Structural Studies of MUC1 Core Related Peptides

by Martin J. Scanlon, B.Sc., M.Sc.

Thesis submitted to the University of Nottingham

for the degree of Doctor of Philosophy,

May, 1993.

PAGE NUMBERING AS

ORIGINAL

Table of Contents

CHAPTER THREE
Solution Conformations of Linear Peptides

CHAPTER FOUR
Materials and Methods

ABSTRACT

Polymorphic epithelial mucins are large complex glycoproteins which consist of a single polypeptide backbone, a large domain of which is usually made up of degenerate tandem repeats. One such molecule MUC1 is expressed at the surface of human mammary cells and is developmentally regulated and aberrantly expressed in tumours. These mucins have been identified as the target antigens for a number of murine monoclonal antibodies raised against a variety of immunogens including human milk products and breast tumour extracts. The anti-MUC1 antibodies have been shown to display tumour reactivity and have been used both as agents for imaging and in immunoassays to assess tumour burden and response to therapy. A number of these antibodies have been found to define epitopes within the tandem repeat of the protein core of the MUC1.

Hydropathicity calculations and secondary structure predictions on the twenty amino acid tandem repeat of the MUC1 core protein have identified a hydrophilic domain Pro-Asp-Thr-Arg-Pro-Ala-Pro, which has a high probability of turn formation. It is within this domain that the epitopes of the anti-MUC1 antibodies are found. High field n.m.r. studies undertaken on an antigenic twenty amino acid peptide corresponding to the tandem repeat sequence of MUC1 in dimethyl sulphoxide have identified the presence of a type-I β-turn in the region Pro-Asp-Thr-Arg. This turn overlaps the epitopes of all of the MUC1 antibodies characterised to date.

In an attempt to identify more precisely the conformational requirements for binding to two anti-MUC1 antibodies, HMFG1 and HMFG2, the solution structures of several MUC1 core related peptides in dimethyl sulphoxide have been investigated. All of the peptides studied have been found to contain either β-turns or modified turns which overlap their hydrophilic epitope domains. While these observations may provide some explanation for the observed reactivity of the different peptides, they give little insight into the precise structural requirements for antibody binding. Due to

the flexibility of linear peptides in solution it is not possible to define the side chain conformations which are crucial to the processes of antibody recognition.

In order to define the bound conformations of antigenic peptides using n.m.r., it is necessary to undertake experiments in the presence of antibody. Initial experiments have been performed in order to determine the conformation of the MUC1 core related twenty amino acid peptide when bound to the anti-mucin antibody C595. In addition DNA coding for the variable domains of C595 has been cloned and sequenced in order to facilitate both expression of recombinant antibody binding fragments and the modelling of the binding site. These studies should provide a clearer understanding of the structural basis of antibody recognition of the peptides, and may give an insight into the specificity of anti-MUC1 antibodies for malignant cells.

ACKNOWLEDGEMENTS

PUBLICATIONS

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

Scanlon M.J., Price M.R. & Tendler S.J.B. (1990) Immunological and biophysical studies on the peptide repeat of polymorphic epithelial mucins. *Br. J. Cancer*, **62** , 544-545.

Price M.R., Briggs S., Scanlon M.J. & Tendler S.J.B. (1990) Identification and molecular expression of monoclonal antibody defined epitopes on the peptide repeat of polymorphic epithelial mucins. *Br. J. Cancer* , **62**, Suppl. XII, 17.

Scanlon M.J., Price M.R. & Tendler S.J.B. (1990) Structural studies on the twenty amino acid peptide relating to the tandem repeat of human tumour associated polymorphic epithelial mucin. *Br. J. Cancer* , **63**, Suppl. XX, 20.

Price M.R., Briggs S., Scanlon M.J., Tendler S.J.B. Sibley P.E.C. & Hand C.W. (1991) The mucin antigens, what are we measuring? *Dis. Markers*, **9**, 205-212.

Scanlon M.J., Morley S.D., Jackson D.E., Price M.R. & Tendler S.J.B. (1992) Structural and computational investigations of the conformation of antigenic peptide fragments of human polymorphic epithelial mucin. *Biochem. J.*, **284** 137-144.

Scanlon M.J., Price M.R. & Tendler S.J.B. (1992) Peptide epitope binding and fluorescence quenching with the anti mucin antibody C595. *Cancer Lett.*, **63**, 199-202.

Scanlon M.J., Price M.R. & Tendler S.J.B. (1993) Modified β-turns in MUC1 core related peptides. *Int. J. Oncology*, In the press.

The work described in Chapter nine was undertaken in collaboration with Mr. Graeme Denton of Nottingham University.

CHAPTER ONE

ANTIBODY-ANTIGEN INTERACTIONS

1.1 Basic Definitions

Antibodies, or immunoglobulins, are a group of glycoproteins which are present in the serum and tissue fluids of higher animals. Their production is induced when the host's lymphoid system comes into contact with immunogenic foreign substances. Antibodies have a bi-functional role in the immune system. One part is responsible for recognition and binding of infectious agents, while the second activates responses which cause their neutralisation. Antibodies have a specific affinity for the foreign substance that elicited their production.

Antigens are foreign macromolecules capable of inducing antibody formation. The specificity of an antibody is directed against a particular site on the antigen called the antigenic determinant or epitope. In the following discussion only protein and peptide antigens will be considered, as the vast majority of structural data available describes antibodies complexed to these materials.

1.2 Antibody Classification

The basic structure of all immunoglobulins is a unit consisting of two identical pairs of polypeptides termed the heavy- (H) and the light- (L) chains. The chains are linked by disulphide bridges to form a Y-shaped molecule (Figure 1.1). Antibodies are divided into five classes on the basis of the type of H-chain and the number of Y-shaped units they contain (Table 1.1), namely IgG, IgM, IgA, IgD and IgE. These differ from one another in size, charge, amino acid composition and glycosylation. In addition to the five types of H-chain, there are two classes of light chain κ and λ. There are no restrictions on which types of H- and L-chains can form antibodies, however any given immunoglobulin molecule will contain only one type of H-chain and one type of L-chain.

Within a class of antibodies there may also be slight differences in the H-chain sequence which give rise to subclasses. Thus, the murine IgG class of molecules contains four subclasses: IgG1, IgG2a, IgG2b and, IgG3. Differences between the

subclasses within a given immunoglobulin class are less pronounced than the differences between the various classes.
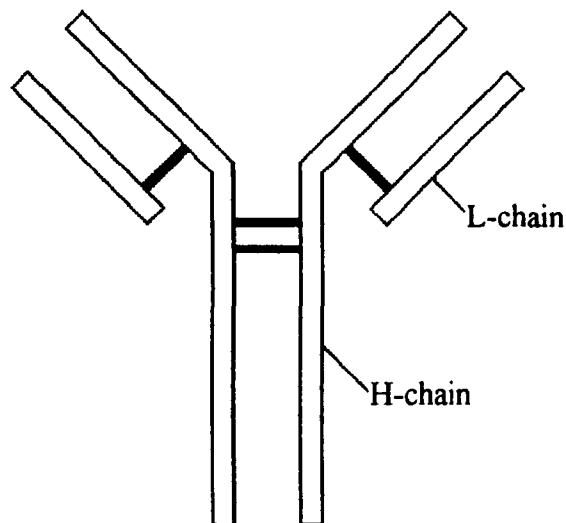
Figure 1.1   Basic structure of the "Y-shaped" antibody subunit.   The bold lines represent interchain disulphide bridges.
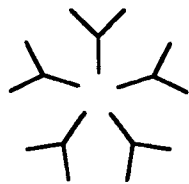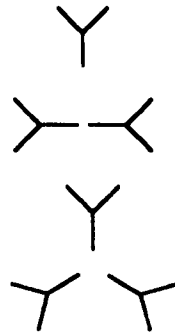
1.3 Antibody Structure

The discussion below focuses on IgG antibodies, which contain only one structural Y-shaped unit, and are the most abundant in normal serum, however, the important structural features of IgG molecules are also present in the Y-shaped units of the other antibody classes.   In immunoglobulins belonging to the IgG class, the H-chain contains approximately 440 amino acid residues and has a molecular weight around 50,000.   L-chains are approximately 220 residues in length and have a molecular weight of around 25,000.   The whole antibody molecule therefore has a molecular weight of 150,000.

IgG antibodies can be split by the protease papain into three components (Porter, 1959 ; Figure 1.2).   Two of these are identical and combine with antigen to form a stable complex.   These are called Fab (fragment antigen binding).   The third component does not combine with antigen, but can be obtained in a crystalline form,

and is termed the Fc (fragment crystallisable). The region between the Fab and the Fc is called the hinge. This segment allows lateral and rotational movement of the Fab arms to facilitate antigen binding.

Table 1.1   Classes of antibody.

Class	IgG	IgM	IgA	IgD	IgE
Heavy Chain	γ	μ	α	δ	ε
Light Chain	κ or λ	κ or λ	κ or λ	κ or λ	κ or λ
Molecular Formula	$\gamma_2\kappa_2$ or $\gamma_2\lambda_2$	$(\mu_2\kappa_2)_5$ or $(\mu_2\kappa_2)_5$	$(\alpha_2\kappa_2)_n$ or $(\alpha_2\lambda_2)_n$*	$\delta_2\kappa_2$ or $\delta_2\lambda_2$	$\varepsilon_2\kappa_2$ or $\varepsilon_2\lambda_2$
Structure					

* n = 1, 2 or 3

Within the antibody molecule there are intrachain disulphide bonds which form loops in the polypeptide chain. These loops fold to form separate domains which contain approximately 110 amino acid residues (Edelman, 1970 ; Figure 1.3). L-chains contain two domains. The N-terminal domain (V_L) shows considerable sequence variation from antibody to antibody, while the C-terminal domain (C_L) is constant in chains of the same type. Sequence variations within the C_L domain account for the differences between κ and λ L-chains. H-chains are made up of a variable domain at the N-terminus (V_H), and three constant domains (C_H1, C_H2 and C_H3). Within a class

of antibody there is considerable sequence homology between the three constant domains.

Within the variable domains of antibodies there are certain sequences having very diverse amino acid sequences (Wu & Kabat, 1970). These regions display variability both in sequence and in the number of amino acid residues they contain.

papain
cleavage
site

Fab fragments

Fc fragment
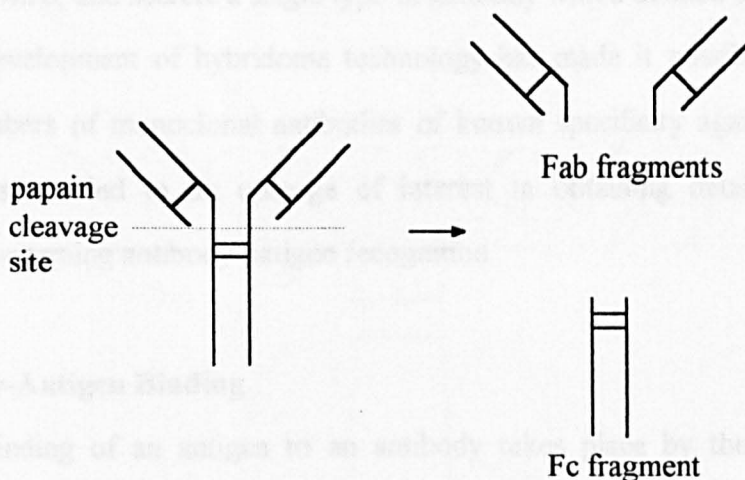
Figure 1.2   Proteolysis of IgG by papain.

Early crystallographic studies of complexes between antibodies and their ligands revealed that these segments come together to produce a continuous hypervariable surface which forms the combining site of the antibody molecule (Amzel *et al*, 1974 ; Padlan *et al*, 1976 ; Amzel & Poljak, 1979). It has long been thought that the specificity of antibody-antigen interactions resides in complementary structures being present on the two molecules (Ehrlich, 1900). Consequently, these hypervariable domains are referred to as the complementarity determining regions (CDRs). The smallest fragment of the antibody which retains its binding capacity for antigen consists of a heterodimer of the V_H and V_L domains and is called Fv (fragment variable).

1.4 Monoclonal Antibodies

Normal serum contains many different types of antibodies that are specific for a wide spectrum of antigens. The use of such a mixed population of immunoglobulins makes it difficult to obtain reproducible sera, and to define antibody specificities.

In 1975, Kohler and Milstein developed a method for immortalising a single antibody secreting B-cell by fusing it to a myeloma cell (a type of B-cell tumour). These hybrid cells (hybridomas), unlike normal plasma cells, can be maintained indefinitely *in vitro*, and secrete a single type of antibody with a defined specificity.

The development of hybridoma technology has made it possible to produce unlimited numbers of monoclonal antibodies of known specificity against any given antigen. This has led to an upsurge of interest in obtaining detailed structural information concerning antibody-antigen recognition.

1.5 Antibody-Antigen Binding

The binding of an antigen to an antibody takes place by the formation of multiple non-covalent bonds between the antigen and amino acids of the binding site. The attractive forces (hydrogen bonds, electrostatic, van der Waals and hydrophobic) involved in these interactions are weak in comparison with covalent bonds. The binding affinity of an antibody for its antigen relies on there being sufficient complementarity between the binding site and the epitope for multiple interactions to occur.

Non-covalent bonds are critically dependent on the distance (d) between the interacting groups. The force is proportional to $1/d^2$ for electrostatic interactions, and to $1/d^7$ for van der Waals forces. Thus the interacting groups must be close in molecular terms before these forces become significant. In order to overcome thermodynamic disruption of the interactions, several non-covalent bonds must form almost simultaneously. Consequently the antigenic determinant and the combining site must have highly complementary structures for binding to occur.

Figure 1.3 Domain structure of IgG antibodies. Disulphide bridges are indicated by bold lines. C and N indicate the C- and N-termini of the polypeptide chains.

1.6 Structural Studies

A fundamental step in characterising the structural basis of antibody-antigen interactions is the identification of the antigenic determinant of the protein. The observation that certain antibodies induced by a native protein do not react with the denatured form whereas others raised against peptide fragments of the primary sequence cross react with the intact protein led to the definition of two classes of antigenic determinant (Atassi, 1975). Sequential or linear epitopes consist entirely of residues which are local in the polypeptide sequence, while conformational or assembled epitopes are composed of residues, distant in the sequence, which are brought together by folding of the protein. It is thought that around 90% of proteins contain conformational epitopes (Atassi *et al*, 1978 ; Barlow *et al*, 1986). A number of predictive methods are now available which attempt to identify epitopes from certain features of the primary structure of the protein. Parameters such as hydrophilicity (Hopp & Woods, 1981 ; Kyte & Doolittle, 1982), flexibility (Karplus & Schultz, 1985) or a combination of these (Jamieson & Wolf, 1988) have been used. None of these predictive algorithms attempt to correlate the antigenicity with the three dimensional structure of the protein, and possibly for this reason, they have not been greatly successful (Pellequer *et al*, 1991).

In cases where the antibody recognises a sequential epitope it is possible to define the antigenic determinant using solid phase peptide synthesis techniques (Geysen *et al*, 1987). This is achieved by producing a set of overlapping peptides of a given length which are homologous with the sequence of the antigen. Antibodies can be tested for their ability to bind the peptides using standard immunological methods. This approach has been used to identify a number of sequential epitopes (Geysen *et al*, 1987 ; O'Sullivan 1990 ; Xing *et al*, 1990, 1991), which have been shown to consist of between 3-8 residues.

Having identified the location and boundaries of the epitope, it is possible to further probe the contribution of the individual residues in the epitope to the overall binding. This is done by producing a set of peptides, based on the defined epitope, in

which each residue in turn is systematically replaced by all of the other 19 naturally occurring amino acids (Geysen *et al*, 1987). If the antibody binding ability of the peptide is lost when the original residue is replaced by amino acids of similar character, then this residue is thought to contribute directly to the interaction. When the binding ability of the peptide is essentially independent of the particular residue present in a certain position, it is assumed that the side chain of this amino acid is not in direct contact with the paratope of the antibody. Such replacement analysis has been successfully applied to determine the fine specificity of a number of MAbs which define sequential epitopes on a breast carcinoma associated mucin (Briggs *et al*, 1991 ; Xing *et al*, 1991).

The problem of defining conformational epitopes is more complex. Early attempts to identify conformational determinants relied on enzymic hydrolysis of the protein to identify successively smaller fragments which retained binding activity. Crumpton & Williamson (1965) localised some of the antigenic sites on myoglobin by obtaining proteolytic fragments which could inhibit antibody binding to the native protein. More recently, site directed mutagenesis has been used to define residues which are critical for antibody binding. Smith-Gill *et al* (1982) found that replacement of Arg^{68} by lysine in hen egg-white lysosyme led to a 1000-fold reduction in binding to the anti-lysosyme antibody HyHEL-5. However, these methods are extremely time consuming, and it is not always apparent whether the loss of binding activity results from the substitution of a residue which is essential for antibody recognition, or if the substitution causes a major change in the structure of the antigen.

1.7 X-ray Crystallography

Recently, a number of crystal structures of Fab fragments bound to protein antigens have been solved (reviewed in Davies *et al*, 1990). These provide a direct method for identifying those residues on both antibody and antigen which are involved in binding, and hence allow precise determination of the antigenic determinant. In addition they provide information regarding the degree of complementarity in the

complex, and in cases where structural data is available for the free antigen, address the question of to what extent conformational changes in the binding region contribute to the interaction.

Of the high resolution structures currently determined, three involve complexes with lysosyme (Amit *et al*, 1986 ; Sheriff *et al*, 1987 ; Padlan *et al*, 1989), and two with influenza virus neuraminidase (Colman *et al*, 1987 ; Tulip *et al*, 1989). In the case of lysosyme, a high resolution crystal structure of the free protein is also available (Ramandhan *et al*, 1989).

The Fabs in the three anti-lysosyme complexes describe epitopes that are almost entirely non-overlapping, which together cover nearly half of the surface area of the protein. Therefore it appears that any part of the protein surface which is accessible is potentially antigenic. In all three structures the interface at the combining site covers an area of around 750Å^2, which is sufficient to accommodate approximately 15-18 residues from the antigen, and the epitopes defined by the antibodies are discontinuous. In the crystal structure of the HyHEL-5 complex (Sheriff *et al*, 1987), the Arg^{68} residue was contained in the epitope and was found to interact with four of the six CDRs. In the Fab-lysosyme complexes studied to date, there appears to be no major conformational change in either the antibody combining site or the backbone atoms of the antigen on binding. However, larger changes have been observed for side chains, sometimes involving rotations of aromatic ring systems (Padlan *et al*, 1989), presumably corresponding to an induced fit of the two species.

The Fab-neuraminidase complexes form a slightly larger interface than those observed with lysosyme. Again the epitopes are discontinuous, but they involve 17-20 residues from the antigen and form a surface area of around 850Å^2.

In all five of the structures, the antigen interacts with either five or six of the antibody CDRs, and all of the combining sites contain a relatively large number of aromatic residues. Van der Waals contacts, hydrogen bonds and salt bridges contribute to the stabilisation of all of the complexes. A striking feature of the interaction of these Fabs with their protein antigens is the degree of structural

complementarity between the two surfaces. In all five cases this results in an almost total exclusion of water from the interface. It is believed that the entropic gain caused by the water exclusion contributes significantly to the stabilisation of the complex (Kauzmann, 1959).

To date, relatively few Fab-protein structures have been determined at high resolution. This is partly to do with the fact that the Fab fragments, which are generally obtained by proteolytic cleavage with papain (Porter, 1959), exhibit considerable microheterogeneity. However with the development of molecular biology technology it is now possible to express recombinant Fab (Better *et al*, 1988) and Fv (Riechmann *et al*, 1988) fragments of antibodies. These methods of protein production have the distinct advantage that extremely homogeneous preparations of antibody fragments can be obtained. With improvements in the determination of Fab-protein complex structures from crystallographic data (Cygler & Anderson, 1988) this should lead to the solution of a larger number of structures of antibody fragments, both free and bound, which will facilitate a more thorough investigation of the structural requirements for antibody-antigen binding.

1.8 Definition of Epitopes

There is an apparent discrepancy in size between the epitopes defined by X-ray crystallography, which contain 15-22 amino acid residues, and those identified by peptide scanning methods, which involve less than eight. However, an important distinction has to be drawn between the two techniques. The stereochemical information obtained from crystallographic data is entirely structural and gives no indication of the relative contribution of individual atomic interactions to the overall binding. The use of synthetic peptides identifies those residues which are directly involved in the binding interaction. Neighbouring residues are thought to be important in the intact protein to enable the contact residues to adopt their final conformation, although they do not contribute in a positive way to complex formation (Geysen *et al*, 1987). This view is supported by molecular modelling calculations which suggest that

only a small number of amino acids, as few as 3-5, contribute the majority of the binding energy of the interaction (Novotny *et al*, 1989).

Consequently, the requirements for antibody binding can be described in terms of either a functional epitope, which contains those residues which interact strongly with the combining site, or a structural epitope, which includes all those residues which are spatially close to the binding site in the complex.

1.9 Nuclear Magnetic Resonance Spectroscopy

Nuclear magnetic resonance (n.m.r.) spectroscopy is a powerful tool for the investigation of the conformation and dynamics of small proteins and peptides in solution. Using two-dimensional (2D) n.m.r. techniques it is now possible to obtain structures at a resolution comparable to that achieved by X-ray crystallography of proteins up to a molecular weight of around 15,000. Although antibody-protein systems are too large for study by these methods, it is still possible to obtain a good deal of information, at somewhat lower resolution on the nature of the interactions within these complexes where the antigen involved is either a peptide (Arata *et al*, 1986 ; Anglister *et al*, 1988), or a small protein (Paterson *et al*, 1990). Even in these cases, to obtain the simplest n.m.r. spectrum possible, it is advantageous to work with the smallest easily obtainable fragment of antibody which retains its affinity for antigen, usually the Fab. A variety of different methods have been used to study these interactions (Anglister *et al*, 1988 ; Kato *et al* 1989 ; Cheetham *et al*, 1991 ; Tsang *et al*, 1991), which yield information both on the identity of those residues in the antigen with which the antibody interacts and on the different strengths of interaction in the binding region.

1.9.1 High Affinity Antibody-Peptide Complexes

Due to the large size difference between antibodies (150,000 M.W.) and peptides (1000-3000 M.W.), their respective n.m.r. spectra are very different. Antibody proton resonances are broad with line widths >20Hz, whereas those of

peptide protons are narrow <5Hz. In a high affinity complex, the peptide is in slow exchange between the bound and free states, and those residues of the antigen which interact strongly with the combining site become immobilised on binding. Thus in an n.m.r. spectrum of the complex, the resonances of the strongly interacting peptide residues assume the line width characteristics of antibody resonances, and are considerably broadened compared to those in the free peptide. Such changes in linewidth can be observed in a 2D-correlated spectroscopy (COSY) experiment (Aue *et al*, 1976), since the cross peak intensity in the COSY experiment is dependent upon the line width (Weiss *et al*, 1984). Hence it is possible to correlate cross peak intensities of resonances with the mobility of the residues from which they derive (Figure 1.4).

Cheetham *et al* (1991) have used this method to identify the interacting residues of a 28-residue peptide derived from hen egg white lysosyme, with the Fab of an anti-peptide antibody, and observed a number of strongly interacting residues. In the intact lysosyme molecule, for which an X-ray crystal structure is available (Handoll, 1985), the residues identified as being tightly bound in the complex, although distant in the primary sequence, are grouped together in the tertiary structure to form a hydrophobic projection on the surface of the molecule. Consequently this method provides a means of identifying the degree of mobility of residues in a peptide antigen bound within the combining site of a high affinity antibody. Since this mobility can be correlated with the nature of the peptide interactions with the antibody in the complex, it offers an insight into the conformational and dynamic features of antibody specificity.

1.9.2 Low Affinity Antibody-Peptide Complexes

For low affinity complexes, where the peptide is in fast exchange between the bound and free states, those residues which interact with the antibody combining site can be identified using transferred nuclear Overhauser effect (TRNOE) spectroscopy (Balaram *et al*, 1972). By acquiring 2D TRNOE spectra at different molar ratios of Fab : peptide, it is possible, by subtraction, to obtain a spectrum in which all the cross

peaks represent magnetisation transfers from antibody combining site to bound peptide protons (Anglister *et al*, 1988). This provides information on the distances between protons of the antibody and peptide in the bound state. The number of NOEs observed in such experiments is generally insufficient to define the conformation of the complexes. Therefore, distance constraints derived from the n.m.r. data are used in conjunction with molecular modelling to obtain a three dimensional structure of the antibody bound peptide.
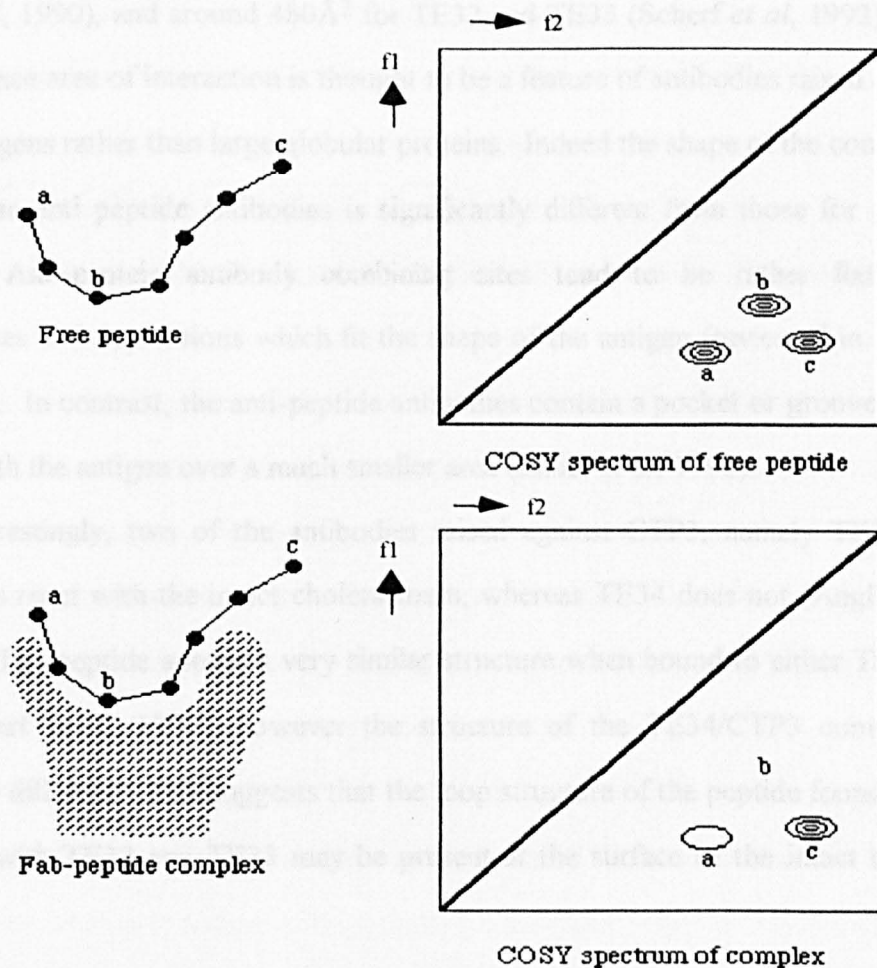
Figure 1.4 Schematic illustration of the study of high affinity Fab-peptide complexes using 2D COSY. In the complex, (a) represents a weakly bound residue; (b) a tightly bound residue; and (c) an unbound residue. The effect of binding on the intensity of the COSY cross peaks is shown. f1 and f2 represent the two frequency axes in the n.m.r. spectrum.

These methods have been used to study a number of complexes involving a peptide derived from cholera toxin, CTP3, with Fab fragments of three antibodies, TE32, TE33 and TE34, raised against it (Anglister *et al*, 1988, 1989, 1990 ; Levy *et al*, 1989 ; Zilber *et al*, 1990). In all three complexes the peptide folds into a loop on binding to the antibody fragment, and interacts with four or five of the CDRs. This increases the surface area of contact between antibody and antigen, thereby potentially increasing the binding energy. The surface area of the interaction is 380Å^2 for TE34 (Zilber *et al*, 1990), and around 480Å^2 for TE32 and TE33 (Scherf *et al*, 1992). The smaller surface area of interaction is thought to be a feature of antibodies raised against peptide antigens rather than large globular proteins. Indeed the shape of the combining sites of these anti peptide antibodies is significantly different from those for protein antigens. Anti-protein antibody combining sites tend to be rather flat, with protuberances and depressions which fit the shape of the antigen (reviewed in Wilson *et al*, 1991). In contrast, the anti-peptide antibodies contain a pocket or groove which interacts with the antigen over a much smaller area (Zilber *et al*, 1990).

Interestingly, two of the antibodies raised against CTP3, namely TE32 and TE33, cross react with the intact cholera toxin, whereas TE34 does not (Anglister *et al*, 1988). The peptide adopts a very similar structure when bound to either TE32 or TE33 (Scherf *et al*, 1990), however the structure of the TE34/CTP3 complex is significantly different. This suggests that the loop structure of the peptide found in the complexes with TE32 and TE33 may be present at the surface of the intact cholera toxin.

1.9.3 Isotope Edited N.M.R. Studies

Recent advances in methodology have extended the applicability of n.m.r. spectroscopy to proteins with a molecular weight of up to 30,000 (Clore & Gronenbaum, 1992). With the development of recombinant DNA technology, which has enabled the expression of Fv fragments of antibodies in *Escherichia coli* (Huston

et al, 1988) or myeloma cells (Reichmann *et al*, 1988), it should now be possible to obtain high resolution structures of Fv and Fv - peptide complexes from n.m.r. data.

Early studies of an engineered anti-lysosyme Fv fragment confirmed that unambiguous resonance assignments could be obtained for proteins of this size by isotopically enhancing the protein with ^{15}N and ^{13}C labels (Wright *et al*, 1990). More recently, Takahashi *et al* (1992) have investigated the conformational changes in the combining site of an anti-dansyl monoclonal antibody upon antigen binding. In this case the antigen interacts strongly with only one of the six CDRs, and although considerable changes occur in the quaternary structure of this region upon antigen binding, it is not yet clear to what extent this is a general phenomenon, or whether it is a special case for antibodies binding to small molecules.

Isotope edited techniques have also been used successfully to investigate the conformation and dynamics of peptide antigens when bound to antibody (Tsang *et al*, 1990, 1992). By labelling the peptide with ^{15}N and ^{13}C it is possible to selectively observe the resonances of peptide protons in the combining site. Tsang *et al* (1992) have shown that a peptide derived from myohemerythrin adopts a helical backbone conformation in the combining site of an anti-peptide Fab (B13A2). In the X-ray crystal structure of myohemerythrin, the region corresponding to the peptide is also helical (Fieser *et al*, 1987). This may explain the observed cross-reactivity of B13A2 with the intact protein. In addition, it has been reported that the free peptide in aqueous solution has a propensity to form a nascent helix (Dyson *et al*, 1988). Thus it appears that on binding to antibody the peptide is stabilised in a helical conformation.

1.10 Concluding Remarks

The ability to characterise the structure and dynamics of antibody combining sites and antigens, both free and in complexes, is of fundamental importance to an understanding of antibody specificity. Improvements in the methodology, both for obtaining binding fragments of antibodies, and in determining their conformation, have enabled new insights to be gained into the structural aspects of antibody-antigen

complexes. The increased availability of engineered Fv fragments and the further refinement of these techniques should facilitate a greater understanding of the molecular basis of antibody-antigen interactions.

CHAPTER TWO

POLYMORPHIC EPITHELIAL MUCINS

2.1 General Characteristics of Mucins

Many epithelial cells are covered by a protective secretion which serves as a selective barrier between the plasma membrane and the environment of the cell. This secretion, usually called mucus, is a highly hydrated gel comprising a large number of components. The predominant macromolecules in this secretion are the mucins.

The mucins can be divided into two general categories. Secretory mucins constitute the major component of the viscous gels which cover the mucosal secretions of respiratory, gastro-intestinal and reproductive tracts. The secretory mucins have the ability to form intermolecular disulphide bridges resulting in the formation of oligomeric structures, which confer on the mucus its protective and lubricative properties (Allen, 1983). Membrane-bound mucins are structurally similar, however they exist in a monomeric form and their polypeptide backbone has a hydrophobic segment which serves to anchor the molecule in the plasma membrane (Gendler *et al*, 1990). The precise function of these membrane bound mucins remains unclear.

2.2 Physical and Chemical Properties of Mucins.

Mucins are high molecular weight glycoproteins which consist of a non-globular polypeptide backbone (Eckhardt *et al*, 1987) to which oligosaccharides are attached via the hydroxyl groups of serine and threonine. The central portion of the protein core contains regions which are especially rich in serine and threonine residues, virtually all of which are glycosylated, while the terminal parts of the backbone consist of a more normal distribution of amino acids. All mucins have a similar amino acid composition. They are rich in aliphatic amino acids, but contain a low percentage of aromatic and sulphur containing residues (Herp *et al*, 1979 ; Carlstedt *et al*, 1983). Serine, threonine, alanine, glycine and proline particularly dominate the backbone, comprising around 50% of the amino acids.

The majority of the sugars on mucins are attached to the hydroxyl oxygen of serine and threonine residues via N-acetyl galactosamine (GalNAc) (Figure 2.1a). The oligosaccharide chains vary greatly in length and consist of up to 20 residues (Gendler

et al, 1988). The pattern of glycosylation is non-uniform, with some regions being heavily glycosylated, while others are virtually free of carbohydrate chains (Roberts, 1976). Five different monosaccharide units are commonly found, namely N-acetyl galactosamine, N-acetyl glucosamine (GlcNAc), galactose (Gal), fucose (Fuc) and sialic acid (SA). In addition the mucin contains variable amounts of sulphate, which is normally attached to Gal, GalNAc or GlcNAc. The presence of the sialic acid and sulphate moieties confer a negative charge on these molecules, which are thought to be important in maintaining the extended rod-like conformations seen in the mature mucins (Section 2.7). N-linked oligosaccharides have been shown to be present in some mucins (Denny & Denny, 1982 ; Hilkens & Buijs, 1988). These N-linked glycans are normally attached through the amide nitrogen of asparagine to GlcNAc (Figure 2.1b). However these represent only a minor fraction of the total carbohydrate content.
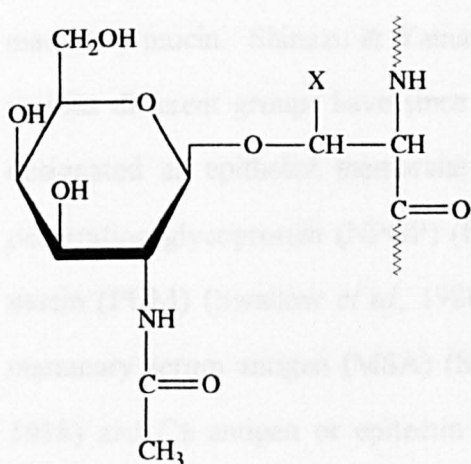
Figure 2.1a   N-Acetyl Galactosamine linked to a serine (X=H) or threonine (X=CH$_3$) residue.
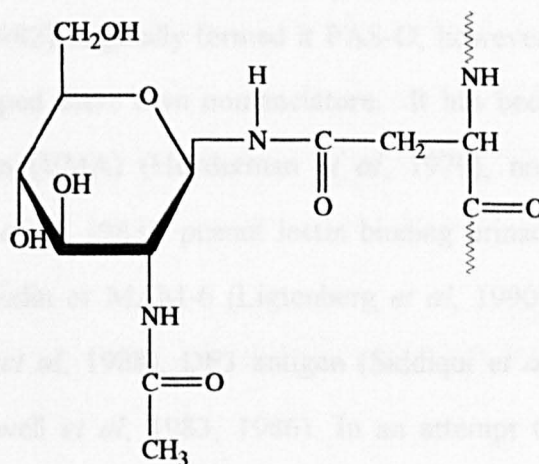
Figure 2.1b   N-Acetyl Glucosamine linked to an asparagine residue.

