SECONDARY STRUCTURES OF HEXA-L-ALANINE HOST PEPTIDES SUBSTITUTED WITH L-LEUCINE AND L-ISOLEUCINE

J.P.W.G. BURKE

A dissertation submitted in fulfillment of the requirements for the degree of Master of Science at the University of the Witwatersrand

Johannesburg, January 1991

DECLARATION

I declare that this dissertation is entirely my own work and has not been presented for any other degree at any other university.

J.P.W.G. Burke

Jonathan Burke

ABSTRACT

TFA.H.leu(ala)$_5$.OH, TFA.H.(ala)$_5$leu.OH, TFA.H.ile(ala)$_5$.OH and TFA.H.(ala)$_5$ile.OH (TFA=trifluoroacetic acid counter ion) were purified and their structures were analysed with infrared, NMR and circular dichroism spectroscopy.

TFA.H.leu(ala)$_5$.OH, TFA.H.(ala)$_5$leu.OH, TFA.H.ile(ala)$_5$.OH and TFA.H.(ala)$_5$ile.OH formed extended ß-structure in the solid phase. In aqueous solution there was evidence from proton NMR spectra that definite structure was present. In alcoholic solution TFA.H.leu(ala)$_5$.OH formed structures with chiroptical properties similar to those seen in hydrophobic pockets. TFA.H.(ala)$_5$leu.OH and TFA.H.(ala)$_5$ile.OH formed ß-structures. TFA.(ala)$_5$leu.OH formed antiparallel ß-sheets approximated to be 4 to 20 strands in width, which decreased in size as the alcohol became less polar. TFA.H.ile(ala)$_5$.OH formed "single ß-chains". TFA.ile(ala)$_5$.OH formed "statistical coils" in alcoholic solution. In TFE/ethanol solutions TFA.H.(ala)$_5$ile.OH formed the most stable ß-structure followed by TFA.H.(ala)$_5$leu.OH and TFA.H.ile(ala)$_5$.OH.

ACKNOWLEDGEMENTS

I wish to thank:
Dr A. Nel and Dr R.L. Jones for their patience and friendship, the members of the Department of Biochemistry, University of the Witwatersrand, Dr G. Osthoff, Dr F. Carlsson and Dr R. Clarke of the CSIR at Pretoria for their time and the use of equipment, my family for their support and my mother for reading the manuscript.

CONTENTS

ABBREVIATIONS

The IUPAC-IUB three letter codes for amino acids are used

C-terminal	carboxy-terminal
CD	circular dichroism
-C=O	carbonyl groups
DCC	dicyclohexylcarbodiimide
DCM	dichloromethane
DMF	dimethylformamide
HOBt	1-hydroxybenztriazole
HPLC	high performance liquid chromatography
IR	infrared
-NH or N-	amino group
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
SPPS	solid phase peptide synthesis
t-Boc	tertiary-butyloxycarbonyl
TFA	2,2,2 trifluoracetic acid
TFE	trifluoroethanol
TLC	thin layer chromatography

LIST OF TABLES

Page numbers are in brackets

LIST OF FIGURES

Page numbers are in brackets

Appendix figures:

1. INTRODUCTION

1.1 Historical background

Early in the twentieth century Emil Fischer placed an upper limit of 30 to 40 residues on the size of any covalent polymer.[1] The result was an acceptance of Otswald's colloid theory of proteins to explain their obvious large sizes. In a paper read in 1925 to a meeting in Zürich, Staudinger was shouted down for claiming to have produced a synthetic polymer of 1'00 residues. The proponents of the covalent theory became few and far between. The few, among them Cohn and Edsall, opposed the current stream of thought because of results from osmotic molecular weight measurements. Cohn predicted that Svedberg, then the most ardent colloid proponent, would find discreet bands of protein in his newly constructed analytical centrifuge. Svedberg on the other hand predicted a heterogeneous distribution of particles in line with the colloid theory. The results proved the covalent theory to be correct and was recorded in history as a significant triumph of science.

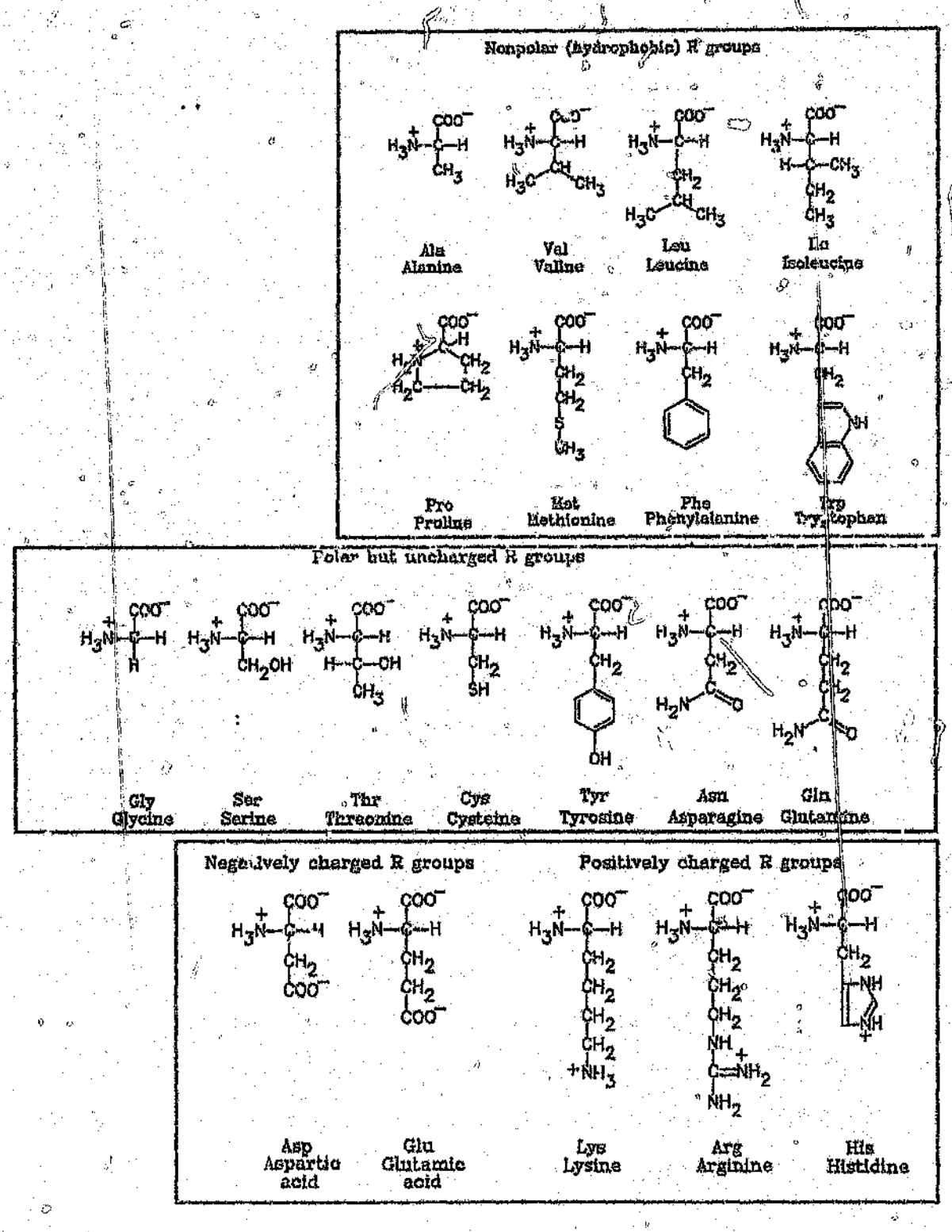
Next was the elucidation of the nature of protein molecules. The separation and identification of the 20 amino acids was completed after 1940 with the development of paper and ion exchange chromatography by Martin and Synge, and Stein and Moore respectively. A decade later Sanger elucidated the first amino acid sequences using the α and β chains of insulin. The three- dimensional mapping of proteins was the

next logical step. Perutz and Kendrew started work on the structures of haemoglobin and myoglobin in 1936. The break was made when Perutz inserted atoms of mercury into haemoglogin to aid his analysis of X-ray diffraction patterns. Despite this, the three-dimensional modelling of haemoglobin required greater computer power which appeared only in 1952.

Pauling and Corey in California in the meantime had measured the dimensions of the peptide bond and predicted that α-helix and ß-sheet structures were stable enough to exist in proteins. Donohue then showed the α-helix to be the most stable and their predictions were proved correct with the appearance of the structures of haemoglobin, myoglobin and silk. The stage was set for a logical and clear understanding of proteins based upon the assumption that underlying their diversity were smaller units of secondary structure.

Anfinsen showed that a protein's unique tertiary structure is dictated by its amino acid sequence.[2] A theory to explain the mechanism by which proteins fold spontaneously from the information stored in the amino acid sequence was the next undertaking. The mechanism will be part of a larger theory that explains the structure and actions of molecules in general. The macromolecular basis of life places such a theory in a fundamental position. Protein structure studies have attracted a variety of people and skills. One reason

Figure 1 The twenty naturally occuring amino acids, classified according to hydrophobicity.[6]

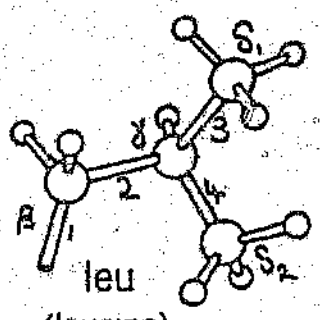


for this may be that the techniques used for structural elucidation, nuclear magnetic resonance, circular dichroism and X-ray crystallography, are at their limits in protein work.

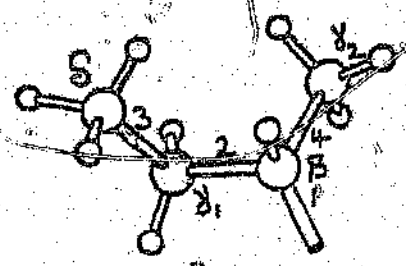
1.2 The amino acids

The building blocks of proteins are the amino acids. Figure 1 shows the twenty that occur naturally. They all have a common skeleton centered at the Cα; to this is attached a carbonyl group, an amino group, a proton and a side chain. The side chain is the only variant. Every amino acid except glycine is asymmetric and is designated either D or L; based upon glyceraldehyde as the reference compound[3]. Only the L-form is found in proteins.[3] Threonine and isoleucine show stereoisomerism about the C_β and this is labelled according to the Ingersoll, Kahn and Meyerhof method. The chirality imposes a strict sense of handedness upon the structures that are built up from the amino acids. The atoms of the side chains are named from the Cα as per the convention of Edsall[3] (figure 2). Glycine is unique in that it has no chirality and can form structures with a left- and right-handed sense. The hydrogen side chain also allows it to be used where chains meet and space is limited or where flexibility is required. Examples can be found where α-helices pack together, in turn conformations and in hinge regions between protein domains.[4,5] At the opposite end of

Figure 2 The Edsall convention for the naming of amino acid side chain atoms and bonds.[5] A heavy atom is any atom not hydrogen. Starting at C_α, the first heavy atom is labelled Beta, the second Gamma, the third Delta and so on. The naming of bonds between heavy atoms follows in numerical order. Branches are numbered according to the size of the branch; the branch with the most atoms being labelled 1. Where branches are similar, a clockwise rule is used. Looking down the previous bond, the first branch reached in a clockwise rotation from 12 O'clock is labbeled 1. L-leucine and L-isoleucine are labelled as examples.



leu
(leucine)



ilu
(isoleucine)

the scale proline is the most stereoconformed amino acid. In proteins the phi angle is stuck at -70° and the amino terminal is unable to hydrogen bond.[4] Proline is found where the peptide chain needs to be turned on itself.[5,6,7] If all the amino acids could be placed in such well defined categories, protein structure would be no more than a matter of common sense. However the rules of geometry or thermodynamics used to pigeon-hole amino acid behaviour have always failed because proteins can accommodate stereo non-conformity or apparent energetic instability to a surprising degree. This is because, through evolution, the twenty amino acids have been selected without the loss of essential functions or the addition of redundancies. Thus each amino acid serves a multiplicity of purposes.

Figure 1 shows the amino acids classified according to their polarities. The peptide bond is the same throughout any protein and it is the side chains that produce a protein's unique properties. Kauzmann[8] proposed the oil drop theory to explain how the hydrophobicities of the side chains organise a protein. The hydrophobic classification of the amino acids is useful for predicting their positions in proteins. Classifications according to steric constraints are also made. Amino acids aggregated in the centre of a protein have to pack against each other efficiently.[4] Inefficient packing may oppose the hydrophobic interaction in aqueous solvents. At physiological pH, the amino acids lysine and arginine

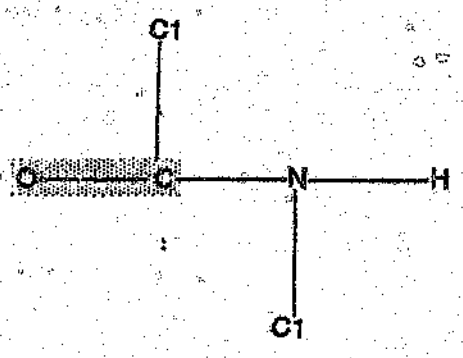
will be positively charged. Aspartic acid and glutamic acid are negatively charged. These amino acids are found at the surface of proteins where they help solvation or they are buried as part of a catalytic site in the protein. The environment of a side chain in a cell or a protein can be very different from the bulk phase of the solvent. These microenvironments are difficult to measure or approximate and this has been a major stumbling block in some protein research.

1.3 Peptide structure

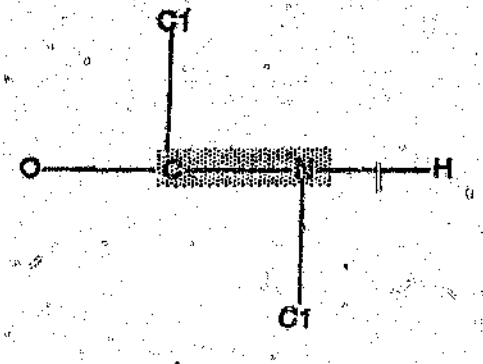
The information for a protein sequence is stored in a linear fashion on DNA. It is translated into protein via RNA on the cell ribosome. The primary structure of any peptide is defined as its amino acid sequence and no protein has yet been found in nature that is branched. Thus primary structures can be found by sequencing proteins or DNA. Conventionally sequences are written starting at the amino terminal.

The local structures found in proteins that are stabilised by the interactions of neighbouring amino acids are defined as secondary. The three dimensional shape and the interactions of amino acids distantly related on the primary structure are defined by the tertiary structure. Finally the association of proteins with other proteins, ligands and prosthetic groups is defined by the quaternary structure. In the four definitions of structure there is an implied

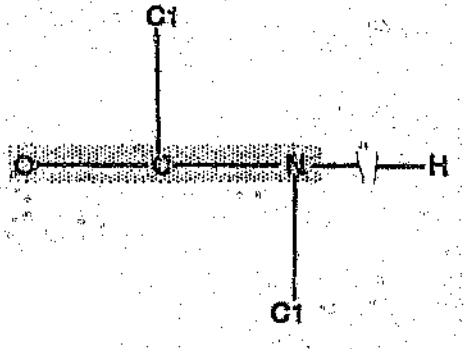
Figure 3 (a), (b) and (c) are diagramatic representations of the peptide bond. C1 represents the C_α atoms. In (a) the nitrogen has an unused electron pair in a 2p orbital. The oxygen 2p orbital is hybridised forming a C-O double bond represented by the shaded area. There is thus rotation about the C-N bond only. In (b) the situation is similar except that the nitrogen electrons are hybridised and oxygen has a lone pair in its 2p orbital. The C-O single bond can rotate. In (c) the resonance structure thought to occur is drawn. Both the oxygen and nitrogen lone pairs are delocalised across the bond (shaded area). There is no rotation about C-O or C-N.[4,6] The most stable arrangement of the C_α atoms is in the Trans configuration.[8]



(a)



(b)



(c)

Figure 4 The dimensions of the planar peptide bond. The C_α atoms are Trans to one another. Bond lengths are in Å units and bond angles in degrees.[4]

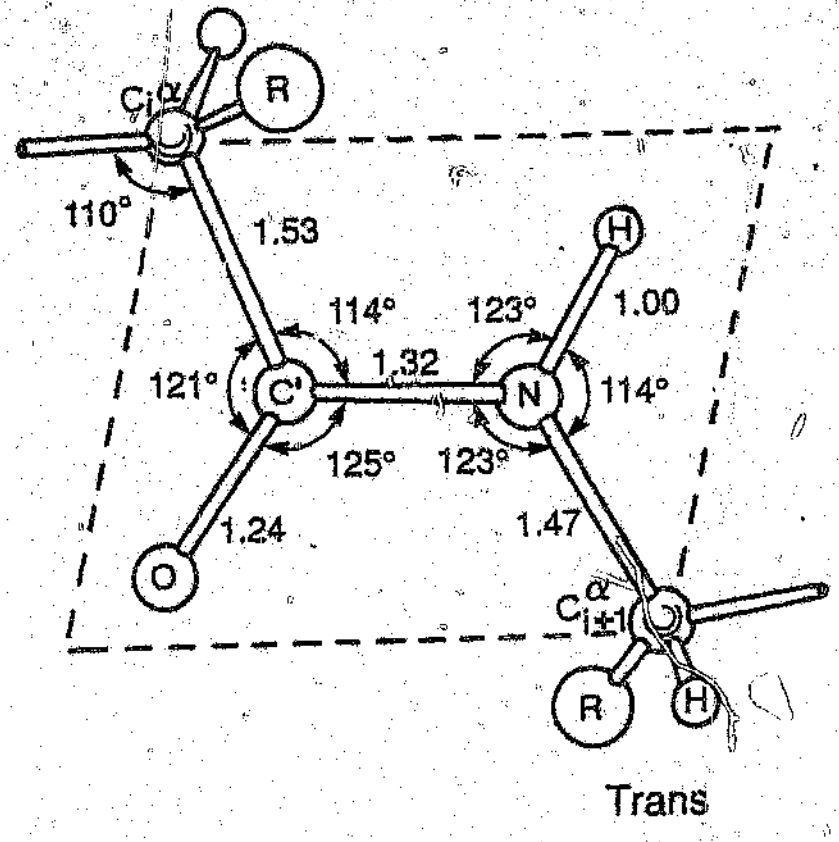
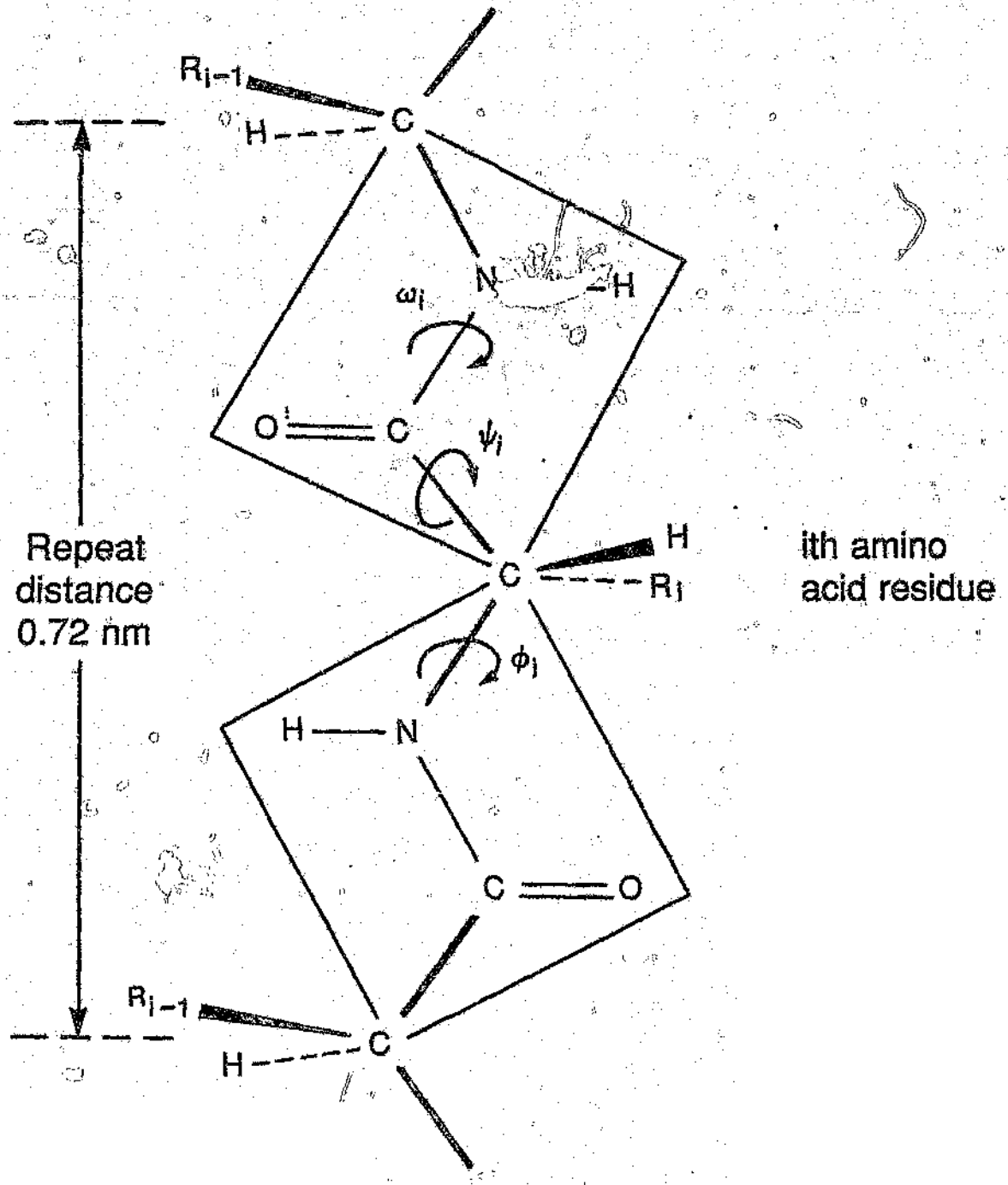


Figure 5 The phi and psi conventions for the rotation of planar peptide bonds around the C_α atom. By the current convention phi and psi are assigned the value of 180°.[4]



hierarchy; the primary and secondary structures deal with sequence and local interactions while the tertiary and quaternary structures deal with the three-dimensional structure of a protein.

1.4 Peptide bond

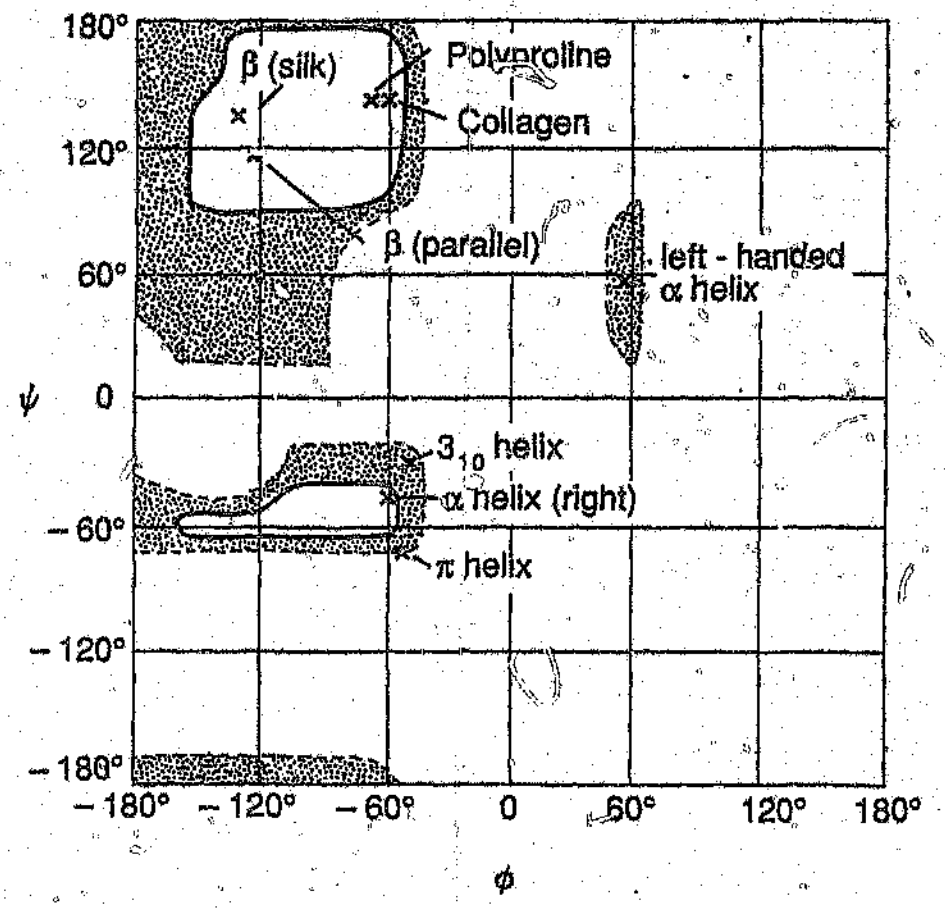
The peptide bond is formed by a condensation reaction between the amino group of one amino acid and the carboxy terminal of another. The peptide bond is planar due to the delocalisation of electrons (figure 3). Pauling and Corey provided the first set of standardised bond lengths and angles for the peptide bond[6,7] (figure 4). Across the peptide bond the Cα of residue n and the Cα of residue n+1 are Trans to one another (figure 4). The Trans configuration has been calculated to be about 8KJ mol^{-1} more stable than the Cis form.[15] There is also a 20Kcal mol^{-1} energy barrier to rotation. Despite this, measurements have shown that proline can form the Cis bond with a probability of 30 percent.[8] Thus thermodynamic stability is not the only requirement for the existence of the Trans form. The peptide bond is normally assumed to be planar, although there are arguments for slight distortions of the plane.[20] Generally a non-planar peptide bond is assumed when such an assumption brings better agreement between energy calculations and the measured evidence. This assumption is supported by nuclear magnetic resonance spin-spin coupling constants with cyclic peptides and high resolution X-ray results.[6] In general it

is assumed that the peptide bond is Trans with respect to the Cαs and is planar. This restriction of movement about the peptide bond only allows two degrees of freedom of rotation about the C_α defined as phi and psi (figure 5). The course of a peptide chain can be mapped out using these two bond angles. Note that in the conventional method of quoting the phi and psi angles, phi is quoted first.

1.4.1 Ramachandran plot

The Ramachandran plot is shown on figure 6.[6,7,10] It plots the rotation of two peptide planes about C_α in a dipeptide. Ramachandran calculated sterically allowed and disallowed values for phi and psi based upon van der Waal's radii. Both residues in the model had a C_β on the side chains. From the examination of X-ray crystallographs it became clear that in proteins the atomic contacts were closer. The effect on the plot of the closer approach of the atoms was to increase the size of the allowed conformational areas. In figure 6 the increased allowed regions are hatched. For a glycine-glycine dipeptide the plot would be symmetrical about phi=0°, psi=0° because it can form the left- and right-handed versions of any allowed conformations with equal ease. The disallowed regions only occur where there is stereochemical hindrance between carbonyl and amino groups. This is the case at -180°,0° and at 0°, -180°. In the former the carbonyls come too close together and in the latter the hydrogens of the -N-H groups interfere. An asymmetric amino acid introduces

Figure 6 The Ramachandran plot of phi versus psi in a dipeptide. The two enclosed, unshaded, irregularly shaped areas occur where there is no steric hindrance. They enclose the values for the α-helix and the β-sheet. The shaded areas are sterically hindered structures found in proteins by X-ray crystallography and left handed helices.



assymetry into the plot. The right-handed conformations are possible while the left-handed conformations are generally not. This is because rotation about phi causes the C_β of the second residue to interfere stereochemically with the carbonyl of the first residue.

Similarly by rotation about psi the amino group of residue n+1 interferes with the side chain of residue n. In figure 6 two strips are disallowed by these factors; they lie between phi values of 80° and 180° and psi values of -80° and -180°. The remaining quadrant in the top left of the plot has three allowed regions. Two large regions represent the β and right-handed helical sections and one smaller left-handed helical section. Generally the distributions of angles found in proteins are clustered around the allowed regions.

To maximise the distance between the carbonyl group and side chain, phi can be fixed at -60°. By doing this and rotating psi the whole range of right-handed helices are possible. Thus the plot can be used to predict likely structures.

1.5 Secondary structure

Kendrew said of myoglobin: "The arrangement seems to be totally lacking in the kind of regularities which one anticipates."[11] This was because prior to the low-resolution X-ray structure of myoglobin, predicted protein structures based on analogies from mathematics and chemistry were simple and regular geometrical shapes. Commenting on

haemoglobin, Perutz even asked himself if the search for ultimate truth could really have revealed "so hideous and visceral-looking an object?"[6] Since then the regularities of protein structure have become better appreciated. The accumulation of knowledge of protein structure has required a whole new vocabulary for its anatomy and taxonomy. The diversity of proteins is bewildering and a structural classification was not possible without a knowledge of secondary structure. Regular secondary structures were the first to be class[illegible]ulin and Corey in 1951 produced the first α-hel[illegible]nd [illegible]set models[13]. This was followed by the structure of the collagen helix of Ramachandran.[74] Other secondary structures are all variations on these themes and include structures such as turns and pockets. The designation of secondary structures in proteins is partly subjective. Each new protein X-ray structure produces differences in the basic patterns of secondary structures. For this reason it has not been possible to provide a standard set of definitions. The problem is that the definitions of bond angles and rules of application have to be flexible enough to account for the deviations seen in proteins while being strict enough to effectively exclude non-conforming structures. The formulation of such rules has been suggested several times and is ongoing.[18]

1.5.1 Helices

The α-helix is the classic secondary structure[7,6,4,10] and is the most stable.[16] It forms part of a series of helices defined by their repeating phi and psi angles. The exclusive use in nature of L-amino acids forces a right-handed chirality in almost all helical structures. Setting phi at about -60° maximises the distance between the amino acid side chains and the carbonyl group in a peptide bond. Rotating psi produces all the right-handed helices as the Ramachandran plot shows (figure 6). Figure 7 depicts the α, 3_{10} and π helices.

1.5.1.1 α-Helix

α-Helices can be up to 35 residues in length. The amino acids L-alanine, L-glutamate, L-leucine and L-methionine tend to favour α-helix formation. L-proline is disruptive to the helix as are C_β branched amino acids. In proteins α-helices deviate from the ideal; axes can bend up to 20° and the hydrogen bonding pattern is sometimes disrupted. L-proline has a restricted phi angle of -70° and forces a turn in the helix. It cannot hydrogen bond through its N-H group and can be accommodated only in the first four amino acids. The α-helix often alters at the carbonyl terminal where it may tighten up to a 3_{10} helix. Another frequent occurrence at a helix terminal is a left-handed conformed glycine.[6]

In proteins α-helices often appear on the surface. In this position one side is in the external medium and one in the hydrophobic interior of the protein and the sequence has an amphilic nature. To achieve this the sequence shows a periodicity with respect to the side chain hydrophobicity of 3 to 4 residues.[6,17]

1.5.1.2 3_{10}, α_{II} and π Helices

The π and 3_{10} helices are drawn on figure 7. The 3_{10} helix is more tightly coiled and less stable than the α-helix. In proteins they are short and appear at the C-terminii of α-helices where they force a tight turn.[6] Often the 3_{10} helix and α-helix conformations mix and form a α_{II} helix.[6]

The π-helix, another structure predicted by Pauling and Corey in 1951[12], has not been observed in many actual proteins[7,6,10]

1.5.1.3 The left-handed and collagen helices

Holding phi at -60° and rotating psi from -30° towards 0° brings the carbonyls of the peptide backbone into contact with the N-H groups of the next peptide bond and is unstable.[4] The next stable structure is a left-handed helix at 120°. Polyproline forms a left-handed helix at this value (figure 8). This structure in a single strand is called an ε-helix[6]. Only a few examples have been found in globular proteins but it forms the basis of the protein collagen. The important property of the left-handed helix is that the carbonyls and N-H groups face outwards and perpendicular to

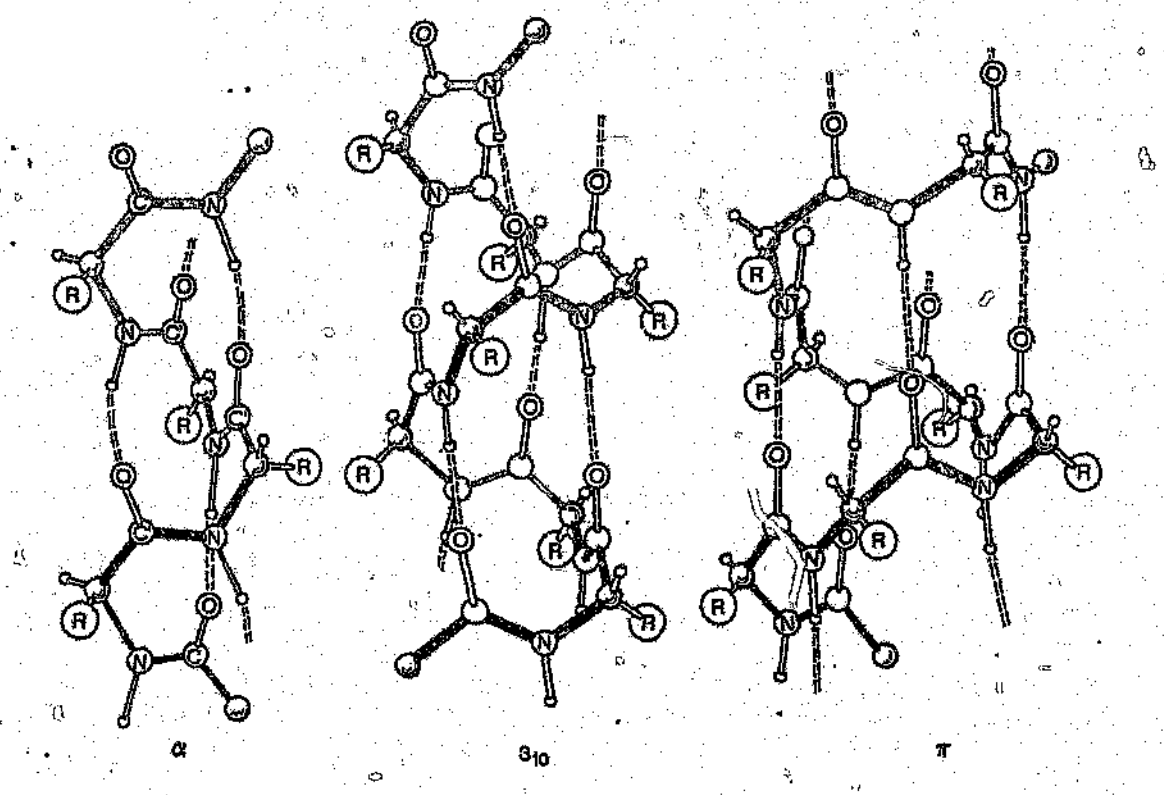


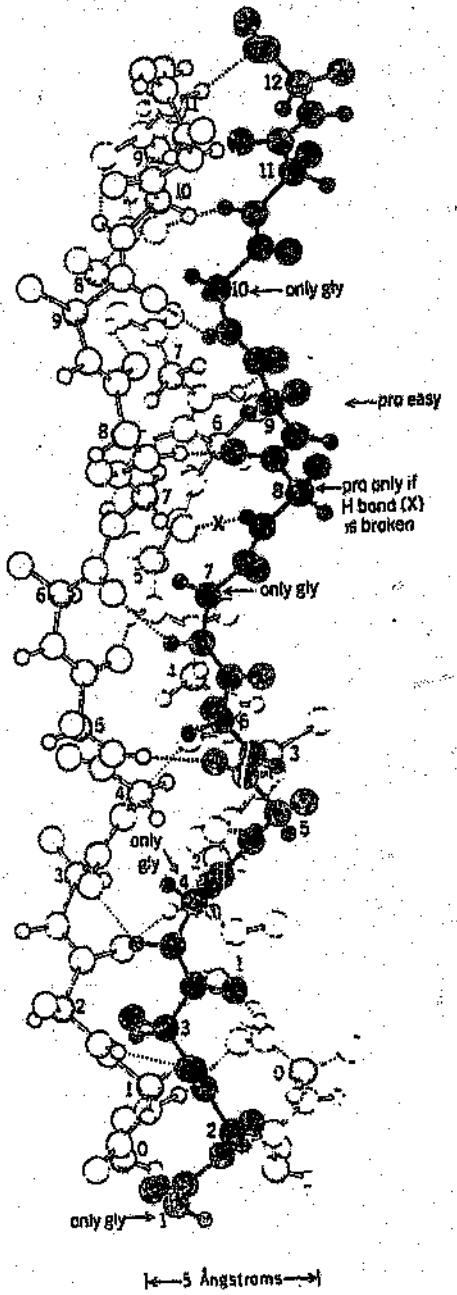
Figure 7 The helices can be described by the number of amino acid residues incorporated into one turn, the distance traversed by a turn or the product of these two values, called the pitch.[7] The α-helix is the most stable of the helical structures. It is characterised by 13 atoms between each hydrogen bond and 3.6 amino acid residues per turn (also called the 3.6_{13} helix). The hydrogen bonds are parallel to the helical axis and linear which maximises their strength.[6] 3_{10} helix hydrogen bonds are joined by 10 atoms and the helix turns on its axis with 3 residues. The hydrogen bonds face outwards and are not aligned with the helical axis.[6] The π helix hydrogen bonds are joined by 16 atoms (4.4_{16} helix). This structure is unstable and rarely seen in proteins.[6]

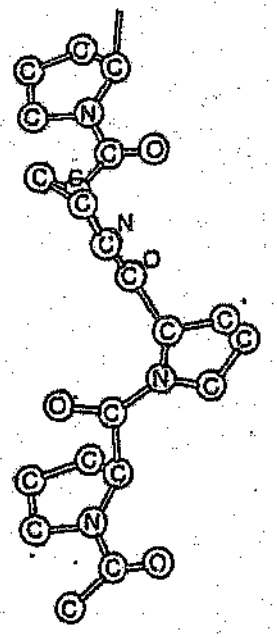
the helix axis. The orientation of the groups allows polyproline chains to form interchain hydrogen bonds. In collagen three chains interact in a right-handed triple helix (figure 8). This forms a strong structure used for a variety of functions including tendons and skin in mammals.

