A STUDY OF SECONDARY STRUCTURE PREDICTIVE METHODS FOR PROTEINS AND THE RELATIONSHIP BETWEEN PHYSICAL-CHEMICAL PROPERTIES AND ENZYMATIC ACTIVITY OF SOME ASPARTYL PROTEINASES

by

RICKEY YOSHIO YADA

B.Sc.(Agr.), The University of British Columbia, 1977
M.Sc., The University of British Columbia, 1980

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

Department of Food Science

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

May 1984

Department of __Food Science__

The University of British Columbia
1956 Main Mall
Vancouver, Canada
V6T 1Y3

Date __June 5, 1984__

ABSTRACT

In the first of two studies, four algorithms were assessed for their ability to accurately determine protein secondary structure from circular dichroism spectral data in relation to X-ray data. The majority of the algorithms examined showed highly significant ($P<0.001$) correlation coefficients for α-helix and β-sheet determination while the correlation coefficients using the simplex optimization algorithm with $\bar{n}$ = variable (in the presence of concanavalin A) were significant at $P<0.05$. None of the algorithms examined showed significant ($P>0.05$) correlation coefficients for β-turn determination. Significant ($P<0.05$) correlation coefficients were obtained for random coil determination from both the simplex optimization algorithm (in the absence of concanavalin A) and the algorithm of Chang et al. (1978). The simplex optimization algorithm with $\bar{n}$ = 10.4 compared favourably to the method of Chang et al. (1978) in the absence of concanavalin A, and may serve as an alternative algorithm to solve least squares. The algorithm of Provencher and Glöckner (1981) demonstrated the greatest versatility for secondary structure determination of the four algorithms examined.

In the second study, the relationship between physical-chemical properties and the enzymatic activity of aspartyl proteinases was investigated. Using the diagonal plot method, pepsin and chymosin showed the highest degree of primary sequence homology while active site regions were highly homologous between the aspartyl proteinases examined. Secondary structure prediction methods indicated that chymosin, pepsin, penicillopepsin and _Mucor_ _miehei_ proteinase had relatively high proportions

of β-sheet, with active site aspartic acid residues located in β-turn regions. Near-UV and far-UV CD spectral analysis indicated that changes in spectra occurred in the neutral to alkaline pH range. Secondary structure determination from far-UV CD spectral data demonstrated that the β-sheet fraction of the aspartyl proteinases decreased above pH 6.3. The milk-clotting to proteolytic activity ratio of the aspartyl proteinases decreased with increasing pH. Principal components derived from various structural and intrinsic parameters allowed for the classification of aspartyl and non-aspartyl proteinases. Regression of the milk-clotting to the proteolytic activity ratio on the principal components indicated that a high ratio may be partially dependent on relatively high hydrophobicity, β-sheet and low charge.

TABLE OF CONTENTS

PART I. Analysis of various algorithms for the determination of secondary structure from CD data

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

ACKNOWLEDGEMENTS

PART I. *Analysis of various algorithms for the determination of secondary structure from CD data*

INTRODUCTION

One of the most intriguing problems in protein chemistry has been the determination of protein structure. The structure of a protein is a major factor which, along with the hydrophobic and electronic parameters, aids in the prediction of functional properties (Stuper et al., 1979). X-ray diffraction analysis is the most accurate means for determining the three-dimensional structure of proteins, however, only a small number of proteins have been analyzed using this method due to constraints of time, cost and the inability of certain proteins to form suitable crystal forms.

Levitt and Chothia (1976) proposed the description of a protein's three-dimensional structure to include the classification of globular proteins based on the predominant type of secondary structure. Ptitsyn and Finkelstein (1980) suggested that because the majority of proteins belonging to a given class (i.e. α-helix, β-sheet or α-helix/β-sheet) have only a limited set of distinct topologies, many topological features may be simply determined by secondary structure. Therefore, the prediction of secondary structure is an integral step in predicting the overall three-dimensional structure of a protein (Ptitsyn and Finkelstein, 1983).

Various statistical methods for the determination of secondary structure have been developed on the basis of primary sequence data or

optical properties (i.e. circular dichroism (CD) and optical rotatory dispersion (ORD) spectral data). One of the most popular methods for the determination of secondary structure from primary sequence data, due to its simplicity and accuracy, was proposed by Chou and Fasman (1978b). However, the primary sequences of many food-related proteins have not been determined which limits the use of the Chou and Fasman procedure. The optical properties of the protein polypeptide backbone have been extremely useful for examining protein secondary structures in solution since the ORD and CD spectra of a protein are considered to be a reflection of the secondary structure fractions (Chang et al., 1978; Hennessey and Johnson, 1981). In recent years the analysis of CD data has predominated, and a variety of theoretical and empirical techniques have been proposed in order to analyze the CD spectrum of protein for the determination of secondary structure.

In the present study four different algorithms were compared for their ability to accurately determine protein secondary structure in relation to X-ray data. The methods compared were those of Siegel et al. (1980), Chang et al. (1978), Provencher and Glöckner (1981) and a procedure utilizing the simplex optimization algorithm of Morgan and Deming (1974). The methods that utilize reference spectra for the determination of secondary structure use a form of multiple regression analysis to solve the least squares algorithm. The simplex procedure was written in attempts to utilize the simplex algorithm as an alternative method to solve the least squares algorithm for the determination of the secondary structure fractions.

LITERATURE REVIEW

A. RELATIONSHIP BETWEEN OPTICAL PROPERTIES AND PROTEIN STRUCTURE

Conformational studies of biological macromolecules in solution have been a major tool in the search for relationships between structure and function. For many years there was no objective method for evaluating the three dimensional arrangements of molecular groups within the macromolecular structure, therefore, studies of macromolecules in solution were generally restricted to the description of gross hydrodynamic properties (Beychok, 1966).

During the mid 1950's, Moffitt (1956) and Moffitt and Yang (1956) established quantitative relationships between the optical parameters of a protein and its conformation. A great stimulus was given to the study of the optical properties of proteins when synthetic polypeptides became available (Bamford et al., 1956; Katchalski and Sela, 1958) which were capable of assuming the helical conformation predicted by Pauling et al. (1951) and Pauling and Corey (1951). The synthesis of these polypeptides greatly aided protein chemists since the secondary structure of these simpler polymers could be established by several independent physical methods (Jirgensons, 1969). On the basis of the correlation between the optical rotatory dispersion and X-ray diffraction pattern it became possible to describe the content of helix in proteins and polypeptides.

In recent years, optical rotatory dispersion (ORD) and its related phenomenon, circular dichroism (CD), have been valuable tools in studying the conformation and conformational changes of proteins in solution.

Both ORD and CD are manifestations of the same property of molecules having asymmetric groupings of atoms. ORD is a dispersion phenomenon (i.e. exhibiting a difference in refractive index for right and left circularly polarized light) and is exhibited at frequencies (wavelengths) far from, as well as near, those frequencies characteristic of the electronic transitions which are responsible for the optical activity. CD is an absorptive phenomenon (i.e. absorbing left and right circularly polarized light differently) and is observable only at frequency intervals where absorption occurs. It is therefore possible to locate the positions and signs of bands in circular dichroism with more certainty than is possible in an ORD spectrum (Foss, 1963).

Due to the intimate nature of the relationship, ORD and CD can be analytically interconverted by the use of a mathematical expression known as the "Kronig-Kramers transform". The nature of this mathematical relationship has been dealt with by several authors (Moffitt and Yang, 1956; Yang, 1956; Moffitt and Moscowitz, 1959).

B. OPTICAL PROPERTIES OF MOLECULAR GROUPS OF PROTEINS

The peptide group is the fundamental chemical and structural unit of all proteins, thus polypeptides and proteins may be regarded as peptide polymers. Each peptide group contains at least one asymmetric centre at the α-carbon atom, except in the case of glycine. Although the asymmetric centres contribute to the optical properties of proteins, the greatest influence on the optical properties of proteins is the specific arrangement of these peptide groups in three-dimensional space (Blout, 1971).

Since the optical activity of any molecule is related to its absorptive properties, it is necessary to identify the groups responsible for the absorptive properties found in proteins. Donovan (1969) summarized the absorption maxima and extinction coefficients for these groups. From this summary it is apparent that most of the absorption due to peptide chains lies in the ultraviolet region of 185 to 300 nm, with the strongest absorption appearing below 230 nm. It should also be noted that there are intense protein absorption bands which lie in the vacuum ultraviolet region, i.e. below 185 nm. Several researchers (Brahms et al., 1977; Bush et al., 1981; Hennessey and Johnson, 1981; Hennessey et al., 1982) have used the vacuum UV region for CD analysis of various proteins to yield useful information.

Research by Okabayashi et al. (1968) and Deker et al. (1970) studying the rotatory properties of oligomeric peptides, has shown that under certain conditions large increases in optical activity occur when the number of peptide units is between five and twelve. It was concluded that the observed increase in optical activity with added chain length was due to the formation of periodic structures which exhibit inherent optical activity.

Studies with homopolypeptides have demonstrated that an important ordering influence is exercised by ionic detergents on amino acid residues bearing an opposite charge (Hamed et al., 1983). Cationic homopolypeptides readily become ordered in the presence of dodecyl sulfate. Poly (L-ornithine) (Grouke and Gibbs, 1967; Satake and Yang, 1973) and poly (L-arginine) (McCord et al., 1977) form an ∝-helix, poly

(L-histidine) adopts a pleated sheet structure and poly (L-lysine) forms either a pleated sheet or helix depending on the conditions employed (Sakar and Doty, 1966; Mattice and Harrison, 1976).

Work by Madison and Schellman (1970) on model amides has indicated that when two peptide groups are held rigidly with respect to one another, large increases in optical activity are observed. It could be concluded that large rotations observed with proteins are caused when the peptide groups in proteins are held with a fixed geometry, either periodic or aperiodic.

The side chain groups of proteins which have significant absorption in the ultraviolet region and contribute to the optical activity, are the aromatic side chains of tryptophan, tyrosine and phenylalanine, and the disulfide groups of cystine residues (Townend et al., 1967; Strickland, 1974). In addition, small contributions to the optical activity may result from the presence of carboxylate and ammonium groups. The presence of metals in certain proteins can modify the absorptive and optical properties of the molecular groups with which the metals are associated (Nakano and Yang, 1981). In general, the contributions of the above mentioned groups to the optical spectra of proteins are small compared with the contributions of the peptide groups for two reasons: the relative numbers of such groupings are small; and, the inherent optical activity of these groups is lower than that of the peptide group (Blout, 1971). The optical properties of the aromatic amino acid residues, however, may in some cases be used to obtain important information about the molecular conformation and interactions of

peptide chains of proteins (Blout, 1971). Strickland (1974) presented an excellent treatise of the contribution to near-UV CD spectra by aromatic groups.

C. PROTEIN STRUCTURE

Linderstrom-Lang and Schellmann (1959) proposed a scheme whereby proteins could be distinguished into four levels of structural organization, namely the primary, secondary, tertiary and quaternary structures. These terms refer to the amino acid sequence, the regular arrangements of the polypeptide backbone, the three-dimensional structure of the globular protein, and the structures of aggregates of globular proteins, respectively. Anfinsen et al. (1961) and Anfinsen (1973) in renaturation experiments with ribonuclease, showed that the primary amino acid sequence contained all the structural information necessary for protein conformation, and postulated that the relationship between the levels of structural organization was dependent upon one another with elements at a lower level determining the elements of higher levels. Undoubtedly, the tertiary structure of biologically active proteins and polypeptides is one of the main factors for the high degree of specificity of reactions seen in vivo (Fasman, 1980).

X-ray diffraction analysis is the best method for the determination of tertiary structure, however, many proteins such as histones, membrane and ribosomal proteins as well as many food related proteins do not yield suitable crystal forms (Chou and Fasman, 1978b; Nakai, 1983). In the absence of three-dimensional structural information, however, one

can still learn much about the nature of the relationship between struc-
ture and function by examining lower levels of structure.

Since essentially all information concerning structure and func-
tion of a protein reside in the primary structure, a study of homology
or similarity between proteins may be important from the viewpoint of
comparative biochemistry. If homologies exist between two protein
sequences over a number of amino acid residues, a structural and func-
tional correlation may exist between these two proteins (Kubota et al.,
1981). Gibbs and McIntyre (1970) proposed the "diagonal-match" method
in which the two primary amino acid sequences are recorded along adja-
cent sides of a rectangular matrix. Within the body of the rectangle
every match is recorded. The method allows for rapid and simple detec-
tion of repetitions in the amino acid sequences of various proteins.
Beynon (1982) wrote a program in BASIC for the Apple® computer to carry
out such a task.

Various attempts have been made to predict secondary structure
from primary amino acid sequence (Guzzo, 1965; Kotelchuk and Scheraga,
1969; Ptitsyn and Finkelstein, 1970; Nagano, 1973; Kabat and Wu, 1973a
and 1973b; Chou and Fasman, 1978a and 1978b; Garnier et al., 1978; Palau
et al., 1982). In this connection, one of the most promising methods
has been the one proposed by Chou and Fasman (1978b). The method is
simple and easy to use and is reported to have reasonable accuracy when
compared to results obtained from X-ray data. However, some of the
rules set forth by the authors are qualitative rather than quantitative,
and are thus open to interpretation. As a result various authors

(Burgess and Scheraga, 1975; Garnier et al., 1976) have had only limited success using the Chou and Fasman method. Pham (1981) wrote a computer program for the Chou and Fasman (1978b) method in attempts to clarify some of the ambiguities in that method and found that the results were in good agreement with X-ray data.

The importance of hydrophobic forces in protein structure as well as in protein-protein interaction has been well documented in the literature (Kauzmann, 1959; Tanford, 1962; Bigelow, 1967; Jones, 1975; Ponnuswamy et al., 1980). Several methods have been suggested for the prediction of secondary structure based on hydrophobicity. Rose (1978) proposed a method whereby β-turns in various globular proteins were determined using a "hydrophobicity profile" technique. In this method the free energy of transfer for each amino acid, as defined by Nozaki and Tanford (1971), in the primary sequence is plotted against residue number; β-turns were identified in the profile as being regions of low energy. Rose and Roy (1980) suggested a procedure where protein-protein contact density about each residue was plotted against residue number. Contact density is defined as the "number of protein atoms other than hydrogen within a sphere of radius r about each α-carbon but excluding intra-residue atoms". The method was used to predict nucleation sites of secondary structure.

In a similar approach, Kyte and Doolittle (1982) wrote a computer program "that progressively evaluates the hydrophilicity and hydrophobicity of a protein along its amino acid sequence" using a "hydropathy"

scale. "Hydropathy" is derived from an amalgamation of experimental observations derived from the literature. Cid et al. (1982) suggested a simple method which allows for the prediction of secondary structure of proteins based on the examination of patterns exhibited when the "bulk hydrophobicity" (Ponnuswamy et al., 1980) of each amino acid residue in the primary amino acid sequence is plotted against residue number. Ptitsyn and Finkelstein (1983) derived an algorithm which took into account both local interactions inside each chain region and long-range interactions between different regions of the protein molecule. This algorithm was then used as a means for protein secondary structure prediction. Results from "blind" predictions (made before the X-ray structure became available) were later found to compare favourably to results obtained from X-ray analysis.

D. SECONDARY STRUCTURE

Secondary structures are regular arrangements of the backbone of the polypeptide chain without reference to side chain types or conformations, and are stabilized by hydrogen bonds between peptide amide and carbonyl groups (IUPAC-IUB Commission, 1970; Schulz and Schirmer, 1978). Four basic structures comprise the secondary structure fraction: these being the α-helix, β-sheet, β-turn and random coil.

1. α-helix

The α-helix is the most stable secondary structure found in proteins and was first described by Pauling et al. (1951). In this

structure there are 3.6 amino acid residues per turn. The alkyl (R) groups of the amino acids extend outward from the backbone of the helix such that the spacing per residue is approximately 1.5Å. A single turn of the helix extends 5.4Å along the axis. The ∝-helical arrangement of the peptide chain is favoured since it permits the formation of intra-chain hydrogen bonds between successive turns of the helix; hydrogen bonds form between the CO group of residue n and NH group of residue n+4. Since each peptide bond of the polypeptide chain participates in hydrogen bonding, the ∝-helix is maximally hydrogen bonded. With natur-ally occurring L-amino acids, either right- or left-handed helical coils can occur, however, the right-handed helix is significantly more stable (Lehninger, 1970). In all naturally occurring proteins examined so far, the ∝-helix has been right-handed.

2. β-sheet

Concomitantly with the ∝-helix, Pauling and Corey (1951) postu-lated that parallel and anti-parallel β-sheets are suitable regular hydrogen bonded structures for polypeptide chains. In this structure the peptide backbone forms a zig-zag pattern with the R groups of the amino acids extending above and below the peptide chain. All peptide bonds are available for hydrogen bonding thus allowing for maximum cross-linking between adjacent peptide chains and resulting in high stability (Pham, 1981).

Two distinctive patterns of hydrogen bonding can occur with β-strands depending on their interaction in a parallel or anti-parallel

orientation. The anti-parallel sheet has hydrogen bonds which are perpendicular to the strands, and has narrowly spaced bond pairs alternating with widely spaced pairs. Parallel sheet has evenly spaced hydrogen bonds which angle across between the strands (Richardson, 1981). A pure parallel sheet, a pure anti-parallel sheet, or a mixed sheet with some strand pairs parallel and some anti-parallel can be formed. However, mixed sheets are not readily formed since slightly different peptide orientations are required for the two types of hydrogen bonding (Richardson, 1977). Parallel β-sheets and the parallel portions of mixed sheets are generally buried within the protein structure with α-helices protecting them on both sides. On the other hand, anti-parallel sheets typically have one side exposed to the solvent and the other side buried which results in an alternation of side chain hydrophobicity in the amino acid sequence.

3. β-turn

β-turns (also known as reverse turns, β-bends, hairpin bends, 3_{10} bends, etc.) are structural features of peptides and proteins involving four consecutive amino acid residues and are located in regions where the peptide chain folds back on itself by 180°. An intramolecular hydrogen bond is formed between the C=O of residue n and the N-H of residue n+3. About one third of the residues in globular proteins may be involved in β-turns (Crawford et al., 1973; Zimmerman and Scheraga, 1977). Since these structures are thought to constitute a substantial proportion of the surface residues of proteins, it is likely that

certain β-turns provide recognition sites for the initiation of various functional reactions (Smith and Pease, 1980). Venkatachalam (1968) first recognized β-turns from theoretical conformational analysis and described three types based on dihedral angles. In addition, Lewis et al. (1973) defined five additional types of β-turns. From these eight types of β-turns it was possible to describe all cases (with the exception of the α-helix) where the α-carbons of residues n and n+3 are less than 7Å apart (Chou and Fasman, 1977).

4. Random coil region

When describing protein secondary structure for the purpose of description, prediction, spectroscopic characterization, etc., the classifications are α-helix, β-sheet, β-turn and random coil. Random coil is defined as being those portions of secondary structure that do not fall into the first three classifications (Richardson, 1981). Although random is used as an adjective in association with coil it has been shown from X-ray crystallographic studies that some portions described as random coil are not random or disordered. These regions have been shown to be "highly organized and firmly held in place as the repeating secondary structures - they are simply harder to describe" (Richardson, 1981). In addition to regions of random coil that are highly organized, there are areas in proteins that are highly disordered. These areas are either entirely absent on electron density maps or appear less dense and more spread out than other regions of the protein.

E. QUANTITATIVE DETERMINATION OF PERIODIC CONFORMATIONS OF PROTEINS

Once the relationship between the optical parameters of the α-helix of proteins and polypeptides and the data from X-ray diffraction analysis had been established, a logical progression was then to attempt to use ORD and CD data to make quantitative estimates of secondary structure fractions (i.e. α-helix, β-sheet, β-turn and random coil). It is now generally accepted that the ORD and CD spectra are a direct reflection of protein secondary structure (Hennessey and Johnson, 1981).

Various attempts using a variety of theoretical and empirical techniques have been used to quantitate ORD and more recently, CD data. Early attempts were made by Greenfield et al. (1967) to fit experimental ORD spectra of proteins to reference ORD curves of poly(L-lysine), containing varying amounts of α-helix, β-structure and random coil segments. Results indicated that the reference curves were useful in predicting gross polypeptide and protein structure, however, interpretation of the ORD data became difficult when the polypeptide or protein contained aromatic groups, disulfide bridges or prosthetic chromophores. Magar (1968) proposed the L_2 norm method (least squares) for minimizing the variance between observed and computed ORD curves for the determination of α-helix, β-form and random coil.

The inherent problems of ORD curve analysis, coupled with advances in technology has resulted in the analysis of CD spectra for the determination of secondary structure becoming prevalent. Greenfield and Fasman (1969) used circular dichroism curves of poly(L-lysine)

containing varying amounts of α-helix, β-sheet and random coil segments to determine protein secondary structure. CD curves of several proteins (e.g. myoglobin, lysozyme, etc.) whose three-dimensional structures had been determined from X-ray data were fitted by a linear combination of the three reference structures. The computed curves were found useful in predicting protein structure. If the protein in question possessed a high degree of secondary structure, the agreement between the calculated and the X-ray diffraction determined structure was extremely good, however, if the protein was largely non-regular the results were less satisfactory.

In the method of isodichroic points (i.e. the point of crossing of two of the three CD spectra), Myer (1970) estimated the fractions of the three forms in proteins from the CD measurements at three wavelengths corresponding to the three isodichroic points. Poly(L-lysine) was again used as a model for the three reference forms. It was found that the method yielded satisfactory results for various proteins when compared to other procedures, for example Greenfield et al. (1967) and Magar (1968), however, in comparison to X-ray diffraction data the results obtained were fair. The use of synthetic polypeptides such as poly (L-lysine) to estimate secondary structure fractions in globular proteins has lead to several criticisms. Synthetic polypeptides do not resemble real proteins whose structural elements differ from ideal models; helical polypeptides of high molecular weight such as deprotonated poly(L-lysine) are unlike short helical segments in a protein whose CD spectrum is chain-length dependent (Chen et al., 1972). In addition, the conformations assumed by the β-polypeptides may have a

different CD spectrum and/or magnitude under different conditions (Greenfield and Fasman, 1969; Li and Spector, 1969; Chang et al., 1978).

Saxena and Wetlaufer (1971), in order to avoid the aforementioned problems, used the CD spectra of three proteins (ribonuclease, lysozyme and myoglobin) of known secondary structure to derive the reference ellipticities for the three structural fractions (α-helix, β-structure and random). The method gave results that were comparable to diffraction estimates, and were as good as, or better than, analyses of protein CD spectra based on polylysine. Similar studies were carried out by Chen and Yang (1971) and Chen et al. (1972) where five reference proteins (myoglobin, lysozyme, lactate dehydrogenase, papain and ribonuclease) were used to calculate the reference ellipticities of α-helix, β-sheet and random coil at the different wavelengths by a least squares method. In a subsequent paper, Chen et al. (1974) increased the number of reference proteins to eight and introduced an expression to account for the chain-length dependence of helices, which reportedly lead to "a more reasonable determination of fractions of helix". Woody and Tinoco (1967) found that the optical properties of helices were dependent on the chain-length of the helix.

Early studies regarding the analysis of CD spectra considered only the α-helix, β-sheet and random coil fractions. Chang et al. (1978) modified the algorithm of Chen et al. (1974) to include the β-turn fraction and increased the protein reference base to fifteen. A similar approach was used by Bolotina et al. (1980) where five proteins were used as a reference base. In attempts to extract more information from

the CD spectrum, Bolotina et al. (1979) defined new CD reference spectra to include both parallel and anti-parallel β-sheet as well as α-helix, β-turn and random coil. Satisfactory results were obtained when the calculated secondary structure fractions were compared to data obtained from X-ray diffraction analysis.

F. CURVE FITTING ANALYSIS OF CD SPECTRA

In attempts to determine the individual secondary structure fractions through the analysis of CD spectra, a variety of techniques have been used. Early studies (Greenfield and Fasman, 1969) visually compared the experimental curve with synthetic curves containing various amounts of α-helix, β-sheet and random coil, however, only approximations could be made using this technique. A mathematical approach was taken in subsequent studies where most researchers (Saxena and Wetlaufer, 1971; Barela and Darnall, 1974; Grosse et al., 1974; Bannister and Bannister, 1974; Chang et al., 1978; Brahms and Brahms, 1979; Bolotina et al., 1979; Siegel et al., 1980) used least squares analysis to solve the reference ellipticities of the secondary structure fractions and to determine the secondary structure fractions.

In a different approach Baker and Isenberg (1976) presented a method whereby the circular dichroism spectra was analyzed by employing integrals over the data. The α-helix, β-sheet and random coil contents of the proteins were calculated from such integrals. It was shown that the analyzed α-helix content was usually reliable, β-sheet somewhat less reliable and the random coil values were least reliable when compared to X-ray data.

Hennessey and Johnson (1981) presented a method for predicting secondary structure from the CD spectrum where eight types of secondary structure were considered: helix; parallel and anti-parallel β-strand; types I, II and III β-turn; all other β-turns combined; and "other" structures. The method was based on the eigenvector technique of multi-component matrix analysis (Lloyd, 1969) using a set of fifteen protein CD spectra and one polypeptide (poly(L-glutamic acid)) CD spectrum over the range of 178 and 260 nm in order to generate a set of orthogonal CD spectra for use as a reference basis. The eight types of structure were determined from the basis CD spectra as a function of the protein CD spectrum. Results compared favourably with X-ray data.