Secretory mucins are distinct from the membrane bound molecules in that they contain cysteine residues in the terminal parts of the protein core which allow them to oligomerise through the formation of disulphide bridges (Allen, 1983). In contrast membrane bound molecules exist as monomers, and contain a hydrophobic trans-membrane domain which anchor the molecule in the membrane (Gendler *et al*, 1990).

2.3 Human Mammary Mucin

In recent years there has been considerable interest in a group of human epithelial mucins. This has primarily been due to the fact that antibodies directed to specific epitopes on these molecules have been widely used for both *in vivo* and *in vitro* diagnosis of cancer (Table 2.1). One such mucin, derived from mammary epithelia has been widely investigated.

2.4 Nomenclature

There is considerable disagreement in the literature as to the name of the mammary mucin. Shimizu & Yamauchi (1982) originally termed it PAS-O, however, various different groups have since developed there own nomenclature. It has been designated as epithelial membrane antigen (EMA) (Heyderman *et al*, 1979), non penetrating glycoprotein (NPGP) (Ceriani *et al*, 1983), peanut lectin binding urinary mucin (PUM) (Swallow *et al*, 1986), episialin or MAM-6 (Ligtenberg *et al*, 1990), mammary serum antigen (MSA) (Stacker *et al*, 1988), DF3 antigen (Siddiqui *et al*, 1988) and Ca antigen or epitectin (Bramwell *et al*, 1983, 1986). In an attempt to resolve the confusion the name polymorphic epithelial mucin (PEM) was suggested (Gendler *et al*, 1988), as this was believed to describe the biochemical characteristics of the molecule. However, since the discovery of a number of other mucins with similar properties a new terminology has been developed. The family of mucins are now designated MUC, followed by a number, which reflects the order in which the gene coding for the protein was sequenced (Gum *et al*, 1990). Thus, the mammary

mucin is MUC1, since this was the first mucin for which partial cDNA clones were obtained and sequenced.

Table 2.1   Clinical applications of anti-MUC1 antibodies.

Technique	Application	Reference
Immunohistology	Identification of cancer cells in biopsy specimens	reviewed in Zotter *et al*, 1988
Radioimmunoassay	Diagnostic and prognostic indicator for advanced breast cancer	reviewed in Kenemans *et al*, 1988
Radio-immunoscintigraphy Radio-guided surgery	Imaging of ovarian tumours	Symonds et al, 1992

2.5 Biochemical Aspects of MUC1

In normal tissues MUC1 is found on the apical borders of highly specialised secretory epithelial cells, however in malignancy it appears to be up-regulated and aberrantly expressed (Burchell *et al*, 1987), and has been shown to be present in elevated levels in the serum of patients with advanced breast cancer (Price *et al*, 1988). Consequently, antibodies raised against the mucin have a number of applications in the management of cancer patients (Table 2.1).

MUC1 exhibits the general features of a mucin glycoprotein, having a high molecular weight, and containing a high proportion of carbohydrate, which is O-linked to serine and threonine residues in its single polypeptide backbone.   The protein has been shown to exhibit extreme size polymorphism which is demonstrable at the DNA (Swallow *et al*, 1986) and protein (Gendler *et al*, 1988) level.   Full length cDNA clones coding for the polypeptide backbone of MUC1 have now been obtained and sequenced (Gendler *et al*, 1990), and the protein core contains a number of distinct regions (Figure 2.2).   The bulk of the core consists of varying numbers of a highly conserved twenty amino acid tandem repeat sequence, the tandem repeat continues

into the terminal domains of the protein core becoming increasingly less degenerate. The N-terminal sequence has a hydrophobic signal peptide, which is probably cleaved in the mature mucin and the C-terminal domain contains a hydrophobic membrane spanning sequence and a cytoplasmic tail (Gendler *et al*, 1990). The basis of the polymorphism lies in the expression of different numbers of the tandem repeat. Analysis of genomic DNA from a variety of individuals has revealed that MUC1 may contain between 25-125 copies of the tandem repeat (Gendler *et al*, 1988).

Figure 2.2 Diagram of the MUC1 core protein showing the N-terminal signal peptide (▨), tandem repeat region (■), C-terminal transmembrane domain (▨) and cytoplasmic tail (☰). N and C refer to the N- and C-termini of the polypeptide chain.

The amino acid composition of the protein core is dominated by the composition of the tandem repeat (Table 2.2), and is typical of a mucin glycoprotein. Proline, alanine, glycine, serine and threonine account for around 60% of the amino acids. All of the cysteine residues in the backbone are contained within the membrane spanning domain, consequently MUC1 does not have the capacity to form disulphide-linked oligomeric structures (Gendler *et al*, 1990). The tandem repeat sequence (Figure 2.3) contains 25% serine and threonine residues, which gives this portion of the molecule the potential to be highly glycosylated. Potential O-glycosylation sites are also present outside the tandem repeat region, and five putative N-glycosylation sites have been identified in the C-terminal sequence (Gendler *et al*, 1990).

Table 2.2   Amino acid composition of the translated DNA sequence of MUC1

containing 42 tandem repeats.

Amino Acid	Composition (%)	Amino Acid	Composition (%)
Alanine	16.1	Leucine	2.7
Arginine	4.2	Lysine	0.8
Asparagine	1.4	Methionine	0.3
Aspartic acid	4.5	Phenylalanine	1.3
Cysteine	0.2	Proline	19.0
Glutamic acid	1.0	Serine	12.2
Glutamine	1.3	Threonine	13.5
Glycine	8.9	Tryptophan	0.2
Histidine	4.5	Tyrosine	1.0
Isoleucine	0.9	Valine	6.0

The pattern of glycosylation of MUC1 is extremely heterogeneous, with carbohydrate chains differing in both length and sugar composition. Unlike N-linked glycans, where oligosaccharides are attached to asparagine residues in the sequence Asn-X-Ser/Thr, where X can be any amino acid other than proline or aspartic acid (Marshall, 1972), no consensus sequence for O-glycosylation has been identified. However it has been suggested that the serines and threonines should be adjacent to either themselves or one another in order to be glycosylated (Timpte et al, 1988) and that proline must reside near to glycosylation sites (Takahashi et al, 1984 ; Briend et al, 1981). Four of the five serine and threonine residues in the tandem repeat comply with these requirements (Figure 2.3). Recently, it has been shown that the mucins expressed by pancreatic and ovarian tissues have the same polypeptide backbone as the mammary mucin. However the pancreatic mucin has reported molecular weights of up to 1,000,000 and is more extensively glycosylated (Lan et al, 1987). Consequently it appears that the glycosylation is tissue specific and may not be determined entirely on the basis of the protein sequence.

$$\overset{\bullet\ \bullet}{P\ D\ T\ R\ P\ A\ P\ G\ S}\ \overset{}{T}\ A\ P\ P\ A\ H\ G\ V\ \overset{\bullet\ \bullet}{T\ S}\ A$$

1   20

Figure 2.3   Sequence of the twenty amino acid tandem repeat indicating probable O-glycosylation sites (•).

It is thought that MUC1 is expressed in an under-glycosylated form by tumours. Analysis of the carbohydrate chains of normal milk mucin have been found to contain between 8-14 oligosaccharides (Hanisch *et al*, 1989), whereas those of mucin from a breast cancer cell line consisted of as few as two (Hull *et al*, 1989). Further indirect evidence for this has come from the observed pattern of reactivity of monoclonal antibodies raised against either deglycosylated mucin, or peptides relating to the tandem repeat. These antibodies have been shown to react with mucin from tumour cells but not the normal mucin (Girling *et al*, 1989 ; Perey *et al*, 1992), supporting the idea that in the cancer-derived material, aberrant glycosylation results in the exposure of epitopes in the MUC1 protein core (Figure 2.4). Clearly a knowledge of the precise determinants identified by the anti-MUC1 antibodies may allow identification of tumour associated epitopes and could yield information regarding the pattern of glycosylation of mucin obtained from different sources.

2.6 Epitope Mapping of Anti-MUC1 Antibodies

MUC1 has been identified as the target antigen for a number of murine monoclonal antibodies raised against a variety of immunogens such as human-milk products (Taylor-Papadimitriou *et al*, 1981 ; Burchell *et al*, 1987) and breast tumour extracts (Ellis *et al*, 1984). Some of the antibodies appear to recognise tumour associated epitopes and there has been considerable interest in characterising these determinants as this may provide a clearer understanding of the specificity of the antibodies for malignant cells.

A number of anti-MUC1 antibodies recognise epitopes within the carbohydrate chains of the mucin (O'Sullivan, 1990). A thorough consideration of the carbohydrate epitopes expressed by MUC1 is beyond the scope of this thesis, and has been comprehensively reviewed elsewhere (Zotter *et al*, 1988). It has been found that many antibodies react with deglycosylated mucin preparations (O'Sullivan, 1990), indicating that their epitopes are contained within the protein core of MUC1. Hence it appears that even in the intact mucin, which is extensively glycosylated, regions of the polypeptide backbone are accessible to antibody binding. Since the sequence of the MUC1 core protein was reported (Gendler *et al*, 1988), it has been possible to define the precise determinants of a number of these antibodies.

Table 2.3   Epitopes of a number of anti-MUC1 monoclonal antibodies.

Antibody	Epitope	Reference
BC1, BC2, BC3	APDTR	Xing *et al*, 1991
SM3	PDTRP	Burchell *et al*, 1989
		Price *et al*, 1990b
HMFG1	PDTR	Burchell *et al*, 1989
HMFG2, EMA M8,	DTR	Burchell *et al*, 1989
RINA 9/22, RINA 5/2		Price *et al*, 1990b
		Taylor-Papadimitriou *et al*, 1991
BrE-2,BrE-3	TRP	Taylor-Papadimitriou *et al*, 1991
onc-M15	TRPA	Burchell *et al*, 1989
Ca-2	TRPAP	Price *et al*, 1990b
NCRC-11	RPA	Price *et al*, 1990b
C595,789/91,F36/22	RPAP	Burchell *et al*, 1989
		Price *et al*, 1990a
		Taylor-Papadimitriou *et al*, 1991

Using a series of tethered synthetic peptides it has been shown that several antibodies recognise epitopes contained within the tandem repeat domain of the MUC1

protein core (Xing *et al*, 1990 ; Briggs *et al*, 1992). The presence of multiple copies of the tandem repeat in the intact mucin give this region the potential to be highly immunogenic. The epitopes consist of 3-5 amino acid residues within a discreet region of the repeat sequence, APDTRPAPG (Table 2.3). Interestingly, all of the residues contain the central arginine residue of this sequence, however, it appears that they can be subdivided into two major categories. Ten of the antibodies contain the motif DTR within their epitopes. Interestingly, all of these antibodies were prepared against either human milk derived products or breast glycoprotein. The majority of the remaining antibody epitopes contain the sequence RPA. These antibodies were raised against either purified mucins or tumour tissue. Whether the accessibility of the different immunogenic epitopes varies in mucin derived from different sources has yet to be determined.

It is clear that the interaction of antibodies with as few as three sequential residues of the tandem repeat is fundamental to the process antigen recognition and binding. Indeed it has been shown that it is possible to replace individual amino acids contained in the epitopes of some antibodies with structurally unrelated residues without affecting the binding (Xing *et al*, 1990 ; Briggs *et al*, 1991, 1992). Hence it appears that not all of the residues defined in the epitope are directly involved in the binding process. However, estimates of the size of antibody binding sites obtained from X-ray crystallography (Section 1.7) and n.m.r. data (Section 1.9) suggest that additional factors are almost certainly involved in antibody recognition of the MUC1. Residues flanking the defined epitopes and carbohydrate side chains present in the intact mucin are likely to impose conformational restraints on the determinant which affect its presentation and recognition by antibody. It has been shown that substitution of the proline residue in the tetrapeptide PDTR by any other residue results in a loss of binding to the antibody HMFG2 (Briggs *et al*, 1991), even though the proline is not contained within the defined epitope (Table 2.3). This highlights the importance of presentation of the epitope in the correct conformation for antibody binding.

It is not yet known to what extent the carbohydrate side chains present in the intact mucin contribute to the overall binding. Since the immunodominant sequence in the tandem repeat contains one possible O-glycosylation site and is flanked by several others it is likely that glycosylation of these residues will restrict the accessibility of this sequence to antibodies. Aberrant or under-glycosylation of the carcinoma-derived mucin is believed to result in more of the protein core becoming accessible. This is thought to account for the observed tumour specificity of the antibody SM3 which defines the epitope PDTRP (Gendler *et al*, 1990). It has been suggested that the threonine residue in the immunodominant sequence cannot be glycosylated in the intact mucin since this residue is contained in the epitopes of a number of anti-MUC1 antibodies. However, replacement experiments based on the technique of Geysen *et al* (1987) have revealed that substitution of the threonine by almost any other amino acid has little effect on the binding of antibodies HMFG1 and HMFG2 (Briggs *et al*, 1991). Consequently, it may be possible to accommodate a carbohydrate chain at this position.

Further studies on the structure of the mucin and peptides based on the tandem repeat sequence may provide a clearer insight into the conformational requirements for antibody recognition and binding.

2.7 Structural Studies of MUC1

Due to the size and complexity of the intact MUC1 molecules, attempts to define their structure in detail have been only partially successful. Electron microscopy studies of MUC1 molecules have suggested that the mucin is rod-shaped (Slayter & Codington, 1973), random coils (Rose *et al*, 1984) or single non-linear extended strands (Bramwell *et al*, 1986). It is not clear to what extent the sample preparation in these studies affects the observed structure, however it appears that the molecules exist in extended conformations. More recently, images of MUC1 obtained from ovarian carcinoma ascitic fluid have been imaged using scanning tunnelling microscopy (STM) (Roberts *et al*, 1992). Unlike electron microscopy, STM has the advantage that

biomolecules can be directly observed under ambient conditions. The ovarian carcinoma MUC1 were found to be rod-shaped molecules of mixed length which aggregate along their long axis (Roberts *et al*, 1992). Aggregation of membrane bound mucin in this manner has previously been suggested as the mechanism by which the MUC1 molecules protect the underlying epithelia (Bramwell *et al*, 1983).

Although these studies provide information on the surface topography of the mucin molecules, the resolution of the techniques is not sufficient to probe the structural features of the antibody binding domain found in the tandem repeat. Furthermore the large size and heterogeneity of the intact mucin make it unsuitable for investigation by more conventional techniques such as X-ray crystallography and n.m.r. spectroscopy. Consequently a number of structural investigations have been carried out on polypeptide sequences derived from the MUC1 protein core.

Hydropathicity calculations (Kyte & Doolittle, 1982) performed on a 60 amino acid polypeptide corresponding to three copies of the MUC1 tandem repeat sequence (Figure 2.3) indicated the presence of a hydrophilic domain extending over seven residues, PDTRPAP (Price *et al*, 1990a). These results suggest that this region is likely to be expressed at the surface of the intact MUC1 molecule where it would be accessible to antibody binding. The hydrophilic domain overlaps the identified immunogenic region of the MUC1 core protein. Secondary structure predictions (Chou & Fasman, 1978) indicate that within the hydrophilic domain there is a high potential for β-turn formation (Price *et al*, 1990a). It has been suggested that secondary structure motifs such as β-turns may be involved in antibody recognition of antigens (Williamson *et al*, 1986 ; Tendler, 1990 ; Dyson *et al*, 1990). N.m.r. studies on peptides based on the tandem repeat sequence have identified the presence of a number of secondary structure features including a type I β-turn in the hydrophilic domain which overlaps the epitopes of all the protein-core-reactive anti-MUC1 antibodies characterised to date (Tendler, 1990 ; Scanlon *et al*, 1992). Further investigations of the structure of peptides based on the MUC1 tandem repeat sequence, which have been shown to react with anti-MUC1 antibodies (Xing *et al*,

1991 ; Briggs *et al*, 1991, 1992 ; Scanlon *et al*, 1992), may provide an insight into the conformational requirements for antibody recognition and a clearer understanding of the basis of the tumour specificity of some of the antibodies.

CHAPTER THREE

SOLUTION CONFORMATIONS OF LINEAR PEPTIDES

3.1 Overview

Until relatively recently it was generally believed that short linear peptides existed as random structures in solution. Early theoretical calculations (Brant *et al*, 1967) and experimental methods using spectroscopic techniques (Epand & Scheraga, 1968 ; Hermans & Puett, 1971) confirmed this view. However an apparent exception to this rule was discovered when Brown & Klee (1971) showed that under certain conditions the S-peptide, the N-terminal 20 residues of ribonuclease, gave a circular dichroism (CD) spectrum which indicated a low population of α helix. Subsequently a whole range of peptides have been found to exhibit local conformational preferences in solution (reviewed in Wright *et al*, 1988). It is now widely accepted that peptides exist in solution as a number of rapidly interconverting conformers. If one of these conformations is sufficiently long lived it may be possible to detect it by a number of techniques. A variety of spectroscopic methods have been used to investigate the solution conformations of linear peptides including circular dichroism (CD), vibrational spectroscopy (Fourier transform infra-red (FTIR), Raman) and n.m.r. If the polypeptide backbone is assumed to consist of planar *trans* peptide bonds linked via the α-carbon atom, then possible conformations of a polypeptide can be described by the angles ϕ, ψ and χ (Figure 3.1).
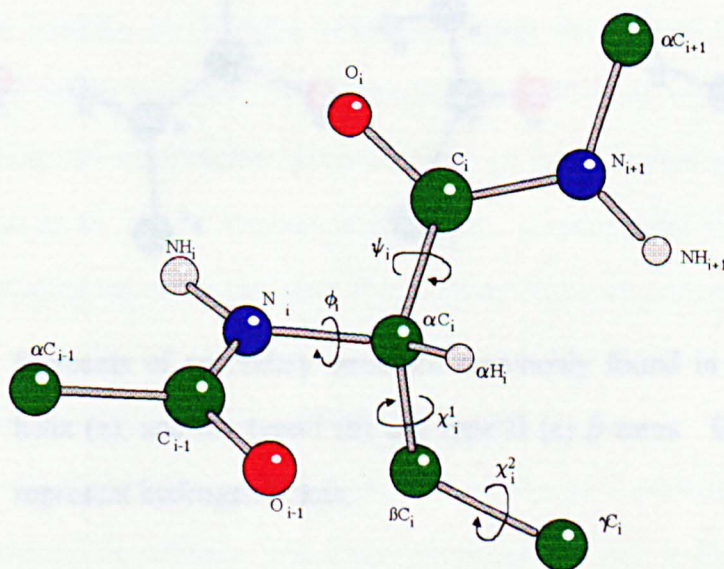
Figure 3.1   Standard nomemclature for atoms and bond angles in peptides.

Peptides in solution normally adopt a limited range of structures, these are the random coil, - a term which implies that no single conformation is particularly stable - the α helix, and the type-I and type-II β-turns (Figure 3.2).
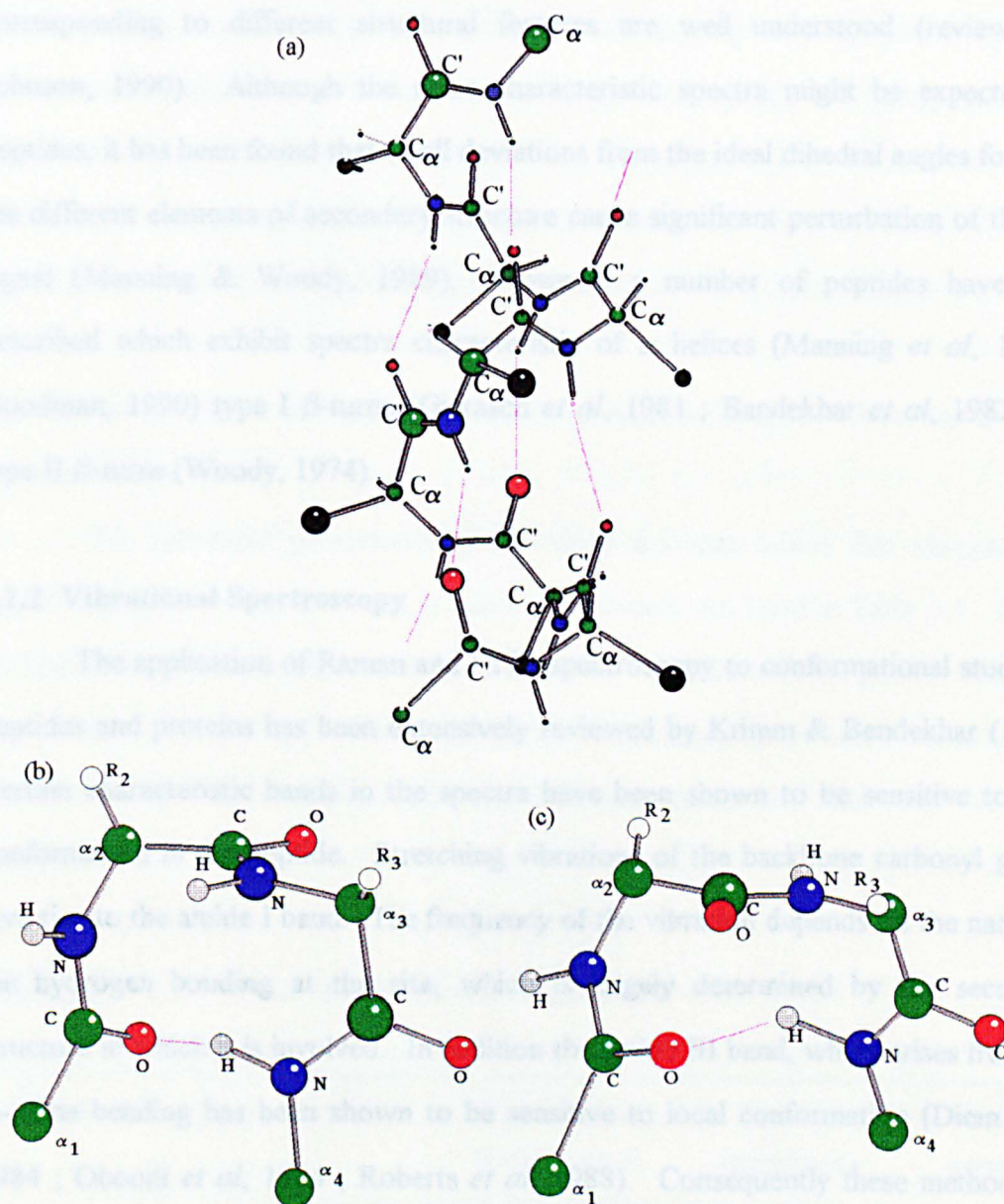
Figure 3.2 Elements of secondary structure commonly found in peptides, the α helix (a), and the type-I (b) and type-II (c) β-turns. Dotted pink lines represent hydrogen bonds.

3.1.1 Circular Dichroism

CD spectroscopy is a widely used technique for determining elements of secondary structure present in polypeptides. For large proteins characteristic spectra corresponding to different structural features are well understood (reviewed in Johnson, 1990). Although the same characteristic spectra might be expected for peptides, it has been found that small deviations from the ideal dihedral angles found in the different elements of secondary structure cause significant perturbation of the CD signal (Manning & Woody, 1989). However a number of peptides have been described which exhibit spectra characteristic of α helices (Manning *et al*, 1977 ; Goodman, 1990) type I β-turns (Gierasch *et al*, 1981 ; Bandekhar *et al*, 1982) and type II β-turns (Woody, 1974).

3.1.2 Vibrational Spectroscopy

The application of Raman and FTIR spectroscopy to conformational studies of peptides and proteins has been extensively reviewed by Krimm & Bendekhar (1986). Certain characteristic bands in the spectra have been shown to be sensitive to local conformation in the peptide. Stretching vibrations of the backbone carbonyl groups give rise to the amide I band. The frequency of the vibration depends on the nature of the hydrogen bonding at the site, which is largely determined by the secondary structure in which it is involved. In addition the amide III band, which arises from NH in-plane bending has been shown to be sensitive to local conformation (Diem *et al*, 1984 ; Oboodi *et al*, 1984 ; Roberts *et al*, 1988). Consequently these methods can provide information regarding elements of secondary structure present.

3.1.3 Nuclear Magnetic Resonance Spectroscopy

N.m.r spectroscopy is probably the most widely used technique for elucidation of peptide structures in solution. The major advantage of n.m.r. is that the information it provides is site-specific. An important difference between n.m.r. and the other forms of spectroscopy lies in the timescales of the different techniques. The various forms of

optical spectroscopy can distinguish conformations which interconvert on a sub-picosecond timescale, therefore they are capable of providing information on the relative populations of different elements of secondary structure present in a peptide in solution. N.m.r. spectroscopy is only capable of detecting exchange processes which occur on the millisecond-to-second timescale, and consequently it is rarely possible to deduce the populations of the different conformations directly from n.m.r. data. However if one of the conformers in the conformational ensemble sampled by the peptide is particularly stable, then it may be possible to characterise using n.m.r.

3.2 Characterisation of Peptide Structure by Nuclear Magnetic Resonance Spectroscopy

The important parameters of the three isotopes which are ubiquitous in peptides and are widely observed in n.m.r. experiments are listed in Table 3.1. Due to its high natural abundance the ^1H nucleus can be detected with far greater sensitivity than either ^{13}C or ^{15}N. Consequently the vast majority of work relies on the observation of ^1H resonances and the following discussion relates to ^1H-n.m.r. spectroscopy.

Table 3.1   Nuclear properties of selected isotopes.

Isotope	Spin (I)	Observation frequency at 11.74T (MHz)	Natural abundance	Relative sensitivity
^1H	1/2	500.0	99.98	1.00
^{13}C	1/2	125.7	1.11	1.8×10^{-4}
^{15}N	1/2	50.7	0.37	3.8×10^{-6}

The determination of peptide conformations in solution relies upon the interpretation of spectral parameters such as chemical shift (δ), coupling constants (J)

and nuclear Overhauser effects (NOEs). The main problem in interpreting these parameters is that the data obtained reflects a population weighted average of all the conformations present in the peptide. Since the different parameters are averaged in a non-linear way, it is not usually possible to deduce the nature of all of the constituent conformers from the averaged data. For example, the intensity of a NOE between two nucleii depends not on the internuclear distance (r), but on r^{-6}, thus a minor conformer can still give a strong NOE even after conformational averaging. Therefore when attempting to identify a preferred structure from the n.m.r. data, it is necessary to adopt a cautious approach, acquiring as much information as possible and ensuring that evidence for a particular conformation comes from as many different sources as possible. The following section contains a brief description of those ^{1}H n.m.r. parameters which are fundamental to the problem of structure determination.

3.2.1 Chemical Shift

The chemical shift (δ) of a proton resonance depends upon its chemical environment. Thus the position of a resonance in the spectrum is characteristic of the type of proton that it is associated with (Figure 3.3). In an unstructured peptide, protons have chemical shifts which depend upon the amino acid type and values of these so-called random coil shifts have been tabulated (Wüthrich, 1986). Deviations of the chemical shift from these values may suggest the presence of a preferred structure in the peptide, although they provide no indication of what that structure may be.
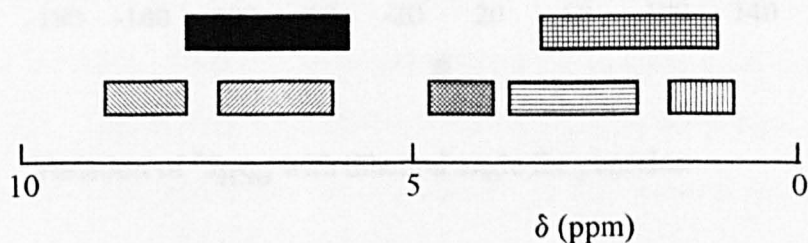
Figure 3.3   Typical resonance positions of protons in common amino acid residues. backbone NH; side chain NH; α; β; methyl; aromatic ring; other aliphatic.

3.2.2 Temperature Dependence of Amide Proton Chemical Shifts

The amide proton temperature coefficient ($\Delta\delta/\Delta T$) is the rate at which an amide proton resonance moves upfield with increasing temperature. Solvent exposed amide protons have a coefficient of 6-8 ppb/K. A lower value suggests that the proton is shielded from solvent in some way, and is widely used as a measure of the extent of hydrogen bonding (e.g. Williamson *et al*, 1986 ; Dyson *et al*, 1988a ; Tendler 1990). Since the different elements of secondary structure have characteristic patterns of hydrogen bonding (Figure 3.2) this parameter can provide information regarding preferred conformations in a peptide.

3.2.3 Coupling Constants

The value of the vicinal coupling constant (^3J) varies in a complicated fashion depending upon the angle between the two protons (Pardi *et al*, 1984 ; Figure 3.4).
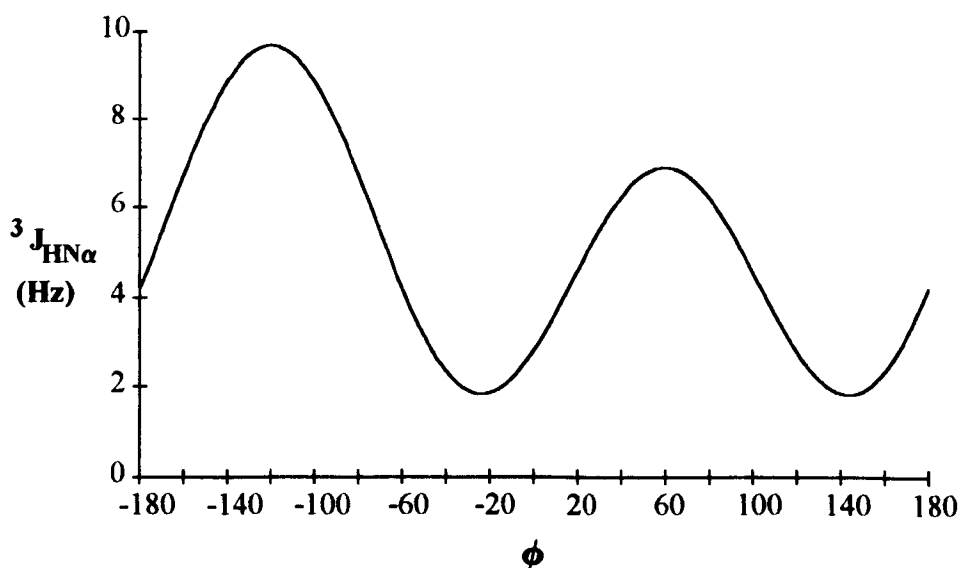
Figure 3.4   Variation of $^3J_{HN\alpha}$ with dihedral angle for peptides.

The coupling constant between the amide and α protons ($^3J_{HN\alpha}$) of an amino acid residue varies with the angle ϕ (Figure 3.1). Since the backbone conformation of a peptide can be described by the angles ϕ and ψ the coupling constant can provide

valuable structural information. However it must again be stressed that since rotations about single bonds, and hence changes in the angle between two protons, occur on a very much faster timescale than the n.m.r. experiment, the measured coupling constant is an average value. So although there are characteristic values of $^3J_{HN\alpha}$ for particular secondary structures (Table 3.2), even if the observed value suggest the peptide is unstructured it does not preclude the presence of minor populations of stable conformers.

Table 3.2   Characteristic angles and coupling constants in secondary structures.

Structure	Residue (i+1)			Residue (i+2)		
	ϕ	ψ	$^3J_{HN\alpha}$	ϕ	ψ	$^3J_{HN\alpha}$
α helix	-57	-47	3.9			
random coil			6.5-8.5			
type-I β turn	-60	-30	4.6	-90	0	7.9
type-II β turn	-60	120	4.6	80	0	6.2

3.2.4 Nuclear Overhauser Effects

The NOE is the most important parameter in defining structure as it provides direct information on the distance between two protons. If one of the proton resonances is pre-irradiated, it causes a change in the intensity of the resonance line observed for the other. Such a change in the intensity of a signal in an n.m.r. spectrum is called a NOE. The intensity of a NOE is highly sensitive to the distance between the two protons involved, and varies as r^{-6}, where r is the distance between the two protons. The observation of a NOE between a given pair of protons in a linear peptide indicates the presence of at least a threshold population of conformers in which these protons are separated by less than approximately 3.5Å (Dyson *et al*, 1988a, 1988b). The distances between two hydrogen atoms A and B located in the amino acid residues

at positions i and j in the sequence of a polypeptide is conventionally denoted $d_{AB}(i,j)$ as shown in Table 3.3. In practice the distances between certain protons in the peptide are of particular interest, and these are denoted as shown below:

$$d_{\alpha N} - \text{distance between } \alpha H_i \text{ and } NH_{i+1}$$

$$d_{NN} - \text{distance between } NH_i \text{ and } NH_{i+1}$$

$$d_{\beta N} - \text{distance between } \beta H_i \text{ and } NH_{i+1}$$

Within the elements of secondary structure commonly found in peptides there are characteristically short interproton distances (Table 3.3). Thus an analysis of the pattern of NOEs provides direct information on conformations present in a peptide. However as the NOE data contains contributions from a variety of different conformers, great care has to be taken in the interpretation of the data to ensure that the observed NOEs result from a single conformation.

Table 3.3   Short distances in secondary structures. Those distances which are short enough for an NOE to be observed in a linear peptide are underlined.

Distance (Å)	α helix	Type-I β-turn	Type-II β-turn
$d_{\alpha N}$ (i,i+1)	<u>3.5</u>	<u>3.4</u>	<u>2.2</u>
$d_{\alpha N}$ (i+1,i+2)		<u>3.2</u>	<u>3.2</u>
$d_{\alpha N}$ (i,i+2)	4.4	3.6	<u>3.3</u>
$d_{\alpha N}$ (i,i+3)	<u>3.4</u>	3.1-4.2	3.8-4.7
$d_{\alpha N}$ (i,i+4)	4.2		
d_{NN} (i,i+1)	<u>2.8</u>	<u>2.6</u>	4.5
d_{NN} (i+1,i+2)		<u>2.4</u>	<u>2.4</u>
d_{NN} (i,i+2)	4.2	3.8	4.3

3.3 Sequence Specific Assignments

Before any of the parameters described above can be interpreted, it is necessary to assign the individual resonance lines in the spectrum to specific protons in the peptide. The assignment procedure relies upon the analysis of scalar coupling patterns and NOEs (Wüthrich, 1986). In n.m.r. spectra of peptides the observation of ^1H-^1H couplings is limited to protons which are separated by two covalent bonds (geminal couplings, ^2J) or three covalent bonds (vicinal couplings, ^3J). As the protons on adjacent residues in a peptide are separated by a minimum of four covalent bonds, scalar ^1H-^1H couplings are only observed between protons within the same amino acid residue. A group of resonances which are connected by scalar couplings are referred to as a spin system. Thus the protons of each amino acid residue constitute one or several spin systems.