1.5.2 ß-Sheets

ß-Sheet structures were better predicted than the helices because the X-ray patterns showed repeating distances along the fibres of an extended structure. The possibilities were therefore reduced.[6] The basic structure of the ß-sheet was predicted to explain the structure of keratin fibres. In 1951 Pauling and Corey produced their models of antiparallel and parallel ß-pleated sheet[6, 13] (figures 9a and 9b). Side groups on adjacent strands extend to the same side of the sheet and are in close contact. Sheets show a statistical preference for matching amino acids; hydrophobic groups pair, opposite charges on side chains pair and branched C_β residues pair with unbranched C_β. The ß-sheet conformation phi and psi angles can be found in the top left quadrant of the Ramachandran plot (figure 6). The energy minimum for the ß-sheet is shallow and this may explain why a strand of ß-sheet can contain up to 15 residues while α-helices can be 35 residues in length.[6] The antiparallel sheets are more stable than the parallel sheets. Parallel sheets usually contain large numbers of strands and are found in protein cores protected from the aqueous medium of a cell.[6] The

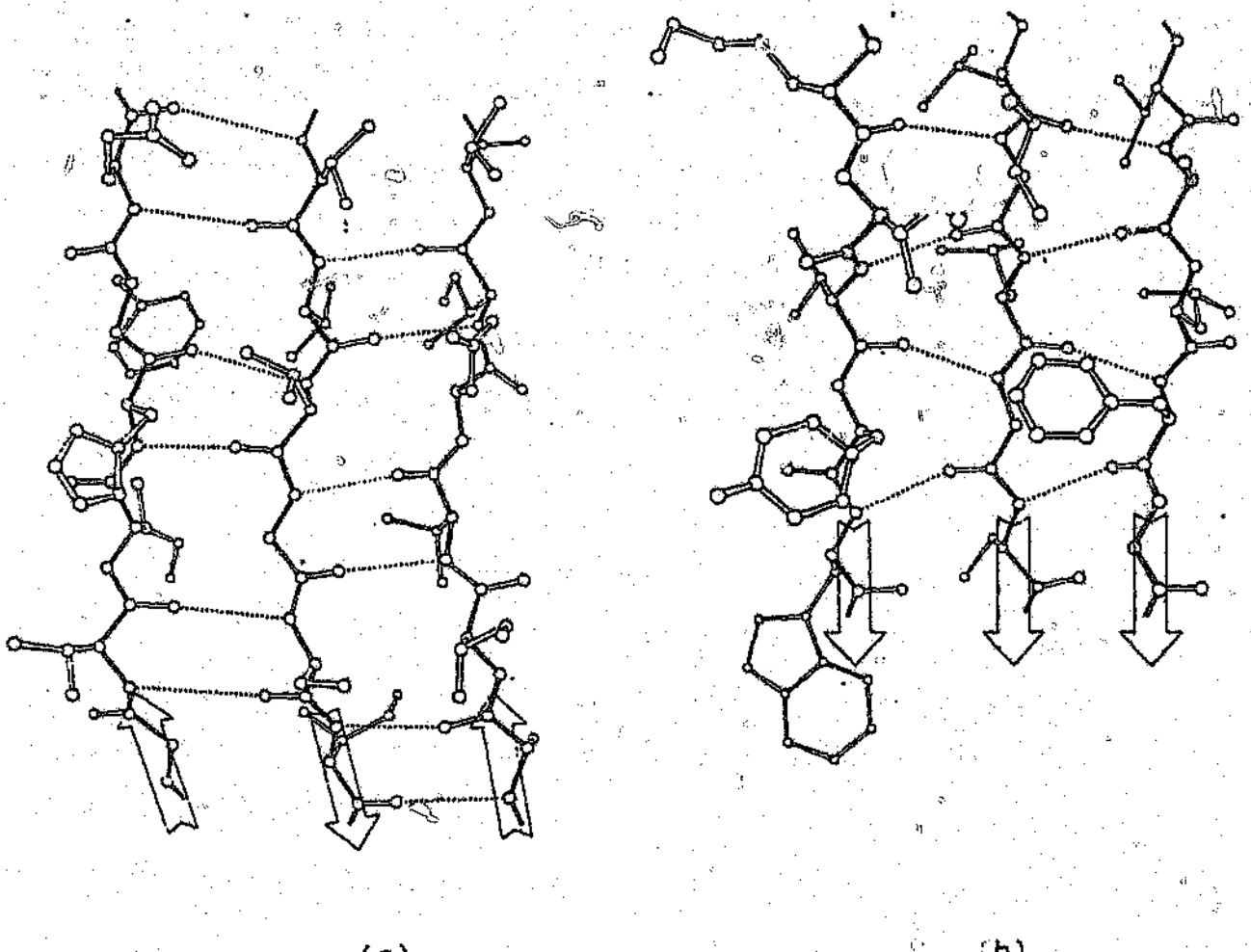
Figure 8 Collagen (a) is made of three left-handed helices similar to the polyproline helix (b). The carbonyl and amine groups face outwards in the left-handed helices and the peptide chains can hydrogen bond with each other. The three helices are wound in a right-handed triple helix. In (a) the amino acid positions 1, 4 and 7 are in close contact with neighbouring chains and there is only room for a glycine residue. Positions 3, 6 and 9 can accommodate proline as there is no requirement for the amine group to hydrogen bond. Positions 2, 5 and 8 can accept a variety of amino acids.[4] This model explains the repeating amino acid sequence of -(glycine-X-proline)- found in collagen (X is any amino acid).[4]





(b)

Figure 9 Antiparallel and parallel ß-sheets are named according to the orientation of the adjacent peptide strands. In (a) and (b) the direction of the peptides (N to C-terminal) is marked by an arrow. Antiparallel ß-sheet is characterised by pairs of widely spaced hydrogen bonds alternating with closely spaced pairs. In (a) the amino acids on the top surface (coming out of the plane of the paper) are hydrophilic, for solvation in the medium, and the bottom surface is hydrophobic.[6] The parallel ß-sheet hydrogen bonds are evenly spaced but slanted. Parallel sheets are usually buried inside a protein core and the amino acids are therefore predominantly hydrophobic.[6]



(a) (b)

steric requirements of parallel sheets are more stringent than antiparallel. This explains why the distributions of phi and psi angles of parallel sheets in proteins are smaller than the antiparallel distributions.

Mixed sheets are rare, probably because of the different phi and psi angles required. In proteins, parallel and the parallel portions of mixed sheets are normally buried under an α-helix.

Antiparallel sheets are exposed to the solvent on one surface. To accommodate this they exhibit a periodicity of 2 in the hydrophobicity of the sequence[6].

1.5.3 Structures with non-repetitive phi and psi angles

Regions in a protein that defy the label of helix or ß-sheet have been classed under the general description of coil. This arbitrary assignment is misleading as the structures are not the same and are often very important in a protein's function. Non-repetitive structures are by definition regions where the phi and psi angles are not regularly repeated.[6] The position, nature and values of the phi and psi angles of each amino acid are far more important in these structures than in repetitive structures where the overall nature of the sequence is the overriding factor.[3] Turns, bulges, hydrophobic pockets, hairpin bends, ordered but unclassified structures, regions of random structural fluctuation and unknown structures are all classified as non-repetitive structure. The first three can be given more

accurate descriptions; the last two are discussed under the heading of random coil.

1.5.4 Turns

A turn is defined as a site where the polypeptide chain reverses its direction. ß-Turns do this with four amino acids and gamma turns require three.[5] Turns are less conspicuous in X-ray analysis than helices because the phi and psi angles are not repetitive. In globular proteins about one-third of the sequence is taken up by turns. Turns are polar, compact structures in which the amino acid side chains face outwards. They combine α-, ß-, left-handed glycine conformations and various hydrogen bonding patterns to give an enormous range of structures even with 3 or 4 residues.[6] Venkatachalam[18], in 1968, predicted turns by using the limits for atomic contact radii set by Ramachandran.[7] Hydrogen bonds were defined as contacts between amide nitrogen and oxygen atoms within 2.6Å and 3.2Å units where the hydrogen bond angle deviated no more than 30° from linear. These definitions had been used for the successful modelling of helices.[18] Conformations with hydrogen bonds separated by three amino acids were calculated. Three types were predicted and labelled I, II and III. The stability of turns has since been shown to be great enough for the requirement of the hydrogen bond to be dropped[3] and they are often found on the surface of proteins hydrogen bonded to solvent molecules. Nemethy and Scheraga have shown that the

standard I, II, III turn definitions require some changes in their dihedral angles to accommodate the hydrogen bond.[19] The three types have mirror images denoted by: I', II' and III'. Type I and II turns differ by a rotation of 180° in the second psi angle (figure 10). The type III turns adopt the same phi and psi angles as a 3_{10} helix. Turns can also be defined in terms of the direction changes of the peptide backbone that they effect.[18] The phi and psi and vector definitions are equivalent and the choice depends upon the type of analysis.

Of the secondary structures, turns have the most fastidious steric requirements with respect to their sequence. Most important to a turn is the change of direction that Trans proline forces on the peptide backbone. Proline forces a change in direction that coincides with the hydrogen bond formed.[3] Types I, II and III invariably place proline in position 2 and the type II' turn places it at position 3 (figure 10).[6]

Turns are pivotal in protein structure. They mark the ends of sheets and helices and they join secondary structural elements together. Their location on the protein surface means that they are solvated. Turns define a set of co-ordinates on the outer surface of a protein between which the sheets and helices ricochet around the protein.[3]

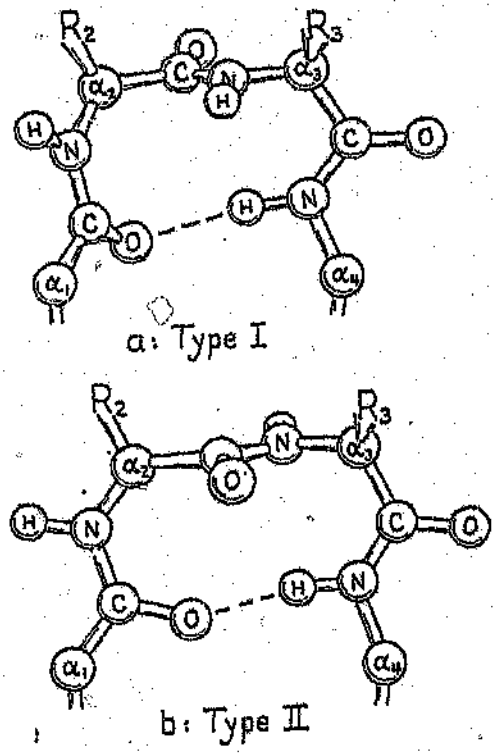


Figure 10 Types I (a) and II (b) ß-turns differ by a rotation of 180° in the peptide bond between residues 2 and 3. These structures are typical of the ß-turns in that they reverse the direction of the polypeptide chain within 4 amino acids.[6] ß-turns are compact structures and place definite steric demands on their amino acids. Types I and II turns usually have a proline at position 2 and glycine at position 3.[6]

1.5.5 ß-Bulges

The ß-bulge is a region between two intermolecularly hydrogen bonded ß-strands where two or more residues on one strand are opposite a single residue on the other.[6] The extra amino acids cause a bulging which results in a bend in the ß-conformation. The angle depends upon the number of residues in the bulge.

1.5.6 Coil structures

The term coil structure is applied to anything that does not fit the label of helix or sheet. Coils can be classified into two groups: those that are genuinely random and fluctuating and those that have an unclassified ordered structure such as the turns and pockets. The former fluctuating or random structures are termed random coil. This type of structure has been modelled by the behaviour of peptides in 6M guanidine hydrochloride.[6] Another term, statistical coil, refers to structures seen in solutions that produce conformational averaging.[19,20] The assumption is that in the solution there are populations of recognised structures.

Genuine random coil can be recognised from X-ray crystallographs by an absence of any electron density or the presence of diffuse electron densities.[6] The cause is due to motion of the regions over the time course of a day or so. The motion can be random or is caused by alternation between different conformations. Some side chains, such as

isoleucine and lysine, show a reduction in their expected electron densities caused by rapid movements from a stable conformation. The regions of actual mobility and disorder may be greater in the solubilised peptide as the crystallisation introduces a bias to the result. The disordered regions are functional as shown by haemoglobin X-ray work.[21] In the transition between the oxy and deoxy forms, the salt bridges in the latter are ordered but in the former they are disordered. However it is not always possible to run X-ray crystallographs to identify these regions and their identification can be a problem.

1.5.7 Factors affecting secondary structure

1.5.7.1 Solvent

In solution, proteins are solvated and the nature of the solvent is important. Solvents that can compete with the peptide backbone hydrogen bonds can disrupt helices, pH alters the charge balance of the side chains and apolar solvents can disrupt the lipophilic core of a protein.

1.5.7.2 Water

Water exists as a liquid at 25°C because of its ability to hydrogen bond with itself.[22] This produces an extended network that holds the water molecules together. The hydrogen bonds are transient, lasting in the order of a tenth of a billionth of a second, and are constantly forming and breaking.[23] The solvation of proteins in an aqueous medium is a complex process and one over which there is

currently much debate. Some of the water molecules surrounding a protein in solution can be considered a part of its structure. Water affects the conformations of exposed side chains, the stability of the terminii of secondary structures, the lipophilic/hydropholic partition of the residues during folding and some active sites.

Those waters that are actually bound to the protein are considered part of the structure. X-ray diffraction pinpoints the occupancy of water at particular sites by measuring the probability of finding a water molecule at a particular site. Predicting the water occupied sites with the Monte Carlo statistical model shows a good correlation with X-ray measured sites[23]. Waters found in a protein core are used as packing to fill space. The water molecules have a high energy relative to the bulk phase solution but they lower the overall energy of the protein. Hydrogen bonds between the carbonyl groups of a protein and water are stronger than those between water and the amino groups.[6] The first hydration layer around a protein is more ordered than the second; by the third or greater layers the bulk phase order is seen[24].

1.5.7.3 Side chains

The uniqueness of a protein is attributable to the primary sequence. In turn the uniqueness of each amino acid results from its side chain. For 19 of the amino acids Levitt[25] has found a weak but significant preference for certain

conformations. Arginine is the odd one out. On the basis of steric hindrances and electrostatic interactions the preferences are to be expected.

The ß-branched and aromatic amino acids, valine, isoleucine, threonine, phenylalanine, tyrosine and tryptophan all favour ß-sheet conformations. These side chains are bulky but they can be accommodated into ß-sheet.[6] The short chain polar residues, serine, aspartic acid and asparagine, prefer reverse turns. This is because they can hydrogen bond back to the chain and are able to remain on the protein surface. Glycine and proline are also commonly found in turns as the low steric hindrance of the glycine allows a left-handed conformation and proline can reverse the chain direction. The rest all favour α-helix. This group of amino acids, except for arginine, has short apolar chains.[23]

The inverse trend of the amino acid preferences is also explainable in the same terms. The ß-sheet group is too bulky for the α-helix or can disrupt the hydrogen bonding. Glycine is too conformationally unconstrained to be able to stabilise a helix and proline is completely disruptive. Proline only forms one hydrogen bond; this is disruptive to ß-sheet. Charged groups in a ß-sheet could force strands apart by electrostatic repulsion. Apolar side chains cannot occur at the surface of proteins, thus excluding them from turns. The aromatic nature of histidine makes it apolar, while the charge gives it polarity and it shows little

propensity for any structure. Cysteine can also be polar and apolar and it has a low statistical probability of being found in turns. Unfortunately these trends are no more than that. Evolution has whittled down the number of amino acids required to the bare minimum with the result that the maximum structural potential is gained from each one.

To summarise: "In general, the impression one takes away from this kind of examination is that the protein is balancing so many factors at the same time that there are always ways to compensate for any individual problem. The studies of individual parameters uncover only weak regularities in spite of the strength of the overall packing constraints."[6]

1.6 Secondary structural prediction methods

A secondary structure has to fit into one of the broad categories now recognised: helix, turn, sheet or coil. The first three are based upon ideal models with limited bond angles governing their definitions. The X-ray crystallographic measurements of secondary structures show a great deal of variation from the ideal. The problem of classification then arises. Currently the designation is done in a number of ways. Kabsch and Sanders[18] provided a comprehensive set of rules based upon hydrogen bonding patterns and phi and psi angles. This works well for regular structures but breaks down at the often irregular ends of a structure. It also suffers from the fact that slight

inaccuracies in the angles measured can greatly affect the result. Levitt[25] includes the inter-C_α distances to reduce the level of doubt. For some work the designation of secondary structure is qualitative. When a prediction method is evaluated it is only valid for the definitions within which it has been set up. Despite the problems a prediction method is needed because DNA technology has produced sequences for which there is no X-ray structure. The situation is worsened by the fact that the crystallisation of proteins is a slow and difficult process.

Predictive methods depend upon statistical analysis, information theory, stereochemical theory and statistical mechanical calculations. The latter three originate from the work by Liquori et al[26] and Ramachandran[7] and assume that proteins fold to a free energy minimum. By using the techniques encompassed by each field the aim is to predict the lowest energy conformation. The results of the energy calculations have proved so far to be unfruitful for two reasons: the volume of calculation required for even simple molecules and difficulties with the application of the quantum theory to chemical problems. The most productive method so far has been the statistical approach. The frequencies of amino acids in particular secondary structures are measured and the information ultimately produces a value for the propensity of an amino acid for a given structure. It has already been pointed out that these

statistical correlations produce only general trends and therefore rules have to be set up to apply the statistical information. This is the way that the Fasman and Chou predictive methods operate.[27,28]

1.7 Tertiary structure

Biology stands on a threshold that cannot be crossed without first discovering the forces that fold and stabilise proteins and other macromolecules. Much of the work in molecular biology relies upon a hit-and-miss approach because of our inability to predict the behavior of macromolecules. Without such predictive abilities major academic advances in medical technology and biotechnology will not be forthcoming. With it the design of proteins and the manipulation of molecular structure will be possible. For these reasons the study of protein stability and folding has attracted huge interest.[10]

Measurements of the stability of whole proteins are based upon the energies of disruption of native proteins.[29] A denaturing solution of a protein contains essentially two populations of molecular species, the native and denatured. The number of intermediate forms has been found to be low. The folding and stabilisation of a protein is therefore a cooperative process. The non-covalent forces responsible are discussed.

1.7.1 The hydrophobic effect

The work by Kauzmann showed how the amino acids of a globular protein partition in an aqueous solution to produce a central hydrophobic core surrounded by a hydrophilic exterior[30]. When a protein comes into contact with water the hydrophobic amino acid side chains aggregate and force the hydrophilic side chains into the solvent. Water is excluded from the centre which becomes predominantly apolar. This provides an initial organising force for folding. This "oil drop" theory[4,31] is driven by entropy. The insertion of a hydrophobic side chain into water forces a break in the water's hydrogen bonded structure and the solvent molecules form an ordered cage or clathrate structure around the hydrophobic group. The positive enthalpy change from the breaking of hydrogen bonds is offset by a negative enthalpic change due to the setting up of London forces between the apolar groups and water.[32] The attractive forces of apolar molecules for each other is similar in magnitude to the attraction of apolar groups for water. The driving force for the removal of the hydrophobic side chain is therefore not enthalpic. Instead the aggregation of the hydrophobic side chains increases the entropy of the water by destroying the clathrate structure. The aggregation moves to minimise the surface area of contact with water. For every side chain removed from water the increased entropy results in the release of about 8KJ mol^{-1} of free energy.[32] A corollary of

this is that the stability of a protein should depend in part on the data gained from simple experiments that determine the solubility of hydrocarbons in water. These simple studies have been extended to measure the free energy of transfer of amino acids from water to apolar organic solvents.[33,32] The work by Baldwin[34] on this correlates well with the positions of amino acids in a protein as determined from X-ray crystallography. There are however numerous examples of hydrophobic amino acids that occur on protein exterior surfaces. More work is required if direct free energy measurements from solution studies are to be extrapolated to proteins[32].

1.7.2 Electrostatic interactions

The electrostatic interactions in proteins cover a wide range of effects with differing spatial and geometric dependencies. Of these, the best characterised are salt bridges, hydrogen bonds and van der Waal's forces.[32] There is evidence that the pattern of close packing seen inside the hydrophobic core of a protein is dictated by the necessity to fill the core to remove water and to ensure the asymmetric distribution of electrons around nucleii. The asymmetric distributions cause enthalpically favourable London forces in a protein core where the atoms are in close proximity. These interactions are electrostatic in nature but exist between normally apolar amino acids. They are stronger than van der Waal's forces and are enthalpically

equivalent to hydrogen bonds. They are categorised as weakly polar. The electrostatic forces in a protein are the most important after the covalent linking of the primary structure.[30]

The strength of an electrostatic interaction depends upon its enthalpic favourability. Direct measurement is often not possible, nd quantum statistical equations are used.

1.7.2.1 Charge-charge interactions

The carbonyl and amino terminals of a protein are nearly fully ionised at physiological pH and are charged. Similarly the side chains of aspartic acid, glutamic acid, cysteine, tyrosine, lysine, arginine and histidine can carry charges.[22] These amino acids are found mainly at the surface of a protein where they interact with each other or with water. Tyrosine, arginine, lysine and glutamic acid have apolar side chains that end in charged groups. To accommodate this the apolar chain is buried beneath the protein surface and the charged group is extended into the solvent. Charged groups inside the core of a protein are found within 4Å units of another opposite charge or are hydrogen bonded. At active sites in peptides the charged groups form important ligands such as the binding of 2,3 diphosphoglycerate in deoxyhaemoglobin. The force of attraction between two charges is inversely proportional to the distance between them and is spatially anisotropic.[22]

1.7.2.2 Dipoles

The size of a dipole moment along a bond depends upon the difference in the electronegativities of the two atoms.[32] The Pauling scale gives a comparative measure of this. The common atoms of proteins have the following values:[32] hydrogen 2.13, carbon 2.55, sulphur 2.53, oxygen 3.45 and nitrogen 2.98. Therefore all the covalent bonds in a protein have a permanent dipole, albeit small. The magnitudes of the dipole moments in proteins vary between $0.35\epsilon^-$ and $0.10\epsilon^-$. The lower values are found in the aliphatic chain bonds and the higher values are found in the polar side chains.[32] The aromatic side chains occur at values around $0.15\epsilon^-$. The existence of the bond polarities mean that the aromatic groups have quadropole moments similar to benzene.

1.7.2.3 The interactions of charges with dipoles

The existence of two opposite charges close to one another causes a dipole. The separation of the amino and carboxyl terminal charges in an α-helix sets up a dipole along the length of a peptide. In addition the dipoles from each hydrogen bond are aligned and the result is a large net dipole. Counteracting this there are usually positive and negatively charged side chains found at the carboxyl and amino terminals respectively to stabilise the helix.[30]

In a protein the strength of interaction of dipoles and charges is spatially anisotropic with respect to the dipole moment. To maximise the interaction therefore the geometry

of the dipole and charge are fixed, which in turn is a means of spatially organising the two[32].

1.7.2.4 The interactions of dipoles

1.7.2.4.1 Hydrogen bonding

The hydrogen bond is stable and disruption requires approximately 10 to 30 kJ mol^{-1} and is about ten times more stable than other dipole-dipole interactions.[35] The hydrogen atom has a sufficiently positive charge to displace electrons from the acceptor and there is some covalent character to the bond formed. Hydrogen bonds using a -C-H donor group have been identified in neutron scattering and X-ray crystallographs of proteins.[6] Normally the donor is a hydroxyl or amino group. The most common acceptor in proteins is the carbonyl of the peptide bond. A linear hydrogen bond is the most stable arrangement. The dipole strength of the bond varies with the Cosine of the angle of deviation from linear. A hydrogen bond bent by 25° has 90% of the strength of a linear hydrogen bond.[32] Therefore structures, like the α-helix and antiparallel ß-sheet, whose hydrogen bonds can be non-linear, are still stable. The angle at 25° allows an oxygen atom to accept two hydrogen bonds and this is often seen.

An analysis of hydrogen bonds found in proteins has led to a series of generalisations about them.[32] The list is as follows:

1) Almost all groups capable of hydrogen bonding, do so.

2) In secondary structures the main chain carbonyls and amino groups are bonded to each other.

3) Polar side chains are capable of forming more than one hydrogen bond. Aspartic acid, glutamic acid, glutamine and asparagine can form up to four; arginine can form five. Side groups incorporating these amino acids can then form extended networks of hydrogen bonds.

4) Interactions between the side chains and the main chain usually involve serine, threonine, aspartic acid and asparagine. These hydrogen bonds are often used to tie up free peptide bond groups at the terminii of α-helices.

5) In helices, the interactions of side chains with the main chain are dominated by serine and threonine which form a hydrogen bond to main chain carbonyls.

6) Side chains at the amino terminal of an α-helix more frequently form hydrogen bonds than those at the carboxyl terminal.

7) Main chain carboxyl groups show a preference for accepting two hydrogen bonds. They are angled at 120°.

8) In order to accommodate the second hydrogen bond the carbonyl groups are tilted.

9) Some α-helices are found in proteins bend. The hydrogen bonds on the outside of the bend are longer than those on the inside.

10) In ß-sheet the hydrogen bond deviations from linear are similar to those found in helices. There is a significant

decrease in the number of double bonding arrangements around carbonyls.

11) The internal and solvent hydrogen bonding in a protein is an integral part of the structure.

Hydrogen bonds are the strongest and most obvious dipole-dipole interactions seen in proteins but they are not the only possibilities. The strength of interaction depends upon the distance between the centres of the two dipole moments involved and the angle at which they relate to each other.

1.7.2.4.2 Weak polar interactions

The weak polar interactions seen in proteins span a wide range that depends upon different distance and geometrical parameters. The protein core is stabilised by the hydrophobic effect[4]. This brings atoms into close proximity with one another in the core. The proximity of the atoms then allows the possibility of a myriad of London forces[52] between normally apolar side chains. This produces an attractive multipole moment between the atoms. Added to this are van der Waals forces and the interactions of aromatic groups. To maximise the stability to be gained from these effects the participating groups must be spatially organised. Thus the packing of these side chains is not random. The conclusion is that the weak electrostatic forces that occur between normally apolar side chains dominate.

They are not random and will direct the precise position of packing in the core[32].

1.8 Folding of the peptide chain

Anfinsen showed that the primary sequence provides all the information required to fold a protein into its native state.[2] This was done by denaturing ribonuclease A and then allowing it to renature in a solution similar to its normal physiological environment. Reversible denaturation is a property shown by most proteins tested so far. The next, and as yet unsolved, question is how the folding process occurs. The question is allied to that of the prediction of tertiary structure.

There is no single picture of protein folding that fits all of the evidence. Any discussion of folding can at best only present the trends in current thinking. To summarise what is a hotly debated issue: "In short, protein folding has been defined as a spontaneous kinetically oriented process with the steps: local folding, followed by long range interactions, and then rearrangement. The question of disulphide bond formation remains still to be answered."[36] If a native structure lies at an energy minimum, either global or local, folding might be a random search through the possible conformations to a low energy state[2,44]. Proteins renature in time periods that vary from seconds to hours. A random search under thermodynamic control through all the conformers possible does not fit this time scale. A

small protein has about 10^{148} possibilities to search through[38] and would take millenia to fold. The conclusion is that folding is under kinetic control. Global and local minima have different kinetic requirements. To fold to a local minimum a specific pathway is required as the vast number of potential conformers allows a protein enough space to fold to several minima without any preference for a particular one. The native structure is thought therefore to lie at a global free energy minimum.[38] A kinetic mechanism then would demonstrate one or several pathways. In both cases there should be identifiable intermediates. In metabolic studies it is possible to stop a pathway at a particular point so that the concentration of an intermediate can be increased. In protein folding the problem has been to identify intermediates as they are transient, in very low concentration and unstable. It is possible to monitor folding to a limited extent.[39] Techniques include the addition of fluorescent markers along the sequence in order to measure changes in distances between residues[40] and changing the pH to arrest the folding. From such studies folding has been shown to occur in two stages: a fast initial phase and a rate limiting slow stage[41]. Folding bridges the gap between the interactions of neighbouring amino acids in secondary structures and the longer range interactions of the folded protein. It would be expected that there should be a hierarchical process

involved whereby over time successively more complex interactions are directed.

The currently accepted hypothesis is that the first event in folding is either a hydrophobic collapse or the formation of local secondary structural initiation sites.[39]

1.8.1 Early events in protein folding

1.8.1.1 Hydrophobic collapse

The aggregation of the hydrophobic side chains is a two-step process. The first is non-specific and is analogous to micelle formation. The second is a specific organisation of the side chains into a molten globule state. The nature of the hydrophobic interactions has been investigated with carbonic anhydrase. The structures observed were transient and their half lives, in the order of seconds, were measured. Initially hydrophobic clusters were observed, $t_{1/2}$ of 0.03 seconds. These aggregate to a rigid hydrophobic core, $t_{1/2}$ of 140 seconds, and the native structure appears at 600 seconds.[49] The molten globule state is a compact hydrophobic structure which shows extensive secondary structure when investigated with circular dichroism.[39] It is as compact as the native protein but there is a greater amount of fluctuation. The protein is still different from its native state even with the large secondary structural content of the molten globular state. These results can be related to the specific nature of weak apolar side chain interactions discussed above.

1.8.1.2 Initiation sites

Potential hydrogen bonding groups, in particular the carbonyl and amide groups, in protein secondary structures have to compete with water for their hydrogen bonds. It was thought that the cooperative effect of a whole protein was the factor that allowed the backbone to win this competition.[44] The discovery of marginally stable secondary structures in small peptide fragments was an unexpected result. Small peptide fragments can fold into α-helices, β-sheets and turns.[39,45] They are marginally stable and in equilibrium with random coils.[45] With this revelation it is now envisaged that along the unfolded protein, sections are capable of autonomous secondary structural formation. These flickering regions can form the initiation sites for further folding.[39,38] Chance collisions for instance increase their local stability thereby allowing further growth. This model is sometimes called the framework model[39]. The small secondary structures may not be native but they would be able to direct folding. The secondary structural nature of the initiation sites means that initially the short-range interactions between neighbouring amino acids are dominant.[46] The model allows early folding pathways to be non-specific. Several pathways reduce the importance of any one amino acid and, from an evolutionary point of view, limit the damage done by mutations.[47]

The chief support for the framework model is the fact that small peptide fragments are capable of forming secondary structures. The next question is what secondary structures provide good initiation sites? In isolation the α-helix is less stable than other structures.[39] Conversely the stability of early structures may not be important as they are transient. A low stability also allows incorrect structures to be reversed.[45] The β-turn is an excellent candidate as it directs long-range interactions by reversing the chain.[45] Turns require no interactions longer than between four neighbouring amino acids and they are stable.[30] They occur at the junction between secondary structures in proteins and it is not difficult to see how a two-stranded antiparallel sheet could be initiated by a turn. Type III turns have phi and psi angles identical to those of a 3_{10} helix and there is not a large jump in energy between that and an α-helix.[6] The β-turns are pivotal as they sit at a crossroads between the sheets and helices. They also have fastidious steric requirements such as the correct positioning of prolines. This makes them vulnerable to mutations.[47]

The hydrophobic cluster is also a possible initiation candidate. The specificity of aggregation would vary dependent upon the amino acid sequence.[40] Matheson and Scheraga have calculated that four or five residues are the

minimum number for a hydrophobic pocket to initiate folding.[48]

1.8.1.3 Specific interactions

The molten globule and the framework model imply that protein folding is dictated by a limited number of non-specific interactions that limit a protein's conformational space. It is possible to envisage a more specific mechanism[39] where the interactions would be directed for example by the correct disulphide bond formation. The low frequency of occurrence of disulphide bridges and the rate limiting step for their correct formation have ruled out such models.[39]

1.8.1.4 Conclusions about early events in protein folding

The enthalpy changes of a denaturing protein, assuming that this is the same as folding in reverse, have shown that during the folding of small proteins the number of intermediates is small. Thus folding must be a cooperative process.[49] Larger proteins denature in several discrete cooperative blocks. Separating out the individual contributions from each type of interaction is not possible. The low population of the intermediates in the folding process observed does not mean that a simple two-state model is correct. Measurements of the folding intermediates with nuclear magnetic resonance and proteases show that they are different from the native and denatured states.[46]

Overall the evidence for the early events in folding dictate that there is one path or a limited number of pathways, the interactions between amino acids are initially local and non-specific and result in the formation of a molten globular state and or flickering secondary structures.[48] The mechanisms are by no means mutually exclusive and it is possible that a hydrophobic collapse provides a surface upon which flickering initiation structures can form, such as an amphilic helix.[39,17]

1.8.2 Late events in protein folding

The early folding events occur in times of the order of seconds or less. There is a final late phase that has been shown to be rate limiting.[50] The evidence for this comes from denaturation experiments. The slow phase corresponds to a transition among two or more denatured states. The transitions are thought to be the result of Cis-Trans isomerism at proline residues.[50,51] The rate of Cis-Trans interconversions can vary by a factor of ten depending upon the amino acid found on the proline amino terminal. The rate in actual globular proteins is unknown.[51] The evidence suggests that an increase in steric hindrance from the amino terminal side chain increases it.