Provencher and Glöckner (1981) developed a method in which the CD spectrum of a protein was analyzed directly as a linear combination of the circular dichroism spectra (190 to 240 nm) of sixteen proteins whose secondary structures were determined from X-ray data. The technique avoids the problem of trying to determine reference CD spectra characteristic of each of the individual secondary structure fractions as is required when using least squares fitting analysis. The method is based on a simple constrained statistical regularization and employs quadratic programming. When the calculated α-helix and β-sheet fractions of the various proteins analyzed were compared to X-ray diffraction results, accuracy was very good. Less accurate results were obtained for β-turn and remainder fractions.

G. SIMPLEX OPTIMIZATION

Presently, one of the main goals of research is to develop and establish mathematical methods to efficiently search for optimum experimental conditions (Nakai, 1982). One of the most popular methods for finding these conditions is the sequential simplex method of Morgan and Deming (1974). The usefulness of this technique lies in its ability to handle several variables at a time in a straight forward manner. This multifactor technique was originally developed by Spendley et al. (1962) as an alternative to the time consuming and complex evolutionary operation (EVOP) method of Box (1957). Nelder and Mead (1965) modified the technique to allow for the rapid search of an optimal response and Morgan and Deming (1974) later refined the procedure to allow for even faster searches.

The simplex optimization method has been extensively used in analytical chemistry and several reviews have been published regarding its application in this field (Deming and Parker, 1978; Kowalski, 1980; Frank and Kowalski, 1982). However, very few applications of the simplex method have been cited with regards to food science. Dols and Armbrecht (1976) were one of the first to propose and outline the simplex optimization procedure for application to food related problems. van de Voort et al. (1979) applied the simplex optimization technique to calculate the molecular weights and relative concentration of the individual proteins in a mixture of three food related proteins using sedimentation equilibrium data. Results obtained were in good agreement with the theoretical values. Fujii and Nakai (1980) used

simplex optimization to transform food related data for linearization and obtained good results when calculated and model curves were compared. Li-Chan et al. (1979) using simplex optimization, optimized the conditions for the covalent attachment of lysine to wheat gluten. Recently, Nakai (1982) carried out an extensive survey of various optimization techniques for use in food product and process development and proposed that a form of the simplex optimization technique (modified super-simplex algorithm) could be very useful in optimizing food processing and analysis.

METHODS

The CD spectra of sixteen proteins of known secondary structure (as determined by X-ray analysis) were used in the comparison of various algorithms for determination of secondary structure fractions. The algorithms included those of Siegel et al. (1980), Chang et al. (1978), Provencher and Glöckner (1981) and a technique which utilizes the simplex algorithm as proposed by Morgan and Deming (1974). The CD spectra of the following proteins were obtained from the CONTIN computer program of Provencher (1980) for use as the reference data base: adenylate kinase, α-chymotrypsin, carboxypeptidase A, concanavalin A, cytochrome C, elastase, insulin, lactate dehydrogenase, lysozyme, myoglobin, nuclease, papain, parvalbumin, ribonuclease A, ribonuclease S and trypsin inhibitor.

A. ALGORITHM OF SIEGEL ET AL. (1980)

The method of Siegel et al. (1980) allows for the prediction of the fraction of helical structure of an uncharacterized protein. The computer program was translated from DEC (Digital Equipment Corporation) standard BASIC to Fortran IV and appears in Appendix 1. Thirteen mean residue ellipticities at wavelengths between 210 and 240 nm (i.e. 210, 213, 216, 218, 220, 222, 224, 227, 229, 231, 234, 237 and 240 nm) were entered into the program to determine the proportion of helix. The amount of β-sheet is calculated indirectly, as it is linearly dependent on the quantity of α-helix present. No β-turns or unordered fractions

can be calculated using the method. The following equations were used for the calculation of α-helix and β-sheet fractions:

$$f_H = \frac{[\Theta_\lambda] - C_\lambda}{\Theta_{H(\lambda)}} \qquad \text{Eq. 1}$$

$$f_\beta = -0.729 \, f_H + 0.583 \qquad \text{Eq. 2}$$

where f_H = fraction of α-helix

       $[\Theta_\lambda]$ = observed mean residue ellipticity

       C_λ = correction factor

       f_β = fraction of β-sheet

       $\Theta_{H(\lambda)}$ = mean residue ellipticity of a hypothetical protein that is 100% helix.

B. ALGORITHM OF CHANG ET AL. (1978)

Chang et al. (1978) used a form of multiple regression to solve the least squares in order to determine the secondary structure fractions. This method is based on the assumption that the mean residue ellipticity [Θ] at any wavelength is the sum of the structural elements (α-helix, β-sheet, β-turn and unordered) of the protein molecule. In order to calculate the secondary structure fractions the following equation (Eq. 3) was used:

$$[\Theta]_\lambda = f_H [\Theta]_H^\infty (1 - {}^k/_n) + f_\beta [\Theta]_\beta + f_t [\Theta]_t + f_R [\Theta]_R \qquad \text{Eq. 3}$$

The [Θ] values on the right hand side of the equation represent the reference values for helix (H), β-sheet (β), β-turn (t) and unordered form (R), respectively. The f values correspond to the fractions of the

structural elements. The chain length dependence of the α-helix was accounted for by the inclusion of k which is a wavelength dependent constant, and n̄, the average number of amino acid residues per helical segment.

Since the program of Chang et al. (1978) was not available, a computer program was written in Fortran IV using subroutines from the UBC Matrix (1979) manual to solve the least squares problem (Appendix 2). The method implemented was one which solved overdetermined systems (i.e. more equations than unknowns) and was basically a multiple regression method which allowed for linear constraints. This program utilized the reference ellipticities for the various secondary structure fractions obtained from J. T. Yang (unpublished data). Preliminary results indicated that the secondary structure fractions obtained using the computer program were nearly identical to those reported by Chang et al. (1978) and therefore, the latter results were used for comparative purposes in the present study.

C. ALGORITHM OF PROVENCHER AND GLÖCKNER (1981)

The computer program of Provencher (1980) which implements the method of Provencher and Glöckner (1981) was adapted to the Amdahl 470 V/8 computer and tested with sample data to ensure that the program was running correctly. In their method Provencher and Glöckner (1981) determined the secondary structure fractions of each protein after removing that protein's CD spectrum from the reference data base (i.e. "blind analysis"). Attempts to duplicate the results of Provencher and Glöckner (1981) by modification of the data base were not pursued, due

to the complexity of the computer program. Therefore, the secondary structure fraction values as reported by the authors were used in the comparison with those determined by X-ray data. The following equations were used for the calculation of the various secondary structure fractions:

$$y(\lambda) = \sum_{j=1}^{N_\gamma} \gamma_j R_j(\lambda) \qquad \qquad \textbf{Eq. 4}$$

$$f_i = \sum_{j=1}^{N_\gamma} \gamma_j F_{ji} \qquad i = 1, \ldots, N_f \qquad \textbf{Eq. 5}$$

$$\sum_{k=1}^{N_y} [y(\lambda_k) - y_{obsd}(\lambda_k)] + \alpha \sum_{j=1}^{N_\gamma} (\gamma_j - \frac{1}{N}) = \text{minimum} \qquad \textbf{Eq. 6}$$

$$\sum_{i=1}^{N_f} f_i = 1 \qquad f_i \geq 0 \qquad \textbf{Eq. 7}$$

where $y(\lambda)$ = mean residue ellipticity

λ_k = wavelength k

γ_j = a proportionality factor

N_γ = the number of proteins used in the data base

$R_j(\lambda)$ = mean residue ellipticity of protein j at wavelength λ

f_i = fraction of secondary structure

F_{ji} = fraction of secondary structure i in protein j

N_f = number of secondary structure classes

N_y = number of measured mean residue ellipticities

α = regularizer

D. SIMPLEX-LEAST SQUARES METHOD

The final method examined used the simplex algorithm of Morgan and Deming (1974) to determine secondary structure fractions in order to obtain a least squares solution for Eq. 8. The technique was therefore termed the simplex-least squares method.

$$\min = \sum([\Theta]_{obs} - (f_H[\Theta]_H^\infty(1-{}^k/_{\bar{n}}) + f_\beta[\Theta]_\beta + f_t[\Theta]_t + f_R[\Theta]_R))^2$$

Eq. 8

The right hand side of Eq. 3 (Chang et al., 1978) was incorporated into Eq. 8 for the least squares calculation. The reference ellipticities for the various secondary structure fractions, as well as the wavelength dependent factor (k) at 1 nm intervals from 190 to 240 nm were obtained from J. T. Yang (unpublished data). Since the helix reference ellipticity used by Chang et al. (1978) is dependent on both k and $\bar{n}$, a program was written which calculates the helix reference ellipticities for various $\bar{n}$ values and appears in Appendix 3.

Two versions of the simplex-least squares method were used in the study: one in which $\bar{n}$ was held constant at 10.4 and the other where $\bar{n}$ was varied according to the protein examined. The value of 10.4 is an average based on 18 proteins of known secondary structure (Chang et al., 1978). In order to calculate the secondary structure fractions, 51 mean residue ellipticities (190 to 240 nm) were entered into the program. A listing of the program is found in Appendix 4.

E. CORRELATION COEFFICIENTS

In order to assess the accuracy of the algorithms the proportion of secondary structure fractions determined by each method was regressed against the proportions determined from X-ray diffraction analysis (obtained from Chang et al., 1978). The linear regression program for the Monroe 1880 Programmable Calculator (Litton Business Systems Inc., Orange, NJ) was used to calculate the correlation coefficient (r). The r value was then compared to a table of critical values (Zar, 1974) to test for significance.

RESULTS AND DISCUSSION

The results of the secondary structure determination using the four algorithms are compared in Table 1 to the X-ray diffraction data reported by Chang et al. (1978).

A. ALGORITHM OF SIEGEL ET AL. (1980)

The method of Siegel et al. (1980) was the most limited of the algorithms examined for use in structure-function studies since only α-helix and β-sheet fractions were determined. In comparison, four secondary structure fractions (α-helix, β-sheet, β-turn and unordered) were determined by the other algorithms. Examination of the equations utilized by Siegel et al. (1980) (Eq. 1 and 2) indicates that the α-helix content of the test protein is calculated directly from the CD spectra while the β-sheet fraction is calculated indirectly and is a function of the α-helix content. Equation 1 was derived by determining the fit of the ellipticities of 16 reference proteins to their X-ray structures at 13 selected wavelengths using multiple linear regression analysis. However, the authors reported that only the mean residue ellipticity ($\Theta_H(\lambda)$) significantly ($P<0.01$) described the observed ellipticity. The relationship between α-helix and β-sheet (Eq. 2) was formulated by regressing the fraction of helix on the fraction of β-sheet using 60 proteins as a data base. No indication was given by the authors as to the nature of these proteins.

Although only ellipticities from 210 nm to 240 nm were entered into the algorithm, the method showed surprisingly good accuracy for

Table 1. A comparison of X-ray secondary structures to predicted secondary structures of selected proteins using various algorithms.

Protein	Helix						β-sheet						β-turn					Random				
	1	2	3	4	5	6[a]	1	2	3	4	5	6	1	3	4	5	6	1	3	4	5	6
Myoglobin	.79	1.09	.80	.86	.78	.72	.00	.00	.00	.00	.05	.11	.05	.02	.00	.02	.03	.16	.18	.14	.16	.14
Lysozyme	.41	.41	.32	.45	.31	.33	.16	.29	.29	.21	.31	.33	.23	.08	.26	.07	.04	.20	.31	.08	.32	.31
Ribonuclease A	.23	.28	.21	.26	.22	.22	.40	.38	.39	.44	.36	.35	.13	.10	.11	.12	.11	.17	.13	.19	.31	.31
Papain	.28	.46	.29	.27	.28	.28	.14	.25	.00	.05	.01	.01	.17	.15	.31	.15	.15	.41	.56	.36	.56	.56
Lactate dehydrogenase	.45	.53	.45	.40	.41	.42	.24	.20	.18	.22	.29	.21	.06	.13	.13	.08	.12	.25	.24	.26	.23	.25
α-chymotrypsin	.09	.14	.05	.09	.06	.09	.34	.48	.53	.29	.50	.43	.34	.02	.22	.03	.05	.23	.40	.40	.41	.44
Concanavalin A	.02	.15	.25	.08	.19	.62	.51	.47	.46	.41	.61	.00	.09	.20	.15	.12	.24	.38	.09	.36	.08	.14
Cytochrome C	.39	.42	.44	.33	.43	.48	.00	.27	.00	.09	.00	.00	.24	.28	.17	.28	.24	.37	.28	.41	.29	.28
Nuclease	.24	.39	.30	.32	.23	.22	.15	.30	.21	.25	.38	.39	.18	.12	.14	.05	.05	.43	.37	.29	.35	.34
Insulin	.51	.43	.46	.49	.40	.43	.24	.27	.22	.23	.37	.34	.12	.19	.27	.12	.12	.13	.13	.00	.11	.10
Parvalbumin	.62	.59	.49	.58	.38	.37	.05	.15	.00	.00	.32	.35	.17	.26	.00	.12	.11	.16	.25	.42	.19	.18
Carboxypeptidase A	.37	.31	.45	.43	.45	.41	.15	.36	.00	.15	.00	.06	.26	.37	.25	.36	.35	.22	.18	.16	.19	.19
Trypsin inhibitor	.28	.29	.07	.21	.08	.09	.33	.37	.32	.28	.29	.27	.03	.00	.23	.02	.02	.36	.61	.29	.61	.61
Adenylate kinase	.54	.54	.46	.44	.46	.45	.12	.19	.26	.15	.26	.29	.19	.08	.20	.08	.07	.15	.20	.21	.20	.19
Ribonuclease S	.26	.28	.24	.25	.19	.19	.44	.38	.33	.37	.46	.46	.13	.14	.16	.08	.07	.17	.29	.23	.27	.27
Elastase	.07	.03	.00	.04	.00	.01	.52	.56	.46	.49	.46	.44	.26	.07	.14	.08	.08	.15	.47	.32	.46	.47

[a]Methods:
 1 = X-ray (Chang et al., 1978).
 2 = Siegel et al. (1980).
 3 = Chang et al. (1978).
 4 = Provencher and Glöckner (1981).
 5 = Simplex-least squares $\bar{n}$ = 10.4.
 6 = Simplex-least squares $\bar{n}$ = variable.

both α-helix and β-sheet fractions as indicated by the significant cor-
relation coefficients (P<0.001) (Table 2). These results were obtained
despite the fact that a number of CD phenomena occur below 210 nm,
including: the 193 nm maximum of α-helix, the 198 nm maximum of β-sheet
and the 202 nm maximum and the 190 nm minimum of the β-turn. Several
authors (Brahms et al., 1977; Brahms and Brahms, 1980; Hennessey and
Johnson, 1981) using vacuum ultraviolet CD have observed new character-
istic bands for the secondary structure fractions. The inclusion of
these bands into algorithms which analyze CD data may improve the deter-
mination of protein secondary structure.

In order to examine the versatility of each of the methods
examined, correlation coefficients were calculated in the presence and
absence of concanavalin A. If a method can accurately determine secon-
dary structure fractions, regardless of the protein examined, then the
correlation coefficient should not be greatly affected by removing that
protein from the analysis. Concanavalin A is known to give anomalous
secondary structure results when CD spectral data are analyzed by some
methods (Chang et al., 1978; Bolotina et al., 1979). Removal of secon-
dary structure values that deviate largely from X-ray data, as may be
the case with concanavalin A, should increase the correlation coeffi-
cient; ideally, the secondary structure fraction predicted should equal
that determined by X-ray diffraction analysis. Using the method of
Siegel et al. (1980) the correlation coefficients for α-helix and
β-sheet were not greatly affected by the removal of concanavalin A.
Therefore, the method showed good ability to determine α-helix and
β-sheet fractions (Table 1).

Table 2. Correlation coefficients between calculated secondary structure fractions and those obtained from X-ray data.

Methods	Secondary structures			
	Helix	β-sheet	β-turn	Random
Siegel et al. (1980)				
+ con A[1]	.9139[a]	.8305[a]	N.A.[2]	N.A.
− con A[1]	.9158[a]	.8116[a]	N.A.	N.A.
Chang et al. (1978)				
+ con A	.8833[a]	.8526[a]	.2126	.5260[c]
− con A	.9299[a]	.8329[a]	.2629	.5571[c]
Provencher and Glöckner (1981)				
+ con A	.9650[a]	.9377[a]	.3030	.4944
− con A	.9634[a]	.9346[a]	.2967	.4572
Simplex-least squares $\bar{n} = 10.4$				
+ con A	.8799[a]	.7718[a]	.3450	.3950
− con A	.9080[a]	.7173[b]	.3611	.5994[c]
Simplex-least squares $\bar{n} = $ variable				
+ con A	.5858[c]	.3697	.2027	.4317
− con A	.9025[a]	.6602[b]	.3236	.5808[c]

[1] Correlation coefficients calculated in the presence of concanavalin A (+ con A) and in the absence of concanavalin A (− con A).

[2] N.A. = not applicable.

[a] Significant at $P<0.001$.

[b] Significant at $P<0.01$.

[c] Significant at $P<0.05$.

It is interesting to note, however, that the α-helix fraction of myoglobin predicted by the method of Siegel et al. (1980) differed markedly from the X-ray data (Table 1). Proteins such as myoglobin which are predominantly helical in nature (Levitt and Chothia, 1976) may require the inclusion of the 193 nm band in the analysis of CD spectral data for the determination of α-helix.

Although some of the CD spectra used by both Siegel et al. (1980) and Chang et al. (1978) to generate the reference ellipticities were obtained from Yang (e.g. myoglobin, lysozyme, ribonuclease A, etc.), the secondary structure fractions used to generate the reference ellipticities differed for some of the proteins. For example, for myoglobin Siegel et al. (1980) used 87.6 as the percentage of helix, while Chang et al. (1978) used 79.0. Siegel et al. (1980) obtained the secondary structure fractions for the various proteins from the data of Levitt and Greer (1977), whereas Chang et al. (1978) obtained the secondary structures from the X-ray diffraction literature. The differences in the secondary structure fractions used by the two groups may explain why Siegel et al. (1980) were only able to find a significant relationship. between observed ellipticity and helix. However, the methods (e.g. Chang et al., 1978; Bolotina et al., 1980) which assume that the observed ellipticity is a linear function of the secondary structures have not used multiple regression analysis to examine the possibility that only certain secondary structure fractions significantly describe the observed ellipticity. In view of the significant correlations for both helix and β-sheet fractions, it would nevertheless appear that the

use of a limited number of wavelengths between 210 and 240 nm, as proposed by Siegel et al. (1980), is sufficient to determine these two secondary structure fractions for the proteins examined.

B. ALGORITHM OF CHANG ET AL. (1978)

The algorithm of Chang et al. (1978) showed relatively good ability to determine secondary structure fractions for the proteins examined with the exception of concanavalin A. X-ray analysis has shown concanavalin A to contain 2 percent α-helix while the method of Chang et al. (1978) predicted 25 percent (Table 1). The discrepancy noted for concanavalin A may stem from the choice of reference spectra of the various structure fractions. According to Chen et al. (1974) inappropriate reference spectra for the secondary structures used to examine a protein may result in incorrect structure determinations. The choice of reference spectra is difficult due to the number of structural variants for each of the secondary structure fractions within the protein molecule. The reference spectrum for each fraction becomes "at best a statistical average of numerous variants" (Chang et al., 1978), so that for example, the helix fraction represents α-helix as well as 3_{10}-helix.

Examination of the correlation coefficients (Table 2) in the presence and absence of concanavalin A indicated that the method could determine α-helix and β-sheet fractions with a high degree of accuracy, in relation to X-ray data. A relatively accurate determination of the unordered fractions was also obtained. The β-turn fractions were the least correlated ($P > 0.05$) of the secondary structure fractions. The

β-turn fraction, although consisting of only four residues, has a large number of structural variants (Venkatachalam, 1968; Lewis et al., 1973). This makes the β-turn reference ellipticity difficult to characterize (Chen et al., 1974).

In order to improve the accuracy of methods based on reference spectra for the various secondary structure fractions, one could increase the number of reference proteins, but more importantly, increase the conformational diversity of the reference proteins. However, Grosse et al. (1974) postulated that a combination of these two factors would be most favourable.

C. ALGORITHM OF PROVENCHER AND GLÖCKNER (1981)

Unlike the other methods examined in this study which used a form of least squares to solve for the structure fractions, the algorithm of Provencher and Glöckner (1981) analyzed the CD spectrum of a protein directly as a linear combination of the CD spectra (from 190 to 240 nm) of the sixteen proteins in order to determine the secondary structure fractions. The authors stated that using such a method would avoid the problem of defining single reference ellipticities for each secondary structure.

Highly significant correlation coefficients (P<0.001) were found for both helix and β-sheet fractions (Table 2) and were the highest among the methods tested. The β-turn and unordered fractions, however, were not significantly (P>0.05) correlated to X-ray data, possibly due to the limited variety of reference protein spectra in the data base.

It was the contention of the authors that the accuracy of the method would increase as the variety of the reference proteins increases. As seen from Table 2, the correlation coefficients generated for each of the secondary structure fractions were not affected to a large extent by the removal of concanavalin A from the analysis. Thus the method was able to predict the secondary structure fractions of concanavalin A with a reasonable amount of accuracy (Table 1) which may indicate the versatility of this algorithm.

The method of Provencher and Glöckner (1981) was the most statistically unbiased of the algorithms examined since a "blind analysis" was carried out by the authors where the protein being examined was removed from the reference spectra prior to secondary structure determination. Preliminary results obtained in the present study revealed that secondary structure determinations from the CD spectrum of a protein, while keeping that protein in the reference data base, were nearly identical to results from X-ray analysis. This would be expected since the method basically matches the CD spectrum being analyzed to one in the reference data base. For example, when the CD spectrum of concanavalin A was analyzed the method predicted 0.03 helix, 0.48 β-sheet, 0.10 β-turn and 0.39 unordered fraction, while from X-ray 0.02 helix, 0.51 β-sheet, 0.09 β-turn and 0.38 unordered fractions were determined.

D. SIMPLEX-LEAST SQUARES METHOD

The final method was one in which the simplex optimization algorithm of Morgan and Deming (1974) was used to minimize the least squares

difference between the observed and calculated mean residue ellipticities (Eq. 8). Numerous papers have been published regarding the determination of secondary structure from CD data with the majority using least squares to solve for the structure fractions. The least squares problem has in turn been solved with multiple regression techniques (Chang et al., 1978; Bolotina et al. 1980).

The simplex optimization algorithm was investigated as an alternative method for solving the least squares problem. The method is an empirical iterative strategy in which the calculations are simple and the decisions are formalized (Morgan and Deming, 1974), and no assumptions are made about the mathematical function in which simplex is to be employed (Leggett, 1977).

The concept of the simplex algorithm can be more easily understood with the following discussion. A simplex is represented by a geometric figure defined by one more dimension than the number of variables to be determined (Deming and Parker, 1978). Since helix, β-sheet, β-turn and unordered fractions were to be determined, the simplex in this study would be represented by a pentagon. The initial step in the simplex algorithm is the generation of the starting simplex matrix which has the dimensions n+1 by n in standard matrix notation (where n is the number of original variables); the matrix is represented by the following, where rows represent vertices and columns represent secondary structure fractions:

$$
\begin{array}{cccc}
0 & 0 & 0 & 0 \\
p & q & q & q \\
q & p & q & q \\
q & q & p & q \\
q & q & q & p
\end{array}
$$

where   0 = lower boundary value

$$p = \frac{1}{n\sqrt{2}} \left\{ (n-1) + \sqrt{n+1} \right\} \qquad \textbf{Eq. 9}$$

$$q = \frac{1}{n\sqrt{2}} \left\{ \sqrt{n+1} - 1 \right\} \qquad \textbf{Eq. 10}$$

The matrix thus defines the proportion of each secondary structure fraction to be entered into Eq. 8.

In the original procedure set forth by Spendley et al. (1962) the range for the variables p and q used in the starting simplex were transformed to achieve a range from zero to one. When dealing with secondary structure fractions, however, this criterion is already met since the proportion of any one fraction cannot be less than zero or greater than one in relation to the total structure of a protein. Normalization of the data generated in the simplex was done in order to meet the constraint that the sum of the secondary structure fractions would total one. This constraint is based on the assumption that the CD spectrum of a protein represents the sum of the independent contributions of n types of secondary structure (Brahms and Brahms, 1979; Hennessey and Johnson, 1981). The mean residue ellipticity at each wavelength is represented by a linear combination of n reference spectra S_i, which are assumed to be representative of the secondary structures:

$$[\Theta]_\lambda = \sum_{i=1}^{n} f_i S_i \qquad \textbf{Eq. 11}$$

where n in most cases equals 4, representing helix, β-sheet, β-turn and unordered fractions.

To obtain the normalized value, the secondary structure fractions for each vertex were totalled and then the individual secondary structure fractions for that vertex were divided into the total (Eq. 12 and 13). In this way the sum of the individual secondary structure fractions for each vertex would equal one.

$$X = \sum_{i=1}^{n} f_i \qquad \text{for } i=1 \text{ to } 4 \qquad \qquad \text{Eq. 12}$$

$$\overline{f}_i = f_i / X \qquad \qquad \text{Eq. 13}$$

where   X = sum of the secondary structure fractions

f_i = secondary structure fraction

$\overline{f}_i$ = normalized secondary structure fraction

The normalized values are entered into Eq. 8. The squared differences between the observed mean residue ellipticity and the calculated mean residue ellipticity using the normalized secondary structure fractions at each vertex are then summed over the wavelength range (190 to 240 nm). The responses at the five vertices are compared to identify the worst (W) and best (B) values (Figure 1). Since a minimization of Eq. 8 is required, the worst vertex would be one which generated a large squared difference value and conversely, the best vertex would be one which generated a small value. The pentagon used in this study may actually be condensed to the triangle BNW, as illustrated in Figure 1, by having N (next-to-worst vertex) represented by the average of the responses excluding the best and worst responses.

