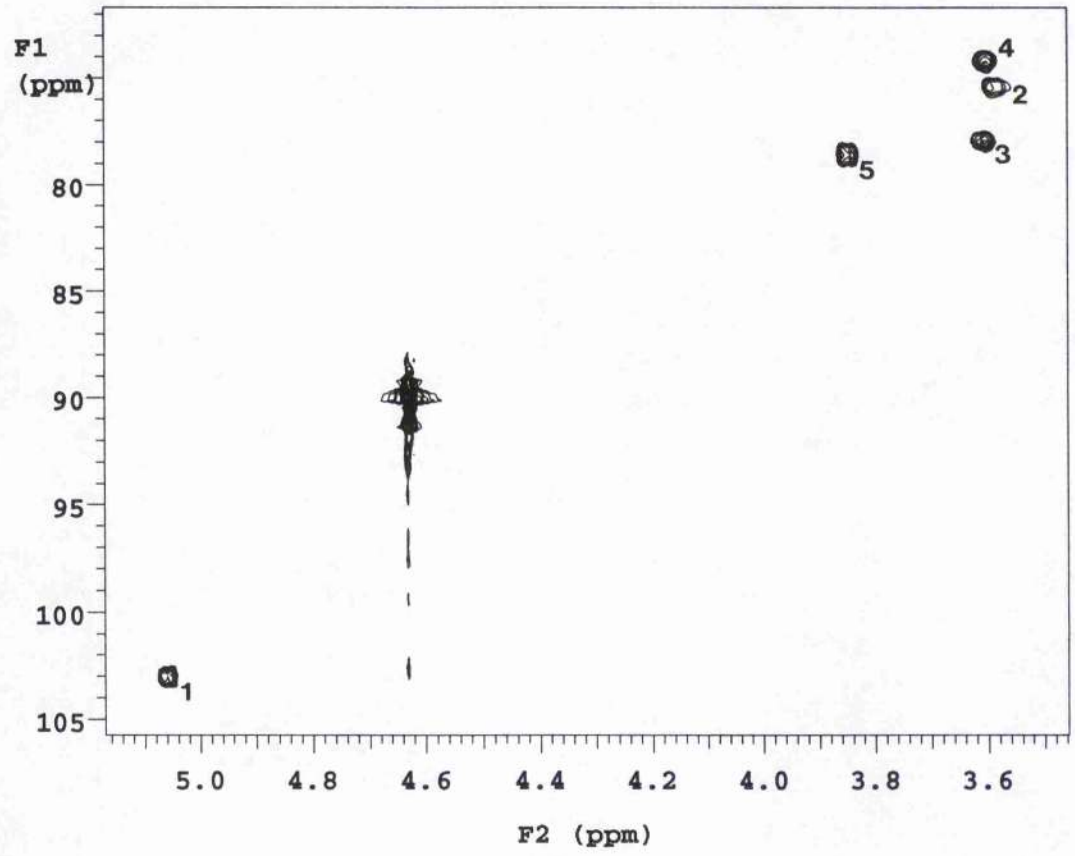


Figure 5.5 ^{1}H-^{13}C HSQC of estrone-3-[^{13}C-U]-glucuronide in D_2O, ^{13}C and proton assignments for the glucuronic acid moiety.

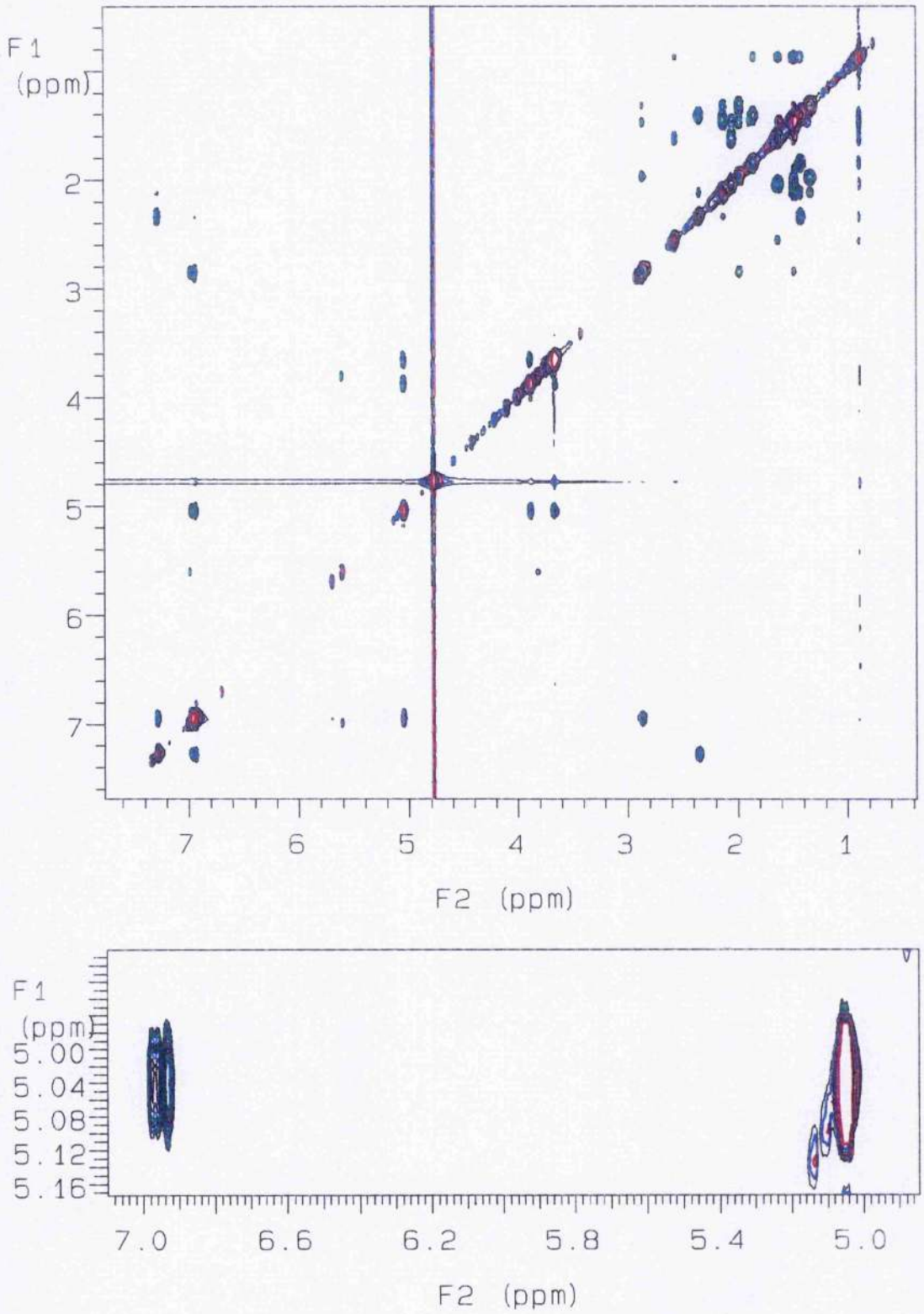


Figure 5.6 ^{1}H-^{1}H ROESY spectrum of E3G in D_2O, showing the two inter-residue ROEs. (Also detail of section of spectra in the boxed area).

A restrained dynamical simulated annealing and molecular dynamics simulation strategy was employed with the inter-residue constraints obtained from the ROESY spectrum (Homans and Forster, 1992; Rutherford and Homans, 1994). The torsional fluctuations around ϕ and ψ are shown in figure 5.7, it is seen essentially that the glycan exhibits limited torsional fluctuations about the glycosidic linkage. In order to assess the validity of this dynamic behaviour, the ROESY connectivities measured experimentally were compared with those predicted from the MD simulations. The latter were determined by performing a full-relaxation matrix analysis calculation (Forster, 1991) using internuclear distances averaged over the timecourse of the MD simulations. The correlation time for overall tumbling of the molecule was determined by varying this parameter during the calculation until the predicted ratio of the diagonal peak volume for the C1 proton resonance glucuronic acid moiety and the crosspeak between the C-1 and C-5 protons resonances equalled that determined experimentally. The absolute theoretical crosspeak intensity was then used to convert the experimental intensities to absolute ROE values. The theoretical value from the restrained dynamics simulations predict much smaller ROEs than those observed, from which it is clear that the predicted dynamical behaviour is not consistent with the experimentally observed ROEs. This can be explained by a simple model building study, which indicated that no single conformation about the glycosidic linkage was consistent with the inter-residue ROEs of the observed intensity, suggesting considerable motional averaging about this linkage.

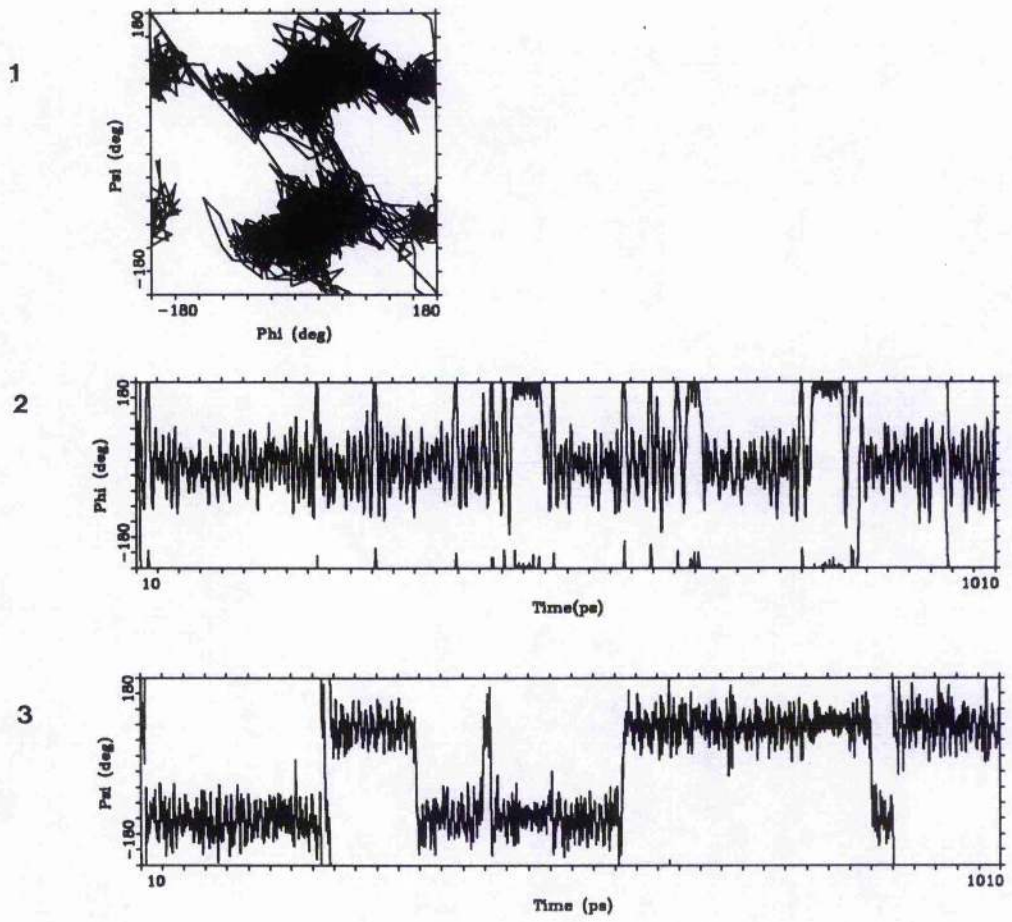


Figure 5.7 Restrained dynamics simulations on minimum energy structure from simulated dynamical annealing.

1. Plot of ϕ and ψ during dynamics run.
2. Plot of angle ϕ over the timecourse of the dynamics.
3. Plot of angle ψ over the timecourse of the dynamics.

In an attempt to quantify the extent of this motion, a restrained dynamical simulated annealing and molecular dynamics simulation strategy was employed essentially as described, but using time-dependent distance restraints in a manner recently described by Torda *et al.* (1990). Two families of structures from the simulated annealing procedure, corresponding to values of ϕ and ψ about the glycosidic linkage of +18 °, -84° and +28°, +92°, and each of these structures were used as input of independent 500 ps molecular dynamics simulations *in vacuo*. The torsional fluctuations about ϕ and ψ over the timecourse of these simulations are shown in figure 5.8. It is seen essentially that the glycan exhibits rapid torsional fluctuations about the glycosidic linkage. The fluctuations about ψ are particularly large, extending over the full 360 °, i.e. complete freedom of motion. In order to assess the validity of this dynamic behaviour, the ROESY connectivities measured experimentally were again compared with those predicted from the MD simulations, performed as before. A comparison between the experimental and theoretical value is included in table 5.1, from which it is clear that the predicted dynamical behaviour is consistent with the experimentally observed ROEs. In order to validate further the predicted dynamic behaviour of the glycan, the magnitudes of the trans-glycosidic three bond heteronuclear coupling constant ($^3J_{H1\text{-}C3'}$) and homonuclear ^{13}C-^{13}C coupling constants ($^3J_{C2\text{-}C3'}$, $^3J_{C1\text{-}C2'}$, $^3J_{C1\text{-}C4'}$) were determined from the ^{13}C NMR spectra of E3G and estrone-3-U-^{13}C-glucuronide respectively, are shown in table 5.1. Quantitatively, the value of the $^3J_{H1\text{-}C3'}$ = 3.4 Hz predicted from the MD simulation using the appropriate Karplus equation is consistent with the experimental value of 3.7±0.4 Hz.