1.9 Prediction of tertiary structure

Quantum theory can be used to predict the conformations of small molecules, such as ammonia, in the vapour phase by finding the lowest energy conformation. A computer program

does a global energy search.[52] The quantum mechanical operators used in physics have to be altered because the time scales for the movements of nuclei are far longer than those for sub-atomic particles. The quantum mechanical approach involves huge numbers of calculations even for small molecules, making it impossibly long for use with proteins. For instance if three phi and psi angles are allowed per residue, a 149 amino acid sequence has 148^{20} conformers to search through[34] without considering the side chain angles. At a rate of a nanosecond per trial it would take approximately 10,100 years to try every one. Two approaches to reduce the number of calculations are used: the latest state-of-the-art energy function can be built into an algorithm or an energy function based upon the tertiary structure of a protein can be modified. In both cases success has been limited. In the second case a structure is being used to predict itself in the hope of finding the key to calculations for unknown structures. For both there are multiple energy minima found close to the native conformation and choosing between them is a problem.[39]

The notion of randomness has obsessed modern physics for some time[53] because it poses the question: does prediction depend upon our own limitations or upon some principle that makes it impossible? Simon de Laplace used Newton's laws to say that prediction would be possible forever. Newton's laws

are deterministic with everything in its place. Quantum mechanics and the Heisenberg Principle have introduced the concept of probability and randomness. The notions of chaos and randomness have become unifying principles in science.[53] It was proposed by Dirac that with quantum mechanics the whole of chemistry could be reduced to simpler physical terms[54]. This has not yet happened and is probably not possible because in science the number of axioms for a theory are reduced to the minimum by Occam's razor. To effect the transition from physics to chemistry it was necessary to modify the standard Hilbert geometry of quantum mechanics. This change had to include concepts alien to physical calculations such as electronegativity. A lesson can be learnt here in biochemistry. The use of Occam's razor has excluded important biochemical concepts from chemistry, such as hydrophobicity and solvent effects. In biochemistry the search is on for a theory to predict protein structures, from the primary sequence, that can encompass all possible solvents, pH values and temperatures. An energy minimising computer program, based upon quantum mechanics, to achieve this goal will only be possible when quantum mechanical principles are adapted to biochemistry.[54] This approach is the one that should eventually prevail. In the absence of this theory, the continued experimental approach is the only option available.

1.10 Relevance of polypeptide models

The speed and simplicity of producing short peptides using solid phase peptide synthesis makes them ideal experimental models. They can be used to measure the specific contributions of short- and medium-range interactions, they are small enough to allow extensive energy calculations and they can be used to test folding models.[88] The initial hypothesis was that short peptides were incapable of forming organised structures in aqueous solution as the competition with water for hydrogen bonds was thought to be too great.[44] This has been found to be untrue and ß-turns have been found in peptides with as few as 4 amino acids[44]. α-Helices and ß-conformations are all found in a range of solvents with a variety of sequences. Cyclic peptides were originally used as these peptides have no carboxyl or amino terminal effects. The limitation is that there are conformational constraints on the side chain geometries that are inappropriate for protein models. Linear oligotides are now used.[85] Small linear peptides are flexible molecules and easily interconvert between structures.[30] This can be controlled to some degree by changing the physico-chemical environment. One application of polypeptide models is in guest-host experiments.[44]

1.10.1 Guest-host experiments

Homo-oligopeptides were the first linear peptides to be analysed[44]. The analysis of the results was facilitated by

the regularity of the chemical groups being monitored. An offshoot of this was the guest-host experiment. In these experiments a homo-oligopeptide is chosen and its inherent secondary structure under specific conditions is measured. A guest amino acid is then substituted into the sequence and any changes measured.[44] "The insertion of the guest amino acid is expected to result in a conformational transition, depending on the structural preference of the guest residue."[63]

L-alanine oligopeptides are useful hosts because this amino acid has been found in left- and right-handed helices and ß-sheets.[63] L-alanine with L-valine is generally free from side reactions in solid phase peptide synthesis.[58] It has the disadvantage that in aqueous solutions the solubility of polyalanine is low. For this reason many workers use amino and carboxyl terminal solubilising groups. For poly-L-alanine of 7 amino acids or more α-helix and ß-sheets have been seen. The solvent, concentration, pH, solubilising group and temperature all have an effect.[55,57] The ß-structures are responsible for the low solubility noted for these peptides. In aqueous and polar solvent, the polarity of the solvent causes aggregation of the peptide to ß-structure and a decrease in solubility. The association of the peptide chains is disrupted by small amounts of acid[56] and fluoroalcohols.[55] Short peptides are unstable and there

is an equilibrium set up between random coil and ordered structures in solutions.

1.11 Leucine and isoleucine discussion

L-isoleucine and L-leucine are both hydrophobic amino acids with aliphatic side chains. They both have the same molecular weights and differ in the position of the methyl group. This difference affects the structures in which the two occur. Leucine is found in α-helix and ß-sheet[59] and statistically in proteins it has a preference for the helix. The substituted C_β of the isoleucine makes it a helix breaker[59] and it is found in ß-sheets.

The insertion of leucine and isoleucine as guest amino acids in a poly-L-alanine host peptide causes the formation of ß-structures.[58] Poly-L-alanine aggregates into ß-sheets at concentrations as low as 0.5mg/ml.[60] Complete disruption of the ß-structure occurs at 0.01mg/ml and partial disruption starts at 0.1mg/ml.[60] The insertion of a guest amino acid into poly-L-alanine results in a decrease in the ß-structural propensity of the peptide. This occurs regardless of the intrinsic ß-forming potential of the guest.[61] In terms of ß-forming propensities, L-alanine is stronger than L-leucine which is stronger than L-isoleucine.[61] If short range interactions between the amino acids were the dominant factors in ß-formation, the reduced ß-forming potentials of the guest-host peptides would not be expected to drop. The conclusion is that the

short range interactions are not dominant. The relative sizes of the side chains,[58] probably due to the steric constraints imposed on the chain by aggregation,[6] are important considerations.

1.12 The aims of this work

The importance of an understanding of the nature of protein folding has been stressed. The framework model of folding[39,58] relies upon initiation sequences of amino acids that form incipient secondary structures. A candidate for initiation sites is the hydrophobic pocket.[48] Previous work in this laboratory on a series of hexa-L-alanine host peptides substituted with L-valine showed that TFA.H.val(ala)$_5$.OH formed a hydrophobic pocket in methanolic solutions. Other findings were that the C-terminal substitutions of guest amino acids were better than N-terminal substitutions at forming ß-structures.[62] This work aims to extend the study with a series of homologous L-alanine hexapeptide guest amino acids substituted with L-leucine and L-isoleucine. These amino acids form a series of hydrophobic aliphatic side chains with L-valine. They differ in size and the position of the side chain C_β branching.

Another aim of the work was to develop a method for producing deuterated L-alanine by the reduction of L-cysteine with Raney-Nickel that had adsorbed deuterium. The deuterated L-alanine is a useful tool for identifying interactions with L-alanine in a synthetic peptide. For

example it could be incorporated into L-valine-L-alanine$_5$, where deuterated L-alanines could be used to identify the intramolecular interactions stabilising the hydrophobic pocket by means of a nuclear Overhauser experiment.

2. DISCUSSION OF THE METHODS

2.1 Solid phase peptide synthesis

The synthesis of peptides in solution often requires the crystallisation of the intermediates at each step.[63] The method can be laborious and time-consuming. It is used to produce amounts of peptide in the order of kilograms. In 1963 Merrifield revolutionised peptide synthesis with his solid phase synthesis method.[64] "The fundamental premise of solid phase peptide synthesis is that amino acids can be assembled into a peptide of any desired sequence while one end of the chain is attached to an insoluble support".[64] The peptide is then cleaved from the support and liberated into solution. The solid phase synthesis intermediate steps involve washes in solvents with different polarities. Changing the polarities of the solvents swells and shrinks the resin support which is used to purify and wash the peptide. Solid phase peptide synthesis provides speed, convenience and the possibility of automation. The process can produce in the order of tens of grams efficiently. For analytical work only milligrams are required and solid phase peptide synthesis is therefore ideal. The scheme for solid phase peptide synthesis (figure 11) is split into esterification, deprotection, coupling and cleavage steps. The coupling and deprotection are repeated until the sequence is finished. Finally the peptide is cleaved from

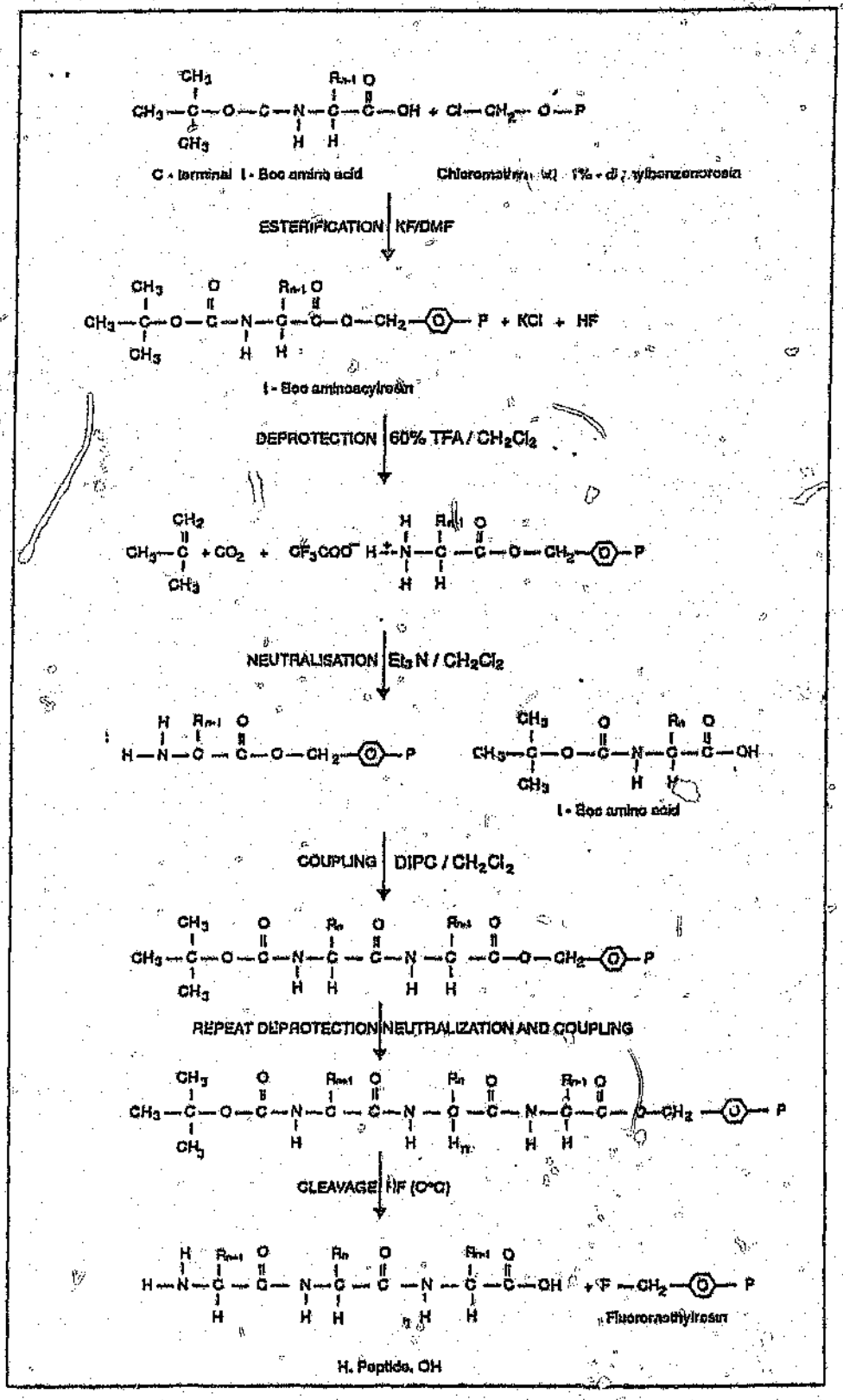


Figure 11 An overall view of solid phase peptide synthesis using t-Boc amino acids.[64] p= polystyrene-divinylbenzene copolymer resin; R= amino acid side chain; DIPC= N,N-diisopropylcarbodiimide. DCC can be used instead of DIPC.

the resin and purified using HPLC.

The chemistry of solid phase peptide synthesis dictates that the ester link between peptide and resin, and the side chain blocking groups must be more stable than the N-terminal protection group (figure 11). This ensures that these bonds are not labile during N-terminal deprotection. The relative strengths of the three will be chosen according to the requirements of the synthesis.[64] The lability of the ester link depends upon the electron density around the α-carbonyl oxygen. The choice of an ester is determined by the length of peptide and the amino acids. Longer peptides require a stronger ester linkage. The two most commonly used N-terminal protecting groups are fluorenylmethoxycarbonyl (Fmoc) and tert-Butyloxycarbonyl (t-Boc). Fmoc is convenient because cleavage of the peptide from the resin can occur in trifluoracetic acid (TFA), needed to remove the t-Boc protecting group.[64] The disadvantage is the increased cost of the Fmoc synthesis. t-Boc is the most common protecting group but cleavage requires more stringent conditions.[63,64]

The solid support has to show chemical stability to the reagents used. It also has to have enough physical stability for manipulation, filtration, and continuous swelling and shrinking in the solvents.

Solid phase peptide synthesis by its nature is amenable to systemisation and this reduces the chances of error. The major loss of peptide yield in liquid phase synthesis was

due to the transfer of material between containers.[63,64] Solid phase peptide synthesis is carried out in one container which reduces this loss. However there is the danger of loss at each step by incomplete reaction and side reactions. A small accumulated loss at each step in a synthesis can result in a large loss overall.[64] Thus the extent of each coupling reaction is checked at each step. A major loss can be avoided by ensuring that the reagents are dry and that hydrolysis of peptide bonds does not occur.[64]

2.1.1 Solid phase peptide synthesis using tert-Butyloxycarbonyl amino acids

Solid phase peptide synthesis (SPPS) using t-Boc amino acids was the method of choice in the work reported here and is discussed in more detail. t-Boc synthesis uses a polystyrene resin crosslinked with 1% divinyl benzene. Nuclear magnetic resonance has shown that in this state the core of the beads is equivalent to the bulk solution phase and there is no evidence of macroscopic insolubility.[63,66] Some peptides precipitate upon cleavage from the resin implying a positive solvation effect caused by the resin.[66] The polystyrene polymer requires activation before the first amino acid can be attached. This is done by chloromethylation. The extent of methylation affects the resin's application. Peptides of 15 residues or less are best synthesised with 1.8 millimoles of chloromethylated sites per gram.[65] For longer peptides a decrease in this substitution is required.[65] The

chloromethyl reaction uses chloromethyl methyl ether, a strong carcinogen, and commercially available chloromethylated resins are normally used. The resin is stored below room temperature as the chloromethyl groups can react to form interchain bridges which increase the crosslinking factor.

2.1.1.1 Esterification of the first amino acid to the polymeric support

The first amino acid is attached to the chloromethylated sites in the resin (figure 11). The reaction is a nucleophilic attack on the partially positive carbon of the chloromethyl by the amino acid's α-carboxylic acid. The reaction is carried out in dimethyl formamide to promote nucleophilic reactions. Potassium fluoride is added as a catalyst.[64] The presence of free chloromethyl groups on the resin promotes side reactions and it is important to ensure a high level of esterification.[65] The substitution level can be monitored with the Gisin Test.[67] A poor percentage substitution of the chloromethyl groups requires that the resin be subjected to esterification again. Once complete the N-terminal deprotection of the first amino acid can begin.

2.1.1.2 Deprotection of the amino terminal of the peptide

The deprotection is effected with a solution of TFA in DCM. The resin is initially washed in the acid solution for a short time to ensure that there is no macroscopic dilution in the beads. Ethanedithiol is added to scavenge for tert-butyl carbocations.[66] The deprotection leaves the TFA salt of the amino terminal and a neutralisation step is required. A tertiary amine such as triethylamine (TEA) is used. This leaves the amino terminal free for coupling to the next amino acid. Two side reactions occur during and after deprotection (figures 12 and 13). These side reactions both reduce the yield of peptide. If the trifluoroacetylation (figure 12) occurs on a residue near the end of the synthesised sequence the resultant incomplete peptide may be difficult to separate later from the desired product. Figure 13 shows the reaction that terminates peptides by the formation of diketopiperazines.[64] The reaction rate is increased in the presence of weak acids and is particularly a problem after the coupling of the first amino acid. To reduce the risk of these side reactions: the deprotection reaction time is minimised, the acid is flushed out before the addition of the base, the acid is dissolved in DCM because DMF promotes the side reactions and complete neutralisation is ensured by rinsing with TEA first to stop any dilution of the base.[64] These steps are necessary to avoid large losses that result from the accumulation of

small losses at each step. To ensure complete deprotection a small sample of the resin is tested qualitatively with the Kaiser test.[69] The qualitative approach is adequate because it is sensitive to 0.5% of the available amino groups.

2.1.1.3 The coupling of amino acids

To activate the carboxyl group of the incoming amino acid it is converted to an O-acyl-isourea by treatment with DCC, a symmetrical anhydride or an active ester (figure 14).[65] The acylureas produced are sparingly soluble in DMF and insoluble in DCM.[65] DMF is a better solvent than DCM even though it promotes N-acylurea formation, useless to the coupling.[64] Diisopropylcarbodiimide is becoming more widely used because it is a liquid and not solid at room temperature, its acylureas are soluble in DMF and it has a similar reactivity to DCC.[65]

To ensure complete coupling excesses of the amino acid and carbodiimide are added. The solvent volumes are kept to a minimum to maximise the reactant concentration. When the O-acyl-isourea pathway does not give a quantitative reaction the active ester path with 1-Hydroxybenztriazole (HOBT) is commonly used (figure 14). In general racemisation is a slow process and over several hours is not a problem in the syntheses. If difficulties are experienced and a coupling requires reaction times above eight hours, the reaction is restarted with fresh reactants.

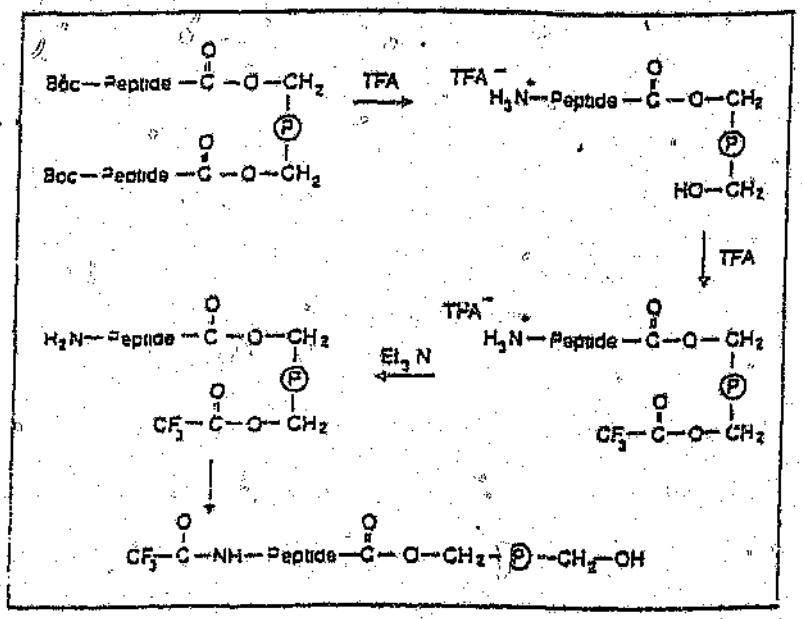


Figure 12 The termination of peptides by trifluoroacetylation during deprotection.[64] P= polymeric support; EtN_3= tertiary amine.

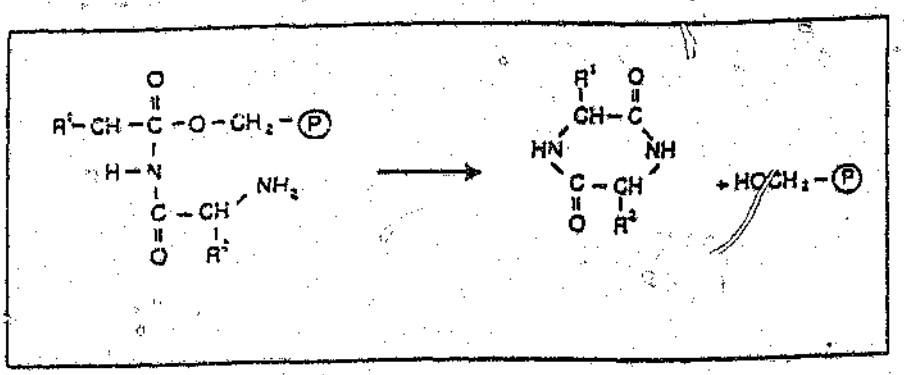


Figure 13 The cleavage of peptides by the formation of a diketopiperazine.[64] p= polymeric support; R= amino acid side chain.

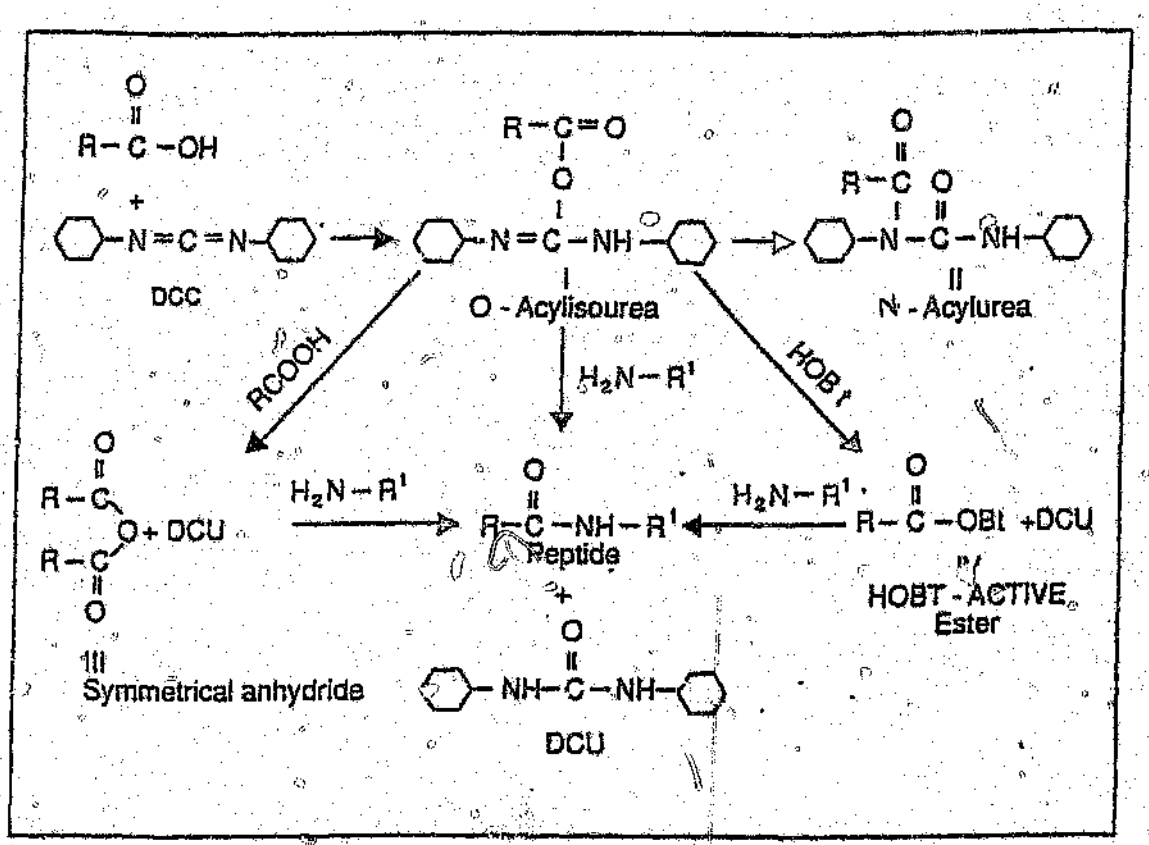


Figure 14 The DCC coupling scheme.[64] DCU= dicyclohexylurea.

The symmetrical anhydride coupling route (figure 14) is thought to occur to some extent in normal DCC couplings as a side reaction. To promote it the stoichiometry of the added reactants is changed. The reaction is not always suitable for use in DMF but when used can take in the order of twenty minutes for some couplings.[64]

To reduce yield losses all coupling must ideally proceed to 100% completion. A 2.5 times excess of amino acid is added to ensure this.[64] The Kaiser ninhydrin test[68] is used to ensure that the free amino groups detected after the deprotection are all reacted.[68] In the event that a coupling does not reach completion after several attempts, the unprotected peptides are acetylated thus terminating their continued growth.[64] Deletion peptides occur when an N-terminal amino group misses a coupling step because it was not deprotected or it was trifluoroacetylated (figure 12). Deletion peptides differing by one or two amino acids from the desired product can be difficult to separate from the product. The achievement of a complete coupling reaction is the single most important step in the synthesis.[65,66]

To produce a sequence the deprotection and coupling steps are repeated using the desired amino acids. It should be noted that, although a routine procedure, solid phase peptide synthesis is not without its difficulties. Syntheses have many complicating aspects such as unfavourable conformations of the intermediate peptides.[64] Once

synthesised the last amino terminal is deprotected, the peptide is washed in DCM and ethyl ether and dried in a stream of nitrogen. The final stage of the synthesis is to cleave the peptide from the resin.

2.1.1.4 Cleavage of the peptide from the resin

The cleavage agent is liquid hydrogen fluoride (HF). Despite the potential hazard it is the most convenient manner of cleaving peptides from resins as it removes side chain blocking groups at the same time.[69,65,64] A full description of the HF treatment is given in the appendix. In outline the reaction vessel with the resin and a stirrer are cooled with dry ice in methanol to -2°C. The pressure is decreased and a mixture of dry HF with 10% anisole is condensed over into the resin. The anisole is an aromatic compound that traps carbocations. The hydrogen fluoride treatment lasts for 45 minutes at-2°C.[65,64,69] The HF is removed under vacuum and disposed of.

The free peptide and resin mixture is washed in a sintered glass funnel with ethyl acetate to remove the anisole and the peptide precipitates out of solution. It is dissolved and washed through the sinter with TFA and acetonitrile.

2.2 Purification of peptides with high pressure liquid chromatography

The crude peptide contains deletion peptides and impurities from the synthesis. A short, easily synthesised peptide that showed no incomplete couplings can be 70-90% pure after

cleavage. The degree of purity required will depend upon the use.[65] For conformational work maximum purity is required and high pressure liquid chromatography (HPLC) is used to achieve this.[64]

HPLC columns have a high sample capacity and are capable of withstanding pressures of 3000 bar to 4000 bar.[64] The technique is fast and efficient and has a high resolution in analytical and preparative systems. The number of theoretical plates of a column is inversely proportional to the size of the particles. HPLC solid phase particle sizes are 5-10μm in diameter and the size distribution is small. The low binding capacity of silica is improved by the attachment of aliphatic hydrocarbons.[65]

Silica is unstable in solutions above pH 8, the bonded phase can bind sample contaminants very strongly, acid chlorides erode the steel in the pumps and tubing, the support phase pores are easily blocked and the high pressures used change the column packing markedly over time.[65] To maximise the resolution of a column, constant recalibration is essential, all solvents and samples must be filtered before use, solvent pH must be monitored, strongly binding contaminants such as aromatic compounds must be removed, a guard column must be attached to remove large particles and the system must be flushed between runs.[65,64]

The separation of small hydrophobic peptides has been found by Imoto and Yamada[70] and Dunlap et al[71] to be optimal on a

reverse phase C_{18} column using a non-eluting 0.1% TFA/water (buffer A) solution with an eluting solution of 60% acetonitrile in 0.1% TFA/water (buffer B). The TFA is added as a counter ion for the peptide. The TFA and acetonitrile are lyophillisable and the water can be removed by freeze drying. This makes TFA the ideal choice of counterion.[64] The column eluant is monitored at 190nm which detects the peptide bond. The solvent system is transparent at this wavelength.[70] The basis for separation on reverse phase systems is hydrophobicity.[70] A gradient is run to determine the percentage of buffer B for elution. This separates the contaminants from the cleavage process and deletion fragments. In the elution of short peptide sequences that gave complete couplings, the largest peak is usually the product. The percentage of acetonitrile at which the major peak elutes, determined from the percentage of buffer B and denoted X, is used for an isocratic run according to the method of Imoto and Yamada.[70] The first isocratic run is at an acetonitrile concentration of X-2%. This is varied slightly to optimise the resolution. The technique can separate diastereoisomers and deletion peptides.[68] An advantage is that the system is then ready to be scaled up to a preparative column.[68]

Preparative HPLC systems are a trade-off between the superior resolution of small low-capacity analytical columns and larger less sensitive preparative columns. The best

arrangement is one where the analytical and preparative column bonded phases and lengths are as similar to each other as possible. The isocratic conditions optimised on the analytical column can then be applied directly to the preparative system. The product can be tested for purity on an analytical HPLC column, by thin layer chromatography, infrared and nuclear magnetic resonance.

2.3 *Nuclear magnetic resonance*

Nuclear magnetic resonance (NMR) measures the absorption of radio frequency electromagnetic radiation by a sample placed in a magnetic field. The result is a plot of the intensity of the absorption against the frequency. Certain nucleii are responsible for the observed absorptions.

All nucleii contain charges and by spinning they give rise to magnetic dipoles or quadrupoles, designated μ. A spin number, I, is calculated by summing the number of protons and neutrons.[72] An even number makes I equal to an integer or 0, an odd number makes I a half integer and when the number of protons and the number of neutrons are both even I is 0. The isotopes: ^{1}H, ^{19}F, ^{13}C and ^{31}P have I=1/2 and a uniform distribution of charge which produces a dipole. For nucleii with I equal to or greater than 1 the charges are not uniformly distributed and they produce a quadrupole moment. The value of I determines the number of orientations a nucleus can assume in a magnetic field and equals 2I+1.[72]

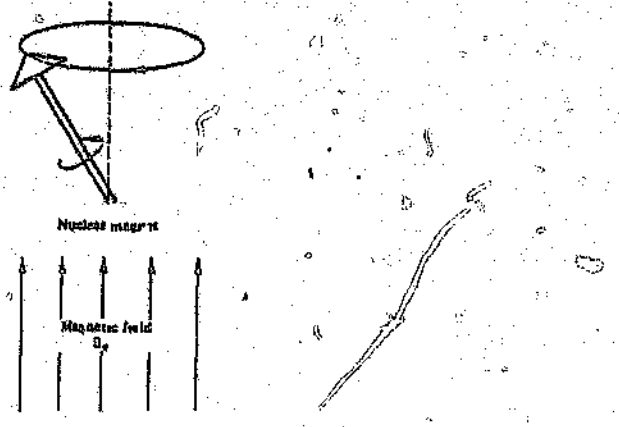


Figure 15 The precession of a nucleus in a magnetic field. A precessing nucleus can absorb radio frequency electromagnetic energy, transmitted at right angles to B_o,[61] provided the frequency of procession equals the radio frequency

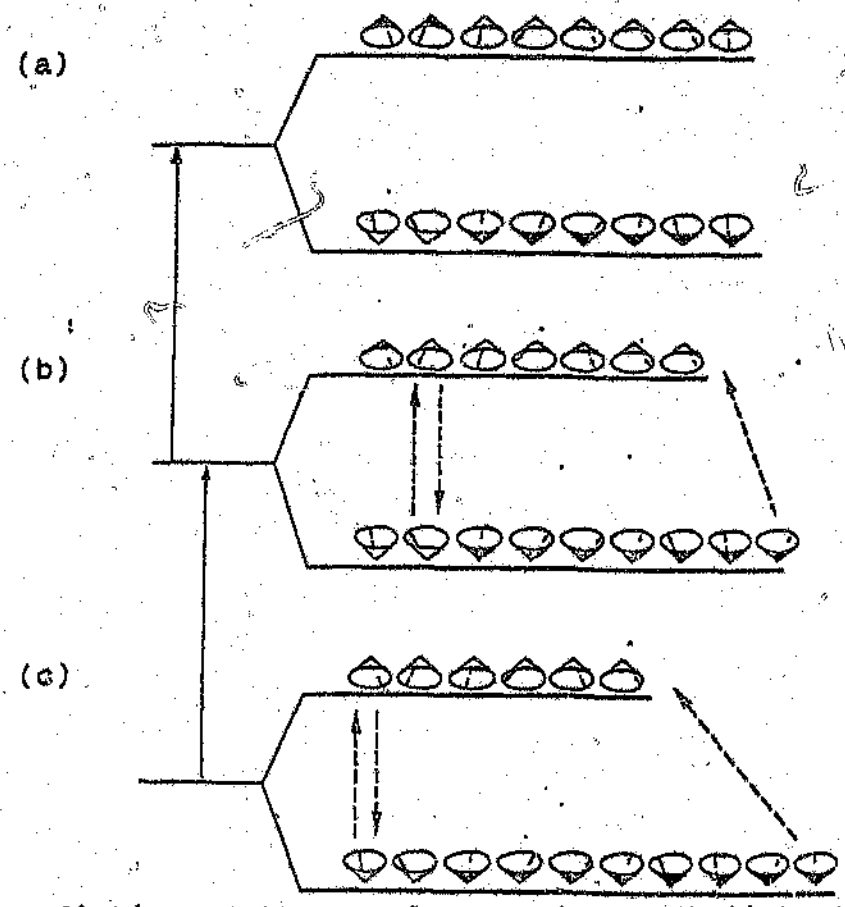


Figure 16 Irradiation of the sample promotes nucleii to the upper energy state. In (c) no more energy can be absorbed. To return nucleii to the lower energy states and therefore to return the system to (b) and (a), energy must be dissipated. Some is transffered to the surroundings (spin lattice relaxation). Nucleii transfer their energy to other nucleii (spin-spin relaxation) and the nucleii swop places as shown by the arrows on the left.[61]

Proton NMR is the most common choice in protein work because of the abundance of the ^1H and its ubiquitous presence in organic chemistry. This discussion will concentrate on this branch of NMR as it pertains to the work presented.