W - WORST VERTEX

B - BEST VERTEX

N - NEXT - TO - WORST VERTEX

$\bar{P}$ - CENTROID OF N - B VERTICES

C_r - CONTRACTION OF R

C_w - CONTRACTION OF W

R - REFLECTION VERTEX

E - EXPANSION VERTEX

Figure 1. Possible simplex moves.

Once the best and worst responses are identified, operations such as reflections, expansions or various contractions are carried out in order to reposition the simplex on the response surface. The first operation is the generation of a reflection or the mirror image of the worst location which is calculated according to the following relationship:

$$\text{Reflection} = \overline{P} + 1.0 \ (\overline{P}-W) \qquad\qquad \textbf{Eq. 14}$$

where $\overline{P}$ is the centroid point, $\overline{P} = \dfrac{M}{n}$

M = the sum of all fractions within a simplex except that fraction identified with the worst location.

The reflection value for each structure fraction is entered into Eq. 8 in order to calculate the response (i.e. the squared difference). A comparison of the responses (i.e. the initial simplex plus the reflection) is then conducted and the appropriate operation taken according to the rules set forth by Morgan and Deming (1974) as outlined in Figure 2. The formulae for the various operations are given in Table 3. The major principle of the simplex algorithm is to replace those conditions which elicit the worst response; once completed, a new simplex is formed and the procedure starting with identification of the best and worst response locations is initiated again. The procedure is repeated until the responses elicited by the simplex do not differ by more than some predetermined value.

Results from Table 2 indicated that when $\overline{n}$ was anchored at 10.4 the correlation coefficients for the various secondary structure fractions were generally higher than when $\overline{n}$ was varied, the only exception being the unordered fraction with $\overline{n}$ = 10.4 in the presence of
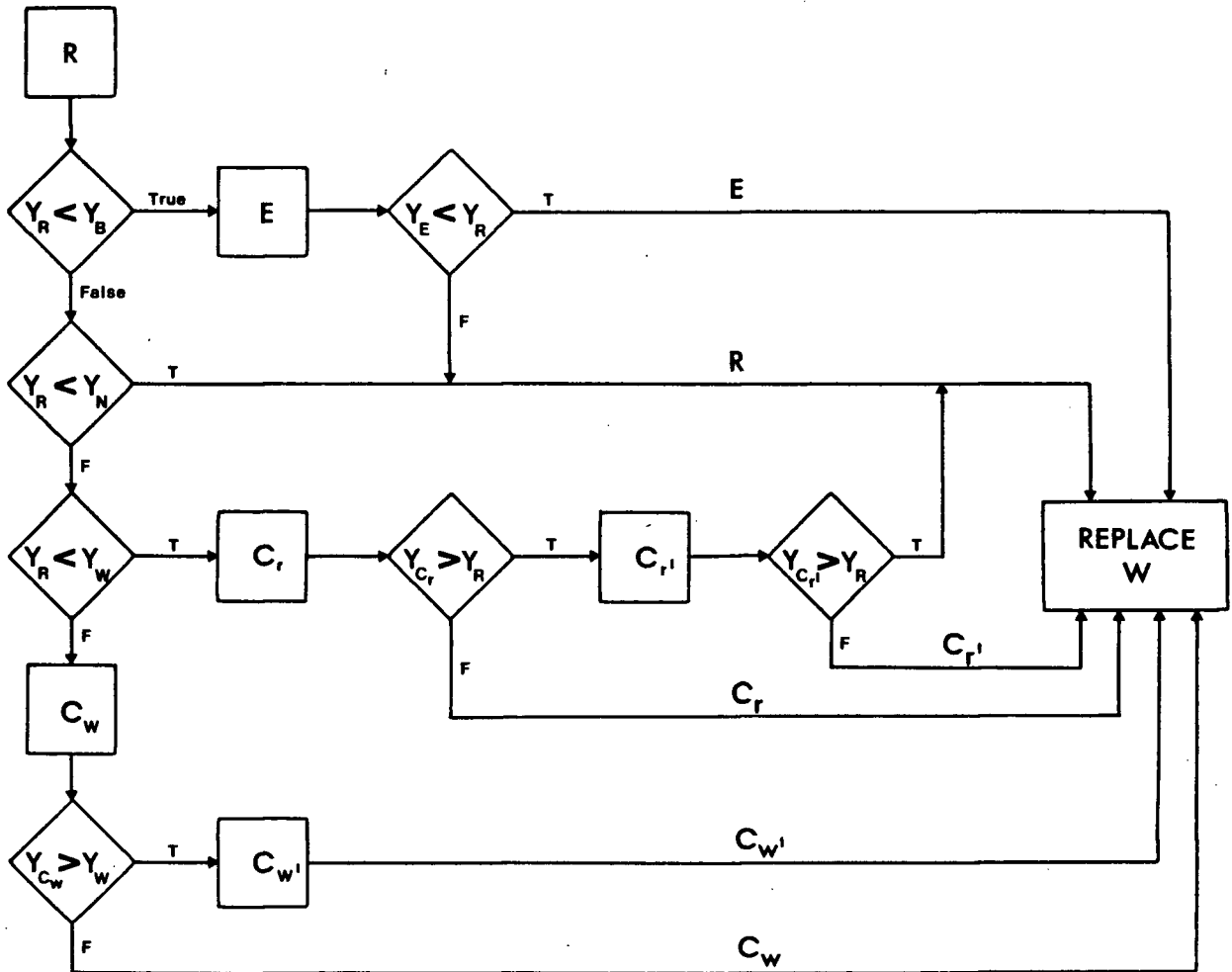
Figure 2. Flow chart of the simplex minimization. Calculation of the new vertex, $V = \overline{P} + \kappa(\overline{P}-W)$.

Table 3. Operational calculations used in the simplex optimization algorithm.

Reflection (R)   **Eq. 14**

$$R = \overline{P} + 1.0(\overline{P}-W)$$

Expansion (E)   **Eq. 15**

$$E = \overline{P} + 2.0(\overline{P}-W)$$

Reflection contraction (C_r)   **Eq. 16**

$$C_r = \overline{P} + 0.5(\overline{P}-W)$$

Worst contraction (C_w)   **Eq. 17**

$$C_w = \overline{P} - 0.5(\overline{P}-W)$$

Massive reflection contraction (C_r')   **Eq. 18**

$$C_r' = \overline{P} + 0.25(\overline{P}-W)$$

Massive worst contraction (C_w')   **Eq. 19**

$$C_w' = \overline{P} - 0.25(\overline{P}-W)$$

concanavalin A which was slightly lower than $\bar{n}$ = variable (+ con A). Examination of the results from Table 1, however, indicated that with the exception of concanavalin A, the structures predicted did not differ greatly with $\bar{n}$ = 10.4 and $\bar{n}$ = variable. The correlation coefficients calculated in the absence of concanavalin A (for both $\bar{n}$ = 10.4 and $\bar{n}$ = variable) were very similar to one another (the helix, β-sheet and unordered fractions were significant at the P<0.001, P<0.001 and P<0.05 levels, respectively). As with the other methods examined, no significant correlation coefficients (P>0.05) were obtained for the β-turn fractions. The effect on the correlation coefficients of removing values that deviate largely from one another was clearly illustrated in the case where $\bar{n}$ is varied. The r values for all the fractions increased with the removal of concanavalin A which was poorly estimated by the simplex method.

Provencher and Glöckner (1981) found that the inclusion of an empirical chain-length dependent factor ($\bar{n}$) for the helix fraction into their algorithm decreased the accuracy of their results, despite the fact that other authors (Woody and Tinoco, 1967; Madison and Schellman, 1972) have found that the optical activity of a helical polypeptide is chain-length dependent.

In general, the simplex method with $\bar{n}$ anchored at 10.4 gave results comparable to those of Chang et al. (1978) (where $\bar{n}$ was also set at 10.4). Since both the Chang et al. (1978) and the simplex method used the same reference protein spectra, any difference in results for the secondary structure fractions for the various proteins would be

attributed to the method used to solve least squares. Since the results were comparable this would indicate that the simplex optimization algorithm could be used as an alternative method to multiple regression in solving least squares. The simplex method may have produced more accurate results if different reference spectra had been generated which were more representative of the major structural classes. In order to generate new reference spectra one would have to increase the number of proteins examined and increase the variety of structures included within a class, as previously discussed.

Overall, the simplex method showed poor versatility (large changes in r values with the removal of concanavalin A from the regression calculation). The change in r values was less extensive when $\bar{n}$ = 10.4 as compared to $\bar{n}$ = variable.

E. LIMITATIONS OF CD ANALYSIS

No method to date has been able to determine the secondary structure fractions of proteins examined (i.e. "blind analysis") with one hundred percent accuracy, which would imply that some information necessary for understanding the relationship between the CD spectrum and secondary structure fractions is lacking, or that some of the assumptions being made are incorrect. For example, it is assumed that the CD spectral range examined (190 to 240 nm) yields all the necessary information to determine protein secondary structure. The use of vacuum ultraviolet CD may yield some interesting information that was previously lacking. Brahms and Brahms (1980) extended the analysis down

to 165 nm and found that the β-turn was characterized by the presence of strong bands in the 182 to 189 nm region, as well as those bands observed at 202 and 225 nm. In the case of α-helix, the authors observed that in addition to the well known 222 nm and 210 nm minima and the 193 nm maxima, there was also a shoulder at 174 nm. The inclusion of these new amide transitions could lead to a more detailed characterization of the circular dichroism of proteins.

In addition to the absence of the various types of structures within the major structural classes, all algorithms to date have neglected the effect of non-peptide chromophores below 250 nm. Although generally disregarded due to their small contribution to the total optical activity, circular dichroism bands arising from aromatic side chains and disulfide linkages may be important for the CD analysis of some proteins. Proteins such as avidin and acid DNase are known to have strong circular dichroism bands in the region of 225 to 250 nm (Sears and Beychok, 1973).

One of the major assumptions made in most of the recent algorithms used to analyze CD data, is that the CD spectrum results from the contribution of the optical activities of the four major classes of secondary structure (i.e. helix, β-sheet, β-turn and unordered fractions). Although this assumption has produced some fairly accurate results, especially in the determination of helical and β-sheet fractions, it may be overly simplistic in light of the various types of structure that exist within each major class of secondary structure. Due to the inherent variability of structure in proteins it is unlikely that single

reference spectra, which represent the four major structures, can be used to calculate the secondary structures of all proteins correctly and thus discrepancies between X-ray and CD results would be expected.

Hennessey and Johnson (1981) presented a method in which eight types of secondary structure were considered: helix; parallel and anti-parallel β-strands; types I, II and III β-turn; all other β-turns combined; and other structures. The authors found that there was good correlation between the X-ray structures and the structures obtained from the analysis of the CD spectrum, although a total agreement for all proteins was still not achieved. The underlying assumption of all CD studies is that the structure determined for a protein from X-ray data is the same as that determined by circular dichroism analysis although the protein is in two different states: solid (crystal) state for X-ray diffraction and an aqueous solution for CD. Chen et al. (1972) stated that this assumption may be reasonable since protein crystals prepared for X-ray diffraction analysis contain considerable amounts of water. Drenth (1981) reported that 50 percent of the interstitial spaces between protein molecules in good size crystals are occupied by water molecules. However, one cannot discount the possibility that changes in protein conformation do occur when crystals are dissolved in an aqueous solution. Such changes in protein conformation may contribute to the lack of total agreement between CD and X-ray results.

Closer agreement between X-ray and CD data may also result from an improvement in the accuracy of the CD data. Hennessey and Johnson (1982) identified and examined the types of errors that were associated

with the analysis of protein circular dichroism spectra for secondary structure determination. Three errors were related to the operation of the spectropolarimeter and included wavelength synchronization, spectral bandwidth and scan speed; three additional errors were experimental and included intensity adjustments and two sources of baseline shift. Since the CD spectral data used in this study were obtained from Provencher (1980), any error in the CD data would be consistent for all the algorithms examined.

In methods which assume that the observed ellipticity is a linear function of the secondary structure fractions, the possibility exists that interactions may occur between the various secondary structure fractions. The initial model treats each secondary structure as a separate entity, however, this may be in error if such interactions exist. Gayle and Bennett (1978) examined the consequences of model departures on the resolution of multicomponent spectra and found that interactions may be present among the independent variables, thus yielding incorrect results.

CONCLUSIONS

The objective of the present study was to compare four different algorithms for their ability to accurately determine protein secondary structure in relation to X-ray data. The majority of methods examined showed highly significant ($P<0.001$) correlation coefficients in relation to α-helix determination when compared to X-ray data; the simplex optimization method for $\bar{n}$ = variable in the presence of concanavalin A was significant at the $P<0.05$ level. The methods of Chang et al. (1978), Siegel et al. (1980) and Provencher and Glöckner (1981) also showed highly significant ($P<0.001$) correlation coefficients for β-sheet determination, although these coefficients were slightly lower than those determined for α-helix. The correlation coefficients for β-sheet obtained for the simplex optimization method were significant at the $P<0.05$ level with the exception of r for $\bar{n}$ = variable (+ con A) which was found to be not significant ($P>0.05$). None of the algorithms examined were able to produce significant correlation coefficients ($P>0.05$) when the β-turn fraction was determined. Random coil determination resulted in significant ($P<0.05$) correlation coefficients from the simplex optimization method in the absence of concanavalin A and from the method of Chang et al. (1978). In the absence of concanavalin A, the simplex optimization method with $\bar{n}$ = 10.4 compared favourably to the method of Chang et al. (1978) and thus could serve as an alternate algorithm to solve least squares. Of the four methods examined, the method of Provencher and Glöckner (1981) showed the least change in the

correlation coefficients when concanavalin A was removed from the analysis, thereby demonstrating the greatest versatility for secondary structure determination.

PART II. *A physical-chemical study of aspartyl proteinases as related to enzymatic activity*

INTRODUCTION

The clotting of milk by proteolytic enzymes during the cheese-making process represents one of the oldest operations in food technology. References to cheese have been found on Sumerian cuneiform tablets dating from 4000 B.C. (Hofmann, 1974). Traditionally, chymosin (rennin) a proteolytic enzyme prepared from calf stomach, has been used for cheese-making; however, due to increases in cheese consumption and shortages of chymosin, alternative enzymes have been examined to replace chymosin. These replacements have included such proteinases as porcine pepsin, chicken pepsin and enzymes produced from microbial sources (Kay and Valler, 1981).

The proteinases which are successfully used for cheese-making belong to the class of enzymes known as aspartyl proteinases. These enzymes are characterized by the presence of two aspartic acid residues in the active site as well as having a general optimal activity in the pH range of 1.5 to 5.0, depending on the particular enzyme and substrate combination being studied (Voynick and Fruton, 1971; Dalgleish, 1982). In addition, sequence and structural homology have been demonstrated for a number of the aspartyl proteinases (Foltmann and Pedersen, 1976; Hsu et al., 1977).

The success of an enzyme used in cheese-making lies not only in its ability to clot milk, but in the relationship between milk-clotting ability and general proteolytic ability (Dalgleish, 1982). Milk-

clotting ability is defined as the specific hydrolysis of the Phe_{105}-Met_{106} bond of κ-casein necessary for the initiation of milk-clotting (Dalgleish, 1982). Milk-clotting enzymes suitable for cheese-making should combine high clotting activity with low proteolytic activity (Visser, 1981). The enzymes trypsin and papain, which are representative of serine and sulfhydryl proteinases, respectively, are able to clot milk, but due to their high general proteolytic ability the clot is rapidly degraded.

Even within the aspartyl proteinase class, the enzymes show dissimilarity in their action. Chymosin has been demonstrated to have a high milk-clotting to proteolytic activity ratio, whereas porcine pepsin has a low ratio. This difference in activity may be the result of subtle conformational differences (Visser, 1981).

The ability to quantitate relationships between structure and function of a molecule is a problem that has continually perplexed scientists. It is generally accepted that the function of a protein cannot be understood until its structure is known (Barry et al., 1974). Hydrophobic, electrostatic and steric forces are three parameters that affect structure and may be used to predict function (Stuper et al., 1979). Although previous researchers have measured various physical and structural properties of some aspartyl proteinases, no attempts have been made to quantitate the nature of the relationship between structure and function. The major objectives of the present research were as follows: firstly, to measure various structural and intrinsic properties of some aspartyl, as well as some non-aspartyl proteinases; secondly, to attempt to classify the proteinases based on these properties

using multivariate statistical techniques (e.g. principal component analysis); and thirdly, to regress the milk-clotting to proteolytic activity ratio on the principal components generated from the structural and intrinsic parameters in order to identify parameters important for this functional property. By identifying these parameters it may be possible in the future to modify enzymes to achieve a desired function in food processing.

LITERATURE REVIEW

A. ASPARTYL PROTEINASES

1. Classification and sources

The majority of the known proteinases have been classified into four distinct classes according to the nature of the groups involved in the catalytic process (Hartley, 1960; Voynick and Fruton, 1971). These four enzyme classes include the serine, sulfhydryl, metallo- and aspartyl proteinases. The serine proteinases such as chymotrypsin, trypsin, elastase, subtilisin, etc. are characterized by the presence of reactive seryl and histidyl residues at the catalytic site. In the sulfhydryl proteinases such as papain, ficin and bromelain, a cysteinyl and a histidyl residue are involved in catalysis. The metalloproteinases require a metal ion for activity and a variety of metal ions are associated with this class of enzymes. Such associations can include iron (e.g. cytochrome oxidase, catalase, peroxidase), copper (lysyl oxidase), zinc (NADP-linked dehydrogenases, carboxypeptidase) and manganese (arginase) (Lehninger, 1982). The aspartyl proteinases, orignally known as the acid proteinases, generally have maximal activity in the acid pH range and have one or more catalytically important carboxyl groups in their active centres.

It was suggested that the name acid proteinases be changed to aspartyl proteinases in keeping with the enzyme classification scheme based on the nature of the specific group(s) located at the active site (Hofmann, 1974; Foltmann and Pedersen, 1976; Visser, 1981). This suggestion is valid if one considers the example of renin, an endopeptidase

found in the kidney, which has an optimal pH of about 6 although its catalytic residues and mechanism belong to the aspartyl proteinases (Hofmann, 1974). Several catalytic features are shared by the aspartyl proteinases, such as inhibition by pepstatin (a pentapeptide-like compound isolated from Streptomyces) as well as inhibition by the active centre inactivators diazoacetyl-norleucine methylester (DAN) and 1,2 epoxy-3-(ρ-nitrophenoxy) propane (EPNP). Inhibition by DAN and EPNP has demonstrated that two catalytically important aspartic acid residues are present in most of these enzymes (Tang, 1976 and 1979).

The aspartyl proteinases have been isolated from a wide variety of sources including: gastric sources (e.g. pepsin, chymosin, gastricsin); microbial sources (e.g. proteinases from Mucor miehei and Mucor pusillus); lysosomes (e.g. cathepsin D and E); kidney (renin); protozoan (e.g. proteinase from Tetrahymena pyriformis); as well as plant sources (proteinase from Sorghum vulgare).

2. Gastric aspartyl proteinases

Probably the best known and most well characterized of the aspartyl proteinases are those isolated from gastric sources such as the enzymes pepsin and chymosin. It is generally accepted that the gastric proteinases may be classified into three main groups, pepsin (EC 3.4.23.1), gastricsin (EC 3.4.23.3) and chymosin (EC 3.4.23.4) (Foltmann, 1981). Only the two more important enzymes, pepsin and chymosin, will be discussed here.

(a) <u>Pepsin</u>

The term pepsin is used for the gastric proteinases formed from the inactive zymogen pepsinogen. This conversion of inactive pepsinogen to active pepsin occurs as a result of limited proteolysis that cleaves forty-eight residues from the N-terminal end of pepsinogen. Although multiple forms of pepsin exist, the name pepsin is generally used in the literature in reference to porcine pepsin A, the most extensively studied form of the pepsins.

The primary amino acid sequence of porcine pepsin has been determined and consists of 327 amino acid residues (Tang et al., 1973; Moravek and Kostka, 1974). Examination of this primary sequence indicates that pepsin is unusual because of its exceptionally low content of basic amino acids. The sequence contains one lysine, one histidine and two arginine residues.

The X-ray crystal structure of pepsin has been described at 2.7Å resolution (Andreeva et al., 1976). In this study it was found that the molecule consisted of two lobes separated by a cleft, and the active site aspartic acid residues (Asp_{32} and Asp_{215}) were located within the cleft. It was also found that a majority of the residues were involved in the β-sheet conformation.

The proteolytic ability of pepsin has been well studied. The enzyme has broad side chain specificity and can cleave many types of peptide bonds. In general, the sensitive bonds are present in dipeptide units containing at least one hydrophobic amino acid residue such as phenylalanine, tyrosine, leucine or methionine; however, a number of

exceptions to this generalization have been noted (e.g. scission of Glu-Asn bond in the A chain of bovine insulin; Glu-Lys and Asp-Pro bonds in the β-chain of human hemoglobin A) (Hill, 1965; Fruton, 1970).

Porcine pepsin is irreversibly inactivated at pH values above 6.0 (Fruton, 1970). It has been suggested that hydrogen bonds associated with the carboxyl groups are broken at pH values greater than 6 (Edel-hoch, 1958a). Upon alkaline denaturation, changes in viscosity have been observed which would indicate that the protein is unfolded to a linear polyelectrolyte (Edelhoch, 1957). Below pH 6.0 pepsin may be stablized by hydrophobic bonding since it is unaffected by heating to 60°C or by treatment with 4M urea or 3M guanidium chloride (Perlmann, 1959; Blumenfeld et al., 1960).

(b) Chymosin

Chymosin is the predominant milk-clotting enzyme from the fourth stomach of the calf (Foltmann, 1966) and is found in the form of rennet, the crude extract from the calf stomach. The first attempts to isolate the enzymatically active component of rennet were published in 1840 by Deschamps (as cited in Foltmann, 1966), at which time Deschamps named the enzyme chymosin. It was suggested by Foltmann (1971) that the name chymosin be used for the pure enzyme, instead of the name rennin, to avoid confusion with the proteolytic enzyme renin isolated from the kidney.

Chymosin, like most gastric proteinases, is initially secreted as an inactive precursor, and in this state is called prochymosin. The conversion of prochymosin to the active enzyme takes place through a

limited proteolysis which removes forty-two residues from the N-terminal segment of the peptide chain (Foltmann, 1981). This limited proteolysis results in a molecular weight reduction of approximately 14 percent (Foltmann et al., 1977).

The primary amino acid sequence of chymosin has been determined and contains 323 amino acid residues (Foltmann et al., 1977). Attempts to prepare crystals appropriate for X-ray diffraction analysis have been unsuccessful (Bunn et al., 1971; Jenkins et al., 1976), therefore, no exact secondary or tertiary structural information is available.

Compared to the other aspartyl proteinases, chymosin is distinguished by having a high milk-clotting to proteolytic activity ratio, a property that has been fully exploited in cheese-making. Like pepsin, chymosin is highly specific for peptide bonds adjacent to and preferably between a pair of hydrophobic residues (Fox, 1981). However, studies with various substrates (oxidized β-chain of insulin and ribonuclease) have indicated that chymosin is much more limited in its proteolytic properties (Bang-Jensen et al., 1964). The rapid and specific proteolytic cleavage of κ-casein by chymosin during the initial step of the milk-clotting process has been the subject of numerous investigations (Bang-Jensen et al., 1964; de Koning, 1968; Jollès et al., 1968; Raymond et al., 1973; Visser et al., 1976 and 1977; Raap et al., 1983). Studies (Visser et al., 1976 and 1977) with synthetic peptide substrates for chymosin have indicated that the residues around the sensitive Phe_{105}-Met_{106} bond of κ-casein, as well as chain length are important factors for hydrolysis of κ-casein. On the basis of these results and results

obtained with pepsin (Raymond and Bricas, 1979), it was concluded that chymosin made higher demands upon the structure of its substrate than pepsin, which was therefore responsible for its (chymosin) limited proteolytic behaviour (Visser, 1981).

Foltmann (1959) investigated the stability of the enzymatic activity of chymosin as a function of pH and found that stability was good between pH 5.8 and 6.3. At pH values above 6.5 enzymatic activity decreased rapidly with increasng pH. Mickelson and Ernstrom (1963) reported similar results in that maximum stability of chymosin was maintained from pH 4.6 to 6.5. Schober et al. (1960) found that solutions of chymosin were inactivated at pH 7.0. The inactivation was accompanied by an increase in ninhydrin reaction, which was concluded by the authors to be the result of autolysis. Foltmann (1966) suggested that the autolytic decomposition of chymosin seen at pH 7.0 was the result of the chymosin molecule unfolding or rearranging "in such a way that as a substrate it is easily accessible to the slight proteolytic activity which is present even at pH 7.0". Andren and de Koning (1982) examined the effect of temperature and pH on the milk-clotting ability of chymosin and found that as pH and/or temperature were increased, clotting activity decreased.

3. Microbial proteinases

In recent years the supply of chymosin has decreased while the demand for cheese has increased. Accordingly, there has been an active search for enzymes with high milk-clotting and low proteolytic

activities which could be used as substitutes for chymosin (Fox, 1969; Kay and Valler, 1981). Recently, a number of microbial proteinases have been used commercially for the production of cheese. These enzymes, like the gastric enzymes pepsin and chymosin, are classified as aspartyl proteinases. Unlike the gastric proteinases which are initially secreted as zymogens, no zymogens for microbial aspartyl proteinases have been found (Foltmann and Pedersen, 1976).

(a) Mucor pusillus var. Lindt proteinase

Arima et al. (1967) isolated a milk-clotting proteinase from Mucor pusillus var. Lindt after an extensive search of approximately 800 strains of micro-organisms. Iwasaki et al. (1967a) found that the enzyme had a milk-clotting to proteolytic activity ratio similar to that of chymosin which made it suitable as a chymosin substitute.