3.3.1 Spin System Nomenclature

Amino acid spin systems are conventionally denoted by a series of upper case letters corresponding to protons with different chemical shifts (Pople *et al*, 1959). The proton which resonates at the highest field is listed first, and the remaining spins are placed in order from the periphary towards the αH. Brackets indicate that the spin network is branched and those spins in the parentheses are listed in the direction away from the αH. Neighbouring letters in the alphabet indicate strongly coupled spins where the chemical shift difference between the two resonances ($\Delta\delta$) is comparable to the coupling constant (J). Conversely non-neighbouring letters indicate weakly coupled spins ($\Delta\delta \gg J$). Thus for proline which has the spin system $A_2(T_2)MPX$ (Table 3.4) the two γH are equivalent and resonate at the highest field and are denoted A_2. The spin network is branched as the γH are coupled to both the βH and the δH. If weak coupling is assumed throughout then the degenerate δH are labelled T_2, in parentheses as they are further away from the αH, and MPX represent the non-degenerate βH and the αH respectively. In the aromatic rings of Phe and Tyr there are two pairs of symmetry related protons, 3H and 5H, and 2H and 6H, which have the

same chemical shift but different couplings with other spins in the aromatic system. Such chemical shift equivalence is denoted by AA'. The spin systems for the non-exchangeable protons in the different amino acid residues are listed in Table 3.4.

Table 3.4   Side chains (R) of the twenty common amino acids, with three-letter and one-letter notation and spin systems for the non-labile protons. The structure for Pro includes the backbone CαH and N atoms.
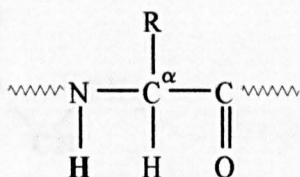
Amino acid residue. R is the side chain (see below). Labile protons are shown in bold type

Pro, P
A$_2$(T$_2$)MPX

Phe, F
AMX + AA'MM'X

Tyr, Y
AMX + AA'XX'

Trp, W His, H
AMX + A(X)MP + A   AMX + AX
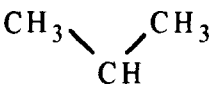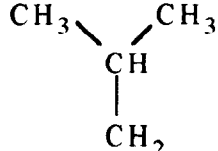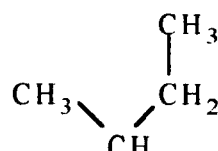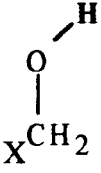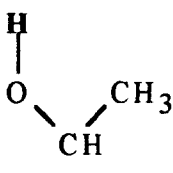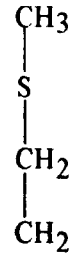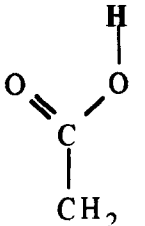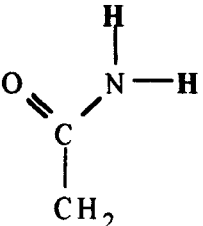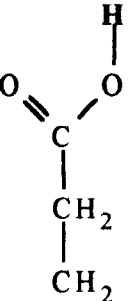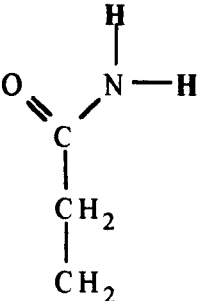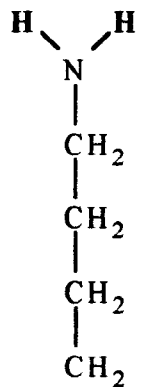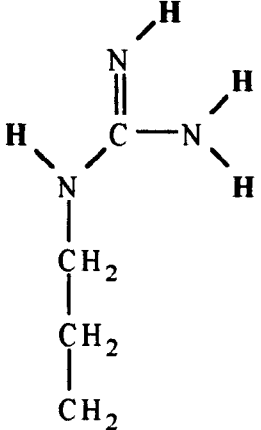
Table 3.4 (Continued).

H	CH_3	Val, V	Leu, L	Ile, I
Gly, G	Ala, A			
AB	A_3X	A_3B_3MX	A_3B_3MPTX	$A_3MPT(B_3)X$

Ser, S — AMX

Thr, T — A_3MX

Cys, C — AMX

Met, M — $AM(PT)X + A_3$

Asp, D — AMX

Asn, N — AMX

Glu, E — $AM(PT)X$

Gln, Q — $AM(PT)X$

Lys, K   Arg, R

$A_2(F_2T_2)MPX$ $A_2(T_2)MPX$

Figure 3.5 COSY and TOCSY connectivity diagrams for the spin systems of non-labile residues of the common amino acid residues. Cross peaks are shown on one side of the diagonal only. Open circles represent those cross peaks observed in a COSY experiment, while both filled and open circles are observed in TOCSY spectra.

3.3.2 Correlated Spectroscopy

The first stage in the assignment of peptide resonances is to identify the different spin systems present. This can normally be achived by using two-dimensional correlated spectroscopy (COSY) (Aue *et al*, 1976) and/or total correlated spectroscopy (TOCSY or HOHAHA) (Braunschweiler & Ernst, 1983). Spectra acquired in these experiments show correlations between coupled spins. Different spin systems give rise to characteristic patterns of connectivity (Figure 3.4), and with a

knowledge of values of the random coil chemical shifts (Wüthrich, 1986), it is normally possible to assign all of the resonances to protons of a particular amino acid type from these data.

3.2.3 Sequential Assignments

Sequence -specific assignments depend on the observation of through-space $d_{\alpha N}$ interactions (NOEs) between the α proton of one amino acid and the amide proton, or in the case of proline the δ-protons, of the next. For L-amino acids, this distance varies between approximately 2.20 and 3.55Å (Wüthrich, 1986), and so these connections can normally be seen for all residues in the peptide. Using this information it is possible to identify groups of spin systems corresponding to unique sequences in the primary structure of the peptide. Thus sequence specific assignments are obtained by matching peptide segments identified from the NOE data to the independently determined amino acid sequence.

NOEs can be measured using either normal longditudinal cross relaxation pathways with one-dimensional (1D) and 2D-NOESY experiments (Jeener et al, 1983 ; Bodenhausen et al, 1984)), or via the rotating-frame NOE, called the ROE, in 1D CAMELSPIN (Bothner-By et al, 1984) or 2D-ROESY (Bax & Davis, 1985) experiments. A major factor affecting the size of the NOE is the tumbling rate of the interproton vectors concerned, which is largely determined by the size of the peptide. For linear peptides in non aqueous solvents it is often the case that the tumbling rate is such that it is necessary to measure ROEs.

3.4 Conformation of Antigenic Peptides

A number of investigations have shown that antigenic peptides have a tendency to adopt folded structures in solution (e.g. Dyson et al, 1985, 1986, 1988a, 1988b ; Williamson et al, 1986 ; Waltho et al, 1989 ; Tendler, 1990 ; Scanlon et al, 1992). In several cases the ordered region overlaps the defined epitope to which the antibody binds (Williamson et al, 1986 ; Tendler, 1990 ; Scanlon et al, 1992), suggesting that

these structural features my play an important role in antibody recognition of the peptide. Indeed it has been suggested that the antigenicity of a peptide is directly related to its ability to adopt a stable conformation in solution (Dyson *et al*, 1992).

The choice of solvent for n.m.r. studies is crucial to obtaining a meaningful result, as peptides often adopt different structures in different solvents. Most peptides are normally found in aqueous environments, and water is therefore widely used. However the binding domains of antibodies have been found to contain a high proportion of aromatic residues and are therefore less polar (Section 1.7), consequently, a less polar solvent may give a more relevant result. Infact Dyson *et al* (1988b) found that in aqueous solution a 19 residue peptide corresponding to the C-helix of myohemerythrin adopted a number of rapidly interconverting turn-like structures which they termed a nascent helix. These structures were stabilised into a regular helical conformation in water/trifluoroethanol mixtures. More recently it was shown that the same peptide adopted a helical conformation when bound to the Fab fragment of an anti-peptide antibody (Tsang *et al*, 1992). Therefore it appears that the use of hydrophobic solvents which more closely resemble the environment of the antibody binding site, may be valuable in determining the likely conformation of peptides complexed to antibody. Further studies of the conformation of antigenic peptides, both free and in complexes, may help to provide an understanding of the molecular basis of antibody specificity.

CHAPTER FOUR

MATERIALS AND METHODS

4.1 Reagents

All reagents were either of analytical grade or, for those described in Section 4.3.10, molecular biology grade.

4.1.1 Aldrich Chemical Co., Gillingham, Dorset, U.K.

Acetic ($^2H_3$) acid

Acetic ($^2H_3$) acid sodium salt

Deuterium oxide (D_2O)

Dimethyl formamide (DMF)

$^2H_6$-dimethyl sulphoxide (d_6-dmso)

Hydrogen peroxide (H_2O_2)

Sodium dodecyl sulphate (SDS)

Tetramethyl silane (TMS)

3-Trimethylsilyl propane sulphonic acid - sodium salt (TPSS)

4.1.2 BDH Biochemical Reagents, Poole, Dorset, U.K.

Acrylamide

Anhydrous disodium hydrogen orthophosphate

2-Azino-di-(3-ethyl benzthiazoline-6-sulphonic acid)-diammonium salt (ABTS)

Bis-acrylamide

Bromophenol blue

Chloroform

Citric acid

Ethanol (100%) - diluted where appropriate

Glacial acetic acid

Glucose

Glycerol

Hydrochloric acid (HCl)

Isopropanol

Methanol

Sodium azide

Sodium chloride (NaCl)

Sodium hydroxide (NaOH)

Sucrose

Xylene cyanol FF

4.1.3 Sigma Chemical Co., Poole, Dorset, U.K.

3-Amino-9-ethyl-carbazole (3-AEC)

Ammonium persulphate

Boric acid

Bovine serum albumin (BSA)

Casein

Ethidium bromide

Ethylene diamine tetra acetic acid (EDTA)

Light mineral oil

Ovalbumin

Papain

Phenol chloroform isoamyl alcohol

Polyoxyethylenesorbitan monolaurate (Tween 20)

Sigmacote™

N,N,N',N'-tetramethylethylene diamine (TEMED)

Thiomersal

Trihydrous sodium acetate

Tris-(hydroxymethyl) amino methane (Tris base)

4.1.4 Pharmacia, Uppsala, Sweden.

Random oligotide hexamers (N_6-hexamers)

Sephacryl S-200

Sepharose 4B-Protein A

4.1.5 USB Cambridge Bioscience, Cambridge, U.K.

Sequinase™ DNA sequencing kit

Urea

4.1.6 Peptides

Peptide P(1-20) was the kind gift of Celltech Ltd. (Slough, Berks., U.K.) the remainder of the peptides described were obtained from the Biopolymer Synthesis and Analysis Unit (Department of Biochemistry, QMC, Nottingham, U.K.). Their purity was established to be >98% by reversed phase h.p.l.c.

4.2 Buffers

All buffers were prepared using de-ionised double distilled water.

4.2.1 Phosphate Buffered Saline (PBS) - pH 7.3

PBS tablets were obtained from Oxoid (Basingstoke, Hants). A 150mM solution was prepared by dissolving ten tablets in one litre of water. The solution was filtered through a Whatmann 0.2μm nitrocellulose membrane (Scientific Lab Supplies, Nottingham, U.K.).

4.2.2 PBSA - pH 7.3

Where a preservative was necessary, sodium azide was added to PBS buffer to give a final concentration of 0.02% (w/v).

4.2.3 PBS/Tween - pH 7.3

Tween 20 was added to PBS buffer to give a final concentration of 0.1% (v/v).

4.2.4 PBSA - pH 8.0

Phosphate buffer at pH 8.0 was prepared by taking one litre of PBSA prepared as described above, and adjusting the pH to 8.0 with 1M NaOH.

4.2.5 Papain Activation Buffer - pH 6.5

The activation buffer was prepared by dissolving one PBS tablet and 74.4mg of EDTA in 50ml of water. Subsequently, mercaptoethanol (70μl) was added and the pH adjusted to 6.5 with 0.1M HCl. The volume was made up to 100ml with water.

4.2.6 Digestion Buffer - pH 6.5

The digestion buffer was prepared by dissolving one PBS tablet and 74.4mg of EDTA in 90ml of water. The pH was adjusted to 6.5 with 0.1M HCl and the volume made up to 100ml with water.

4.2.7 Casein Buffer - pH 7.3

Casein (100mg) was dissolved in one litre of PBS by stirring overnight at room temperature. If the buffer was to be kept for more than 24h thiomersal was added to a final concentration of 0.005% (w/v) .

4.2.8 Sodium Acetate (Stock Solution)

For a 1M stock solution, trihydrous sodium acetate (13.608g) was dissolved in 80ml of water and the solution adjusted to the desired pH (usually 5.0) by the dropwise addition of glacial acetic acid. This solution was made up to 100ml and diluted to the required concentration prior to use.

4.2.9 Tris/HCl (Stock Solution)

For a 1M stock solution Tris base (121.14g) was dissolved in 900ml of water. The pH of the solution was adjusted to the desired value by addition of 2M HCl and

the solution was made up to one litre with water. Prior to use the stock solution was diluted to the required concentration.

4.2.10 SDS Sample Loading Buffer

Proteins were applied to SDS gels in non reducing sample buffer. This was prepared as follows:

Component	(ml)
0.5M Tris/HCl (pH 6.8)	1.0
glycerol	0.8
10% (w/v) SDS	1.6
0.05% bromophenol blue	0.2
water	4.4
total	8.0

4.2.11 Deuteroacetate buffer - pH 4.4

A stock solution of 0.1M sodium acetate (A) was prepared by dissolving 85mg of acetic-($^2H_3$)-acid sodium salt in 100ml of 90% H_2O/10% D_2O or D_2O alone. A stock solution of 2M acetic acid (B) was prepared by mixing 0.5ml acetic-($^2H_3$)-acid with 4.44ml of 90% H_2O/10% D_2O or D_2O. 1.88ml (A) plus 150µl (B) gave a solution of deuteroacetate buffer in either 90% H_2O/10% D_2O or D_2O at pH 4.4.

4.2.12 Citrate-Phosphate Buffer - pH 4.0

Citric acid (4.53g) and anhydrous disodium hydrogen phosphate (4.06g) were dissolved in water (450ml). The pH was adjusted to 4.0 by dropwise addition of 1M HCl and the volume made up to 500ml with water. The buffer was stored at 4°C for up to one month.

4.2.13 Tris-Boric Acid-EDTA (TBE) - pH 8.0

A 5 x stock solution of TBE was prepared by dissolving Tris base (54g), boric acid (27.5g) and EDTA (3.7g) in water (500ml). The buffer was sterilised by autoclaving and diluted prior to use.

4.2.14 DNA Sample Loading Buffer

A 5 x stock solution of the DNA sample loading buffer was prepared by dissolving bromophenol blue (0.25g) and xylene cyanol FF (0.25g) in 40% (w/v) sucrose solution (100ml).

4.2.15 Solution I

Tris base (3.02g) and EDTA (0.372g) were dissolved in water (80ml) and the pH adjusted to 8.0 by the dropwise addition of 0.1M HCl. Finally, glucose (0.45g) was added and the solution made up to 100ml with water.

4.2.16 Solution II

Solution II was prepared freshly by dissolving SDS (1g) in a NaOH solution (0.2M, 100ml)

4.2.17 Solution III

Potassium acetate (29.45g) was dissolved in water (60ml), glacial acetic acid (11.5ml) added and the volume made up to to 100ml with water.

4.2.18 LB-Medium (Luria-Bertani Medium)

The medium was prepared by dissolving bacto-tryptone (10g, Difco, E. Molesey, Surrey, U.K.), bacto-yeast extract (5g, Difco) and NaCl (10g) in 950ml of water. The pH was adjusted to 7.0 by dropwise addition of 5M NaOH, and the volume made up to 1 litre. The medium was sterilised by autoclaving.

4.3 Methods

4.3.1 Enzyme-Linked Immunosorbent Assay : Epitope Mapping Studies With Synthetic Peptides

The minimum sequence required for binding of monoclonal antibody C595 was determined using a series of fifteen overlapping peptide heptamers tethered at their C-termini to polypropylene pin supports. The peptides were related in sequence to the twenty amino acid tandem repeat of the MUC1 core protein (Gendler *et al*, 1988) (Table 4.1).

Table 4.1 Sequence of the MUC1 tandem repeat, and synthetic peptide heptamers used in ELISA.

	1 20
	G S T AP P AHGVT S AP DT R P AP GS T AP P AHGVT S AP DT R
1	P P AHGVT
2	P AHGVT S
3	AHGVT S A
4	HGVT S AP
5	GVT S AP D
6	VT S AP DT
7	T S AP DT R
8	S AP DT R P
9	AP DT R P A
10	P DT R P AP
11	DT R P AP G
12	T R P AP GS
13	R P AP GS T
14	P AP GS T A
15	AP GS T AP

Each peptide overlapped by six amino acids, so that the heptamers covered a region starting at position 12 of one repeat and continuing through to position 12 of the next. Thus the peptides scanned through the previously identified antibody binding domain of the MUC1 core protein (Xing *et al*, 1990 ; Briggs *et al*, 1991). Antibody binding to the pins was assessed in an enzyme linked immunosorbent assay (ELISA), as originally described by Geysen *et al* (1984). ELISAs were performed in quadruplicate on Falcon 96-well microtitre plates (Becton Dickinson, CA, USA). The pins were arranged in such a way that during each stage of the assay they could be immersed in separate wells on the plate containing either antibody or reagent (Figure 4.1).
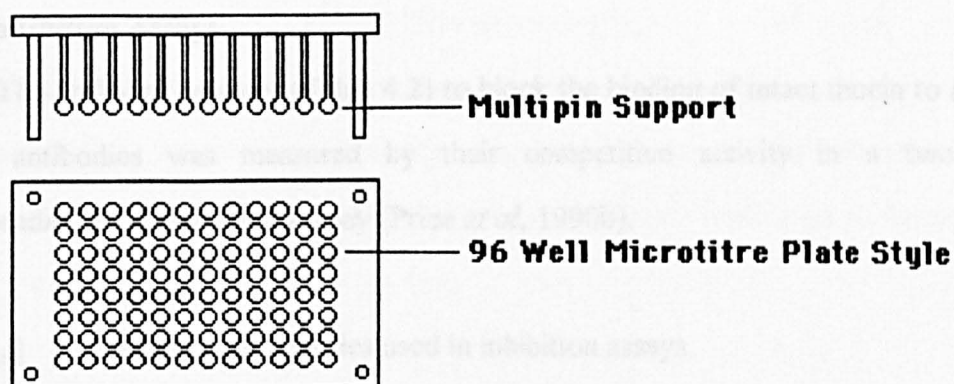
Figure 4.1   Configuration of epitope mapping pins for ELISA performed on 96-well microtitre plates.

The immobilised peptides were incubated for 1h in a solution of 1% (w/v) ovalbumin and 1% (w/v) BSA in PBS/Tween containing 0.02% (w/v) sodium azide, in order to block non-specific adsorption sites. The pins were then incubated overnight at 4°C in 175μl/well of C595 (10mg/ml) in PBSA. After washing (4 x 10min) in PBS/Tween at room temperature with agitation, the pins were incubated for 1h at room temperature in 150μl/well of horse radish peroxidase (HRP) conjugated rabbit-

anti-mouse immunoglobulins (DAKO Ltd., High Wycombe, Bucks.) diluted 1:500 in PBS containing 1% (w/v) BSA. The pins were then washed (4 x 10min) in PBS/Tween at room temperature with agitation, and incubated in the dark with 150µl/well of 0.05% (w/v) ABTS in citrate phosphate buffer pH 4.0, containing 30ml hydrogen peroxide (120 vol) per 100ml of buffer. Incubation was stopped by removing the pins from the wells when sufficient when sufficient colour had developed. The absorbance in the wells at 405nm was read using an Anthos Labtec Instruments 2001 plate reader.

Positive and negative controls for the assay were included in the block of pins in the form of two tetrapeptides, PLAQ and GLAQ. These were tested simultaneously for reactivity with an anti-PLAQ antibody (Cambridge Research Biochemicals, Northwich, Cheshire, U.K.) using the same protocol as described above.

4.3.2 Inhibition Assays

The ability of peptides (Table 4.2) to block the binding of intact mucin to anti-MUC1 antibodies was measured by their competitive activity in a two-site immunoradiometric (sandwich) assay (Price *et al*, 1990b).

Table 4.2   Sequences of peptides used in inhibition assays.

Peptide	Sequence
P(1-20)	PDTRPAPGSTAPPAHGVTSA
P(1-11)	PDTRPAPGSTA
P(7-20)	GSTAPPAHGVTSA
D-peptide	APDTRPAPG
E-peptide	APETRPAPG

The capture antibody (10µl at 10µg/ml in PBSA) or PBSA alone was adsorbed onto the walls of Terasaki Microtest plates (Becton Dickinson) for 18h at 4°C. The wells were aspirated and washed four times with casein buffer, then incubated for 1h at 25°C in casein buffer to block non-specific binding. Portions (5µl) of the peptides at various concentrations (1mg/ml - 10µg/ml) plus casein buffer (5µl) were added to the wells and the plates incubated for 2h at 37°C. The plates were aspirated and the wells washed four times with casein buffer. Portions (5µl) of the peptide solutions at the same concentrations as in the previous incubation plus purified mucin (5µl in PBSA, approx 10µg/ml as estimated from antigen-antibody titrations (Price et al, 1990b)) were added to the wells and the plates incubated for a further 30min at 4°C. The plates were aspirated and washed four times with casein buffer. ^{125}I-labelled NCRC-11 antibody (as the radioactive tracer to detect mucin captured on the solid phase) was added to the wells at 10µl per well (corresponding to 1.5 x 10^5 c.p.m). After incubation for 1h at 20°C, the wells were aspirated and the plates washed six times with casein buffer. The wells were separated with a band saw and the radioactivity in each well was determined using an LKB γ-radiation counter. Tests were performed in triplicate and the mean ± s.d. was recorded.

4.3.3 Radio-Immunoassay

The capacity of C595 antibody and its Fab fragments to bind to urinary mucin was measured using an indirect radioisotopic antiglobulin assay (Price et al, 1990b). Purified urinary mucin (10µl at approx 10µg/ml) was added to the wells of Terasaki microtest plates and air dried overnight at 4°C. Non specific adsorption sites were blocked by incubation with casein buffer for 1h at 25°C. The wells were aspirated and then incubated with either whole C595 antibody, its Fab fragments (10µl at 10µg/ml in casein buffer in each case) or cell free supernatant (P3NS0, 10µl) for 1h at 25°C. The wells were aspirated and washed four times with casein buffer before being incubated with ^{125}I-labelled F(ab')$_2$ fragments of rabbit anti-mouse immunoglobulins (DAKO) at 105 c.p.m./10µl per well (1h at 25°C). After washing six times with casein buffer the

wells were separated with a band saw and the radioactivity in each well was determined using an LKB γ-radiation counter. Tests were performed in quadruplicate and the mean ± s.d. was recorded.

4.3.4 Papain Digestion of C595

C595 antibody, purified from tissue culture supernatants (Price *et al*, 1991) was digested using a column of immobilised papain. This was prepared by mixing excess papain with Affi-Gel 10 (1ml, BioRad, Herts, U.K.). The papain was activated with phosphate buffer (150mM) at pH 6.5 containing mercaptoethanol (10mM) and EDTA (2mM). Mercaptoethanol was removed from the column with phosphate buffer (10ml, 150mM) at pH 6.5 containing EDTA (2mM). Antibody C595 (5.28mg) in digestion buffer was cycled through the column at 10ml/h. Aliquots (10µl) were removed from the column over a period of 24h and the extent of antibody digestion determined by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) and sensitive silver staining. The optimum digestion time was taken to be the point at which whole antibody could no longer be detected on the SDS-PAGE gel using the silver stain. Fab preparations were undertaken as described above except that the proteolysis was allowed to proceed for 18h after which time the crude digest was collected, the residue washed from the column with digestion buffer (2ml), and the fragments purified.

4.3.5 Purification of Fab Fragments

Fab fragments of C595 were purified by passing the digest sequentially through a Sepharose 4B-protein A column and a Sephacryl S-200 gel exclusion column at 50ml/h in PBSA at pH 8.0. Before purification the proteolytic digest was dialysed overnight into PBSA at pH 8.0 in order to improve binding of the Fc containing antibody fragments to the protein A. Fractions were collected and assayed for protein content by measuring their absorbance at 280nm on a Pye Unicam SP8-400 UV/Visible spectrophotometer. Fab containing fractions were pooled, dialysed into

PBSA at pH 7.3 concentrated, and analysed by Western blotting. The final Fab concentration was determined using $E_{1\%}(280nm) = 15.3$ (Hudson & Hay, 1989).

4.3.6 SDS-Polyacrylamide Gel Electrophoresis

SDS-PAGE (Laemmli, 1970) gels were run on a Pharmacia Phastsystem apparatus according to the manufacturers instructions. Protein samples were diluted with an equal volume of non-reducing SDS loading buffer and loaded onto 12.5% polyacrylamide gels. After electrophoresis, the gels were developed in the Phastsystem development unit using a sensitive silver stain (Oakley *et al*, 1980) according to the manufacturers instructions. The molecular weights of proteins were estimated by comparison of their migration with that of the following standards obtained from BRL, Uxbridge, Middlesex.

Protein	Molecular Weight
myosin (H-chain)	200,000
phosphorylase B	97,000
bovine serum albumin	68,000
ovalbumin	43,000
carbonic anhydrase	29,000
β-lactoglobulin	18,400
lysosyme	14,300

4.3.7 Western Blotting

Western blots were performed on a Pharmacia Phastsystem using the supplied semi-dry electrophoretic transfer apparatus according to the manufacturers instructions. After transfer the nitrocellulose membranes (Scientific Lab Supplies) were blocked with casein buffer for 2h at room temperature then incubated for with

horse radish peroxidase conjugated sheep-anti-mouse-kappa-light-chain antibodies (DAKO, diluted 1/1000 in casein buffer) for 1h at room temperature. The substrate was prepared by mixing 3-AEC (1.6ml of a stock solution at a concentration of 0.4% (w/v) in DMF) with sodium acetate buffer (50mM, 38ml at pH 5.0) and PBS/Tween (2ml). Hydrogen peroxide (120 vol., 10μl) was added to the substrate buffer immediately prior to development.

4.3.8 Fluorescence Quenching

Fluorescence measurements were carried out at room temperature on a Perkin Elmer LS5 luminescence spectrophotometer. The optimum excitation and emission wavelengths was determined by recording excitation and emission spectra of the Fab over the range 260-400nm. The excitation wavelength used was 282nm with a slit bandwidth of 2.5nm, and the fluorescence intensity was recorded at an emission wavelength of 362nm with a 10nm bandwidth. The concentration of Fab used in the quenching experiments was 3×10^{-7} M. The maximal quenching value was determined using a 30-40 molar excess of peptide P(1-20) to ensure complete saturation of the binding sites. No correction for peptide concentration was required as P(1-20) contains no chromophoric group that absorbs light in the wavelength range 280-380nm. Titrations were performed by adding 1μl aliquots of a 1mM solution of the peptide to 1ml of a solution of the Fab in the appropriate buffer. Corrections for dilution were applied by performing blank titration in which 1μl aliquots of buffer were added to 1ml of a similar solution of Fab, and these data were subtracted from the corresponding peptide titration values.

4.3.9 N.M.R. Spectroscopy

For experiments undertaken in d_6-dmso, samples contained approximately 4mM peptide and 0.09mM tetramethylsilane (tms). ^{1}H-n.m.r. experiments were undertaken at 400MHz, 500MHz and 600MHz on Bruker AM400, AM500 and AMX600 spectrometers respectively. Over the range of approximately 2-10mM, no

concentration dependent change in the chemical shift of the resonances was observed for any of the peptides. Chemical shift values (δ), accurate to 0.01ppm are reported relative to the tms resonance ($\delta = 0.00$). In order to prepare the D-and E-peptides in their protonated states, the peptides (2mg) were dissolved in water (2ml) and the pH adjusted to 2.0 with HCl (0.1M). After lyophilisation, the peptides were dissolved in d_6-dmso (0.5ml).

For the aqueous experiments P(1-20) (3.0mg) was dissolved in deuteroacetate buffer (0.54ml) in 90% H_2O/10% D_2O or D_2O alone and a trace amount of 3-(trimethylsilyl)-1-propane-sulphonic acid sodium salt (TPSS) was added as an internal reference. For experiments performed in D_2O the peptide was lyophilised twice from D_2O to remove the exchangeable protons then redissolved in D_2O deuteroacetate buffer prior to investigation by n.m.r. spectroscopy. For the aqueous experiments, chemical shift values are reported relative the TPSS resonance ($\delta = 0.00$).

4.3.9.1 One-Dimensional Spectra

Experimental parameters used for the acquisition of 1D spectra at 400MHz are summarised in Table 4.3. Experiments undertaken at higher field strengths were recorded with proportionately larger sweep widths.

Table 4.3 Experimental parameters used in 1D n.m.r. experiments at 400MHz.

Sweep width (Hz)	Size before zero filling	Size after zero filling	Resolution (Hz/pt)
5000	16k	32k	0.305

For the temperature study, spectra were recorded between 293K and 318K with 5K increments. A twenty minute period was allowed for temperature

equilibration between each of the experiments. The residual water signal was suppressed by presaturation at all times except during the acquisition period. In cases where signal overlap made it impossible to assign the amide proton resonances from the 1D spectrum alone, HOHAHA spectra were acquired at different temperatures.

4.3.9.2 Two-Dimensional Spectra

2D experiments were typically recorded at 303K into 512 x 2k files. A summary of the acquisition parameters used in the different experiments is presented in Table 4.4.

Table 4.4   Summary of experimental parameters used in 2D n.m.r. experiments at 400MHz.

Parameters	COSY	HOHAHA	NOESY	ROESY
Sweep width in F2(Hz)[a]	5000	5000	5000	5000
Sweep width in F1(Hz)[a]	2500	2500	2500	2500
Matrix size (F1 x F2) before zero-filling	512x1k	512x2k	512x2k	512x2k
Matrix size (F1 x F2) after zero-filling	1kx2k	1kx2k	1kx2k	1kx2k
Mixing times (ms)		40-120	100-600	100-200
Window functions for 2DFT (F1/F2)[b]	S/S	S/S	S/S	S/S
Shifts of window function in fractions of π (F1/F2)	1/2	1/2	1/2	1/2

[a] The sweep widths quoted are for experiments performed at 400MHz. Spectra acquired at higher field strengths had proportionately higher sweep widths.

[b] S stands for sine-bell squared.

The 2D experiments were performed in phase sensitive mode using time proportional phase incrementation (TPPI). For the HOHAHA experiment (Braunschweiler & Ernst, 1983) isotropic mixing was achieved using a MLEV17 pulse sequence. In the ROESY experiments (Bax & Davis, 1985) a field strength of 1-2kHz was used for spin locking. Spin-locking was achieved either by using a DANTE pulse sequence at 400MHz or a continuous wave (CW) pulse at 500MHz and 600MHz. ROESY and NOESY (Bodenhausen *et al*, 1984) spectra were acquired at least twice using either different transmitter frequencies or frequency offsets to ensure that amide to amide cross peaks did not arise as a result of incoherent Hartmann-Hahn magnetisation transfer.

4.3.10 Molecular Biology

4.3.10.1 RNA Extraction

595/102 (NCRC-48) hybridoma cells (1×10^7) were harvested by centrifugation ($3000g$ for 5 min) on a Microcentaur microcentrifuge. The supernatant was removed by gentle aspiration, and the cellular pellet vortexed briefly before resuspension in RNAzol™ B solution (1ml, Biogenesis Ltd., Bournemouth, U.K.). The RNA was solubilised by passing the lysate a few times through a pipette.

Chloroform (0.1ml) was added to the homogenate and the sample vortexed vigourously for 15s before being left to stand on ice for 5min. The resulting suspension was separated by centrifugation ($3,000g$ for 15min at 4°C). Addition of chloroform separates the homogenate into two phases. The lower blue phenol/chloroform phase and the aqueous upper phase. RNA is retained exclusively in the aqueous phase whereas the DNA and proteins remain in the interphase or the organic layer.

The aqueous phase was gently removed and transferred to a fresh Eppendorf tube and an equal volume of isopropanol added. The sample was left on ice for 15min to allow the RNA to precipitate. Centrifugation ($3,000g$ for 5min at 4°C) recovered

the RNA as a yellow/white pellet. The supernatant was removed by aspiration and the pellet washed twice by addition of 75% ethanol (1ml), vortexing (15s) and subsequent centrifugation (1750g for 8min at 4°C). After washing the supernatant was removed by aspiration, and the pellet air-dried for 15-20min. The pellet was resuspended in 50μl of sterile distilled water and stored at -20°C.

4.3.10.2 Quantitation of Nucleic Acids

The concentration of nucleic acid was determined spectrophotometrically. Samples (5μl) were diluted to 1ml in water and optical densities recorded at 260nm and 280nm using quartz cuvettes on a Pye Unicam SP8-400 UV/Visible spectrophotometer. The quantity of DNA or RNA was calculated using the following equations (Sambrook *et al*, 1989)

$$[DNA, \mu g/ml] = absorbance\ (260nm)\ x\ 50$$

$$[RNA, \mu g/ml] = absorbance\ (260nm)\ x\ 40$$

The purity of the samples was assessed assuming that the A(260nm)/A(280nm) for pure double stranded DNA = 1.8 and for RNA = 2.0 (Sambrook *et al*, 1989).

4.3.10.3 Reverse Transcription

RNA extracted from the 595/102 hybridoma cells was quantitated spectrophotometrically. Total RNA (10μg) was resuspended in sterile distilled water (11μl) in a 0.5ml Eppendorf tube. To this was added 10 x PCR Buffer (2μl, NBL, Cramlington, Northumberland, U.K.), 5mM dNTPs (4μl), N_6-hexamers (2μl) and Superscript™ RNase H- Reverse Transcriptase (RT, 1μl, Gibco BRL, Uxbridge, Middlesex, U.K.). The mixture was subjected to the following temperature cycle on a Pharmacia ATAQ Gene Controller PCR block.

10min @ 25°C - annealing of random primers

60min @ 37°C - reverse transcription

10min @ 95°C - denaturation of reverse transcriptase

After the final incubation, the tube was transferred immediately to ice. cDNA generated in this way was used directly for the PCR reactions without further purification. Negative controls for reverse transcription were set up in the form of tubes to which an extra 1 μl of water was added in place of the RT enzyme.

4.3.10.4 Amplification of DNA

To cDNA (1 μl) or the negative control (1 μl) generated as described above in a 0.5ml Eppendorf tube was added 5mM dNTPs (4 μl), 50pM each of the appropriate primers (0.5 μl from a 100pM/μl stock), 10 x PCR buffer (10 μl) and sterile distilled water (80 μl). The following primers were used for amplification of the heavy- and light-chain DNA respectively.