Figure 15 shows that a proton can align its magnetic dipole with or against the applied magnetic field. There is a slight overpopulation of the low energy state. As the magnetic field is increased the energy gap between the +1/2 state and the -1/2 state increases.

In a magnetic field the nucleus precesses at a particular frequency (figure 15). In a NMR machine electromagnetic radiation in the form of radio waves is passed through the sample which is kept in a constant applied magnetic field. The magnetic component of the radio waves is at right angles to the applied field Bo (figure 15). At certain wavelengths, the frequency of oscillation of the associated magnetic field will equal the precessional frequency which produces an absorption maximum.[72] As the sample absorbs energy the population of nucleii in the high energy state will become filled (figure 16). There are two mechanisms for transferring the nucleus from the high to the low energy state. In spin-lattice relaxation the energy is passed to the molecular lattice and the efficiency of the process is proportional to the half life, T1, of the high energy state. The spectral line width is inversely proportional to T1.[72] The line widths in gases and liquids are acceptable but in

crystals and solids, which have long T1 values, the line widths are too narrow. The second energy transfer is between nucleii, called spin-spin relaxation, and its efficiency is also governed by a half-life T2. Solids show a broadening of the bands from this process. Samples are most efficiently analysed in solution as this optimises the T1 and T2 effects. For proton spectroscopy, proton transparent solutions are required so that only a spectrum of the sample is measured. High purity deuterated solvents are used but inevitably they produce strong proton impurity peaks from non-deuterated molecules. Deuterated water or carbon tetrachloride are the most common solvents as they are cheap and relatively non-volatile.

The position of a spectral line is proportional to the frequency at which an absorption maximum has occurred. First therefore a standard must be chosen and generally this is tetramethyl silane. It is inert, magnetically isotropic, volatile, soluble in organic solvents and the protons are strongly shielded. It thus has a sharp peak upfield from the point where most sample protons would show. By convention, samples are measured as a displacement from the standard. A peak 60Hz from the standard peak in a 60MHz field has a chemical shift value of 1 part per million. The chemical shift value is constant regardless of the applied magnetic field[72] and depends upon the extent of shielding or deshielding of the nucleii by the electrons. Two protons

found in identical environments will have the same chemical shift value.

The electron density surrounding a nucleus depends upon its electronegativity and the proximity of other groups.[72] In NMR the presence of the magnetic field induces electron movements that oppose the applied field.[72] A high electron density will reduce the size of magnetic field experienced by a nucleus. This in turn affects the frequency of absorption and the position of peaks. Tertiary and secondary conformations bring different chemical groups into close proximity and the patterns of shielding and deshielding are complex. In a fluctuating random structure the chemical shift values are an average of all the different chemical environments.[73] Any deviation from the average values implies some definite structure. The large number of factors that alter the chemical shift and the ability of high resolution NMR to discriminate the differences, make it an invaluable tool.

Nucleii within three atomic bonds of one another are capable of coupling their spins. This leads to anything from simple first order patterns to complicated second order patterns of interfering groups.[72] The advantage this lends to analysis is that for a known compound the spin-spin coupling pattern can be predicted. This information can then be used to assign peaks in a spectrum. The shifts actually seen in the positions of peaks give information about the chemical

environments of those groups which can be related to conformation.[72]

The nuclear Overhauser effect (NOE) is a means of identifying protons in close proximity to one another brought together by folding.[73] The NOE is observed when protons are within 5Å units of one another.

Two- and three- dimensional NMR techniques have been developed to map protein structure and overcome the problems of one-dimensional NMR. Modern NMR aims to produce X-ray crystallographic-like maps of proteins and this has been accomplished for a few small peptides.[73]

2.4 Ultraviolet and visible light circular dichroism

"Circular dichroism provides information about the average state of secondary structure existing in solution based upon chromophore analysis".[29] Proteins have three ultraviolet chromophores[72]: the peptide bonds, aromatic side chains and disulphide bonds. The transitions of the resonating chromophoric electrons result in optical activity. The carbonyl groups (-C=O) have electrons that vibrate in an asymmetric environment.[64] There are two transitions associated with the -C=O groups. Near 190nm the electrons are lifted from a π molecular orbital to a π^* antibonding orbital and this is designated as π-π^*. Near 220nm the transition is from a non-bonding orbital to a π^* antibonding orbital, designated n-π^*.[64] The transitions are affected by the conformation of the peptide bond. The aromatic and

disulphide bridge chromophores absorb between 280nm and 300nm. The peptides used in this work did not contain aromatic groups or disulphide bonds and a discussion of these chromophores is therefore omitted. Figure 17 shows how left and right circularly polarised light is absorbed differentially by the peptide bond.[64] Circular dichroism (CD) measures the difference between the absorption of the

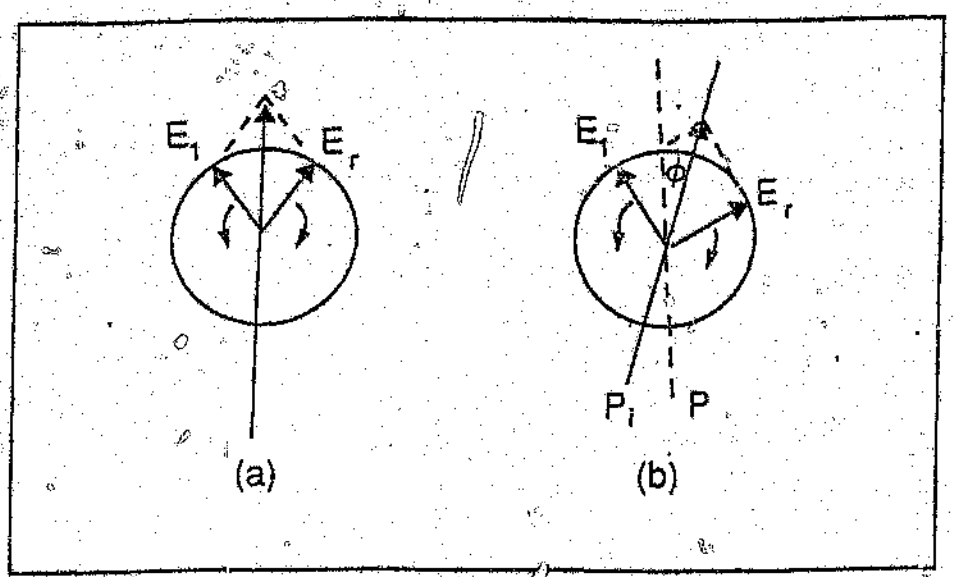


Figure 17 E_l and E_r are the electric vectors of circularly polarised light. In an optically inactive medium (a) E_l and E_r travel at the same velocity and the vector $E_l + E_r$ yields unchanged plane polarised light in the plane P. In an optically active medium containing assymetric molecules (b) the vectors E_l and E_r travel at different velocities. The resulting vector is rotated by ϕ and the light is elliptically polarised.[74]

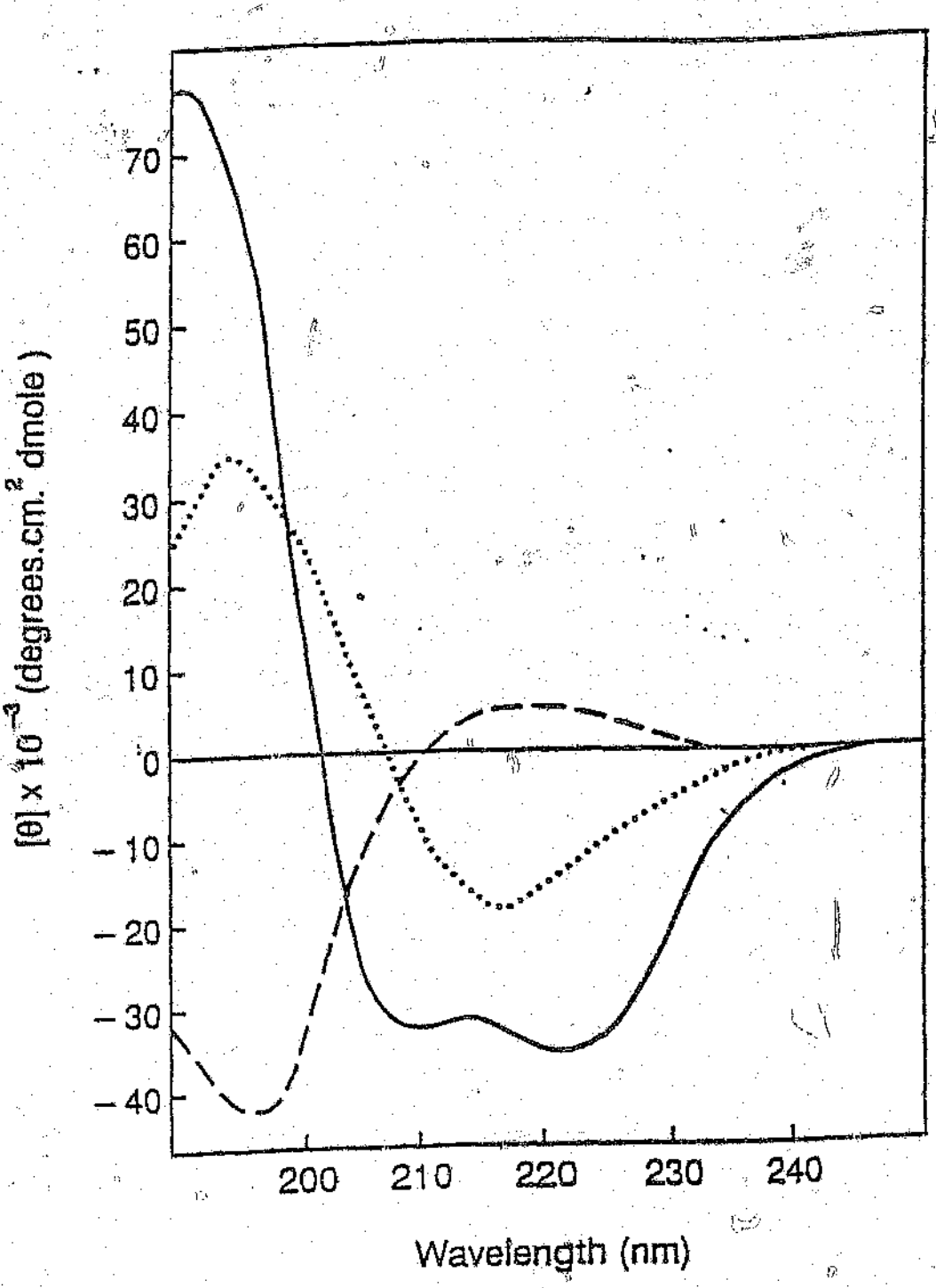


Figure 18 The circular dichroism of α-helix (———), β (·····) and random conformations (– – –).[76]

left and right components of circularly polarised. This effect causes elliptical polarisation.[74,75] CD measurements are converted to a standard value known as the molar ellipticity.[74] CD bands are sensitive to the peptide bond's structural environment and are used to measure the conformation of peptides in solution. CD measurements are therefore an average of all the secondary structures present in a solution.[29] Ellipticities can be positive or negative as the degree of absorbance of the left can be greater or less than that of the right. The change in ellipticity through a positive maximum at a lower wavelength than the next negative minimum is defined as a positive cotton effect.[74] The change through a negative minimum in ellipticity at a lower wavelength than the next positive maximum is defined as a negative cotton effect.[74]
Different secondary structures have been studied with CD.[74,75] The studies were based upon the ability of poly-L-lysine to undergo conformational changes in solutions of different pH values.[76] Greenfield and Fasman generated a set of theoretical CD curves for α-helix, β-sheet and random coil.[76] The theoretical curves fit the experimental curves of poly-L-lysine well. Greenfield and Fasman also computed curves for peptide solutions where there was no one predominant conformation.[76] These curves are used as references for identifying CD spectra.

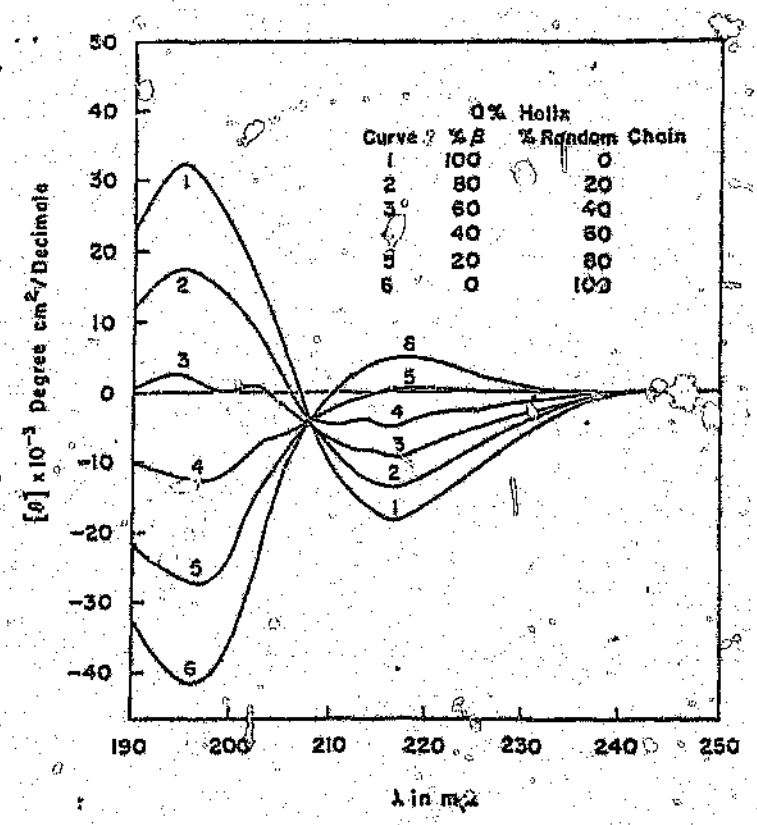


Figure 19 The calculated CD curves of "statistical coil" containing different proportions of β-structure and random coil.[76]

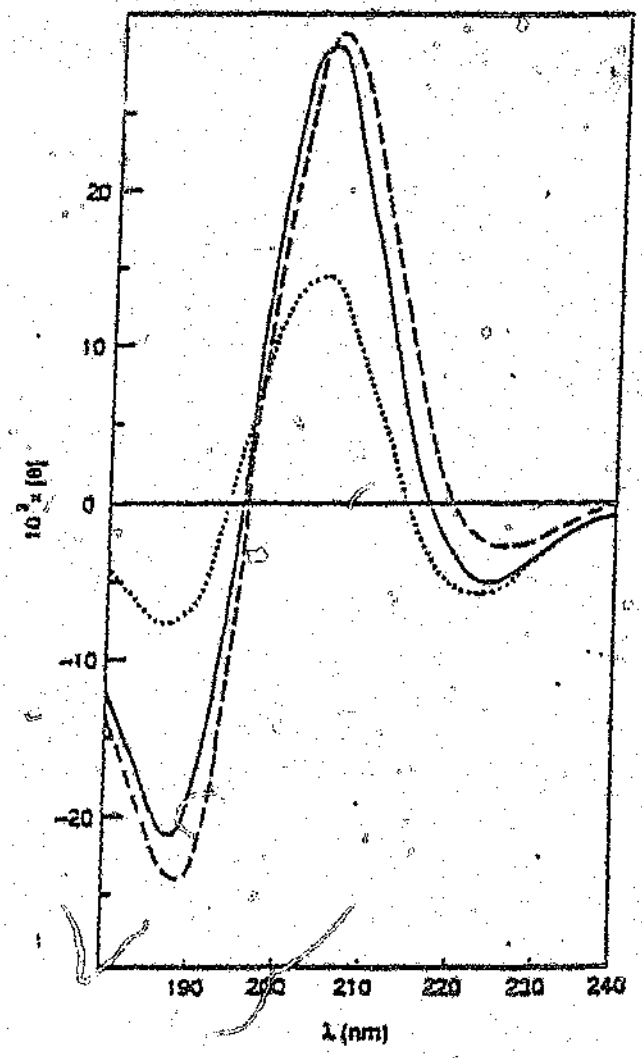


Figure 20 The calculated CD curves for β-turns,[18,82] Type I Type II (– – –) and Type III (·····)

Ellipticity values reported for particular conformations

Table 1

Conformation	Wavelength nm	Ellipticity deg.cm²/dmole	Reference
α-helix	191	76000	75
	222	−36000	78
	208	−33000	78
β-structure, possibly antiparallel	191	25000	76
	191	77000	78
	215	−17900	76
	217	−18000	78
Random coil	197	−42000	78
	217	4600	78
	238	−150	76,78

The characteristic α-helix CD curve (figure 18) has a split π-π^* transition with a positive peak at 191nm and negative minimum at 205nm.[64,75,42] The minimum at 220nm is due to the n-π^* transition.[64,78] The ellipticity values and the exact positions of Cotton effects will be influenced by the chain length and solvent.[78] Despite the variations of the ellipticity values (table 1), the positions of the maximum and minima of the reference curve are a reliable means of identifying CD curves of α-helices.

ß-structures, unlike the α-helix, are not restricted to one or even a few states. They vary with respect to their chain orientations, length, sheet chirality and the number of sheets.[79,80] This variation means that the optical properties of ß-structures are not constant. The use of a reference curve is therefore limited in its application.[79]

Poly-L-lysine, used to test the reference ß-structure curve (figure 18), forms an antiparallel sheet structure.[76] Table 1 details the values of ellipticities seen in the CD of ß-structures.

The random coil assumed by poly-L-lysine is attained at a pH where the side chains of the amino acids are charged and the peptide is rigid.[76] Poly-L-lysine in this state produces a CD spectrum that fits the computed random coil curve[76] (figure 18). "Statistical coil" describes CD curves that result from an average of several structures in solution.[6] The assignment of CD spectra to "statistical coil" or random

coil because the results do not fit the recognised secondary structural categories must be done with care. The danger is that by designating structures as random coil or "statistical coil" other unclassified structures will be missed. Greenfield and Fasman have published a series of curves made up of different relative proportions of the α-helix, ß-structure and random coil curves (figure 19).[76] These curves can be used to identify "statistical coil". The less well described and documented structures such as hydrophobic pockets or hairpin bends have no reference CD curves for their identification. The structures themselves are not regular and a reference CD curve is therefore not possible; each situation must be treated separately. This is illustrated by an example in which an α-helix-like CD curve was shown to be due to a ß-hairpin structure joining two antiparallel ß-strands.[77] CD results therefore require confirmation from other techniques such as NMR and IR. CD spectra for ß-turns have been measured and calculated (figure 20).[62]

2.5 Infrared spectroscopy

The most useful infrared frequencies for organic chemistry are between 666cm^{-1} and 4000cm^{-1}. The term frequency is applied to the unit cm^{-1} as it is the reciprocal of the wavelength and therefore directly proportional to the frequency.[78] Infrared radiation causes two effects, the vibration of particular groups within a molecule and the

rotation of molecules in space. In the solid phase only vibration occurs. In solution molecules can also rotate and vibrate. The energies involved are quantised and produce a spectrum. Vibration refers to either the stretching or bending of bonds.[63] Stretching is a rhythmical movement within a bond that changes the interatomic distances. Bending changes the angle between two bonds centred around a common atom or is the movement of a group of atoms within a molecule with respect to that molecule. In infrared spectroscopy only those stretching modes that cause a change in the dipole moment of the molecule are observed. Stretching causes an oscillating electric field that can interact with infrared radiation. The frequency of absorption is proportional to the relative masses of the atoms involved, the bond force constants and the geometry of the atoms.[72] Given that each atom has three degrees of freedom, the number of vibrations and rotations in a molecule are vast.[72] The effective number is reduced because many are equivalent or do not cause infrared absorption. The overlapping and coupling of vibrations through bonds produces a spectrum containing bands rather than sharp peaks.[73]

For the purposes of calculation it is assumed that the stretching and bending of a bonds is analagous to the stretching and bending of a spring with a weight at each end. By using this model and applying Hooke's law, the

energies of vibrations within a bond can be calculated. From this the absorbance bands in infrared spectra can be assigned to specific chemical groups. The complexity of a spectrum makes it unique for a particular compound or chemical group[73] and infrared spectra are used for identification.[72]

Infrared transitions are sensitive to the environment in which chromophores are placed[72] and conformational changes cause shifts in the position of infrared bands. The infrared spectra of proteins and peptides are dominated by the peptide bond amide group. The bands that they produce are unparalleled in their effectiveness as detectors of hydrogen bonds and their positions are affected by the conformational environment of the chromophores.[44,72]

The assignment of the amide bands to particular chromophores and secondary structures in proteins relies upon evidence from N-substituted amino acids spectra,[44,65] the cross correlation of NMR determined structures and the deuteration of groups containing exchangeable protons.[72] The amide chromophores are named as follows: A(1600-1500cm^{-1}), B(3100cm^{-1}), I(1700-1600cm^{-1}), II(1600-1500cm^{-1}), II (1300-1200cm^{-1}), IV (800-600cm^{-1}), V(750-600cm^{-1}) and VI(600cm^{-1}). Tables 2, 3 and 4 summarise the secondary structures assigned to positions of these bands.

The amide I and V bands are the most conformationally sensitive.[66] The amide V band is used to discriminate

α-helix from random coil as both appear at the same wavelengths in the amide I band. A conclusion about a structure requires the overall evidence from several of the bands and correlation with other techniques such as NMR and CD. As well as the amide bands there are chromophores associated with the side chains of the amino acids (table 5).

Amide A and B frequencies characteristic of various secondary structures

Table 2

Amide band		Secondary structural conformation assignment	Reference numbers
A 1/cm	B 1/cm		
3450	3100	non-hydrogen bonded -NH groups	83
3300-3400		hydrogen bonded -NH groups	44,83
3290-3300		-NH groups in alpha helices	83
3280-3300		intermolecularly hydrogen bonded B-structures. The band frequency decreases with increasing hydrogen bond strength.	83,49,59
3360		B-turns with 4-1 hydrogen bonds	59
3370		random coil	83

Amide I and II frequencies characteristic of various secondary structures

Table 3

Amide band		Secondary structural conformation assignment	Reference numbers
I /cm	II /cm		
1680–1700	1500 or greater	non–hydrogen bonded and weakly hydrogen bonded –C=O groups	83
1600–1650	1590	hydrogen bonded –C=O groups	83
1687p 1655m 1638s		ß–type I turns	59
1660 1602s	1550–1570	ß–turns	59
1650–1663	1540–1550p 1516s	α–helices	86,83,77, 79
1625–1635	1520–1526	intermolecularly hydrogen bonded ß–conformations	83,49,59, 60,
	1540–1550p 1530s	ß–pleated sheet	79
1663–1642	1553–1586	antiparallel ß–sheet	49
1635p 1690s	1530	parallel ß–sheet	86,49,60, 79
1660–1665	1535	random coil	59,60,77, 79

Amide III and V frequencies characteristic of various secondary structures

Table 4

Amide band		Secondary structural conformation assignment	Reference numbers
III ν/cm	V ν/cm		
1290–1330		β–turns	62
1230–1663	610–620	α–helices	77,79
1210–1250	700	intermolecularly hydrogen bonded β–conformations	62
	705–720	β–conformations	79
	712	parallel β–sheet	87
1240–	650–	random coil	77

Non-amide infrared chromophores found in peptides

Table 5

Frequency /cm	Chromophores	Reference numbers
1450	alanyl methyl stretch	62
1360–1390 1663	–NH in plane bending and stretch mode of –CH groups	85
1328	alpha proton bending	85
1140–1170	alanyl methyl rock	85
1130–1150	backbone and side chain stretch	85
1050–940	–CH rock	85
800	rocking modes	85

2.6 Deuterated amino acids as structural analysis tools

Infrared and NMR are two major analytical techniques in peptide conformation studies. Much of the protein NMR information comes from proton spectra. Deuterium is not paramagnetic and therefore does not produce a signal.[78] The presence of deuterium in amino acid residues in known positions would affect the multiplicities seen in a spectrum. A study of the difference between the normal a the deuterated spectrum is a means of assigning peaks in a protein. At points of close contact a deuterium would change an NOE result.[73] Another use of deuterated amino acids in peptides is in infrared work. The positions of absorption bands are affected by changes in the force constant of chromophoric bonds.[72] The deuteration of chromophores will slightly affect the position of a band by changing the bond force constant. This is a way of identifying infrared bands.[87] Deuterated amino acids therefore provide a general tool for structural analyses.

3. EXPERIMENTAL WORK

3.1 Materials

The tertiary-butyloxycarbonyl L-amino acids, trifluoracetic acid, S-benzyl-L-cysteine and 1-hydroxybenztriazole were supplied by Fluka Inc. Chloromethylated polystyrene 1% divinyl benzene copolymer resin (1% crosslinked containing 0.9meq Cl^{-1}/g), 2,2,2 trifluoroethanol and N,N-dicyclohexylcarbodiimide were purchased from Sigma Chemical Co. Isopropanol was purchased from BDH Laboratory Reagents Inc. Technical grade dichloromethane, analytical grade dimethylformamide, acetonitrile and triethylamine were purchased from Saarchem Pty Ltd. The DCM was distilled over potassium carbonate. 4Å molecular sieve, was purchased from Merck Chemical CO. 45 µm cellulose acetate and teflon filters for HPLC were supplied by Millipore Corp. Camphorsulfonic acid was supplied by Dr. Fritz Carlsson, CSIR, Pretoria. The highest commercial grades of solvents were used for the spectroscopy: methanol was supplied by BDH Laboratory Reagents inc, ethanol, propanol and butanol were supplied by Merck Chemical Co. The spectroscopic solvents were dried with desiccated Na_2SO_4. The ethanol was refluxed with magnesium and iodine according to the method of Vogel.[92] Thin layer plates (TLC) were purchased from Merck Chemical Co. The reverse phase C_{18} HPLC columns were from Brownlee Labs.

3.2 Solid phase synthesis of L-methionine-L-alanine$_5$

The SFPS method of Marki et al[63] was used for the synthesis of L-methionine-L-alanine$_5$ (met(ala)$_5$). This method is a variation of the Merrifield system[65,64] that uses t-Boc amino acids. 2.76g of chloromethyl resin, a 2.5 times excess of t-Boc-L-alanine dissolved in DMF and a 5 times excess of KF were placed into a round bottomed flask, stirred for four hours and then left overnight at 80°C. After cooling the resin was washed in:

1) 80ml DMF
2) 50ml 50% DMF/distilled water
3) 120ml distilled water
4) 30ml 50% methanol/water
5) methanol 3 times
6) DCM 3 times
7) parts 5 and 6 were repeated three times.
8) The resin was dried under vacuum and stored and the extent of the esterification was monitored by the Gisin test[67] and found to be 6.6meq/g. This was a 66% substitution of the 1meq Cl^{-1}/g resin and the coupling could proceed.

3.2.1 The Gisin Test

The test used was a modification of the Gisin test.[67] It differed from the original in that methanol washes to shrink the resin were included.

1) A sample of the resin was removed and weighed, typically 10 to 20mg were used. The resin was placed into a sintered glass funnel and swollen with DCM.

2) The resin was neutralised with 5% diisopropylethylamine in DCM for 1 minute and this step was repeated. The resin was then washed with DCM for 5 minutes and rinsed three times in methanol.

3) The resin was reacted with a 0.1M picric acid solution in DCM for 1 minute. The picric acid treatment was repeated. The excess picric acid was washed away by rinsing alternatively with DCM and methanol. This was continued until the rinses were clear. These washes were discarded.

4) The reacted picrate was eluted by washing the resin alternatively with diisopropylethylamine solution (step 2) and methanol. The washes were pooled and the volume recorded, normally 5 to 10ml were collected. A 1M diisopropylethylamine-picrate solution has an extinction coefficient of 0.145 at 358nm and this was used to calculate number of esterified sites on the resin. Each esterified site on the resin reacts with one picrate ion and therefore the stoichiometry of the reaction is 1:1. DCM has a high absorbance at 358nm and the readings were done with dilutuions of the picrate solution in 95% ethanol such that the final concentration of DCM did not exceed 20%.

Table 6 shows the amounts of reagents used and the times for each coupling. For DCC couplings a 2.5 times molar excesses

of t-Boc amino acid was dissolved in the minimum volume of DMF. The amino acid was added to the resin with a 2.5 times molar DCC excess (added as a 1M solution in DCM) and enough DCM to slurry the resin. The reaction mixture was shaken for four hours. At the end of this time the extent of reaction was monitored with the Kaiser test.[68] A 1-hydroxybenztriazole coupling was required (table 6) for which a 5 times excess of hydroxybenztriazole (HOBT) was used. The deprotection method involved the following washes:

1) DCM twice

2) methanol 3 times

3) parts 1 and 2 were repeated

4) DCM 3 times

5) 60% TFA with 5% 1,2, ethanedithiol in DCM for 5 minutes

6) The TFA solution was flushed out and replaced with fresh TFA solution. The reaction vessel was shaken for half an hour.

7) 5% 1,2,ethanedithiol in isopropanol twice

8) The resin was neutralised with 10% triethylamine in DCM

9) methanol 3 times

10) DCM 3 times

11) parts 9 and 10 were repeated.

The finished peptide was deprotected and dried under vacuum. It was cleaved from the resin with HF (see appendix for method).[69] The HF cleaved peptide and resin were washed with

Coupling Step	Mass of t-Boc amino acid g	Volume of 1M DCC ml	Mass of HOBT g	Coupling time hrs	Chloranil test colour	Reaction product
1	0.86	4.6		4	yellow	Boc-2(ala)-resin
2	0.86	4.6		4	yellow	Boc-3(ala)-resin
3	0.86	4.6		4	yellow	Boc-4(ala)-resin
4	0.85		1.26	8	yellow	Boc-5(ala)-resin
5	1.14	4.6		4	yellow	Boc-6(ala)-resin
6	1.16		1.17	8	yellow	Boc-5(ala)met-resin

Table 6 The synthesis of t-Boc-L-met(L-ala)$_5$ by solid phase peptide synthesis. HOBT and t-Boc amino acids were added in a 5 times molar excess over the number of esterified sites, DCC was added in a 2.5 times excess.

Peptide	1) mmoles of crude peptide	2) mmoles of purified peptide	Percentage yield
Leu(ala)	0.86	0.36	41.8
(Ala) leu	0.73	0.13	18.3
Ile(ala)	0.94	0.18	19.2
(Ala) Ile	0.57	0.11	18.9

Key
1) HF.H.Peptide.OH
2) TFA.H.Peptide.OH

Table 7 Yields of peptides based upon the weight of the freeze dried product after purification with HPLC.

ethyl acetate in a sintered glass funnel to remove the anisole. The cleaned peptide was then dissolved and washed through the sinter with TFA and several acetonitrile washes. The solution was rotavaporated and freeze-dried. 0.61g of the dry crude peptide was collected. The maximum yield possible was 0.95g and therefore the yield was 64%. The peptide was ready to be purified by HPLC.

3.3 Purification of peptides by HPLC

A Spectra Physics HPLC solvent delivery system with a reverse phase C_{18} column were used for the HPLC. The column eluant was monitored at 190nm by a Biorad Uvichord. A non-eluting solution of 0.1% TFA in water and an eluting solvent of 60% acetonitrile in 0.1% TFA was used according to the method of Imoto and Yamada.[70]

The peptides purified were: HF.H.met$(ala)_5$.OH, HF.H.leu$(ala)_5$.OH, HF.H.$(ala)_5$leu, HF.H.ile$(ala)_5$.OH and HF.H.$(ala)_5$ile.OH. The leucine and isoleucine peptides had been previously synthesised[69]. The peptide yields after purification are given in table 1. The purity of the peptides was assessed from the collective results of the HPLC, NMR and infrared.

3.4 Nuclear magnetic resonance Spectra

One dimensional proton NMR spectra of the TFA.H.met$(ala)_5$.OH, TFA.H.leu$(ala)_5$.OH, TFA.H.$(ala)_5$leu.OH, TFA.H.ile$(ala)_5$.OH and TFA.H.$(ala)_5$ile.OH were recorded on a Bruker AC200 spectrometer at 200.13 MHz in D_2O. Assignments

Concentrations of peptides used in the CD spectroscopy

Table 8

Solutions	Peptide concentration in mg/ml			
	TFA.H.Leu-$(ala)_5$.OH	TFA.H.$(ala)_5$-leu.OH	TFA.H.Ile $(ala)_5$.OH	TFA.H.$(ala)_5$-ile.OH
methanol	0.57	0.60	0.62	0.56
ethanol	0.62	0.59	0.55	0.53
propanol	0.62	0.64	0.66	0.65
butanol	0.52	0.55	0.53	0.55
25% TFE/ethanol	0.68	0.56	0.67	0.51
50% TFE/ethanol	0.55	0.64	0.63	0.60
75% TFE/ethanol	0.60	0.56	0.63	0.60
TFE	0.53	0.53	0.61	0.60
Aqueous solutions				
water	0.60	0.54	0.53	0.50
0.1M phosphate bu				
ph 4.9	0.62	0.57	0.54	0.62
pH 6.2	0.59	0.58	0.58	0.55
pH 8.2	0.60	0.57	0.65	0.62
pH 9.1	0.58	0.61	0.52	0.56
0.1M sod.carb.	0.63	0.59	0.57	0.62
pH 9.2	0.55	0.58	0.50	0.62
pH 10.0	0.58	0.55	0.61	0.52
0.1M sod.acetate	0.59	0.61	0.63	0.54
pH 5.0	0.70	0.53	0.65	0.60

Concentrations of Peptides Used in the Liquid Phase Infrared Spectroscop

Table 9

Solutions	Peptide concentration in mg/ml			
	TFA.H.Leu-$(ala)_5$.OH	TFA.H.$(ala)_5$-leu.OH	TFA.H.Ile $(ala)_5$.OH	TFA.H.$(ala)_5$-Ile.OH
methanol	0.67	0.65	0.75	0.47
ethanol	0.61	0.49	0.65	0.74
propanol	0.78	0.56	0.55	0.56
butanol	0.80	0.49	0.71	0.52
25% TFE/ethanol	0.60	0.46	0.61	0.67
50% TFE/ethanol	0.78	0.58	0.84	0.65
75% TFE/ethanol	0.72	0.61	0.74	0.54
TFE	0.67	0.57	0.58	0.71

were made using the chemical shift values and spin coupling systems of the amino acids published by Wuttrich.[73] A further discussion of the NMR analysis is given in the results and discussion section.