No primary amino acid sequence data has been published for Mucor pusillus proteinase; however, based on amino acid composition data, the enzyme has between 277 and 281 residues (Arima et al., 1970). D-galactose and D-glucosamine have also been found to be associated with the proteinase (Etoh et al., 1979).

Yu et al. (1970) examined various synthetic peptides in order to determine the hydrolytic specificity of the proteinase and found that the specificity was similar to that of pepsin and chymosin. Peptide units containing at least one hydrophobic amino acid residue were found to be the most susceptible to hydrolysis.

Iwasaki et al. (1967a and b) found that at pH values greater than 6.5 there was a rapid drop in both milk-clotting and proteolytic

activity. This loss of activity at pH values near neutrality is consistent with that observed for other aspartyl proteinases (Kay and Valler, 1981).

(b) Mucor miehei proteinase

In 1970, Ottesen and Rickert (1970a) published a paper on the isolation and partial characterization of an acid protease from the fungus Mucor miehei. The authors found that like the proteinase produced from Mucor pusillus var. Lindt, Mucor miehei proteinase was also able to clot milk and therefore, resembled pepsin and chymosin. From inhibition studies using DAN and EPNP it was concluded that M. miehei proteinase belonged to the aspartyl proteinases (Rickert and McBride-Warren, 1977).

Compositional analysis has shown that M. miehei proteinase is a glycoprotein with about 370 amino acid residues (Ottesen and Rickert, 1970a and 1970b; Rickert and Elliott, 1973; Rickert and McBride-Warren, 1974). The proteinase has approximately 6 percent carbohydrate which is composed of hexosamine and neutral hexoses (Ottesen and Rickert, 1970a and 1970b).

Bech and Foltmann (1981) recently published a partial primary structure of M. miehei proteinase and found that a large number of amino acids in the N-terminal domain of the molecule were similar to that of calf chymosin. The researchers also found that amino acid residues that would be expected to participate in the catalytic mechanism were found in the same invariant positions as for other members of the aspartyl proteinases.

Rickert (1970), using the β-chain of oxidized insulin as a substrate, found that only those bonds involving aromatic amino acid residues were hydrolyzed by the proteinase. Sternberg (1972) working with synthetic substrates also found that peptide bonds having an aromatic amino acid (on the N-terminal side of the peptide bond) were hydrolyzed by Mucor miehei proteinase. Enzymatic activity for the proteinase is maximal between pH 3.0 and pH 6.0; outside this range activity is lost at a rapid rate (Ottesen and Rickert, 1970a).

(c) Endothia parasitica proteinase

A chymosin-like proteinase is produced by the fungus Endothia parasitica (Tam and Whitaker, 1972). This proteinase catalyzes both a specific proteolysis leading to the clotting of milk, as well as a general proteolysis of proteins (Whitaker, 1970).

No primary amino acid sequence data for E. parasitica proteinase has been published to date although Whitaker (1970) found that the molecule contains approximately 330 amino acid residues.

Preliminary work on the X-ray crystallographic data of the proteinase was carried out by Moews and Bunn (1970) and was later refined by Jenkins et al. (1975) and Jenkins et al. (1976). Jenkins et al. (1976) described the structure of the proteinase as consisting of two lobes which were separated by "a deep and extensive cleft" which formed a hydrophobic pocket. Within the cleft the two active site aspartic acid residues were found and were in close proximity to one another. As with pepsin and penicillopepsin, whose crystal structures

have been determined, X-ray data indicated that a large proportion of the secondary structure consists of β-sheet structures (Jenkins et al. 1975; Jenkins et al., 1976). This observation was confirmed by the circular dichroism analysis of the proteinase where three different algorithms were used to analyze the CD spectra (Jenkins et al., 1976).

Hydrolysis of the oxidized β-chain of insulin showed that E. parasitica proteinase had a specificity for the hydrophobic regions similar to that of pepsin, although peptide maps of digests of the oxidized β-chain of insulin by E. parasitica proteinase and chymosin showed more extensive proteolysis by the former (Whitaker, 1970; Williams et al., 1972).

E. parasitica proteinase shows maximum stability at pH 3.8 to 4.5. Below pH 2.5 loss in activity is associated with an increase in ninhydrin reactive groups which may be attributed to autolysis. Above pH 6.5 activity is lost rapidly and above pH 8.0 activity is lost almost instantaneously; no increase in ninhydrin reactive groups is associated with this loss of activity (Whitaker, 1970).

(d) Penicillopepsin

Pencillopepsin is the acid proteinase produced by the mold Penicillium janthinellum at pH values less than 4. The enzyme is produced once mycelial growth has ceased and sporulation has begun (Sodek and Hofmann, 1970b).

The primary amino acid sequence for the proteinase has been determined (Hsu et al., 1977; James and Sielecki, 1983) and consists of

323 amino acid residues. The amino acid composition is somewhat comparable to that of pepsin in that the number of the various residues (i.e. hydrophilic, hydrophobic) are similar. X-ray diffraction analysis at 1.8Å resolution (James and Sielecki, 1983) has shown that the polypeptide folds into 2 hydrophobic domains with a cleft separating the two. As with pepsin and E. parasitica proteinase, the active site aspartic acid residues were found in the cleft with each domain contributing one aspartyl residue. Secondary structure analysis of the proteinase based on the X-ray analysis indicated that more than two-thirds of the residues (218/323) were involved in parallel or anti-parallel β-sheets (Sodek and Hofmann, 1970c; James and Sielecki, 1983).

Using the oxidized β-chain of insulin as a substrate, penicillopepsin showed a hydrolytic specificity similar to pepsin (Sodek and Hofmann, 1970a; Mains et al., 1971). Peptide bonds which had a hydrophobic amino acid in the P_1' position (as defined by Berger and Schechter, 1970) were preferentially cleaved by penicillopepsin (Mains et al., 1971). To date no information pertaining to milk-clotting activity has been reported for this proteinase.

pH stability studies (Sodek and Hofmann, 1970b) have shown that maximum stability is between pH 2.2 and 6.6; beyond this range there is a rapid drop in activity. The stability, however, is both time and temperature dependent.

(e) Aspergillus saitoi proteinase

Yoshida (1956) isolated a proteinase from the mold Aspergillus

saitoi which was capable of activating trypsinogen and chymotrypsinogen A at acidic pH values.

No primary amino acid sequence data for this proteinase has yet been published, however, from the amino acid composition data it is estimated that the proteinase contains between 283 and 289 amino acid residues. The proteinase contains a high amount of the hydroxy amino acids serine and threonine (Ichishima and Yoshida, 1965).

Conformational studies using optical rotatory dispersion (ORD) have indicated that A. saitoi proteinase contains little if any α-helix (Ichishima and Yoshida, 1966a and 1967). In the infrared spectrum of the deuterium-exchanged proteinase an amide I band was located at 1632 cm^{-1}, which would suggest that the enzyme contains anti-parallel β-structure (Ichishima and Yoshida, 1966a).

Little is known about the hydrolytic specificity of A. saitoi proteinase other than the activation of trypsinogen by cleavage of a lysine-isoleucine bond, resulting in an active enzyme (Gabeloteau and Desnuelle, 1960). In work with synthetic substrates, it was found that Aspergillus saitoi proteinase had hydrolytic properties that were "unique among the well known proteinases" (Yoshida and Nagasawa, 1956b). Fukumoto et al. (1967) examined the milk-clotting ability of various aspartyl proteinases and found that Aspergillus saitoi protein-ase was inactive.

The proteinase shows good stability in the pH range of 2 to 6, however, beyond this range activity is rapidly lost (Yoshida and Nagasawa, 1956a; Ichishima, 1970).

B. ENZYMATIC COAGULATON OF MILK

The enzymatic coagulation or clotting of milk by certain proteinases is a process that may be divided into three stages: primary, secondary and tertiary stages. During the primary stage the bond between Phe_{105}-Met_{106} of κ-casein in the casein micelle is specifically attacked to yield two peptides: glycomacropeptide and para-κ-casein. These two peptides have very different properties. Glycomacropeptide or caseinomacropeptide, which includes residues 106 to 169, is hydrophilic and soluble and will diffuse away from the micelle subsequent to κ-casein attack. Para-κ-casein, on the other hand, which includes residues 1 to 105, is very hydrophobic and remains associated with the micelle. The progressive hydrolysis of κ-casein during the primary stage results in alterations of casein micelle properties and subsequently leads to aggregation. The aggregation of casein micelles is the secondary stage of the milk-clotting process. The primary and secondary stages of milk-clotting are not readily distinguishable since the aggregation of the casein micelles is initiated before the complete hydrolysis of κ-casein by the milk-clotting proteinases (Mackinlay and Wake, 1971; Dalgleish, 1982). The tertiary stage of the milk-clotting process is the least clearly defined of the three stages, however, it includes syneresis, the expulsion of water by the curd arising from a structural rearrangement after clot formation, and non-specific proteolysis of the various casein components in the clot (Dalgleish, 1982).

Although the satisfactory clotting of milk is dependent on the cleavage of κ-casein at or very near the Phe_{105}-Met_{106} bond, there is

also a need for a more general interaction between enzyme and substrate. This interaction is necessary to promote the correct binding which allows for the hydrolysis of κ-casein at the appropriate site (Dalgleish, 1982). Most enzymes which successfully clot milk belong to the aspartyl proteinases, however, the specific hydrolysis of κ-casein by these enzymes is not entirely attributable to the particular specificity of the enzymes themselves (Dalgleish, 1982). Studies (Hill, 1969; Trout and Fruton, 1969; Schattenkerk and Kerling, 1973; Powers et al., 1976; Visser et al., 1976; Visser et al., 1980) concerning the interaction of enzymes and peptide substrates resembling the $Phe_{105}-Met_{106}$ area of κ-casein have shown that composition, sequence and length of the substrate are important determinants regarding enzyme-substrate interaction prior to κ-casein hydrolysis.

C. MULTIVARIATE ANALYSIS

One of the primary goals of scientific research is the collection of data. It is now possible with modern technology to generate large amounts of data wherein several parameters are measured for a single sample. However, once these data are collected one must be able to critically evaluate the data so that useful information can be extracted. The use of multivariate data analysis techniques allows for the efficient simplification and interpretation of many different variables simultaneously such that maximum information and minimum noise are obtained (Derde and Massart, 1982). Gower (1982) defined multivariate analysis as those statistical methods concerned with either the analysis

of data on many variables or the display of results in many dimensions, or both. By using multivariate techniques the conclusions made are more precise than those obtained from the individual input data since the effect of random noise is decreased; this is similar to averaging over replicate measurements (Martens, 1982). Multivariate analysis techniques have been extensively used in analytical chemistry, medicine and the social sciences. Recently, a number of papers have also appeared where the use of such techniques (e.g. stepwise linear discriminant analysis, principal component analysis, etc.) have been applied to food research. Several authors (Kowalski, 1980; Frank and Kowalski, 1982; Martens and Russwarm, 1982) have reviewed the use of multivariate analysis techniques in food research.

1. Principal component analysis

Principal component analysis (PCA) is a data transformation technique that is concerned with the total variance of the variables (Daultry, 1976). This technique reduces the number of original variables to a smaller number of new variables or components that explain as much of the observed variance as in the original variables. The intent is to explain the maximum amount of observed systematic variation in the data with the fewest possible components. The principal components derived are mutually uncorrelated and are weighted according to the amount of total variance they represent, such that the first component has the largest variance, the second component accounts for as much of the remaining variance as possible while being uncorrelated with the first, and so on (Daultry, 1976; Hoffman and Young, 1982).

There are four main steps in principal component analysis which include the following: firstly, the correlation or covariance matrix is generated; secondly, component loadings are obtained; thirdly, the component scores are computed; and fourthly, two dimensional plots of various principal component pairs are constructed. A basic outline of the procedure is given in Figure 3. Data for N objects (i = 1 to N) for which R variables (j = 1 to R) are measured can be represented by a data matrix [O]. The measurements for each variable are first standardized (matrix [A]) so that no single variable dominates the analysis (Martens, 1982). A correlation matrix [B] is then calculated for the R variables. Eigenvalues and eigenvectors for the correlation matrix are then determined. The eigenvalues generated are an indication of the amount of variance in the data that is accounted for by the principal component associated with that eigenvalue. Generally, only principal components with eigenvalues greater than or equal to one are used in the analysis. In order to obtain the component loading matrix [L], the eigenvector matrix [E] is multiplied by the square root of the eigenvalues $[\Lambda]^{1/2}$. The component loadings describe the correlation between each variable and each component. Squaring the elements of the component loading matrix produces the proportion of the variance that each variable contributes to each component. Variables that have high loadings on a component tend to be highly correlated to one another while variables which do not have similar loadings are less highly correlated. In order to calculate the principal component scores for each object i in the various principal components, the vector of the

Figure 3. Outline of the procedure for principal component analysis.

$$[O] = \begin{matrix} X_{11} & X_{12} & \cdots & X_{1R} \\ X_{21} & \bullet & & \bullet \\ \bullet & \bullet & & \bullet \\ \bullet & \bullet & & \bullet \\ \bullet & & & \\ X_{N1} & & \cdots & X_{NR} \end{matrix}$$

Original data matrix [O]
    Objects i = 1 to N
    Variables j = 1 to R

$$[A] = \begin{matrix} S_{11} & S_{12} & \cdots & S_{1R} \\ S_{21} & \bullet & & \bullet \\ \bullet & \bullet & & \bullet \\ \bullet & \bullet & & \bullet \\ \bullet & & & \\ S_{N1} & & \cdots & S_{NR} \end{matrix}$$

Standardized data matrix [A]
    Objects i = 1 to N
    Variables j = 1 to R

$$[B] = \begin{matrix} C_{11} & C_{12} & \cdots & C_{1R} \\ C_{21} & \bullet & & \bullet \\ \bullet & \bullet & & \bullet \\ \bullet & \bullet & & \bullet \\ \bullet & & & \\ C_{R1} & & \cdots & C_{RR} \end{matrix}$$

Correlation matrix [B]

Solve

Eigenvalue matrix [Λ] and Eigenvector matrix [E]

Component loading matrix [L]
$$[L] = [E] * [\Lambda]^{1/2}$$
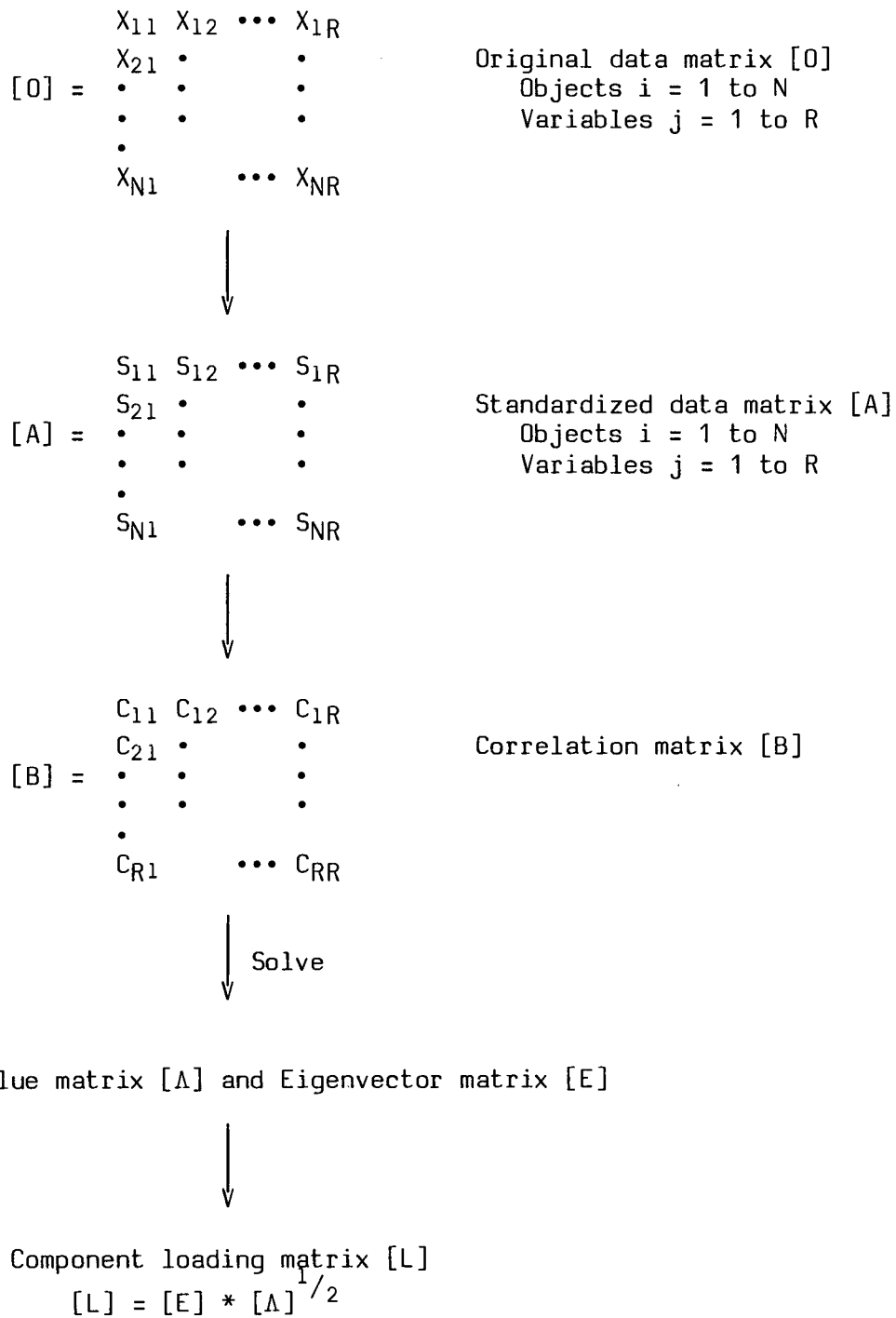
Figure 3. (continued)

$$
[L] = \begin{matrix}
PC_{11} & PC_{12} & \cdots & PC_{1R} \\
PC_{21} & \bullet & & \bullet \\
\bullet & \bullet & & \bullet \\
\bullet & \bullet & & \bullet \\
\bullet & & & \\
PC_{R1} & & \cdots & PC_{RR}
\end{matrix}
$$

The loadings for each principal component are represented by the column vectors of the matrix.

$$
[V] = \begin{matrix}
PC_{11}{}^2 & PC_{12}{}^2 & \cdots & PC_{1R}{}^2 \\
PC_{21}{}^2 & \bullet & & \bullet \\
\bullet & \bullet & & \bullet \\
\bullet & \bullet & & \bullet \\
\bullet & & & \\
PC_{R1}{}^2 & & \cdots & PC_{RR}{}^2
\end{matrix}
$$

Explained variance matrix [V]

Principal component scores matrix [Y]

$$[Y] = [A] * [E]$$

$$
[Y] = \begin{matrix}
PCS_{11} & PCS_{12} & \cdots & PCS_{1R} \\
PCS_{21} & \bullet & & \bullet \\
\bullet & \bullet & & \bullet \\
\bullet & \bullet & & \bullet \\
\bullet & & & \\
PCS_{N1} & & \cdots & PCS_{NR}
\end{matrix}
$$

The principal component scores (PCS) for each object i in each principal component j are represented by the values in each column vector.

Two dimensional plots of the various principal components

standardized measurements for the various R variables for object i are multiplied by the eigenvector matrix [E]. Once the component scores (matrix [Y]) have been calculated, two dimensional plots of various pairs of principal components may be plotted in order to characterize the various N objects. Since two dimensional plots are an integral part of principal component analysis (PCA), PCA is referred to as a linear display method (Derde and Massart, 1982).

Recently, a number of papers have appeared in the literature regarding the use of PCA in food related problems. In 1979, Aishima published a series of papers (Aishima, 1979a; 1979b; 1979c; Aishima et al., 1979) in which principal component analysis was used as an objective means for the differentiation of soy sauces based on gas chromatographic profiles. Kwan and Kowalski (1980) used principal component analysis on both objective chemical measurements and subjective sensory evaluation data from various Pinot Noir wines. The authors found that it was possible to correlate the principal components generated from the two groups of measurements (chemical and sensory), and reported that this type of approach would eventually lead to "a better understanding of the direct stimulus-response mechanism". Martens (1982) published a paper outlining the principles of principal component analysis and then cited an example of its application which involved African finger millet samples grown under various growth conditions. Sixteen amino acids were measured in the millet samples and by plotting the first two principal components it was possible to show the millet's response to increasing nitrogen fertilization. Wold et al. (1982) used principal component

analysis for the analysis of gas chromatographic (GC) data from fresh and stored swedes (rutabagas) and found it was possible to separate the two based on their GC profiles. Other applications of principal component analysis as related to food research and data analysis have been compiled by Martens and Russwurm (1982).

MATERIALS AND METHODS

A. MATERIALS

Proteinases used in this study were obtained from a variety of sources. Porcine pepsin (pepsin A, EC 3.4.23.1) two times crystallized and chymosin (rennin, EC 3.4.23.4) were obtained from Sigma Chemical Co. (St. Louis, MO). Mucor miehei proteinase and Endothia parasitica proteinase were generous gifts from Pfizer Canada (Kirkland, PQ). Mucor pusillus var. Lindt proteinase was a generous gift from Dr. S. Iwasaki (Meito Sangyo Co., Ltd., Tokyo, Japan). Aspergillus saitoi proteinase was obtained from Calbiochem (San Diego, CA). Crystals of penicillo-pepsin were kindly supplied by Dr. T. Hofmann (University of Toronto, Toronto, ON). The non-aspartyl proteinases, namely papain (two times crystallized), trypsin (two times crystallized from bovine pancreas) and α-chymotrypsin (three times crystallized from bovine pancreas) were obtained from Sigma Chemical Co. (St. Louis, MO). The purity of the proteinases was determined using the SDS polyacrylamide gel electrophoresis method of Laemmli (1970). All proteinases used in the present study were shown to be at least 80 percent pure.

The fluorescent probe cis-parinaric acid (9,11,13,15-cis, trans, trans, cis-octadecatetraenoic acid) was purchased from Molecular Probe Ltd. (Junction City, OR). The sodium salt of 1-anilino-8-napthalene sulfonic acid was obtained from Eastman Kodak (Rochester, NY). The protein carrier 3,3'-dimethylbiphenyl used for zeta potential

measurements was a product of Aldrich Chemical Co., Inc. (Milwaukee, WI). Fluorescamine was purchased from Chemical Dynamics Corp. (South Plainfield, NJ).

Unless otherwise stated, reagent grade chemicals were used throughout the study. Glass distilled, deionized water was used in the preparation of all solutions and buffers.

B. DIAGONAL PLOT METHOD

In order to determine primary amino acid sequence homologies between proteinases, the diagonal plot method of Beynon (1982) was used. Briefly, in this method the two sequences are aligned along the normal axes and a point is drawn in the enclosed area where two residues are identical. A computer program was written in Fortran IV to perform the comparison of primary amino acid sequences of two proteinases (Appendix 5).

C. SECONDARY STRUCTURE PREDICTION

For those proteinases with known primary amino acid sequences, secondary structures were predicted using a computerized version of the Chou and Fasman method (1978b) as modified by Pham (1981) which was run on an Amdahl 470 V/8 computer. The primary sequence was converted into a numerical sequence which was then entered into four separate computer programs to determine regions of α-helix, β-sheet, β-turn and overlap regions (i.e. regions where α-helix and β-sheet regions overlap). A predicted secondary structure was drawn on the basis of the computer output.

In addition to the Chou and Fasman method (1978b), secondary structure was also predicted using the hydrophobicity profile method of Cid et al. (1982). A computer program was written in Fortran IV for this method (Appendix 6). The primary amino acid sequence was again digitized for use in the computer program.

D. CIRCULAR DICHROISM

CD spectra were measured using a JASCO J-500A spectropolarimeter (Japan Spectroscopic Co., Ltd., Tokyo, Japan) under a constant nitrogen flush at 20°C. The instrument was calibrated by a two-point calibration technique at wavelengths 290.5 and 192.5 nm using a 600 mg/L solution of d-10-camphorsulfonic acid (Chen and Yang, 1977). The corresponding molar ellipticities, $[\Theta]$, are 7800 and -15600 deg cm^2 dmol^{-1}, respectively.

1. Sample preparation

Proteinases were dissolved in either phosphate or acetate buffer ($^\Gamma/_2$ = 0.01) depending on the pH of the assay. Acetate buffer was used for pH values 5.0, 5.3 and 5.8, while phosphate buffer was used for pH values 6.3, 7.0 and 8.0. The protein solutions were then filtered through a 0.45 μm Millex-HA filter (Millipore Corporation, Bedford, MA) prior to CD analysis. Analyses were carried out within 1.0 h of sample preparation.

2. Far-UV spectra (190 to 240 nm)

Between 190 and 240 nm a 1.0 mm cell was used with a protein concentration of approximately 0.1 mg/mL in the appropriate buffer. The spectral data were reported in terms of [θ], the molar ellipticity per residue (mean residue ellipticity). The molar ellipticities were not corrected for the refractive index factor. In order to obtain [θ] data the signal from the photomultiplier of the spectropolarimeter was digitized into an analogue reading via a digital voltmeter (Schlumberger-Solartron A220). The signal was then transferred to a Texas Instruments Silent 700 data acquisition terminal via a data transfer unit (Schlumberger-Solartron) and a hard copy of the data was obtained. Each protein solution was measured a minimum of three times. The baseline spectrum for each protein sample was obtained by running the appropriate buffer under the identical conditions used for the sample. The average digitized signal at 1 nm intervals from 240 to 190 nm was entered into a computer program (Appendix 7) which allowed for baseline correction and subsequently converted the data into [θ] based on the following equation:

$$[\theta]_{MRW_\lambda} = \frac{\theta * MRW}{10 * d * c} \qquad \textbf{Eq. 20}$$

where λ = wavelength

θ = observed ellipticity in degrees

MRW = mean residue weight

c   = concentration in gm/mL

d   = pathlength in cm

The values for MRW were calculated from the ratio of molecular weight to total number of residues in the protein. The values for molecular weight, total number of residues and MRW are summarized in Table 4. Secondary structure fractions for each of the protein samples were determined using the constrained regularization procedure of Provencher and Glöckner (1981) by entering the [θ] at 1 nm intervals (from 240 to 190 nm) into the computer program.