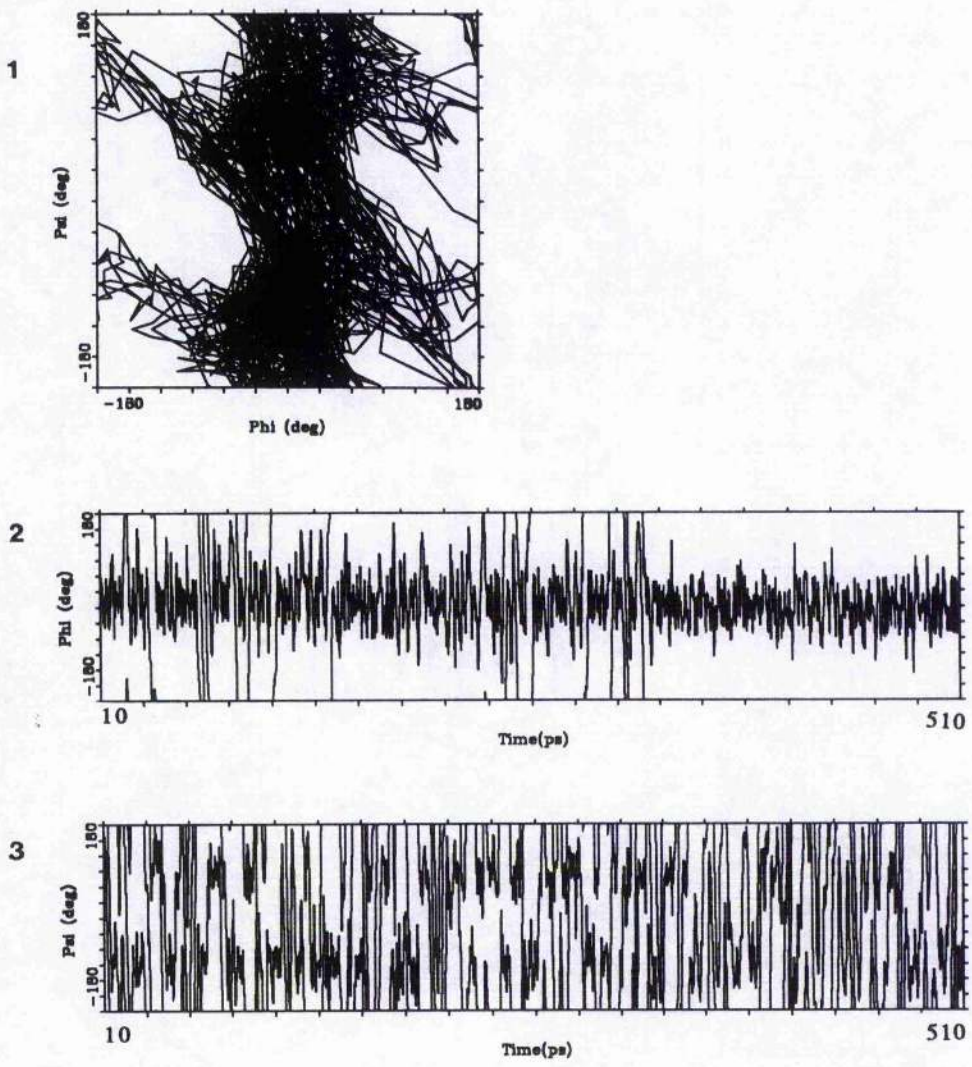


Figure 5.8 (A) XPLOR time-averaged restrained dynamics simulations on minimum energy structure from 1st family of structures.

1. Plot of ϕ and ψ during dynamics run.
2. Plot of angle ϕ over the timecourse of the dynamics.
3. Plot of angle ψ over the timecourse of the dynamics.

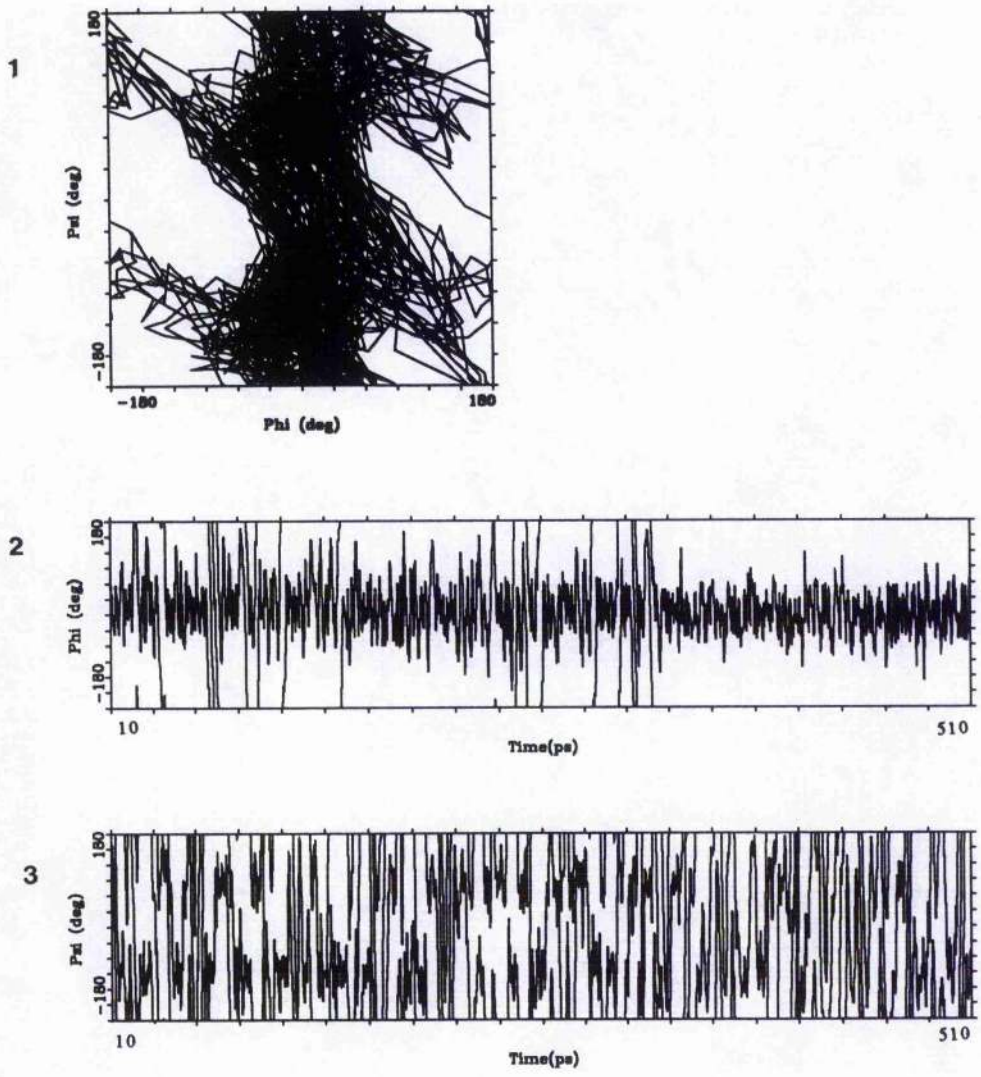


Figure 5.8 (B) XPLOR time-averaged restrained dynamics simulations on minimum energy structure from 2nd family of structures.

1. Plot of ϕ and ψ during dynamics run.
2. Plot of angle ϕ over the timecourse of the dynamics.
3. Plot of angle ψ over the timecourse of the dynamics.

Atom pair/ Dihedral angle	Crosspeak Intensity	Absolute ROE (%)[1] Exptl.	Theor.	^{3}J (Hz) Exptl.	Theor.
GlcU-H1 diagonal	19.4	-	-		
GlcU-H1-GlcU-H5	-2.5	7.2	7.2[2]		
GlcU-H1-Est-H2′	-1.6	4.5	3.9		
GlcU-H1-Est-H4′	-1.7	4.8	3.9		
H1-C1-O1-C3′				3.7[3]	3.4
C2-H1-O1-C2′				2.9	
C1-O1-C3′-C2′				2.6	
C1-O1-C3′-C4′				2.3	

1 calculated with τ_c = 0.195 ns

2 estimated error in experimental ROE measurements = ± 20 %.

3 estimated error in experimental ^{3}J values = ± 0.4 Hz.

Table 5.1 Experimental versus predicted ROEs (τ_m = 500 ms) and coupling constants determined for estrone-3-glucuronide and estrone-3-[^{13}C-U]-glucuronide in free solution.

The homonuclear ^{13}C-^{13}C coupling constants could not be quantified exactly in conformational terms due to the absence of a suitable Karplus parametrisation for couplings of this type. However, $^{3}J_{C2\text{-}C3'}$ is large, consistent with a conformation for the dihedral C2-C1-O-C3' which is approximately trans (150°) for an average value of ϕ = 30° over the timecourse of the simulation. The couplings $^{3}J_{C1\text{-}C2'}$ and $^{3}J_{C1\text{-}C4'}$ are also large, and are equal with experimental error. Couplings of this size could not arise if the glycan existed predominantly in each of the two minimum energy conformations predicted from the simulated annealing calculations, since the value of ψ in each of these minima (90° and -90°) would give rise to small (<1.2 Hz) couplings (Vuister and Bax, 1993). Rather the molecule must exist in solution for significant periods of time in a conformation where C1 and C2' and C1 and C4' are trans or eclipsed. Taken together, these data are consistent with with a model involving fast internal motion which is restricted about ϕ but essentially unrestricted about ψ.

5.3 Conclusions

The use of the pseudo-random structures and simulated annealing has given two favourable conformations for E3G in solution. In trying to describe the dynamical motion of the glycan in solution with the use of experimental ROEs, initially proved unsuccessful with restrained molecular dynamics. The magnitude of the two ROEs cannot be due to the molecule existing predominantly in these two favourable conformers and that there most be motional averaging. The use of a time averaged restraints in the MD simulation, and the output of this simulation

subjected to a full relaxation matrix calculation, gave theoretically obtained values for cross peak volumes and coupling constants which matched well with those obtained experimentally. The $^{3}J_{CH}$ coupling calculated from this simulation also compared favourably with experimental data, and the $^{3}J_{CC}$ couplings, although not quantifiable with a suitable Karplus parameterisation, measured experimentally agreed qualitatively with the predicted solution behaviour.

The high degree of flexibility around the torsional angles, particularly the full 360° rotation around ψ, and the close contacts between C1 with C2′ and C4′ arising from this rotation explain the strong ROEs observed for both couplings.

Chapter 6.

Structure of Estrone-3-glucuronide in Complex with Fv E3G myc.

Abstract

By use of heteronuclear (^{13}C, ^{1}H) NMR methods, the three-dimensional structure of the glycoconjugate estrone-3-glucuronide has been probed in association with an anti-E3G antibody Fv fragment. The antibody selects a conformation from the solution repertoire which does not correspond to either of the two lowest energy conformations of the free glycan, and the internal energy of the glycan in the bound state is estimated to be at most ~ 15 kcal.mol^{-1} higher than the global minimum energy conformation. The glucuronide undergoes a stacking interaction with an aromatic ring in the binding site, and both ring current shifts and nuclear Overhauser effects computed from the predicted conformation are in good agreement with preliminary X-ray data on a related complex.

6.1 Introduction

A prerequisite for a detailed understanding of the molecular basis of the interaction of a carbohydrate with its protein receptor is a high resolution three dimensional structure of the complex. Several important crystallographic studies have been described recently which illustrate in detail the precise nature of certain carbohydrate-protein interactions in the solid state (reviewed by Cambillau, 1995). In contrast, high resolution structural studies of glycan protein interactions in solution using nuclear magnetic resonance have been conspicuous by their absence. Solution studies have important potential advantages since a comparison of the solution structure of the free ligand with that of its bound state is more meaningful, and moreover the dynamics of the system are accessible.

In cases where the affinity of a ligand protein receptor is low and the 'off'

rate of the ligand is within the appropriate timescale, the bound state conformation is accessible in principle by measurement of nuclear Overhauser effects in the exchanging system. While these measurements commonly known as transferred nuclear Overhauser effects (TRNOEs) (Bothner-By and Gassend, 1972; Albrand *et al.*, 1979; Clore and Gronenborn, 1982,1983; Campbell and Sykes, 1993; Ni, 1994), have been applied to a variety of glycan-protein interactions (Glaudemans *et al.*, 1990; Bevilacqua *et al.*, 1990; Bundle *et al.*, 1994; Weimar and Peters, 1994; Andrews *et al.* 1995; Asensio *et al.* 1995; Scheffler *et al.*, 1995), they cannot however be replied upon to provide any information on the nature of the interaction of the ligand with the protein, since ligand-protein TRNOEs are observable with difficulty (Arepalli *et al.*, 1995) or not at all. Paradoxically, TRNOEs can only be quantified accurately with knowledge of the architecture of the protein binding site, since the presence of spin-diffusion in the molecular complex requires a multi-spin full relaxation matrix analysis (London *et al.*, 1992). Exclusion of nuclear spins derived from the protein can lead to erroneous conclusions regarding the bound state conformation of the ligand.(Glaudemans *et al.*, 1990; Arepalli *et al.*, 1995)

In cases where the affinity of the glycan for the receptor is high, the bound state conformation of the ligand cannot in general be determined from TRNOE measurements, since these become vanishingly small as the rate of exchange becomes very slow. However in this exchange regime it is possible in theory to obtain substantial information on the nature of the ligand-protein interaction using conventional NOE measurements. Resonances derived from the ligand can be distinguished from those of the protein by isotope enrichment of either the protein or the ligand with

a stable isotope, together with application of isotope-editing (Weber *et al,* 1991)or isotope-filtration (Petros *et al.,* 1992) techniques respectively. To my knowledge, the value of this approach has not been previously examined in high resolution structural studies of carbohydrate-protein interactions. In the present work, we examine the practical value of isotope-edited NMR techniques in probing the bound state conformation of the glycoconjugate estrone-3-glucuronide (E3G) in association with an anti-E3G antibody Fv fragment. Furthermore, we access the the extent to which the architecture of the binding site of the protein can be predicted with these methods in the absence of high resolution structural data on the protein moiety.