3.5 Ultraviolet circular dichroism spectroscopy

The CD spectra of TFA.H.leu(ala)$_5$.OH, TFA.H.(ala)$_5$leu.OH, TFA.H.ile(ala)$_5$.OH and TFA.H.(ala)$_5$ile.OH between 190nm and 260nm were recorded on a Jasco Model J20 Recording Spectropolarimeter. The spectra were run in water, 0.1M buffer solutions at pH values between 5 and 10, methanol, propanol, butanol and a series of TFE solutions in ethanol. The peptide concentrations for each solution are given in table 4. The readings taken were transferred to a spreadsheet and the molar ellipticities calculated. The plots of ellipticity against wavelength were produced by running a 5 point smoothing program according to the method of Savitzky and Golay.[90]

3.6 Infrared spectroscopy

The infrared spectra were recorded on a Jasco A-202 Spectrometer. Solid phase KBr disc spectra were recorded for TFA.H.leu(ala)$_5$.OH, TFA.H.(ala)$_5$leu.OH, TFA.H.ile(ala)$_5$.OH and TFA.H.(ala)$_5$ile.OH. KBr and KCl cells were used for the liquid phase spectra. The liquid phases were measured in methanol, ethanol, propanol, butanol and a series of TFE/ethanol solutions. The peptide concentrations of each solution are shown in table 9.

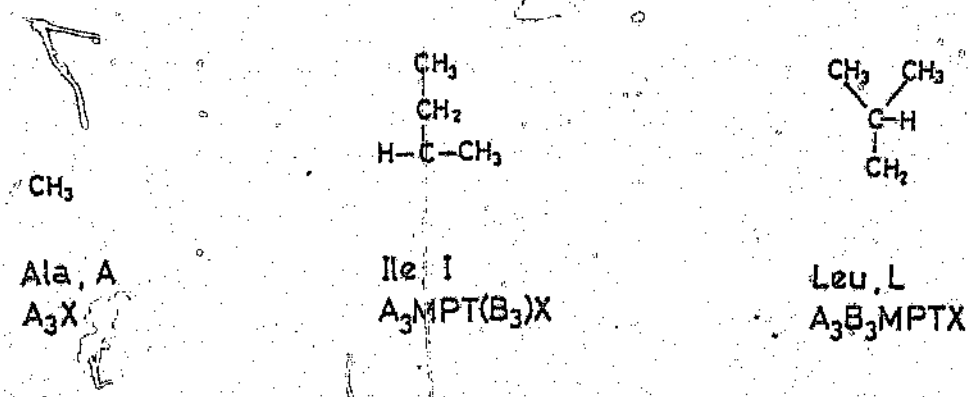


Figure 21 The spin systems of the non-labile protons in L-alanine, L-leucine and L-isoleucine.[73]

3.7 The Raney nickel reduction of cysteine to alanine

This experiment aimed to reduce cysteine with deuterium to produce a monodeuterated alanine. This would label any particular alanine in the hexapeptide and permit its identification in the NMR spectrum.

A desulphurisation procedure for phthaloyl amino acids was published by Balenovic and Fles.[91] The technique had to be adapted for amino acids and t-Boc amino acids.

The Raney Nickel was made according to the method published by Vogel.[92] It was decided to desulphurise S-benzyl-L-cysteine according to the method of Balenovic and Fles.[91] 0.004 moles of the cysteine derivative were dissolved in 100ml of a refluxing medium and refluxed for four hours with 5g of Raney Nickel. Initially, and according to the published method, 100ml of absolute ethanol were used. It was found however that the best results were obtained with dilute ammonia solution in water. The reflux mixture was left at room temperature overnight. It was then filtered and washed with 50cm^3 of absolute ethanol. The filtrate and washes were mixed, rotavaporated and freeze-dried. The reaction products was analysed by thin layer chromatography (TLC) using a 5.5/3/1.5 n-butanol/acetic acid/water solvent system.[93] The amino acids were detected by spraying the plates with ninhydrin and then heating them in an oven at 50°C. Rf values of standards and the reaction products were recorded.

4. RESULTS AND DISCUSSION

4.1 The analytical HPLC of the leucine/alanine and isoleucine/alanine pepitdes showed a single peak for each one. The infra red and NMR results showed there to be no sign of impurities in the peptides produced. The purification of each peptide was therefore sucessful.

4.2 Nuclear magnetic resonance results

The D_2O ^1H NMR spectra of TFA.H.leu(ala)$_5$.OH, TFA.H.(ala)$_5$leu.OH, TFA.H.ile(ala)$_5$.OH and TFA.H.(ala)$_5$ile.OH (tables 10, 11, 12, 13) are evidence for the purity of the peptides as there are no impurity peaks and the peaks correspond to those expected.[72] The nomenclature for the side chains of leucine, isoleucine and alanine is given in figure [21].[73]

The chemical shifts of the peptide protons (tables 10, 11, 12, 13) were assigned using the values given by Wuttrich.[73] Where complex order multiplicities occured in the spectra such as the isoleucine $^1H_\beta$ in table 12, the pattern was analysed and the chemical shifts were given as the average of the two central peaks.[72] This was not possible for some multiplicities, for example the isoleucine $^1H_\alpha$, where the pattern was not discernible from the spectrum. In these cases the upper and lower limits of the combined peak representing the multiplicity was measured and is given as the chemical shift.

Table 10 TFA.H.Leu(ala)$_5$.OH NMR Results. For the published random coil chemical shift values see reference 73.

Chemical Shift ppm	Published random coil values ppm	Shielding deshielding ppm	Assignment	Multiplicity Description	Protons
4.37	4.35	-0.02	Ala alpha-H	Quartet	5
4.34	4.35	0.01	Ala alpha-H	Quartet	
4.28	4.35	0.07	Ala alpha-H	Quartet	
4.26	4.35	0.09	Ala alpha-H	Quartet	
3.99	4.38	0.39	Leu alpha-H	Doublet	0.5
	4.38	4.38	Leu alpha-H	Doublets	0.5
1.6-1.8	1.65(P,T)	-0.15 max	Leu M, P, T	Broad Complex	3
	1.64(M)		Protons	Pattern	
1.42	1.39	-0.03	Ala -CH	Doublet	3
1.37	1.39	0.02	Ala -CH	Doublet	6
1.38	1.39	0.01	Ala -CH	Doublet	6
0.96	0.94	-0.02	Leu B protons	Doublet	3
0.93	0.90	-0.03	Leu A protons	Doublet	3

Table 11 TFA.H.(ala)$_5$Leu.OH NMR Results. For the published random coil chemical shift values see reference 73.

Chemical Shift ppm	Published random coil values ppm	Shielding deshielding ppm	Assignment	Multiplicity Description	Protons
4.34	4.35	0.01	Ala alpha-H	Quartet	4
4.32	4.35	0.03	Ala alpha-H	Quartet	
4.28	4.35	0.07	Leu alpha-H	Doublet of	1
				Doublets	
4.07	4.38	0.31	Ala alpha-H	Quartet	1
1.6-1.8	1.65(P,T)	-0.15 max	Leu M, P, T	Broad complex	3
	1.65(M)		Protons	pattern	
1.52	1.39	-0.13	Ala -CH	Doublet	3
1.38	1.39	0.01	Ala -CH	Doublet	6
1.37	1.39	0.02	Ala -CH	Doublet	6
0.92	0.94	0.02	Leu B protons	Doublet	3
0.87	0.90	0.03	Leu A protons	Doublet	3

Table 12 TFA.H.Ile(ala)$_5$.CH NMR Results. For the published random coil chemical shift values see reference 73.

Chemical Shift ppm	Published Random coil shift value ppm	Deshielding/ shielding ppm	Assignment	Multiplicity Description	number protons
4.40	4.35	-0.05	Ala -H	Quartet	5
4.29	4.35	0.06	Ala -H	Quartet	
4.27	4.35	0.08	Ala -H	Quartet	
3.84	4.23	0.39	Ile -H	1st order	1
				Doublet	
1.95	1.90	-0.05	Ile T	Complex Band	1
				multiplicity 16	
1.50	1.19	-0.31	Ile P	Multiplicity 16	1
1.42	1.39	-0.03	Ala -CH	Doublet	9
1.38	1.39	0.01	Ala -CH	Doublet	3
1.37	1.39	0.02	Ala -CH	Doublet	3
1.3-1.1	1.48	0.47-0.18	Ile M	multiplicity 16	1
0.95	0.89	-0.06	Ile B	Doublet	3
		0			
0.94	0.95	0.01	Ile A	Doubled doublet	3

Table 13 TFA.H.(ala)$_n$Ile.OH NMR Results. For the published random coil chemical shift values see reference 73.

Chemical Shift ppm	Published Random coil shift value ppm	Deshielding/ shielding ppm	Assignment	Multiplicity Description	number protons
4.29	4.35	0.06	Ala -H	Quartet	2
4.27	4.35	0.08	Ala -H	Quartet	2
4.04	4.35	0.31	Ala -H	Quartet	1
		0			
4.28	4.23	-0.05	Ile -H	Doublet	1
		0			
1.9	1.9	-1.9	Ile-T	Complex Band	1
		0		multiplicity 16	
		0			
1.6-1.7	1.19	0.41-0.51	Ile P	Multiplicity 16	1
				Complex Band	
1.49	1.39	-0.1	Ala -CH	Doublet	3
1.35	1.39	0.04	Ala -CH	Doublet	12
1.26	1.48	0.22	Ile M	multiplicity 16	1
				8 Peaks visible	
0.90	0.95	0.05	Ile B	Doublet seen	3
0.87	0.89	0.02	Ile-A protons	Double	3
				Doublet	

Generally the chemical shift values for the protons in the four peptide NMR spectra are close to the random coil values (tables 10, 11, 12, 13).81 This suggests that those protons are not incorporated into any secondary structures. There are however notable exceptions to this trend. In the spectra of the C-terminally substituted peptides (TFA.H.$(ala)_5$leu.OH and TFA.H.$(ala)_5$ile.OH) an alanine $^1H\alpha$ is deshielded by 0.31ppm (tables 11 and 13). In the spectra of the N-terminal substituted peptides, TFA.H.leu$(ala)_5$.OH and TFA.H.ile$(ala)_5$.OH, the isoleucine and leucine ^1H α are both shielded by 0.39ppm (tables 10 and 12). Therefore the shielding or deshielding effects are associated with the position of the substitution of the guest amino acid. The shielding or deshielding could be caused by the formation of structure implying that the position of the guest affects the structure formed. The shifts are however small and may not be significant. Other shielding effects include the shielding of the isoleucine M, P and T protons in TFA.H.$(ala)_5$ile.OH and the M proton in TFA.H.ile$(ala)_5$.OH and are due to the formation of some structure.

In the NMR spectrum of TFA.H.met$(ala)_5$.OH two peaks, at 3.7ppm and 3.?ppm, were evidence for impurities in the sample. Each peak, as measured from the integration, represented one proton and they were therfore due to significant contamination of the sample. There are a great

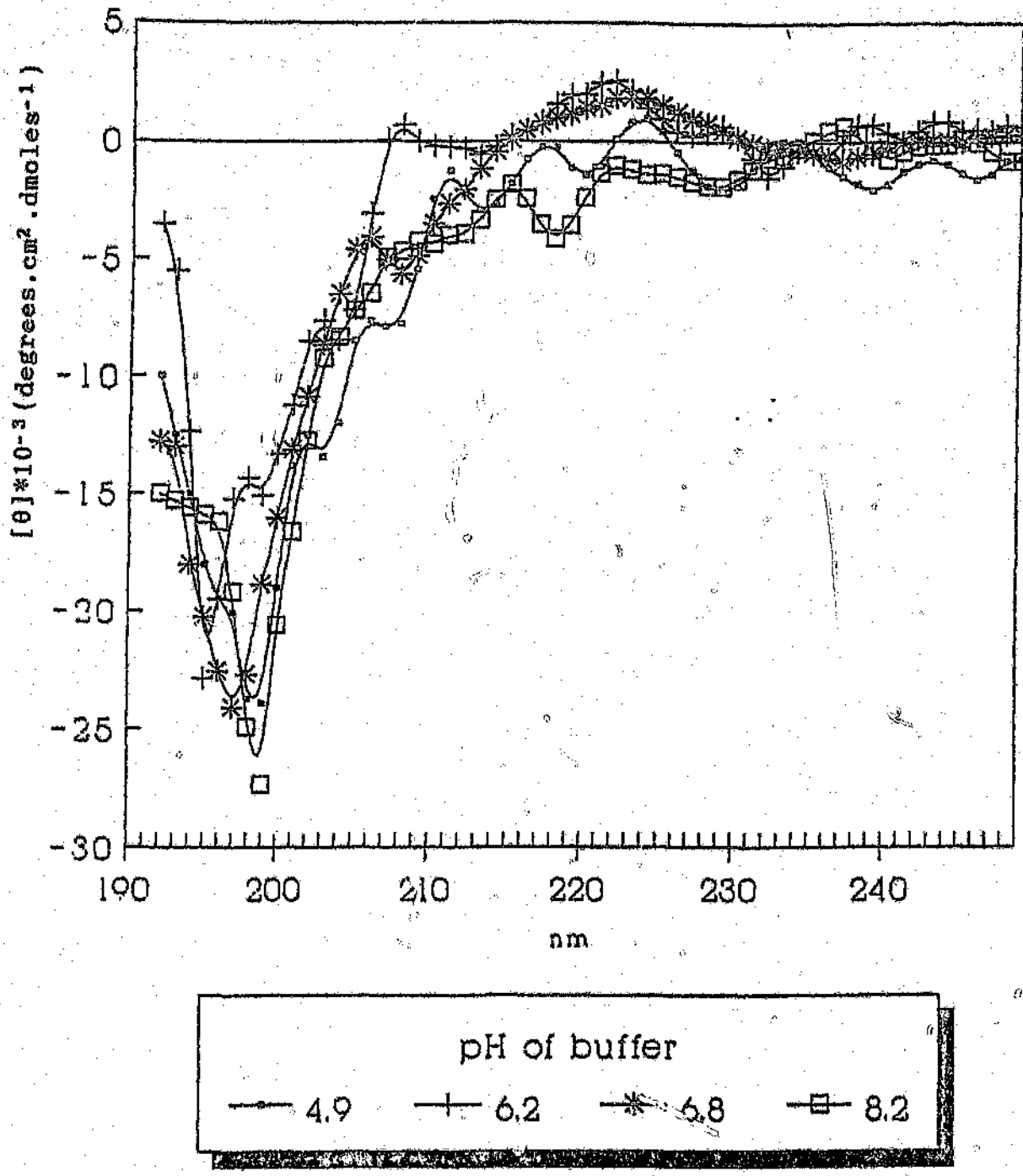


Figure 22 TFA.H.leu(ala)5.OH in 0.1M potassium phosphate buffers of varying pH values.

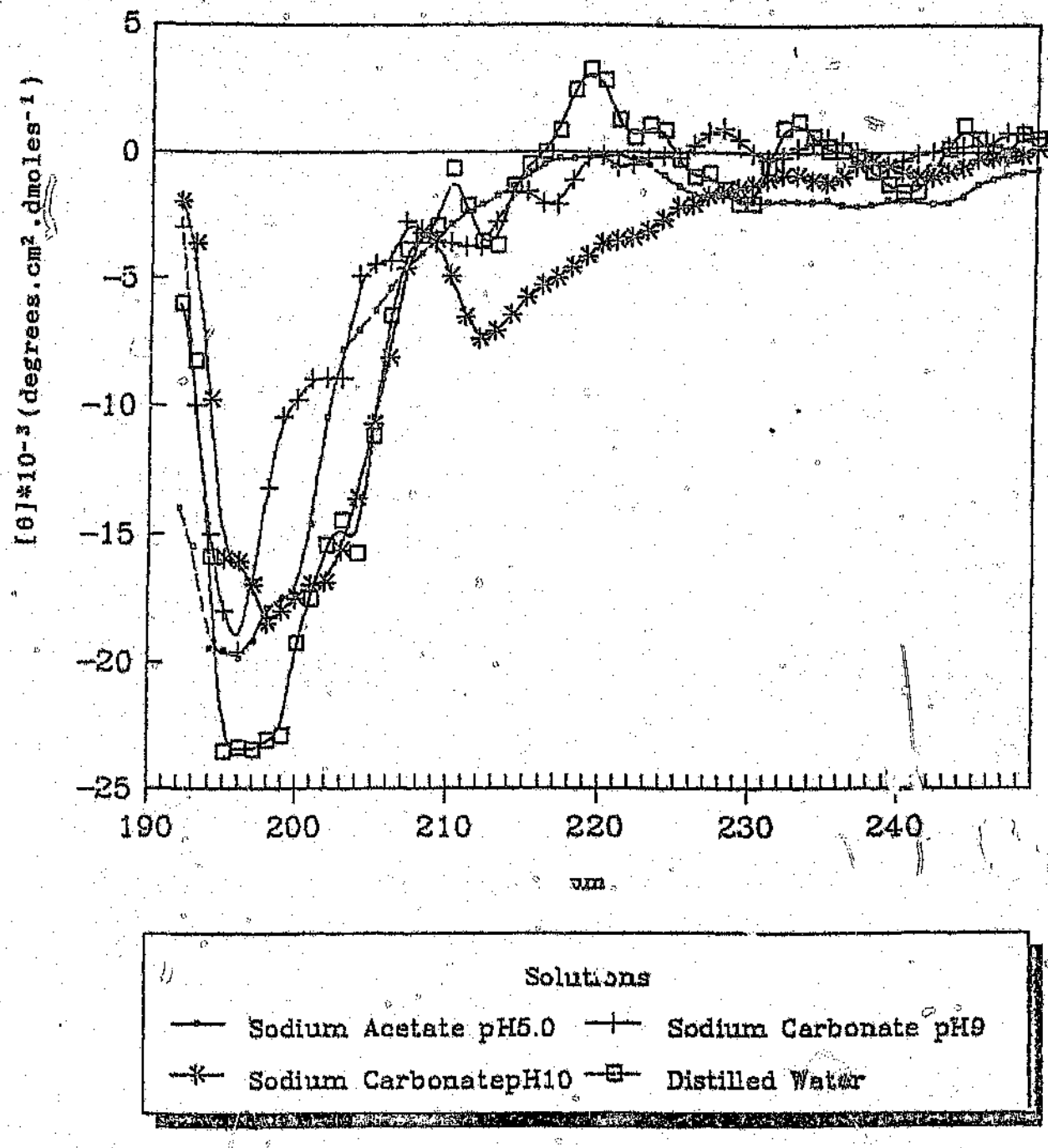


Figure 23 TFA.H.leu(ala)s.OH in different 0.1M buffers of varyi
pH values.

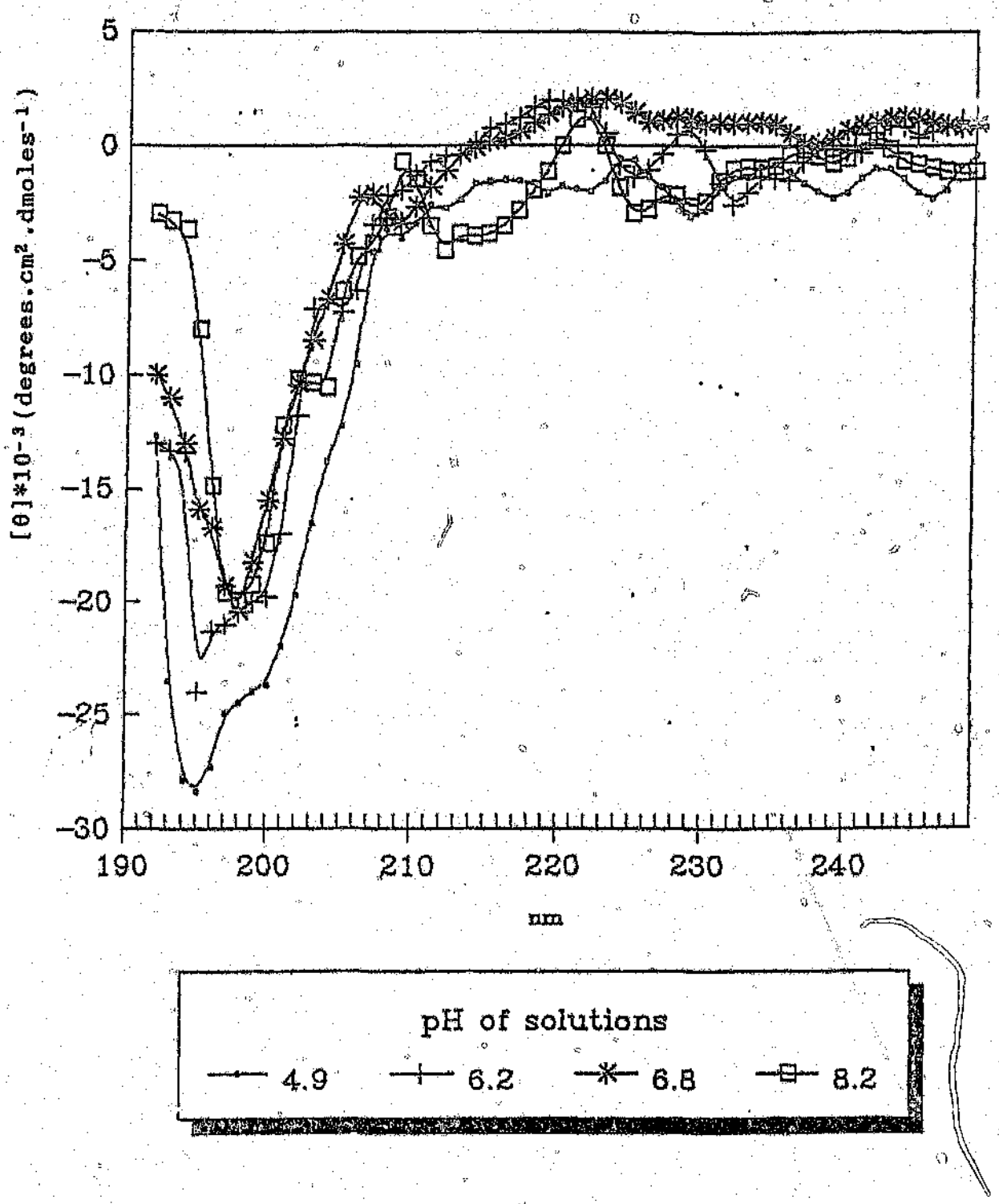


Figure 24 TFA.H.(ala)₅leu.OH in 0.1M potassium phosphate buffers of varying pH values.

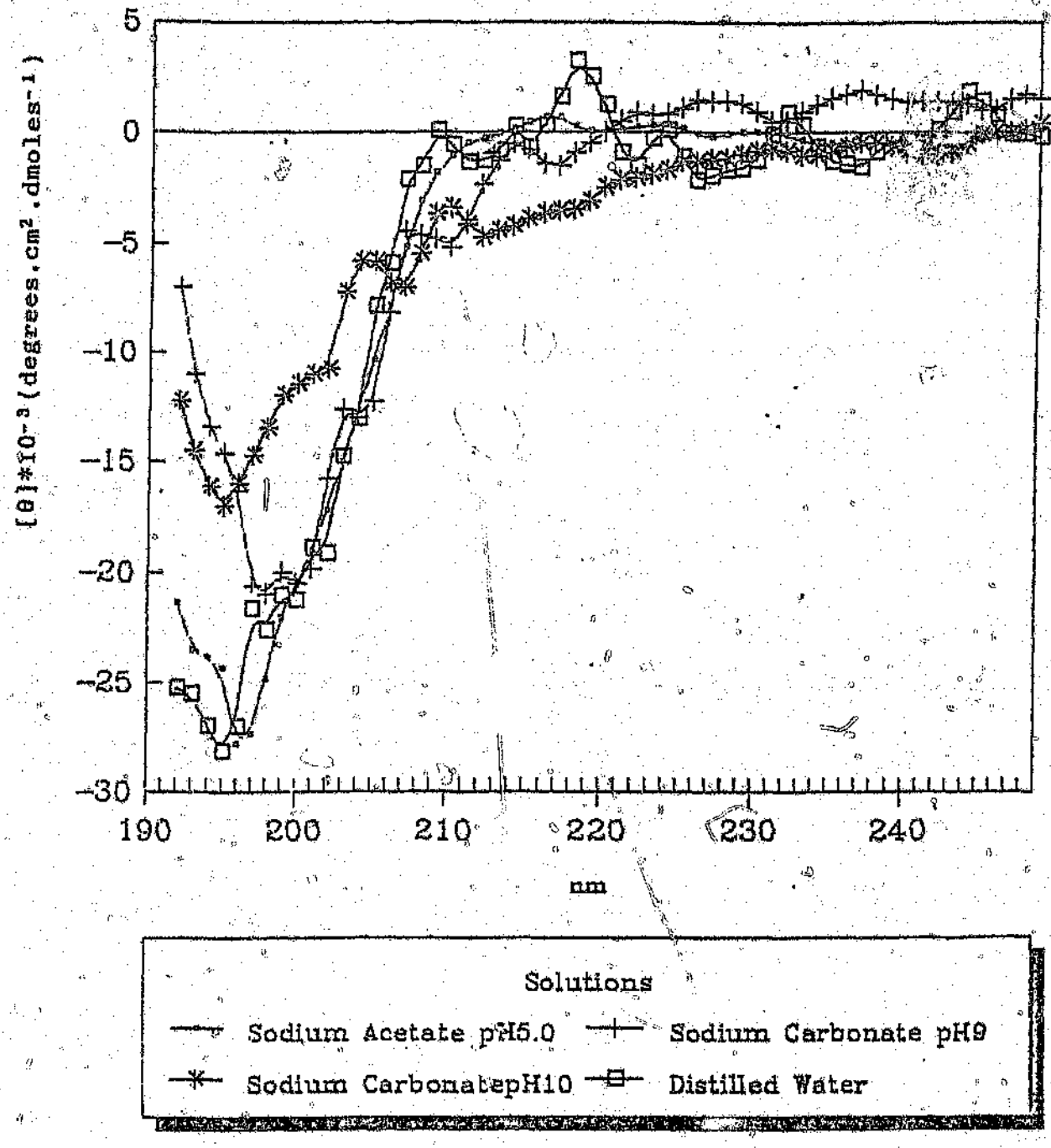


Figure 25 TFA.H.(ala)$_5$leu.OH in different 0.1M buffers of varying pH$_6$ values.

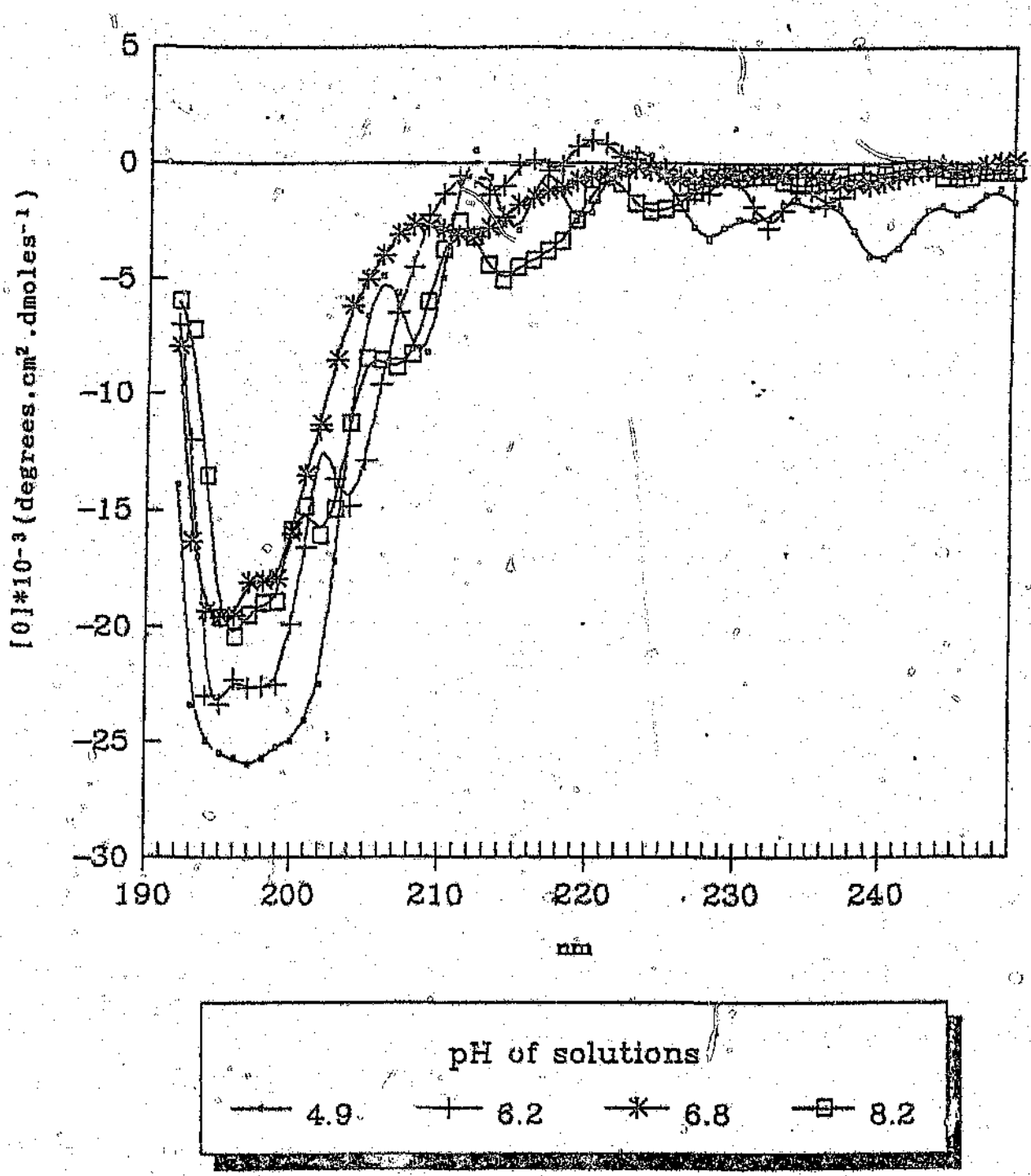


Figure 26 TFA.H.ile(ala)$_5$.OH in 0.1M potassium phosphate buffers of varying pH values.

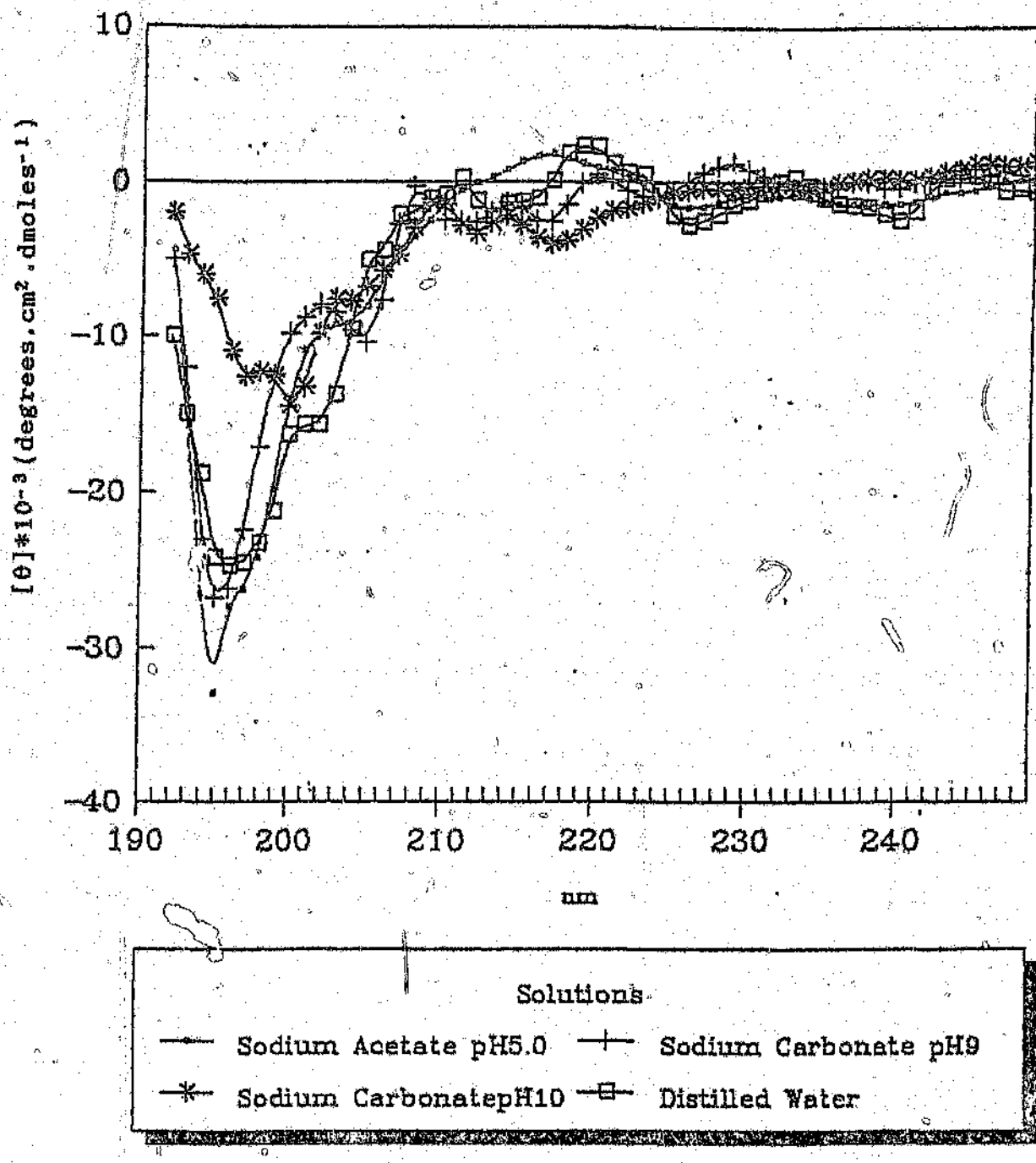


Figure 27 TFA.H.ile(ala)$_5$.OH in different 0.1M buffers of varying pH values.