<div align="center">

Heavy chain variable region

VH1 FOR-2 5'-TGAGGAGACGGTGACCGTGGTCCCTTGGCCCC-3'

VH1 BACK 5'-AGGTCCAGCTGCAGGAGTCTGG-3'

Kappa chain

3'-PRIMER 5'-GCGCCGTCTAGAATTAACACTCATTCCTGTTGAA-3'

5'-PRIMER 5'-CCAGTTCCGAGCTCGTTGTGACTCAGGAATCT-3'

</div>

Finally, the Amphitaq™ DNA polymerase enzyme (1 μl, Promega, Southampton, Hants, U.K.) was added using a positive displacement pipette. The mixture was overlaid with light mineral oil (50 μl) and subjected to 30 cycles of the following temperature profile.

45s @ 94°C - denaturation of double stranded DNA

45s @ 56°C - annealing of primers to DNA template

90s @ 72°C - strand elongation.

After completion of the temperature cycles, the tubes were transferred to ice, analysed by agarose gel electrophoresis and then stored at -20°C.

4.3.10.5 Agarose Gel Electrophoresis

Agarose gels (1-2%) were prepared by dissolving powdered NuSieve™ agarose (Flowgen, Sittingbourne, Kent, U.K.) in the required volume of TBE buffer using either a water bath at 95°C or by heating in a microwave oven. Once completely dissolved, the agarose solution was allowed to cool to approximately 50-55°C and then poured into the appropriate electrophoresis tank and the comb inserted. Ethidium bromide (0.2 mg/ml) was incorporated into the gel prior to pouring. Once set, the comb and gel forming plates were removed and the gel was placed in an electrophoresis chamber (Pharmacia) and covered with TBE buffer pH 8.0.

Prior to loading samples were diluted 4:1 with DNA loading buffer. Electrophoresis was performed at a constant voltage of between 100-150 V and continued until the faster dye front (bromophenol blue) had migrated almost the full length of the gel. DNA was visualised by illumination of the gel using a Spectroline TM-312A ultra violet transilluminator and photographs recorded using a Polaroid DS-34 direct screen instant camera. Molecular weight markers were non-methylated bacteriophage ϕX174 DNA digested with restriction endonuclease Hae III (NBL, Cramlington, Northumberland, U.K.).

4.3.10.6 Ligation of PCR Fragments

Ligations were performed using a TA Cloning™ Kit (Invitrogen, San Diego, CA, U.S.A.) according to the manufacturers instructions. PCR fragments were ligated into the PCR™ II vector directly after amplification without further purification. A rough estimate of the amount of DNA present after amplification was obtained by visualising the PCR product on a 2% agarose gel. Ligations were set up between 1:1 to 1:3 molar ratio of PCR™ II vector to PCR product using 50 ng of vector per reaction.

4.3.10.7 TA Cloning™ Transformation

Transformation of competent *E.coli* supplied with the TA Cloning Kit was undertaken according to the manufacturers instructions. After the transformation reactions were complete, aliquots (25µl or 100µl) were spread onto agar plates containing kanamycin (50µg/ml) and X-Gal (Boehringer-Mannheim, Lewes, E.Sussex, U.K. ; 50µl from a 20mg/ml stock solution), prepared as described in Section 4.3.10.8.

4.3.10.8 Preparation of Agar Plates

Agar plates were prepared by adding bacto-agar (15g, Difco) to one litre of L-Broth according to the protocol described in Sambrook *et al* (1989). Plates used for selection of transformed *E.coli* were coated with X-Gal (50µl from a 20mg/ml stock solution) and allowed to air dry prior to spreading the bacterial culture.

4.3.10.9 Bacterial Culture

General culture techniques were taken from Sambrook *et al* (1989). Transformed cells were selected from white colonies on the agar plates and transfrred to LB-Medium (5ml) using a sterile needle. The cultures were grown overnight on a gyratory shaker (200rpm, 37°C). The plasmid was purified from the bacterial culture using a modified alkaline lysis method.

4.3.10.10 Minipreparation of Plasmid DNA

Plasmid DNA isolated using this method was routinely used for electrophoretic analysis, restriction endonuclease digestion and as a template in DNA sequencing reactions. The procedure is a modification of the rapid alkaline lysis method of Ish-Horowicz & Burke (1981). Samples (1.5ml) were removed from overnight cultures (5ml) of bacteria and centrifuged (3,000g for 5min) in a MSE Centaur centrifuge. The pellet was resuspended in 100µl of Solution I. To this suspension was added 200µl of Solution II and the tube agitated gently to ensure complete lysis of the bacteria. To the lysed solution was added 150µl of Solution III and the tube shaken vigourously for

5s before incubation on ice (5 min). Bacterial debris was pelleted by centrifugation (3,000*g* for 5min) and the clear supernatant removed to a clean tube. The nucleic acid was purified by the addition of an equal volume of phenol/chloroform. The mixture was vortexed vigourously, and the two layers separated by centrifugation (3,000*g* for 5min) on a microcentrifuge. The upper layer containing the plasmid was carefully removed to a fresh tube. Nucleic acid was precipitated by addition of 100% ethanol (720μl) and pelleted by centrifugation (13,000 rpm for 5min). The pellet of crude nucleic acid was washed once with 70% ethanol, air dried, and resuspended in sterile distilled water (50μl).

4.3.10.11 Restriction Endonuclease Digestion

Purified plasmid DNA was tested for incorporation of the appropriate PCR fragment by digestion with EcoR I (Cambridge Bioscience, Cambridge, U.K.) in 10x High Buffer (Cambridge Bioscience). The plasmid (1μl from stock obtained by minipreparation) was digested in a volume of 10μl for 2h at 37°C.

4.3.10.12 DNA Sequencing Reactions

Di-deoxy sequencing reactions (Sanger *et al*, 1977) were performed using a Sequinase™ DNA sequencing kit according to the manufacturers instructions. Purified plasmid DNA (5μg) was dissolved in sterile distilled water (30μl). The plasmid was denatured by addition of 2M NaOH (5μl) followed by incubation at room temperature for 5min. The DNA was then precipitated by the addition of ice cold 100% ethanol (120μl) and 1M sodium acetate buffer pH 5.0 (15μl). After incubation for a further 5min at room temperature the plasmid was pelleted by centrifugation (3,000*g* for 10min) on a microcentrifuge. The supernatant was decanted and the tube inverted to allow the DNA precipitate to air dry. The plasmid contains a priming sequence at each end of the cloning site. These primers, T7 and Sp6 (NBL) were used for the reactions. After the sequencing reactions were complete, stop solution (5μl) was added to each of the tubes and mixed by pipetting. The samples were either

analysed immediately by polyacrylamide gel electrophoresis or stored for later analysis at -20°C.

4.3.10.13 DNA Sequencing Gels

DNA sequencing gels were cast between the plates of a Bio-Rad SequiGen™ sequencing apparatus with an integral plate/chamber (IPC). Prior to casting the gel, the glass plates were extensively cleaned with soap and water, tap water, distilled water and finally ethanol and allowed to air dry. The removable plate of the IPC was then coated with Sigmacote™ in order to facilitate the subsequent removal of the gel from the plate. The solution was spread evenly over the plate and allowed to dry. The plates were then assembled according to the manufacturers instructions using 0.4mm spacers. Acrylamide for use in the sequencing apparatus was prepared freshly as required and contained acrylamide (5.7% w/v) bis-acrylamide (0.3% w/v) and urea (48% w/v) in 1 x TBE. Ammonium persulphate (25% w/v, 100µl) and TEMED (100µl) were added to the acrylamide mix (20 ml). The solutions were mixed by swirling and then poured briskly into the casting tray to seal the bottom of the gel. Once the plug had set, 85 ml of the acrylamide mix was used to form the gel itself. The acrylamide was de-gassed using a Buchner flask for 5min before addition of ammonium persulphate (10% w/v, 0.5ml) and TEMED (25µl). The solutions were mixed thouroughly, transferred into a 50ml syringe and injected carefully between the glass plates which were inclined at an angle of approximately 45° to the horizontal to facilitate pouring. Immediately after the gel was poured a flat 0.4mm spacer was placed into the acrylamide on the gel top such that it intruded into the gel by approximately 10mm. This allowed the formation of a flat gel surface essential to the effective use of the sharks tooth comb during electrophoresis. The gel was then laid flat and allowed to polymerise at room temperature for 3h. Once the gel had set, the plates were assembled in the electrophoresis apparatus according to the manufacturers instructions. The buffer chambers were filled with 1 x TBE and the gel pre-electrophorised at 2,000 V to heat it to the required operating temperature (50°C).

The sharks tooth comb was then inserted so that the tips protruded into the gel by approximately 2 mm. Sequencing reaction mixtures were heated to 75°C for 5min prior to loading to denature any secondary structure and loaded into the wells (2µl/well) in the order G, A, T, C. Electrophoresis of the samples was then allowed to proceed at a constant power of 55 W for 2-6h. After electrophoresis was complete, combs were removed and the IPC was carefully separated from the remaining plate and gel. The gel was then immersed in a fixative solution containing glacial acetic acid (10% v/v) and methanol (10% v/v) for 15min. The gel was transferred onto a large sheet of Whatmann 3MM filter paper (Scientific Lab Supplies) and dried on a vacuum gel drier (85°C for 75min) prior to autoradiography.

4.3.10.14 Autoradiography

Dried gels containing ^{35}S radioactivity were placed in contact with pre-flashed Kodak XAR-5 X-ray film (Sigma Chemical Co.). The gel and film were sealed in a light-tight aluminium autoradiography cassette (Sigma Chemical Co.) and exposed for varying lengths of time. Autoradiographs were developed in Polycon™ developer (KJP, Nottingham, U.K.) (diluted 1:4) for 15min, washed briefly in water and fixed in Perfix™ fixer (KJP) (diluted 1:4). The autoradiographs were washed in tap water and air dried.

4.3.10.15 Reading of Sequencing Gels

Autoradiographs of sequencing gels were read manually. The gels were read independently by at least two people and then loaded onto the Daresbury Seqnet computer. Sequences were aligned using the program Gap (Needleman & Wunsch, 1970) and any discrepancies checked against the original autoradiograph. Similar DNA sequences were identified using the program FASTA (Pearson & Lipman, 1988). After translation, similar protein sequences contained in the Swissprot database were identified with the program PFASTA (Pearson & Lipman, 1988).

CHAPTER FIVE

BINDING STUDIES

5.1 Introduction

In order to characterise the molecular basis of antibody-antigen interactions, it is necessary to precisely define the epitope with which the antibody reacts. A number of murine monoclonal antibodies for which MUC1 is the target antigen have been found to define epitopes within the tandem repeat of the protein core. The antibody C595, raised against affinity purified urinary mucin (Price et al, 1990a), has been tested for reactivity with a series of synthetic peptides in an attempt to accurately map the precise determinant with which it reacts.

5.2 Epitope mapping with synthetic peptide heptamers

Antibody binding to the immobilised peptides was detected in the form of colour development in the corresponding well of a microtitre plate. The results of the ELISAs for C595 are presented graphically in Figure 5.1.
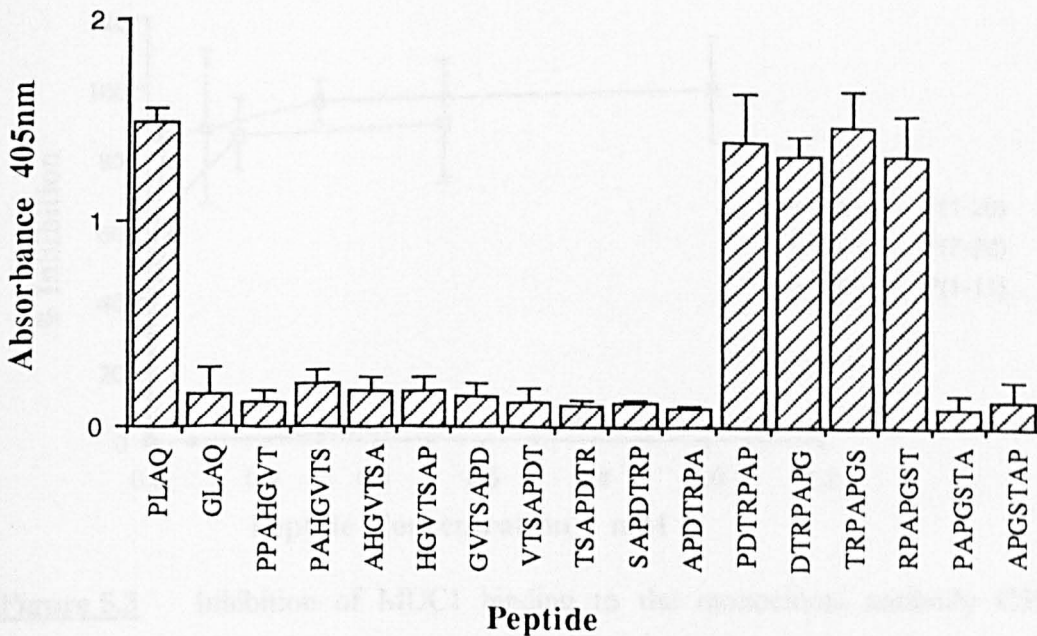
Figure 5.1   Antibody C595 binding to synthetic peptides based on the tandem repeat sequence of the MUC1 protein core.

C595 bound to four sequential peptides, from PDTRPAP to RPAPGST. The common sequence shared by these peptides is Arg-Pro-Ala-Pro (Figure 5.2). This

represents the minimum structure required for binding of the antibody. Hence C595 reacts with the tandem repeat domain in the protein core of MUC1.

P D T **RPAP**
D T **RPAP** G
T **RPAP** G S
RPAP G S T

Figure 5.2 Minimum structure required for binding of monoclonal antibody C595.

5.3 Inhibition Assays

The inhibition of binding of MUC1, affinity purified from urine (Price et al, 1990b), to the antibody C595 by peptides P(1-11), P(1-20) and P(7-20) is presented graphically in Figure 5.3.
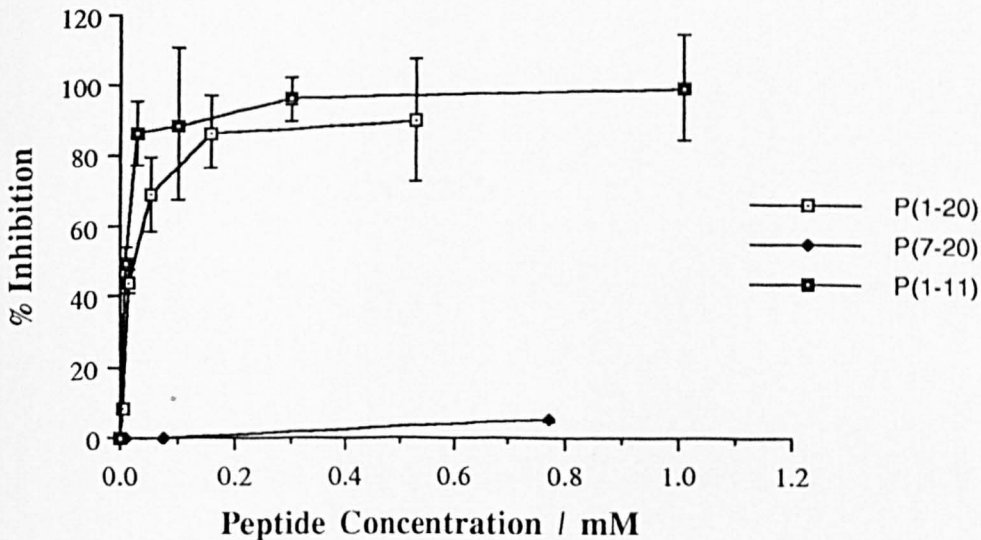
Figure 5.3 Inhibition of MUC1 binding to the monoclonal antibody C595 by synthetic peptides in solution.

Peptides P(1-11) and P(1-20) showed significant inhibition of antigen binding compared to P(7-20), which failed to affect this binding. These results clearly demonstrate that peptides containing the identified epitope for this antibody, Arg-Pro-

Ala-Pro, can inhibit the binding of the anti-mucin monoclonal antibody C595 to the intact urinary mucin.

5.3 Discussion

The epitope mapping studies clearly show that the monoclonal antibody C595 reacts with a determinant Arg-Pro-Ala-Pro which is contained within the tandem repeat domain of the protein core of MUC1. Furthermore, the inhibition assays demonstrate that peptides based on the tandem repeat sequence and the intact MUC1 molecules compete for the same binding site on this antibody. Therefore, investigations of the conformations adopted by these peptides in solution may provide an insight into the structural requirements for antibody binding to the mucin.

CHAPTER SIX

CONFORMATIONAL STUDIES OF P(1-20)

determine whether a salt bridge between the side chains of the aspartic acid and arginine residues plays a role in the stabilisation of the observed type-I β-turn in P(1-20) in d$_6$-dmso, the exchange characteristics of the low field resonances have been investigated.
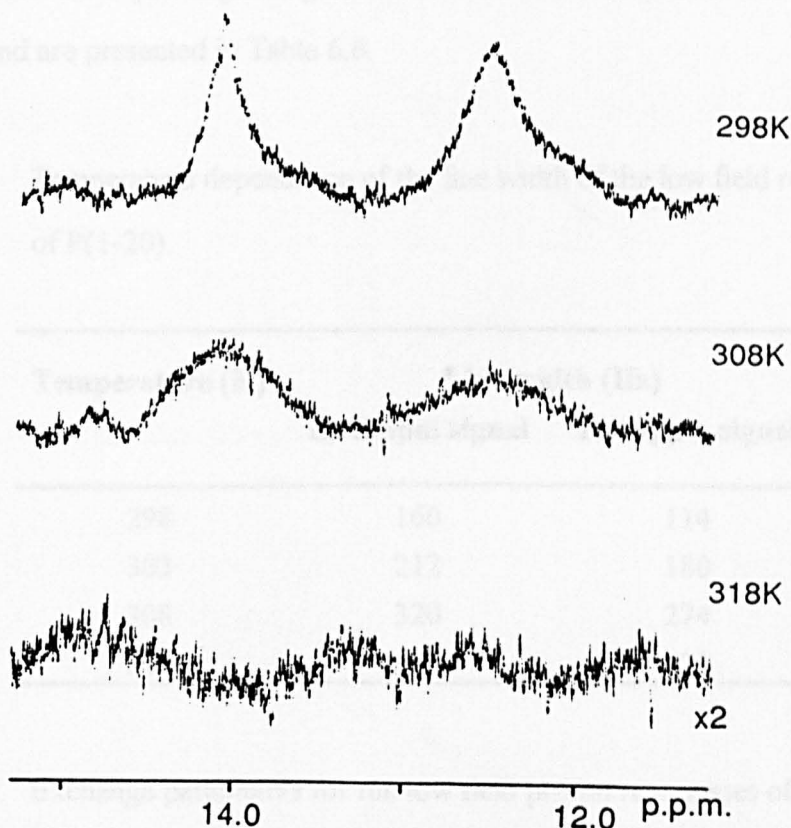
Figure 6.5   Temperature dependence of the line widths of the two low field proton resonances observed in the ^{1}H-n.m.r. spectrum of P(1-20).

Changes in the line width of the low field resonances of P(1-20) as a function of temperature over the range 293-318K are presented in Table 6.5. For exchangeable protons it is normally observed that at low temperature the line width is determined by the overall tumbling rate of the molecule under investigation, while at higher temperatures the exchange contribution is more important (Bevan *et al*, 1985). These two contributions have opposite temperature dependencies. The exchange dominated region of the temperature dependence is clearly observed in Figure 6.5. However since

dmso freezes at 291K, it is not possible to lower the temperature sufficiently to reach the region where the line width is dominated by the tumbling rate of the peptide. These data can be analysed quantitatively if it is assumed that the exchange rate shows Arrhenius behaviour as a function of temperature. The dynamics of the exchange process and the corresponding energies of activation have been calculated using such an analysis and are presented in Table 6.6.

Table 6.5 Temperature dependence of the line width of the low field resonances of P(1-20).

Temperature (K)	Line width (Hz)	
	12.42ppm signal	14.03ppm signal
298	160	114
303	212	180
308	320	274
318	470	430

Table 6.6 Exchange parameters for the low field proton resonances of P(1-20).

Chemical Shift (ppm) (303K)	Exchange Rate (s⁻¹) (303K)	Activation Energy (kJ/mol)
12.42	713	54.0
14.03	562	60.5

The exchange rates and activation energies of each of the low field resonances in P(1-20) are consistent with the presence of a proton involved in an ionic-type interaction. The observation of two resonances with similar populations and dynamics suggests that if this signal arises as a result of a salt bridge between the carboxylate side chain of Asp[2] and the guanidinium group of Arg[4], then it can exist in two

conformations. A single low field resonance with similar dynamic properties has been observed in P(1-11) indicating that in this peptide only one conformation is present (Scanlon *et al.* 1992).
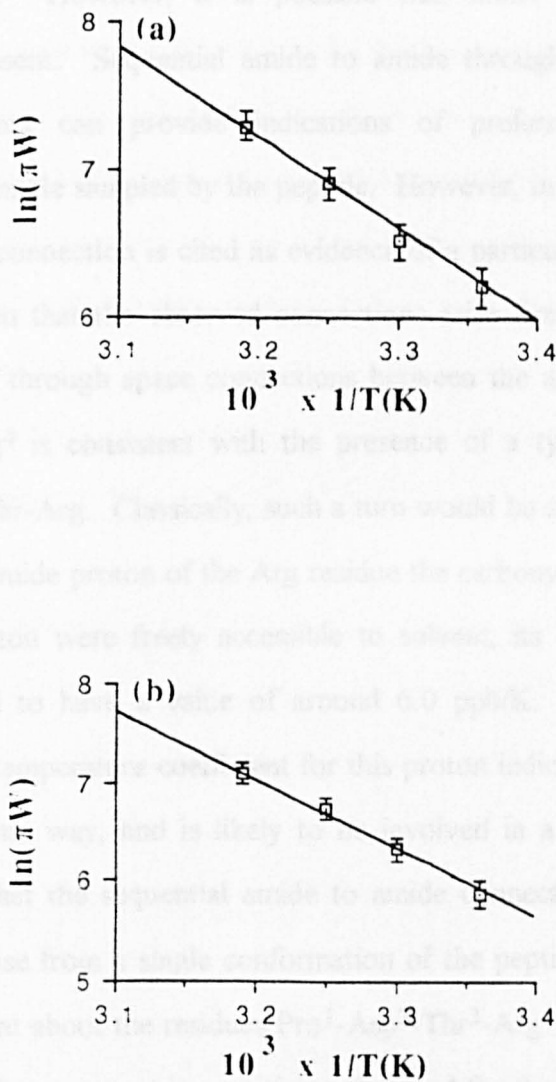
Figure 6.6   Temperature dependence of the line width of the low field resonances of P(1-20) presented as an Arrhenius plot [ln ($\pi\Delta$W) vs. 1/T]. (a) shows the temperature dependence of the signal which resonates at 14.03ppm at 303K, and (b) the 12.94ppm signal. The points are experimental data, the error bars indicating the estimated uncertainty in the line width ($\pm$5Hz). The lines were calculated for exchange processes having the parameters described in Table 6.6.

6.4 Discussion

The presence of strong $d_{\alpha N}$ connections in ROESY spectra along the entire backbone of P(1-20) indicates that much of the peptide exists in extended or random coil conformations. However, it is possible that minor populations of stable conformers are present. Sequential amide to amide through space connections in ROESY experiments can provide indications of preferred structures in the conformational ensemble sampled by the peptide. However, in cases where more than one through space connection is cited as evidence of a particular conformation, great care has to be taken that the observed connections arise from the same conformer. The observation of through space connections between the amide protons of Asp[2]-Thr[3] and Thr[3]-Arg[4] is consistent with the presence of a type I β-turn about the residues Pro-Asp-Thr-Arg. Classically, such a turn would be stabilised by a hydrogen bond between the amide proton of the Arg residue the carbonyl oxygen of the Pro. If the Arg amide proton were freely accessible to solvent, its temperature coefficient would be expected to have a value of around 6.0 ppb/K. The observation of a significantly lower temperature coefficient for this proton indicates that it is protected from solvent in some way, and is likely to be involved in a hydrogen bond. This strongly suggests that the sequential amide to amide connections observed between Asp-Thr-Arg do arise from a single conformation of the peptide and that a type I β-turn is indeed present about the residues Pro[1]-Asp[2]-Thr[3]-Arg[4]. Although a low value of the $^3J_{HN\alpha}$ coupling constant is sometimes observed for the residue at position 2 of a type I β-turn, the observed value for the Asp residue in P(1-20) is typical of an extended conformation in the peptide. However it must be stressed that the n.m.r. data represents a time weighted average of all the conformations sampled by the peptide. The ROESY data indicates that much of the peptide exists in random coil conformation and the β-turn represents only a minor conformation. It is therefore not surprising that the Asp $^3J_{HN\alpha}$ coupling constant has a value which reflects the major population in the conformational ensemble of the peptide.

al, 1988a ; Tendler, 1990). Indeed it has recently been suggested that the immunogenicity of peptides may be directly related to their propensity to form ordered structures in solution (Dyson, 1992). The identified structure in P(1-20) directly overlaps the epitopes of a number of anti-MUC1 monoclonal antibodies, suggesting that the salt-bridge locked type-I β-turn may be present on the surface of the intact MUC1 molecule where it has a role in the immune recognition of the mucin.

Although the salt bridge locked type I β-turn provides an attractive model for the interaction of MUC1 related peptides with antibodies subsequent data have shown that even if this structural featureis present, it is not an absolute requirement for recognition of the peptides.

6.1 Introduction

Peptides based on the tandem repeat sequence of the MUC1 protein core compete for the same binding site on the monoclonal antibody C595 as the intact mucin (Section 5.2). Consequently studies of the conformation of the peptides in solution may provide a model for investigating the recognition phenomena associated with the interaction of MUC1 with specific monoclonal antibodies. Previous studies on an eleven amino acid fragment of the tandem repeat sequence P(1-11) (Table 4.2) have revealed the presence of a type-I β-turn in the peptide between the residues Pro[1] and Arg[4] (Tendler, 1990). This turn overlaps the identified epitopes of several of the anti-MUC1 antibodies and may be involved in antigen recognition and binding.

In order to characterise further the solution structure of MUC1 related peptides, high field n.m.r. studies have been performed on the twenty amino acid peptide P(1-20) (Table 4.2) corresponding to the entire tandem repeat sequence. It has been found that antibody binding sites tend to contain a relatively high proportion of aromatic residues (Section 1.7). Therefore it is likely that the peptides, when docking into the binding site of C595 enter a more hydrophobic environment than that found in aqueous solution. The n.m.r. studies of P(1-20) were undertaken in d_6-dmso both because this solvent facilitates the acquisition of good quality n.m.r. data, and as it was believed to resemble more accurately the environment of the peptides at the surface of the antibody.

6.2 Sequence Specific Assignments

The first stage in the assignment of the proton resonances in the n.m.r. spectrum of P(1-20) to specific amino acid protons is the identification of the different spin systems present in the peptide. Spin systems were assigned on the basis of chemical shift and connectivity in one-dimensional and two-dimensional COSY and HOHAHA experiments using the methods described in Section 3.3. The amide to aliphatic region of a HOHAHA experiment is shown in Figure 6.1. Due to the presence of several residues of the same type in the peptide it was not possible to
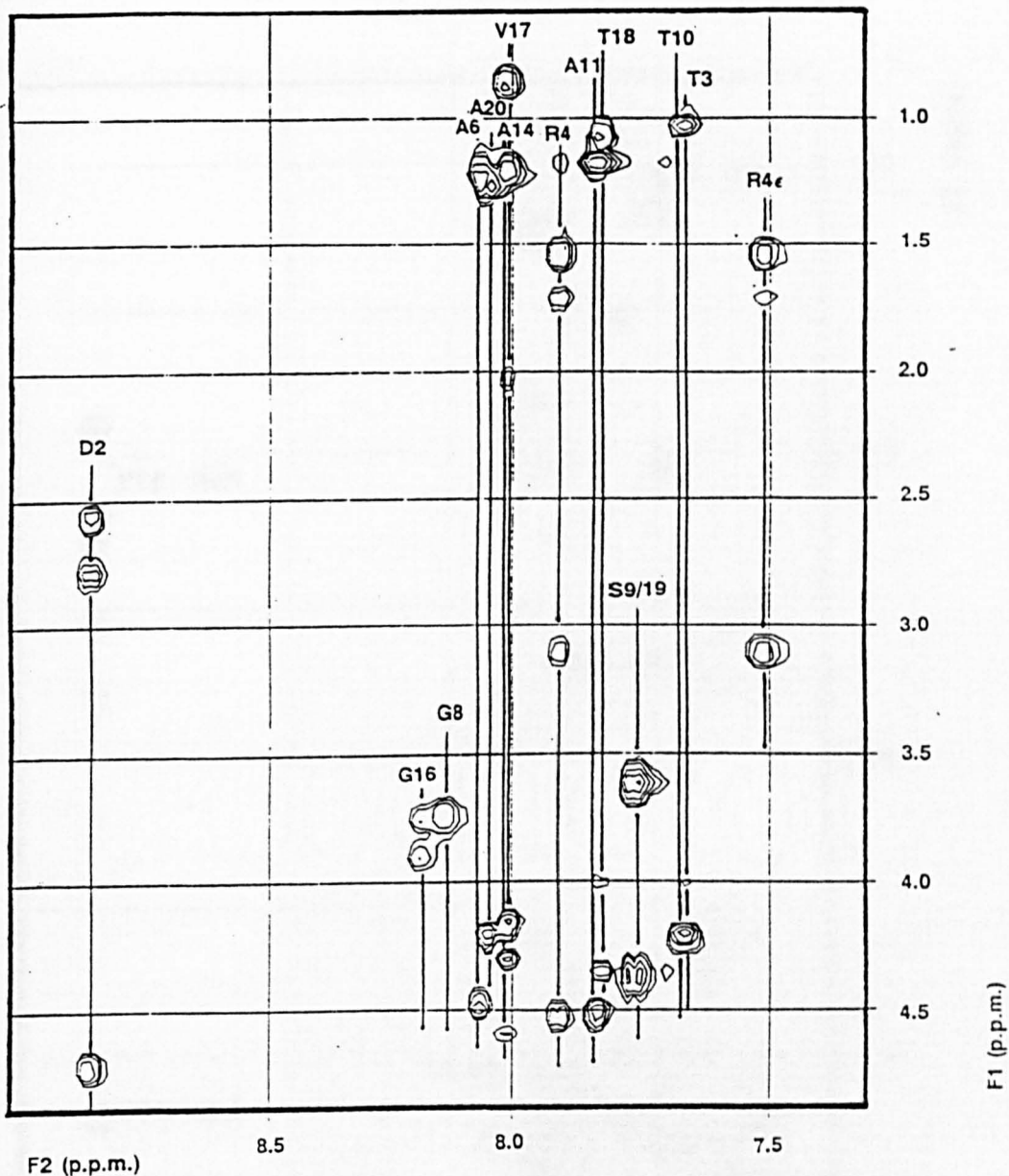
Figure 6.1   Amide to aliphatic region of the ^{1}H HOHAHA spectrum of P(1-20) in

d$_6$-dmso. The spectrum was recorded at 303K on a 400MHz

spectrometer with a mixing time of 80ms. The capital letters are the

amino acids given in the one-letter notation and the numbers refer to

the position of the residue in the sequence.

specifically assign the resonances from the spin system data alone. Consequently,

sequence specific assignments were based on the observation of nuclear Overhauser

effects (NOEs) between the resonances of protons on adjacent residues. As a result of

the tumbling rate of the peptide, no through space connections were observed in
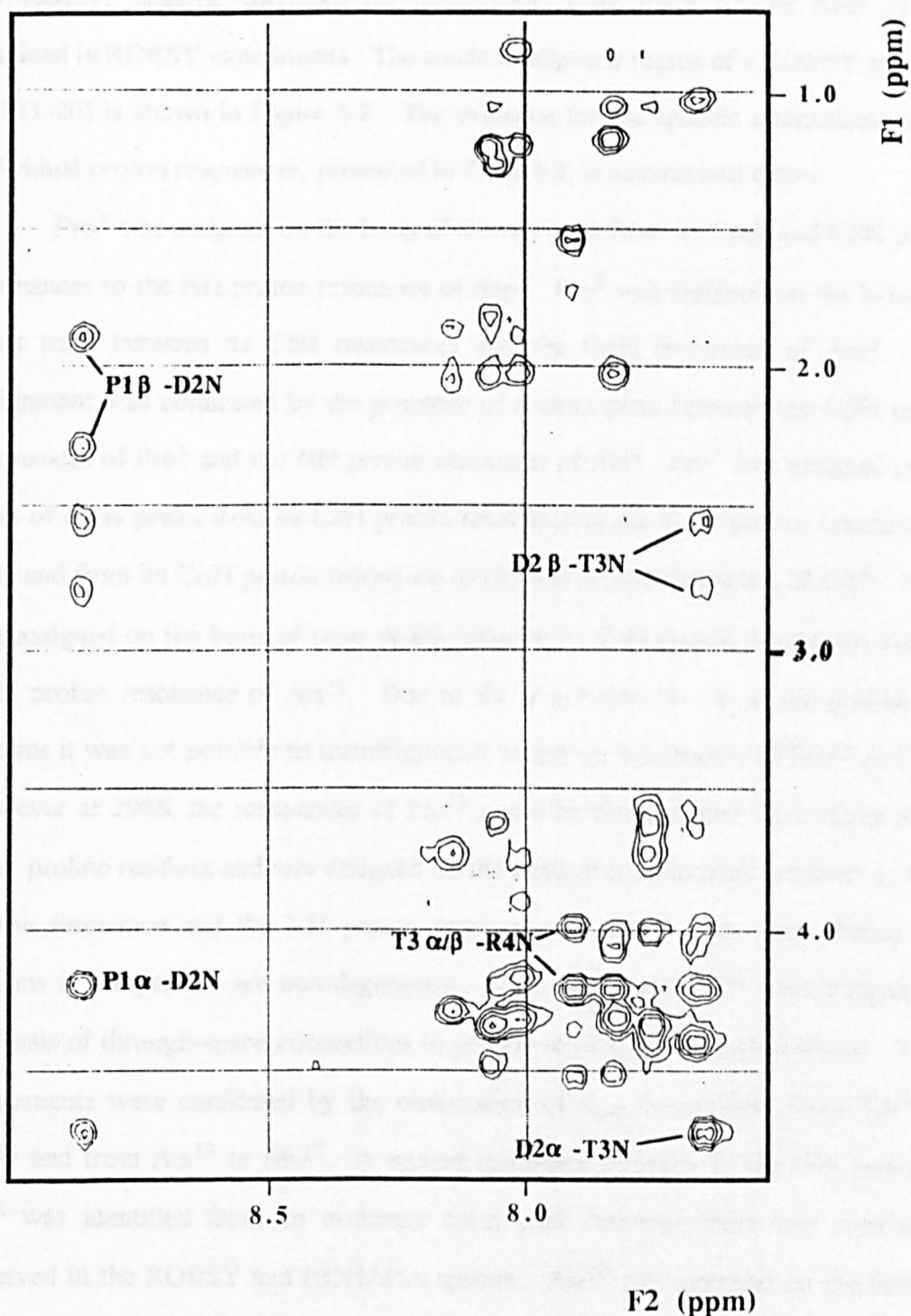
Figure 6.2 Amide to aliphatic region of the ^{1}H ROESY spectrum of P(1-20) in d_6-dmso. The capital letters are as defined in Figure 6.1. Annotation of cross peaks between the resonances of Pro^1-Asp-Thr-Arg^4 illustrates the assignment strategy.