6.2 Materials and Methods

NMR Measurements.

Two-dimensional HCCH-COSY experiments were acquired using the three-dimensional pulse scheme described by Bax *et. al.,* (1990) keeping the first (proton) acquisition time constant. Data were acquired with spectral widths of 3000 Hz and 7000 Hz and with 1024 and 64 complex points in the ^{1}H and ^{13}C dimensions, respectively. A total of 256 increments were acquired per t1 increment, resulting in a total acquisition time of ~ 15 hours. The free procession delays in the sequence were as follows: τ = 1.5 ms, Δ = 2,15 ms, $\delta 2 = \delta 1$ = 1.1 ms. Heteronuclear ^{1}H - ^{13}C single quantum correlation.

Heteronuclear ^{1}H-^{13}C single quantum correlation (HSQC) data were acquired with spectral widths of 3000 Hz and 10000 Hz and with 1024 complex points and 128 complex points in the proton and carbon

dimensions respectively. Sixteen transients were acquired per t1 increment, and the data were adopised with cosine-bell weighting function, followed by zero-filling once in each dimension prior to Fourier transformation. In HSQC experiments and NOESY-HSQC experiments on the complex of estrone-3-[U-^{13}C]-glucuronide with anti-E3G Fv were performed on a 1:1 mixture at a concentration of ~1 mM on estrone-3-[U-13C]-glucuronide in 100 mM phosphate buffer (pH 7.0) prepared in D_2O (99.96%).

Molecular Modelling and Dynamics.

For the unrestrained and restrained models the molecular modelling package DISCOVER was used (Biosym Technologies Inc.), as described in the previous chapter.

The program MDNOE was used to calculate a full relaxation matrix of proton spins from the history or trajectory file from dynamics simulations, as described in Chapter 5.

Ring Current Calculations.

These RCCAL calculations and parameterisation described in this section were performed by Kothandaraman Seshadri.

The ring current shifts in the sugar ring of E3G bound to Fv were computed on the basis of preliminary X-ray crystal data an a related complex (S. E. V. Phillips, personal communication) using the program RCCAL. (Perkins, 1982). RCCAL only contains parameters for the 20 standard amino acids, so the parameters for glucuronic acid were introduced. Optimisation of the geometry of bound E3G with respect to

measured ring current shifts was achieved by a grid-search optimisation procedure. Grid points were generated for a stacked glucuronic acid - tyrosine aromatic ring system with a step size of 0.2 Å at seven positions in each of the mutually orthogonal directions giving $7^3 = 343$ points. The sugar ring was placed at each of these points and the ring-current shifts were computed using RCCAL. In order to arrive at the best position , a root mean square deviation of the shifts was computed relative to the experimental shifts of the five ring protons, for each of those positions. The position that had the minimum deviation from the experimental shifts was utilised as the predicted bound state conformation.

6.3 Results and Discussion.

Bound-state conformation of E3G.

Study of the bound-state conformation of E3G was addressed using the use of ^{13}C isotope-edited NOESY experiments (Weber *et al.*, 1991). This approach allows the selective observation of NOEs to and from the protons directly bonded to a ^{13}C nucleus. By use of estrone-3-[U-^{13}C]-glucuronide, the technique therefore potentially enables the direct observation of inter-residue NOEs from the C1 through C5 protons of the glucuronic acid moiety to the steroid and to amino acid residues within the binding site. A prerequisite for these studies is the need to obtain resonance assignments for the glycan in the bound state. This is not trivial since these are are large chemical shift changes on binding in both the proton and carbon spectra, together with substantial line broadening, as can be seen in the ^{13}C - ^{1}H HSQC spectrum of a 1:1 complex of estrone-3-[U-^{13}C]-glucuronide with Fv (figure 6.1).

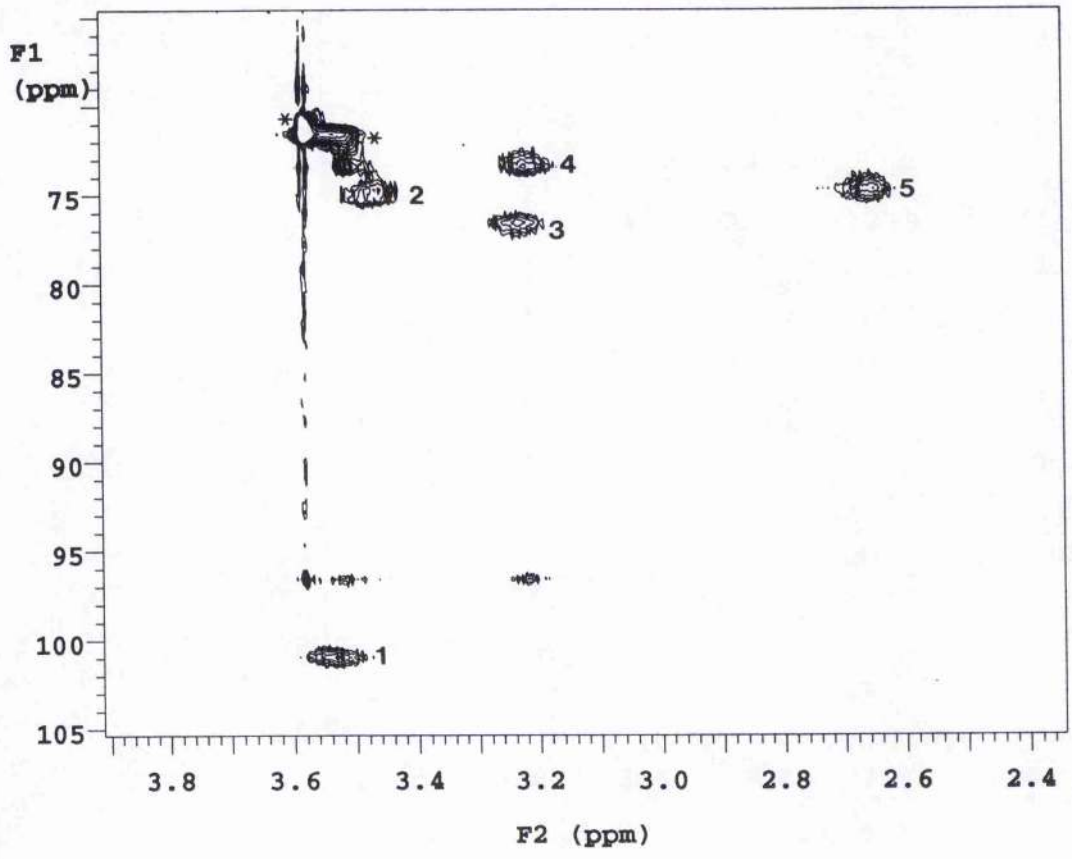


Figure 6.1 ^{1}H-^{13}C HSQC spectrum of the Fv/estrone-3-[^{13}C-U]-glucuronide complex in 100 mM phosphate buffer in D_2O, showing assignments for the glucuronic acid moiety in the bound state. Low molecular weight contaminants are marked '*'

In view of the large linewidths, which ruled out the application of coherence transfer methods based upon ^{1}H-^{1}H couplings, we utilised the a technique known as HCCH-COSY originally designed for for the resonance assignment of sidechains in proteins (Bax *et al.*, 1990), which relies on much larger one bond ^{13}C-^{13}C and ^{1}H-^{13}C couplings. In two-dimensional form, sensitivity was sufficient to allow resonance assignment of C1 through C5 non-exchangable protons of the glucuronic acid moiety (figure 6.2). These assignments were then utilised in the interpretation of a ^{13}C isotope-edited NOESY experiment on the complex (figure 6.3). Crosspeaks in this spectrum corresponding to intra residue NOEs were within the glucuronic acid moiety could readily be assigned, but however the assignment of inter-residue NOE connectivities was not straightforward due to the absence of resonance assignments for the steroid moiety in the bound state and for the amino acid sidechains in the binding site. In order to overcome this difficulty, various NOE connectivities to the steroid were therefore simulated in terms of the possible bound-state configurations of E3G arising from torsional variations about the glycosidic linkage, using a full relaxation matrix analysis and an estimate of the rotational correlation time of 15 ns, derived by fitting this parameter to the observed ratio of the C5 proton diagonal peak and the intra residue C5 - C1 crosspeak intensities. No single conformation could be found which satisfied the observed NOE connectivities either quantitatively or qualitatively. For example, no conformation about the glycosidic linkage of E3G predicted an NOE of the observed intensity from the C5 to two other resonances at 6.01 ppm and 7.09 ppm in the isotope-edited NOESY spectrum, assuming that both these resonances derive from the estrone moiety.

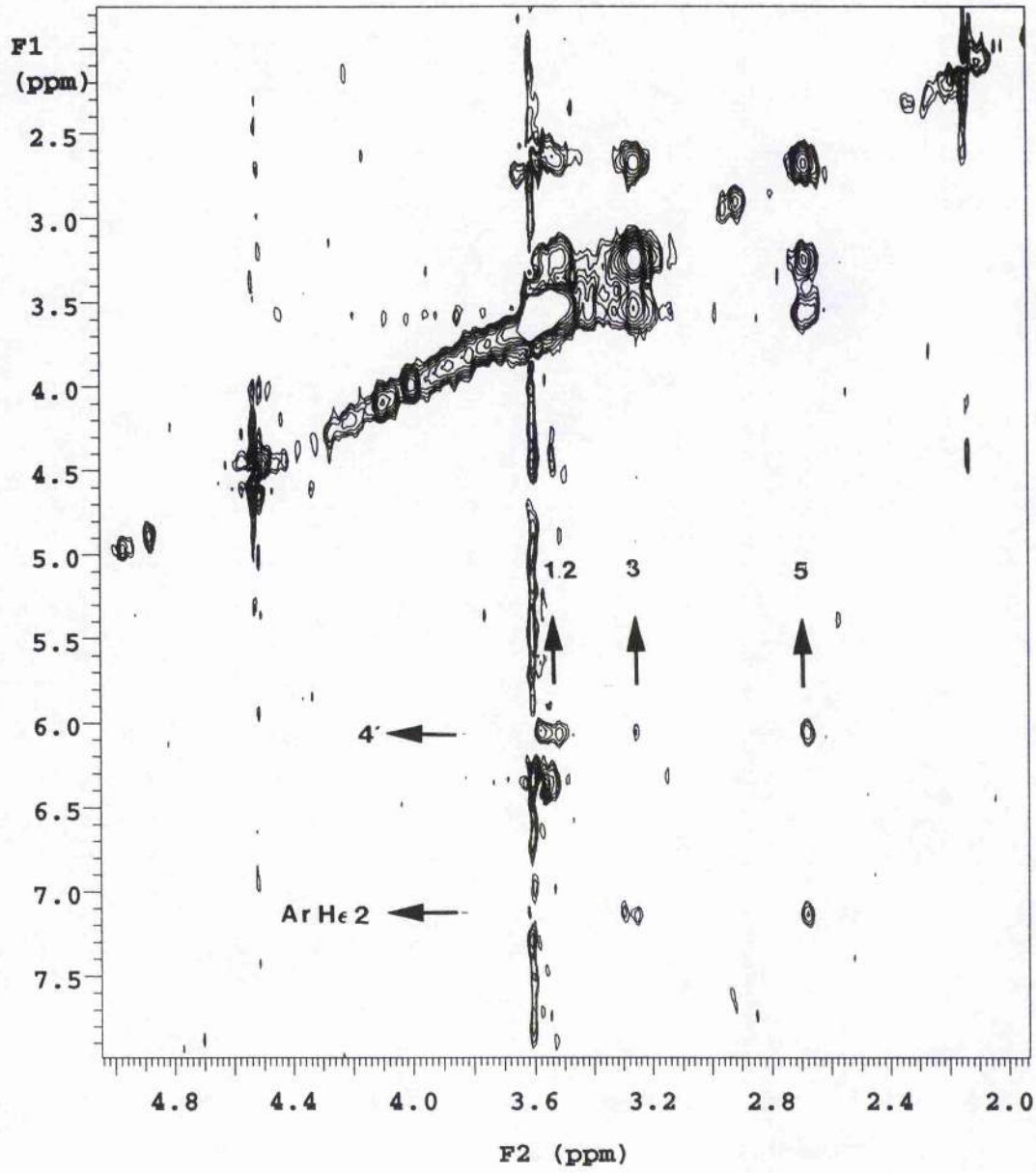


Figure 6.2 ^{1}H-^{1}H NOE-HSQC spectrum of the Fv/estrone-3-[^{13}C-U]-glucuronide complex in 100 mM phosphate buffer in D_2O, showing the inter-residue NOEs and ligand protein NOEs to Tyr 33.