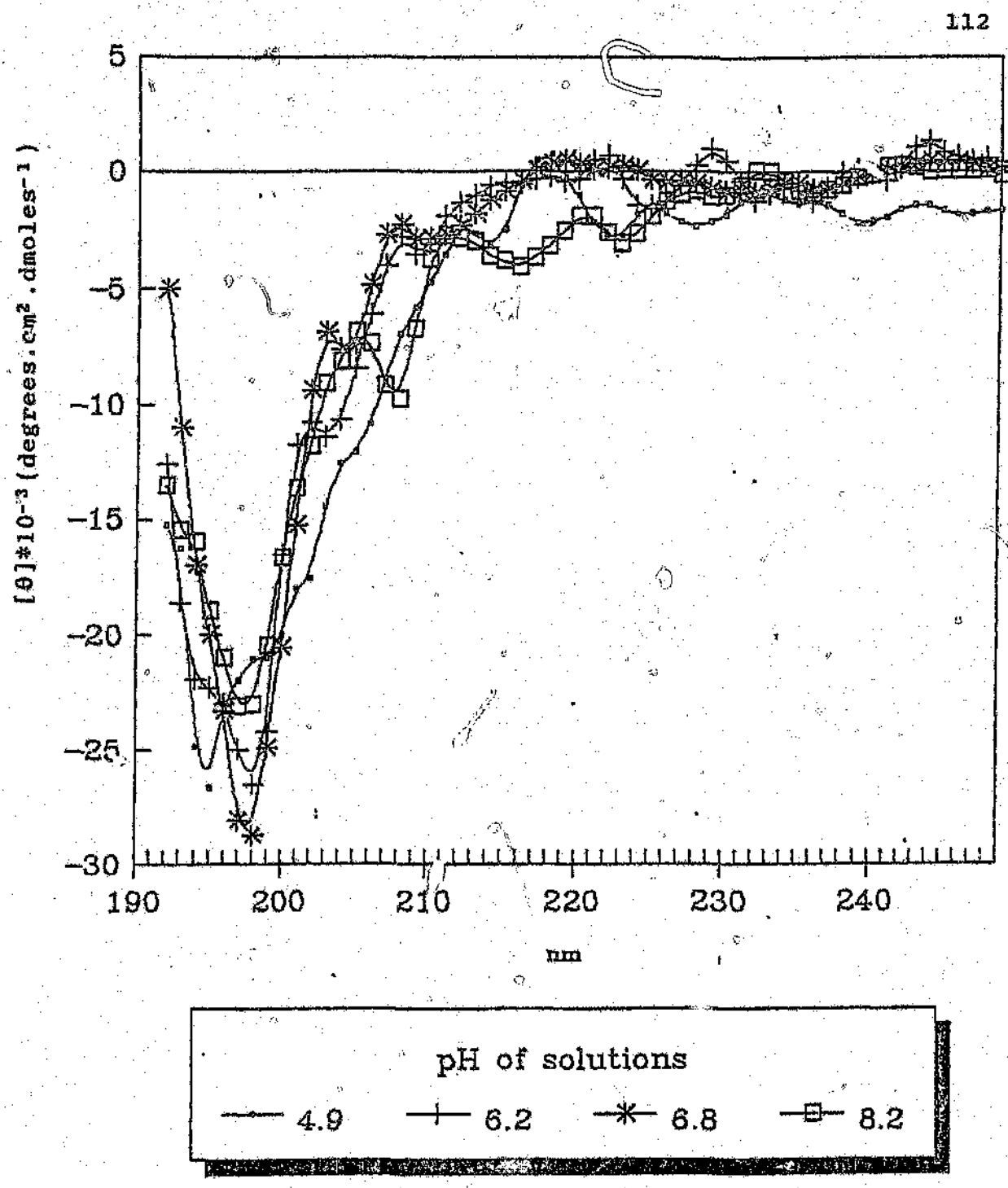


Figure 28 TFA.H.(ala)sile.OH in 0.1M potassium phosphate buffers of varying pH values.

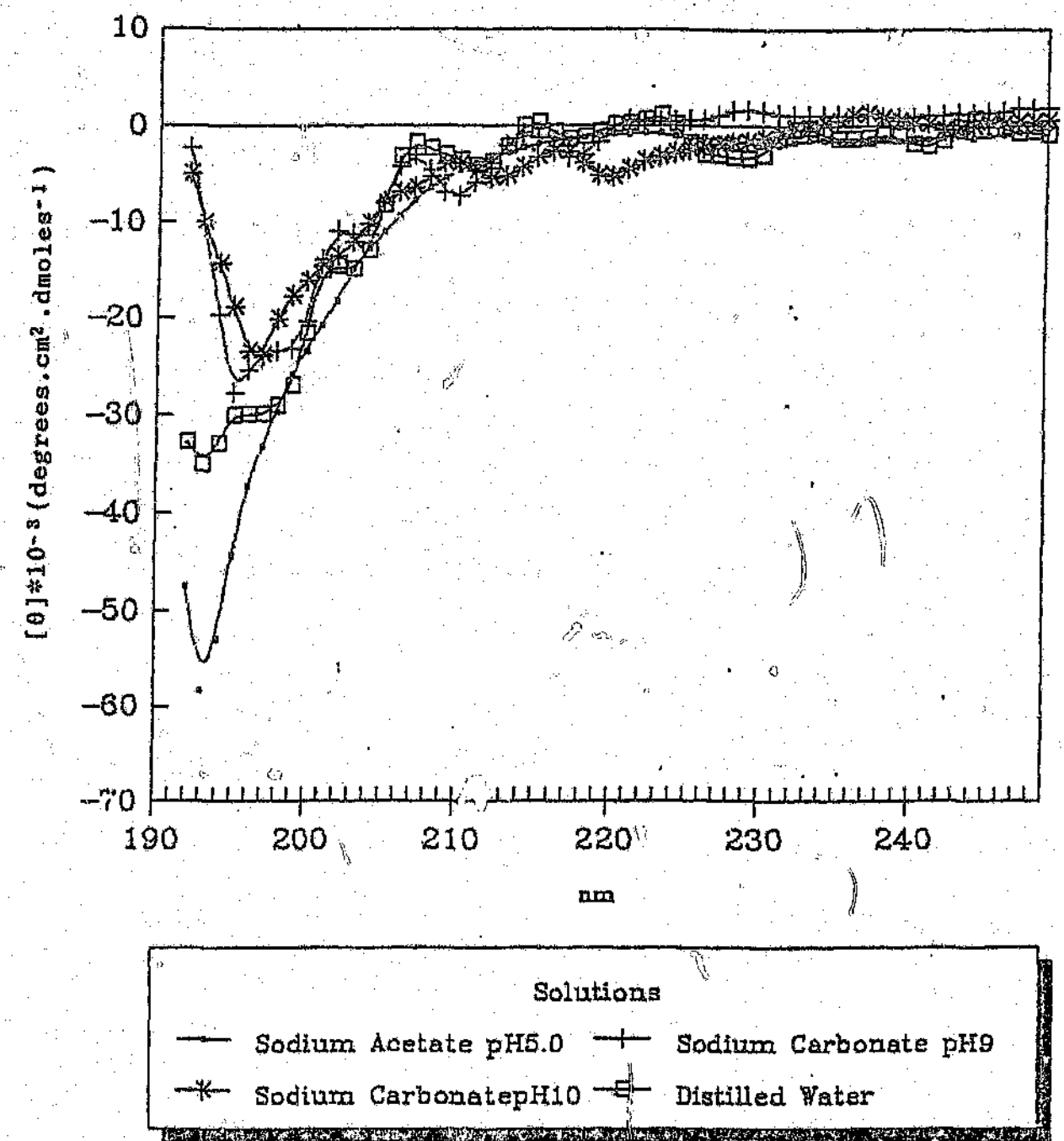


Figure 29 TFA.H.ile(ala)5.OH in different 0.1M buffers of varying pH values.

many possible impurities that could explain these peaks, such as -OH groups or a -CH group attached to an amide.[72] They show that the purification was not successful. The peptide proved to be very difficult to dissolve even after the addition of a few drops of HCl or ammonia to the acetonitrile HPLC buffer. This insolubility made the purification difficult.

4.3 Ultraviolet circular dichroism results

4.3.1 Aqueous solution results

The aqueous solution CD curves (figures 22 to 29) all show negative π-π^* transitions at 195nm to 200nm. The n-π^* transitions are positive or close to zero. These characteristics are typical of peptides in a random or "statistical coil".[94,76,64] Basing an analysis upon the Greenfield and Fasman curves for statistical coils, figure 19,[76] most of the peptides in aqueous solution formed "statistical coils" with 20% to 40% ß-structure and the rest random coil. The only exception is the curve for TFA.H.(ala)$_5$leu.OH in 0.1M sodium carbonate, pH 10.0 (figure 25) which formed a "statistical coil" of 20% α-helix and 80% random coil. The NMR spectra showed, section 4.1, that the peptides formed random coils althgough there was some evidence that some structure may have formed in D_2O. The water CD results confirm that random coils were the predominant structures.

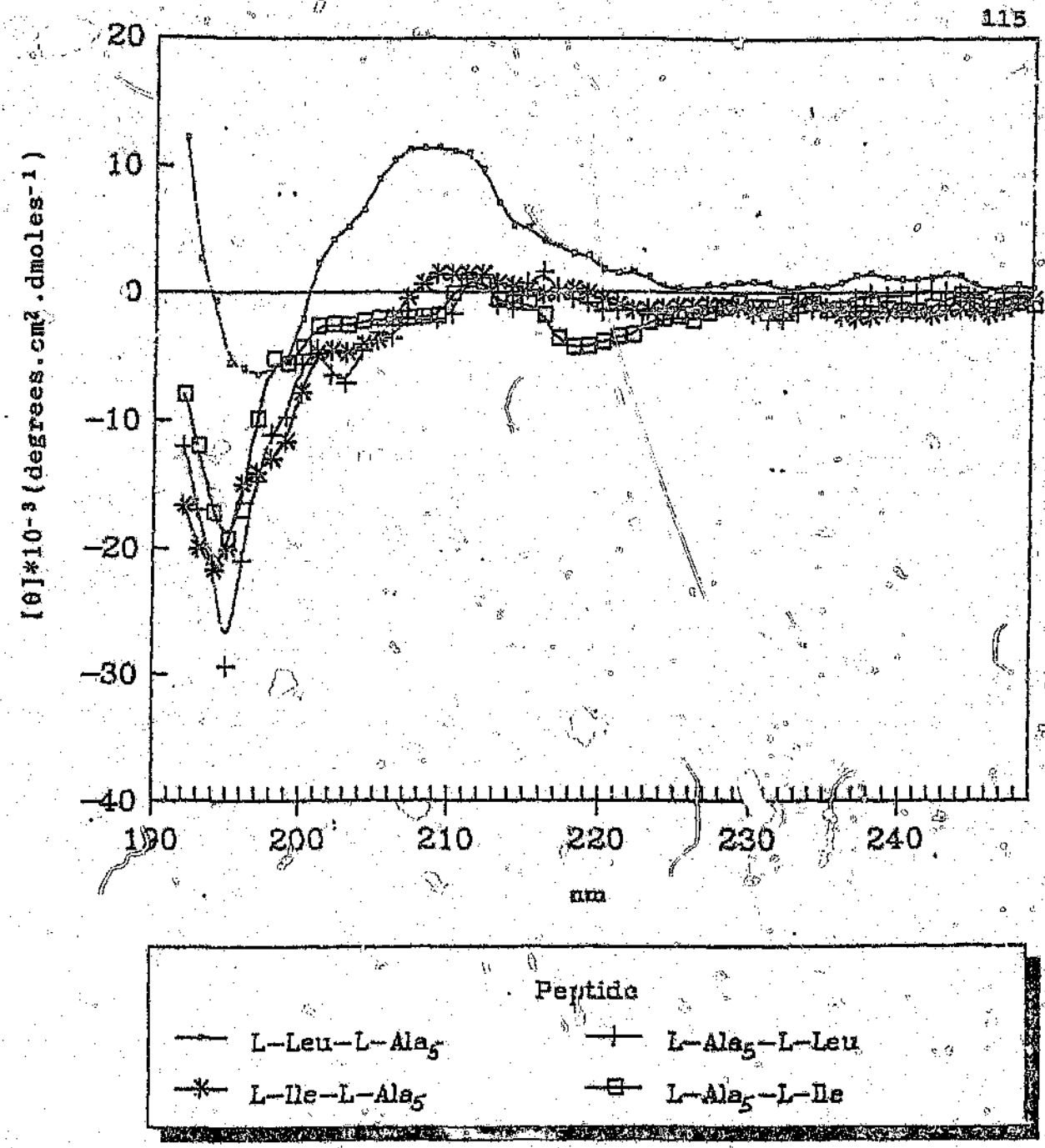


Figure 30 Methanol CD spectra of TFA.H.leu(ala)$_5$.OH (L-leu-L-ala$_5$), TFA.H.(ala)$_5$leu.OH (L-ala$_5$-L-leu), TFA.H.ile(ala)$_5$.OH (L-ile-L-ala$_5$) and TFA.H.(ala)$_5$ile.OH (L-ala$_5$-L-ile)

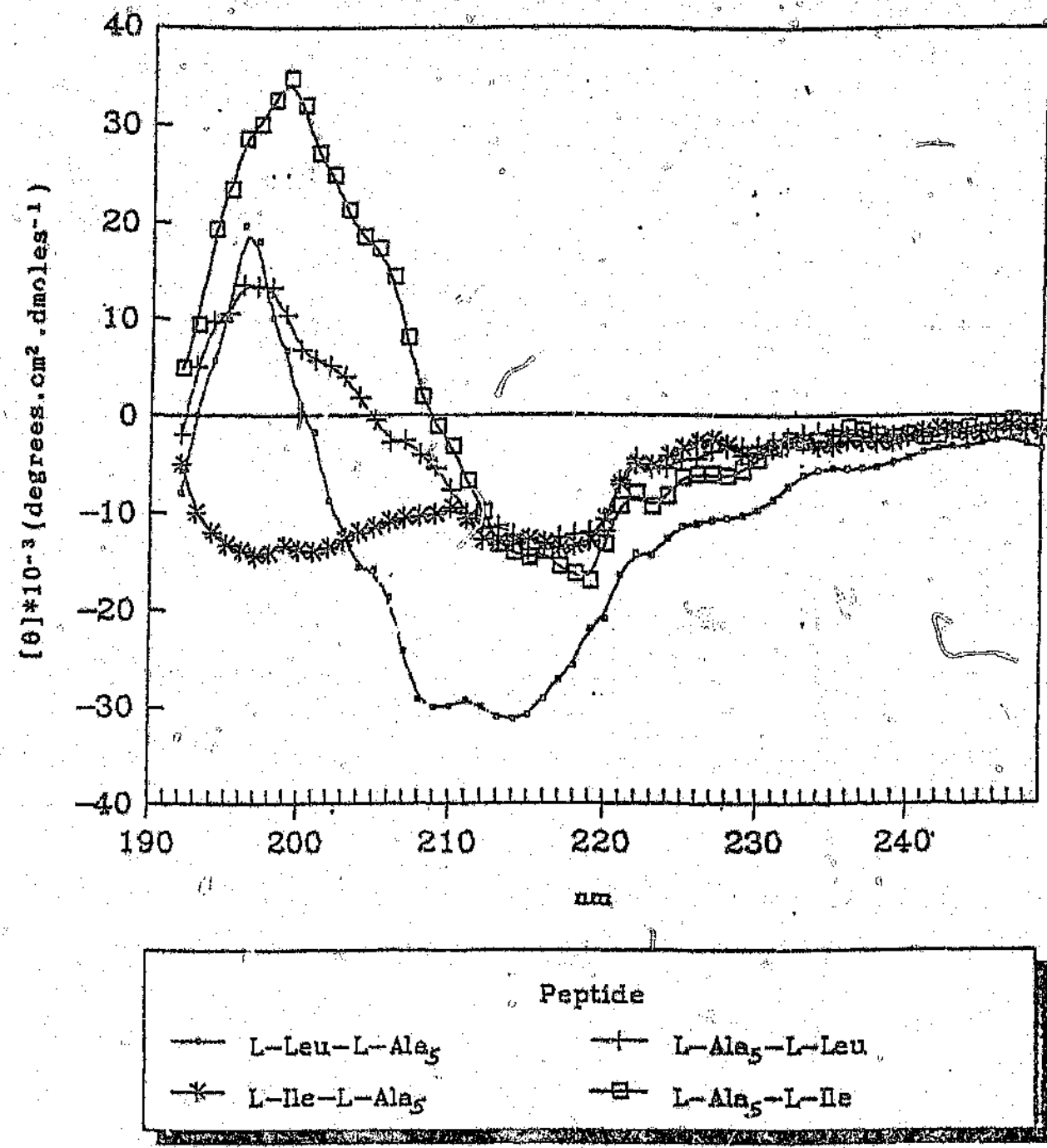


Figure 31 Ethanol CD spectra of TFA.H.leu(ala)5.OH (L-leu-L-ala5), TFA.H.(ala)5leu.OH (L-ala5-L-leu), TFA.H.ile(ala)5.OH (L-ile-L-ala5) and TFA.H.(ala)5ile.OH (L-ala5-L-ile)

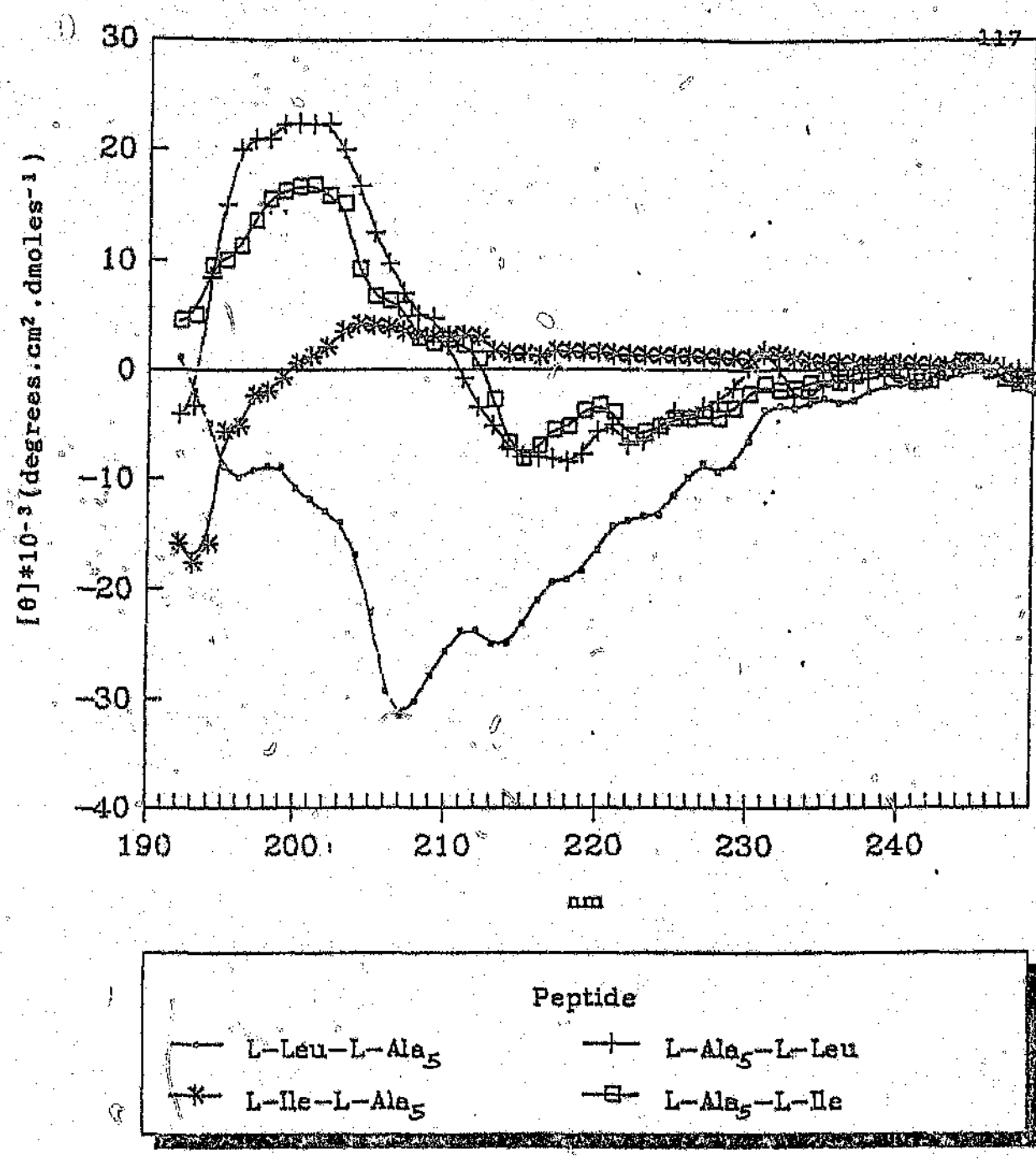


Figure 32 Propanol CD spectra of TFA.H.leu(ala)5.OH (L-leu-L-ala5), TFA.H.(ala)5leu.OH (L-ala5-L-leu), TFA.H.ile(ala)5.OH (L-ile-L-ala5) and TFA.H.(ala)5ile.OH (L-ala5-L-ile)

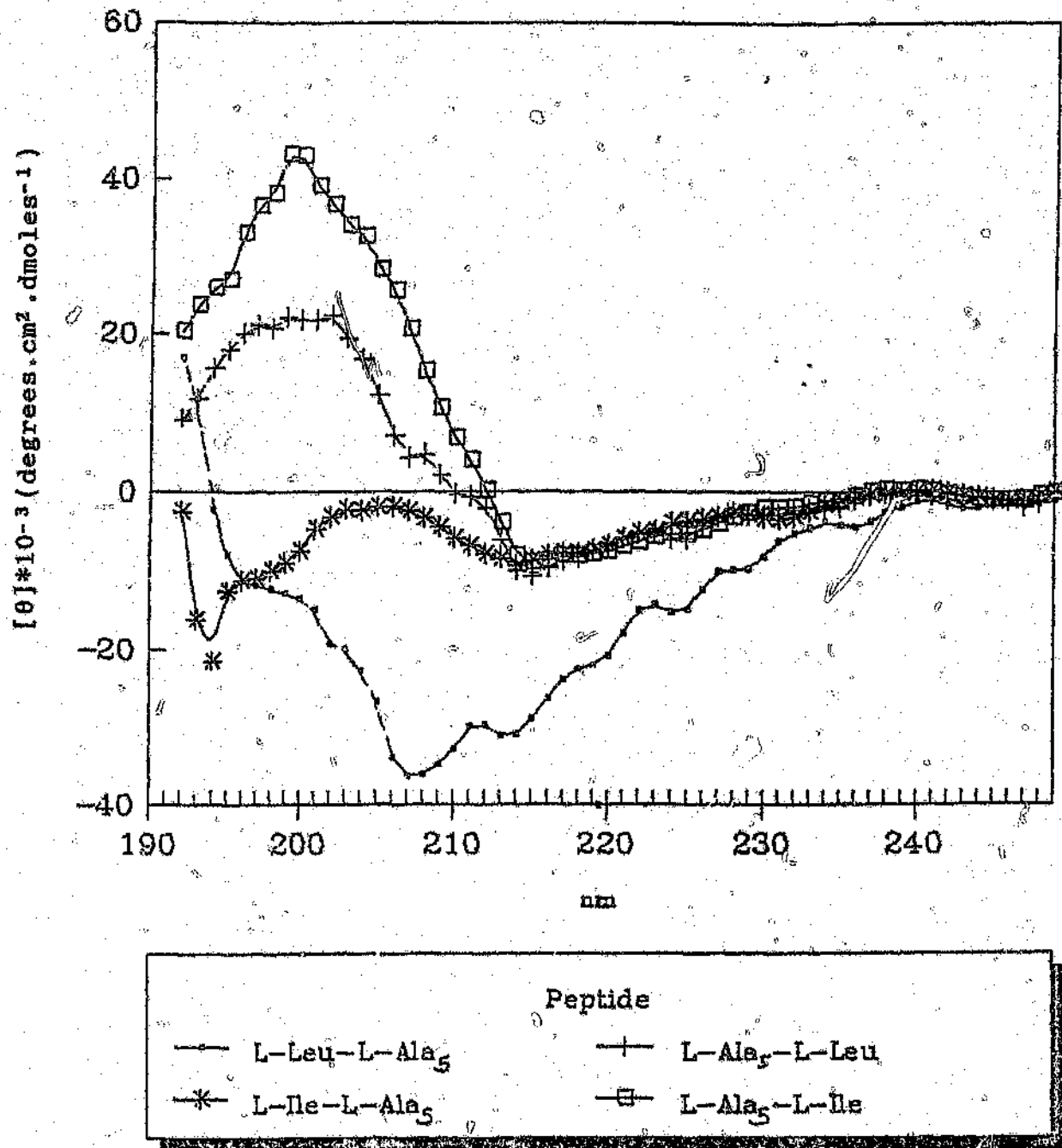


Figure 33 Butanol CD spectra of TFA.H.leu(ala)$_5$.OH (L-leu-L-ala$_5$), TFA.H.(ala)$_5$leu.OH (L-ala$_5$-L-leu), TFA.H.ile(ala)$_5$.OH (L-ile-L-ala$_5$) and TFA.H.(ala)$_5$ile.OH (L-ala$_5$-L-ile)

Table 14 The percentage of β-structure in the "statistical coils" found in the alcoholic and TFE/ethanol CD studies. The percentages were based upon the Greenfield Fasman reference curves.[79]

Solution	Peptides		
	TFA.H.(ala)$_5$.leu.OH	TFA.H.(lo.(ala)$_5$.OH	TFA.H.(ala)$_5$.ile.OH
Methanol	20%	20%	20%
Ethanol	70–80%	20–40%	100%
Propanol	>80%	20%	80%
Butanol	>80%	20–40%	100%
TFE/Ethanol			
25%	80–100%	80–90%	100%
50%	see discussion	see discussion	see discussion
TFE	40%	see discussion	see discussion

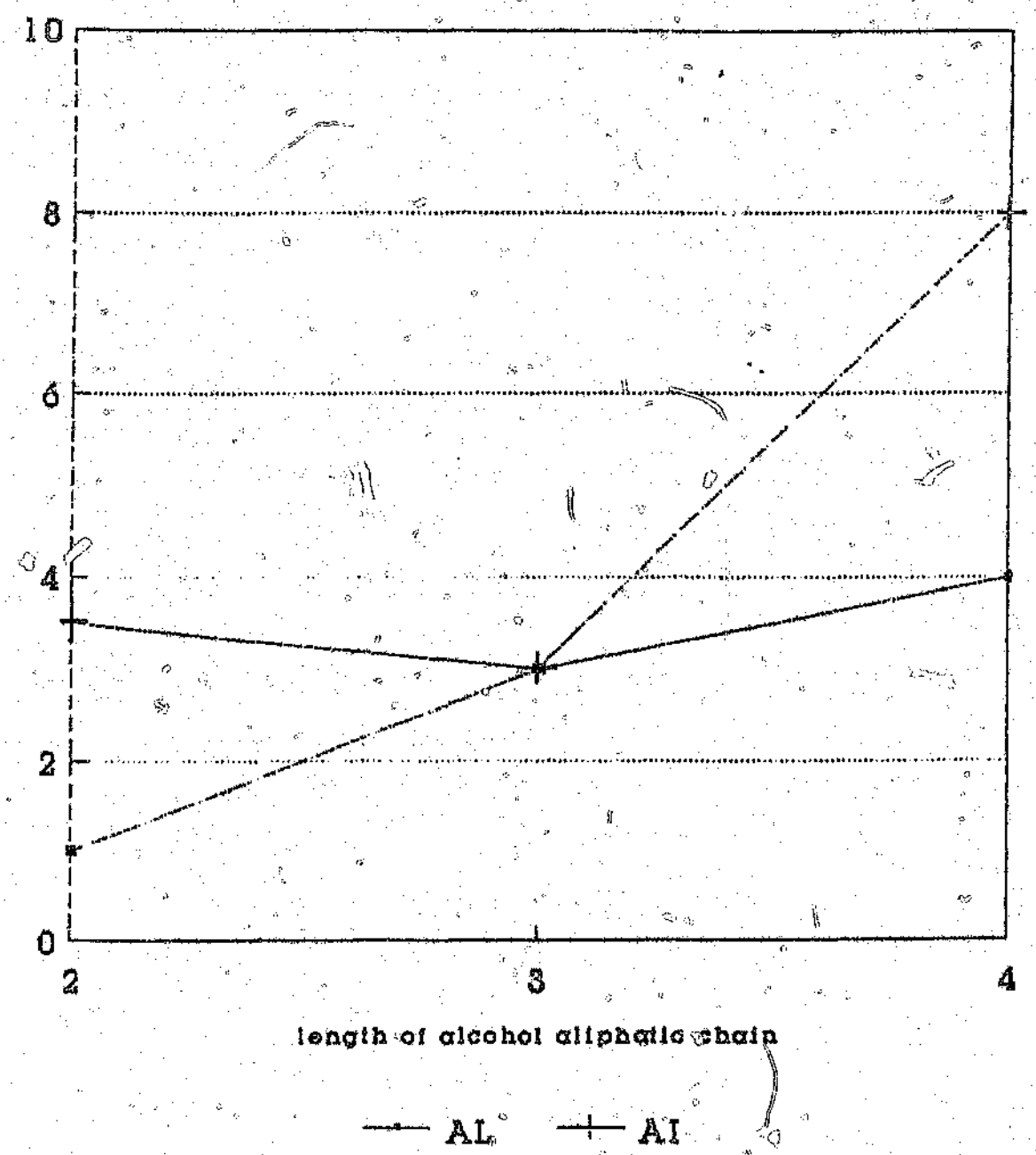


Figure 34 The π-π*/n-π* CD ratios for TFA.H.(ala)₃leu.OH (AL), TFA.H.(ala)₃ile.OH (AI) in ethanol, propanol and butanol.

4.3.2 Alcoholic solution results

The CD curves of the alcoholic solutions (figures 30 to 33) fall into two groups: the TFA.H.(ala)$_5$ile.OH, TFA.H.ile(ala)$_5$.OH and TFA.H.(ala)$_5$leu.OH results conform to the reference spectra for either secondary structures (figure 18) or "statistical coils" curves (figure 19).[76] The TFA.H.leu(ala)$_5$.OH results differ from those of the other three peptides.

In methanol the TFA.H.(ala)$_5$leu.OH, TFA.H.ile(ala)$_5$.OH and TFA.H.(ala)$_5$ile.OH CD curves are all similar (figure 30). They are characterised by strong negative π-π^* transitions with ellipticities between -20,000 and -27,000 deg.cm2/mol-1. The n-π^* transitions around 210nm are close to zero. These results are closest in ellipticity values and shape to a statistical coil curve with a mixture of 20% β-structure and 80% random coil (figure 19).[76] The TFA.H.leu(ala)$_5$.OH curve (figure 30) differs from the other three altogether and does not fit the curves for the recognised secondary structures or "statistical coils".[64,76,78] The structure is therefore unclassified and distinct from that assumed by the other three peptides.

In ethanol, propanol and butanol, the TFA.H.(ala)$_5$ile.OH and TFA.H.(ala)$_5$leu.OH CD curves (figures 31, 32 and 33) show strong positive π-π^* transitions and weakly negative n-π^* transitions. Their curves are similar to those published for β-structures.[64,78] Based upon the value of the π-π^* transition

the TFA.H.(ala)$_5$ile.OH is the strongest ß-structure former of the two in ethanol and propanol. The deviations of these curves from the reference ß-structure CD can be explained in two ways.[76] The curves could be due to "statistical coils" of random coil and ß-structure mixtures.[76] From figure 19, the percentage ß-structure of TFA.H.(ala)$_5$leu.OH would be in ethanol 70% to 80% and in propanol and butanol in excess of 80%. In TFA.H.(ala)$_5$ile.OH the ß-structure percentage would be 100% in ethanol and butanol and about 80% in propanol. The other explanation relies upon the assumption that as ß-sheets grow in the number of strands they incorporate, there is a decrease in the magnitude of the π-π^* transition and an increase in the magnitude of the n-π^* transition.[79] The π-π^*/n-π^* ratio for TFA.H.leu(ala)$_5$.OH and TFA.H.(ala)$_5$ile.OH increases as the alcohol aliphatic chain length increases (figure 34). For both peptides, therefore there is the possibility that the sheet sizes decrease as the alcohol polarity decreases. The TFA.H.(ala)$_5$ile.OH size decreases are greater than TFA.H.(ala)$_5$leu.OH and this would imply that its ß-sheets are smaller. Based upon this, TFA.H.(ala)$_5$leu.OH formed multi-stranded antiparallel ß-sheets of between four to twenty strands in ethanol (table 1).[79] In ethanol the TFA.H.(ala)$_5$ile.OH formed smaller or "single extended ß-chains".[79]

In the alcohol solutions the TFA.H.ile(ala)$_5$.OH CD curves (figures 31, 32 and 33) fit "statistical coil" curves of

ß-structure and random coil (figure 19).[76] In ethanol the "statistical coil" contains about 40% ß-structure, in propanol and butanol the amount of ß-structure is reduced to between 20% and 40%.

The TFA.H.leu(ala)$_5$.OH methanol CD curve was not classified and was due to an unknown structure. The TFA.H.leu(ala)$_5$.OH CD curve in ethanol (figure 31) shows a positive π-π^* transition, similar to ß-structures, and two negative transitions at 208nm and 215nm, similar to α-helical CD.[64,78] The curve does not fit the published "statistical coil" curves for mixtures of ß-structure and α-helix.[76] In propanol and butanol (figures 32 and 33) there is an increased accentuation of the 208nm minimum and the positive π-π^* transition is lost. Thus in ethanol, propanol and butanol, as in methanol, the TFA.H.leu(ala)$_5$.OH CD differs from the three other peptides and the reference curves. The double minima in ethanol is similar to that seen in previous work on TFA.H.val(ala)$_5$.OH.[62] TFA.H.val(ala)$_5$.OH was found to form a hydrophobic pocket in aqueous and methanol solutions (Dr. A. Nel in print, Biopolymers 1991). The propanol spectra of TFA.H.val(ala)$_5$.OH and TFA.H.leu(ala)$_5$.OH are also similar.[62] L-leucine and L-valine are both branched at the gamma and beta side chains carbons respectively. This similarity in their structures could allow them to form similar secondary structures. Therefore the TFA.H.leu(ala)$_5$.OH results could be due to the formation of

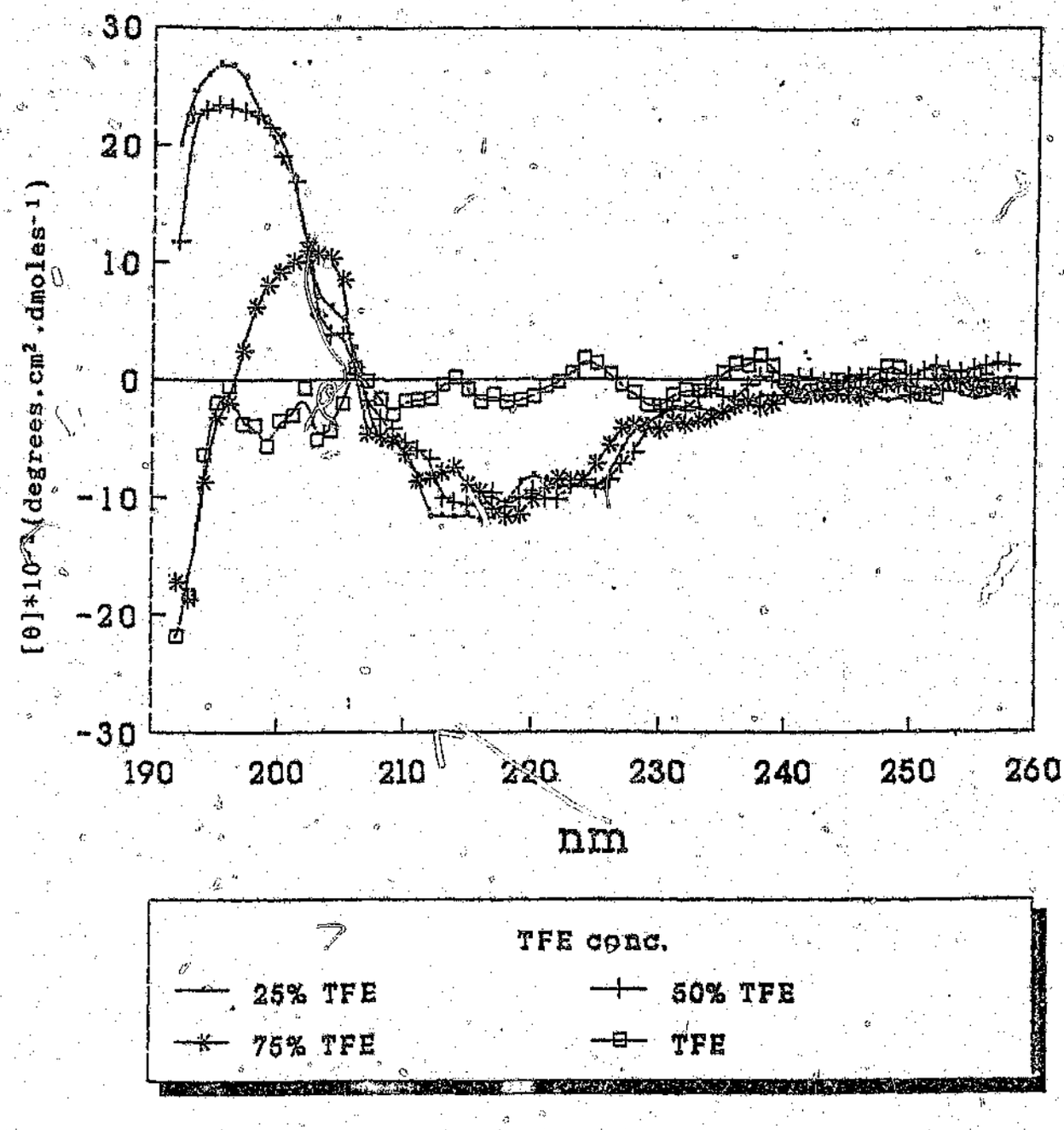


Figure 35 CD spectra of TFA.H.leu(ala)₅.OH in TFE/ethanol mixtures.