2D NOESY spectra, therefore the assignments were made on the basis of data obtained in ROESY experiments. The amide to aliphatic region of a ROESY spectrum of P(1-20) is shown in Figure 6.2. The evidence for the specific assignments of the individual proton resonances, presented in Table 6.1, is summarised below.

Pro[1] was assigned on the basis of a cross peak from its $C\alpha H$ and $C\beta H$ proton resonances to the NH proton resonance of Asp[2]. Pro[5] was assigned on the basis of a cross peak between its $C\delta H$ resonances and the $C\alpha H$ resonance of Arg[4]. This assignment was confirmed by the presence of a cross peak between the $C\beta H$ proton resonances of Pro[5] and the NH proton resonance of Ala[6]. Pro[7] was assigned on the basis of cross peaks from its $C\delta H$ proton resonance to the $C\alpha H$ proton resonance of Ala[6] and from its $C\alpha H$ proton resonance to the NH proton resonance of Gly[8]. Pro[12] was assigned on the basis of cross peaks between its $C\delta H$ proton resonances and the $C\alpha H$ proton resonance of Ala[11]. Due to the degenerate nature of the proline spin systems it was not possible to unambiguously assign the resonances of Pro[13] at 303K. However at 298K the resonances of Pro[13] could be distinguished from those of the other proline residues and was assigned on the basis of a cross peak between its $C\alpha H$ proton resonance and the NH proton resonance of Ala[14]. The four alanine spin systems in the peptide are non-degenerate. Ala[6], Ala[11] and Ala[14] were assigned on the basis of through-space connections to proline residues as described above. These assignments were confirmed by the observation of $d_{\alpha N}$ connections from Thr[10] to Ala[11] and from Ala[14] to His[15]. A second resonance position for the NH proton of Ala[6] was identified from an exchange cross peak between these two resonances observed in the ROESY and HOHAHA spectra. Ala[20] was assigned on the basis of cross peaks between its NH proton resonance and the $C\alpha H$ and NH proton resonances of Ser[19]. Thr[3] and Thr[10] were assigned on the basis of $d_{\alpha N}$ connections to their amide proton resonances from the $C\alpha H$ proton resonances of Asp[2] and Ser[9] respectively. These assignments were confirmed by the presence of cross peaks between the $C\alpha H$

Table 6.1   Chemical shifts (ppm) of the assigned amino acid residues for P(1-20)

determined at 303K.

Residue	NH	αH	βH	Others
Pro[1]	8.43	4.18	2.25, 1.86	γH 1.86 δH 3.23, 3.18
Asp[2]	8.85	4.71	2.79, 2.55	
Thr[3]	7.70	4.19	3.96	γH 1.03 OH 4.86
Arg[4]	7.93	4.51	1.70, 1.53	γH 1.53 δH 3.10 εNH 7.52, 7.13
Pro[5]	-	4.33	2.00,1.75	γH 1.83 δH 3.65,3.50
Ala[6] (min)	8.07 9.17	4.48	1.17	
Pro[7]	-	4.30	1.98, 1.77	γH 1.83 δH 3.63
Gly[8]	8.16	3.75,3.70		
Ser[9]	7.77	4.31	3.59, 3.51	
Thr[10]	7.72	4.18	4.03	γH 1.02 OH 4.82
Ala[11]	7.84	4.50	1.17	
Pro[12]	-	4.54	2.11, 1.73	γH 1.89 δH 3.67, 3.44
Pro[13]*	-	4.46	2.14, 1.73	γH 1.73 δH 3.40
Ala[14]	8.04	4.15	1.17	
His[15]	7.94	4.46	3.05, 2.95	2H 6.58 4H 7.14
Gly[16]	8.18	3.89, 3.73		
Val[17]	8.03	4.28	2.20	γH 0.86
Thr[18]	7.83	4.33	3.98	γH 1.04
Ser[19]	7.78	4.38	3.63, 3.56	OH 5.05
Ala[20]	8.06	4.19	1.25	

*Assignments determined at 298K

and CβH proton resonances of Thr[3] to the NH proton resonance of Arg[4] and from the CαH and CβH proton resonances of Thr[10] to the NH proton resonance of Ala[11]. Thr[18] was therefore assigned by a process of elimination. The two glycine residues were assigned on the basis of $d_{\alpha N}$ connections from their CαH proton resonances to the NH proton resonances of Ser[9] and Val[17] respectively. Similarly, the two serine residues were assigned on the basis of $d_{\alpha N}$ connections from their CαH proton resonances to the NH proton resonances of Thr[10] and Ala[20]. The Asp[2] residue was assigned on the basis of connections observed from its amide proton resonance to the CαH proton resonance of Pro[1]. The low field position of the Asp[2] amide proton resonance is typical for the second residue in a linear peptide (Aumelas *et al*, 1989) and is due to the electron withdrawing effect of the positively charged N-terminus. Arg[4], His[15] and Val[17] were assigned on the basis of $d_{\alpha N}$ connections observed from their amide proton resonances to the CαH proton resonances of Thr[3], Ala[14] and Gly[16] respectively. The full assignments for P(1-20) are presented in Table 6.1.

6.3 Structural Studies

With a knowledge of the sequence specific assignments, it is possible to investigate a number of parameters in order to determine whether any structural features are present in the peptide.

6.3.1 Amide Proton Temperature Coefficients

The temperature coefficients of the amide protons of P(1-20) were determined between 293K-313K. The variation of chemical shift with temperature was found to be linear for all of the amide protons (Figure 6.3), suggesting that no major conformational change takes place in the peptide over this temperature range. The amide proton temperature coefficients are presented in Table 6.2.

The Asp[2] amide proton has the lowest temperature coefficient in the peptide at 2.6ppb/K, indicating that this proton is shielded from solvent and is likely to participate in a well formed hydrogen bond. In addition the amide protons of Arg[4], Ser[9], His[15]

and Ser[19], with temperature coefficients below 4.6 ppb/K appear to have restricted access to solvent and may also be involved in hydrogen bonding.
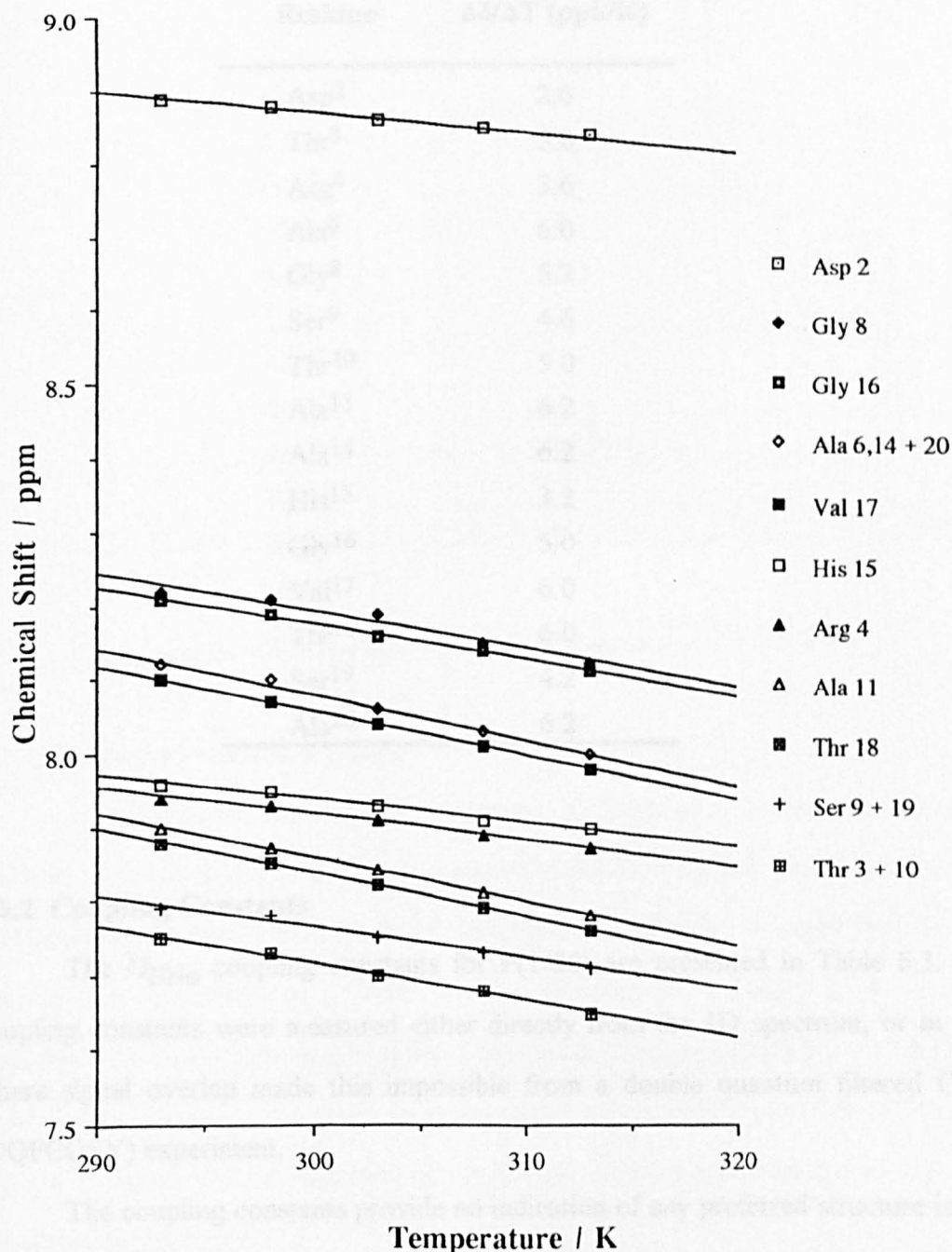
Figure 6.3 Variation of the chemical shifts of the amide proton resonances of P(1-20) in d₆-dmso. Lines drawn on the Figure are computer generated fits of the data obtained by the least squares method.

Table 6.2   Temperature coefficients for the amide protons of P(1-20).

Residue	$\Delta\delta/\Delta T$ (ppb/K)
Asp[2]	2.6
Thr[3]	5.0
Arg[4]	3.6
Ala[6]	6.0
Gly[8]	5.2
Ser[9]	4.6
Thr[10]	5.0
Ala[11]	6.2
Ala[14]	6.2
His[15]	3.2
Gly[16]	5.0
Val[17]	6.0
Thr[18]	6.0
Ser[19]	4.2
Ala[20]	6.2

6.3.2 Coupling Constants

The $^3J_{HN\alpha}$ coupling constants for P(1-20) are presented in Table 6.3. The coupling constants were measured either directly from the 1D spectrum, or in cases where signal overlap made this impossible from a double quantum filtered COSY (DQFCOSY) experiment.

The coupling constants provide no indication of any preferred structure in P(1-20). However, from the assignments of the resonances of protons in the peptide it is clear that more than one conformation is present - for example the amide proton of Ala[6] resonates at two positions. So although the values of the $^3J_{HN\alpha}$ coupling constants are characteristic of a random coil peptide, this observation does not rule out the possibility of minor conformations of preferred structures in the peptide.

Table 6.3 $^3J_{HN\alpha}$ coupling constants (Hz) for P(1-20) in d_6-dmso at 303K.

Residue	$^3J_{HN\alpha}$ (Hz)
Asp[2]	8
Thr[3]	9
Arg[4]	7
Ala[6]	8
Ser[9]	7
Thr[10]	7
Ala[11]	8
Ala[14]	7
His[15]	8
Val[17]	10
Thr[18]	8
Ser[19]	8
Ala[20]	8

6.3.4 Nuclear Overhauser Effects

In addition to enabling the assignment of resonances to specific amino acid protons, through space connections observed in the 2D ROESY experiments provide information regarding the structure of the peptide in solution. A cross peak between two proton resonances in a linear peptide indicates the presence of a significant population of conformers in which those two protons are separated by <3.5Å (Dyson *et al*, 1988a, 1988b). The amide to amide region of a ROESY spectrum of P(1-20) is shown in Figure 6.4. A summary of the dipolar connections observed in P(1-20) is presented in Table 6.4.

The pattern of connectivity observed between the amide protons of Asp[2]-Thr[3]-Arg[4] is consistent with the presence of a type-I β-turn being present in P(1-20) (Section 3.2.4). A similar turn has previously been identified in the peptide P(1-11) (Tendler, 1990). In addition, a computational study undertaken on the peptide Pro-

Asp-Thr-Arg-Pro identified a type-I β-turn about the residues Pro-Asp-Thr-Arg as the

major low energy conformer (Scanlon *et al*, 1992).

Figure 6.4   Amide to amide region of the ^{1}H ROESY spectrum of P(1-20).

Projections of the 1D spectrum are shown along both axes. The capital

letters are as defined in Figure 6.1.

Table 6.4   Summary of through-space connections observed in peptide P(1-20).

P D T R P A P G S T A P P A H G V S T A
$d_{\alpha N}$
d_{NN}
$d_{\beta N}$

The pattern of connectivity observed between the amide protons of Gly[8]-Ser[9] and Thr[10]-Ala[11] may indicate further structure in this region. Secondary structure calculations on the tandem repeat sequence have previously predicted a turn about the residues Gly[8]-Ser[9]-Thr[10] (Price *et al*, 1990a). However the distance constraints imposed by the observed NOEs are not sufficient to definitely characterise any particular structural feature.

6.3.5 Ionic Stabilisation of the β-turn in P(1-20)

The resonance line width in an n.m.r. spectrum is correlated with the value of the transverse relaxation time T_2.

$$\Delta W = 1/2 \; \pi T_2$$

where ΔW is the half width of the resonance line at half height.

The value of T_2 is characteristically affected by exchange processes taking place in the sample. Consequently, monitoring the effect of temperature on the resonance line width can provide thermodynamic information regarding dynamic processes in the molecule under investigation.

In the spectrum of P(1-20) at 303K there are two very low field resonances at 12.42 and 14.03 ppm (Figure 6.5). As a result of the broadness of the resonance lines, no connections were observed to these signals in COSY and HOHAHA experiments, with the result that they could not be assigned to specific protons. However, the low field nature of these resonances suggests that they may arise from protons involved in an ionic interaction. Simulated annealing calculations using a modified Metropolis Monte-Carlo algorithm have indicated that in the peptide Pro-Asp-Thr-Arg-Pro the major low energy conformer is stabilised by a salt bridge between side chains of the Asp and Arg residues (Scanlon *et al*, 1992). Furthermore it has been shown that in the n.m.r. spectrum of P(1-20) in which the guanidinyl group of the Arg residue was blocked, no low field signals were observed (Scanlon *et al*, 1992). In order to

CHAPTER SEVEN

STRUCTURAL STUDIES OF MUC1 CORE RELATED PEPTIDES

7.1 Introduction

HMFG1 and HMFG2 are murine monoclonal antibodies raised against preparations of human milk fat globule membranes which define epitopes within the tandem repeat of the protein core of MUC1 (Burchell *et al*, 1989). The minimum sequence required for HMFG1 binding is Pro-Asp-Thr-Arg, while HMFG2 defines the epitope Asp-Thr-Arg. The fine specificity of these antibodies has been investigated (Briggs *et al*, 1991) by testing their ability to bind to a series of tetrameric peptides based on the sequence Pro-Asp-Thr-Arg, according to the replacement net strategy developed by Geysen *et al* (1987) (Section 2.6). These studies have revealed that certain amino acids within the tetramer may be replaced with little or no loss of binding while the presence of others is essential (Briggs *et al*, 1991). HMFG2 has an absolute requirement for the Asp in this sequence. Replacement of this residue with any other amino acid leads to a loss of antibody binding activity. In the case of HMFG1, the replacement of Asp with its homologue Glu still allows sequence recognition to occur (Briggs *et al*, 1991). For HMFG1, the Arg residue is essential for antibody binding. With HMFG2, however, the Arg residue may be replaced by a number of other amino acids including Phe, Tyr and Leu, and binding activity is maintained. Interestingly, for HMFG2, although sequence recognition still occurs when the Arg is replaced by Leu, substitution of Arg for Ile removes the binding capacity of the peptide (Briggs *et al*, 1991).

In an attempt to identify a structural basis for these observations, the conformations of a number of peptides have been investigated by high field n.m.r. spectroscopy (Table 7.1). A type-I β-turn has previously been identified in the peptides P(1-11) (Tendler, 1990) and P(1-20) (Chapter 6) about the residues Pro-Asp-Thr-Arg. This structural feature directly overlaps the epitopes of both HMFG1 and HMFG2 and is thought to be important in recognition of these peptides by the antibodies.

Table 7.1   Sequences of peptides, based on the tandem repeat sequence of MUC1,
investigated by high field n.m.r. spectroscopy.

Peptide	Sequence									
D-peptide		A^1 P	D	T	R	P	A	P	G^9	
E-peptide		A^1 P	E	T	R	P	A	P	G^9	
I-peptide		A P	D	T	I	P	A	P	G^9	
L-peptide		A^1 P	D	T	L	P	A	P	G^9	
V-peptide	V^1 T	S A	P	D	T	R	P	A	P	G^{12}

Detection of β-turns by n.m.r. spectroscopy traditionally relies on three types of evidence. Firstly, a hydrogen bond between the amide proton at position 4 in the turn and the carbonyl oxygen of the residue at position 1 can be detected by a lowered amide proton temperature coefficient for the residue at position 4. This hydrogen bond, however, is not an absolute requirement for turn formation (Zimmerman & Scheraga, 1977 ; Chou & Fasman, 1977). Secondly, characteristically short proton-proton distances associated with the turn may be detected through the observation of NOEs. Thirdly, type-I and type-II turns may have low values of the $^3J_{HN\alpha}$ coupling constant for the residue at position 2 of the turn (Section 3.2.3). In the previously studied MUC1 core related peptides, the n.m.r. data indicated that a significant population of extended conformations were present. This was reflected in the observation of coupling constants characteristic of unstructured peptides. Consequently, evidence for β-turn formation was based on information from the first two sources.

These studies are directed towards ascertaining whether fine differences in the specificity of anti-MUC1 antibodies can be understood at the molecular level in terms of preferred structures adopted by the peptides.

7.2 HMFG1-Binding Peptides

7.2.1 Inhibition Assays

The ability of the D- and E-peptides to block the binding of intact urinary mucin to the anti-milk-fat-globule-membrane antibody HMFG1 was determined in a two-site radioimmunometric assay (Section 4.3.2). The inhibition of binding is presented graphically in Figure 7.1.
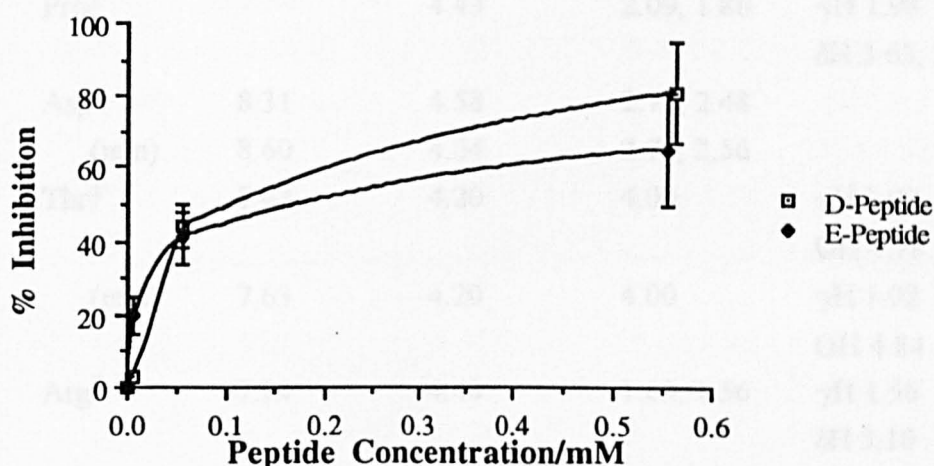
Figure 7.1   Inhibition of MUC1 binding to HMFG1 by synthetic peptides.

Both D- and E-peptides showed significant inhibition of antigen binding. These results confirm that replacement of the Asp residue with Glu in the tandem repeat sequence of the MUC1 core protein still allows recognition of the sequence by the antibody HMFG1.

7.2.2 N.m.r. Studies

In an attempt to identify any structural similarity between the D- and E-peptides, they have been investigated by high field n.m.r. spectroscopy. These studies were undertaken with the peptides dissolved in d_6-dmso. Assignment of the proton resonances in the two peptides was carried out using the methods previously described

in detail in Chapter 3 and Chapter 6. The full assignments for both peptides are presented in Table 7.2.

Table 7.2a   Resonance assignments for the D-peptide in d_6-dmso at 303K.

Residue	NH	αH	βH	Others
Ala[1]	8.05	4.20	1.35	
Pro[2]		4.43	2.09, 1.86	γH 1.98
				δH 3.65, 3.49
Asp[3]	8.31	4.58	2.73, 2.48	
(min)	8.60	4.64	2.76, 2.56	
Thr[4]	7.43	4.20	4.00	γH 1.00
				OH 4.78
(min)	7.63	4.20	4.00	γH 1.02
				OH 4.84
Arg[5]	7.94	4.48	1.69, 1.56	γH 1.56
				δH 3.10
Pro[6]		4.34	2.03, 1.82	γH 1.82
				δH 3.66, 3.53
Ala[7]	8.03	4.48	1.20	
Pro[8]		4.31	2.03, 1.85	γH 1.85
				δH 3.55
Gly[9]	8.01	3.78, 3.71		
(min)	8.30	3.71		

In both D- and E-peptides a number of residues resonate in two positions (Table 7.2). It is possible that the doubling of these proton resonances arises as a result of *cis/trans* isomerisation of one or more of the proline residues in the peptides. However, there is considerable signal overlap in the proline spin systems in both peptides, and no direct evidence for the presence of a *cis* proline conformation was observed. This suggests that both of these peptides exist in a number of distinct conformations.

Table 7.2b   Resonance assignments for the E-peptide in d_6-dmso at 303K.

Residue	NH	αH	βH	Others
Ala[1]	8.06	4.16	1.33	
(min)	7.91	4.22	1.17	
Pro[2]		4.41	2.10, 1.82	γH 1.90
				δH 3.63, 3.44
Glu[3]	8.17	4.28	1.93, 1.73	γH 2.29
(min)	8.44	4.35	1.93, 1.73	γH 2.29
Thr[4]	7.55	4.21	3.93	γH 1.00
(min)	7.71	4.23	3.98	γH 1.02
				OH 4.96
Arg[5]	7.95	4.50	1.71, 1.54	γH 1.54
				δH 3.07
				εNH 7.43
(min)	7.76	4.37	1.60, 1.46	γH 1.46
				δH 3.07
				εNH 7.49
Pro[6]		4.34	2.10, 1.84	γH 1.90
				δH 3.63, 3.51
Ala[7]	8.04	4.45	1.20	
(min)	8.39	4.49	1.24	
Pro[8]		4.29	2.02, 1.82	γH 1.90
				δH 3.54
Gly[9]	8.01	3.73		
(min)	8.33	3.71		

In the previously studies peptides P(1-11) and P(1-20), the amide proton of the Asp[2] residue has been found to resonate at a low field position (Tendler, 1990 ; Chapter 6). This has been explained as being due to electron withdrawing effect of the terminal ammonium group which causes deshielding of the second residue amide proton in linear peptides (Aumelas *et al*, 1989). Interestingly, the amide proton of the Asp residue in its minor conformation in the D-peptide and that of the Glu residue in its

minor conformation in the E-peptide both resonate at unusually low field positions. Since the Asp/Glu is the third residue in the D-/E-peptides this cannot arise as a result of deshielding by the N-terminal ammonium group and may be indicative of structure in this region. In order to further investigate the structure of the peptides in solution, the solvent accessibility of the amide protons was determined. The amide proton temperature coefficients for both peptides are presented in Table 7.3.

Table 7.3a   Amide proton temperature coefficients for the D-peptide.

Residue		Temperature Coefficient $(\Delta\delta/\Delta T)$ ppb/K
Ala[1]		4.0
Asp[3]	(maj)	5.8
	(min)	3.5
Thr[4]	(maj)	3.6
	(min)	2.8
Arg[5]		4.7
Ala[7]		6.6
Gly[9]	(maj)	6.2
	(min)	6.5

Table 7.3b   Amide proton temperature coefficients for the E-peptide.

Residue		Temperature Coefficient $(\Delta\delta/\Delta T)$ ppb/K
Ala[1]		4.5
Glu[3]	(maj)	5.1
	(min)	3.5
Thr[4]	(maj)	4.3
	(min)	5.6
Arg[5]		5.4
Ala[7]		6.2
Gly[9]	(maj)	6.0
	(min)	6.4

As shown in Table 7.3, the least solvent accessible proton in the D-peptide is that of the Thr in the minor conformation. In addition Asp(min), Thr(maj) and Arg appear to be protected from solvent. In the case of the E-peptide the amide proton temperature coefficients are generally higher, and apart from the Glu(min) and Thr(maj) resonances, there is no indication of the amide protons being involved in hydrogen bonding.

Short distances between protons in β-turns can be detected through NOEs. In the case of the D-peptide, no cross peaks were observed in conventional NOESY experiments, and consequently NOE data were obtained from ROESY spectra. For the E-peptide, however, cross peaks were obtained in NOESY spectra, indicating a fundamental difference in the tumbling rate of the two peptides. The amide to amide regions of NOE spectra for both peptides are presented in Figure 7.2.
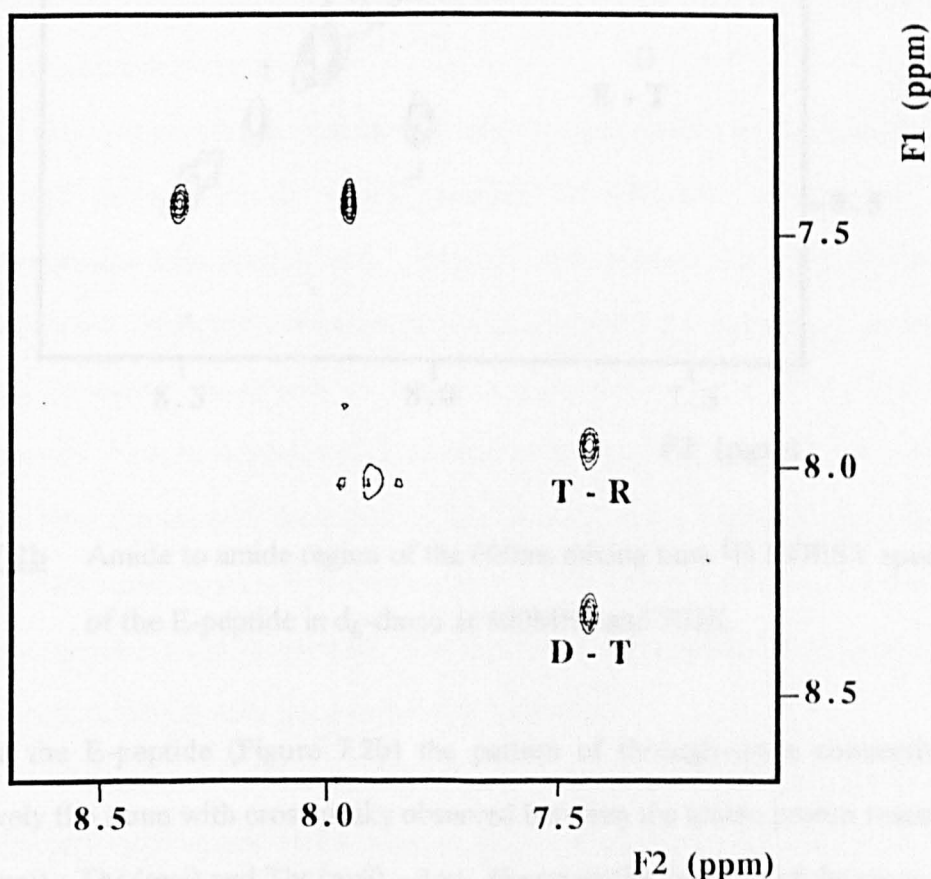
Figure 7.2a   Amide to amide region of the 200ms mixing time ^{1}H ROESY spectrum of the E-peptide in d_6-dmso recorded at 600MHz and 303K.

As shown in Figure 7.2a, strong amide to amide connections are observed for the D-peptide between the resonances of Asp(maj)-Thr(maj) and Thr(maj)-Arg. This observation in conjunction with the lowered temperature coefficient of the Arg amide proton resonance, suggests that a type-I β-turn is present in the D-peptide about the residues Pro-Asp-Thr-Arg.
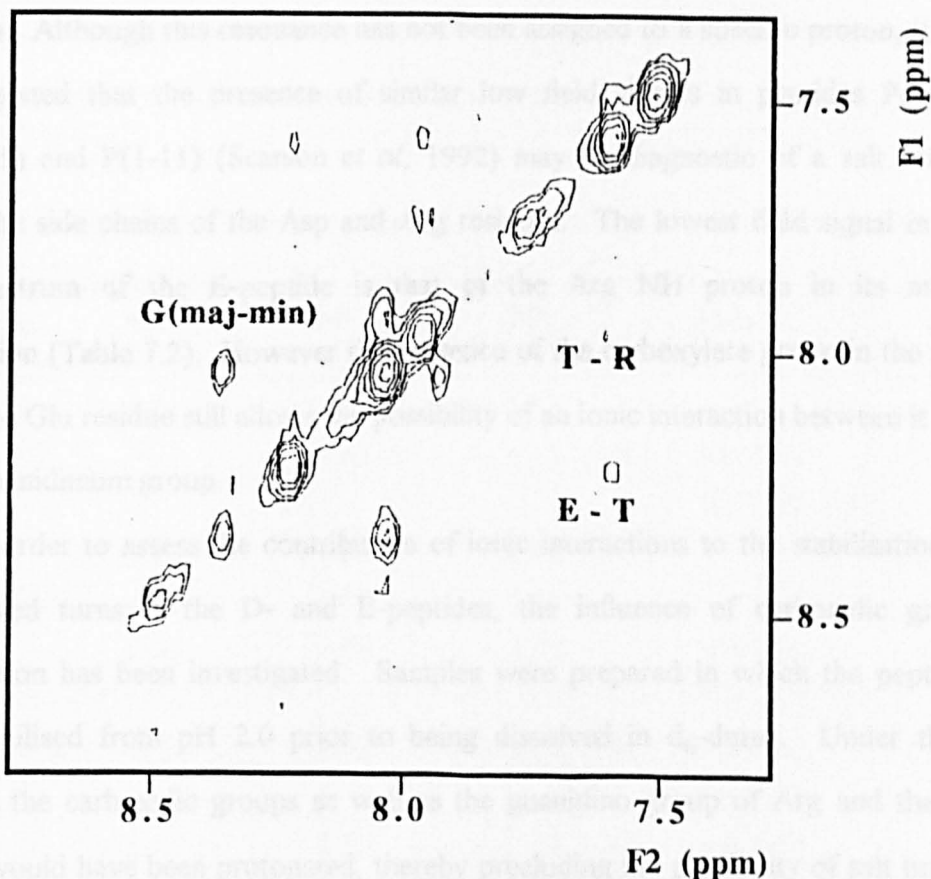
Figure 7.2b   Amide to amide region of the 600ms mixing time ¹H NOESY spectrum of the E-peptide in d₆-dmso at 400MHz and 303K.

In the E-peptide (Figure 7.2b) the pattern of through-space connectivity is qualitatively the same with cross peaks observed between the amide proton resonances of Glu (maj) - Thr (maj) and Thr (maj) - Arg. However the intensity of the cross peaks is significantly lower, indicating that either the distance between the amide protons in the E-peptide is greater, or the turn is present at a lower population. The presence of a

less populated or less well formed turn would explain the high value of the Arg amide proton temperature coefficient which shows no indication of hydrogen bonding.

7.2.3 Role of Ionic Interactions in the Stabilisation of MUC1 Core Related Peptides

In the n.m.r. spectrum of the D-peptide there is a single low field resonance at 12.46 ppm. Although this resonance has not been assigned to a specific proton, it has been suggested that the presence of similar low field signals in peptides P(1-20) (Figure 6.5) and P(1-11) (Scanlon *et al*, 1992) may be diagnostic of a salt bridge between the side chains of the Asp and Arg residues. The lowest field signal in the n.m.r. spectrum of the E-peptide is that of the Arg NH proton in its minor conformation (Table 7.2). However the presence of the carboxylate group in the side chain of the Glu residue still allows the possibility of an ionic interaction between it and the Arg guanidinium group.

In order to assess the contribution of ionic interactions to the stabilisation of the observed turns in the D- and E-peptides, the influence of carboxylic group deprotonation has been investigated. Samples were prepared in which the peptides were lyophilised from pH 2.0 prior to being dissolved in d_6-dmso. Under these conditions the carboxylic groups as well as the guanidino group of Arg and the N-terminus would have been protonated, thereby precluding the possibility of salt bridge formation. For the samples described in Section 7.2.2 and Chapter 6, the side chains of the Asp/Glu residues were in the carboxylate (COO⁻) state, while the Arg side chain and amino terminus were protonated. The resonance assignments for both peptides in the carboxylic (COOH) state are presented in Table 7.4.

Differences in the chemical shift of proton resonances in both peptides in the carboxylic and carboxylate states suggest that the conformations adopted by the peptides are sensitive to changes in pH. For the D-peptide carboxylic sample, the Asp residue was found to resonate in two positions, indicating the presence of more than one stable conformation. The amide proton of the Asp residue in its minor

conformation was no longer found to resonate at the unusually low field position observed in the D-peptide in the carboxylate state. However, a low field resonance was still observed at 12.46 ppm, suggesting that this resonance is not indicative of salt bridge formation. The presence of a similar low field resonance in Asp-containing peptides irrelevant to the MUC1 protein core sequence (G.M. Davies, unpublished results) suggests that this signal may arise from the carboxylic proton of the Asp residue. For the E-peptide in the protonated state, a single resonance position is observed for each proton in all of the amino acid residues.

Table 7.4a Resonance assignments for the D-peptide in the carboxylic state.

Residue	NH	αH	βH	Others
Ala[1]	8.12	4.17	1.21	
Pro[2]		4.42	2.17, 1.84	γH 1.84 δH 3.68, 3.47
Asp[3]	8.40	4.69	2.92, 2.60	
(min)	8.41	4.60	2.74	
Thr[4]	7.46	4.20	3.98	γH 1.00 OH 4.85
Arg[5]	8.00	4.49	1.70, 1.53	γH 1.53 δH 3.10
Pro[6]		4.36	2.10, 1.88	γH 1.88 δH 3.67
Ala[7]	8.12	4.47	1.36	
Pro[8]		4.32	2.06, 1.88	γH 1.88 δH 3.57
Gly[9]	8.07	3.74		

Conformational analysis of the two peptides has been carried out on the basis of amide proton solvent accessibility and through-space connectivity in nuclear Overhauser enhancement experiments. The amide proton temperature coefficients are presented in Table 7.5.

Table 7.4b   Resonance assignments for the E-peptide in the carboxylic state.