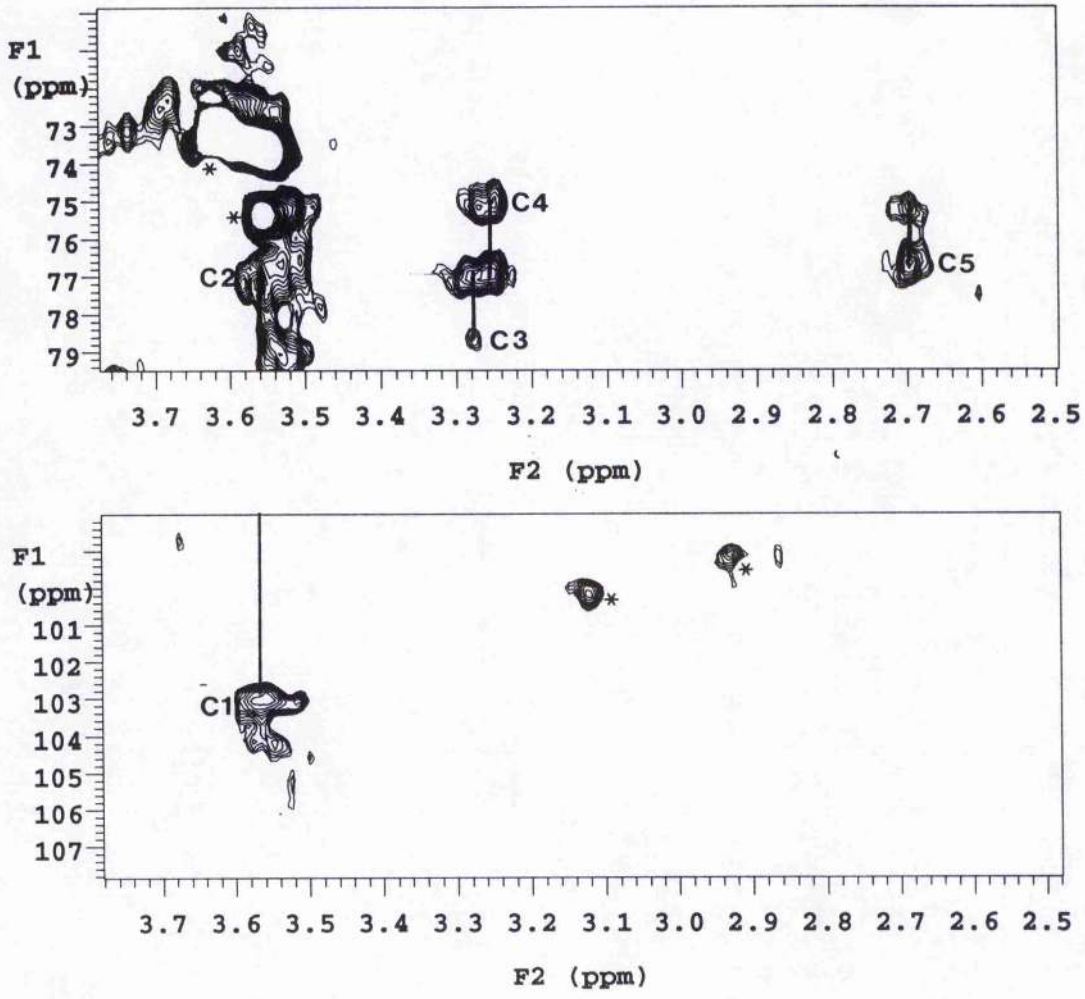


Figure 6.3 HCCH-COSY spectrum of the Fv/estrone-3-[^{13}C-U]-glucuronide complex in 100 mM phosphate buffer in D_2O, showing the proton and ^{13}C assignments for the glucuronic acid moiety in the bound state. Low molecular weight contaminants are marked '*'.

It was therefore surmised that certain of the observed inter-residue connectivities were between the glucuronic acid moiety and protein. Clearly either of the resonances at 6.01 and 7.09 could arise from protein, but given that these shifts are in the 'aromatic' region of the proton NMR spectrum of proteins, it was assumed that the relevant NOEs were to an aromatic sidechain in the binding site. Guidance in this assumption was also given by the very large proton resonance shifts of the glucuronic acid moiety observed in the bound state in comparison with those in the free state (table 6.1). Shifts of this magnitude can only arise from ring currents (Perkins, 1982), and the particularly large upfield shifts of the C1, C3 and C5 protons of the glucuronic acid moiety strongly suggested a stacking interaction between the C1, C3 and C5 face and an aromatic sidechain, the observed shifts were interpreted in terms of the distance of each proton in glucuronic acid from the centre of a simple phenolic ring, using a gridsearch and optimisation procedure coupled with the program RCCAL (Perkins, 1982) for computation of the solution of the Johnson-Bovey equation for the various geometries. The steroid moiety was then added to this optimised geometry, and NOEs were simulated for this complex for a variety of geometries obtained by varying the torsion angles ϕ and ψ independently in 30° increments. Good qualitative agreement was obtained between simulated and observed NOEs for a geometry ϕ,ψ = +30°, +30°, and this geometry was further optimised manually to obtain the best quantitative fit with experimental data, giving a predicted bound-state conformation of the ligand with ϕ,ψ = +45°, +24°. The predicted assignments and NOE intensities for this complex together with experimental NOE experimental values are given in table 6.1. It is seen that there is generally good agreement between the simulated and

experimental values.

Atom/Atom pair	Ring Current Shift (ppm) Exptl.	Theor.	NOE (%) Exptl.	Theor.
GlcU-H1	1.51	1.49		
GlcU-H2	0.10	0.22		
GlcU-H3	0.36	0.47		
GlcU-H4	0.38	0.21		
GlcU-H5	1.18	1.30		
GlcU-H5 - GlcU-H1/H2			10.7	7.3/3.4
GlcU-H1/H2 - Est-H4′			10.0	9.2/2.8
GlcU-H3 - Est-H4′			1.8	3.6
GlcU-H5 - Est-H4′			4.4	7.6
GlcU-H5 - Ar-Hε4′			3.8	4.3
GlcU-H3 -AR-Hε4′			2.4	4.2

Table 6.1 Experimental ring current shifts and NOEs derived from estrone-3-[U-^{13}C]-glucuronide in complex with anti-E3G versus theoretical values from predicted bound state conformation shown in figure 6.4.

Atom/Atom Pair	Ring Current Shift (ppm)		NOE (%)	
	Exptl.	Theor.	Exptl.	Theor.
GlcU-H1	1.51	1.34		
GlcU-H2	0.10	0.12		
GlcU-H3	0.36	0.41		
GlcU-H4	0.38	0.37		
GlcU-H5	1.18	1.15		
GlcU-H5 - GlcU-H1/H2			10.7	7.7/3.0
GlcU-H1/H2 - Est-H4′			10.0	8.3/2.0
GlcU-H3 - Est-H4′			1.8	2.6
GlcU-H5 - Est-H4′			4.4	5.0
GlcU-H5 - Ar-Hε4′			3.8	4.2
GlcU-H3 -AR-Hε4′			2.4	3.2

Table 6.2 Theoretical ring current shifts and NOEs derived from X-ray structure of estradiol-3-glucuronide in complex with anti-E3G versus experimental values from derived from estrone-3-[U-^{13}C]-glucuronide in complex with anti-E3G Fv.

The above analysis suggest that the resonance at 7.09 ppm derives from an aromatic sidechain, while that at higher field (6.01 ppm) derives from the C4′ proton of the steroid moiety of E3G. This gives rise to two points worthy of note. First, while this particular pair of assignments has been obtained indirectly, the reverse assignments do not give rise to a pattern of NOE connectivities which are consistent with experiment. For example, there is no valid conformation for bound E3G which would give rise to a measurable NOE from the C2′ proton of the glucuronic acid moiety to the aromatic sidechain, since the latter is on the opposing side of the sugar ring. Second, on grounds of symmetry, it could be postulated that the steroid moiety is rotated by 180° about ψ, and that the NOEs to the C4′ proton could easily well be attributed to the C2′ proton. However, in that case, substantial spin-diffusion which occurs at these rotational correlation times and relatively long NOESY mixing time would result in a measurable NOE to the C1′ proton of the steroid from the C1 proton of the glucuronic acid moiety for example, yet none is observed. We consider the formal possibility that the C1 and C2 protons of the steroid moiety in the bound state. Clearly a definitive set of assignments in the bound state could be obtained from E3G ^{13}C enriched in the steroid moiety but this was not feasible on grounds of availability and cost.

It is clear from the above that the bound state conformation of the glycan is different from either of the minimum energy structures predicted from the dynamical simulated annealing calculations, and it is of value to compare the relative energies of these conformations. The relative energies of the two conformers are 0 kcal.mol^{-1} and 0.2 kcal.mol^{-1} for ϕ,ψ = +18°, –84° and φ,ψ = +28°, +92° respectively, while the relative energy of

the NMR derived bound state conformation (ϕ,ψ = +45°, +24°) = 15.8 kcal.mol^{-1}. It would appear, therefore, that considerable torsional strain is imposed on the glycan in the bound state. However the relative energy of the bound state conformation is only a crude estimate, having been obtained from a single point calculation on the NMR derived conformation. In view of the errors inherent in the derivation of this conformation, it is very possible that the actual relative energy of the glycan in the bound state is lower, and 15.8 kcal.mol^{-1} should be considered as an upper estimate.

During the course of this work we were made aware of preliminary X-ray diffraction data on the complex between anti-E3G Fv and a ligand closely related to E3G, namely estradiol-3-glucuronide (S. E. V. Phillips, personal communication). It is instructive to compare the architecture of the binding site with that predicted from the present work. The crystallographic study reveals that there is is an indeed an aromatic ring (from Tyr 33, V_H) which stacks with the glucuronic acid moiety in the predicted manner, and the ligand binds in a similar conformation to that predicted with ϕ,ψ = +38°, 0° (figure 6.4). The disposition of the sidechain of Tyr 33 in the crystal appears to differ somewhat from that in solution. However this probability reflects the limited precision of the the ring current calculations which, in the absence of structural data in the protein were based upon a single aromatic residue. This conclusion is reinforced by the fact that the relevant ring current shifts predicted by use of the full crystallographic coordinates of the complex are in better agreement with experimental data than those derived above (table 6.2), primarily as a result of ring current contributions from two additional tyrosine residues

in the binding site. Furthermore, NOE simulations involving all protons within the vicinity of the binding site from the crystallographic study are in better agreement values than those determined above (table 6.2).

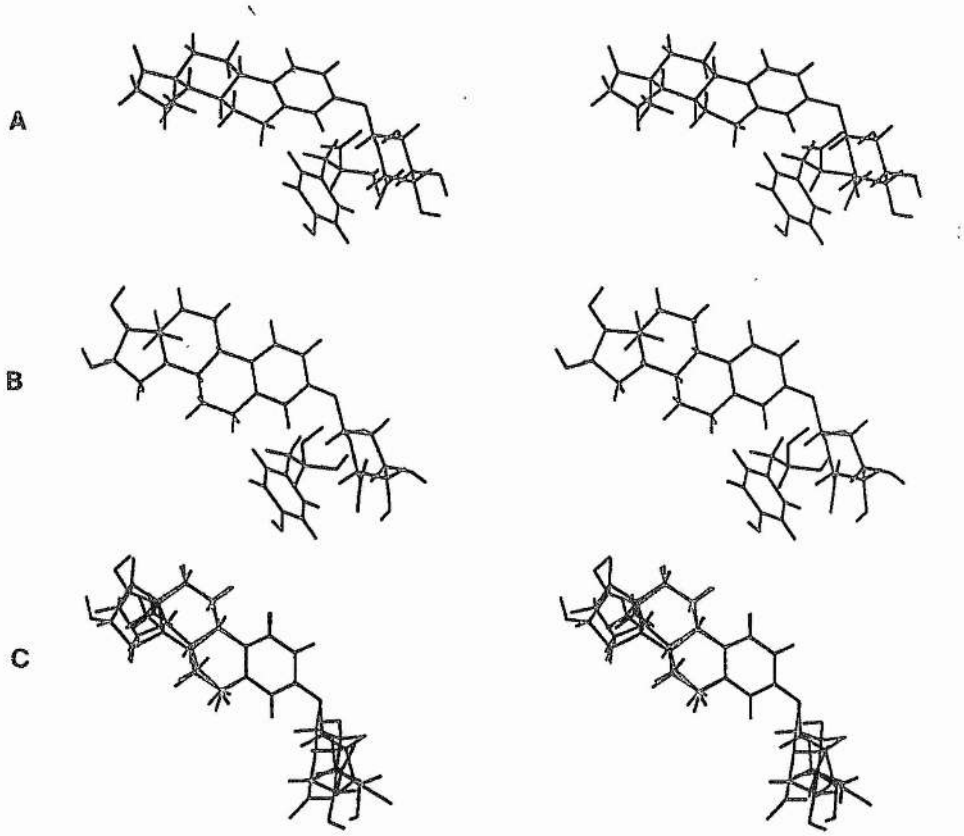


Figure 6.4 Comparison of calculated NMR structure of E3G in the bound state with the stacked tyrosine residue (A), with the X-Ray structure of a similar antigen in a complex with the anti-E3G Fv fragment (B), (C) shows the two structures superimposed.

6.3 Conclusions

The conformation of E3G in the binding site of the anti-E3G Fv does not correspond to either of the two low-energy conformations in solution, and has an internal energy estimated as 15.8 kcal.mole^{-1} higher than the lowest energy solution confirmation. The reduction in configurational energy of the free ligand plus the torsional strain placed on the binding is obtained at the expense of the free energy of binding. Since the affinity is very high, a substantial contribution to the binding energy must derive from other favourable interactions. Moreover, since the affinity of estrone for anti-E3G Fv is substantially lower than E3G a significant contribution to the binding energy must derive from the glucuronic acid moiety. The presence of a stacking interaction between an aromatic sidechain and the C1, C3, C5 face of the glucuronic acid moiety suggests that the hydrophobic interaction might be a dominant factor driving complex formation, as previously suggested by the previous work of Bundle and co-workers on the thermodynamics of oligosaccharide binding to a monoclonal antibody specific for *Salmonella* O-antigen (Sigurskjold *et al.*, 1991; Sigurskjold and Bundle 1992). However it is also probable that hydrogen bond formation involving the hydroxyl groups of the the glucuronic acid moiety and favourable entropic factors from the displacement of bound water in the Fv binding site might contribute to binding energy. This cannot be addressed in the present study, and a deeper insight into the nature of the interaction will require detailed thermodynamic analysis of E3G and analogues bound to Fv in combination with high resolution crystallographic and NMR data on the complex. Such studies are in progress.