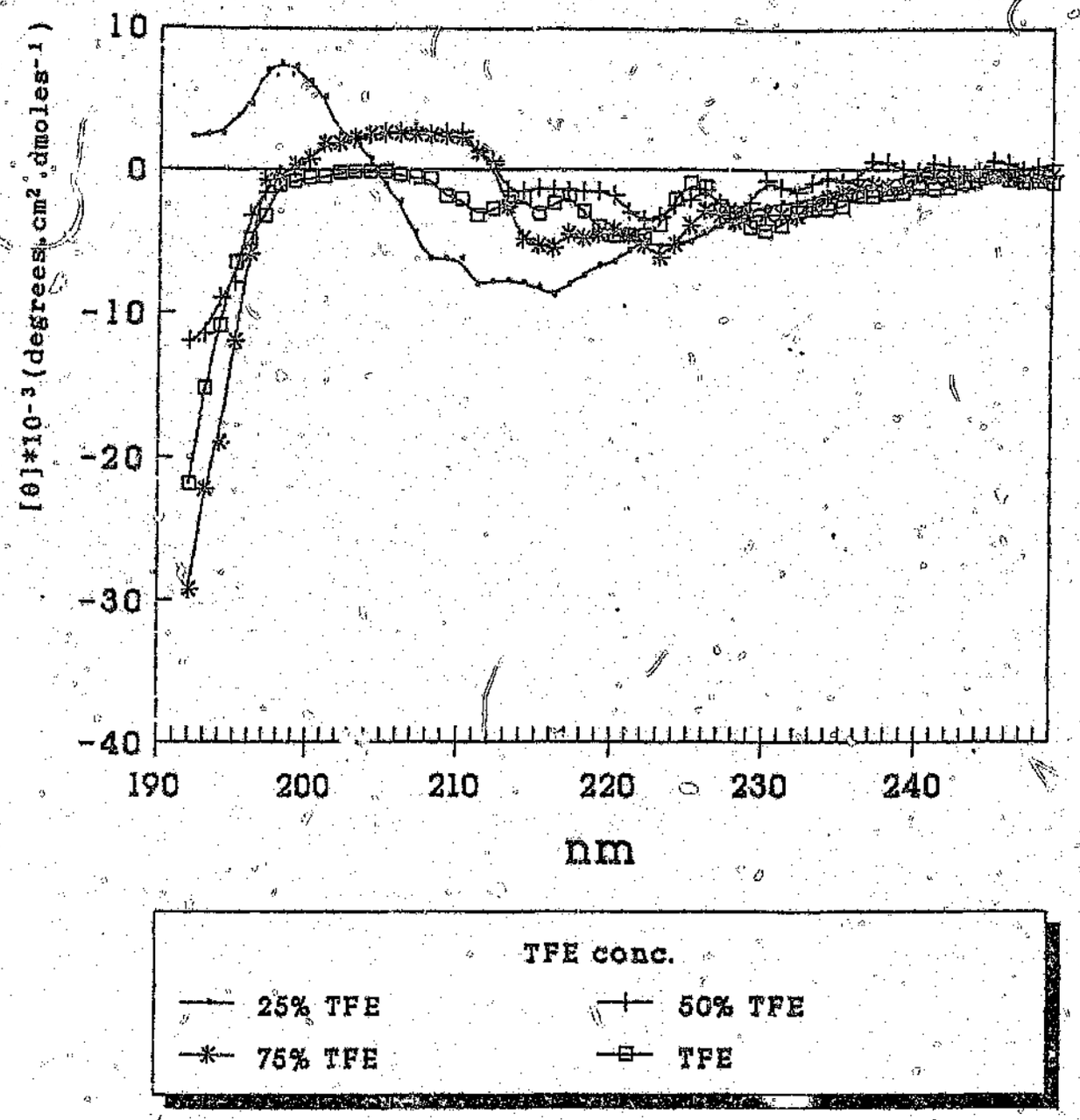


Figure 36 CD spectra of TFA.H.ile(ala)5.OH in TFE/ethanol mixtures.

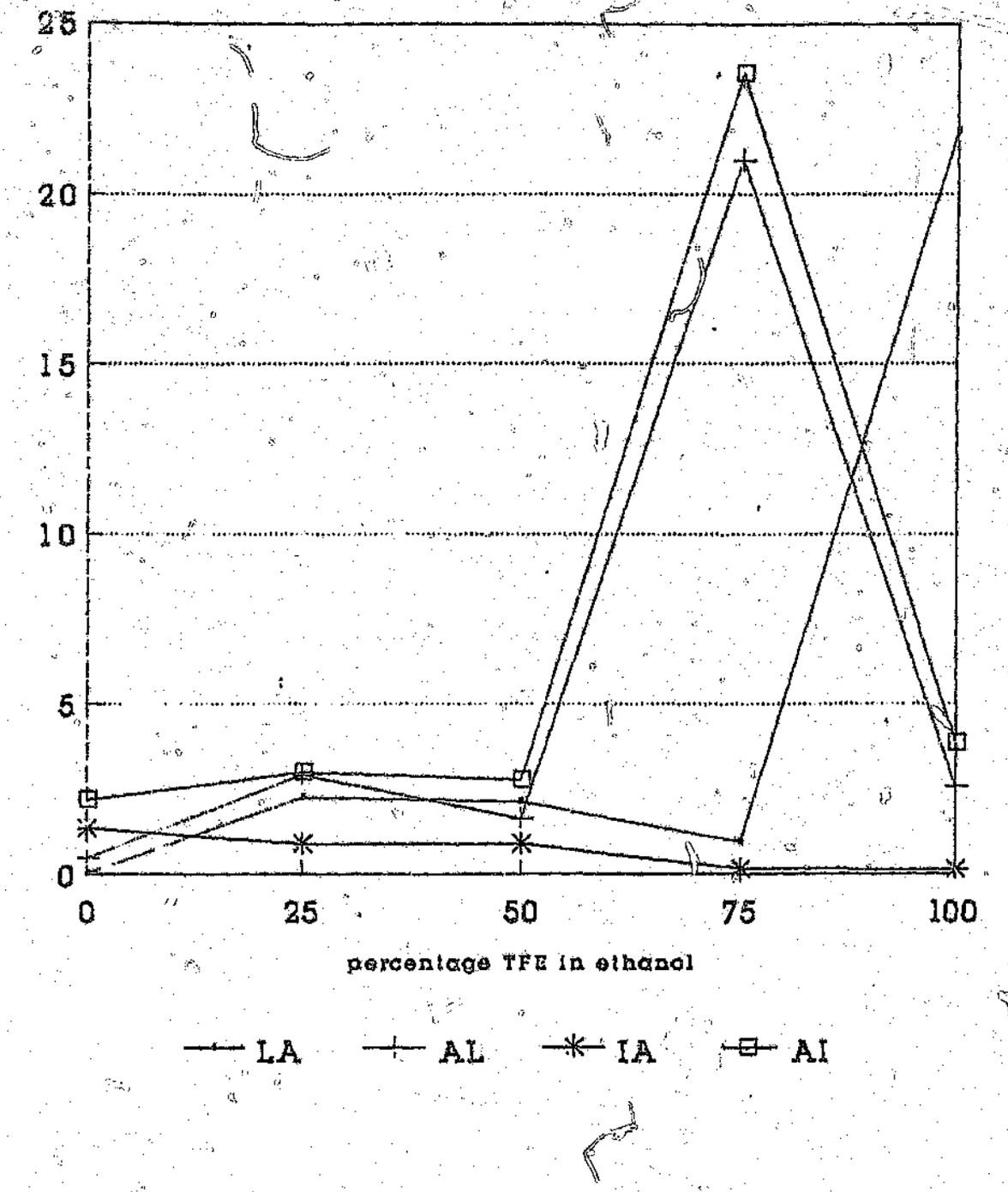


Figure 37 The π-π*/n→π* CD ratios for TFA.H.leu(ala)$_5$.OH (LA), TFA.H.(ala)$_5$leu.OH (AL), TFA.H.ile(ala)$_5$.OH (IA) and TFA.H.(ala)$_5$ile.OH (AI)in TFE/ethanol solutions.

the same type of structure seen in TFA.H.val(ala)$_5$.OH in alcoholic solution.

4.3.3 Trifluoroethanol/ethanol solution CD results

TFE acts as a hydrogen bond donor and solvates peptides. It reduces inter-molecular hydrogen bonding thereby disrupting ß-structures.[58] Therefore the titration of peptide solutions with TFE can be used to measure the stability of ß-structure.

The TFA.H.(ala)$_5$leu.OH and TFA.H.(ala)$_5$ile.OH CD TFE titration curves are similar (appendix). As with their alcohol CD results, there are two means of analysis based upon the composition of "statistical coil" or upon the possible sizes of ß-sheets formed. In ethanol (figure 31) TFA.H.(ala)$_5$leu.OH and TFA.H.(ala)$_5$ile.OH formed ß-structures. In the titrations of ethanol with TFE the CD curves formed looked similar to the "statistical coil" curves of Greenfield and Fasman.[76] Using these curves, the possible percentage secondary structural composition of the "statistical coils" are given in table 14. The 25% and 50% TFE concentration curves for both peptides match the published "statistical coil" curves well.[76] The 75% TFE "statistical coils" can be analysed in two ways. Based upon the π-π^* transition values they would contain 100% ß-structure.75 Based upon the n-π^* transition the ß-content would contain 40% to 20%.76 This analysis also applies to the TFA.H.(ala)$_5$ile.OH TFE curve which could contain 100%

ß-structure based upon the π-π^* transition or 40% based on the n-π^* transition. The TFA.H.(ala)$_5$leu.OH TFE curve fits the 40% ß-structure "statistical coil" curve (figure 19) without any contradiction from the two transitions. Therefore the TFE has less of an effect upon the TFA.H.(ala)$_5$ile.OH ß-structure than the TFA.H.(ala)$_5$leu.OH. The plot of the variation in the π-π^*/n-π^* transition ratios against the TFE concentrations (figure 37) for TFA.H.(ala)$_5$leu.OH and TFA.H.(ala)$_5$ile.OH shows that, excepting 75% TFE, the numerical values of the ratios lie between 1 and 4 for both peptides. The 75% TFE ratios rise above 20 for both TFA.H.(ala)$_5$leu.OH and TFA.H.(ala)$_5$ile.OH. The sudden changes in the magnitudes of the π-π^*/n-π^* transitions between 50% and 75% TFE imply that the ß-structures seen in the CD curves changed radically. The ratio changes are possibly linked to the sizes of ß-sheets and therefore the smallest ß-structures occurred in 75% TFE.[79] Based upon the CD curve (figure 31), TFA.H.(ala)$_5$leu.OH in ethanol formed antiparallel ß-structures with four to twenty strands.[79] In TFE the CD curve changes altogether to that of a "statistical coil" with a high content of random coil (appendix).[76] The CD results show that in TFE the TFA.H.(ala)$_5$ile.OH (appendix) maintained the predominantly ß-structure seen in ethanol, where it formed a "single extended ß-chain". Therefore in TFE concentrations above 50% the TFA.H.(ala)$_5$leu.OH ß-sheets

are solvated and the structure is lost. The TFA.H.(ala)$_5$ile.OH in ethanol formed a "single extended β-chain" or small β-sheets"[79] and therefore solvation by TFE had less effect. Above 75% TFE a change did occur to TFA.H.(ala)$_5$ile.OH as the π-π^*/n-π^* ratio was reduced, while according to the CD curves β-structure was maintained. It is possible that the β-structures in pure TFE and in 25%, 50% and 75% TFE were different.

For both TFA.H.(ala)$_5$leu.OH and TFA.H.(ala)$_5$ile.OH two analyses of the alcohol and TFE series CD curves were possible: one according to the composition of "statistical coils"[78] and the other according to the size of β-sheets.[79] The two are not mutually exclusive but what is plain is that the different possible molecular variations of β-structures make the application of any one method of analysis limited. The unclassified α/β curve of TFA.H.leu(ala)$_5$.OH seen in the ethanol CD (figure 31) is changed by titration with TFE. In the 25%, 50% and 75% TFE solution CD curves (figure 35) the n-π^* double minima disappears. These curves are similar to "statistical coils" with mixtures of random coil and β-structure.[76] If "statistical coils" occurred, the TFA.H.leu(ala)$_5$.OH in 25% and 50% TFE solutions would have comprised 80-100% β-structure and the 75% TFE solution would have contained 60-80% β-content. The pure TFE curve does not follow the "statistical coil" trend and the structures formed were of an unknown nature. The π-π^*/n-π^* ratios for

the TFA.H.leu(ala)$_5$.OH TFE titration CD (figure 37) differ from the C-terminally substituted peptides. If "statistical coil" was present, the interaction of the peptides in 75% TFE was also different. In TFE there was no evidence for ß-structure (figure 35) so the increase in the π-π^*/n-π^* transition was not due to sheet size decreases but some other conformational change. This change could be due to the solvation of a hydrophobic pocket by TFE.

The π-π^*/n-π^* transition ratios of the TFA.H.ile(ala)$_5$.OH TFE titration CD (figure 37) do not vary greatly but this hides significant changes in the CD curves (figure 36). In ethanol TFA.H.ile(ala)$_5$.OH formed a "statistical coil" containing 20% to 40% ß-structure and random coil (table 14). From the CD curve in 25% TFE (figure 36) the ß-content is between 60% and 80%. In 50%, 75% and pure TFE solutions the CD curves (figure 36) show negative π-π^* and positive n-π^* transitions and fit neither the reference secondary structures (figure 18) nor the "statistical coil" curves (figure 19).[76,78] Increasing the TFE concentration above 25% therefore disrupts TFA.H.ile(ala)$_5$.OH ß-structure. In the higher TFE concentrations the structure formed was unknown.

4.4 Infrared results

4.4.1 Solid phase infrared results

Table 15 shows the results of the solid phase KBr infrared spectra for TFA.H.leu(ala)$_5$.OH, TFA.H.(ala)$_5$leu.OH, TFA.H.ile(ala)$_5$.OH and TFA.H.(ala)$_5$ile.OH. The non-amide chromophores are assigned according to table 5. The amide A bands for the four peptides all fall within the range assigned to intermolecularly hydrogen bonded structures.[63,49,89] The band frequency decreases with an increase in the hydrogen bond strength,[63] implying that the leucine/alanine peptides form stronger hydrogen bonds than the isoleucine/alanine peptides. This is consistent with the fact that L-leucine has been found to be a stronger ß-structure former than L-isoleucine.[88] The amide A shoulders at 3425cm^{-1} in the leucine peptides could be due to moisture in the KBr.[72] The amide B bands can be attributed to free amine groups.[63] The amide I, II and III peaks for all four peptides are due to intermolecular hydrogen bonded ß-structures.[49,59,60,63,63] The TFA.H.ile(ala)$_5$.OH and TFA.H.(ala)$_5$ile.OH amide V bands (720cm^{-1}, 700cm^{-1}) can be assigned to extended ß-conformations.[67,79] The TFA.H.leu(ala)$_5$.OH and TFA.H.(ala)$_5$leu.OH amide V band region has peaks at 720cm^{-1}, 700cm^{-1} and 620cm^{-1} (table 15). The 620cm^{-1} peak can be attributed to α-helix[86,84] or skeletal deformation.[85] The absence of an amide II peak near 1540cm^{-1} and shoulder at 1516cm^{-1} for confirmation rules out the

Table 15 Solid Phase KBr Infrared Results of TFA.H.Peptides.OH

Amide Band	LA5 Freq.1/cm	A5L Freq.1/cm	IA5 Freq.1/cm	A5I Freq.1/cm	Band Assignment
A	3425s	3425s	3260p	3260	Intermolecular hydrogen
	3275p	3275p			bonded structure
B	3050-3100p	3050-3100p	3050p	3075p	Free -NH groups
	2950s	2975s	2950p	2975p	-CH stretch
	2925p	2925p	2925p	2950p	-CH antisymmetrical
					stretch
I	1670s	1680s	1680s	1680s	See results and
	1660s	1660s	1650s	1660s	discussion
	1640s	1650s	1620p	1620p	
	1620p	1640s		1610s	
		1620p			
II	1550s	1550s	1570s	1560s	
	1530s	1525p	1540s	1530s	
	1520p	1500s	1520p	1520p	
			1510s	1500s	
	1440p	1450p	1450p	1450p	Alanyl methyl deformations
	1390-	1380p	1370p	1380p	-NH in plane bending mode
	1360p				and -CH of -CH
					symmetrical bending
III	1200p	1200p	1200p	1200p	See results and discussion
	1170p	1170p	1180p	1140p	Alanyl methyl rocking modes
	1140p	1130s	1140p	1050p	backbone and side chain
					stretching
	1050p	1050p	1050p		weak -OH deformation
	960p	960p	960p	960p	Different C-C stretching
					modes
	800p	830s	840p	800p	-CH rocking modes
		800p	780		
V	720p	720p	720p	720p	Skeletal deformation
	700s	700s		700s	
	620p	620p			

Key

s shoulder
p peak

LA5= TFA.H.leu(ala)$_5$.OH
A5L= TFA.H.(ala)$_5$leu.OH

IA5= TFA.H.ile(ala)$_5$.OH
A5I= TFA.H.(ala)$_5$ile.OH

helical possibility.[63] Therefore the amide V bands (720cm^{-1} and 700cm^{-1}) represent ß-conformations.

1690cm^{-1} amide I shoulders, which identify antiparallel ß-sheet,[66,49] are absent from all the spectra (table 15) nor do the peaks correspond to the parallel ß-sheet frequencies.[49] The peptides are probably therefore in some other intermolecularly hydrogen bonded extended ß-structures. The shoulders of the amide bands suggest that there are small populations of other conformations present. The amide I 1660cm^{-1} and amide II 1530cm^{-1} to 1540cm^{-1} shoulders indicate a random or unclassified structure;[77,59] the same amide II shoulders also indicate ß-pleated sheet.[49]

4.4.2 Liquid phase infrared results

4.4.2.1 Alcohol solution results

The liquid phase spectra produced very strong amide I and II bands. The high absorbance of the alcohols and the low concentrations of peptides meant that the øther amide bands did not appear.

The methanol TFA.H.(ala)$_5$leu.OH amide I peak indicates random coil[77] and extended hydrogen bonded ß-structure[83] (table 17). The width of the peak implies that there was a variety of structures present. The amide II peak (1590cm^{-1}) is attributable to hydrogen bonded structure.[83] Except for an amide II shoulder (1560cm^{-1}) and a sharper amide I peak, the methanol TFA.H.leu(ala)$_5$.OH infrared spectrum (table 17) is similar to the TFA.H.(ala)$_5$leu.OH spectrum (table 16). The TFA.H.leu(ala)$_5$.OH therefore formed one predominant structure different from the TFA.H.(ala)$_5$leu.OH.

TFA.H.(ala)$_5$ile.OH and TFA.H.ile(ala)$_5$.OH did not produce usable amide I and II spectra in methanol.

The TFA.H.leu(ala)$_5$.OH ethanol amide I peak (1660cm^{-1}) can be attributed to random coil[59] or α-helix[66,77,79,62] (table 16). The amide II peak (1550cm^{-1}) could be due to α-helix,[63,77] ß-pleated sheet[79] or ß-turns.[59] However there are no shoulders in the spectrum to confirm the presence of these structures. The sharpness of the amide I band implies that a definite structure and not random coil was present. The natures of these hydrogen bonded structures are unknown.

The TFA.H.leu(ala)$_5$.OH propanol and butanol amide I peaks (1620cm^{-1}) show the presence of intermolecularly molecular hydrogen bonded ß-structure (table 5).[62] The amide II peak (1590cm^{-1}) of the propanol solution indicates hydrogen bonding.[63] The analysis of the broad butanol amide II band is the same as that for ethanol. The dominant structures in propanol and butanol, as in methanol and ethanol, are therefore unknown. There is an increase in the number of amide I and II shoulders in propanol and butanol when compared to ethanol and thus an increase in small populations of other structures. These include possible antiparallel ß-sheet[66,49] and random coil.[77,59] These results are consistent with a possible hydrophobic pocket that is partially broken down by the decreased polarity of propanol and butanol.

In the TFA.H.(ala)$_n$leu.OH ethanol spectrum (table 17), the amide II peak (1530cm^{-1}) and the broad amide I band and 1690cm^{-1} shoulder[66,49] can be assigned to antiparallel ß-sheet[64] and/or random coil[77,59,60,79]. In the propanol and butanol spectra (table 17) the amide I peaks (1620cm^{-1} and 1630cm^{-1} respectively) are sharp. They are accompanied by 1690cm^{-1} shoulders and can be attributed to antiparallel ß-sheets.[49,60] This is confirmed in butanol by the amide II peak (1530-1540cm^{-1}).[79] In propanol the amide II peak (1590cm^{-1}) is due to hydrogen bonded structures.[63] The shoulders (1660cm^{-1},1540cm^{-1}) are due to random coil.[77] The

results suggest that antiparallel ß-sheets were present in the alcohol solutions with random coil. This is in keeping with the "statistical coil" implied by the CD spectra. The ethanol, propanol and butanol infrared spectra of TFA.H.(ala)$_5$ile.OH (table 19) amide I peaks (centred at 1620cm^{-1}) indicate that extended intermolecularly hydrogen bonded ß-sheets were present.[65-68,60] The amide II bands (between 1550cm^{-1} and 1560cm^{-1}) are attributable to ß-pleated sheet.[79] Amide I 1690cm^{-1} shoulders are absent from the spectra, implying that sheets were not antiparallel.[79] In ethanol the amide I band is broad indicating a variety of structures. The amide I and II bands indicate the possibility of parallel ß-sheets[49] or random coil.[77] Only very large insoluble parallel sheets are stable[6] and are an unlikely possibility. The predominant structures in ethanol were therefore ß-structures and random coil. The amide I peaks in the propanol and butanol spectra are sharper than in ethanol (table 19) meaning an increase in the relative sizes of the ß-structure population. Each spectra contained shoulders caused by small populations of different conformations which can be assigned structure according to table 3. The results seem to be contradictory as there is extended, intermolecularly hydrogen bonded ß-structure present but no evidence of antiparallel sheets. This supports the CD results which suggested that the TFA.H.(ala)$_5$ile.OH formed "single extended ß-chain".[79]

The amide I peaks of TFA.H.(ala)$_5$leu.OH (table 17) and TFA.H.(ala)$_5$ile.OH (table 18) both show a similar trend. The amide I band in ethanol is wide and starts at 1660cm^{-1}. In propanol and butanol the peaks narrow and are shifted towards 1630cm^{-1}. This is a shift away from random coil[77] towards ß-structure[63] as the alcohol aliphatic chain increases in length. These changes correspond to an increase in the π-π^*/n-π^* ratios of the CD transitions (figure 34) and may be caused by a decrease in the size of ß-sheets[79]. This would be caused by the lowering of the solvent polarity and therefore increased peptide solvation.

The sharp TFA.H.ile(ala)$_5$.OH ethanol amide I peak (1610cm^{-1}) is due to hydrogen bonded carbonyl groups (table 18). The broad amide II peak covers a wide variety of structure.[63,77,49,79] There are no 1690cm^{-1} amide I or 1516cm^{-1} amide II shoulders, which rules out antiparallel ß-sheet[63,49,79] and α-helix.[44,77] ß-turns are ruled out by the amide I band.[89] The peptide therefore formed a hydrogen bonded structure in ethanol; the sharpness of the amide I peak implies a limited variety of conformations. There were small populations of other structures, including random coil,[77,89] indicated by the shoulders. The amide I bands in propanol (1620cm^{-1}) and butanol (1640cm^{-1}) are attributable to intermolecularly hydrogen bonded ß-structures. The amide II peaks (1530cm^{-1}-1550cm^{-1}) and 1690cm^{-1} amide I shoulders are evidence for antiparallel ß-sheets[63,49,79].

Table 16 TFA.H.Leu(ala)$_5$.OH Alcohol Solution Amide I and II Bands

Amide Band	Methanol 1/cm	Ethanol 1/cm	Propanol 1/cm	Butanol 1/cm	25% TFE 1/cm	50% TFE 1/cm	75% TFE 1/cm	TFE 1/cm
I	1690s	1700s	1700s	1690s	1640s	1690s	1690s	1695
	1670s	1680p	1690s	1660s	1620s	1670s	1680s	1670s
	1660p		1660s	1640s	1600+p	1650s	1660s	1660s
	1630s		1640	1620+p		1640s	1620p	1630p
			1620+p			1620–		
						1600+p		
II	1590+p	1540–	1590+p	1570s	1590–	1600–	1590–	1590s
	1560s	1550p	1570s	1550–	1540+p	1570+p	1570+p	1535–
			1550s	1530+p	1530s	1550s	1550s	1525p
			1540s	1520s		1500s	1540s	
V		700						

Table 17 TFA.H.(ala)$_5$leu.OH Alcohol Solution Amide I and II Bands

Amide Band	Methanol 1/cm	Ethanol 1/cm	Propanol 1/cm	Butanol 1/cm	25% TFE 1/cm	50% TFE 1/cm	75% TFE 1/cm	TFE 1/cm
I	1690s	1690s	1690s	1690s	1690s	1690s	1690s	1690s
	1670s	1660–	1660s	1680s	1670s	1670s	1670s	1670s
	1660–	1610p	1640s	1660s	1660–	1660p	1660s	1660s
	1640p		1620+p	1640s	1610p	1640s	1630s	1640s
	1630s		1610s	1630+p		1620+p	1610+p	1630p
				1610s				
II	1590p	1530p	1590+p	1550s	1590+p	1570+p	1580+p	1530p
	1580s	1520s	1570s	1540–	1560s	1550s	1550s	1530s
			1550s	1530p	1530s	1540s	1540s	
			1540s	1520s	1520s			
					1510s			

Key

+ high intensity p peak

w low intensity s shoulder

Table 18 TFA.H.Ile(ala)$_5$.OH Alcohol Solution Amide I and II Bands

Amide Band	Ethanol 1/cm	Propanol 1/cm	Butanol 1/cm	25% TFE 1/cm	50% TFE 1/cm	75% TFE 1/cm	TFE 1/cm
I	1680ws	1690s	1690s	1630p	1660s	1690s	1690s
	1670s	1680s	1680s		1640s	1670s	1670s
	1660s	1660s	1660s		1630p	1660s	1660s
	1610p	1640s	1640+p			1640s	1640–
		1620+p				1630–	1600+p
						1600+p	
II							
	1600–	1550+p	1570s	1580–	1600–	1580+p	1590+p
	1530p	1530s	1550–	1550+p	1570+p	1560s	1560s
			1530p		1550s	1550s	1550s
			1520s		1540s		
V					660wp		

Table 19 TFA.H.(ala)$_5$ile.OH Alcohol Solution Amide I and II Bands

Amide Band	Ethanol 1/cm	Propanol 1/cm	Butanol 1/cm	25% TFE 1/cm	50% TFE 1/cm	75% TFE 1/cm	TFE 1/cm
I	1680ws	1670s	1680s	1630–	1660s	1660s	1680s
	1670s	1660s	1660s	1600p	1640s	1640s	1670s
	1660–	1640s	1630–		1620p	1620+p	1660–
	1620+p	1620+p	1620+p				1600p
II	1550p	1560+p	1550–	1580–	1600–	1590–	1590+p
		1550s	1530p	1540+p	1560+p	1550+p	1560s
		1530s	1500s		1550s	1540s	1550s
					1540s	1530s	

4.4.2.2 Trifluoroethanol/ethanol titrations results

In the TFE/ethanol titration spectra, the amide II peaks of TFA.H.$(ala)_5$leu.OH (tables 17) and TFA.H.$(ala)_5$ile.OH (table 19) are all above 1550cm^{-1} and have a high intensity. This can be attributed to a predominance of hydrogen bonded carbonyl groups.[63] In 25% TFE, the peaks are broad and include the possibility of ß-pleated sheet.[79] The amide I peaks are between 1600cm^{-1} and 1630cm^{-1}. Values below 1620cm^{-1} are due to hydrogen bonded carbonyl groups and near 1630cm^{-1} they are due to ß-sheets.[63,49] The TFA.H.$(ala)_5$leu.OH spectra have 1690cm^{-1} amide I shoulders, due to antiparallel ß-sheet,[79] not seen in the TFA.H.$(ala)_5$ile.OH. The numerous other amide I and II shoulders of the two peptides can be assigned structure using table 3. The 1660cm^{-1} and 1550cm^{-1} shoulders are due to random coil.[77,59]

In the TFA.H.$(ala)_5$leu.OH TFE titration infrared spectra, the amide I and II peak frequencies shift from 1600cm^{-1} and 1590cm^{-1} to 1630cm^{-1} and 1560cm^{-1} respectively as the TFE concentration increases (table 17). The shifts are associated with a decrease in the strength of intermolecular hydrogen bonding.63 The 75% TFE solution is an exception to this. This coincides with an increase of the π-π^*/n-π^* CD transition ratio in 75% TFE (figure 37). Therefore in this solution there is an increase in the hydrogen bond strength associated with a possible decrease in ß-sheet size.[79] The increased hydrogen bond strength may be a result of the

peptide-TFE bonding and would indicate that the TFE solvated peptide chains and disrupted the ß-sheet.

The TFA.H.(ala)$_5$ile.OH amide I and II peaks show no trends in their frequencies with changes in the TFE concentrations. There is an increase in the intensity of the 1620cm^{-1} peak in the 75% TFE/ethanol solution when compared to the 50% TFE solution. This can be attributed to an increase in the population of intermolecularly hydrogen bonded ß-conformations.[63,49] The increased intensity is associated with an increase in the π-π^*/n-π^* CD transition ratios (figure 37). Therefore the causes of the change in the ratio are different from those in TFA.H.(ala)$_5$leu.OH.

In the TFA.H.leu(ala)$_5$.OH TFE/ethanol titration spectra (table 16), the amide I peaks are between 1600cm^{-1} and 1620cm^{-1} and the amide II peaks are above 1560cm^{-1}. This indicates hydrogen bonded carbonyl groups.[63] In TFE the peak is at 1630 and would normally be attributed to intermolecularly hydrogen bonded ß-structure.[63] However as TFE disrupts these interactions the source of the peak is presumably due to some other structure. There are amide I shoulders at 1690cm^{-1} and between 1630cm^{-1} and 1640cm^{-1}, indicating small populations of antiparallel ß-sheet; supported by amide II shoulders near 1530cm^{-1}.[79] The 50% TFE solution amide I and II peaks are broad, implying a variety of structures. These results, with the possible exception of the TFE solution, do not support the idea from the CD

spectra (figure 35) that "statistical coils" were present in the TFE titrations as there is scant evidence of ß-structure. Instead the results may b showing that, as the TFE concentration increases there is solvation of a hydrophobic pocket with the result that the variety of structure present increases.

The amide I and II peaks of the TFA.H.ile(ala)$_5$.OH TFE/ethanol titrations (table 18) are due to intermolecularly hydrogen bonded ß-structure.[83] The 75% TFE and pure TFE 1690cm^{-1} shoulders in conjunction with the amide I peaks imply that antiparallel ß-sheet was present.[66,69,79] There are amide II shoulders at 1550cm^{-1} indicating ß-pleated sheet[79] but no amide II shoulders to confirm the chain orientations.[79] Except for the 25% TFE each spectra showed a variety of structures, indicated by the large number of shoulders, which can bc assigned according to table 3. The structures include random coil[77,89] which is in agreement with the CD curves that implied the presence of "statistical coils" of random coil and ß-structure. The amide II peaks show an increase in frequency as the TFE concentration increases. This implies that the hydrogen bonds formed became stronger. TFE disrupts intermolecular hydrogen bonding and the effect observed must therefore be due to solvent-peptide or intramolecular hydrogen bonds.

4.4.3 Infrared peak intensities

Tables 20 to 23 show the intensities of the major peaks in the solution phase infrared. Included in the tables are those peaks attributed to the amino terminal chromophores, to free carbonyl groups and the amide I intensities at 1620cm^{-1} and 1600cm^{-1}. The amide I values were chosen because the former lie in the region due to free carbonyl groups and the latter are due to hydrogen bonded groups.[63] The ratios of these two are shown for the TFE/ethanol titration series. The intensities were measured in each spectrum from an absorption minimum common to all at 2000cm^{-1}.

In the TFE/ethanol titrations of TFA.H.(ala)$_5$leu.OH, TFA.H.ile(ala)$_5$.OH and TFA.H.(ala)$_5$ile.OH, the 1900cm^{-1} peak intensities decrease as the TFE concentration increases (tables 21, 22 and 23) and there are no 1920cm^{-1} peaks. This implies some similarity in the behaviour of these three peptides. The 1900cm^{-1} trend is not seen in the TFA.H.leu(ala)$_5$.OH TFE series (table 20) and it has 1920cm^{-1} peaks implying that its behaviour was different from the other three peptides. In the series of alcohol solutions, both the TFA.H.(ala)$_5$leu.OH and TFA.H.(ala)$_5$ile.OH 1920cm^{-1} peak intensities decrease from ethanol to butanol. This trend does not occur in the TFA.H.leu(ala)$_5$.OH or TFA.H.ile(ala)$_5$.OH. The CD results for TFA.H.(ala)$_5$leu.OH and TFA.H.(ala)$_5$ile.OH both showed strong β-structures in the alcohol solution series. TFA.H.ile(ala)$_5$.OH was found to

form "statistical coils" with a predominance of random coil and TFA.H.leu(ala)$_5$.OH formed undetermined structures. The differences in the structures seen in the CD results are therefore supported by the patterns of infrared peak intensities.