3. Near-UV spectra (240 to 320 nm)

Between 240 and 320 nm a 10.0 mm cell was used with a protein concentration of approximately 1.0 mg/mL in the appropriate buffer. The CD spectra were expressed in terms of $\Delta\varepsilon$ (difference in molar absorptivity between left and right circularly polarized light). The $\Delta\varepsilon$ was calculated according to the following equation:

$$\Delta\varepsilon = \frac{[\theta]_{MRW} * N}{3300} \qquad \textbf{Eq. 21}$$

where N = total number of residues in the protein.

The use of $[\theta]_{MRW}$ in the near-UV is not correct since CD intensity is a function of only a few amino acid side chains rather than all the amino acids in the protein (Strickland, 1974).

E. BIGELOW AVERAGE HYDROPHOBICITY

Bigelow (1967) formulated an equation (Eq. 22) to calculate average hydrophobicity ($H\Phi_{AVG}$) based on the free energies of transfer of amino acid side chains as previously calculated by Tanford (1962).

Table 4. The molecular weight, number of residues and the mean residue weight (MRW) of the various proteinases.

Proteinase	Molecular Weight	Number of Residues	Mean Residue Weight (MRW)
Chymosin	30700[a]	323[b]	96
Pepsin	35000[c]	327[b]	107
Mucor miehei proteinase	38000[d]	369[e]	103
Mucor pusillus proteinase	30600[f]	281[f]	109
Endothia parasitica proteinase	37500[g]	328[g]	114
Aspergillus saitoi proteinase	34500[h]	289[i]	119
Penicillopepsin	32000[j]	323[k]	99
Papain	20900[l]	212[m]	99
Trypsin	24000[n]	223[o]	108
α–chymotrypsin	21600[p]	214[m]	101

[a] Foltmann (1966).

[b] Foltmann (1981).

[c] Fruton (1970).

[d] Ottesen and Rickert (1970b).

[e] Bech and Foltmann (1981).

[f] Arima et al. (1970).

[g] Whitaker (1970).

[h] Hayashi et al. (1967).

[i] Ichishima and Yoshida (1966b).

[j] Sodek and Hofmann (1970a).

[k] Hsu et al. (1977).

[l] Smith and Kimmel (1960).

[m] Chang et al. (1978).

[n] Walsh and Neurath (1964).

[o] Cunningham et al. (1953).

[p] Sober (1970).

$$H\Phi_{AVG} = \frac{\Sigma H\Phi}{n}$$ Eq. 22

where $H\Phi$ = free energy of transfer

n   = number of amino acid residues.

The calculation was computerized and a listing of the program appears in Appendix 8.

F. HYDROPHOBICITY USING FLUORESCENT PROBES

Protein hydrophobicity was determined using the fluorescent probes cis-parinaric acid and the magnesium salt of 1-anilino-8-napthalene sulfonic acid.

1. cis-Parinaric acid

The method of Kato and Nakai (1980) was employed with slight modifications for the determination of hydrophobicity using cis-parinaric acid (CPA). Protein solutions ranging in concentrations from 0.002% to 0.01% (w/v) in 0.01 M phosphate buffer pH 6.3 were used for the assay. Ten µL of cis-parinaric acid (3.6×10^{-3} M in absolute ethanol containing 10 µg/mL butylated hydroxyanisole to prevent oxidation) were added to a 2 mL aliquot of each sample which was then rapidly vortexed using a Vortex mixer (Thermolyne Corp., Dubuque, IA). The relative fluorescence intensities of the solutions were measured with an Aminco Bowman spectrofluorometer No. 4-8202 (American Instrument Co., Inc., Silver Spring, MD) at an excitation wavelength of 325 nm and an emission wavelength of 420 nm. Prior to the sample readings the instrument was calibrated with a mixture of 2 mL n-decane and 10 µL cis-parinaric acid solution to

obtain a fluorescence intensity reading of 30 percent. "Blank" readings were also obtained for protein solutions that did not contain any fluorescent probe. The fluorescence of these samples was subtracted from the appropriate samples containing fluorescent probe to give the corrected fluorescence values.

2. 1-Anilino-8-napthalene sulfonate

(a) Preparation of ANS

Preparation of 1-anilino-8-napthalene sulfonate (ANS) was carried out according to the method of Weber and Young (1964) with slight modifications. The magnesium salt of 1-anilino-8-napthalene sulfonic acid was prepared from the sodium salt of ANS by precipitation with saturated magnesium acetate. The product was then recrystallized three times from water after filtration of the hot solutions through activated charcoal and Hyflo Super-Cel (Johns-Manville, Etobicoke, ON).

(b) Hydrophobicity determination

The determination of hydrophobicity using 1-anilino-8-napthalene sulfonate was carried out employing the method of Hayakawa and Nakai (1983). Protein concentrations of the same range and at the same pH as those used for CPA were utilized. Ten μL of 8.0×10^{-3} M ANS in 0.01 M phosphate buffer pH 6.3 was added to 2 mL aliquots of each sample and vortexed immediately. Relative fluorescent intensities were measured at an excitation wavelength of 390 nm and an emission wavelength of 470 nm. The fluorometer was calibrated with a mixture of 2 mL methanol

(HPLC grade) and 10 μL ANS solution to obtain a fluorescence intensity reading of 30 percent. Blank readings and corrected fluorescence values were obtained as for the cis-parinaric acid study.

For both fluorescent probes (CPA and ANS) protein hydrophobicity was measured as the initial slope of the curve of percent relative fluorescence versus percent protein. Duplicate determinations were made for each probe.

G. CHARGE RATIOS

Based on amino acid composition data various ratios were calculated for each proteinase. These included the number of charged groups (Asp, Glu, Lys, His, Arg) to total number of amino acid residues, the number of acidic groups (Asp, Glu) to total number, the number of basic groups (Lys, His, Arg) to total number, and the number of acidic groups to number of basic groups.

H. ZETA POTENTIAL

The zeta potential measurements of the various proteinases at different pH values were measured according to the method of Hayakawa and Nakai (1983). Five mL of a 0.05% (w/v) protein solution and 0.15 mL 3,3'-dimethylbiphenyl were emulsified with a Polytron mixer (Brinkmann Instruments, Rexdale, ON) at 2200 rpm for 20 s. A 0.5 mL aliquot of the emulsion was then pipetted into 40 mL of the appropriate buffer. The buffers used were similar to those prepared in Section D-1. The zeta potential reading of the solution was then obtained from a Laser Zee

Meter Model 501 (Pen Kem, Inc., Bedford Hills, NY) and corrected for temperature as suggested by the manufacturer. All measurements were done in duplicate.

I. ACCESSIBLE SURFACE AREA

The accessible surface area (ASA) of each of the proteinases was calculated based on their molecular weights, according to the algorithm (Eq. 23) of Janin (1976):

$$ASA = 11.1 \ M^{2/3} \qquad \text{Eq. 23}$$

where   M = molecular weight of the proteinase.

J. DETERMINATION OF MILK-CLOTTING ACTIVITY

The milk-clotting activity of the various proteinases was determined according to the method of Berridge (1952) as modified by Iwasaki et al. (1967a). Spray-dried skim-milk powder was used as the substrate and was stored desiccated at 4°C. Twelve percent solutions (w/v) of skim-milk powder in 0.01 M $CaCl_2$ were adjusted to various pH values with 0.1 N HCl or 0.1 N NaOH (Iwasaki et al., 1967a) and allowed to equilibrate for 1 h at room temperature. A ten mL aliquot of reconstituted skim-milk was dispensed into a 18 x 150 mm test tube and incubated in a shaking water bath at 30°C for 10 min. The proteinase was diluted with the appropriate buffer to obtain a milk-clotting activity of 4 to 5 min. A 1.0 mL aliquot of the diluted proteinase solution, pre-incubated at 30°C, was pipetted into the skim-milk

solution and mixed by inversion while simultaneously beginning to measure clotting time. The inversion was repeated an additional two times to ensure complete mixing and the test tube was then returned to the shaking water bath. About 30 s before the expected end-point (clot formation) a glass rod was placed into the milk solution and then raised about 4 cm above the level of the milk, keeping the lower end of the rod in contact with the side of the test tube to allow a stream of milk to run from the rod. Clotting time was taken as the moment when the milk film broke into visible particles. Berridge (1945) defined one unit of milk-clotting activity as the amount of proteinase which clots 10 mL of reconstituted skim-milk in 100 s at 30°C. The specific activity was expressed as milk-clotting activity per mg protein.

In order to activate papain for milk-clotting measurement, the proteinase solution was prepared in the presence of 0.05 M cysteine and 0.02 M EDTA (Arnon, 1970).

K. DETERMINATION OF PROTEOLYTIC ACTIVITY

The ability of the proteinases to hydrolyze sodium caseinate was determined using a method described by Green (1972) with slight modifications. The substrate solution contained 1% (w/v) sodium caseinate in either 0.1 M acetate or 0.1 M phosphate buffer depending on the pH. Acetate buffers were used for pH 5.0, 5.3 and 5.8, while phosphate buffers were used for pH 6.3, 7.0 and 8.0. To 2.5 mL of substrate, pre-incubated at 35°C for 10 min, 0.5 mL of enzyme was added and the mixture incubated at 35°C for 10 min; 2.5 mL of 6.6% (w/v) trichloro-

acetic acid (TCA) was then added to give a final TCA concentration of 3% (Ma, 1979). The resulting precipitate was removed by filtration through Whatman No. 2 paper. The filtrate was then analyzed for free amino groups according to the method of Kwan et al. (1983) with slight modifications. To 0.1 mL of filtrate, 0.3 mL of 1 M K_2HPO_4 was first added, followed by 0.15 mL of 0.03% (w/v) fluorescamine in acetone which was added rapidly and directly to the solution and mixed immediately using a Vortex mixer. A volume of 3.0 mL of distilled-deionized water was added to the reaction mixture and the mixture vortexed once again. The fluorescence of the final mixture was measured with an Aminco Bowman 4-8202 spectrofluorometer using an excitation wavelength of 395 nm and an emission wavelength of 480 nm. Blanks were prepared following the same procedure except that TCA was added to the substrate prior to the addition of the enzyme solution. A standard curve was constructed using tyrosine. Proteolytic activity was defined as the amount of tyrosine released per mg of enzyme. Papain was activated as previously described for milk-clotting.

L. PRINCIPAL COMPONENT ANALYSIS

Data collected for the various proteinases was subjected to principal component analysis using the BMDP:4M program (Frane et al., 1981) and run on an Amdahl 470 V/8 computer. Sixteen variables were included: molar ellipticity values at 10 wavelengths from the CD spectra (190, 193, 198, 200, 202, 210, 213, 222, 224 and 225 nm), zeta potential, acidic/basic amino acid ratio, Bigelow average hydrophobicity, accessible surface area, CPA hydrophobicity and ANS hydrophobicity. The ten

wavelengths were selected on the basis of characteristic wavelengths representative of each of the secondary structure fractions (Chang et al., 1978). The α-helix has a characteristic double minimum at 222 and 210 nm and a maximum at 193 nm. The β-sheet has a minimum at 213 nm and a maximum at 198 nm while the β-turn has two maxima at 224 and 202 nm with a strong negative band below 190 nm. The unordered or random coil structure shows a strong negative band near 200 nm and another negative band around 225 nm.

The milk-clotting to proteolytic activity ratio was then regressed on the principal components derived from the sixteen variables using the BMDP: 4R program (regression using principal components).

RESULTS AND DISCUSSION

A. DIAGONAL PLOTS

In the diagonal plot method the primary sequences of the two proteins in question are aligned along the two normal axes with the N-terminal amino acids starting at the origin. The method is simple and highly visual and allows for a rapid comparison of two primary amino acids sequences (Gibbs and McIntyre, 1970). A complete identity in sequence would result in a diagonal line dissecting the plot from corner to corner (Beynon, 1982). The diagonal plots for comparison of primary amino acid sequences between pepsin and chymosin are presented in Figures 4a – 4b, between chymosin and Mucor miehei proteinase in Figures 5a – 5c, between pepsin and M. miehei proteinase in Figures 6a – 6c, between pepsin and penicillopepsin in Figurs 7a – 7b and between chymosin and penicillopepsin in Figures 8a – 8b.