In the context of rational drug design, the derivation of the bound state conformation of a glycan in the absence of structural data on the protein receptor is potentially of great value. In the current study the use of of a ^{13}C-enriched glycan in combination with isotope-filtration methods has been quite successful in realising this goal, as evidenced by the good agreement between the predicted conformation of E3G in complex with the anti-E3G Fv in comparison with the crystal structure in a closely related complex. Certain details of the architecture of the binding site could also be delineated in the solution study. However, this was highly dependent on the fact that significant ring current shifts were measurable in the bound state conformation of the ligand. Since carbohydrate protein interactions examined to date always involve a stacking interaction with an aromatic ring (Cambillau, 1995), ring current shift measurements are potentially a very important tool for probing carbohydrate-protein interactions, particularly in view of the fact that these shifts can be detected in systems in the fast exchange regime (Sauter *et al.*, 1989) as well as in the slow-exchange regime observed in this present study. Conversely, in circumstances where an aromatic residue is not implicitly involved in glycan binding, it will almost certainly be more difficult to define the the disposition of protons in the binding site, even if their presence is implicitly more obvious from crosspeaks in isotope-filtered NOESY experiments. Under these circumstances it may be difficult to define the conformation of the ligand also, since interpretation of intra-ligand NOEs is very dependent on nearby spins within the protein binding site due to efficient spin-diffusion in the relatively large complexes which are likely to be encountered.

While these problems might be over come in part by application of isotope-filtered rotating frame Overhauser effect measurements which are much less sensitive to the effects of spin diffusion, the determination of the complex using ^{13}C, ^{15}N labelled Fv is likely to provide a much more detailed picture of the structure and dynamics of the interaction.

Chapter 7.

Homology Modelling of Fv E3G and a comparison with the X-Ray Crystallographic Structure

Abstract

A computer generated model structure of Fv E3G was built using the modelling package Homology (Biosym Inc.). This was to assess the validity of the modelled structure by comparison with the Fv structure obtained by X-ray crystallography. Although the C^{α} trace suggested a good correlation between the crystal and theoretical structure, closer inspection of the binding site, showed a poor correlation with the position of the tyrosine residues in the X-ray structure (Tyr32, 33, 105), and with the predicted tyrosine residue (Tyr33) position in the NMR structure of the binding site.

7.1 Introduction

As there was not sufficient labelled material for a full structural analysis of the protein by NMR, homology modelling allowed a fast method of obtaining a model structure of the protein of interest; and with the subsequent availability of an X-ray crystallographic structure of the Fv fragment binding a closely related antigen, this would provide a comparison of the two structures and allow the assessment of the validity of the modelled protein structure. Several homology modelled Fv structures have been published (Eigenbrot *et al.*, 1993; Padlan and Kabat, 1988; Smith-Gill *et al.* 1987) also including those which have a comparison with an experimentally determined structure (Eigenbrot *et al.*, 1993; Smith-Gill *et al.* 1987). From the number of available published X-ray and NMR structures, those chosen were the structures that had the highest resolution. The Fv fragments all have homologous amino acid sequences and similar numbers of amino acids, identical in the framework regions, this makes the alignment of the sequences very obvious. This alignment stage is critical because it determines the correspondence between the

amino acids in the reference proteins and the model protein. By comparing the structure of the backbone atoms of the reference proteins, structurally conserved regions (SCRs) between all the reference can be determined, and these coordinates applied to the backbone of the model protein. Non-conserved or variable regions (VRs) can be determined by searching the protein structure database to find suitable backbone structures. Once further refinements are carried out using normal minimisation and dynamical techniques a model protein structure is obtained.

During the modelling procedure, to reduce the vast number of non-bonded terms calculated in modelling a protein of this size a cut off distance for these non-bonded parameters was used. The parameter 'cutoff' is a point from where any non-bond interactions will be ignored. Simply cutting these of at a given point leads to discontinuities in the energy, therefore a switching function is applied that decreases these interactions over a distance ('swtdis'), where the function is reduced to zero at 'cutdis'. This now allows a method of modelling that includes non-bonded parameters that is not too computationally intense.

7.2 Materials and Methods.

The model of Fv 4155 myc was built using the Homology modelling package (Biosym Technologies Inc.). Three reference Fv fragments were used as suitable model proteins. The initial refinement, splice repair, were carried out with this package, but subsequent refinements were carried out using the DISCOVER modelling package using modified input files that were generated from Homology.

Ten pseudo-random geometries were generated by dynamical quenching: The initial minimised structure was subjected to unrestrained dynamics at 900 K for 50 ps, during which the torsional term ϕ was scaled by a factor of 7 to prevent distorted geometries, with a structure written after every 5 ps. Simulated annealing was carried out for each model in the following manner: The structure was minimised with a steepest decent algorithm until the maximum derivative was less than 1 kcal.Å^{-2}, with ϕ scaled by 5; then equilibrated with a thermal path by running dynamics successively for 1 ps at 900 K, 850 K and in steps of 50 K until a temperature of 300 K and then successively for 1 ps in decreasing steps of 10 K. After 1 ps at 10 K a further 1 ps at 5 K was followed by minimisation using a steepest All minimisations and dynamics were generally run with a dielectric constant of 2.0, 'cutoff' of 12.0 Å, a 'cutdis' of 11.0 Å and a 'swtdis' of 1.5 Å.

7.3 Results and Discussion.

Choice of reference proteins.

Because the protein to be modelled is a murine Fv antibody fragment of an immunoglobulin, Fab and Fv fragments could be used as suitable reference proteins. The proteins chosen were all Fv fragments as this was thought to be the best possible reference. Three Fv fragments from the Brookhaven protein data base with the highest resolution and suitable r-values were used. The first was a murine IgG Fv fragment with resolution of 1.8 Å and an r-value of 0.185. Its code in the Brookhaven database is 1VFB (Bhat *et al.*, 1994). The second protein was a murine Fv fragment with resolution of 1.7 Å and an r-value of 0.166, and its code is 1MFA (Zdanov *et al.*, 1994). The third reference protein is a human IgM Fv fragment with 2.3 Å resolution, an r-value of 0.201 and its code is 1IGM (Fan *et al.*, 1992).

Sequence alignment.

This is a critical part of the of the modelling procedure, an incorrect alignment would result in comparing unrelated parts of each protein. From the nature of these proteins function, where the diversity in binding comes from variability of the CDR regions of the proteins. and the framework regions of the proteins have well conserved amino acid sequences with identical numbers of amino acids, the sequence alignment is fairly trivial. The protein alignment for the three reference proteins along with the model protein was carried out using the multiple protein alignment module in the 'Homology' package (a modification of that described by (Schuler *et al.*, 1991). This alignment can be seen in table 7.1.

Assigning structurally conserved regions.

These are determined by comparing the backbone structure of the known proteins. The package allows an amino acid backbone chain to be aligned with the corresponding backbone atoms of one of the other reference proteins to be aligned and a root mean square deviation (RMSD) value calculated. The superimposition is displayed visually and it can be seen easily where the structures begin to be non-compatible structurally. Generally an RMSD value below 1.0 Å indicates a structurally conserved region. This is done throughout the sequence for all combinations of two of the reference proteins so we have a total of three alignments. If a certain part of sequence is structurally conserved in all of these comparisons we assume that this part of sequence is structurally conserved throughout the family and therefore can have coordinates for the backbone atoms assigned to it.

V_L

```
          1   5   10   15   20   25   30   35   40   45   50
E3G       DIELTQTPPSLPVSLGDQVSISCRSSQSLVSNNRRNYLHWYLQKPGQSPK
VFB       DIVLTQSPASLSASVGETVTITCRAS-----GNIHNYLAWYQQKQGKSPQ
MFA       QIVVTQ-ESALTTSPGETVTLTCRSS--TGTVTSGNHANWVQEKPDHLFT
IGM       DIQMTQSPSSLSASVGDRVTITCQAS-----QDISNYLAWYQQKPGKAPE

             55   60   65   70   75   80   85   90   95  100
E3G       LVIYKVSNRFSGVPDRFSGSGSGTDFTLKISRVAAEDLGLYFCSQSSHVP
VFB       LLVYYTTTLADGVPSPFSGSGSGTQYSLKINSLQPEDFGSYYCQHFWSTP
MFA       GLIGDTNNRAPGVPARFSGSLIGDKAALTITGAQPEDEAIYFCALWSNNW
IGM       LRIYDASNLETGVPSRFSGSGSGTDFTFTISSLQPEDIATYYCQQYQNLP

            105  110
E3G       LTFGSGTKLEIK
VFB       RTFGGGTKLEIK
MFA       WIFGGGTKLTVLGQP
IGM       LTFGPGTKVDIKRTVAAPSV
```

V_H

```
          115  120  125  130  135  140  145  150  155  160
E3G       QVQLQESGGGLVNLGGSMTLSCVASGFTFNTYYMSWVRQTPEKTLELVAA
VFB       QVQLQESGPGLVAPSQSLSITCTVSGFSLTGYGVNNVRQPPGKGLEWLGM
MFA       EVQVQQSGTVVARPGASVKMSCKASGYTFTNYWMHWIKQRPGQGLEWIGA
IGM       EVHLLESGGNLVQPGGSLRLSCAASGFTFNIFVMSWVRQAPGKGLEWVSG

          165  170  175  180  185  190  195  200  205  210
E3G       INSDGEPIYYPDTLKGRVTISRDNAKKTLYLQMSSLNFEDTALYYCAR--
VFB       IWGDGNTDY-NSALKSRLSISKDNSKSQVSLKMNSLHTDDTARYYCAR--
MFA       IYPGNSATFVNHKFRAKTKLTAVTSTTTAYMELSSLTSEDSAVYYCAR--
IGM       VFGSGGNTDYADAVKGRFTITRDNSKNTLYLGMNSLRAEDTAIYYCAKHR

            215  220  225  230
E3G       LNYAVYGMDYWGQGTTVTVSS
VFB       --ERDYRLDYWGQGTTLTVS
MFA       -GGHGYYDGYWGQGASLTVS
IGM       VSYVLTGFDSWGQGTLVTVSSGSASAPTL
```

Table 7.1 Fv E3G amino acid sequence alignment with other Fv fragments used in homology modelling (numbers refer only to model protein: Fv E3G).

Assigning coordinates to the SCRs of the model protein.

The coordinates are assigned by assigning backbone coordinates of the model protein to those of a particular SCR of a reference protein which has the best sequence alignment to the model protein. To do this the three reference proteins must be superimposed upon one another, then each of the SCRs in the model protein sequence aligned with the corresponding reference protein. Using the mutation scoring matrix alignment method, in the Homology package, a score is given for each alignment. The model protein SCR then has its backbone coordinates assigned to that of the SCR of the reference protein with the best sequence similarity. This is carried out for all SCRs (table 7.2). The model protein now has coordinates assigned for all the backbone atoms of its SCRs.

Assigning variable region (VR) coordinates.

The loop or variable regions of the model protein (table 7.2) do not show any structural homology to the reference proteins and therefore can not be inferred from the coordinates of these proteins. These coordinates are obtained by searching proteins in the database for structures of the required length that would fit properly into the model protein between the two SCRs. The program gives ten best options with the best overall values for RMSD between the model and database proteins in the linking, or preflex and postflex peptide segments. These regions could be built into the structure and minimised, but this is only seen as an alternative if no suitable loops could be found from the first procedure. In each case suitable loops were found and the one with the lowest RMSD value was chosen as the backbone coordinates to which the model protein loop would be assigned. These results are displayed in table 7.2.

Structurally Conserved Regions.

SCR No.	amino acids	SCR No.	amino acids
V_L		V_H	
1.	2-6	1	114-118
2.	8-26	2.	129-132
3.	38-46	3.	143-151
4.	50-70	4.	156-164
5.	76-93	5.	173-195
6.	102-110	6.	196-210
		7.	219-230

Variable Regions

VR No.	amino acids	RMSD (Å)	VR No	amino acids.	RMSD (Å)
V_L			V_H		
1.	7	0.506	1	119-128	0.348
2.	25-37	0.566	2.	133-142	0.329
3.	47-49	0.513	3.	152-155	0.808
4.	71-75	0.480	4.	165-172	0.769
5.	94-101	0.287	5.	211-218	0.707

Table 7.2 List of structurally conserved and variable regions for both light and heavy chains. VR regions show rms deviation of post- and pre-flex regions of the chosen loop in comparison to that of modelled protein.

Now backbone assignments have been made for all the amino acids in the model protein except for the ends of the chains that are not already defined in a structurally conserved region. These are assigned using the end assignment protocol in the Homology modelling package.

Homology modelled Fv fragment structure refinement.

Now all the backbone atoms have been assigned for the model protein, these can not be considered as final positions. Since the model has been built from several different proteins the splice or junction points between the segments may not be smooth, although precautions such as superimposing all the reference proteins before assigning coordinates and all loops are superimposed in the tail region before assignment, there is no guarantee that the peptide bonds between these sections will be of the proper length or have the correct trans configuration. Another problem is the coordinates of the side chains. These can be inferred from the backbone atoms but since the backbone coordinates come from segments of different sequence there may be overlap problems with the side chains of the amino acids of the model protein in this inferred geometry. Therefore several modelling and energy minimisation routines will be needed to refine the structure. These minimisations can be carried out using the DISCOVER molecular modelling package of which the homology package creates suitable input files to carry out the necessary refinements. These files were edited to make them suitable for modelling Fv fragments as described in section 7.2.

Splice repair.

The homology package was used to generate the necessary files for a DISCOVER energy minimisation. This calculation fixes all the amino acids

not involved in joining to separate segments, but forces the peptide bonds between segments trans (180 °), and to the correct length, then minimised.

Steric overlap.