The ratios of the 1600cm^{-1} to 1620cm^{-1} peak intensities should give an indication of the relative numbers of hydrogen bonded to non-hydrogen bonded carbonyl groups (tables 20, 21, 22 and 23). TFE is expected to increase the amount of solvent-peptide hydrogen bonding.[63] The TFA.H.ile(ala)$_5$.OH and TFA.H.leu(ala)$_5$.OH ratios increase (tables 22 and 29) as the TFE concentration increases. This shows that there is an increase in peptide-TFE hydrogen bonding and therefore the peptide molecules were well solvated by TFE. This is consistent with the idea that TFA.H.leu(ala)$_5$.OH forms a hydrophobic pocket in ethanol which is disrupted by the TFE solvation.

Table 20 TFA.H.Leu(ala)$_2$.OH Alcohol Solution Infrared Peak Intensities

	Amino terminal stretch			Free carbonyl stretch					
Solution	2400-2450cm %abs	2200-2250cm %abs	2050-2100 %abs	1960cm %abs	1920 cm %abs	1900cm %abs	1880cm %abs	1800cm %abs	1760cm %abs
Methanol	50	36	54					4	16
Ethanol	51	21				28			
Propanol	32	21	17		16				
Butanol	64	17	23	7	7		3		
TFE/Ethanol titration									
25% TFE	15	24	26	3		5		14	15
50% TFE	36	28	27	20		23		14	23
75% TFE	41	31				14		17	23
TFE	41	34				30		14	20

	Amide I		
Solu	1680cm %abs	1620-1600cm %abs	Ratio of amide I peaks
Methanol	60	72	
Ethanol	40	58	
Propanol	39	64	
Butanol	43	56	
TFE/Ethanol titration			
25% TFE	24	27	1.1
50% TFE	31	48	1.5
75% TFE	25	35	1.4
TFE	16	25	1.6

Table 21 TFA...(ala)$_n$leu.OH Alcohol Solution Infrared Peak Intensities

Solution	Amino terminal stretch			Free carbonyl stretch					
	2400-2450cm %abs	2200-2250cm %abs	2150-2100 %abs	1980cm %abs	1920 cm %abs	1900cm %abs	1860cm %abs	1800cm %abs	1760cm %abs
Methanol	17	31	26					4	16
Ethanol	20	13	12		17				14
Propanol	35	23	18		16			4	12
Butanol	61	17	23	7	8		2	5	8
TFE/Ethanol titration									
25% TFE	15	24	26	3		5		14	15
50% TFE	42	30	26	15		25		16	24
75% TFE	42	31	29	18		14		16	22
TFE	25	20	18	16			3	12	17

Solution	Amide I		
	1680cm %abs	1620-1600cm %abs	Ratio of amide I peaks
Methanol	53	65	
Ethanol	23	31	
Propanol	53	68	
Butanol	44	75	
TFE/Ethanol titration			
25% TFE	54	71	1.3
50% TFE	28	47	1.7
75% TFE	20	25	1.3
TFE	16	24	1.5

Table 22 TFA.H.ile(ala)$_9$.OH Alcohol Solution Infrared Peak Intensities

Solution	Amino terminal stretch: 2400–2450cm %abs	2200–2250cm %abs	2050–2100 %abs	Free carbonyl stretch: 1950cm %abs	1900cm %abs	1860cm %abs	1800cm %abs	1760cm %abs
Methanol		16	22					
Ethanol	55	37	29		46	14		34
Propanol	35	17	14					
Butanol	53	21	23	8	3	1		
TFE/Ethanol titration								
25% TFE	33		27	13	31		13	23
50% TFE	38	27	28	16	12		16	21
75% TFE	35	25	30	12	7		10	19
TFE	46	29	29	16		5	10	17

Solution	Amide I: 1660cm %abs	1620–1600cm %abs	Ratio of amide I peaks
Methanol	58		
Ethanol		48	
Propanol		40	
Butanol		91	
TFE/Ethanol titration			
25% TFE	22	22	1.0
50% TFE	22	38	1.7
75% TFE	30	50	1.7
TFE	8	17	2.1

Table 23 TFA.H.(ala)$_5$ile.OH Alcohol Solution Infrared Peak Intensities

	Amino terminal stretch			Free carbonyl stretch					
Solution	2400-2450cm %abs	2200-2250cm %abs	2060-2100 %abs	1980cm %abs	1920 cm %abs	1900cm %abs	1860cm %abs	1800cm %abs	1760cm %abs
Methanol		28	40			30			
Ethanol	38	31	33		26				18
Propanol	34	17	14		15				5
Butanol	56	45	21	5	4	3	1	3	5
TFE/Ethanol titration									
25% TFE	42	32	26	11		30		14	27
50% TFE	36	31	25	12		25		14	23
75% TFE	36	29	26	14		13		14	20
TFE	19	10	21	19		2	5	6	12

	Amide I		
Solution	1680cm %abs	1620-1600cm %abs	Ratio of amide I peaks
Methanol			
Ethanol	78		
Propanol	21	48	
Butanol	36	69	
TFE/Ethanol titration			
25% TFE	25	30	1.2
50% TFE	26	42	1.7
75% TFE	19	27	1.4
TFE	54	68	1.3

4.5 The Raney nickel desulphurisation of L-cysteine

In absolute ethanol there was no evidence of desulphurisation of the s-benzyl-L-cysteine, probably due to solubility problems. Several solvents were tried and the most successful was a 5% ammonia in water solution with a 2 times excess of Raney nickel.[91] The Rf values of the products of the desulphurisation, on TLC plates eluted with n-butanol/acetic acid/water (60:20:20 v/v),[93] showed an intense alanine spot and a weaker cysteine spot (table 24). The products appeared as a blue powder due to amino acid complexes with nickel. The crude recovery was 60%. The published recovery for pthaloyl-amino acid desulphurisations was 40%.[91] The desulphurisation of t-Boc-L-cysteine resulted in a product that turned ninhydrin blue implying that deprotection had occurred. There was thus nothing to be gained by pursuing a method for t-Boc amino acid desulphurisation.

The route to be taken for producing deuterated t-Boc-L-alanine was therefore to desulphurise S-benzyl-L-cysteine with deuterated Raney nickel, to separate the L-alanine from the reaction mixture and to bocylate the pure L-alanine.[64] The next, and unfinished step, is the development of a method for separating residual L-cysteine from the L-alanine produced, possibly on an ion exchange column using the different pK values of the amino acids.[22]

Table 24 Raney Nickel Desulphurisation of S-Benzyl-L-Cysteine in Ammoniated Water.
The product of the desulphurisations were run on TLC plates using a 5.5/3/1.5 n-butanol/acetic acid/water solvent system. The plates were run and then sprayed with ninhydrin and developed in an oven at 50°C.93using a 5.5/3/1.5 n-butanol/acetic acid/water solvent system. The plates were run and then sprayed
with ninhydrin and developed in an oven at 50°C.[93]

Sample	
Standards	Rf value
L-alanine	0.3
L-cysteine	0.4
S-benzyl-L-cysteine	
Raney Nickel S-benzyl-L-cysteine desulphurisation	
	0.3
	0.4

The intensity of the product Rf value, 0.3, was much lower than that at 0.4.

5. CONCLUSIONS

There is a wide range of possible ß-structure molecular variations and for this reason the extrapolation of ß-structure results has been shown to be unreliable.[79] The results of the work reported here show this to be true. The results also show that the position and type of amino acid in a sequence and the solvent all affect secondary structure. There is thus a balance between position, structural propensity and environment.

TFA.H.(ala)$_5$leu.OH and TFA.H.(ala)$_5$ile.OH formed more stable ß-sheets than their N-terminally substituted homologues in alcoholic solution. Isoleucine has been reported to be a weaker ß-sheet former in the solid phase than leucine.[50,58] The results show that this is not the case in alcoholic solutions for these peptides as in alcoholic solut TFA.(ala)$_5$ile.OH formed more stable ß-structure than TFA.(ala)$_5$leu.OH. This may be due to steric problems with the incorporation of isoleucine residues into extended aggregated sheets. Possibly thus TFA.(ala)$_5$ile molecules formed smaller ß-structures or even "single extended ß-chains" while TFA.(ala)$_5$leu.OH formed larger intermolecularly hydrogen bonded sheets. The sheets could have been solvated and disrupted by TFE; a process that would have left single chains unaffected. The increased ß-sheet forming ability of the C-terminally substituted peptides, over their N-terminally substituted counterparts,

is best shown by TFA.H.ile(ala)$_5$.OH and TFA.H.(ala)$_5$ile.OH. TFA.H.ile(ala)$_5$.OH formed "statistical coils" with a dominance of random coil while TFA.H.(ala)$_5$ile.OH formed the strongest ß-structures. Therefore TFA.H.ile(ala)$_5$.OH did not benefit from the ß-forming propensity of the isoleucine residue because of its position. The increased stability of the C-terminal susbtitution is in agreement with previous work.[62] It has been postulated that the structural propensity of the amino acids in a sequence wholly determines the secondary structure.[27] The results reported for this work show that the position of the amino acid is important as well.[40]

The balance of amino acid position and structural propensity was further shown by the TFA.H.leu(ala)$_5$.OH and TFA.H.ile(ala)$_5$.OH results. Hexa-L-alanine was found to form ß-sheets in a variety of solvents[62]. Both L-leucine and L-isoleucine at the C-terminal reinforced the ß-structure while TFA.H.leu(ala)$_5$.OH completely disrupted it. TFA.H.leu(ala)$_5$.OH behaved differently from the other three peptides. The CD results were similar to those of previous work with TFA.H.val(ala)$_5$.OH[62] and it is possible that TFA.H.leu(ala)$_5$.OH forms a hydrophobic pocket in polar solvents similar to that found in TFA.H.val(ala)$_5$.OH.62 TFA.H.ile(ala)$_5$.OH disrupts the ß-structure of hexa-L-alanine but forms random coil, not a pocket. The hydrophobic pocket in TFA.H.val(ala)$_5$.OH is stabilised by an

interaction between the isopropyl side chain of the L-valine at position 1 and the L-alanine methyl group at position 5. There is a hydrogen bond between the carbonyl of residue 3 and the amine of 6 (Dr. A. Nel, submitted for publication in Biopolymers, 1991). A model shows that such an interaction is possible with an L-leucine side chain but there are steric problems with an L-isoleucine side chain. This explains why the TFA.H.ile(ala)^5.OH did not behave in the same way as TFA.H.leu(ala)5.OH.

The hydrophobic pocket has been suggested by Matheson and Scheraga[40] as a prime candidate for initiation of protein folding. The spontaneous imposition of order in a structure implies a negative free energy. The Gibbs free energy equation ($\Delta G=\Delta H-T\Delta S$) shows that the resulting decrease in entropy, from the imposition of order, works against this. The hydrophobic pocket brings apolar groups together allowing London forces between them[32] and it brings the peptide bond amines and carbonyls together for hydrogen bonding. The free energy released from these electrostatic interactions provide $-\Delta H$ that may drive the process of initiation.[32] Proteins are synthesised at the ribosomes from the amino terminal and it is logical to expect that initiation structures would be directed by amino acids at the N-terminal. Thus the finding that TFA.H.leu(ala)$_8$.OH, an N-terminally substituted guest-host peptide, possibly forms a hydrophobic pocket supports this theory.

6.REFERENCES

1. Fischer, E. (1917). Z. Physiol. Chem.,

2. Anfinsen, C.B. (1973). "Principles That Govern the Folding of Polypeptide Chains", Science, 181, 223-230.

3. Rose, G.D., Gierasch, L.M., and Smith, J.A. (1985). Advances in Protein Chemistry, 37, 1-109.

4. Dickerson, R.E., and Geiss, I. (1969). Structure and Action of Proteins, 1-43, Benjamin Cummings Publishing Company, New York.

5.Edsall, J.T., Florey, P.J., Kendrew, J.C., Liquori, A.M., Nemethy, G., and Scheraga, H.A. (1966) J. Mol. Biol. 399.

6. Richardson, J.S. (1981). Advances in Protein Chemistry, 34, 167-363.

7. Ramachandran, G.N., and Sasisekharan, V. (1968). Advances in Protein Chemistry, 23, 283-437.

7a) Ramachandran, G.N., ed., in Aspects of Protein Structure, Academic Press, New York, 1963.

8. Ramachandran, G.N., and Mitra, A.K. (1976). Jnl. Mol. Biol., 107, 85-92.

9. Ramachandran, G.N., Lakshminarayanan, A.V., and Kolaskar, A.S. (1973). Biochimica et Biophysica Acta, 303, 8-13.

10. Schellman, J.A., and Schellman, C. (1976). The Proteins, Vol. 11, edited by Neurath, H., 2-111. Academic Press, New York.

11. Kendrew, J.C., Bodo, G., Dintzis, H.M., Parrish, R.G., Wyckhoff, H., and Phillips, D.C. (1958). Nature, 181, 662-666.

12. Pauling, Corey, R.B., and Branson, H.R. (1951).Proc. Nat. Acad. Sci. USA, 37, 205-211.

13 Pauling, L., Corey, R.B., and Branson, H.R. (1951). Proc. Nat. Acad. Sci. USA
, 37, 729-740.

14.Hagler, A.T., Moult, J. (1978) Nature, 272, 222-226.

15. Kabsch, W., and Sanders, C. (1983). Biopolymers, 22, 2577-2637.

16. Donohue, J. (1953). Proc. Nat. Acad. Sci., 39, 470-478.

17. De Gardo, W.F., and Lear, J.D. (1985). J. Am. Chem. Soc., 107, 7684-7689.

18. Venkatachalam, C.M. (1968). Biopolymers, 6, 1425-1436.

19. Nemethy, G., and Scheraga, H.A., Biochem. and Biophys. Res. Comm. (1980). 95, 320-327.

20. Edison, A. (1990). TIBS, 15, 216-217.

21. Mathews, K.M., and van Holde, K.E. (1990). Biochemistry, 1st edition, Benjamin Cummings Company, New York.

22. Lehninger, A.L. (1975). Biochemistry, 2nd edition, 1-122, published by Worth Publishers Inc., New York.

23. Levitt, M. (1978). Biochemistry, 17, 4277-4285.

24. Merrifield, R.B. (1965). Science, 150, 178-185.

25. Levitt, M., and Greer, J. (1977). J. Mol. Biol., 114, 181-293.

26. Liquori, A.M., DeSantis, P., Kovacs, A.L.L., and Mazzarella, L. (1966). Nature, 211, 1039.

27. Chou, P., and Fasman, G.D. (1978). Annual Review of Biochemistry, 47, 251-276.

28. Chou, P.Y., and Fasman, G.D. (1974). Biochemistry, 13(2), 222-245.

29. Privalov, P.L., and Gill, S.J., (1988). Advances in Protein Chemistry, 39, 191-234.

30. Rose, G.D., and Siddhartha, R. (1980). Proc. Nat. Acad. Sci., 77(8), 4643-4647.

31. Kauzmann, W. (1959). Advances in Protein Chemistry, 14, 1.

32. Burley, S.K., and Petsko, G.A. (1988). Advances in Protein Chemistry, 39, 125-189.

33. Ben-Niam, A. (1990). Biopolymers, 29, 567-596.

34. Baldwin, R.L. (1986). Proc. Nat. Acad. Sci., 65, 44-45.

35. Moore, J.W., Davies, W.G., and Collins, R.W. (1978). Chemistry, International Student Edition, published by McGraw-Hill International Book Company.

36. Tayyab, S., (1988). Biochemical Education, 16, 221.

37. Ogushi, M., and Wada, A., (1989). FEBS LETTS, 164, 21-24.

38. Anfinsen, C.D., (1972). Biochemical Jnl., 128, 737-749.

39. Baldwin, R.L. (1989). TIBS, 14, 291-294.

40. Haas, E., McWherter, H., and Scheraga, H. (1988). Biopolymers, 27, 1-21.

41. Levinthal, C. (1968). J. Chim. Phys., 65, 44-45.

42. Goodman, M., Toniolo, C. and Verdini, A.S. (1969). Peptides, North Holland Publishing Company, 207-221.

43. Semistonov, G.V., Rodionova, N.A., Kutyshenko, V.P., Ebert, B.B., Larick, J., and Ptitsyn, O.B. (1987). FEBS LETTS, 224, 9-13.

44. Katchalski, E., Sela, M., Silman, H.I., and Berger A. (1964). The Proteins, Vol. II, 405-542, edited by Neurath, H., Academic Press, New York.

45. Wright, P.E., and Dyson, H.J., and Lerner, R.A. (1989). Biochemistry, 27, 7167-7175.

46. Montelione, G.T, and Scheraga, H.A. (1989). Acc. Biochem. Res., 22, 70-76.

47. Richardson, J.S. (1976). Nature, 261, 552-558.

48. Matheson, R.R., and Scheraga, H.A. (1978). Jnl. Am. Chem. Soc., 4, 819-828.

49. Bonaro, G.M., Palumbro, M., Toniolo, C., and Mutter, M. (1979). Makromolecular Chem., 180, 1293-1304.

50. Brandts, J.F., Halvorson, H.R., and Brennan, M. (1975). Biochemistry, 14, 4953-4963.

51. Grathwohl, C., and Wutrich, K. (1981). Biopolymers, 20, 2623-2633.

52. Crippen, G.M., and Snow, M.E., (1990). Biopolymers, 29, 1479-1489.

53. Chaitin, G. (1990). New Scientist, 24th March, 44-46.

54. Booyens, J. (1990). Proceedings of the University of Pretoria 50th Annivesary Conference, p2.

55. Goodman, M., Toniolo, C., and Nadier F. (1974). Peptides, Polypeptides and Proteins, 308-319, edited by Blout, E.R., Bovey, F.A., Goodman, M., and Lotan, N.L., Wiley Interscience.

56. Bodanszky, M., and Martinez, J. (1981). Synthesis, 333-356.

57. Goodman, M., Rupp, R., and Nadier, F. (1971). Bioorganic Chemistry, 1, 310-328.

58. Mutter, M., and Altmann, K.H. (1985). Int. Jnl. Pep. Res., 26, 373-380.

59. Kawai, M., and Fasman, G.D. (1978). J. Am. Chem. Soc., 100(1), 3630-3631.

60. Toniolo, C., Bonaro, G.M., Palumbro, M., Peggion, E., and Stevens, E.S. (1978). Biopolymers, 17, 1713-1727.

61. Kemp, W. (1986). NMR In Chemistry, A Multinuclear Introduction, 1st edition, Macmillan Education ltd, London.

62. Nel, A., (1990). Synthesis and Conformational Analysis of Series of Short Peptides, doctoral thesis, University of the Witwatersrand.

63. Kent, S.B.H., Mitchell, A.R., Engelhard, M., and Merrifield, R.B. (1979). Proc. Natl. Acad. Sci. U.S.A., 76, 2180.

64. Stewart, J.S., and Young, J.D. (1984). Solid Phase Peptide Synthesis, Pierce Chemical Company Co., Rockford, Illinois, U.S.A.

65. Van Regenmortel, M.H.V., Briand, J.P., Muller, S., and Plaue, S. (1988). Synthetic Peptides as Antigens, 1st edition, Elsevier.

66. Kent, B.H., Mitchell, A.R., Barany, G., and Merrifield, R.B.(1978). Anal. Chem., 50, 155-159.

67. Gisin, B.F. (1972). Anal. Chim. Acta, 58, 248-249.

68. Kaiser, E., Colescott, R.L., Bossinger, C.D., and Cook, P.I. (1970) Anal. Biochem., 34, 595-598.

69. Lenard, J., and Robinson, A.B. (1967). J. Am. Chem. Soc., 89(1), 181-182.

70. Imoto, T., and Yamada, H. (1983). Molecular and Cellular Biochemistry, 51, 111-121.

71. Dunlap III, C.E., Gentleman, S., and Lowney, L. (1978). J. of Chromatography, 160, 191-198.

72. Silverstein, R.M., Bassler, G.C., and Morill, T.C. (1981). Spectrometric Identification of Organic Compounds, 4th edition, John Wiley and Sons.

73. Wuttrich, K. (1986). NMR of Proteins and Nucleic Acids, Wiley Interscience, New York.

74. Jirgensons, B. (1973). Molec. Biol. Biochem. and Biophys., Optical Activity of Proteins and other Macromolecules, 2nd edition, Chapman and Hall Ltd, London.

75. Adler, A.J., Greenfield, N.J., and Fasman, G.D. (1973). Methods in Enzymology, 27, 675-735.

76. Greenfield, N.J., and Fasman, G.D. (1969). Biochemistry, 8(10), 4108-4115.

77. Saudek, V., Stokrova, S., Schmidt, P. (1982). Bioplymers, 21, 2195-2203.

78. Blout, E.R., (1974). Polypeptides and Proteins, 352-357, edited by Blout, E.R., Bovey, F.A., Goodman, M., and Lotan, N.L., Wiley Interscience, New York.

79. Kubota, S., and Fasman, G.D. (1975). Bioplymers, 14, 603-631.

80. Applequist, J. (1982). Bioplymers, 21, 779-795.

81. Azzena, U., and Luisi, P.L. (1986). Bioplymers, 25, 555-570.

82. Woody, R. (1974). Peptides, Polypeptides and Proteins, edited by Blout, E.R., Bovey, F.A., Goodman, M., and Lotan, N.L., 339-349, Wiley Interscience, New York.

83. Mutter, M., Masen, F., Klein, B., Toniolo, C., and Bonora, G. (1983). Bioplymers, 22, 233-240.

84. Jose, C.I., Bandekar, A.A., and Agashe, M.S. (1987). Bioplymers, 26, 1315-1325.

85. Dwivedi, A.M., and Krimm, S. (1982). Bioplymers, 21, 2377-2397.

86. Palumbro, M., Rin, S.D., and Bonaro, G.M. (1976). Makromol. Chem., 177, 1477-1492.

87. Bandekar, J., and Krimm, S. (1988). Bioplymers, 27, 909-921.

88. Marki, W., Spiess, J., Tache, Y., Brown, M., Rivier, J.E. (1981). J. Am. Chem. Soc., 103, 3178-3185.

89. Robson, J. (1987). Synthesis and Conformational Studies of Hexapeptides, honours dissertation, University of the Witwatersrand.

90. Savitzky, A., and Golay, J.E. (1964). Anal. Chem., 36(8).

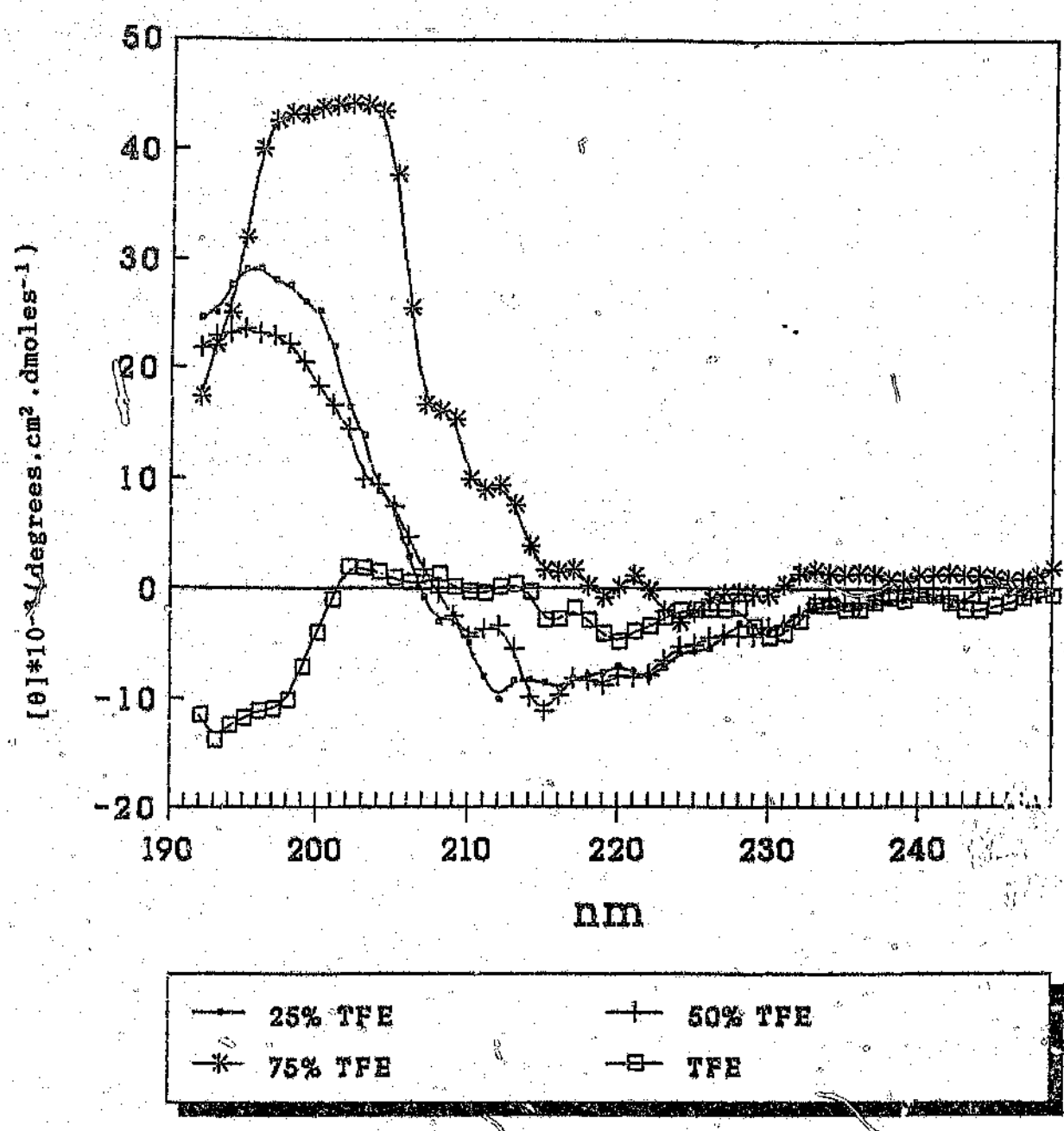
91. Balenovic, K., and Fles, D. (1952). J. Chem. Soc., 2, 2447.

92. Vogel, A.I. (1978). Textbook of Practical Organic Chemistry, 4th edition, Longmans, Green and Company, London.

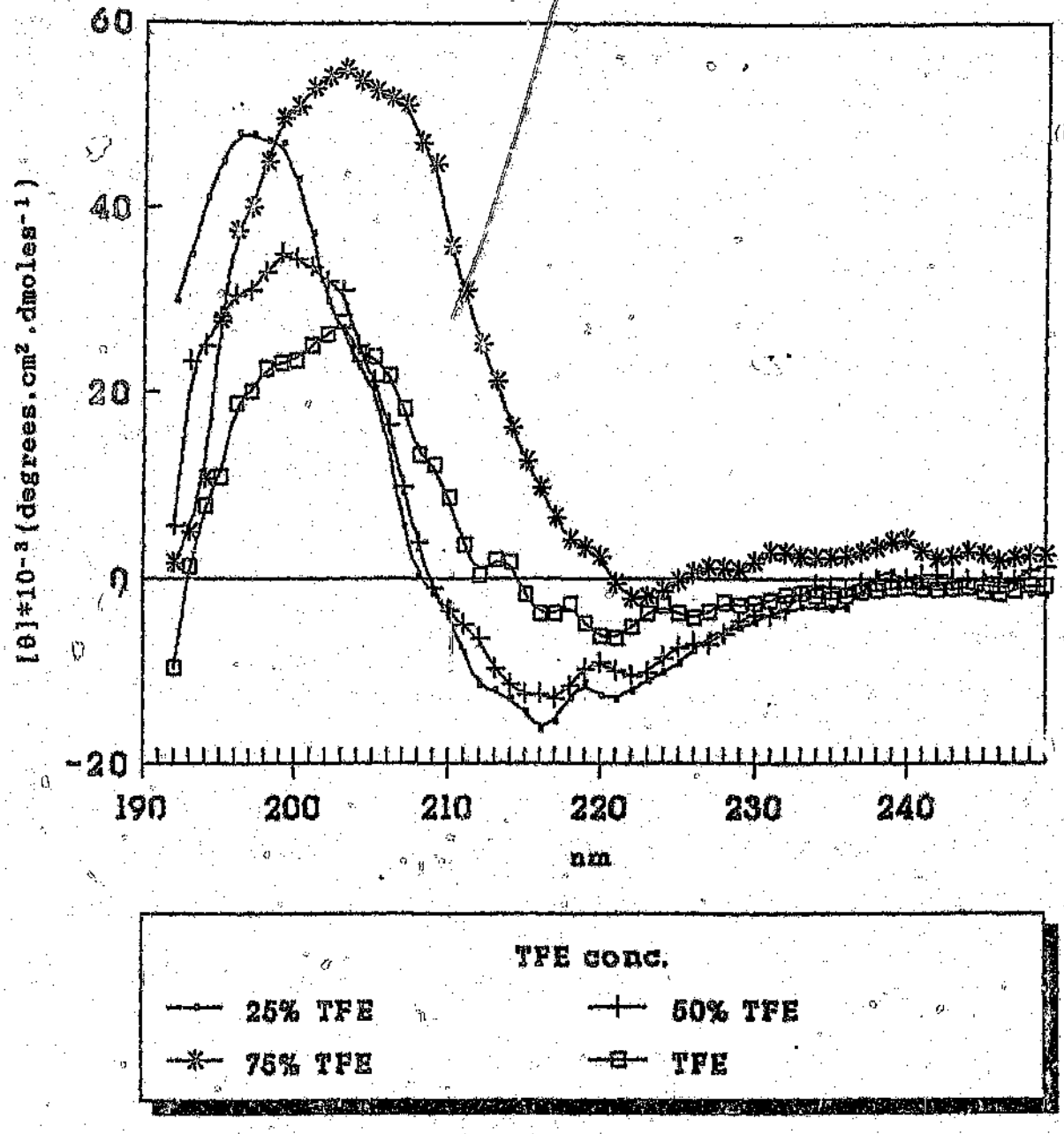
93. Randrath, K. (1963). Thin Layer Chromatography, 1st edition, translated by Libman, D.D., Academic Press, New York.

94. Bush, C.A., Sarkar, S.K., and Kopple, K.D. (1978). Biochemistry, 17(23), 4951-4954.
4951-4954.

7. APPENDIX



Appendix figure 1 CD spectra of TFA.H.(ala)$_5$leu.OH in TFE/ethanol mixtures.



Appendix figure 2 CD spectra of TFA.H.(ala)$_3$ile.OH in TFE/ethanol mixtures.

7.1 Hydrogen Fluoride Cleavage of the Peptide from the Resin

HF is highly corrosive, it reacts exothermically with water and is potentially lethal if inhaled. The HF treatment of the peptido-resin must therefore always be carried out in a fume hood and steps must be taken to ensure that the apparatus and peptido-resin is dry.[64] The apparatus material must be resistant to HF and must be able to withstand low temperatures. Reference 64 lists the commercially available material. Figure 38 shows the apparatus used for HF cleavage.[64] The method detailed refers to figure 38, the bracketed numbers refer to the stopcocks and the letters to the various vessels.

7.1.1 Drying and Distillation of the HF[57]

(1) Was closed and vessels (B) and (C) were isolated with (3) and (4) respectively. A teflon magnetic stirrer and a few grams of CoF_3 were added to (A). CoF_3 was the drying agent for the gas. (A) was screwed into place and (2) was opened to allow manometer readings of the pressure in the HF line. (7) was closed to ensure that the high vacuum pump was, under no circumstances, able to suck water from the tap vacuum. (6) and (5) were opened to allow the water vacuum to suck air and the tap was turned on fully. (5) was then turned so that the vacuum line and (A) were evacuated. Evacuation continued until the pressure was about 630mm of mercury.

(A) was immersed in dry ice and ethanol contained in a thermos. This was started 5 minutes before the HF was distilled over. (2) was turned so that (A) was open to the manometer but closed to the line and the reaction vessels ((B) and (C)). The system was checked for leaks by monitoring the manometer reading, which remained constant. The water pump was vented by turning (5) and then turned off. The HF cylinder was opened and HF was gently distilled over by opening (1) slowly. Enough HF was distilled over for two to three cleavages (about 10ml of HF per 0.2millimoles of peptide[64]). The vacuum was allowed to drop until a slight vacuum was left. The level of the dry ice/ethanol bath was maintained such that (A) remained immersed. (1) was then closed. The HF was ready for use.

(B) and (C) in figure 38 are the reaction vessels. The method will only deal with (B) as the same instructions apply for (C) as well. A teflon stirrer bar and anisole (1.0ml per 0.2millimoles of peptide) were placed into (B) and it was secured onto the HF line. (A) was placed into a 1 litre beaker containing water at room temperature. A magnetic stirrer was placed under this to stir the HF. (B) was then cooled for 5 minutes in a dry ice/ethanol thermos. As a precaution tap (7) was checked to ensure that the high vacuum pump was isolated. The water pump was turned on, (6) opened and (5) closed to the air. (3) was turned to open (B) to the vacuum line and (2) was turned to open the manometer

to (1) and close it to (A). Using the manometer to check, the tap vacuum ran until a strong vacuum was attained. The vacuum was not run for long as the anisole distils over. (5) was turned to isolate (B) and the water vacuum was swithched off. (3) was opened to allow passage of the HF between (B) and (A). (2) was turned slowly to allow the HF to distil across. The HF was not allowed to boil over and (2) was adjusted so that there was a steady drop in the vacuum. 10ml of HF per 0.2 millimoles of peptide were distilled across. (B) was then isolated with (3) and (A) was isolated with (2).

7.1.2 Cleavage of the Peptido-Resin

(B) was immersed in a beaker with crushed ice and water and a magnetic stirrer was used to stir the contents of (B). Periodically a torch was shone through (B) to ensure that the resin was suspended in the HF and that the anisole was melted. The resin was red in the presence of anisole; in the absence of anisole, caused by either its solidifying or distilling off, the resin turns blue.[64] A 45 minute reaction at 0°C was used.

At the end of the time (1) was closed, (5) was opened to connect the water vacuum and (4) and the line was evacuated. (3) was opened slowly to allow the HF to distil off. The resin was aspirated and stirred until it became a cream colour.

(D) was chilled in dry ice and ethanol and (5) was turned to allow air to be sucked into (F). The water was turned off and (6) closed. Once (D) was chilled (7) was opened and (5) turned an 1/8th of a turn to close off the line and access to the air. The high vacuum pump was turned on and (5) opened to the line. The line was then evacuated for 1 hour. (5) was then turned to vent the high vacuum pump and isolate the line. The high vacuum pump was turned off and (7) closed. The line was vented by slowly turning (5) to the air. This process cleared the line of HF and (B). The resin and peptide were removed and the anisole washed away in a sintered glass funnel with ethyl acetate. The line was washed with acetone and could then be used for the next cleavage.

Appendix figure 3 The HF line and apparatus used to cleave peptides, synthesised using t-Boc amino acids, from the resin. See section 7.1 for a discussion of the methods.[54]