Residue	NH	αH	βH	Others
Ala[1]	8.10	4.15	1.37	
Pro[2]		4.44	2.11, 1.85	γH 1.85
				δH 3.65, 3.47
Glu[3]	8.25	4.30	1.92, 1.75	γH 2.30
Thr[4]	7.60	4.21	3.96	γH 1.01
Arg[5]	7.97	4.51	1.71, 1.54	γH 1.54
				δH 3.10
				εNH 7.43
Pro[6]		4.35	2.03, 1.77	γH 1.77
				δH 3.65, 3.55
Ala[7]	8.05	4.47	1.20	
Pro[8]		4.33	2.03, 1.85	γH 1.85
				δH 3.58
Gly[9]	8.05	3.71		

Table 7.5a   Amide proton temperature coefficients for the D-peptide in the carboxylic state.

Residue		Amide proton temperature coefficient ($\Delta\delta/\Delta$T) ppb/K
Asp[3]	(maj)	1.5
	(min)	2.0
Thr[4]		2.4
Arg[5]		5.1
Ala[7]		6.2
Gly[9]		6.5

In the D-peptide the amide proton temperature coefficients of the Asp residue in both its major and minor conformations and that of the Thr residue are very low,

indicating that these amide protons are shielded from solvent, and likely to participate in hydrogen bonding. However, the temperature coefficient of the Arg amide proton is significantly higher when the D-peptide is in the carboxylic rather than carboxylate state. The temperature coefficients of the E-peptide are again generally higher, with the lowest values being observed for the Glu and Thr amide protons.

Table 7.5b Amide proton temperature coefficients for the E-peptide in the carboxylic state.

Residue	Amide proton temperature coefficient ($\Delta\delta/\Delta T$) ppb/K
Glu^3	4.8
Thr^4	3.5
Arg^5	5.3
Ala^7	6.1
Gly9	5.8

For both peptides in the carboxylic state, strong $d_{\alpha N}$ through space connections were observed along the entire backbone. However no sequential d_{NN} through space connections were present in either case.

The absence of d_{NN} connections, in conjunction with the high values for the amide proton temperature coefficient of the Arg residue in both peptides suggests that the β-turn which is present in both peptides in the carboxylate state is no longer formed when the carboxylic groups are protonated. This observation implies that ionic interactions may be important in stabilising the turn. Previously it was thought that low field resonances observed in peptides P(1-11), P(1-20) and the D-peptide may be diagnostic of the formation of a salt bridge. Molecular modelling studies had identified such an interaction between the side chains of the Asp and Arg residues in the pentapeptide Pro-Asp-Thr-Arg-Pro (Scanlon *et al*, 1992). Since this low field

Table 7.6   Assignments of the proton resonances for the V-peptide at 303K in d6-
             dmso at 303K.

Residue	NH	αH	βH	Others
Val[1]	n.d.	3.83	2.13	γH 0.98
Thr[2]	8.43	4.55	4.03	γH 1.13
				OH 5.18
Ser[3]	7.93	4.35	3.63	OH 4.88
Ala[4]	8.07	4.50	1.23	
(min)	8.45	4.55	1.30	
Pro[5]		4.35	2.05, 1.90	γH 1.90
				δH 3.62, 3.57
Asp[6]	8.20	4.57	2.77, 2.50	
(min)	8.61	4.63	2.77, 2.50	
Thr[7]	7.38	4.23	4.03	γH 1.05
				OH 4.75
Arg[8]	7.97	4.50	1.73, 1.58	γH 1.58
				δH 3.13
				εNH 7.60
Pro[9]		4.37	2.05, 1.90	γH 1.90
				δH 3.68, 3.57
Ala[10]	8.10	4.58	1.23	
Pro[11]		4.36	2.05, 1.90	γH 1.90
				δH 3.68, 3.57
Gly[12]	8.05	3.75		
(min)	8.35	3.75		

n.d. = not determined

resonance is observed in the D-peptide in both the carboxylate state and in the carboxylic state, where no possibility for salt bridge formation exists, it cannot arise due to the presence of an ionic interaction in the peptide.

Although the modelling studies indicated that the Asp carboxylate group was involved in a salt bridge with the Arg guanidinium group, the proximity of the Asp residue to the N-terminus, which was protonated in all the n.m.r. experiments described above, means that this cannot be ruled out as the positively charged partner in an ionic interaction. In order to clarify which of these two groups is the more likely to be involved in the salt bridge, the V-peptide was prepared. In this peptide the N-terminus was extended by three residues, thereby reducing the likelihood of salt bridge formation between the N-terminus and the Asp side chain carboxylate group.

The conformation of the V-peptide in d_6-dmso was investigated using high field n.m.r. spectroscopy according to the methods described above. The full assignments for the V-peptide are presented in Table 7.6.

Conformational analysis of the V-peptide has been carried out on the basis of amide proton temperature coefficients and through space connectivity in 2D NOESY experiments as described above. The amide proton temperature coefficients for the V-peptide are presented in Table 7.7.

The temperature coefficients indicate that with the exception of the Thr^7 residue, there is little evidence of hydrogen bonding involving backbone amide protons in the V-peptide.

The pattern of through space connectivity in the V-peptide (Figure 7.3) is similar to that observed for other mucin core related peptides. Strong $d_{\alpha N}$ connections are observed along the entire length of the peptide backbone indicating that much of the peptide exists in extended conformations. However, the pattern of d_{NN}-connectivity observed is somewhat different. As in previously studied peptides d_{NN}-connections are observed between Asp^6(maj)-Thr^7 and Thr^7-Arg^8. In addition further d_{NN} through space connections are present between Ser^3-Ala^4 and

Ala^4-Pro$^5\delta$. A summary of the through space connections observed in the V-peptide is presented in Table 7.8.

Table 7.7 Amide proton temperature coefficients for the V-peptide in d$_6$-dmso.

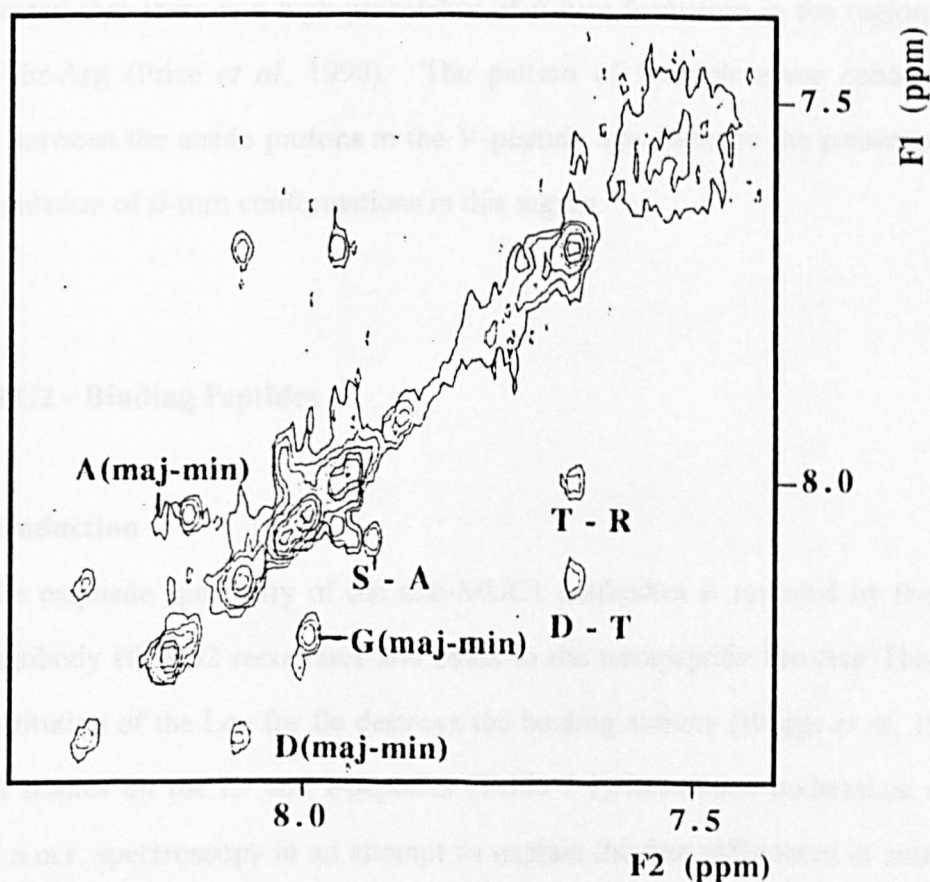| Residue	Amide proton temperature coefficient $(\delta\Delta/\delta T)$ ppb/K
Thr^2	5.3
Ser^3	5.4
Ala^4	7.0
Asp^6	5.9
Thr^7	1.8
Arg^8	5.6
Ala^{10}	6.5
Gly^{12}	6.5

Figure 7.3 Amide to amide region of the 600ms mixing time ^{1}H NOESY spectrum of the V-peptide in d$_6$-dmso recorded at 400MHz and 303K.

Table 7.8   Through space connections observed in the V-peptide.

$$\text{V T S A P D T R P A P G}$$

$\text{d}_{\alpha N}$ ‾‾‾‾‾‾‾‾‾‾‾‾‾‾‾‾‾‾

d_{NN}^{*} ‾‾‾   ‾‾‾

*In the case of the prolines connections were observed to the δ-proton

resonances.

Without confirmatory evidence from other sources, the pattern of through

space amide to amide connectivity is insufficient to unambiguously define β-turn

conformations in the peptide.   Structural predictions on a 60 amino acid peptide

corresponding to three copies of the tandem repeat found in the MUC1 core protein

have indicated that there is a high probability of β-turn formation in the region Ala-

Pro-Asp-Thr-Arg (Price *et al*, 1990).   The pattern of through space connectivity

observed between the amide protons in the V-peptide may indicate the presence of a

mixed population of β-turn conformations in this region.

7.3 HMFG2 - Binding Peptides

7.3.1 Introduction

The exquisite specificity of the anti-MUC1 antibodies is revealed by the fact

that the antibody HMFG2 recognises and binds to the tetrapeptide Pro-Asp-Thr-Leu,

while substitution of the Leu for Ile destroys the binding activity (Briggs *et al*, 1991).

Structural studies on the L- and I-peptides (Table 7.1) have been undertaken using

high field n.m.r. spectroscopy in an attempt to explain the fine differences in antibody

specificity at the molecular level.   Furthermore, substitution of the Arg residue for

Leu/Ile removes the possibility of forming a salt bridge to the Asp side chain and

thereby addresses the question of to what extent such an interaction is prerequisite for turn formation.

7.3.2 Sequence Specific Assignments

The ^{1}H resonances of the L- and I-peptides were assigned on the basis of chemical shift and connectivity in 1D and 2D n.m.r spectra using the methods described above. Sequence specific assignments were obtained from through space connections observed in 2D NOESY spectra. The full assignments for both peptides are presented in Table 7.9.

Table 7.9a   Chemical shifts (ppm) for the L-peptide in d_6-dmso at 303K.

Residue	NH	αH	βH	Others
Ala^1	7.99	4.48	1.18	
Pro^2		4.43	2.09, 1.90	γH 1.90
				δH 3.53
Asp^3	8.33	4.59	2.73, 2.53	
(min)	8.56	4.68	2.78, 2.60	
Thr^4	7.36	4.20	3.99	γH 1.01
(min)	7.60	4.20	3.99	γH 1.01
Leu^5	7.83	4.53	1.43	γH 1.64
				δH 0.88
Pro^6		4.30	2.03, 1.88	γH 1.88
				δH 3.63, 3.50
Ala^7	8.08	4.16	1.34	
(min)	8.38	4.55	1.25	
Pro^8		4.30	2.03, 1.88	γH 1.88
				δH 3.63, 3.50
Gly^9	8.03	3.70		

As with previously studies MUC1 core related peptides, a number of the residues in both the L- and I-peptide were found to resonate in two positions.

However, in the case of the I-peptide a minor conformation of the Pro^2 spin system, presumably corresponding to the *cis* isomer, was observed. The Pro^2(min) resonances

Table 7.9b Chemical shifts (ppm) for the I-peptide in d_6-dmso at 303K.

Residue	NH	αH	βH	Others
Ala^1	7.96	4.50	1.19	
Pro^2		4.45	2.25, 1.81	γH 2.05
				δH 3.51
(min)		4.54	2.18, 1.75	γH 1.90
				δH n.d.
Asp^3	8.30	4.60	2.76, 2.60	
(min)	8.55	4.68	2.71, 2.53	
Thr^4	7.43	4.22	3.99	γH 0.99
(min)	7.64	4.23	4.00	γH 0.99
Ile^5	7.71	4.38	1.75	γH 1.50, 1.08
				γCH₃ 0.89
				δH 0.80
(min)	7.53	4.28	1.63	γH 1.40, 1.01
				γCH₃ 0.80
				δH 0.80
Pro^6		4.33	1.99, 1.81	γH 1.88
				δH 3.75, 3.60
Ala^7	8.04	4.18	1.36	
(min)	8.39	4.48	1.28	
Pro^8		4.41	2.10, 1.88	γH 2.06
				δH 3.68
Gly^9	8.01	3.76		
(min)	8.36	3.73		

of the I-peptide were assigned on the basis of a through space connection from it CαH proton resonance to the NH proton resonance of Asp^3(min). Assuming the Asp^3(min) represents the proportion of peptide conformations containing a *cis* proline at position

2, and Asp^3(maj) reflects the population of *trans* proline conformations, then on the basis of the relative intensity of the amide proton resonances of Asp(maj)/(min) the ratio of *trans/cis* conformations of Pro^2 is approximately 9:1. No evidence has been found for *cis* proline conformations in any of the other MUC1 related peptides studied to date.

In both L- and I-peptides there is a single low field resonance at 12.46 ppm, confirming that this resonance is not diagnostic of a salt bridge interaction involving the Arg guanidinium group.

Structural information regarding the two peptides was obtained from calculation of amide proton temperature coefficients and through space connectivity in 2D NOESY spectra. The amide proton temperature coefficients for both peptides are presented in Table 7.10.

Table 7.10a   Amide proton temperature coefficients for the L-peptide.

Residue		Temperature Coefficient $(\Delta\delta/\Delta T)$ ppb/K
Ala^1		-
Asp^3	(maj)	3.4
	(min)	5.7
Thr^4	(maj)	2.6
	(min)	4.5
Leu^5		4.8
Ala^7	(maj)	6.3
	(min)	7.0
Gly^9		7.1

The temperature coefficients of the Asp(maj), Thr(maj), Thr(min) and Leu amide protons in the L-peptide are all below the value of 6.0 ppb/K expected for solvent exposed amide protons. This suggests that they may be involved in hydrogen

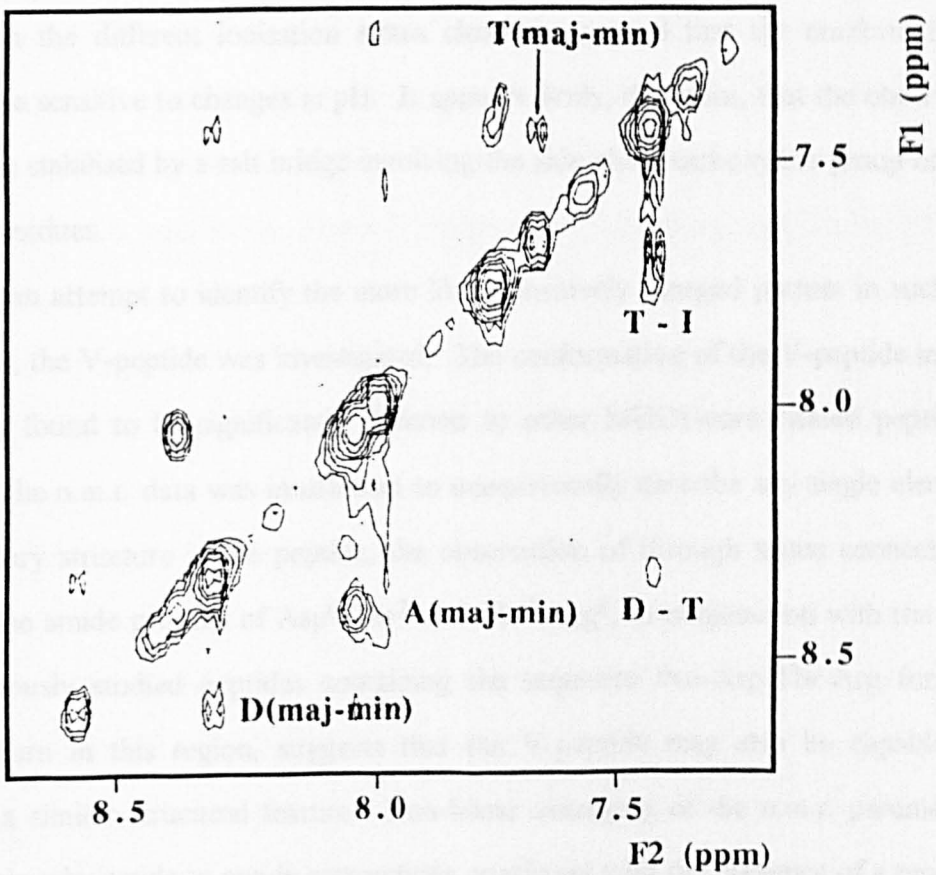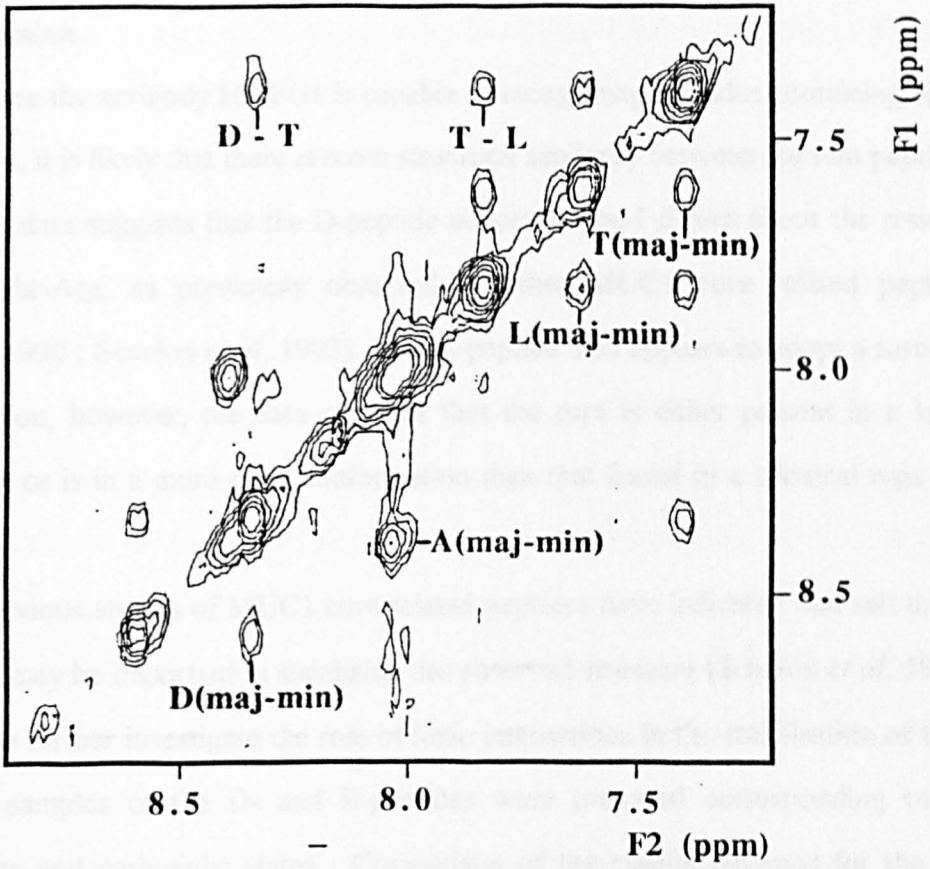bonding. In the I-peptide the amide protons of Asp(maj) and Thr(maj) also appear to be protected from solvent.

Table 7.10b Amide proton temperature coefficients for the I-peptide.

Residue		Temperature Coefficient $(\Delta\delta/\Delta T)$ ppb/K
Ala[1]		
Asp[3]	(maj)	3.7
	(min)	5.5
Thr[4]	(maj)	3.3
	(min)	5.0
Ile[5]	(maj)	5.8
	(min)	5.1
Ala[7]	(maj)	5.1
	(min)	5.5
Gly[9]		7.5

Amide-to-amide connectivities observed for the two peptides are shown in Figure 7.4. The pattern of through-space connectivity observed between the amide protons of Asp(maj)-Thr(maj)-Leu in the L-peptide, in conjunction with the low value of the temperature coefficient of the Leu amide proton strongly suggests that a type-I β-turn exists in this peptide about the residues Pro[2]-Asp-Thr-Leu[5]. In the I-peptide the pattern of connectivity between the amide protons of Asp(maj)-Thr(maj)-Ile(maj) is consistent with the presence of a type-I β-turn, however the intensity of the amide to amide cross peaks in the 2D NOESY experiments is lower than those observed in the L-peptide. In addition, the temperature coefficient of the Ile(maj) amide proton indicates that it is freely accessible to solvent. Consequently if such a turn is present in the I-peptide then it is likely to represent a very much lower population than the turn observed in the L-peptide.

Figure 7.4a   Amide to amide region of the 600ms mixing time ^{1}H NOESY spectrum of the L-peptide recorded at 400MHz and 303K.

Figure 7.4b   Amide to amide region of the 600ms mixing time ^{1}H NOESY spectrum of the I-peptide recorded at 400MHz and 303K.

7.4 Discussion

Since the antibody DO8.Q is unable to recognise peptides containing either Asp or Glu, it is likely that there is some structural requirement in the two peptides. The n.m.r. data suggests that the D-peptide adopts a type I β-turn about the residues Pro-Asp[?]-Ile-Arg, as previously observed for some MUC1 core related peptides (Tendler, 1990; Scanlon et al, 1992). The [?] peptide does appear to adopt a turn-like conformation, however, the data suggests that the turn is either present in a lower population or is in a more open conformation than that found in a classical type I β-turn.

Previous studies of MUC1 core related peptides have indicated that salt bridge formation may be important in stabilising the observed structure (Scanlon et al, 1992). In order to further investigate the role of ionic interaction in the conformation of these peptides, samples of the D- and E-peptides were prepared corresponding to the carboxylate and carboxylic states. Comparison of the titration curves for the two peptides in the different ionisation states clearly showed that both the conformations adopted are sensitive to changes in pH. In particular, both data sets show that the observed β-turns are stabilised by a salt bridge involving the carboxylate side chain of the Asp/Glu residue.

In an attempt to identify the more favourable conformation present in such an interaction, the V-peptide was investigated. The conformation of the V-peptide in d₆-dmso was found to be significantly different to other MUC1 core related peptides. Although the n.m.r. data was unable to unambiguously ascribe any major element of secondary structure to the peptide, the observation of multiple weak connections between the amide proton of Asp[?] [...] coupled with the fact that previously studied peptides containing the sequence Pro-Asp-Thr-Arg form a type I β-turn in this region, suggests that the V-peptide may also be capable of adopting a similar structural motif. While investigation of the n.m.r. parameters may explain why unable to adopt conformations analogous with the presence of a typical β-turn are observed, while the temperature coefficient of the NH2 amide proton

7.4 Discussion

Since the antibody HMFG1 is capable of recognising peptides containing either Asp or Glu, it is likely that there is some structural similarity between the two peptides. The n.m.r. data suggests that the D-peptide adopts a type-I β-turn about the residues Pro-Asp-Thr-Arg, as previously observed in other MUC1 core related peptides (Tendler, 1990 ; Scanlon *et al*, 1992). The E-peptide also appears to adopt a turn-like conformation, however, the data suggests that the turn is either present in a lower population or is in a more open conformation than that found in a classical type-I β-turn.

Previous studies of MUC1 core related peptides have indicated that salt bridge formation may be important in stabilising the observed structure (Scanlon *et al*, 1992). In order to further investigate the role of ionic interactions in the stabilisation of these peptides, samples of the D- and E-peptides were prepared corresponding to the carboxylate and carboxylic states. Comparison of the results obtained for the two peptides in the different ionisation states clearly suggested that the conformations adopted are sensitive to changes in pH. It appears likely, therefore, that the observed β-turns are stabilised by a salt bridge involving the side chain carboxylate group of the Asp/Glu residues.

In an attempt to identify the more likely positively charged partner in such an interaction, the V-peptide was investigated. The conformation of the V-peptide in d_6-dmso was found to be significantly different to other MUC1-core related peptides. Although the n.m.r. data was insufficient to unequivocally describe any single element of secondary structure in the peptide, the observation of through space connections between the amide protons of Asp[6]-Thr[7] and Thr[7]-Arg[8], in conjunction with the fact that previously studied peptides containing the sequence Pro-Asp-Thr-Arg form a type-I β-turn in this region, suggests that the V-peptide may also be capable of adopting a similar structural feature. Non-linear averaging of the n.m.r. parameters may explain why amide to amide connections consistent with the presence of a type-I β-turn are observed, while the temperature coefficient of the Arg amide proton

indicates that it is freely accessible to solvent. In a type-I β-turn it would be expected that this amide proton would be involved in a hydrogen bond with the carbonyl oxygen of Pro resulting in a lower value of its temperature coefficient. The only amide proton in the V-peptide which is protected from solvent is that of the Thr[7] residue. In previously studied MUC1 core related peptides the threonine residue at this position in the tandem repeat sequence has been found to have a lowered amide proton temperature coefficient. On the basis of molecular modelling data (Scanlon *et al*, 1992) it has been suggested that the low value of the temperature coefficient arises due to the presence of an inter-residue hydrogen bond from this amide proton to the side chain carboxyl oxygen of the Asp. However, the value of the Thr[7] amide proton temperature coefficient in the V-peptide is significantly lower than has been observed for any other peptide studied to date. This may suggest that this amide proton is further protected from solvent in some way.

The data obtained for the V-peptide does not clarify the role of ionic interactions in the stabilisation of structural features observed in MUC1 related peptides. It appears that the V-peptide may still form a low population of type-I β-turn about the residues Pro-Asp-Thr-Arg, although the n.m.r. evidence for such a conformation is not conclusive. This may suggest that in the previously studied MUC1 core related peptides an interaction between the negatively charged Asp side chain and the positively charged N-terminus stabilises the observed type-I β-turn. In the V-peptide the distance between the Asp residue and the N-terminus is considerably larger, consequently a similar ionic interaction is less likely. The inability of the V-peptide to form such an interaction may account for the less well defined structure in this region. However, molecular modelling studies suggested that the lowest energy conformation of the peptide Pro-Asp-Thr-Arg-Pro involved a type-I β-turn stabilised by an ionic interaction between the side chains of the Asp and Arg residues (Scanlon *et al*, 1992). Since protonation of the Asp/Glu side chain carboxyl groups in the D-/E-peptides removes the capacity of the peptide to form a β-turn, it is likely that ionic interactions play some part in stabilising the observed structure in these peptides.

Therefore, if the through space connections observed between the amide protons of Asp-Thr-Arg in the V-peptide are indeed indicative of a low population of β-turn about Pro-Asp-Thr-Arg, then it may be that a salt bridge involving the Asp and Arg residues has a role in the stabilisation of this structural element.

Data obtained for the I- and L-peptides indicate that in the case of the L-peptide at least, an ionic interaction between the side chains of residues in the MUC1 core related peptides is not essential for β-turn formation. Surprisingly, the L- and I-peptides appear to adopt somewhat different conformations. The temperature coefficient of the Ile(maj) amide proton in the I-peptide suggests that if a β-turn is indeed present then it is a less well defined structural element than the turn observed in the L-peptide.

The structural differences observed for the D-/E-peptides and the I-/L-peptides may account for the observed difference in the reactivity of anti-MUC1 monoclonal antibodies towards these peptides. However the n.m.r. data provides little indication of what is the most likely conformation of the peptides when interacting with the antibody. In the antibody-peptide complexes it is likely that the majority of the interaction between the molecules arises due to shape complementarity of the peptide surface and the binding site. Elements of secondary structure such as β-turns may be important in orienting the peptide in such a way that antibody binding is possible. However it is likely that the backbone atoms which form the turn will be masked by the side chains of residues in the peptide. Consequently, in order to identify a likely binding structure, it is necessary to characterise the conformations of side chain groups in the peptides. Due to the flexibility of linear peptides in solution it is not usually possible to do this from the n.m.r. data. Indeed, in the case of the MUC1 core related peptides studied to date, it is generally the case that the vast majority through space connections involve only backbone protons. Therefore, in order to unambiguously assign the structural features of the peptides which are important for recognition by anti-MUC1 antibodies it will be necessary to elucidate their conformations when bound in the complex.

CHAPTER EIGHT

TOWARDS INVESTIGATIONS OF THE BOUND
CONFORMATIONS OF P(1-20)

8.1 Introduction

A number of n.m.r. techniques are now available which enable the study of peptides and small proteins in complexes with antibody fragments (Section 1.9). The applicability of a particular method depends fundamentally on the rate at which the peptide antigen exchanges between the bound and free state. Consequently, in order to determine which is the most suitable technique, it is necessary to quantify the affinity of the antibody for its antigen. In addition, due to the size constraints imposed by n.m.r. spectroscopy, it is advantageous to use the smallest easily obtainable fragment of the antibody which retains its binding activity, usually the Fab. Finally, since all of these techniques involve observation of the antigen in exchange between the bound and the free states, it is necessary to assign the proton resonances of the free peptide in the aqueous conditions used for the n.m.r. studies of the antibody-antigen complex.

The following chapter describes the initial steps towards studies of the conformation of peptides based on the tandem repeat sequence of the MUC1 protein core in complex with the anti-mucin antibody C595.

8.2 Proteolytic Digestion of Antibody

The monoclonal antibody C595 was digested on a column of immobilised papain (Section 4.3.4). The optimum digestion time was determined by removing aliquots of material from the column which were analysed by SDS-PAGE on 12.5% polyacrylamide gels. The optimum digestion time was taken to be the point at which whole antibody could no longer be detected on the gel using a sensitive silver stain (Figure 8.1).

The proteolytic digest was purified as described in Section 4.3.5 and Western blotted onto a nitro-cellulose membrane. Development of the membrane with sheep-anti-mouse-kappa-light-chain antibodies revealed a band of molecular weight 50,000 to be present (Figure 8.2), which is consistent with the presence of the Fab fragment of the antibody. The yield of Fab in these studies was 10-12%. Shorter digestion times on the papain column produced lower yields of Fab.
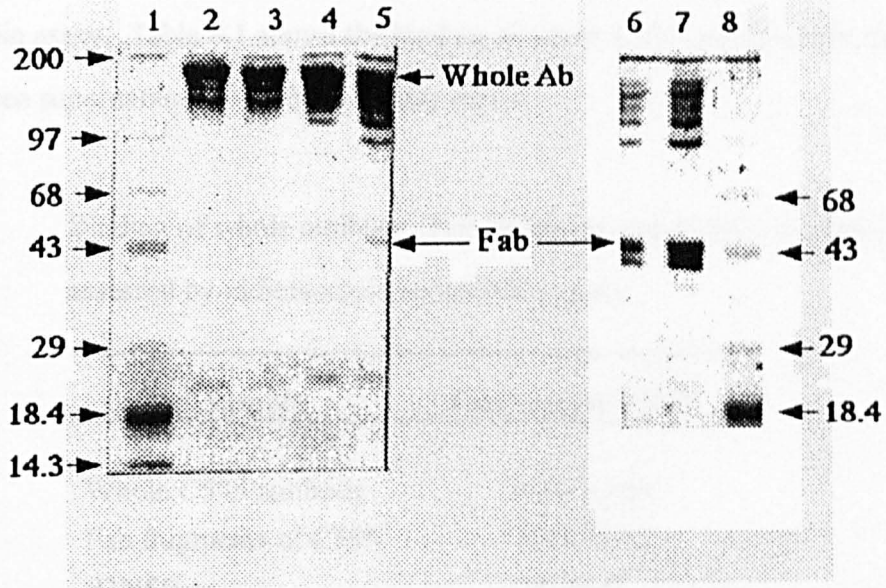
Figure 8.1   Time-course for the digestion of monoclonal antibody C595 on a column of immobilised papain. Lanes 1 and 8 contain molecular weight standards the sizes of which are marked in kDa. Lanes 2 to 7 contain the crude proteolytic digest removed from the column after incubation times of 0, 1h, 2h, 4h, 18h and 24h respectively.
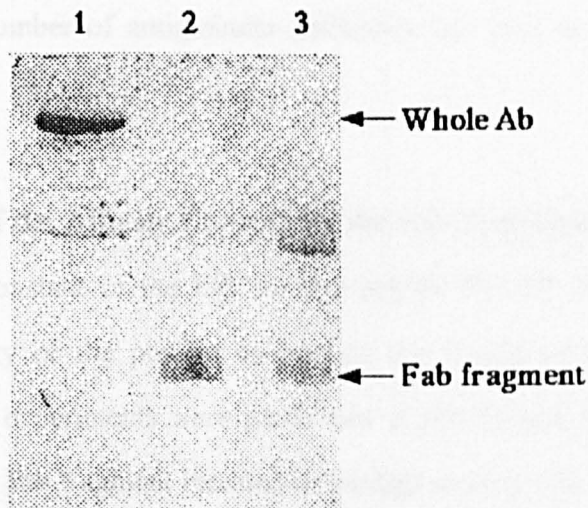

Figure 8.2   Western blot of the papain digest of monoclonal antibody C595 developed with sheep-anti-mouse-kappa-light-chain antibodies. Lane 1, whole C595. Lane 2, purified Fab fragment of C595. Lane 3, crude digest after 18 hours.

The Fab fragment was tested for its reactivity with MUC1 in a radioisotopic antiglobulin assay. Table 8.1 shows the binding of whole C595, purified Fab fragments and cell free supernatant (P3NS0) to urinary mucin.

Table 8.1   Binding of whole antibody, Fab fragments and P3NS0 to urinary mucin assessed by radioisotopic antiglobulin assay.

Sample	Counts/min ± S.D.
Whole C595 antibody	8346 ± 346
Fab fragments of C595	1743 ± 98
P3NS0	81 ± 43

The assay clearly demonstrates that the Fab retains its binding activity for the mucin. The lower absolute value of counts recorded with the Fab compared to the whole antibody reflects the fact that the binding to mucin was detected with radiolabelled $F(ab')_2$ fragments from a polyclonal rabbit-anti-mouse-immunoglobulins antiserum. It is generally accepted that, due to the absence of the Fc region in the Fab fragment, a greater number of antiglobulin antibodies can bind to a whole antibody molecule.