Due to the problems of superimposing side chain coordinates to the backbone coordinates obtained from a protein with a different amino acid sequence it is likely that several side chains will have van der Waal's overlap with other side chains. These overlaps were relieved by an energy minimisation where all the backbone atoms were held in position but where all the side chain atoms were free to move in the energy minimisation, firstly with a less stringent overlap value of 0.8 then a more stringent value of 0.4. This value is the minimum distance between atoms during initialisation of a DISCOVER run.

Variable region minimisation.

The variable region coordinates were taken directly from proteins in the database. The structure of these loops could be highly influenced by the position of other parts of the protein from which that particular variable region was obtained. For this reason an energy minimisation was performed on the model protein, where the variable regions atoms were allowed the freedom to find the minimum energy conformation in the context of the model protein, but all atoms remained fixed.

Modelling of binding site of Fv E3G.

An attempt to determine the amino acid residues involved in binding the E3G antigen was made using the program GRID. This program probes the surface of the molecule with different functional groups, and from the output an electrostatic map can be drawn which shows the most

favourable sites of binding. This was tried with the COO^-, phenyl and CH_3 groups which are all present in the antigen. These results proved inconclusive, with no specific points of binding found for any of the above groups.

Comparison of Homology modelled structure with X-ray crystallography data.

As for the previous chapter the preliminary X-ray crystallography data for the complex of the recombinant protein bound to estriol glucuronide could be used to assess the validity of the modelled protein. Unfortunately there is no similar data for the protein in the unbound state, which would provide a much better reference point for a comparison with the homology structure. To make a valid comparison, the Fv-E3G complex was modelled using the final homology structure with E3G in the position predicted by the X-ray crystallography data, and using the bound conformation of E3G predicted by NMR. From this generated complex, the SCR backbone atoms and the E3G atoms were fixed. The molecule was used to generate 10 pseudo random structures, which were minimised by dynamical simulated annealing. the C^{α} trace of the 10 protein structures are shown in figure 7.1. These show that the structures align very well in the SCRs and in the 3 variable regions involved in binding the Fv fragment V_L VR4; V_H VR2 and VR5, but in other regions there is a large diversity: V_H VR1 and V_L VR3 and VR4.

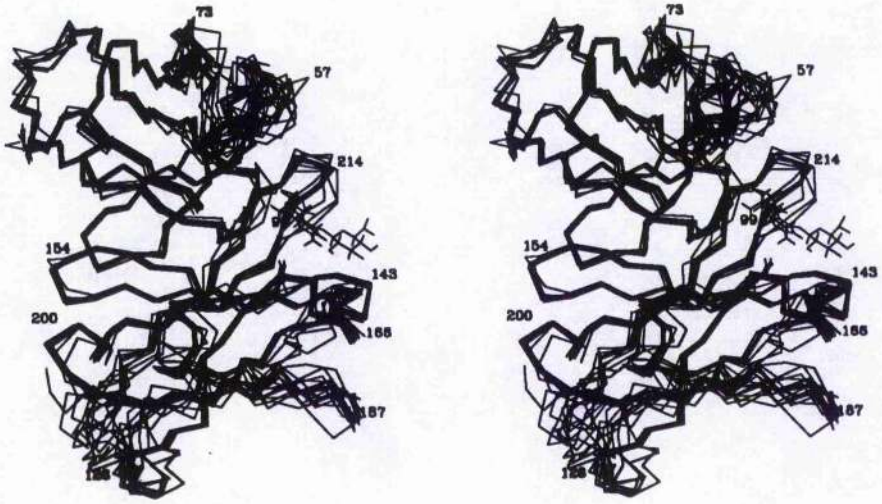


Figure 7.1 Stereo diagram of C_α traces from dynamical simulated annealing of E3G anti E3G Fv complex (amino acid numbers shown for orientation).

When comparing the minimum energy structure with the X-ray protein structure (figure 7.2), superimposition of the two C^α traces gives an RMSD value of 3.085 Å. By inspection the result matches the crystal structure reasonably, although there is a lack of homology for V_H VR3 and VR4. Importantly the areas of the modelled structure most directly affected by the inclusion of the antigen match well with the experimental structure. But because of the conserved nature of Fv structures this is not surprising and does not give much valuable information.

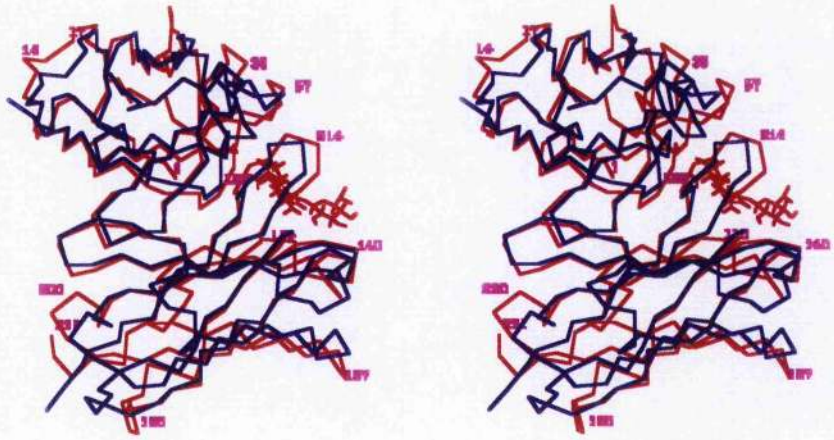


Figure 7.2 Comparison of C^{α} trace of lowest energy annealed structure with X-ray crystal structure C^{α} trace

A much better comparison would be to compare the amino acids directly involved in binding the E3G, especially the tyrosines involved in binding of the glucuronic acid moiety. This would allow a direct comparison with the NMR structure obtained in chapter 6, and give insight into the relative validity of the structure obtained relative to both the NMR and X-ray structures, allowing back calculation of NOEs and ring current shifts. Figure 7.3 shows the 10 annealed structures and the Fv structure superimposed on the glucuronic acid moiety, and for each protein structure residues 144,145 and 213 are displayed, for the X-ray structure these are equivalent to H32, H33 and H101. It can easily be seen from this figure that the modelled structures do not match the experimental structure, and from the position of residue 144 (Tyr33) it is obvious the calculated ring current shifts would bear no relation to experimental values. It is obvious that this approach is not adequate for obtaining an accurate protein structure.

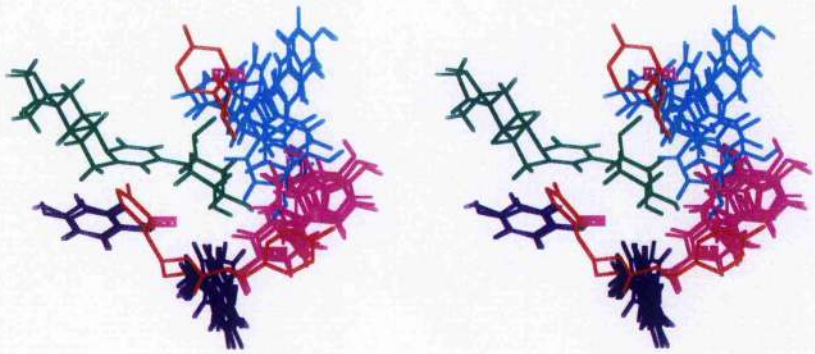


Figure 7.3 Comparison of binding sites for the annealed protein structures (red)with the X-ray crystal structure. Amino acids are numbered for the X-ray structures and the corresponding amino acid side chains for the modelled structures are shown as follows: 32, magenta; 33, blue; and 101, cyan.

7.4 Conclusions

By using the Homology modelling package it has been possible to obtain a protein structure for the Fv fragment without relying on any structural data obtained experimentally. In comparing the C^{α} traces of the minimum energy annealed structure and the X-ray structure of the Fv fragment binding a closely related antigen, a relatively good agreement is obtained with good agreement in secondary structure. Obviously the use of database segments of structures will give more valid structure than that on pure theoretical structure.

A theoretical calculation of the binding site tyrosine side chains was attempted. This could prove as a useful comparison to both X-ray and NMR derived structural information. Firstly a suitable binding site was not found purely from theoretical means, and secondly a poor correlation of the position of the tyrosine side chains involved in binding between

the X-ray and theoretical structures was found. As there was a good range of starting geometries from the pseudo-random structures, and these minimised in each case to either one or two families, it would seem the theoretical calculations are incapable of providing an accurate nature of the binding site. Therefore, if the suitable binding site had been found theoretically, it would appear that the predicted theoretical structure would still be incorrect. This would suggest that in this case some experimental evidence of the binding site would have to be attained in order to produce a valid structure. For example the Fv fragment with labelled tyrosine residues (with NMR assignments and constraints) would allow a more accurate NMR structure of the binding site. Work on a fully labelled Fv fragment produced in yeast is now being undertaken in the laboratory.

Albrand J. P., Birdsall B., Feeney J. and Roberts G. C. K. (1979) *International Journal of Biological Macromolecules* **1** 37-41.

Andrews J. S., Weimar T., Frandsen T. B., Svensson B. and Pinto B. M. (1995) Novel disaccharides containing sulphur in the ring and nitrogen in the interglycosidic linkage - conformation of methyl 5'-thio-4-N-α-maltoside bound to glucoamylase and its activity as a competitive inhibitor. *Journal of the American Chemical Society* **117** 10799-10804.

Arepalli S. R., Glaudemans C. P. J., Daves G. D., Kovac P. and Bax A. (1995) Identification of protein-mediated indirect NOE effects in a disaccharide-Fab' complex by transferred ROESY. *Journal of Magnetic Resonance B* **106** 195-198.

Asensio J. L., Cañada F. J. and Jimenez-Barbero J. (1995) Studies of the bound conformations of methyl α-lactoside and β-allolactoside to ricin-B chain using transferred NOE experiments in the laboratory and rotating frames, assisted by molecular mechanics and dynamics calculations. *European Journal of Biochemistry* **233** 618-630.

Backman I., Erbing B., Jansson P.-E. and Kenne L. (1988) NMR and conformational studies of some 1, 4-linked disaccharides *Journal of the Chemical Society, Perkin Transactions 1.* **4** 889-898.

Barbas C. F., Kang A. S., Lerner R. A. and Benkovic S. J. (1991) Assembly of

combinatorial antibody libraries on phage surfaces. The gene III site. *Proceedings of the National Academy of Sciences of the United States of America* **88** 7978-7982.

Baumann H. (1988) Synthesis of, NMR and conformational studies on some 3-O, 4-O and 3,4, di-O-glycopyranosyl-substituted methyl D-glycopyranosides. *University of Stockholm, Chemical Communications* **9** 1-32.

Bax A., Clore G. M., Driscoll P. C., Gronenborn A. M., Ikura M. and Kay L. E. (1990) Practical aspects of proton carbon carbon proton 3-dimensional correlation spectroscopy of ^{13}C labelled proteins. *Journal of Magnetic Resonance* **87** 620-627.

Better M., Chang C. P., Robinson R. R. and Horwitz A. H. (1988) *Escherichia coli* secretion of an active chimeric antibody fragment. *Science* **240** 1041-1043.

Bevilacqua V. L., Thomson D. S. and Prestegard J. H. (1990) Conformation of methyl β-lactoside bound to the ricin-B-chain - interpretation of transferred nuclear Overhauser effects facilitated by spin simulation and selective deuteration. *Biochemistry* **29** 5529-5537.

Bhat T. N., Bentley G. A., Boulot G., Green M. I., Tello D., Dallacqua W., Souchon H., Schwarz F. P., Mariuzza R. A. and Poljak R. J. (1994) Bound water molecules and conformational stabilisation help mediate an antigen-antibody association. *Proceedings of the National Academy of*

Sciences of the United States of America. **91** 1089-1093.

Bird R. E., Hardman K. D., Jacobson J. W., Johnson S., Kaufman B. M., Lee S.-M., Lee T., Pope S. H., Riordan G. S. and Whitlow M. (1988) Single-chain antigen-binding proteins. *Science* **242** 423-426.

Birnboim H. C. and Doly J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Research* **7** 1513.

Blundell T. L., Carney D., Gardner S., Hayes F., Howlin B., Hubbard T., Overington J., Singh D. A., Sibanda B. L. and Sutcliff M. (1988) Knowledge-based protein modelling and design. *European Journal of Biochemistry* **172** 513.

Blundell T. L., Sibanda B. L., Sternberg M. J. E. and M. T. J. (1987) Knowledge-based prediction of protein structures and the design of novel molecules. *Nature* **326** 347.

Bock K., Arnarp J. and Lonngren J. (1982) The preferred conformation of oligosaccharides derived from complex-type carbohydrate portions of glycoproteins. *European Journal of Biochemistry* **129** 171-178.

Bock K. and Lemieux R. U. (1982) The conformational properties of sucrose in aqueous solution - intramolecular hydrogen bonding. *Carbohydrate Research* **100** 63-74.

Bothner-By A. A. and Gassend R. (1972) Structure determination of a tetrasaccharide: Transient nuclear Overhauser effects in the rotating frame.

Annals of the New York Academy of Sciences **222** 668-676.

Brady J. W. (1986) Molecular dynamics simulations of α-D-glucose *Journal of the American Chemical Society* **108** 8153-8160.

Brady J. W. (1987) Molecular dynamics simulations of β-D-glucopyranose. *Carbohydrate Research* **165** 306-312.

Browne W. J., North A. C. T., Phillips D. C., Brew K., Vanaman T. C. and Hill R. L. (1969) A possible three-dimensional structure of bovine α-lactalbumin based on that of hen's egg-white lysozyme. *Journal of Molecular Biology* **42** 65.

Bundle D. R., Baumann H., Brisson J. R., Gagne S. M., Zdanov A. and Cygler M. (1994) Solution structure of a trisaccharide-antibody complex - comparison with a crystal structure. *Biochemistry* **33** 5183-5192.

Cambillau C. in *New Comprehensive Biochemistry* (eds. Neuberger, A. & van Deenen, L.L.M.) 29-65 (1995).