Examination of the plots indicated that there were different degrees of homology between the pairs of aspartyl proteinases compared, as demonstrated by the various degrees of diagonal completeness. Overall, however, the homologies were relatively high. From the plot between pepsin and chymosin (Figures 4a – 4b) the compilation of the number of asterisks along the diagonal in comparison to the total number of residues indicated that approximately 60 percent of the total amino acid residues were identical in the two sequences (Table 5). Similar comparisons between the other proteinases revealed substantially lower homology values with approximately 24 to 27 percent of the total
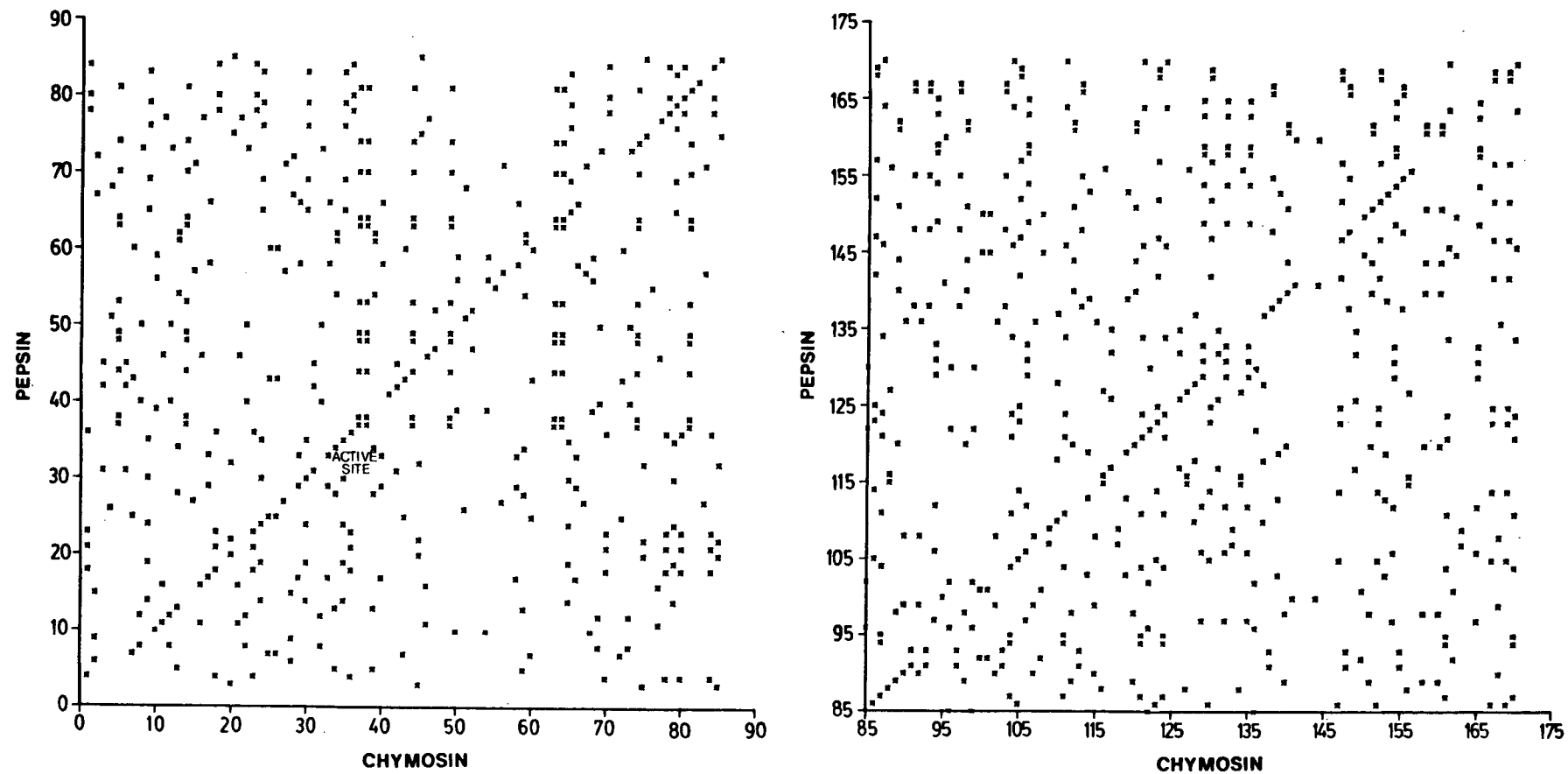
Figure 4a. Diagonal plot of the primary amino acid sequences of pepsin and chymosin for residues 1 to 85 and 85 to 170.
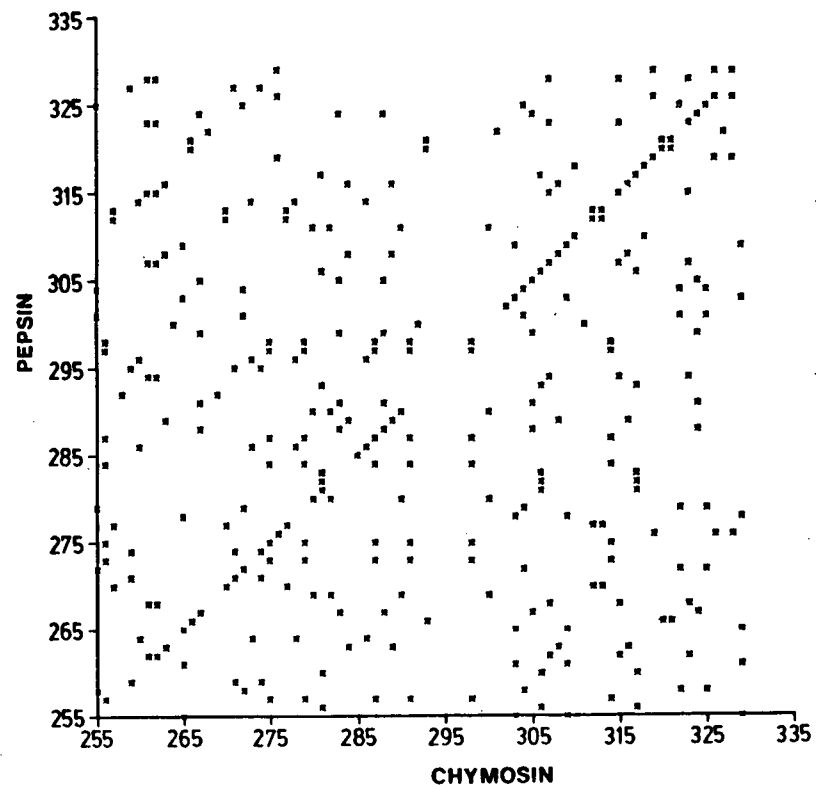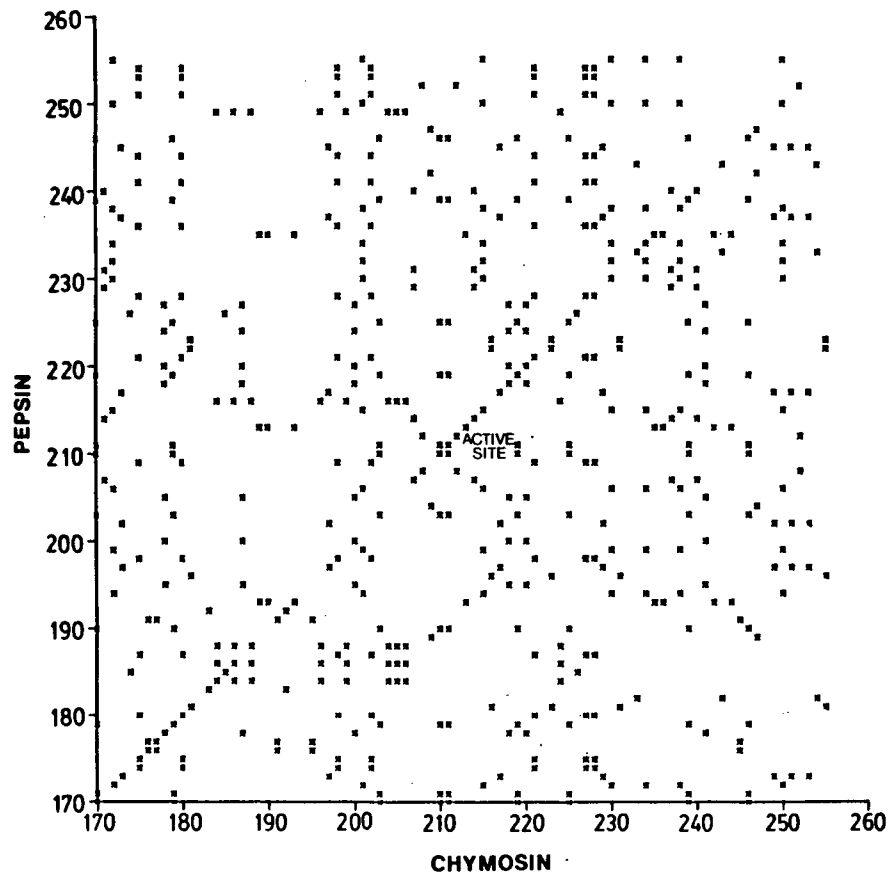
Figure 4b. Diagonal plot of the primary amino acid sequences of pepsin and chymosin for residues 170 to 255 and 255 to 327.
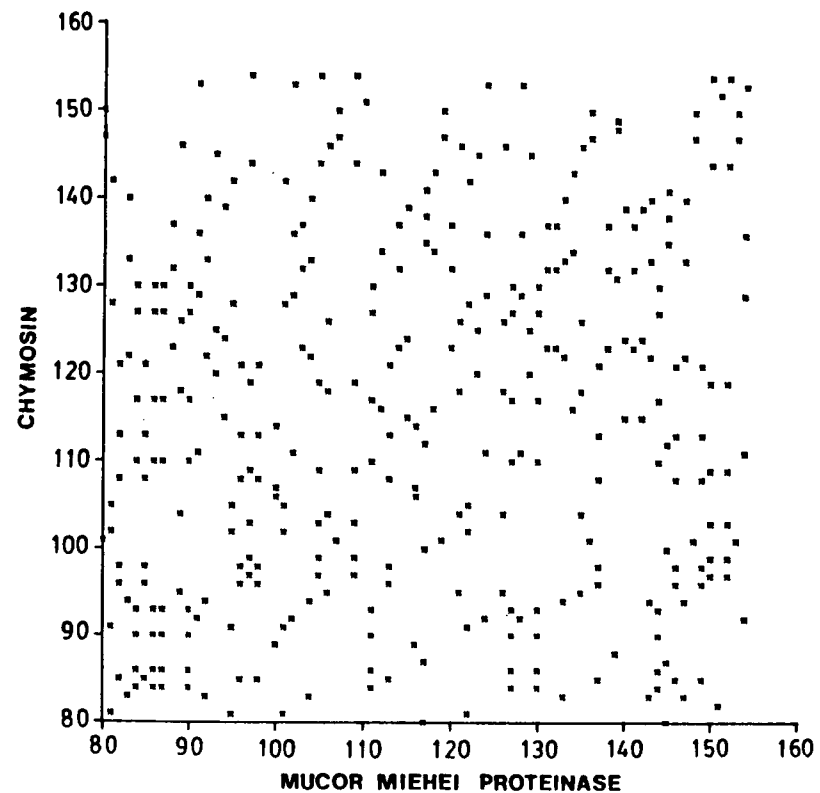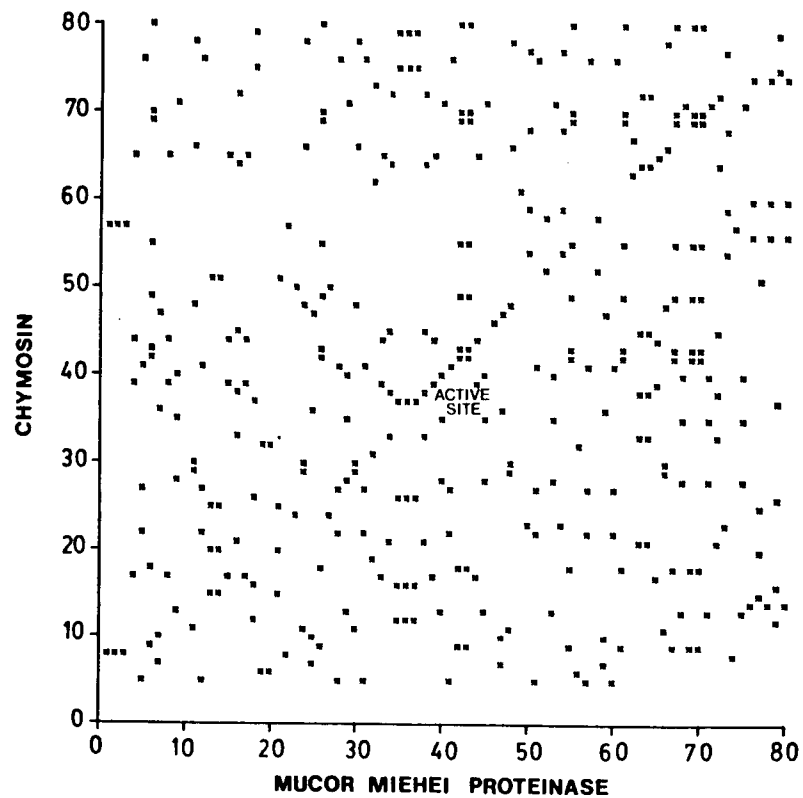
Figure 5a. Diagonal plot of the primary amino acid sequences of chymosin and <u>Mucor</u> <u>miehei</u> proteinase for residues 1 to 80 and 80 to 155.
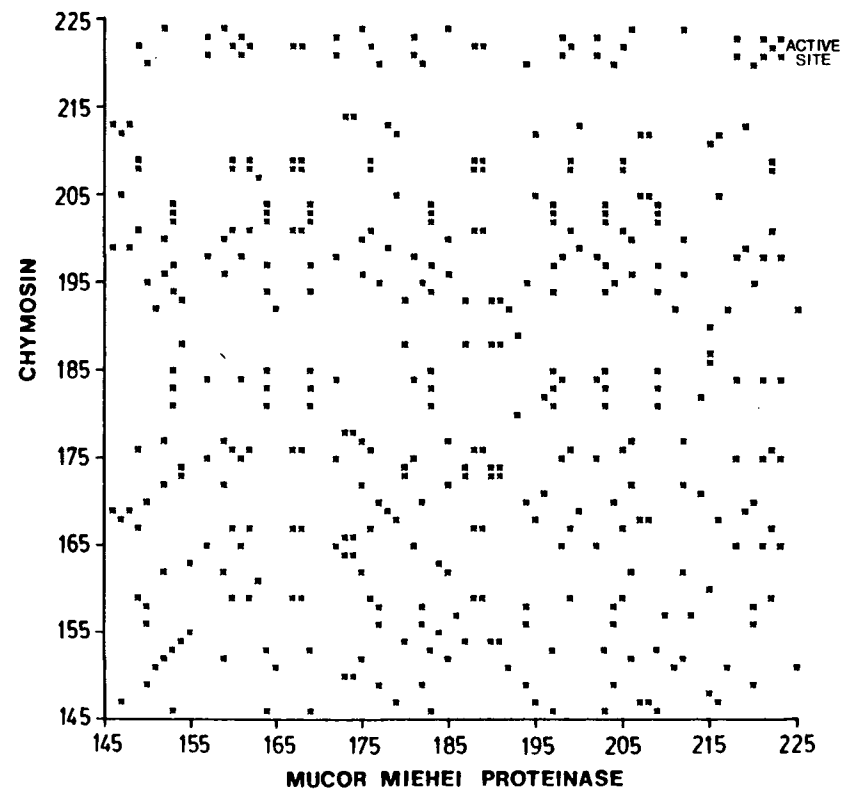
Figure 5b. Diagonal plot of the primary amino acid sequences of chymosin and <u>Mucor</u> <u>miehei</u> proteinase for residues 145 to 225.
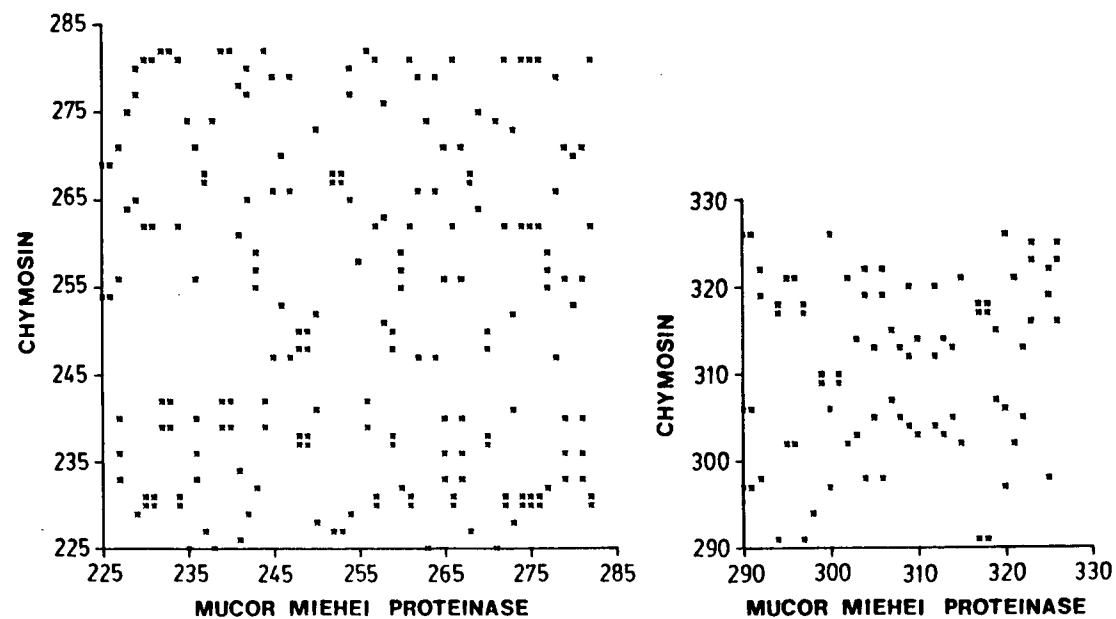
Figure 5c. Diagonal plot of the primary amino acid sequences of chymosin and _Mucor_ _miehei_ proteinase for residues 225 to 285 and for residues 290 to 330.
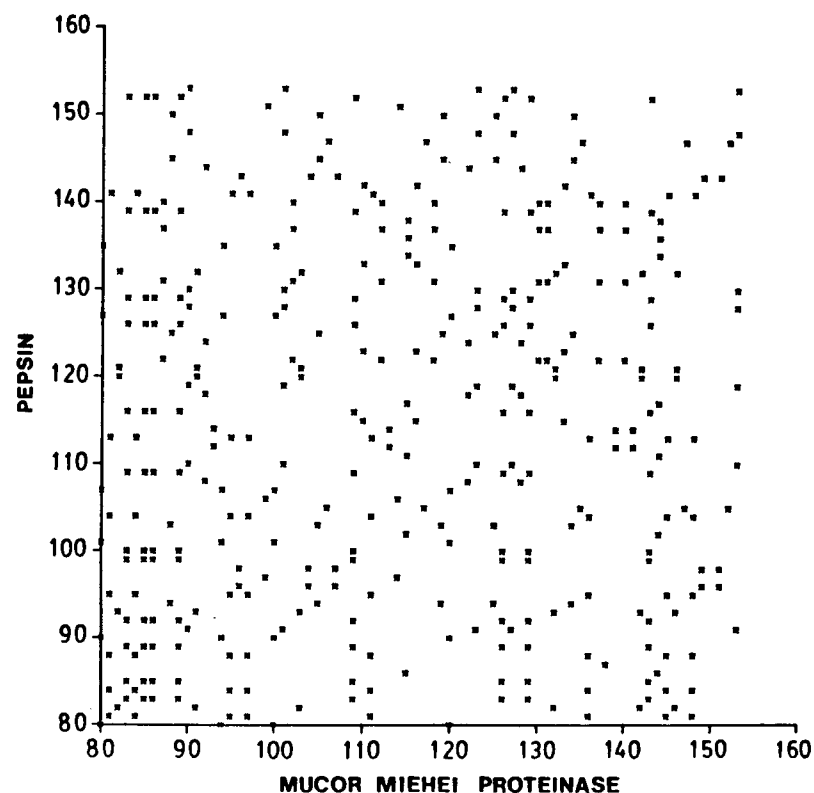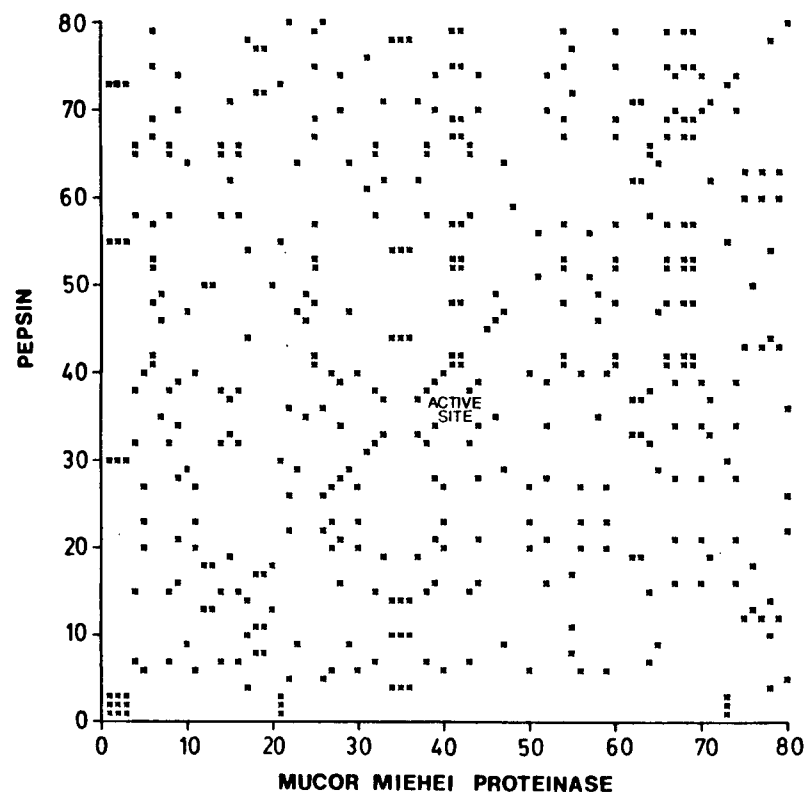
Figure 6a. Diagonal plot of the primary amino acid sequences of pepsin and <u>Mucor</u> <u>miehei</u> proteinase for residues 1 to 80 and 80 to 155.
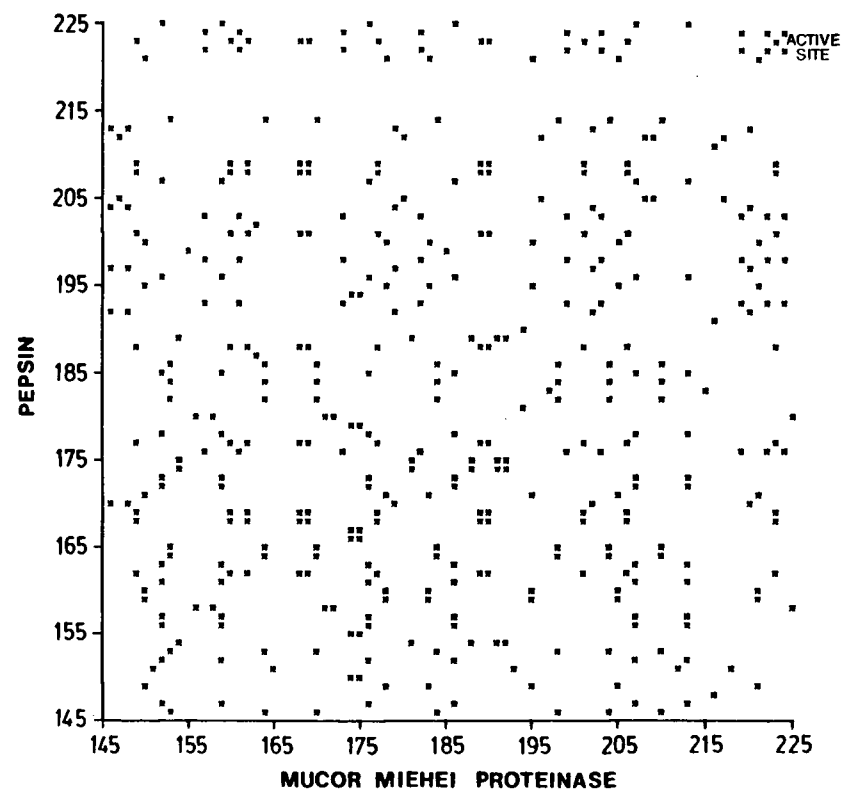
Figure 6b. Diagonal plot of the primary amino acid sequences of pepsin and <u>Mucor</u> <u>miehei</u> proteinase for residues 145 to 225.
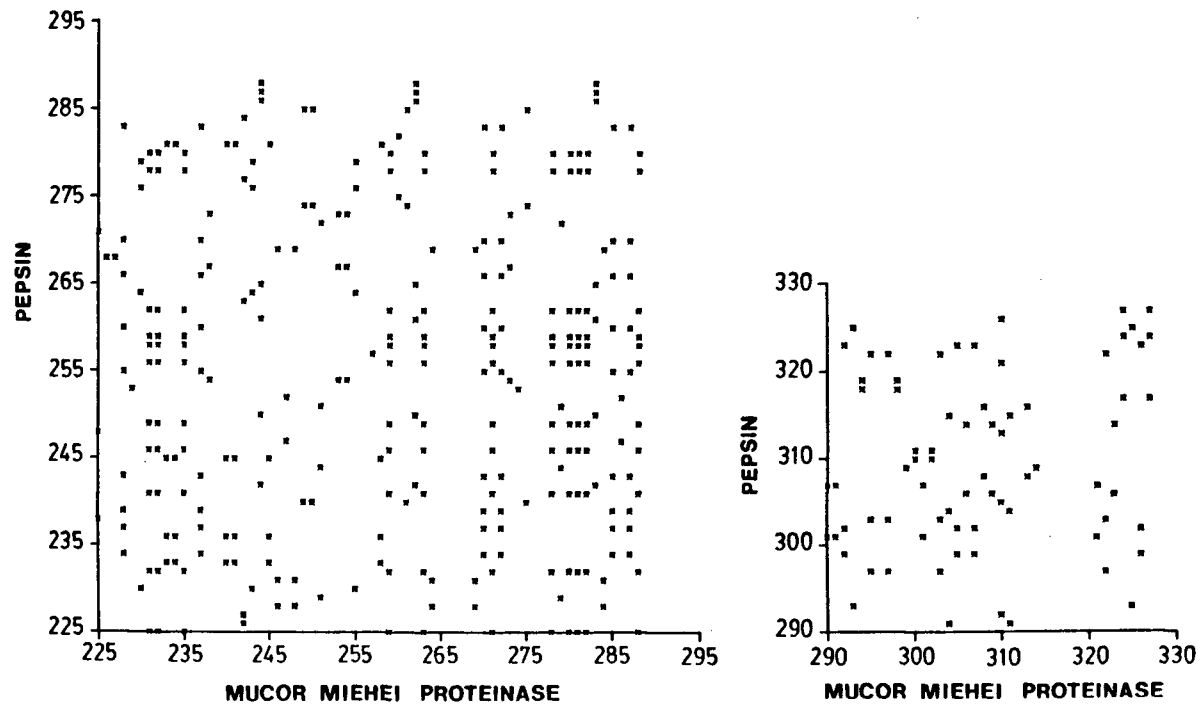
Figure 6c. Diagonal plot of the primary amino acid sequences of pepsin and Mucor miehei proteinase for residues 225 to 290 and 290 to 330.
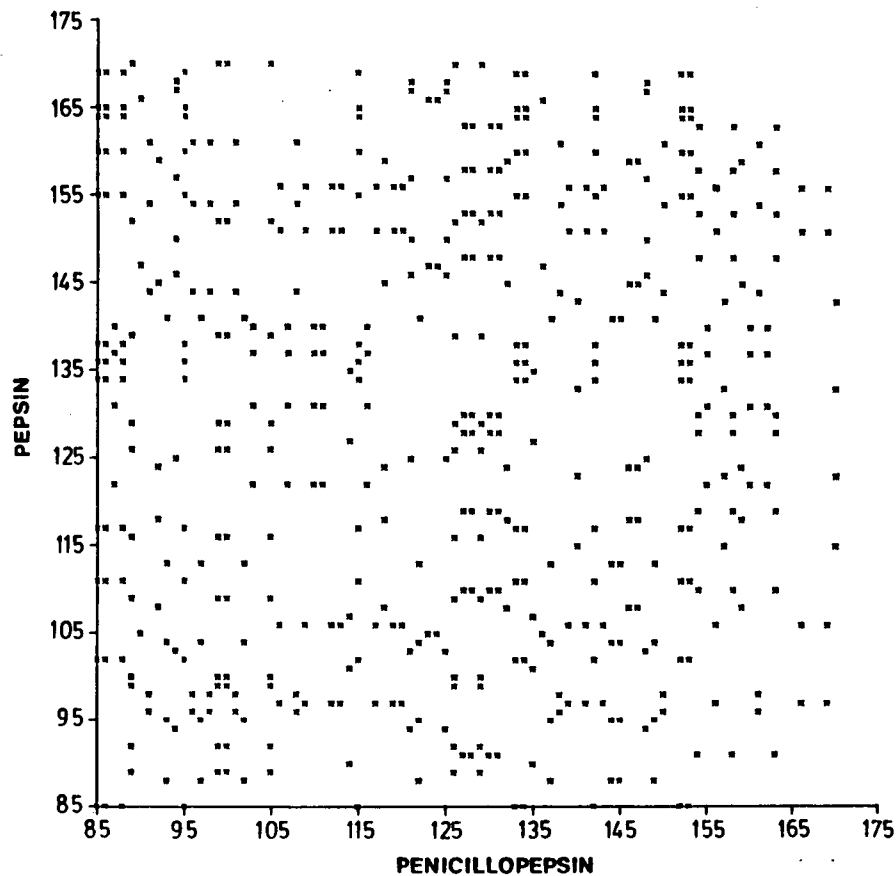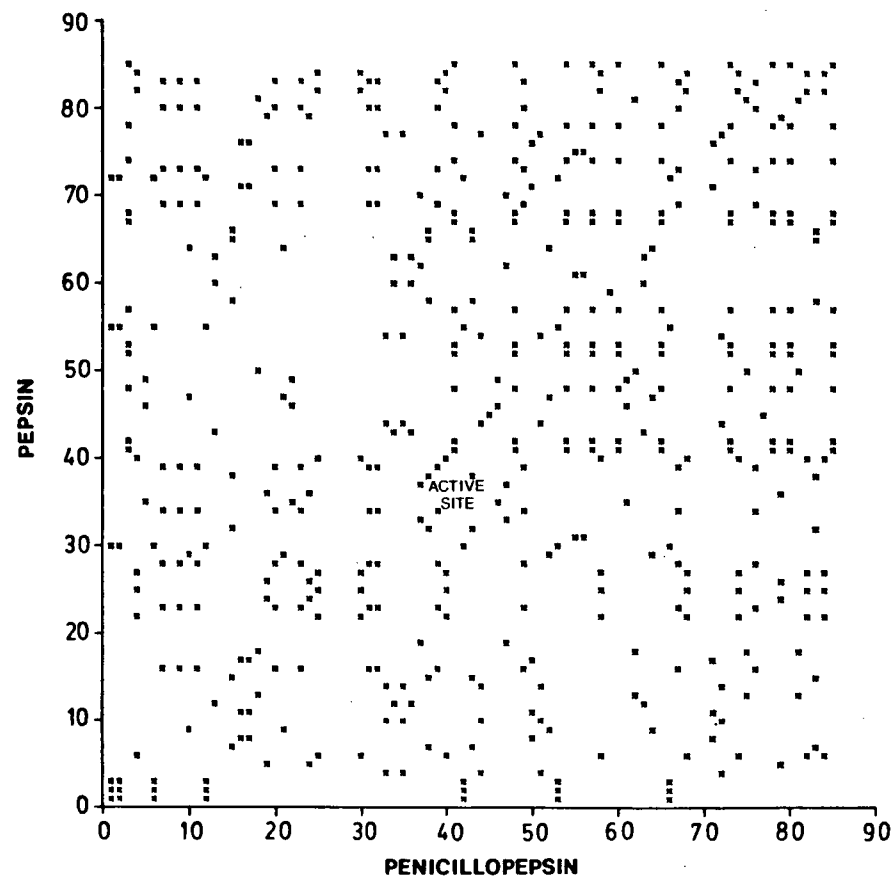
Figure 7a. Diagonal plot of the primary amino acid sequences of pepsin and penicillopepsin for residues 1 to 85 and 85 to 170.
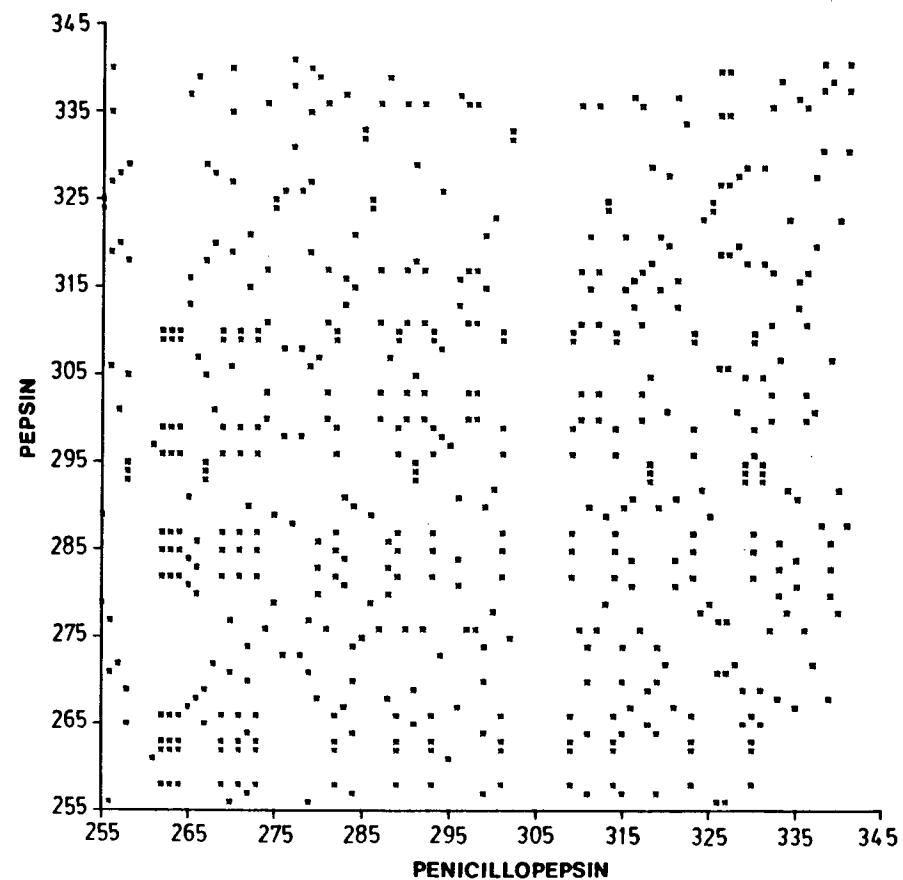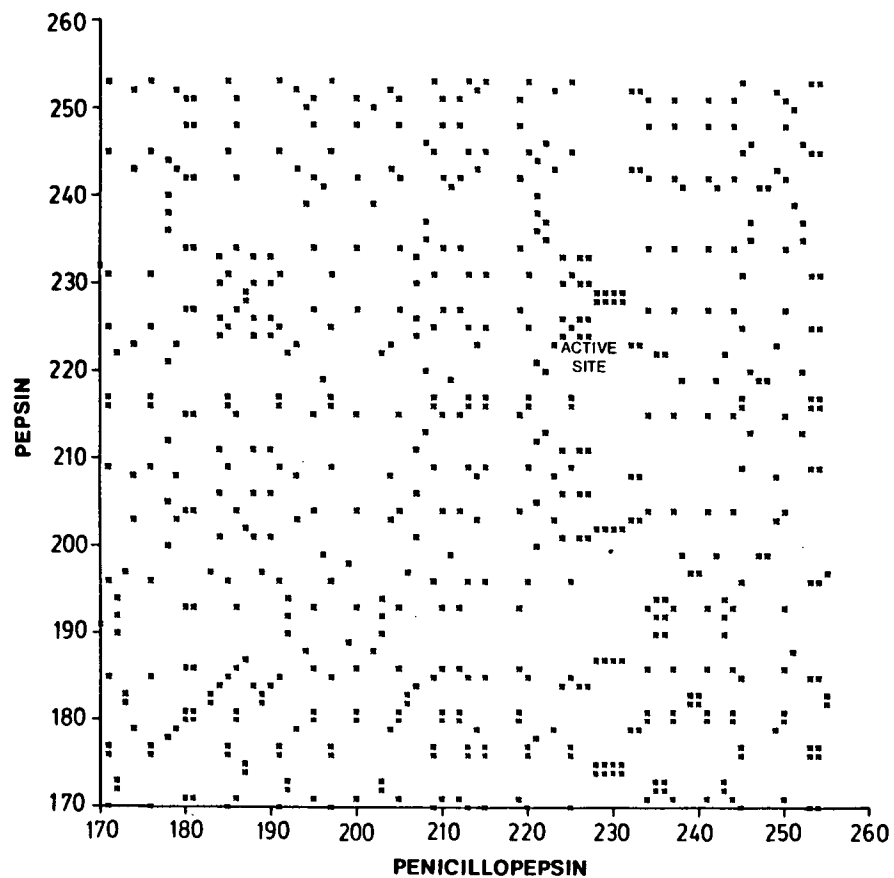
Figure 7b. Diagonal plot of the primary amino acid sequences of pepsin and penicillopepsin for residues 170 to 255 and 225 to 340.
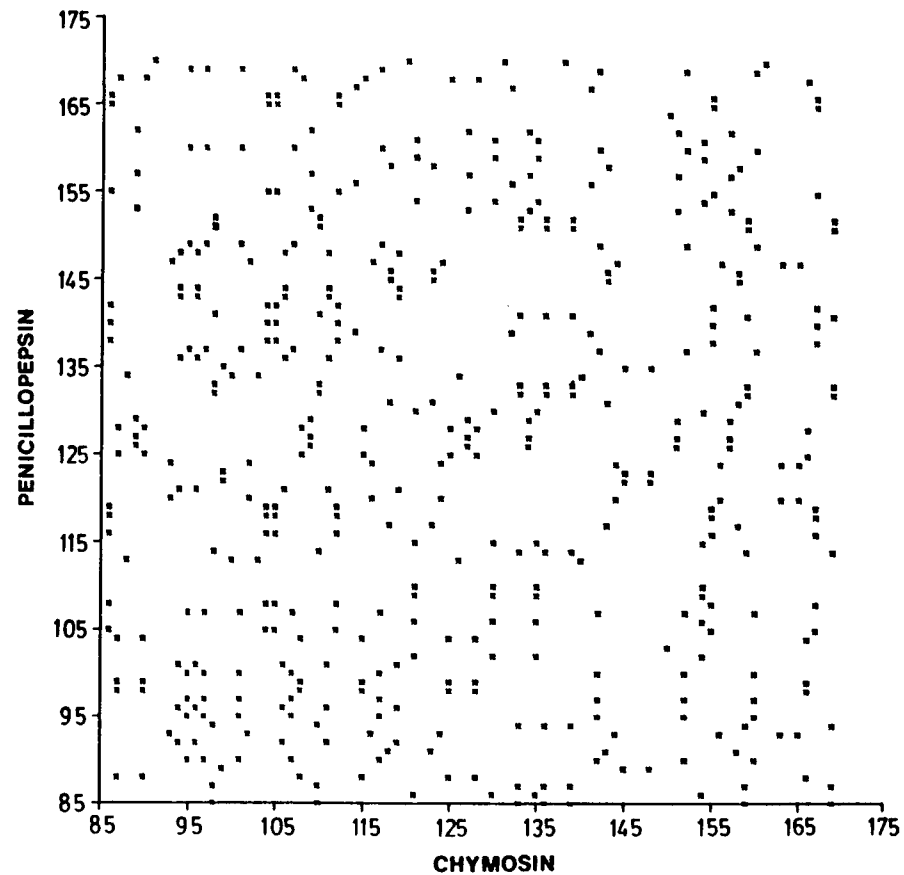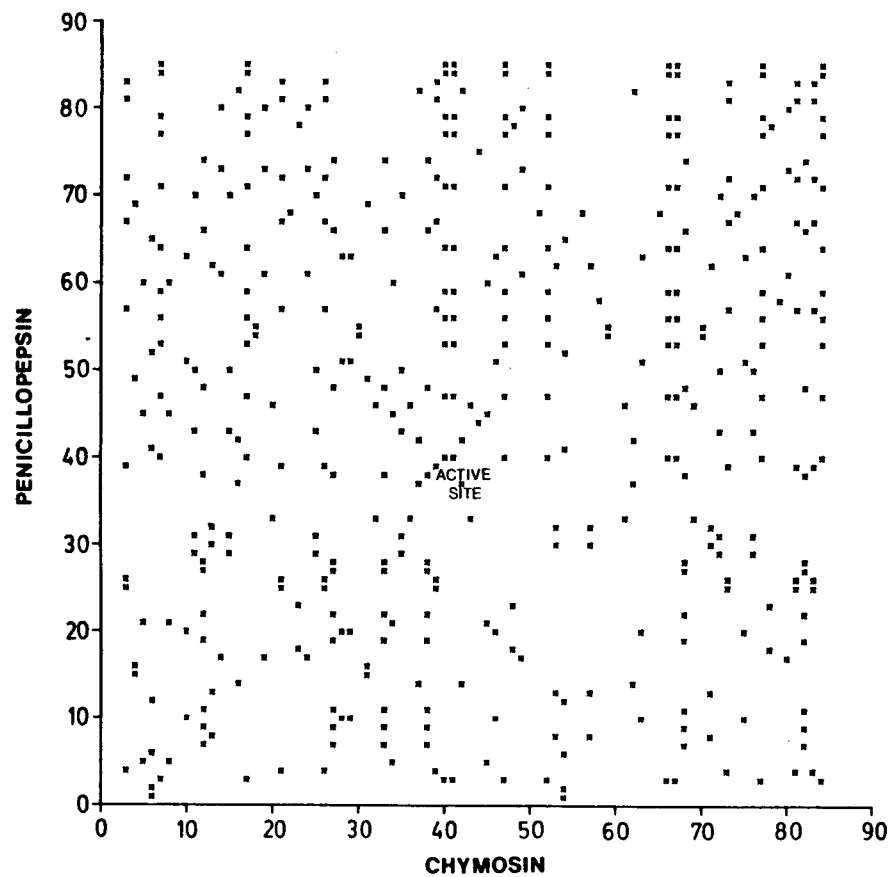
Figure 8a. Diagonal plot of the primary amino acid sequences of penicillopepsin and chymosin for residues 1 to 85 and 85 to 170.
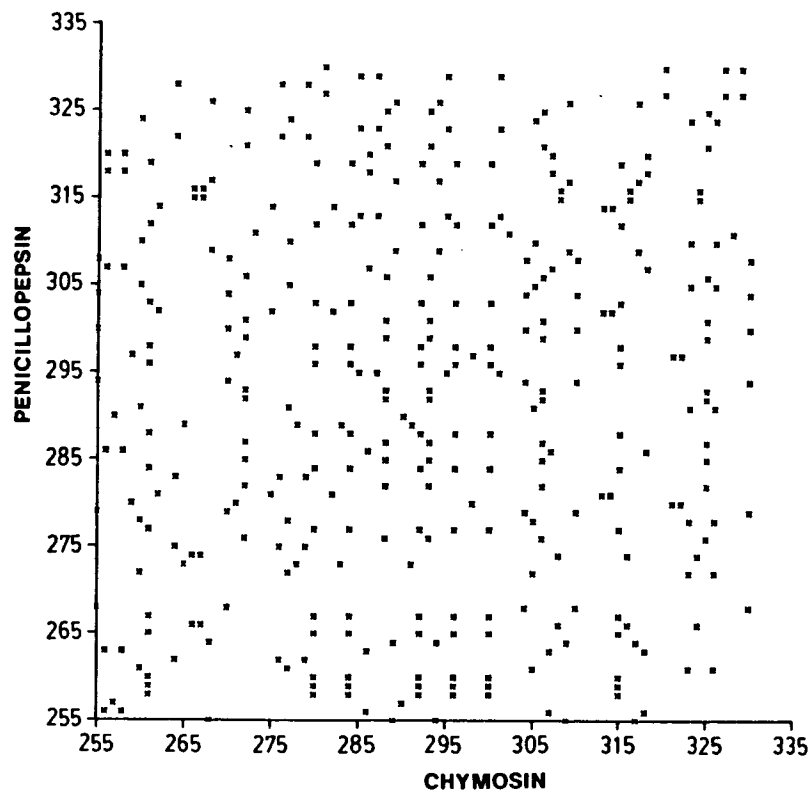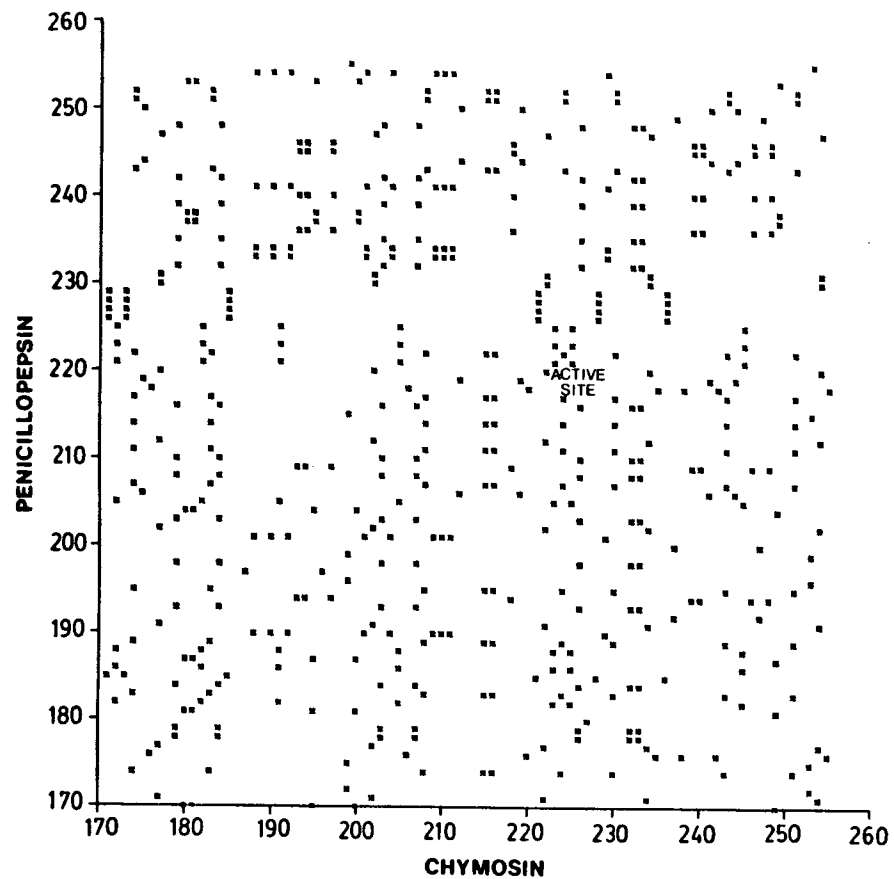
Figure 8b. Diagonal plot of the primary amino acid sequences of penicillopepsin and chymosin for residues 170 to 255 and 255 to 330.