8.3 Determination of the Affinity Constant of the Fab Fragment

The binding constant of the Fab for the peptide P(1-20) was determined by measuring the capacity of the peptide to quench the fluorescence of the antibody fragment. Quenching experiments were performed at pH 7.0 and pH 3.5 in order to determine whether the Fab fragment retained its binding activity at lowered pH. At pH 7.0 in aqueous solution, the exchange rate of the amide protons in linear peptides is such that they cannot usually be observed in n.m.r. experiments. Lowering the pH slows the exchange rate, thereby enabling resonances of amide protons to be observed. Since important structural information regarding the conformation of peptides relies on

the observation of amide proton resonances, it is advantageous to undertake n.m.r. experiments at low pH. Clearly, in studies of antibody-antigen complexes, this is only valid if the antibody retains its specificity for the peptide under these conditions.
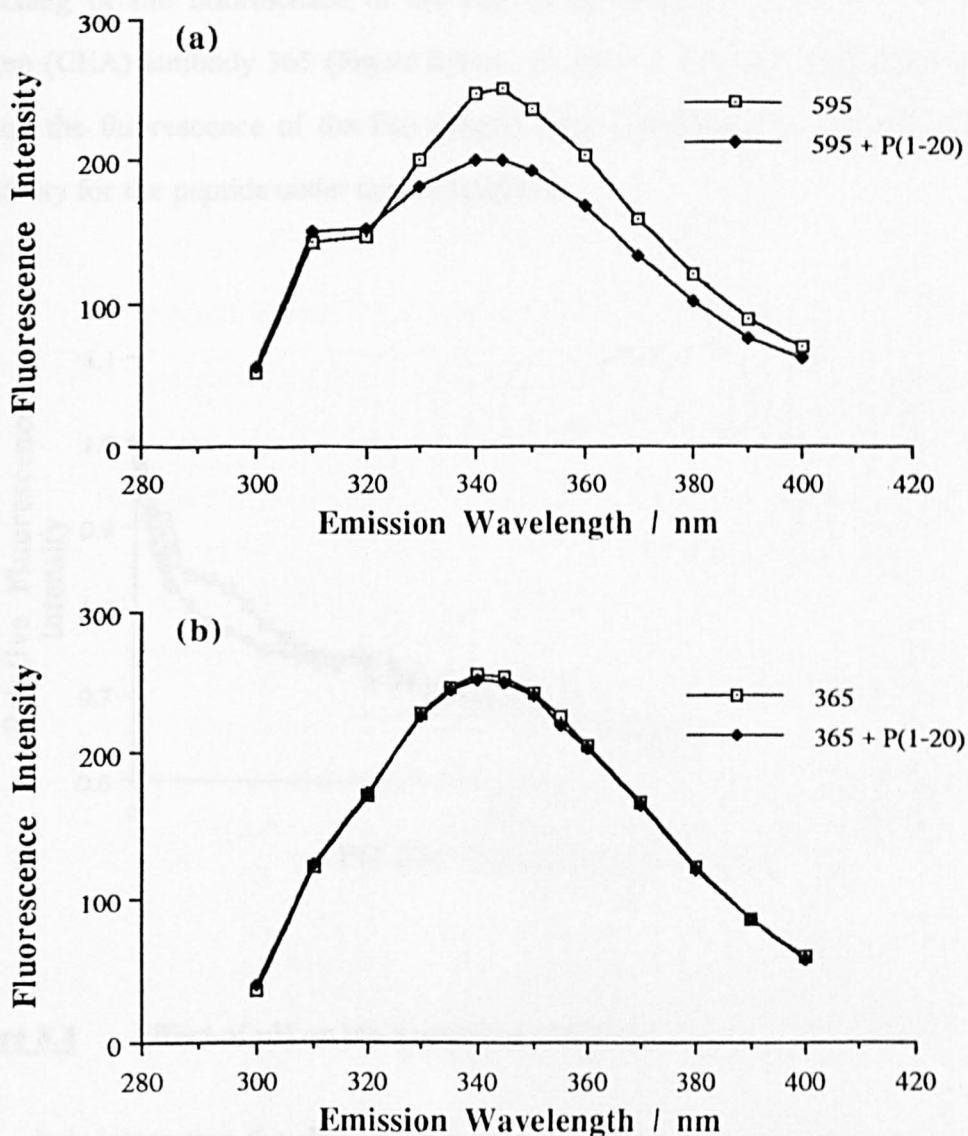

Figure 8.3   Effect of peptide on the fluorescence of antibody Fab fragments. (a) Fab of anti-MUC1 antibody C595. (b) Fab of anti-CEA antibody 365.

At pH 7.0, saturation of the Fab fragment of C595 with a 40-fold molar excess of P(1-20) resulted in a 24% quenching of Fab fluorescence (Figure 8.3a). A similar molar excess of the peptide P(7-20) which does not contain the identified epitope for C595 caused no quenching of Fab fluorescence. Similarly P(1-20) caused no quenching of the fluorescence of the Fab of an irrelevant anti-carcino-embryonic-antigen (CEA) antibody 365 (Figure 8.3b). At pH 3.5, P(1-20) retained its ability to quench the fluorescence of the Fab (Figure 8.4), indicating that the Fab retains its specificity for the peptide under these conditions.
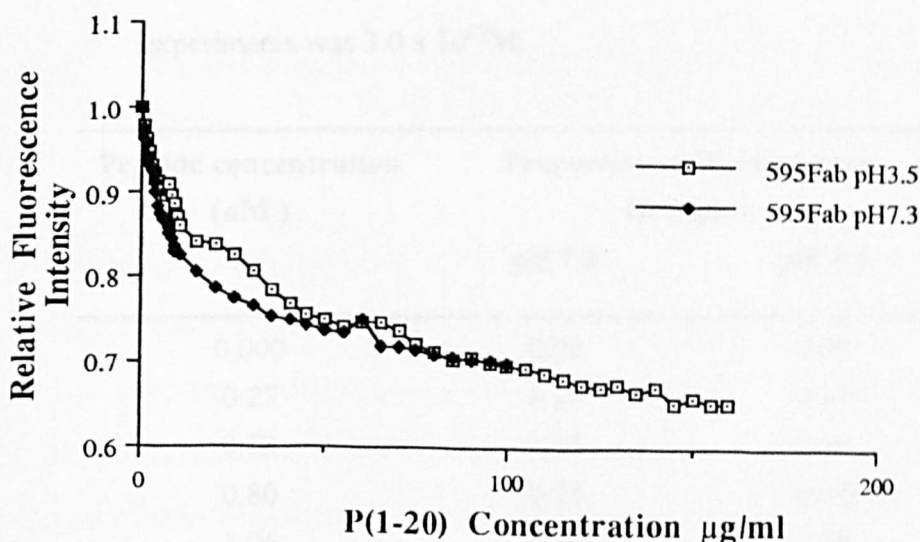
Figure 8.4   Effect of pH on the quenching of Fab fluorescence by P(1-20).

It is interesting that P(1-20) can quench the fluorescence of the C595 Fab to such a large extent. The fluorescence of proteins arises primarily as a result of the tryptophan residues they contain. As (P1-20) contains no chromaphoric group with an absorption overlapping the excitation or emission wavelengths of tryptophan, the quenching cannot occur as a result of fluorescent energy transfer. Furthermore, as the Fab fragments were present in sub-micromolar concentrations in these experiments, the inner filter effect would have been negligible. Therefore the most likely explanation for

the observed quenching is that the peptide interacts with the tryptophan residue(s) in the Fab, causing a significant change in their environment.   This implies that Tryptophan must be involved either at or very close to the binding site of the antibody.

The binding constant of the Fab fragment of C595 for P(1-20) was determined from the fluorescence quenching data. The fractional occupancy of the binding sites at different concentrations was expressed as a proportion of the maximal quenching (Table 8.2).

Table 8.2   Proportion of Fab binding sites occupied at different peptide concentrations.   The concentration of Fab fragments used in these experiments was 3.0×10^{-7}M.

Peptide concentration (μM)	Proportion of Binding Sites Occupied	
	pH 7.3	pH 3.5
0.000	0.00	0.00
0.27	0.10	0.07
0.53	0.15	0.10
0.80	0.18	0.12
1.06	0.25	0.18
1.33	0.30	0.21
1.60	0.35	0.25
1.86	0.42	0.29
2.13	0.51	0.32
2.39	0.63	0.34
2.66	0.73	0.38
2.93	0.77	
3.19	0.81	

The binding constant was determined from this data by Scatchard analysis (Figure 8.5).   The binding constant of the Fab fragment of C595 for P(1-20) was calculated to be 0.26×10^{-6} M^{-1} at pH 7.3 and 0.13×10^{-6} M^{-1} at pH 3.5.
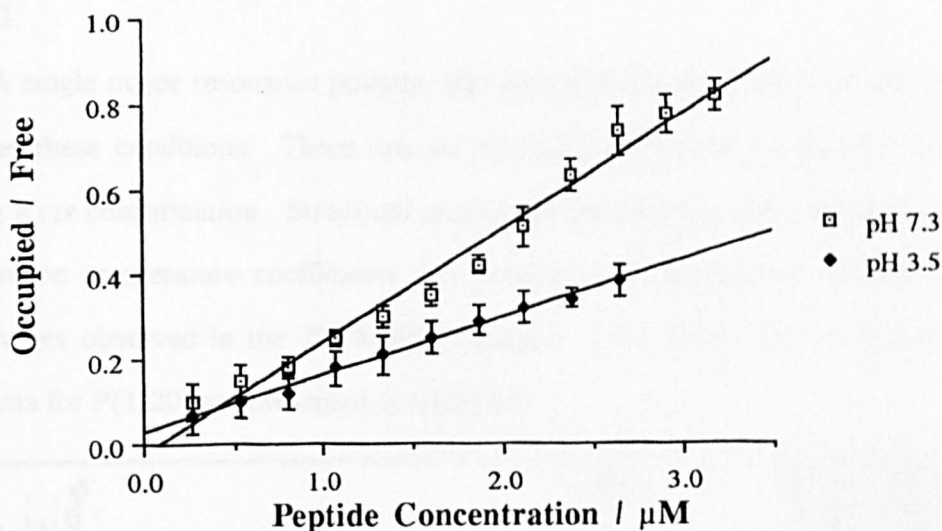
Figure 8.5   Scatchard analysis of the binding of peptide P(1-20) to Fab fragments
of C595.

8.4 N.m.r. Studies of P(1-20) in Aqueous Solution

The ^{1}H resonances of P(1-20) in 100mM deuteroacetate buffer at pH 4.4 were
assigned on the basis of chemical shift and connectivity in 1D and 2D n.m.r. spectra
using the methods described in Chapter 3.   Due to the problems associated with
suppression of the water signal in aqueous n.m.r. spectra experiments were performed
in both D_2O and 90% H_2O/10% D_2O.   Firstly the spin systems of the non-
exchangeable proton resonances were identified from COSY and HOHAHA spectra of
the peptide in D_2O.   The aliphatic region of a HOHAHA spectrum of P(1-20) in D_2O
buffer is shown in Figure 8.6.   Amide proton resonances were assigned from NH-side
chain connectivities in a HOHAHA spectrum performed in 90% H_2O/10% D_2O.   The
amide-to-aliphatic region of a HOHAHA spectrum of P(1-20) in 90% H_2O/10% D_2O
buffer is shown in Figure 8.7.   Sequence specific assignments were based on through-
space connectivities observed in ROESY spectra in 90% H_2O/10% D_2O.   As was the
case with P(1-20) in d_6-dmso, no through space connections were observed in 2D

NOESY spectra of the peptide. The full assignments for P(1-20) are presented in Table 8.3.

A single major resonance position was identified for all of the protons in P(1-20) under these conditions. There was no indication of any of the proline residues adopting a *cis* conformation. Structural analysis of the peptide was undertaken using amide proton temperature coefficients and distance constraints from through space connectivities observed in the 2D ROESY spectra. The amide proton temperature coefficients for P(1-20) are presented in Table 8.4.
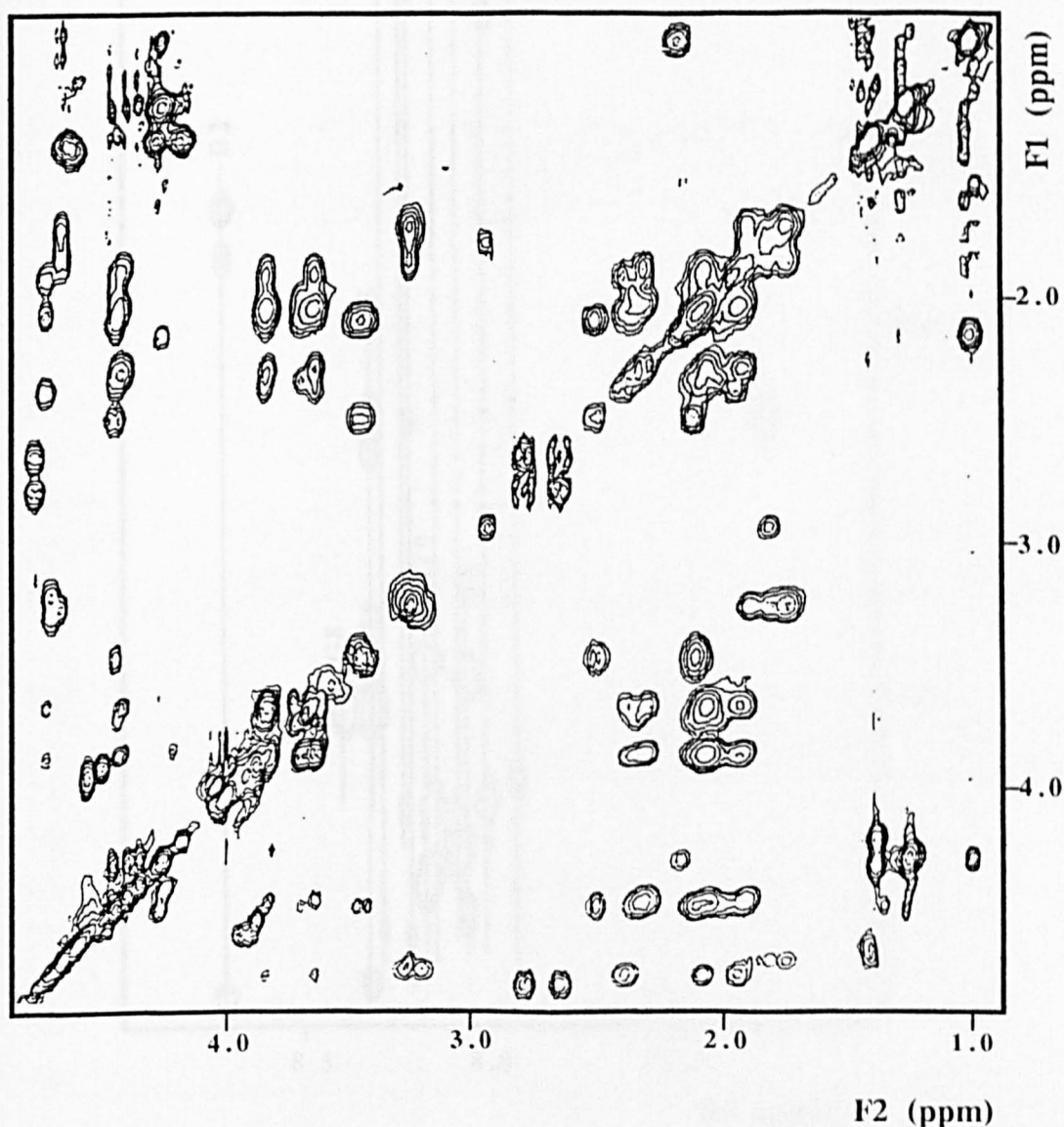
Figure 8.6   Aliphatic region of the 40ms mixing time ^{1}H HOHAHA spectrum of P(1-20) in D$_2$O recorded at 400MHz and 310K.
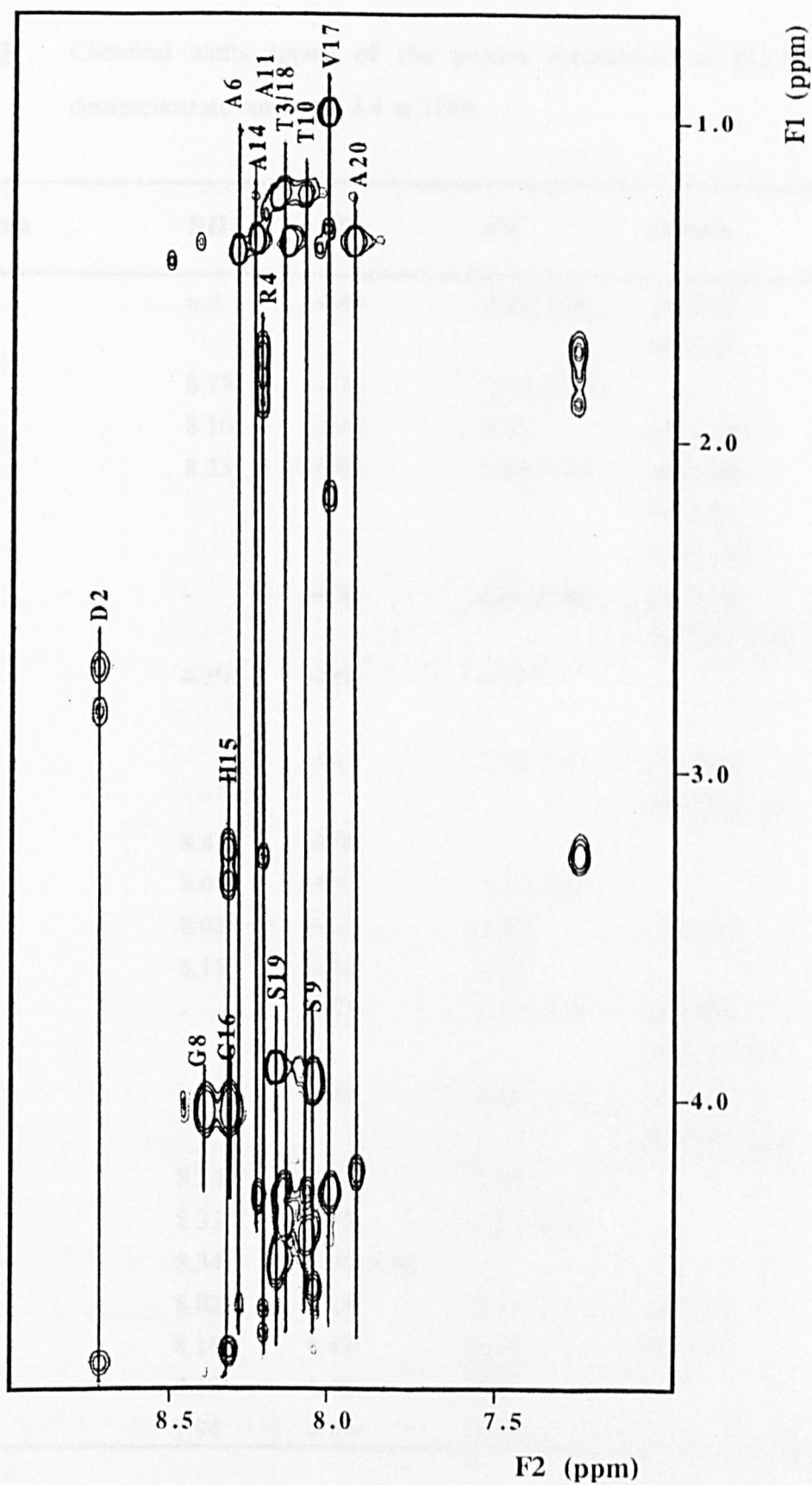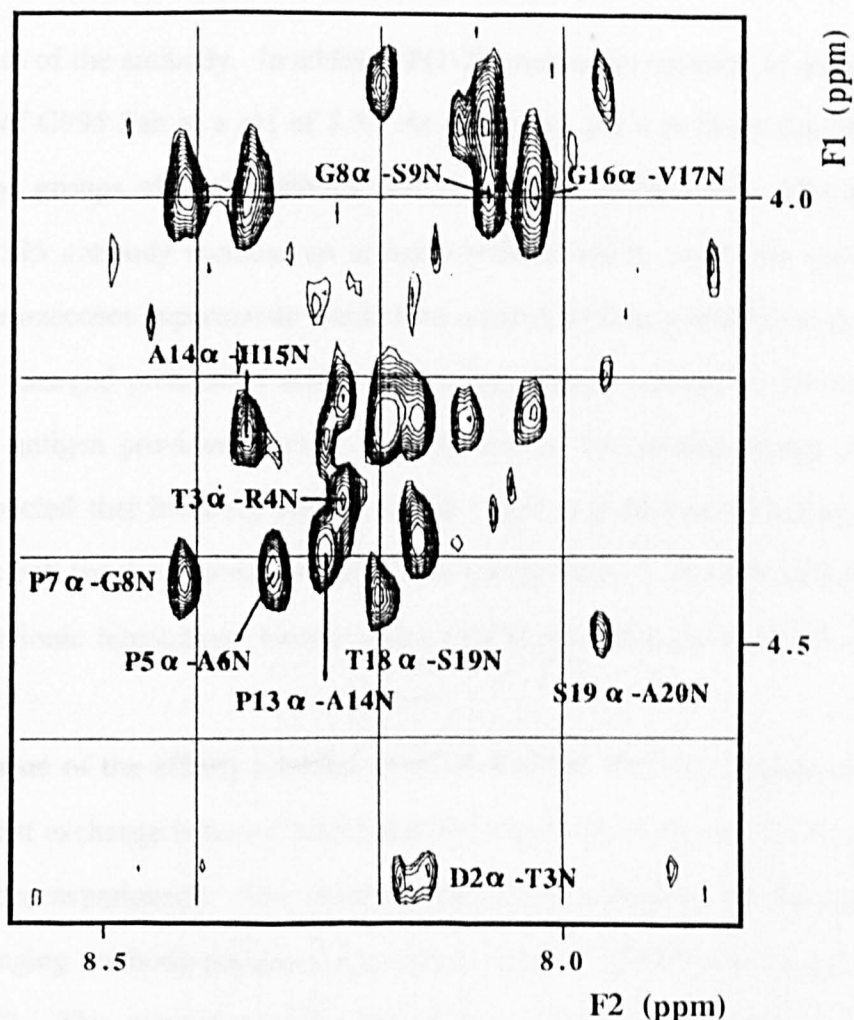
Figure 8.7   Amide to aliphatic region of the 40ms mixing time ^{1}H HOHAHA

spectrum of P(1-20) in 90%H_2O/10%D_2O recorded at 500MHz and

310K.

Table 8.3   Chemical shifts (ppm) of the proton resonances of P(1-20) in deuteroacetate buffer pH 4.4 at 310K.

Residue	NH	αH	βH	Others
Pro[1]	n.d.	4.44	2.47, 2.07	γH 2.07
				δH 3.44
Asp[2]	8.73	4.76	2.79, 2.65	
Thr[3]	8.16	4.32	4.21	γH 1.19
Arg[4]	8.23	4.65	1.85, 1.76	γH 1.68
				δH 3.22
				εNH 7.27
Pro[5]	-	4.42	2.29, 1.88	γH 2.04
				δH 3.82, 3.62
Ala[6]	8.30	4.60	1.38	
Pro[7]	-	4.43	2.32, 1.97	γH 2.07
				δH 3.82, 3.67
Gly[8]	8.41	3.98		
Ser[9]	8.07	4.53	3.92, 3.89	
Thr[10]	8.09	4.37	4.24	γH 1.20
Ala[11]	8.15	4.63	1.36	
Pro[12]	-	4.71	2.35, 1.89	γH 2.04
				δH 3.83, 3.63
Pro[13]	-	4.39	2.26, 1.84	γH 2.02
				δH 3.80, 3.63
Ala[14]	8.24	4.26	1.34	
His[15]	8.33	4.75	3.32, 3.21	
Gly[16]	8.34	4.03, 3.93		
Val[17]	8.02	4.25	2.13	γH 0.95
Thr[18]	8.18	4.43	4.24	γH 1.21
Ser[19]	8.19	4.48	3.86	
Ala[20]	7.94	4.19	1.35	

Table 8.4   Amide proton temperature coefficients for P(1-20) in deuteroacetate buffer pH 4.4.

Residue	Temperature Coefficient $(\Delta\delta/\Delta T)$ ppb/K
Asp^2	6.4
Thr^3	7.8
Arg^4	9.6
Ala^6	8.4
Gly^8	7.6
Ser^9	7.4
Thr^{10}	7.3
Ala^{11}	8.2
Ala^{14}	8.3
His^{15}	8.0
Gly^{16}	7.0
Val^{17}	5.7
Thr^{18}	7.5
Ser^{19}	8.1
Ala^{20}	7.6

The temperature coefficients of all the amide protons in P(1-20) are relatively high and provide no evidence of hydrogen bonding in the peptide. Furthermore, the pattern of through space connectivity observed in the 2D ROESY experiments (Figure 8.8) suggests that the peptide exists in random coil conformations. Strong $d_{\alpha N}$ connections are present along the entire backbone of the peptide while no d_{NN} connectivities are observed. Taken in conjunction, these results suggest that under the conditions used in these n.m.r. experiments P(1-20) adopts no preferred structure in aqueous solution.

Figure 8.8   Amide to aliphatic region of the 200ms mixing time ^{1}H ROESY spectrum of P(1-20) in 90%H_2O/10%D_2O recorded at 600MHz and 310K.

8.5 Discussion

Binding data obtained in the radioisotopic antiglobulin assay (Section 8.2) reveals that the proteolytically obtained Fab fragment of monoclonal antibody C595 retains its specificity for MUC1.   Furthermore, since P(1-20) quenches the native fluorescence of C595 Fab, it is clear that the antibody fragments are capable of recognising and binding to peptides containing the identified epitope for C595.   The fluorescence quenching data also provides information regarding the nature of the antibody binding site.   The fact that P(1-20) is capable of quenching the Fab fluorescence implies that tryptophan residue(s) are involved either at or very close to

the binding site of the antibody. In addition P(1-20) retains its capacity to quench the fluorescence of C595 Fab at a pH of 3.5. At this lower pH it is likely that the side chain carboxyl groups of both antibody and peptide are protonated. The epitope defined by C595 antibody contains an arginine residue which, under the conditions used in the fluorescence experiments would have existed with its guanidinium group in the positively charged protonated state. If a charge-charge interaction between the antibody and antigen provided a major contribution to the binding energy, then it would be expected that lowering the pH would result in a marked reduction in the affinity of the Fab for the peptide. Since this is not the case, it appears unlikely that intermolecular ionic interactions have a major role in the recognition of P(1-20) by C595.

The value of the affinity constant of C595 Fab for P(1-20) suggests that the peptide is in fast exchange between bound and free states under the conditions used in the fluorescence experiments. The most suitable n.m.r. technique for the study of rapidly exchanging antibody-peptide complexes is that of 2D-TRNOE spectroscopy (Section 1.9.2). This method provides information regarding the conformations of bound peptide antigens by detecting magnetisation transferred from bound antibody protons to bound peptide protons. As the peptide is in fast exchange between bound and free states, such transfer of magnetisation is detected in the form of cross peaks between the resonance position for the antibody proton and that of the corresponding peptide proton in the unbound state. Consequently, for interpretation of 2D-TRNOE spectra it is necessary to have obtained resonance assignments for the free peptide under the same conditions used for the n.m.r. experiments on the antibody-peptide complex. Due to the advantages for conformational analysis of recording n.m.r. spectra of aqueous peptides at low pH, and in view of the fact that the C595 Fab fragment retains its specificity for P(1-20) at pH 3.5, it is proposed to undertake 2D-TRNOE experiments on the complex at lowered pH. The proton resonances of P(1-20) have been assigned in deuteroacetate buffer at pH 4.4. Under these conditions there is no indication of the peptide adopting stable conformations. However, the

fluorescence quenching experiments suggest that ionic interactions may not be a significant factor in the binding of C595 Fab to P(1-20). Therefore, it is likely that the peptide, when docking into the antibody binding site enters a less polar environment to that found in aqueous buffer. Consequently it would not be surprising if the peptide were found to adopt a radically different conformation at the binding site of the complex to the random coil populations it adopts in deuteroacetate buffer. Given that P(1-20) has previously been shown to form stable conformations in d_6-dmso, the 2D-TRNOE experiments may also be useful in addressing the question of the suitability of different solvents in investigations of the conformation of antigenic peptides.

Due to the large size of the Fab it is often necessary to use extensive deuteration in order to resolve antibody proton resonances in the spectrum of the complex (Anglister *et al*, 1988). In n.m.r. investigations of antibody-peptide complexes by 2D-TRNOE it is usual to perform several experiments using Fab fragments with different patterns of deuteration. Therefore it is necessary to have an efficient means for producing the Fabs. The yield of C595 Fab obtained by proteolysis with papain is such that this is clearly an unsuitable means of producing Fab fragments for n.m.r. experiments. Furthermore from the SDS-PAGE gels of the antibody digest and Western blots of the Fab it appears that more than one species of Fab is produced by enzymatic proteolysis. Given the complexity of 2D-TRNOE spectra it is essential to have a single Fab species present in these experiments. Therefore before such studies can be undertaken it will be necessary to develop a more efficient means of producing the Fab fragment of C595.

CHAPTER NINE

ANTIBODY SEQUENCING

9.1 Introduction

Improvements in techniques for manipulation of DNA have made it possible to routinely express fragments of antibodies in bacterial cells (Reichmann *et al*, 1988). The production of recombinant Fv and Fab has a number of advantages over the alternative proteolytic methods for obtaining these fragments. Firstly, the recombinant fragments are highly homogeneous, which is essential if they are to be studied either by X-ray crystallography or n.m.r. spectroscopy. Secondly, if isotopic enrichment with n.m.r active nuclei is required, the high yield of antibody fragments obtained in this way is beneficial. The production of a DNA construct which codes for the binding domain facilitates acquisition of sequence data for the antibody. With a knowledge of the antibody sequence it is possible to develop computer models of the binding site. Furthermore, such a construct makes possible the generation of a panel of point mutated antibodies which may have differing specificities for the antigen. These techniques may be of fundamental importance in elucidating the processes of molecular recognition involved in antibody-antigen interactions.

9.2 Amplification of C595 DNA

The DNA coding for the light chain (κ) and the constant chain variable domain (V_H) of antibody C595 was obtained by PCR amplification as described in Section 4.3.10. Figure 9.1 shows an agarose gel of the products of the PCR amplification. Negative controls for the PCR reactions were included in the form of RNA extracts to which no reverse transcriptase was added prior to PCR.

As can be clearly observed from the gel, PCR fragments of the correct molecular weight were obtained for both the V_H-domain and the κ-chain suggesting that the correct portion of DNA was amplified in both cases.

9.3 Cloning of PCR Products

The PCR products were ligated into the PCR™ II vector and selected as white colonies on agar plates coated with X-Gal. Plasmid, prepared from overnight cultures

of transformed *E.Coli*, was tested for incorporation of the appropriate PCR fragment by restriction endonuclease digestion. Figure 9.2 shows the digests of plasmid into which had been incorporated either the V_H-domain or the κ-chain PCR products.
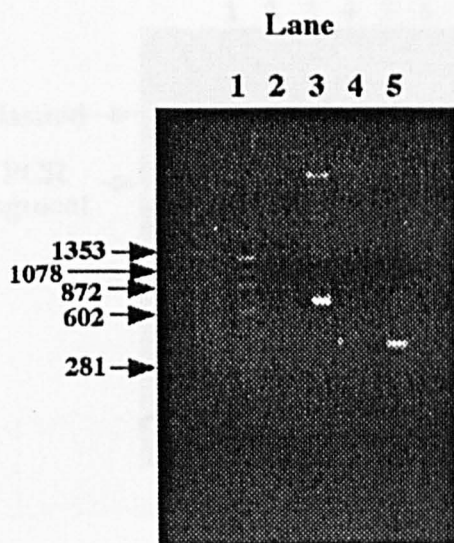
Lane

1 2 3 4 5

1353 →
1078
872 →
602 →

281 →

Figure 9.1   Agarose gel (2%) of the V_H-domain and the κ-chain PCR products for C595. Lane 1 contains the molecular weight standards, the sizes of which are quoted in base pairs. Lane 2 contains the negative control for κ-chain PCR. Lane 3 contains the κ-chain PCR product. Lane 4 contains the negative control for V_H-domain PCR. Lane 5 contains the V_H-domain PCR product.

The agarose gels confirm that the DNA inserted into the vector was of the same size as the corresponding PCR product suggesting that the correct DNA fragment was inserted.

9.4 DNA Sequencing

Autoradiographs of sequencing gels for which the plasmid containing either the V_H-domain or the κ-chain PCR products as inserts are shown in Figure 9.3.

Table 9.1 contains both the DNA and protein sequences for the V_H-domain and the V_L-domain of C595. The CDRs of both domains are underlined.

(a)

(b)
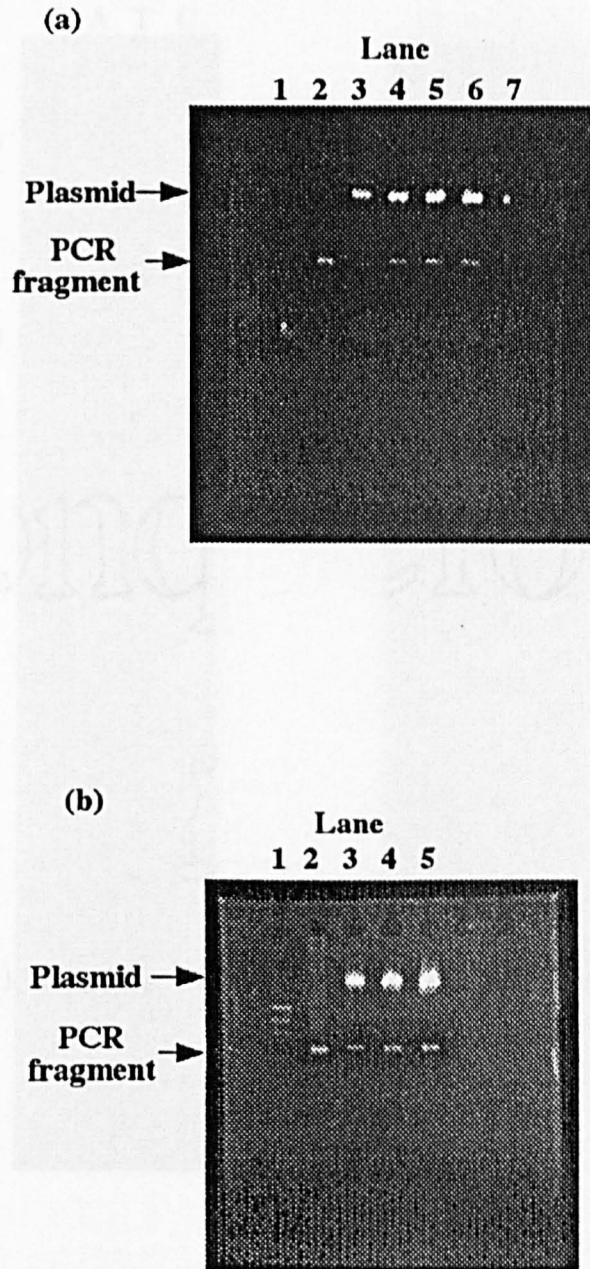
Figure 9.2   Restriction digests of plasmid containing either (a) V_H - or (b) κ-chain DNA insert. In both cases, lane 1 contains the molecular weight standards, lane 2 contains the corresponding PCR product, and the remainder of the lanes contain digests of plasmid preparations.
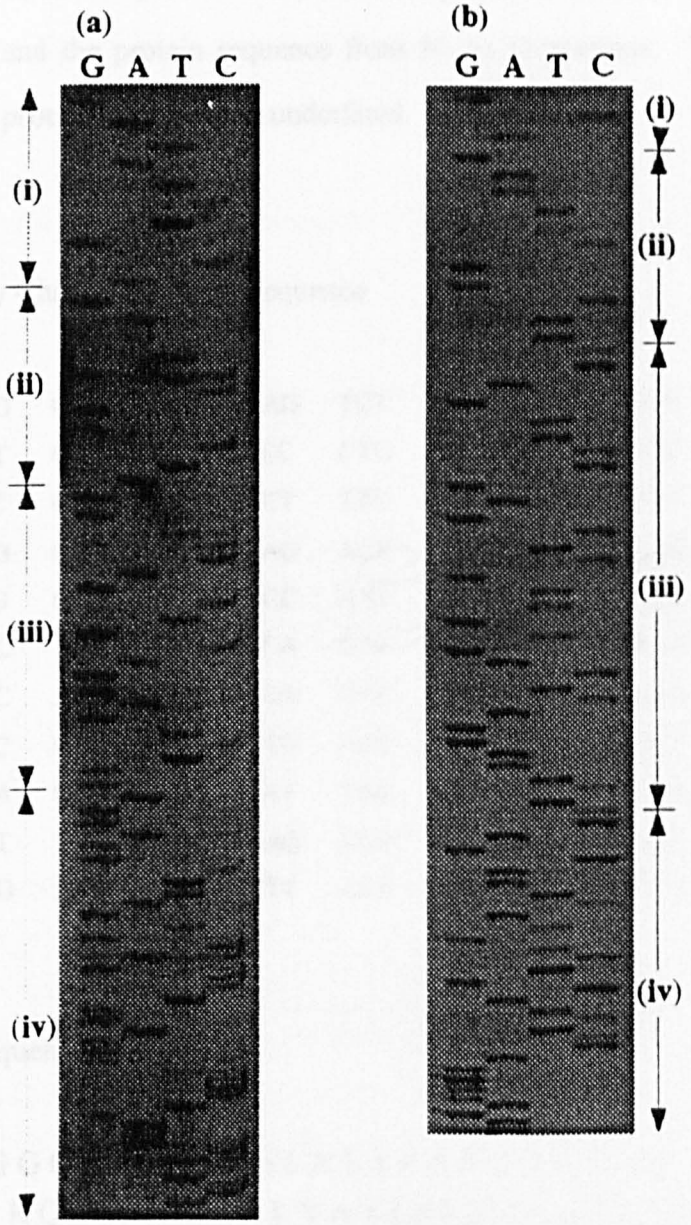
Figure 9.3 Autoradiographs of the 2h DNA sequencing gels of the (a) V_H and (b) V_L domains of antibody C595. The gels have been labelled to show the sequence corresponding to (i) plasmid, (ii) cloning site, (iii) primer and (iv) C595 antibody.