Campbell A. P. and Sykes B. D. (1993) The two-dimensional transferred nuclear Overhauser effect - theory and practice. *Annual Reviews in Biophysics and Biomolecular Structure* **22** 99-122.

Carter P., Presta L., Gorman C. M., Ridgway J. B. B., Henner D., Wong W. L., Rowland A. M., Kotts C., Carver M. E. and Shepard H. M. (1992) Humanisation of an anti-p185^{HER2} antibody for human cancer therapy.

Proceedings of the National Academy of Sciences of the United States of America **89** 4285-4289.

Cavanagh J., Hunter C. A., Jones D. N. M., Keeler J. and Saunders J. K. M. (1988) Practicalities and applications of reverse heteronuclear shift correlation - porphyrin and polysaccharide examples. *Magnetic Resonance in Chemistry* **26** 867-875.

Chaudhary V. K., Queen C., Junghans R. P., Waldmann T. A., Fitzgerald D. J. and Pastan I. (1989) A recombinant immunotoxin consisting of two antibody variable domains fused to *Pseudomonas* exotoxin. *Nature* **339** 394-397.

Chitarra V., Alzari P. M., Bentley G. A., Bhat T. N., Eisele J.-L., Houdusse A., Lescar J., Souchon H. and Poljak R. J. (1993) Three dimensional structure of a heterocyclic antigen-antibody cross-reaction complex *Proceedings of the National Academy of Sciences of the United States of America* **90** 7711.

Clore G. M. and Gronenborn A. M. (1982) Theory and application of the transferred nuclear Overhauser effect to the study of small ligands bound to proteins. *Journal of Magnetic Resonance* **48** 402-417.

Clore G. M. and Gronenborn A. M. (1983) Theory of the time-dependent transferred nuclear Overhauser effect - application to analysis of ligand protein complexes. *Journal of Magnetic Resonance* **53** 423-442.

Clore G. M., Gronenborn A. M., Brunger A. T. and Karplus M. (1985) Solution conformation of a heptadecapeptide comprising the DNA

binding helix-F of the cyclic-AMP receptor protein of *Escherichia coli* - combined use of ^{1}H nuclear magnetic resonance and restrained molecular dynamics. *Journal of Molecular Biology* **186** 435-455.

Constantine K. L., Goldfarb V., Wittekind M., Anthony J., Ng S.-C. and Mueller L. (1992) Sequential ^{1}H and ^{15}N NMR assignments and secondary structure of a recombinant anti-digoxin antibody V_L domain. *Biochemistry* **31** 5033-5043.

Cumming D. A. and Carver J. P. (1987) Virtual and solution conformations of oligosaccharides. *Biochemistry* **26** 6664-6676.

Dais P. and Perlin A. S. (1987) Proton spin-lattice relaxation rates in the structural analysis of carbohydrates. *Carbohydrate Chemistry and Biochemistry* **45** 125-168.

Derome A. E. *Modern NMR techniques for chemistry research.* (Pergamon Press, Oxford, 1987)

Dueñas M., Vázques J., Ayala M., Söderlind E., Ohlin M., Pérez L., Borrebaek C. A. K. and Gavilondo J. V. (1994) Intra- and extracellular expression of an scFv antibody fragment in *E. coli*: Effect of bacterial strains and pathway engineering using GroES/L chaperonins. *Biotechniques* **16** 476-483.

Edmonds M., Vaughan Jr. M. H. and Nakazato H. (1971) Polyadenylic acid sequences in the heterogeneous nuclear RNA and rapidly labelled

polyribosomal RNA of HeLa cells: Possible evidence for a precursor relationship. *Proceedings of the National Academy of Sciences USA* **68** 1336.

Eigenbrot C., Randal A., Leonard P., Carter P. and Kossiakoff A. A. (1993) X-ray structures of the antigen binding domain from three variants of humanised anti-p185^{HER2} antibody 4D5 and comparison with molecular modelling. *Journal of Molecular Biology* **229** 969-995.

Essen L.-O. and Skerra A. (1994) The *de novo* design of an antibody combining site; crystallographic analysis of the V_L domain confirms the structural model *Journal of Molecular Biology* **238** 226-244.

Essig N. Z., Wood J. F., Howard A. L., Raag R. and Whitlow M. (1993) Crystallization of single-chain Fv proteins. *Journal of Molecular Biology* **234** 897-901.

Fan C.-Z., Shan L., Geddat L. W., He X.-M., Gray W.-R., Raison R. L. and Edmundson A. B. (1992) Three dimensional structure of an Fv from a human IgM immunoglobulin. *Journal of Molecular Biology* **288** 188.

Favaloro J., Treisman R. and Kamen R. (1980) Transcription maps of polyoma virus specific RNA: Analysis by two-dimensional nuclease S1 gel mapping. *Methods in Enzymology* **65** 718.

Forster M. J. (1991) Comparison of computational methods for simulating nuclear Overhauser effects in NMR spectroscopy. *Journal of*

Computational Chemistry **12** 292-300.

French A. D. (1988) Rigid and relaxed-residue conformational analyses of cellobiose using the computer program MM2. *Biopolymers* **27** 1519-1525.

French A. D. and Brady J. W. Computer modelling of carbohydrate molecules in *ACS symposia* (American Chemical Society., Washington., 1989)

Freund C., Ross A., Guth B., Plückthun A. and Holak T. A. (1993) Charecterisation of the linker peptide of a single-chain Fv fragment of an antibody by NMR spectroscopy. *FEBS Letters* **320** 97-100.

Genest D. (1989) A Monte-Carlo simulation study of the influence of internal motions on the molecular conformation deduced from two-dimensional NMR experiments *Biopolymers* **28** 1903-1911.

Gidley M. J. and Bociek S. M. (1988) ^{13}C CP/MAS NMR studies of amylose inclusion complexes, cyclodextrins and the amorphous phase of starch granules: Relationship between glycosidic linkage conformation and solid-state ^{13}C chemical shifts. *Journal of the American Chemical Society*

Glaudemans C. P. J., Lerner L., Daves G. D., Kovac P., Venable R. and Bax A. (1990) Significant conformational-changes in an antigenic carbohydrate epitope upon binding to a monoclonal antibody. *Biochemistry* **29** 10906-10911.

Glockshuber R., Malia M., Pfitzinger I. and Plückthun A. (1990) A comparison of strategies to stabilize immunoglobulin Fv-fragments. *Biochemistry* **29** 1362-1367.

Goebel C. V., Dimpfl W. L. and Brant D. A. (1970) The conformational energy of maltose and amylose. *Macromolecules* **3** 644-654.

Grant D. M. and Cheney B. V. (1967) Carbon-13 magnetic resonance. VII: Steric perturbation of the Carbon-13 chemical shift. *Journal of the American Chemical Society* **89**

Greer J. (1980) Model for haptoglobin heavy chain based upon structural homology. *Proceedings of the National Academy of Sciences of the United States of America* **77** 3393.

Greer J. (1981) Comparitive model-building of the mammalian serine proteases *Journal of Molecular Biology* **153** 1027.

Greer J. (1985) Model structure of the inflammatory protein C5a *Science* **228** 1055.

Griffiths A. D., Malmquist M., Marks J. D., Bye J. M., Embleton M. J., McCafferty J., Baier M., Holliger K. P., Gorick B. D., Hughes-Jones N. C., Hoogenboom H. R. and Winter G. (1993) Human anti-self antibodies with high specificity from phage display. *EMBO Journal* **12** 725-734.

Ha S., Giammona A. and Field M. (1988a) A revised potential-energy surface for molecular mechanics studies of carbohydrates. *Carbohydrate*

Research **180** 207-221.

Ha S. N., Madsen L. J. and Brady J. W. (1988b) Conformational analysis and molecular dynamics simulations of maltose *Biopolymers* **27** 1927-1952.

Haasnoot C. A. G., de Leeuw F. A. A. M. and Altona C. (1980) The relationship between proton-proton coupling constants and substituent electronegativities-1. *Tetrahedron* **36** 2873-2792.

Haigh C. W. and Mallion R. B. (1972) *Organic Magnetic Resonance* **4** 73.

Hall L. D. and Johnson L. F. (1969) Chemical studies by ^{13}C nuclear magnetic resonance spectroscopy: Some chemical shift dependencies of oxygenated derivatives. *Journal of the Chemical Society, Chemical Communications* **509-510**

Hall L. D. and Preston C. (1972) A configuration dependence of the longitudinal relaxation times of carbohydrate derivatives. *Journal of the Chemical Society, Chemical Communications* 1319.

Hamer G. K., Balza F., Cyr N. and Perlin A. S. (1978) A conformational Study of methyl-beta cellobioside-d_8 by ^{13}C NMR Spectroscopy: dihedral angle dependene of ^3J in ^{13}C-O-C-^1H arrays *Canadian Journal of Chemistry* **56** 3109-3116.

Harris L. J., Larson L. B., Hasel K. W., Day J., Greenwood A. and McPherson A. (1992) A three-dimensional structure of an intact monoclonal antibody

for canine lymphona *Nature* **360** 369-372.

Holland I. B., Kenny B., Steipe B. and Plückthun A. (1990) Molecular mechanisms of protein secretion: the role of the signal sequence. *Advances in Protein Chemistry* **38** 132-143.

Homans S. W. (1990) A molecular mechanical forcefield for the conformational analysis of oligosaccharides: Comparison of theoretical and crystal structures of Manα1-3β1-4GlcNac. *Biochemistry* **29** 9110-9118.

Homans S. W., Edge C. J., Ferguson M. A. J., Dwek R. A. and Rademacher T. W. (1989) Solution structure of the Glycosylphosphatidylinositol membrane anchor glycan of *Trypanosoma bruciei* variant surface glycoprotein. *Biochemistry* **28** 2881-2887.

Homans S. W. and Forster M. (1992) Application of restrained minimisation, simulated annealing and molecular dynamics simulations for the conformational analysis of oligosaccharides. *Glycobiology* **2** 143-151.

Homans S. W., Pastore A., Dwek R. A. and Rademacher T. W. (1987) Structure and dynamics in oligo-mannose type oligosaccharides. *Biochemistry* **26** 6649-6655.

Hoogenboom H. R., Marks J. D., Griffiths A. D. and Winter G. (1992) Building antibodies from their genes. *Immunological Review* **130** 41-68.

Horri F. in *ACS Symposium Series* (1984)

Horwitz A. H., Chang C. P., Better M., Hellstrom K. E. and Robinson R. R. (1988) Secretion of functional antibody and Fab fragments from yeast cells. *Proceedings of the National Academy of Sciences of the United States of America* **85** 8678-8682.

Imberty A., Tran V. and Perez S. (1989) Relaxed potential energy surfaces of N-linked oligosaccharides: The mannose-α-1-3 mannose case. *Journal of Computational Chemistry.* **11** 205.

Ish-Horowicz D. and Burke J. F. (1981) Rapid and efficient cosmid cloning. *Nucleic Acids Research* **9** 2989.

Johnson C. E. and Bovey F. A. (1958) Calculation of nuclear magnetic resonance spectra of aromatic hydrocarbons. *Journal of Chemical Physics* **29** 1012-1014.

Kabat E. A. *Sequences of proteins of immunological interest.* (NIH publications, Bethedsa, MD., 1991)

Karplus M. (1959) Contact electron-spin coupling of nuclear magnetic moments. *Journal of Chemical Physics* **30** 11-15.

Karplus M. (1963) Vincinal proton coupling in nuclear magnetic resonance. *Journal of the the American Chemical Society.* **85** 2870-2871.

Köhler G. and Milstein C. (1975) Continuous cultures of fused cells

secreting antibody of predefined specificity. *Nature* **256** 495-497.

Laukkanen M.-L., Alfthan K. and Keinänen A. (1994) Functional immunoliposomes harbouring a biosynthetically lipid-tagged single-chain abtibody. *Biochemistry*. **33** 11644-11670.

Lemieux R. U. and Bock K. (1983) The conformational analysis of oligosaccharides by ^{1}H NMR and HSEA calculation. *Archives of Biochemistry and Biophysics* **221** 125-134.

Lemieux R. U., Bock K., Delbaere L. T. J., Koto S. and Rao V. S. (1980) The human conformations of oligosaccharides related to the ABH heman blood group determinents. *Canadian Journal of Chemistry*. **58** 631-653.

Lipari G. and Szabo A. (1982, a) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecule. 1. Theory and range of validity. *Journal of the American Chemical Society* **104**

Lipari G. and Szabo A. (1982, b) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecule. 2. Analysis of experimental results. *Journal of the American Chemical Society* **104** 4559-4570.

London R. E., Perlman M. E. and Davis D. G. (1992) Relaxation-matrix analysis of the transferred nuclear Overhauser effect for finite exchange rates. *Journal of Magnetic Resonance* **97** 79-98.

Madsen L. J., Ha S. N., Trans V. H. and Brady J. W. Molecular dynamics simulations of carbohydrates and their solvation. in *Computer modelling of carbohydrate molecules*. 69-90 (1990).

Marks J. D. Human monoclonal antibodies from V-gene reportoires expressed on bacteria phage. in *Antibody engineering*. (ed. Borrebaeck C. A. K.) 53-58 (Oxford University Press., Oxford., 1995).

Marks J. D., Hoogenboom H. R., Bonnert T. P., McCafferty J., Griffiths A. D. and Winter G. (1991) By-passing immunisation: human antibodies directly from V-gene libraries displayed on phage. *Journal of Molecular Biology* **222** 581.

McCafferty J., Griffiths A. D., Winter G. and Chiswell D. J. (1990) Phage antibodies: filamentous phage displaying antibody variable domains. *Nature* **348** 552-554.

Messing J. (1983) New M13 vectors for cloning. *Methods in Enzymology* **101** 20.