Table 5. Degree of similarity between various pairs of aspartyl pro-
teinases based on the diagonal plot method.

Proteinase pair	Similarity[1] (%)
pepsin – chymosin	57.3
chymosin – Mucor miehei proteinase[2]	24.7
pepsin – Mucor miehei proteinase	25.7
pepsin – penicillopepsin	27.1
chymosin – penicillopepsin	25.3

[1]Similarity (%) was calculated by dividing the number of asterisks that
lay on the 45° diagonal by the total number of residues compared.

[2]Comparisons of M. miehei proteinase with chymosin and pepsin were based
on the partial sequence of M. miehei proteinase.

residues being identical (Table 5). The fact that pepsin and chymosin showed a higher degree of homology than the comparison of pepsin or chymosin with Mucor miehei proteinase or penicillopepsin may be because pepsin and chymosin are gastric enzymes while Mucor miehei proteinase and penicillopepsin are microbial enzymes. Proteinases from similar sources may have higher homologies. In comparison to chymosin-pepsin, the degree of similarity between the other aspartyl proteinases was relatively low, despite the observation that as a class, the aspartyl proteinases have been shown to have very similar three-dimensional structures. Schulz and Schirmer (1978) stated that in the evolution of homologous proteins, important positions such as active sites are almost invariant, however, very similar three-dimensional structures can result from quite different amino acid sequences.

It was interesting to note that areas adjacent to the active site aspartic acid residues (i.e. Asp_{32} and Asp_{215} using the pepsin numbering system) were highly homologous in all proteinase pairs examined. Foltmann and Pedersen (1976) in a comparison of the primary structures of acidic proteinases and of their zymogens, reported that the two catalytically active aspartic acid residues were located in highly identical surroundings.

B. SECONDARY STRUCTURE PREDICTION

It is now generally accepted that the three-dimensional structure of a protein dictates its action (Barry et al., 1974). From renaturation experiments (Anfinsen et al., 1961) it has been hyothesized that

the three-dimensional structure of a protein is a unique function of its amino acid sequence. An important first step in the prediction of the overall three-dimensional structure of a protein may be the prediction of secondary structure (Ptitsyn and Finkelstein, 1983). In this light, two methods were used to determine secondary structure from the primary amino acid sequences, namely the Chou and Fasman (1978b) method as computerized by Pham (1981) and the hydrophobicity profile method of Cid et al. (1982). Secondary structure predictions were limited to those proteinases for which the complete or near complete sequences had been determined. The proteinases examined were chymosin, pepsin, penicillo-pepsin and Mucor miehei proteinase (partial sequence).

The Chou and Fasman (1978b) method is a simple predictive method in which the secondary structure forming potential of the twenty amino acids are used to calculate the α-helix, β-sheet and β-turn potential for any protein segment in the primary amino acid sequence. Empirical rules set forth by the authors aid in locating the specific secondary structures. Figures 9, 10 and 11 are the predicted secondary structures for chymosin, pepsin and penicillopepsin, respectively. The predicted structure of chymosin (Figure 9) revealed a relatively high proportion of β-sheet fractions, with only minor sections of the total secondary structure predicted as α-helix. The active site aspartic acid residues (Asp_{34} and Asp_{216}) were predicted to be in β-turn regions. The predicted secondary structure of pepsin (Figure 10), like chymosin, was dominated by β-sheet fractions. Only one section of α-helix (residues 65 to 70) was predicted. The active site aspartic acid residues (Asp_{32}
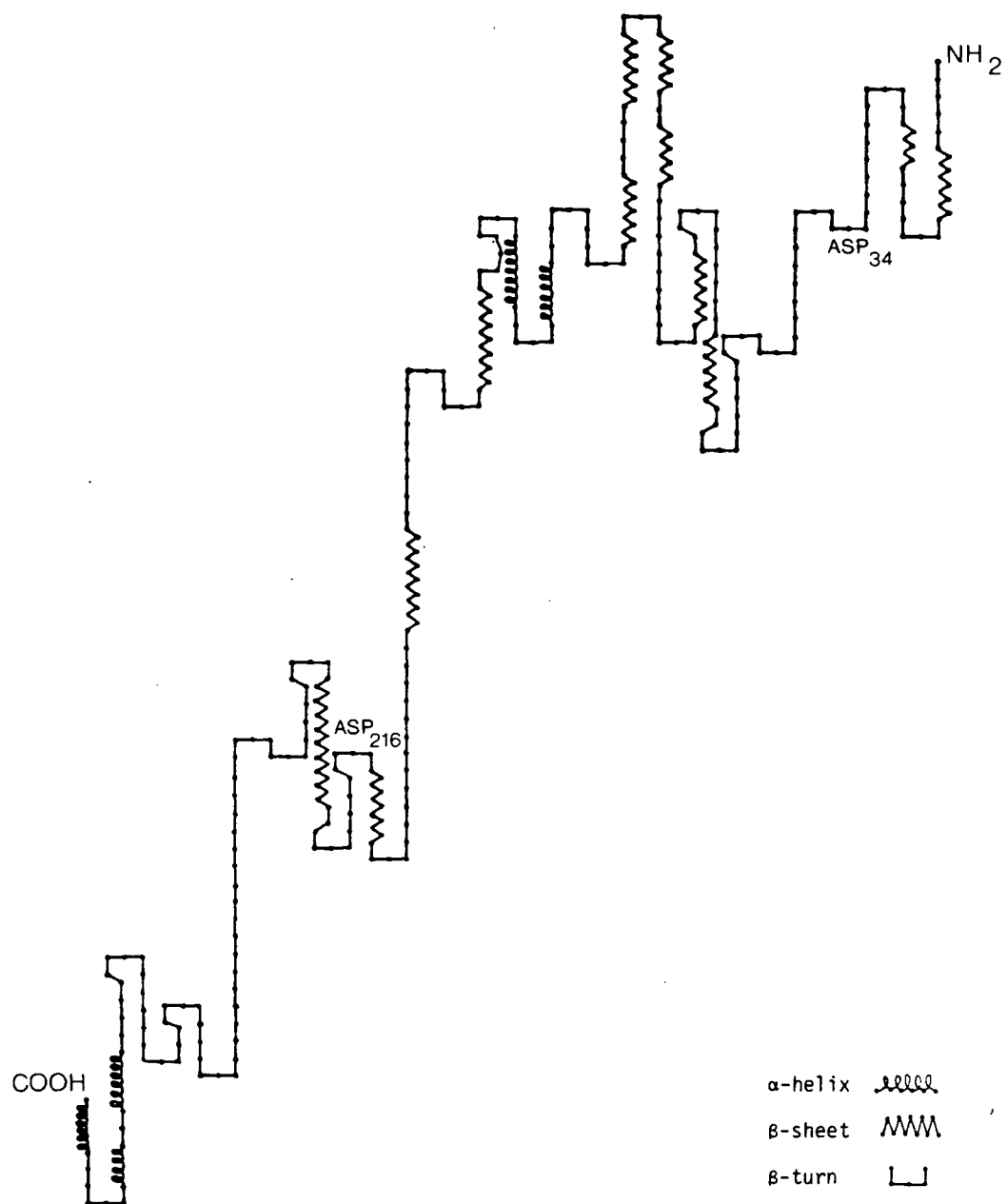
Figure 9. Predicted secondary structure of chymosin using the Chou and Fasman method.
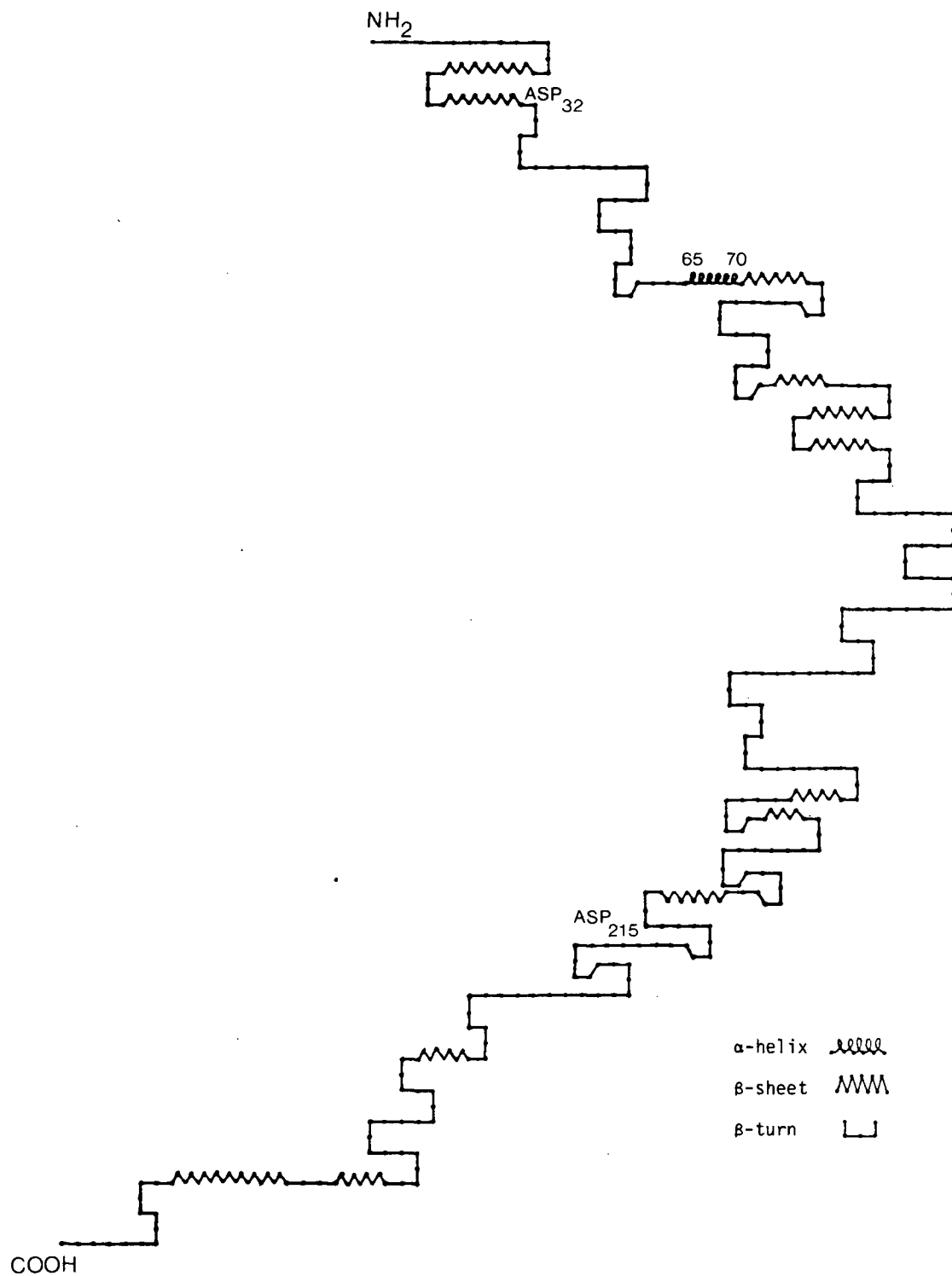
Figure 10. Predicted secondary structure of pepsin using the Chou and Fasman method.
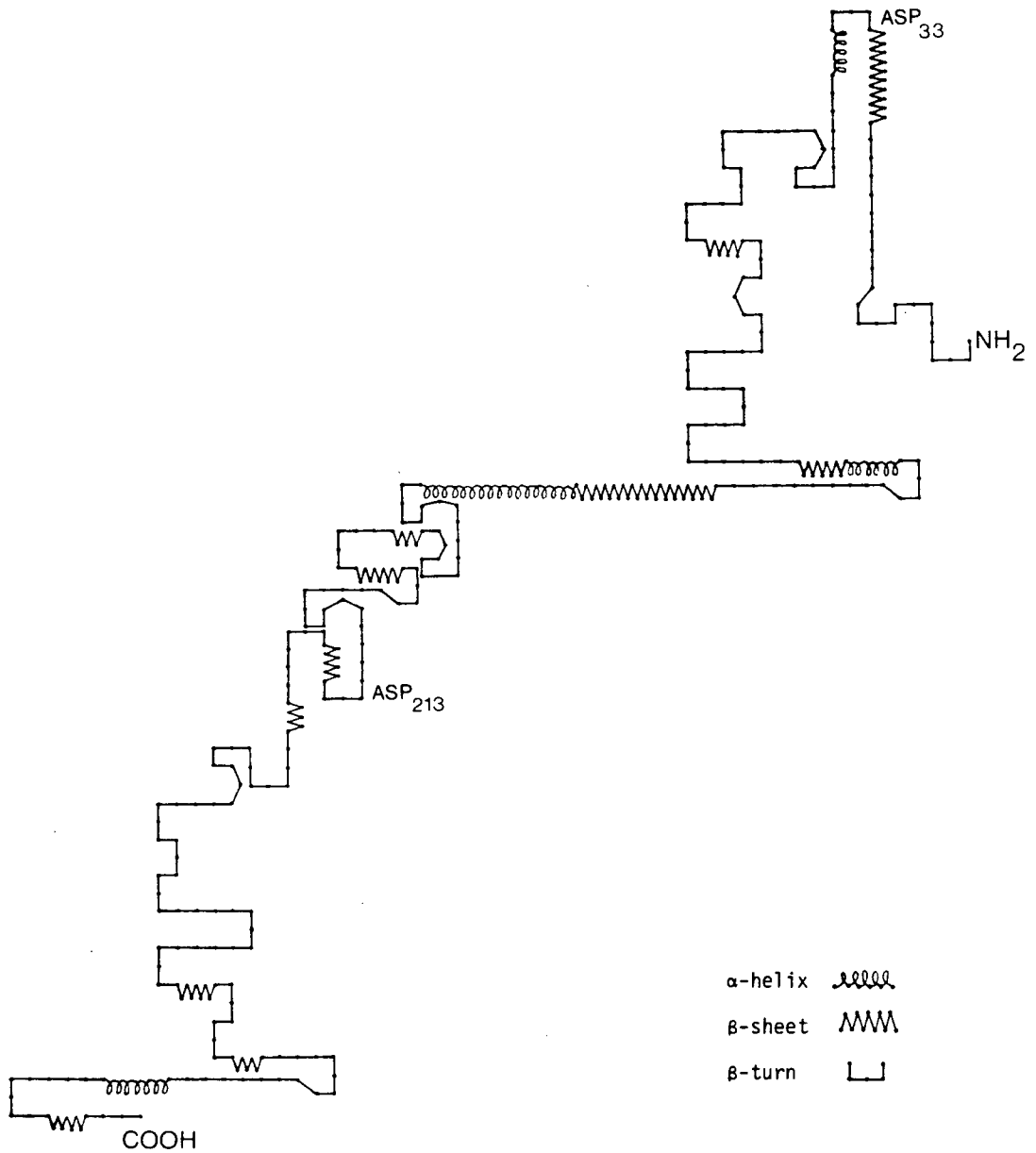
Figure 11. Predicted secondary structure of penicillopepsin using the Chou and Fasman method.

and Asp$_{215}$) were also predicted to be located in β-turn regions. The β-sheet structure was predicted to be the main structural feature of the secondary structure of penicillopepsin (Figure 11). Alpha-helix regions were a minor portion of the total predicted secondary structure although the amount of this secondary structure was slightly higher in penicillopepsin than in the other proteinases examined. Aspartic acid$_{33}$ and Asp$_{213}$, the active site residues, were located in β-turn regions. Hsu et al. (1977) reported that the two active site aspartic acid residues of penicillopepsin were located at the ends of β-strands near β-turn regions.

As was apparent from the above observations, several secondary structure features were common among the three aspartyl proteinases examined. The high proportion of β-sheet observed is consistent with the results obtained from X-ray diffraction analysis for several aspartyl proteinases. Andreeva et al. (1976) examined the X-ray data obtained from pepsin and found that the major structural components of the molecule were β-sheet structures. The crystal structures for the microbial proteinases from Endothia parasitica (Jenkins et al., 1976) as well as from Rhizopus chinensis (Subramanian et al., 1976) have also been elucidated. In both these proteinases the β-sheet structure was found to be a major secondary structure fraction. Recently, James and Sielecki (1983) examined the crystal structure of penicillopepsin at 1.8Å resolution and found that approximately two-thirds of the total amino acid residues of the microbial proteinase have φ, Ψ conformational angles that describe those of parallel or anti-parallel β-sheets.