Table 9.1a   DNA and protein sequences of the variable domain of the heavy chains of antibody C595. The DNA sequence is listed in the direction 5'-3', and the protein sequence from N- to C-terminus. The CDRs in the protein sequence are underlined.

Variable Heavy Chain (V$_H$) DNA Sequence

GTC	CAG	CTG	CTG	GAG	TCT	GGA	GGA	GGC	TTA	GTG
CAG	CCT	GGA	GGG	TCC	CTG	AAA	CTC	TCC	TGT	GCA
GCC	TCT	GGA	TTC	ACT	TTC	AGT	AGC	TAT	GGC	ATG
TCT	TGG	GTT	CGC	CAG	ACT	CCA	GAC	AAG	AGG	CTG
GAG	TTG	GTC	GCA	ACC	ATT	AAT	AGT	AAT	GGT	GGT
AGC	ACC	TAT	TAT	CCA	GAC	AGT	GTG	AAG	GGC	CGA
TTC	ACC	ATC	TCC	AGA	GAC	AAT	GCC	AAG	AAC	ACC
CTG	TAC	CTG	CAA	ATG	AGC	AGT	CTG	AAG	TCT	GAG
GAC	ACA	GCC	ATG	TAT	TAC	TGT	GCA	AGA	GAT	CGG
GAT	GGT	TAC	GAC	GAG	GGA	TTT	GAC	TAC	TGG	GGC
CAA	GGG	ACC	ACG	GTC	ACC	GTC	TCC	TCA		

V$_H$ Protein Sequence

V Q L L E S G G G L V Q P G G S L K L S C A A S G F T F S <u>S</u>
<u>Y G M S</u> W V R Q T P D K R L E L V A <u>T I N S N G G S T Y Y P</u>
<u>D S V K G</u> R F T I S R D N A K N T L Y L Q M S S L K S E D T
A M Y Y C A R <u>D R D G Y D E G F D Y</u> W G Q G G T V T V S S

9.5 Identification of Similar Sequences

The five protein sequences with greatest homology to the identified sequence for the V$_H$ -domain and the V$_L$ -domain of C595 are listed in Table 9.2.

Table 9.1b   DNA and protein sequences of the variable domain of the light chain of

antibody C595. The DNA sequence is listed in the direction 5'-3', and

the protein sequence from N- to C-terminus. The CDRs in the protein

sequence are underlined.

Variable Light Chain (V_L) DNA Sequence

AGT	TCC	GAG	CTC	GTT	GTG	ACG	CAG	GAA	TCT	CCA
GCA	ATC	ATG	TCT	GCA	TCT	CCA	GGG	GAG	AAG	GTC
ACC	ATG	ACC	TGC	AGT	GCC	AGC	TCA	AGT	GTA	AGT
TAC	ATC	GAC	TGG	TAC	CAC	GAG	AAG	TCA	GCG	ACC
TCC	CCC	AAA	AGA	TGG	ATT	TAT	GAC	ACA	TCC	AAA
CTG	GCT	TCT	GGA	GTC	CCT	GCT	CGC	TTC	AGT	GGC
AGT	GGG	TCT	GGG	ACC	TCT	TAC	TCT	CTC	ACA	ATC
AGC	AGC	ATG	GAG	GCT	GAA	GAT	GCT	GCC	ACT	TAT
TAC	TGC	CAG	CAG	TGG	AGT	AGT	AAC	CCA	CCC	ACG
TTC	GGA	GGG	GGG	ACC	AAG	CTG	GAA			

V_L Protein Sequence

S S E L V V T Q E S P A I M S A S P G E K V T M T C <u>S A S S</u>
<u>S V S Y I D</u> W I H E K S A T S P K R W I Y <u>D T S K L A S</u> G V
P A R F S G S G S G T S Y S L T I S S M E A E D A A T Y Y C
<u>Q Q W S S N P P T</u> F G G G T K L E

Of fifty similar sequences identified by database searching all of the
homologous sequences coded for immunoglobulin variable regions of the correct chain
type.   In the case of both the V_H-domain and the V_L-domain the most similar
sequences were murine immunoglobulin. The majority of mismatches between the
C595 and identified homologous sequences occurs within the CDRs .
Furthermore, sequence data obtained for the C_L-domain of C595 showed 100%
homology with a number of other murine C_L-domains (data not shown). Taken in
conjunction these results suggest that the sequence data obtained for C595 is accurate.

Table 9.2 Sequences in the Swissprot database having the greatest homology to the identified protein sequence of (a) the C595 V_H-domain and (b) the C595 V_L-domain.

(a)

Protein	Swissprot Accession Number	Percent Similarity
Mouse Ig heavy chain V-region	HV53-M	90.7
Mouse Ig heavy chain V-region	HV55-M	89.7
Mouse Ig heavy chain V-region	HV54-M	88.7
Mouse Ig heavy chain V-region	HV16-M	72.5
Human Ig heavy chain V-region	HV3G-H	66.9

(b)

Protein	Swissprot Accession Number	Percent Similarity
Mouse Ig kappa chain V-region	KV6F-M	83.8
Mouse Ig kappa chain V-region	KV6H-M	82.9
Mouse Ig kappa chain V-region	KV6I-M	82.9
Mouse Ig kappa chain V-region	KV6J-M	81.9
Mouse Ig kappa chain V-region	KV6G-M	81.9

9.6 Discussion

The fact that the determined protein sequence of the variable domains of C595 show such a high level of homology to other murine immunoglobulin sequences suggests that these data are indeed correct. The observed distribution of tryptophan residues both within and adjacent to the CDRs of C595 is interesting in the light of the ability of MUC1 core related peptides to quench the fluorescence of its proteolytically generated Fab fragments. CDR3 of the light chain contains one tryptophan residue, and there is a further residue directly adjacent to CDR1 of the light chain and CDR1 and CDR3 of the heavy chain respectively. The fact that the peptides quench the

native Fab fluorescence suggests that the environment of one or more of these residues is significantly disturbed upon binding of the peptide.

CDR3 of the heavy chain contains six Asp/Glu amino acid residues which possess side chains which would be negatively charged at physiological pH. The minimum sequence required for C595 to bind to peptides based on the MUC1 tandem repeat sequence is Arg-Pro-Ala-Pro (Section 5.2). Replacement net experiments based on the techniques of Geysen *et al* (1987) have identified that the arginine residue contained within the epitope is essential for antibody recognition of the peptides (Briggs *et al*, 1992). These results may suggest that an ionic interaction between the positively charged side chain of the arginine residue of MUC1 related peptides and the negatively charged heavy chain CDR3 plays an important role in antibody recognition. However, such a hypothesis is difficult to reconcile with the fact that the affinity of C595 Fab fragments for peptide P(1-20) appears remarkably insensitive to pH. More data is required both on the conformation of the antibody combining site and the specific interactions involved in antibody-peptide binding before it will be possible to determine the important events in the processes of molecular recognition.

CHAPTER TEN

GENERAL DISCUSSION AND FUTURE WORK

The MUC1 family of mucins are tumour associated antigens which can express epitopes that are relatively tumour specific. Although MUC1 mucins are also expressed by normal tissues the profile of reactivity with anti-mucin antibodies indicates that there is a difference in the material derived from the two sources. It has been found that the major difference lies in the composition of the carbohydrate side chains and the extent of glycosylation, since the protein cores are identical. Interestingly, however, a number of the antibodies which show preferential tumour reactivity define epitopes within the protein core of the mucin. The explanation for this is thought to be that aberrant glycosylation of the protein core results in the exposure of epitopes which are masked in the normal mucin. Defining the precise conformational requirements for recognition of the mucin by different anti-MUC1 antibodies would provide an appreciation at the molecular level of the structural basis of the observed tumour specificity.

Due to the size and complexity of the mucin molecules, attempts to define their structures in detail have been only partly successful. It has been found, however, that peptides based on the tandem repeat of the MUC1 core protein compete for the same binding site on the antibodies as the intact mucin. Therefore these peptides provide a model for investigating the conformational requirements for binding to the antibodies.

Structural studies on the twenty amino acid peptide P(1-20) based on the tandem repeat of the MUC1 protein core utilising high field n.m.r. spectroscopy have indicated that in dimethyl sulphoxide a type-I β—turn exists about the residues Pro^1-Asp-Thr-Arg^4. This element of secondary structure overlaps the epitopes of all of the anti-MUC1 antibodies studied to date which recognise determinants within the protein core of the mucin. Simulated annealing calculations on this region of the peptide have identified that the lowest energy conformation consists of a type-I β—turn Pro^1-Asp-Thr-Arg^4 which is stabilised by an ionic interaction between the side chains of the Asp and Arg. This low energy conformation fits extremely well with the solution n.m.r. data observed for P(1-20). Similar elements of structure have previously been identified as being important in antibody recognition of antigens. Therefore it appears

likely that this structural motif may be present on the surface of the intact MUC1 where it has a role in the immune recognition of the molecules by anti mucin antibodies.

A number of studies have now been undertaken which attempt to characterise the fine specificity of several anti-MUC1 antibodies by determining their reactivity with immobilised synthetic peptides. It has been shown that antibodies which define similar epitopes in the protein core of the mucin have very different profiles of reactivity with synthetic peptide analogues of the tandem repeat sequence. In an attempt to define a structural basis for these observations, a number of MUC1-core related peptides in dimethyl sulphoxide were investigated using high field n.m.r. spectroscopy. These studies revealed subtle differences in the conformations of the peptides. While the observed differences may account for the profile of reactivity of the peptides with different anti-MUC1 antibodies, they provide little indication of the precise conformational requirements for binding. Consequently, initial experiments have been undertaken with the aim of characterising the bound conformations of the peptides.

A number of n.m.r. techniques are available which allow investigation of the structure of small antigens in complexes with antibody. The suitability of a particular method depends crucially on the affinity of the antibody for the peptide. Fab fragments of the anti-mucin antibody C595 have been obtained by proteolysis. These fragments have been shown to retain their specificity for both the intact mucin and the core related peptides. Furthermore, fluorescence quenching experiments have demonstrated that the interaction of these Fab fragments with P(1-20) involves at least one antibody tryptophan residue, and that it is unlikely that ionic interactions contribute significantly to the binding energy. The affinity of the fragments for the peptide P(1-20) is such that the most suitable technique for investigating its bound conformation is 2D-TRNOE spectroscopy. However, as a result of the amount of Fab required by this method, enzymatic proteolysis is an unsuitable means of production of the antibody fragments. Consequently it is planned to produce recombinant Fab fragments of C595. DNA coding for the variable domains of this antibody has been amplified and sequenced.

The obtained sequence confirms that a number of tryptophan residues are contained both within and adjacent to the CDRs. Interestingly, however, CDR3 of the heavy chain contains seven charged residues, six of which are acidic. This CDR is thought to be the most important for antibody recognition and binding. Since the defined epitope for C595 contains an arginine residue it is surprising that the affinity of the antibody is apparently insensitive to changes in pH.

Studies of the MUC1 related peptides in complex with antibody fragments should provide a clearer understanding of the precise conformational requirements for antibody binding. Work is currently underway in this laboratory to produce recombinant Fab fragments of C595. Initial results suggest that these recombinant Fabs retain their specificity for the mucin. This provides further conformation that the DNA sequence obtained for the antibody was indeed correct. This expression system should allow the production of large quantities of highly homogeneous Fab. It will also facilitate the incorporation of isotopic labels into the protein which would be required for investigation of this system by n.m.r. spectroscopy. With the availability of these Fab fragments it should be possible to determine the conformation of the MUC1 core related peptides in complex with antibody.

Knowledge of the sequence of the variable domains of the antibody enables the calculation models of the binding site. Modelling of the binding site of C595 is being undertaken at present. Such a model in conjunction with data on bound peptide conformations obtained from n.m.r. may provide insight into the conformational aspects of antibody recognition and binding.

To further refine this model new peptides are being produced to which sugars are attached in O-linkage to the serine and threonine residues. These peptides will allow investigation of the effects of glycosylation both of the structure of the peptides and its effect on antibody binding. With this information it may be possible to define the epitopes which appear to be expressed only on the tumour derived material. Such studies are of fundamental interest to investigations both of the nature of MUC1 molecules and the processes of molecular recognition involved in antibody antigen

interactions. In addition they may allow the production of new anti-MUC1 antibodies with improved specificity for the mucin which could be exploited in a clinical context.

BIBLIOGRAPHY

Crumpton M.J. & Williamson J.M. (1965), *Biochem. J.*, **94**, 545-556.

Cygler M. & Andeson W.F. (1988), *Acta Crystallogr., Sect. A.*, **44**, 38-45.

Davies D.R., Padlan E.A. & Sherriff S. (1990), *Ann. Rev. Biochem.*, **59**, 439-473.

Denny P.A. & Denny P.C. (1982), *Carbohydrate Res.*, **110**, 305-314.

Diem M., Oboodi M.R. & Alva C. (1984), *Biopolymers*, **23**, 1917-1930.

Dyson H.J., Satterthwait A.C., Lerner R.A. & Wright P.E. (1990), *Biochemistry*, **29**, 7828-7837.

Dyson H.J., Cross K.J., Houghten R.A., Wilson I.A., Wright P.E. & Lerner R.A. (1985), *Nature (London)*, **318**, 480-483.

Dyson H.J., Cross K.J., Ostresh J., Houghten R.A., Wilson I.A., Wright P.E. & Lerner R.A. (1986), In *Synthetic Peptides as Antigens. CIBA Foundation Symposium* 119, 58-75, John Wiley & Sons, Chichester.

Dyson H.J., Rance M., Houghten R.A., Lerner R.A. & Wright P.E. (1988a), *J. Mol. Biol.*, **201**, 161-200.

Dyson H.J., Rance M., Houghten R.A., Wright P.E. & Lerner R.A. (1988b), *J. Mol. Biol.*, **201**, 201-217.

Dyson H.J., Norrby E., Hoey K., Parks D.E., Lerner R.A. & Wright P.E. (1992), *Biochemistry*, **31**, 1458-1463.

Eckhardt A.E., Timpte C.S., Abernethy J.L, Toumadje A. Johnson W.C. & Hill R.L. (1987), *J. Biol. Chem.*, **262**, 11339-11344.

Edelman G.M. (1970), *Biochemistry*, **9**, 3197-3205.

Ehrlich P. (1900), *Proc. R. Soc. London Ser. B.*, **66**, 424-448.

Ellis I.O., Robins R.A., Elston C.W., Blamey R.W., Ferry B. & Baldwin R.W. (1984), *Histopathology*, **8**, 561-566.

Epand R.M. & Scheraga H.A. (1968), *Biochemistry*, **7**, 2864-2872.

Fieser T.M., Tainer J.A., Geysen H.M., Houghten R.A. & Lerner R.A. (1987), *Proc. Natl. Acad. Sci (U.S.A.)*, **84**, 8568-8572.

Gendler S., Taylor-Papadimitriou J., Duhig T., Rothbard J. & Burchell J. (1988), *J. Biol. Chem.*, **263**, 12820-12823.

Gendler S., Lancaster C.A., Taylor-Papadimitriou J., Duhig T., Peat N., Burchell J., Pemberton L., Lalani E. & Wilson D. (1990), *J. Biol. Chem.*, **265**, 15286-15293.

Geysen H.M., Meloen R.H. & Barteling S.J. (1984), *Proc. Natl. Acad. Sci. (U.S.A.)*, **81**, 3998-4002.

Geysen H.M., Rodda S.J., Mason T.J., Tribbick G. & Schools P.G. (1987), *J. Immunol. Meth.*, **102**, 259-274.

Gierasch L.M., Deber C.M., Madison V., Niu C.H & Blout E.R. (1981), *Biochemistry*, **20**, 4730-4738.

Girling A., Batkova J, Burchell J. Gendler S., Gillet C. & Taylor-Papadimitriou J. (1989), *Int. J. Cancer*, **43**, 1072-1076.

Gum J.R., Hicks J.W., Swallow D.M., Lagace R.L., Byrd J.C., Lampert D.T.A., Siddiki B & Kim Y.S. (1990), *Biochem. Biophys. Res. Comm.*, **171**, 407-415.

Handoll H.H.G. (1985), D. Phil Thesis, University of Oxford.

Hanisch F.G., Uhlenbruck G., Peter-Katalinic J., Egge H., Dabrowski J. & Dabrowski U. (1989), *J. Biol. Chem.*, **264**, 872-883.

Hermans J. & Puett D. (1971), *Biopolymers*, **10**, 895-914.

Herp A., Wu A.M. & Moschera J. (1979), *Mol. Cell. Biochem.*, **23**, 27-44.

Heyderman E., Steele K. & Omerod M.G. (1979), *J. Clin. Path.* **32**, 35-39.

Hilkens J. & Buijs F. (1988), *J. Biol. Chem.*, **263**, 4215-4222.

Hopp T.P. & Woods K.R. (1981), *Proc. Natl. Acad. Sci. (U.S.A.)*, **78**, 3824-3828.

Hull S., Bright A., Carraway K., Abe M., Hayes D. & Kufe D.W. (1989), *Cancer Comm.*, **1**, 261-267.

Huston J.S., Levinson D., Mudgetthunter M., Tai M.S., Novotny J., Margolies M.N., Ridge R.J., Bruccoleri R.E., Haber E., Crea R & Oppermann H. (1988), *Proc. Natl. Acad. Sci. (U.S.A.)*, **85**, 5879-5883.

Ish-Horowicz D. & Burke J.F. (1981), *Nucleic Acids Res.*, **9**, 2898-2998.

Jameson B.A. & Wolf H. (1988), *CABIOS*, **4**, 181-186.

Jeener J., Meier B.H., Bachmann P. & Ernst R.R. (1983), *J. Magn. Reson.*, **53**, 521-528.

Johnson W.C. (1990), *Proteins Struct. Funct. Genetics*, **7**, 205-214.

Karplus P.A. & Schultz G.E. (1985), *Naturwissenschaften*, **72**, 212-213.

Kato K., Matsunaga C., Nishimura Y., Waelchli M., Kainosho M. & Arata Y. (1989), *J. Biochem.*, **105**, 867-869.

Kauzmann W. (1959), *Adv. Protein Chem.*, **14**, 1-63.

Kenemans P., Bast R.C., Yemeda C.A., Price M.R. & Hilgers J. (1988), *Cancer Rev*, **11-12**, 119-144.

Kohler G. & Milstein C. (1975), *Nature (London)*, **256**, 495-497.

Krimm S & Bandekhar J. (1986), *Adv. Prot. Chem.*, **38**, 181-364.

Kyte J. & Doolittle R.F. (1982), *J. Mol. Biol.*,**157**, 105-132.

Laemmli U.K. (1970), *Nature (London)*, **227**, 680-685.

Lan M.S., Bast R.C., Colnaghi M.I., Knapp R.C. Colcher D., Schlom J & Metzgar R.S. (1987), *Int. J. Cancer*, **39**, 68-72.

Levy R., Assulin O., Scherf T., Levitt M. & Anglister J. (1989), *Biochemistry*, **28**, 7168-7175.

Ligtenberg M.J.L., Vos H.L., Genissen A.M.C. & Hilkens J. (1990), *J. Biol. Chem.*, **189**, 463-473.

Manning M.C., Illangasekane M. & Woody R.W. (1977), *Biophys. Chem.*, **31**, 77-86.

Manning M.C. & Woody R.W. (1989), *Biochemistry*, **28**, 8609-8613.

Marshall R.D. (1972), *Ann. Rev. Biochem.*, **41**, 673-702.

Needleman S.B. & Wunsch C.D. (1970), *J. Mol. Biol.*, **48**, 443-453.

Novotny J., Bruccoleri R.E. & Saul F.A. (1989), *Biochemistry*, **28**, 4735-4749.

O'Sullivan C. (1990), Ph.D. Thesis, University of Nottingham.

Oakley B.R., Kirsch D.R. & Morris N.R. (1980), *Anal. Biochem.*, **105**, 361-363.

Oboodi M.R., Alva V.C. & Diem M. (1984), *J. Phys. Chem.*, **88**, 501-505.

Padlan E.A., Davies D.R., Rudikoff S. & Potter M. (1976), *Immunochemistry*, **13**, 945-949.

Padlan E.A., Silverton E.W. Sheriff S., Cohen G.H., Smith-Gill S.J. & Davies D.R (1989), *Proc. Natl. Acad. Sci (U.S.A.)*, **86**, 5938-5942

Pardi A., Billeter M. & Wuthrich K. (1984), *J. Mol. Biol.*, **180**, 741-751.

Paterson Y., Englander S.W. & Roder H. (1990), *Science*, **249**, 755-759.

Pearson W.R. & Lipman D.J. (1988), *Proc. Natl. Acad. Sci. (U.S.A.)*, **85**, 2444-2448.

Pellequer J.L., Westhof E. & Van Regenmortel M.H.V. (1991), *Methods in Enzymology*, **203**, 176-201.

Perey L., Hayes D.F., Miamonis P., Miyako A., O'Hara C. & Kufe D.W. (1992), *Cancer Res.*, **52**, 2563-2568.

Pople J.A., Schnieder W.G. & Bernstein H.J. (1959), *High Resolution N.M.R.*, McGraw-Hill, New York.

Porter R.R. (1959), *Biochem. J.*, **73**, 119-126.

Price M.R. (1988), *Eur. J. Clin. Oncol.*, **24**, 1799-1804.

Price M.R., Hudecz F., O'Sullivan C., Baldwin R.W., Edwards P. & Tendler S.J.B. (1990a), *Mol. Immunol.*, **62**, 795-802.

Price M.R., Pugh J., Hudecz F., Griffiths W., Jacobs E., Symonds I.M., Clarke A.J., Chan W.C. & Baldwin R.W. (1990b), *Br. J. Cancer*, **61**, 681-686.

Price M.R., Sekowski M. & Tendler S.J.B. (1991), *J. Immunol. Meth.*, **139**, 83-90.

Ramanadhan M., Sieker L.C. & Jensen L.H.(1990), *Acta Crystallogr. Sect. B.*, **46**, 63-69.

Riechmann L., Foote J. & Winter G. (1988), *J. Mol. Biol.*, **203**, 825-828.

Roberts C.J., Sekowski M., Davies M.C. Jackson D.E., Price M.R. & Tendler S.J.B. (1992), *Biochem J.*, **283**, 181-185.

Roberts G.M., Lee O., Calienni J. & Diem M. (1988), *J. Am. Chem. Soc.*, **110**, 1749-1752.

Roberts G.P. (1976), *Arch. Biochem. Biophys.*, **173**, 528-537.

Rose M.C., Voter W.A., Brown C.F. & Kaufmann B. (1984), *Biochem J.*, **222**, 371-377.

Sambrook J., Fritsch E.F. & Maniatis T., (1989), *Molecular Cloning*, Cold Spring Harbour Laboratory Press.

Sanger F., Nicklen S. & Coulson R. (1977), *Proc. Natl. Acad. Sci. (U.S.A.)*, **74**, 5463-5467.

Scanlon M.J., Morley S.D., Jackson D.E., Price M.R. & Tendler S.J.B. (1992), *Biochem J.*, **284**, 137-144.

Scherf T., Hiller R., Naider F., Levitt M. & Anglister J. (1992), *Biochemistry*, **31**, 6884-6897.

Sheriff S., Silverton E.W., Padlan E.A., Cohen G.H., Smith-Gill S.J., Finzel B.C. & Davies D.R. (1987), *Proc. Natl. Acad. Sci (U.S.A.)*, **84**, 8075-8079.

Shimizu M. & Yamauchi K. (1982), *J. Biochem.*, **91**, 515-524.

Siddiqui J., Abe M., Hayes D., Shani E., Yunis E. & Kufe D.W. (1988), *Proc. Natl. Acad. Sci. (U.S.A.)*, **35**, 2320-2323.

Slayter H.S. & Codington J.F. (1973), *J. Biol. Chem.*, **248**, 3405-3410.

Smith-Gill S.J., Wilson A.C., Potter M., Prager E.M., Feldmann R.J. & Mainhart C.R. (1982), *J. Immunol.*, **128**, 314-322.

Stacker S.A., Sacks N.P.M., Golder J., Tjandra J.J., Thompson C.H., Smithyman A. & McKenzie I.F.C. (1988), *Br. J. Cancer*, **57**, 298-303.

Swallow D.M., Griffiths B., Bramwell M.E., Wiseman G. & Burchell J. (1986), *Dis. Markers*, **4**, 247-254.

Symonds I.M., Price M.R., Pimm M.V., Perkins A.C., Wastie M.L., Baldwin R.W. & Symonds E.M. (1992), *6 Hamburger Symposium uber Tumormarker* (R. Klapdor ed.), W. Zuckchwerdt Verlag, Munchen, Germany.

Takahashi N., Takahashi Y. & Putnam F.W. (1984), *Proc. Natl. Acad. Sci. (U.S.A.)*, **81**, 2021-2025.

Taylor-Papadimitriou J., Peterson J., Arklic J., Burchell J., Ceriani R.L. & Bodmer W.F. (1981), *Int. J. Cancer*, **28**, 23-29.

Taylor-Papadimitriou J. (1991), *Int. J. Cancer*, **49**, 1-5.

Tendler S.J.B. (1990), *Biochem J.*, **266**, 733-737.

Timpte C.S., Eckhardt A.E., Abernethy J.L. & Hill R.L. (1988), *J. Biol. Chem.*, **263**, 1081-1088.

Tsang P., Fieser T.M., Ostresh J.M., Houghten R.A., Lerner R.A. & Wright P.E. (1990), In : *Frontiers of N.M.R. in Molecular Biology* (Live D., Armitage I. & Patel D. eds.), 63-73, Alan R. Liss New York.

Tsang P., Rance M. & Wright P.E. (1991), *Methods in Enzymology*, **203**, 241-261.

Tsang P., Rance M., Fieser T.M. Ostresh J.M., Houghten R.A., Lerner R.A. & Wright P.E. (1992), *Biochemistry*, **31**, 3862-3871.

Tsang P., Rance M., Fieser T.M., Ostresh J.M., Houghten R.A., Lerner R.A. & Wright P.E. (1992), *Biochemistry*, **31**, 3862-3871.

Tulip W.R., Varghese J.N., Webster R.G., Air G.M., Laver W.G. & Colman P.M. (1990), *Cold Spring Harbour Symp. Quant. Biol.*, **54**, 257-263.

Waltho J.P., Feher V.A., Lerner R.A. & Wright P.E. (1989), *FEBS Lett.*, **250**, 400-404.

Weiss M.A., Eliason J.L. & States D.J. (1984), *Proc. Natl. Acad. Sci (U.S.A.)*, **81**, 6019-6023.

Williamson M.P., Hall M.J. & Handa B.K. (1986), *Eur. J. Biochem.*, **158**, 527-536.

Wilson I.A., Rini J.M., Fremont D.H., Fieser G.G. & Stura E.A. (1991), *Methods in Enzymology*, **203**, 153-176.

Wilson I.B.H., Gavel Y. & von Heijne G. (1991), *Biochem. J.*, **275**, 529-534.

Woody R.W. (1984), In *Peptides, Polypeptides and Proteins.* (Blout E.R., Bovey F.A., Goodman M. & Lotan N. eds.), John Wiley & Sons, New York.

Wright P.E., Dyson H.J. & Lerner R.A. (1988), *Biochemistry*, **27**, 7167-7175.

Wright P.E., Dyson H.J., Lerner R.A., Riechmann L. & Tsang P. (1990), *Biochem. Pharmacol.*, **40**, 83-88.

Wu T.T. & Kabat E.A. (1970), *J. Exp. Med.*, **132**, 211-250.

Wüthrich K. (1986), *N.M.R. of Proteins and Nucleic Acids.* John Wiley & Sons, New York.

Xing P.X., Reynolds K., Pietersz G.A. & McKenzie I.F.C. (1991), *Immunology*, **72**, 304-311.

Xing P.X., Prenzoska J. & McKenzie I.F.C. (1992), *Mol. Immunol.*, **29**, 641-650.

Xing P.X., Reynolds K., Tjandra J.J., Tang X.L. & McKenzie I.F.C. (1990), *Cancer Res.*, **50**, 89-96.

Zilber B., Scherf T., Levitt M. & Anglister J (1990), *Biochemistry*, **29**, 10032-10041.

Zimmerman S.S. & Scheraga H.A. (1977), *Proc. Natl. Acad. Sci. (U.S.A.)*, **74**, 4126-4129.

Zotter S., Hageman P.C., Lossnitzer A., Mooi W.J. & Hilgers J. (1988), *Cancer Rev.*, **11-12**, 55-101.

Allen A. (1983), *TIBS* , **8**, 169-173.

Amit A.G., Mariuzza R.A., Phillips S.E.V. & Poljak R.J. (1986), *Science*, **233**, 747-753.

Amzel L.M., Poljak R.J., Saul F., Varga J.M. & Richards F.F. (1974), *Proc Natl. Acad Sci. (U.S.A.)*, **71**, 1427-1430.

Amzel L.M. & Poljak R.J. (1979), *Ann. Rev. Biochem.*, **48**, 961-997.

Anglister J., Jacob C., Assulin O., Ast G., Pinker R. & Arnon R. (1988), *Biochemistry*, **27**, 717-724.

Anglister J., Levy R. & Scherf T. (1989), *Biochemistry*, **28**, 3360-3365.

Anglister J. & Zilber B. (1990), *Biochemistry*, **29**, 921-928.

Atassi M.Z., (1975), *Immunochemistry*, **12**, 423-438.

Aue W.P., Bartholdi E. & Ernst R.R. (1976), *J. Chem. Phys.*, **64**, 2229-2246.

Aumelas A., Audoussetpuech M.P., Heitz A., Bataille D. & Martinez J. (1989), *Int. J. Pept. Protein Res.*, **34**, 268-276.

Balaram P., Bothner-By A.A. & Breslow E. (1972), *J. Am. Chem. Soc.*, **94**, 4017-4019.

Bandekhar J., Evans D.J., Krimm S., Leach S.J. & Lee S. (1982), *Int. J. Pept. Protein Res.*, **19**, 187-205.

Barlow D.J., Edwards M.S. & Thornton J.M. (1986), *Nature (London)*, **322**, 747-748.

Bax A. & Davis D.G. (1985), *J. Magn. Reson.*, **63**, 207-213.

Better M., Chang C.P., Robinson R.R. & Horwitz A.H. (1988), *Science*, **240**, 1041-1043.

Bevan A.W., Roberts G.C.K., Feeney J. & Kuyper L. (1985), *Eur. Biophys. J.*, **11**, 211-218.

Bodenhausen G., Kogler G. & Ernst R.R. (1984), *J. Magn. Reson.*, **58**, 370-388.

Bothner-By A.A., Stephens R.L., Lee J., Warren C.D. & Jeanloz R.W. (1984), *J. Am Chem. Soc.*, **106**, 811-813.

Bramwell M. E., Wiseman G. & Shotton D.M. (1986), *J. Cell Sci.*, **86**, 249-261.

Bramwell M.E., Bhavanandan V.P., Wiseman G. & Harris H. (1983), *Br J. Cancer*, **48**, 177-183.

Brant D.A., Miller W.G. & Flory P.J. (1967), *J. Mol. Biol.*, **23**, 47-65.

Braunschwieler L.R. & Ernst R.R. (1983), *J. Magn. Reson.*, **53**, 521-528.

Briand J.P., Andrews S.P. Jr., Cahill E., Conway N.A. & Young J.D. (1981), *J. Biol. Chem.*, **256**, 12205-12207.

Briggs S., Price M.R. & Tendler S.J.B. (1991), *Immunol.*, **73**, 505-507.

Briggs S., Price M.R. & Tendler S.J.B. (1993), *Eur. J. Cancer*, **29A**, 230-237.

Brown J.E. & Klee W.A. (1971), *Biochemistry*, **10**, 470-476.

Burchell J., Gendler S., Taylor-Papadimitriou J., Girling A., Lewis A. & Millis R. (1987), *Cancer Res.*, **47**, 5476-5482.

Burchell J., Taylor-Papadimitriou J., Boshell M., Gendler S. & Duhig T. (1989), *Int. J. Cancer*, **44**, 691-696.

Carlstedt I., Lindgren H., Sheehan J.K., Ulmsten U. & Wingerup L. (1983), *Biochem. J.*, **211**, 13-22.

Ceriani R., Peterson J., Lee J., Moncada R. & Blank E. (1983), *Somat. Cell. Genet.*, **9**, 415-427.

Cheetham J.C., Redfield C., Griest R.E., Raliegh D.P., Dobson C.M. & Rees A.M. (1991), *Methods in Enzymology*, **203**, 202-228.

Chou P.Y. & Fasman G.D. (1977), *J. Mol. Biol.*, **115**, 135-175.

Chou P.Y. & Fasman G.D. (1978), *Adv. Enzym.*, **47**, 45-147.

Clore G.M. & Gronenbaum A.M. (1991), *Science*, **252**, 1390-1399.

Colman P.M., Laver W.G., Varghese J.N., Baker A.T., Tulloch P.A., Air G.M. & Webster R.G. (1987), *Nature (London)*, **326**, 358-363.