Morris G. A. and Freeman R. (1979) Enhancement of nuclear magnetic resonance signals by polarisation transfer *Journal of the American Chemical Society* **101** 760.

Mullis K., Faloona F., Scharf S., Saika R., Horn G. and Elrich H. (1986) Specific enzymatic amplification of DNA *in vitro*: The polymerase chain reaction. *Cold Spring Harbour Symposia on Quantitative Biology*. **1** 263-273.

Mulloy B., Frenkiel T. A. and Davies D. B. (1988) Long-range carbon-proton coupling constants: Application to conformational studies of oligosaccharides. *Carbohydrate Research* **184** 39-46.

Munro S. and Pelham H. R. B. (1986) An HSP70-like protein in the ER - identity with 78 kD glucose-regulated protein and immunoglobulin heavy-chain binding-protein. *Cell* **46** 291-300.

Neri D., de Lalla C., Petrul H., Neri P. and Winter G. (1995) Calmodulin as a versatile tag for antibody fragments. *Bio/Technology* **13** 373-377.

Neuhaus D. and Williamson M. *The nuclear Overhauser effect in structural and conformational analysis.* (VCH publishers, London, 1989)

Ni F. (1994) Recent developments in transferred NOE methods. *Journal of Magnetic Resonance* **97** 79-88.

Nishida Y., Hori H., Ohrui H. and Meguro H. (1987) ^{1}H NMR analyses of rotameric distribution of C5-C6 bonds of D-glucopyranoses in solution. *Journal of Carbohydrate Chemistry* **7** 239-250.

Nissim A., Hoogenboom H. R., Tomlinson I. M., Flynn G., Midgley C., Lane D. and Winter G. (1994) Antibody fragments from a 'single pot' phage display library as immunochemical reagents. *EMBO Journal* **13** 692-698.

Noggle J. H. and Schirmer R. E. *The nuclear overhauser effect, chemical*

applications. (Academic Press, New York, 1971)

Norrander J., Kempe T. and Messing J. (1983) Construction of improved M13 vectors using oligodeoxynucleotide directed mutagenesis. *Gene* **26** 101.

Nyyssönen E., Pentillä M., Harkki A., Saloheimo A., Knowles J. K. C. and Keränen S. (1993) Efficient production of antibody fragments by the filamentous fungus *Trichoderma reesei*. *Bio/Technology* **11** 591-595.

Orlandi R., Guessow D. H., Jones P.T. and Winter G. (1989) Cloning immunoglobulin variable domains for expression by the polymerase chain reaction. *Proceedings of the National Academy of Sciences of the United States of America* **86** 3833-3837.

Padlan E. A. and Kabat E. A. (1988) Model-building study of the combining sites of two antibodies to α(1-6)dextran *Proceedings of the National Academy of Sciences of the United States of America* **85** 6885-6889.

Perkins S. J. Application of ring current calculations to the proton NMR of proteins and transfer RNA. in *Biological Magnetic Resonance* (eds. Berliner, L. & Reuben, J.) pp 193-336 (Plenum Press., New York, 1982).

Petros A. M., Kawai M., Luly J. R. and Fesik S. W. (1992) Conformation of 2 nonimmunosuppresive FK506 analogues when bound to FKBP by isotope-filtered NMR. *FEBS Letters* **308** 309-314.

Plückthun A. (1990) Antibody engineering: Advantages from the use of

Escherichia coli expression systems. *Bio/Technology* **9** 545-552.

Plückthun A. and Skerra A. (1989) Expression of functional antibody Fv and Fab fragments in *Escherichia coli*. *Methods in Enzymology* **178** 497-515.

Poljak R. J., Amzel L. M., Avery H., Chen B. L., Phizakerly R. P. and Saul F. (1973) Three-dimensional structure of the Fab-fragment of a human immunoglobulin at 2.8 Å resolution. *Proceedings of the National Academy of Sciences of the United States of America* **70** 3305-3310.

Pople J. A. (1956) Proton magnetic resonance of hydrocarbons. *Journal of Chemical Physics* **24** 1111.

Popple L. and van Halbeek H. (1991, a) ^{1}H Detected Measurements of Long-range heteronuclear coupling constants. Application to a trisaccharide. *Journal of Magnetic Resonance* **92** 636-641.

Popple L. and van Halbeek H. (1991, b) Selective, Inverse Detected Measurements of Long-range ^{13}C, ^{1}H Coupling Constants. Application to a disaccharide. *Journal of Magnetic Resonance* **93** 214-217.

Reichmann L., Weill M. and Cavanagh J. (1992) Improving the antigen affinity of an antibody Fv fragment by protein design. *Journal of Molecular Biology* **224** 913-918.

Ridder R., Schmitz R., Legay F. and Gram H. (1995) Generation of rabbit

monoclonal antibody fragments from a combinatorial phage display library and their production in the yeast *Pichia pastoris*. *Bio/Technology* **13** 255-260.

Rutherford T. J. and Homans S. W. (1994) Restrained vs. free dynamics simulations of oligosaccharides - application to solution dynamics of biantennary and bisected biantennary N-linked glycans. *Biochemistry* **33** 9606-9614.

Saito H. and Ando I. in *Annual Reports on NMR Spectroscopy.* (Academic Press., London, 1989)

Sambrook J., Fritsch E. F. and Maniatis T. *Molecular cloning: A laboratory manual.* (1989)

Saunders J. K. M. and Hunter B. K. *Modern NMR spectroscopy: A guide for chemists.* (Oxford University Press, Oxford, 1987)

Sauter N. K., Bednarski M. D., Wurzburg B. A., Hanson J. E., Whitesides G. M., Skehel J. J. and Wiley D. C. (1989) Hemagglutinins from 2 influenza-virus variants bind to sialic acid derivatives with millimolar dissociation-constants - a 500 MHz proton nuclear magnetic resonance study. *Biochemistry* **28** 8388-8396.

Scarsdale J. N., Ram P., Prestegard J. H. and Yu R. K. (1988) A molecular mechanics-NMR pseudoenergy approach to the solution conformation of glycolipids.

Scheffler K., Ernst B., Katapodis A., Magnani J. L., Wang W. T., Weiseman R. and Peters T. (1995) Determination of the bioactive conformation of the carbohydrate ligand in the E-selectin in sialyl Lewis (X) complex. *Angewandte Chemie - International Edition in English* **34** 1841-1844.

Schuler G. D., Altschul S. F. and Lipman D. J. (1991) A workbench for multiple alignment construction and analysis. *Proteins: Structure Function and Genetics* **9** 180-190.

Shaskov A. S., Lipkind G. M., Knirel Y. A. and Kochetkov N. K. (1988) Stereochemical effects of Glycosylation on the C-13 chemical shifts in carbohydrates. *Magnetic Resonance in Chemistry* **26** 735-747.

Shotton D. M. and Watson H. C. (1970) Three-dimensional structure of tosyl-elastase. *Nature* **225** 811.

Sigurskjold B. W., Altman E. and Bundle D. R. (1991) Sensitive titration microcalometric study of the binding of *Salmonella* O-antigenic oligosaccharides by a monoclonal-antibody. *European Journal of Biochemistry* **197** 239-246.

Sigurskjold B. W. and Bundle D. R. (1992) Thermodynamics of oligosaccharide binding to a monoclonal-antibody specific for a *Salmonella* O-antigen point to hydrophobic interactions in the binding site. *Journal of Biological Chemistry* **267** 8371-8376.

Skerra A. and Plückthun A. (1988) Assembly of a functional immunoglobulin Fv fragment in *Escherichia coli. Science* **240** 1038-1041.

Smith G. P. (1985) Filamentous fusion phage: Novel expression vectors that display cloned antigens on the virion surface. *Science* **228** 1315-1317.

Smith-Gill S. J., Mainhart C., Lavoie T. B., Feldmann F. J., Drohan W. and Brooks B. R. (1987) A three-dimensional model of an anti-lysozyme antibody *Journal of Molecular Biology*

Strohal R., Kroemer G., Wick G. and Kofler R. (1987) Complete variable region sequence of a nonfunctionally rearranged kappa light chain transcribed in the nonsecretor P3-X63-Ag8.653 myeloma cell line. *Nucleic Acids Research* **15** 2771.

Tai M., Mudgett-Hunter M., Levinson D., Wu G., Haber E., Oppermann H. and Huston J. S. (1990) A bifunctional fusion protein containing Fc-binding fragment B of staphylococcal protein A amino terminal to antidigoxin single-chain Fv. *Biochemistry* **29** 8024-8030.

Thogerson H., Lemieux R. U., Bock K. and Meyer B. (1982) Further justification for the exo-anomeric effect. Conformational analysis based on nuclear magnetic resonance spectroscopy of oligosaccharides. *Canadian Journal of Chemistry* **60** 44-57.

Torda A. E., Scheek R. M. and van Gunsteren W. F. (1989) Time-dependent Distance Restraints in Molecular Dynamics Simulations. *Chemical Physics Letters* **157** 289-294.

Torda A. E., Scheek R. M. and van Gunsteren W. F. (1990) Time-averaged

nuclear Overhauser effect distance restraints applied to tendamistat. *Journal of Molecular Biology* **214** 223-235.

Tropp J. (1980) Dipolar relaxation and nuclear Overhauser effects in nonrigid molecules: The effect of fluctuating internuclear distances. *Journal of Chemical Physics* **72** 6035-6044.

Tvaroska I., Hricovini, M. and Petrakova, E. (1989) An attempt to derive a new Karplus-type equation of vincinal proton-carbon coupling constants for C-O-C-H segments of bonded atoms. *Carbohydrate Research* **189** 359-362.

Tvaroska I. (1990) Dependence on saccharide conformation of the one-bond and three-bond carbon-coupling constants. *Carbohydrate Research* **206** 55-64.

Tvaroska I. and Bleha T. (1989) Anomeric and exo-anomeric effects in carbohydrate chemistry. Advances in carbohydrate chemistry and biochemistry. *Biochemistry*. **47** 45-123.

Veregin R. P., Fyfe C. P., Marchessault R. H. and Taylor M. G. (1987) Correlation of ^{13}C chemical shifts with torsional angles from high-resolution 13C CP-MAS NMR studies of crystalline cyclomalto-oligosaccharide complexes, and their relation to the structure of the starch polymorphs. *Carbohydrate Research* **160** 41-56.

Vuister G. W. and Bax A. (1993) Measurement of 2-bond and 3-bond proton

to methyl-carbon J-couplings in proteins uniformly enriched with ^{13}C. *Journal of Magnetic Resonance B.* **102** 228-231.

Wall R. and Kuehl M. (1983) Biosynthesis and regulation of immunoglobulins. *Annual Review of Immunology* **1** 393-422.

Ward E. S., Guesow D., Griffiths A. D., Jones P. T. and Winter G. (1989) Binding activities of a repertoire of single imunoglobulin variable domains secreted from *Escherichia coli. Nature* **341** 575-581.

Weber C., Wider G., von Freyberg B., Traber R., Braun W., Widmer H. and Wütrich K. (1991) The NMR structure of cyclosporine-A bound to cyclophilin in aqueous solution. *Biochemistry* **30** 6563-6574.

Weimer T. and Peters T. (1994) Aleuria-aurantia agglutinin recognises multiple conformations of α-L-Fuc-(1-6)-β-D-GlcNac-ome *Angewandte Chemie - International Edition in English* **33** 88-91.

Weiner S. J., Kollman P. A., Nguyen D. T. and Case D. A. (1986) An all-atom forcefield for simulations of proteins and nucleic acids.

Weiner S. J., Kolman P. A., Case D. A., Singh U. C., Ghio C., Alagona G., Profeta jr. S. and Weiner P. (1984) A new force field for molecular mechanical simulation of nucleic acids and proteins. *Journal of the American Chemical Society* **106** 765-784.

Winter G., Griffiths A. D., Hawkins R. E. and Hoogenboom H. R. (1994)

Making antibodies by phage display technology. *Annual Review of Immunology* **12** 433-455.

Wolfe S., Pinto B. M., Varma V. and Leung R. Y. N. (1990) The Perlin effect - bond lengths, bond strengths, and the origins of stereoelectronic effects upon one-bond C-H coupling constants. *Canadian Journal of Chemistry* **68** 1051-1062.

Wood C. R., Boss M. A., Kenten J. H., Calvert J. E., Roberts N. A. and Emtage J. S. (1985) The synthesis and *in vivo* assembly of functional antibodies in yeast. *Nature* **314** 446-449.

Yan Z.-Y. and Bush C. A. (1990) Molecular dynamics simulations and the conformational mobility of blood group oligosaccharides. *Biopolymers* **29** 799.

Yanisch-Perron C., Vieira J. and Messing J. (1985) Improved M13 phage cloning vectors and host strains: Nucleotide sequences of NR1 *in vitro* and *in vivo*. *Gene* **33** 103.

Young N. M., Mackenzie C. R., Narang S. A., Oomen R. P. and Baenziger J. E. (1995) Thermal stabilisation of a single-chain Fv antibody fragment by introduction of a disulphide bond. *FEBS Letters* **377** 135-139.

Zdanov A., Li Y., Bundle D. R., Deng S. J., MacKenzie C. R., Narang S. A., Young N. M. and Cygler M. (1994) Structure of a single chain antibody variable domain (Fv) fragment complexed with a carbohydrate antigen at 1.7-Angstrom resolution. *Proceedings of the National Academy of*

Sciences of the United States of America **91** 6423-6427.

Zhang L. H., Laughlin M. R., Rothman D. L. and Shulman R. G. (1990) 13C NMR relaxation times of hepatic glycogen *in vitro* and *in vivo*. *Biochemistry* **29** 6815-6820.

zu Putlitz J., Kubasek W. L., Duchêne M., Marget M., von Specht B.-U. and Domdey H. (1990) Antibody production in baculovirus-infected insect cells *Bio/Technology* **8** 651-654.