Although the Chou and Fasman (1978b) method predicted a relatively high proportion of β-sheet in penicillopepsin (approximately 0.2), the amount was substantially lower than that observed by James and Sielecki (1983). Kabsch and Sander (1983) examined various methods used in the prediction of protein secondary structure from the amino acid sequence and found that the Chou and Fasman (1978b) method had only a 50 percent overall prediction accuracy. The difference between results obtained using this prediction method (Chou and Fasman) and those obtained crystallographically have been attributed to: (1) ambiguities of some of the rules set forth by Chou and Fasman, and (2) the variability in the definitions of secondary structure used by crystallographers.

Although the method of Chou and Fasman (1978b) may yield erroneous results one could argue that a single program was used, thus differences in the interpretation would be consistent for all proteins examined; therefore, this method may allow us a means for the relative comparison of the aspartyl proteinases. Furthermore, the prediction of the active site aspartic acid residues for the three proteinases occurring in β-turn regions is reasonable, considering the role of these residues. In order to be involved in the hydrolysis of a peptide substrate (i.e. catalysis), it would be imperative that these residues be on the surface of the enzyme. Kuntz (1972) observed that a substantial number of β-turn fractions were located on the surface of globular proteins.

The type and sequence of amino acids are important for determining the two- and three-dimensional structure of a protein; other important factors are the forces that affect the folding of the polypeptide.

Kauzmann (1959) was one of the first researchers to recognize the importance of hydrophobic forces in affecting protein structure. Cid et al. (1982) proposed a simple method for the prediction of secondary structure by means of a hydrophobicity profile. Basically in this method the bulk hydrophobicity for each amino acid residue is plotted against residue number and the secondary structure fraction is determined based on the profile exhibited.

The hydrophobicity profile method of Cid et al. (1982) was used to predict the secondary structures for chymosin, pepsin, penicillopepsin and Mucor miehei proteinase and the results are presented in Figures 12a-12b, 13a-13b, 14a-14b and 15a-15b. Using this method it was found that characterization of secondary structures other than exposed β-strands or β-turns was difficult, which greatly restricted the capabilities of this method. It should be noted that for Mucor miehei proteinase the region of sequence 160 to 180 should be interpreted with some caution since, although the amino acid composition is known, the exact sequence has not been established. As with the Chou and Fasman method which had predicted β-sheet as a major secondary structure fraction, the hydrophobicity profile method also predicted β-sheet as a major fraction in the four proteinases. The β-sheet fraction is represented by the "zig-zag" pattern (Figures 12 to 15) resulting from alternating hydrophilic and hydrophobic residues. Kanehisa and Tsong (1980) had previously described β-strands on the surface of proteins as alternating hydrophilic and hydrophobic residues.
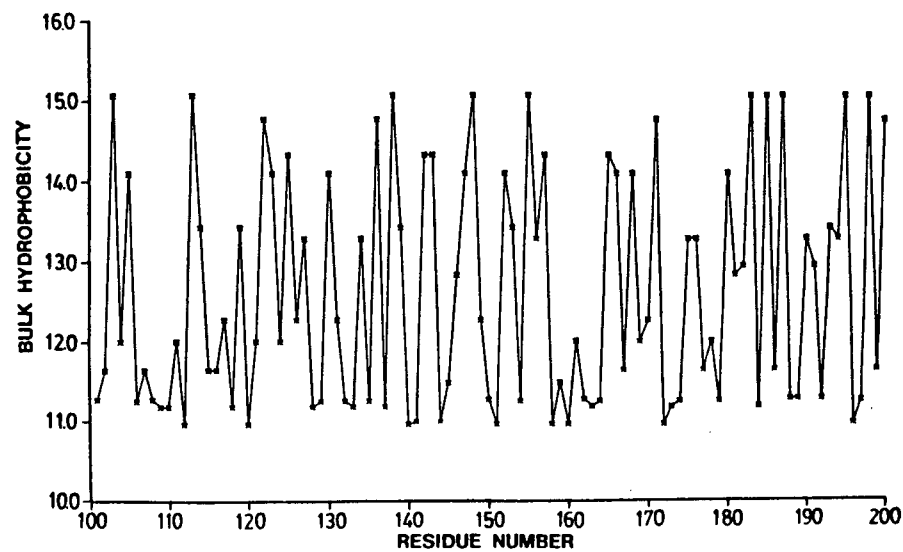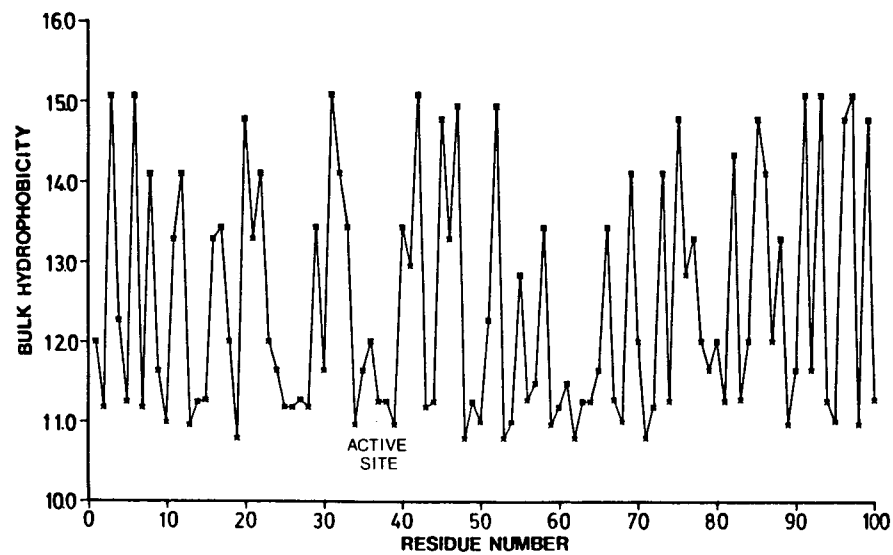
Figure 12a. The bulk hydrophobicity profile of chymosin for residues 1 to 100 and 100 to 200.
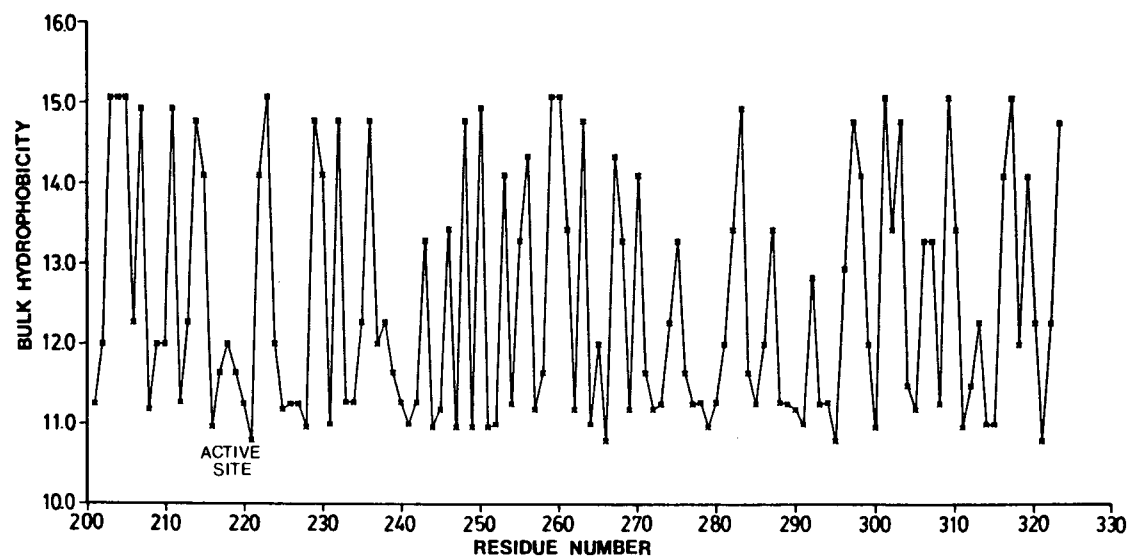
Figure 12b. The bulk hydrophobicity profile of chymosin for residues 200 to 323.
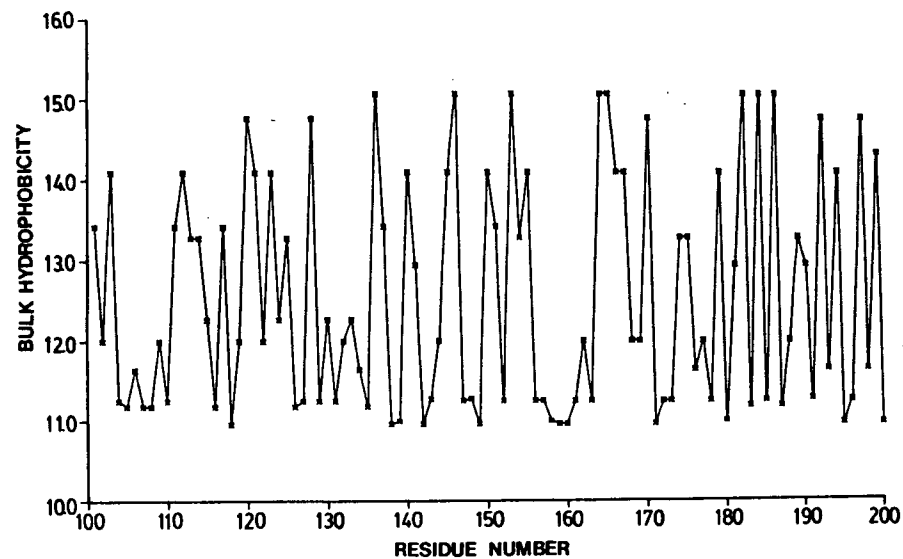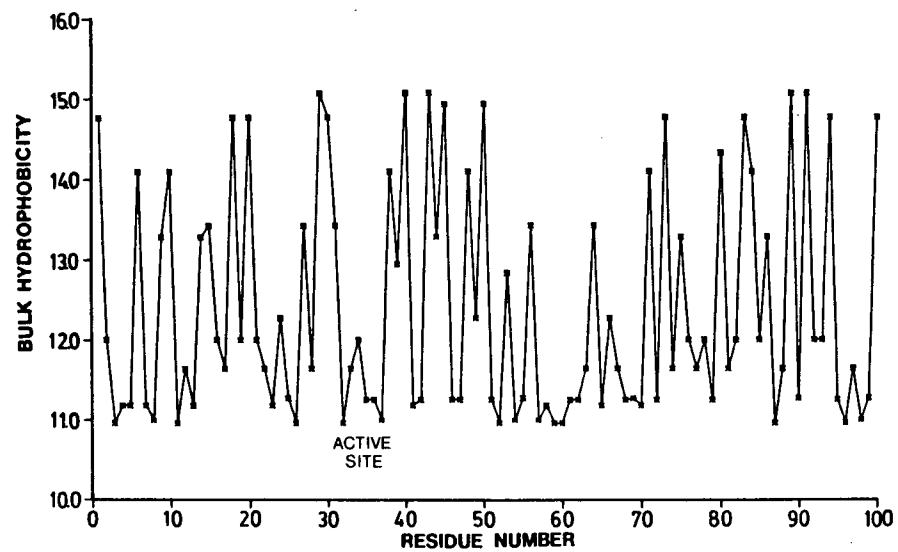
Figure 13a. The bulk hydrophobicity profile of pepsin for residues 1 to 100 and 100 to 200.
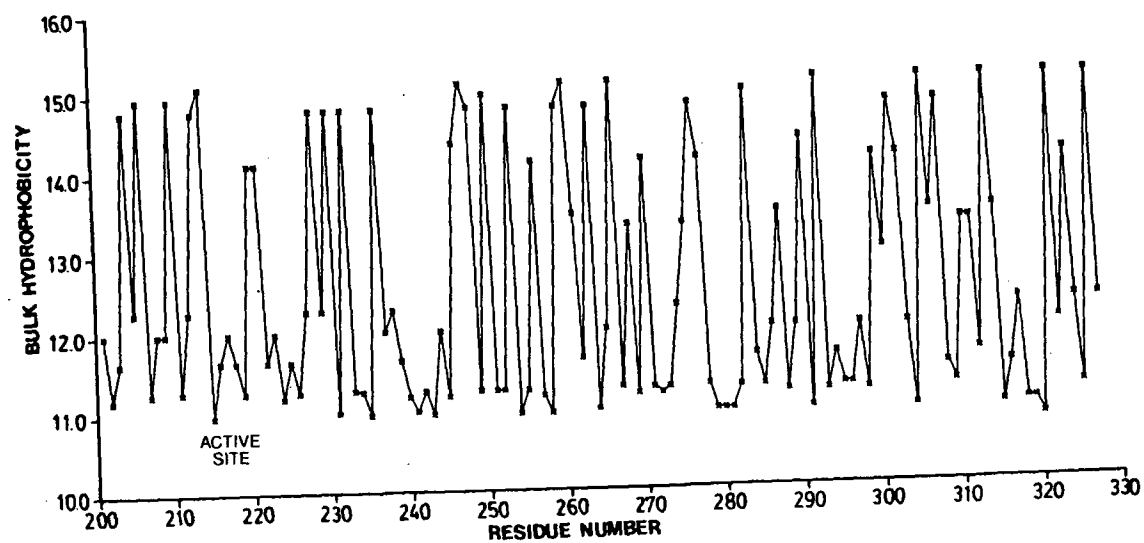
Figure 13b. The bulk hydrophobicity profile of pepsin for residues 200 to 327.
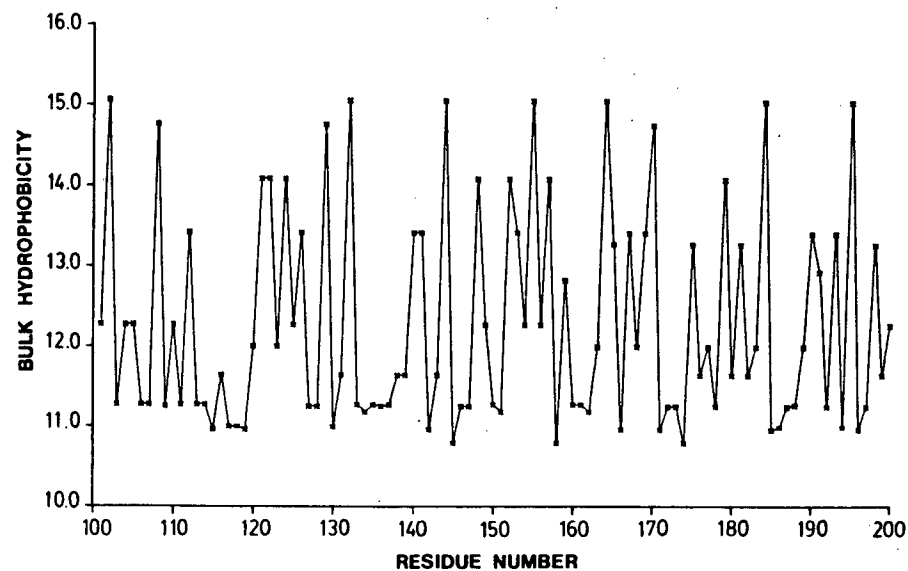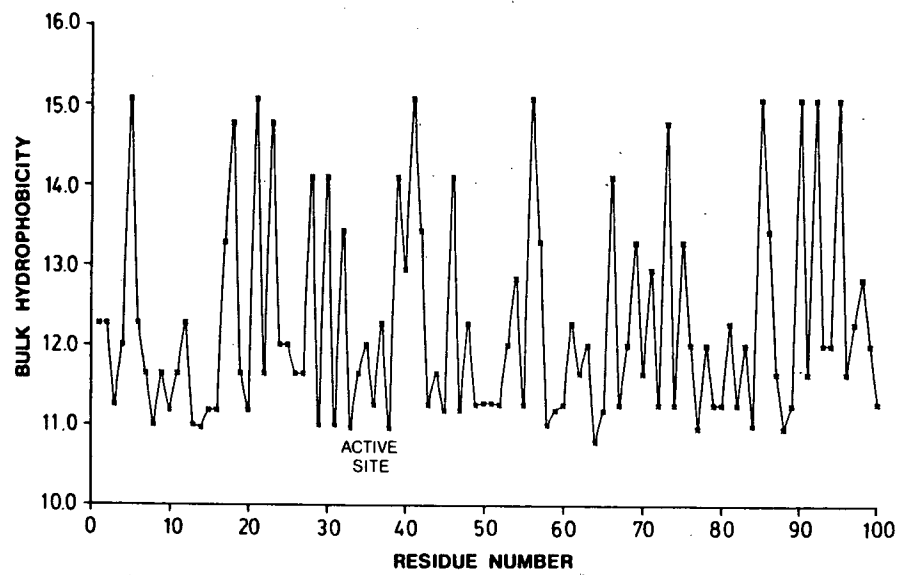
Figure 14a. The bulk hydrophobicity profile of penicillopepsin for residues 1 to 100 and 100 to 200.
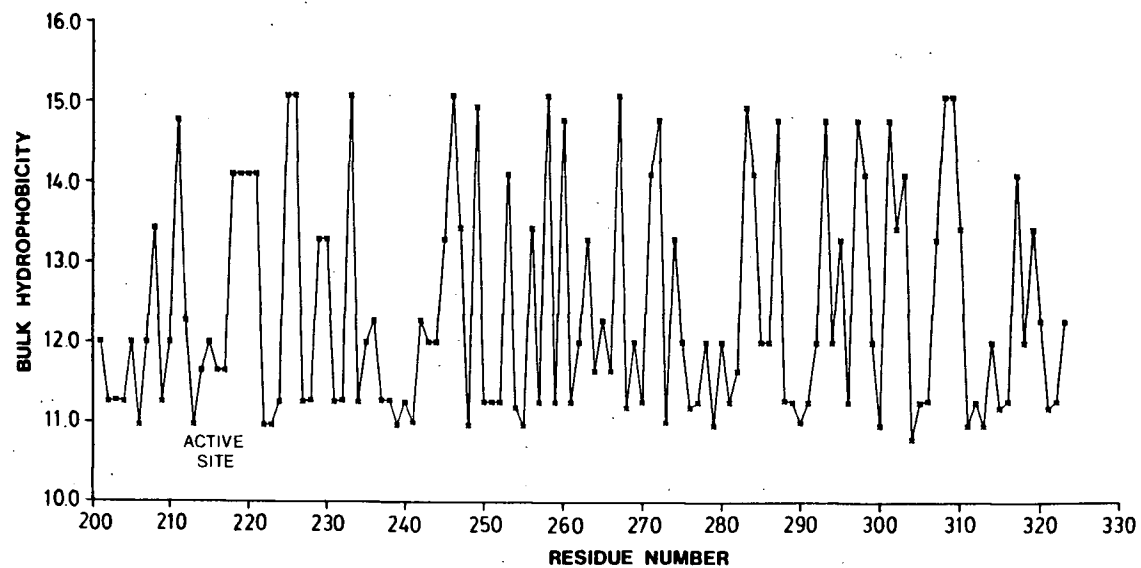
Figure 14b. The bulk hydrophobicity profile of penicillopepsin for residues 200 to 323.
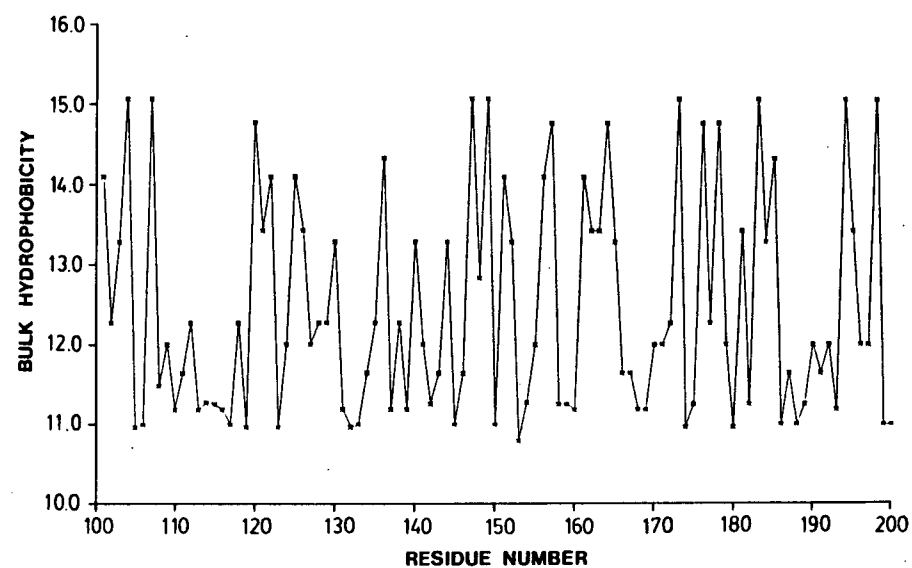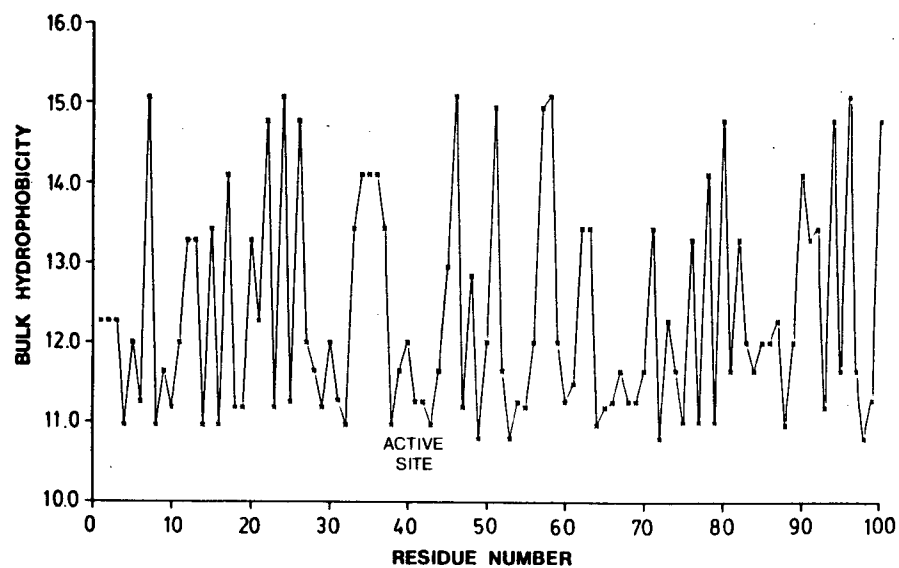
Figure 15a. The bulk hydrophobicity profile of <u>Mucor</u> <u>miehei</u> proteinase for residues 1 to 100 and 100 to 200.
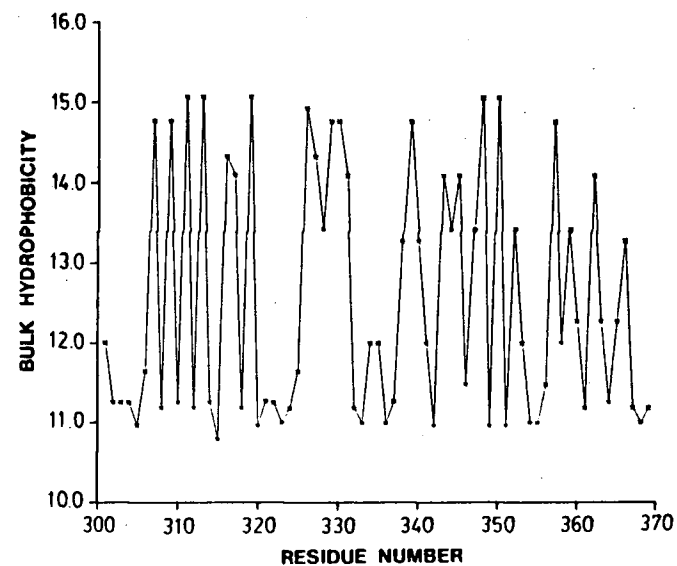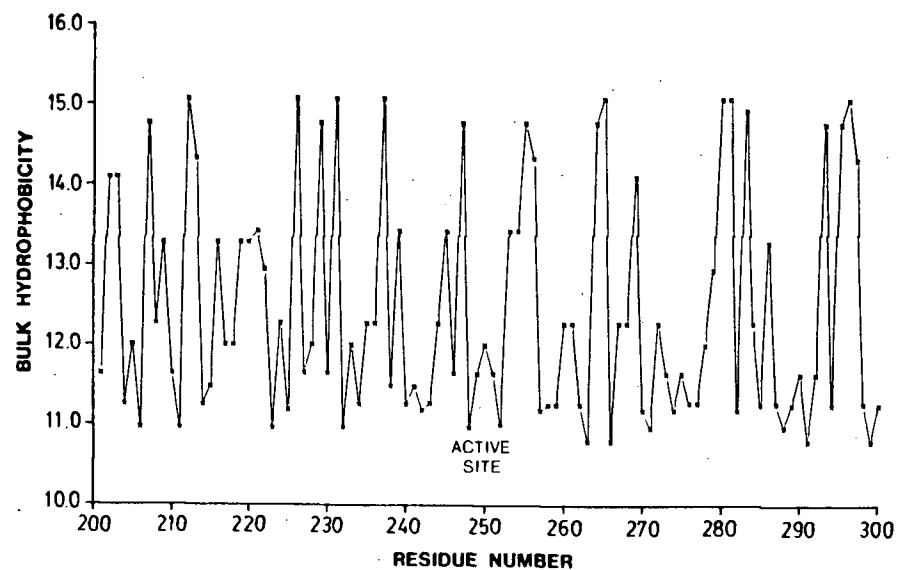
Figure 15b. The bulk hydrophobicity profile of <u>Mucor</u> <u>miehei</u> proteinase for residues 200 to 300 and 300 to 369.

The hydrophobicity profile method revealed that the active site aspartic acid residues for the four proteinases were located in areas of low hydrophobicity (i.e. hydrophilic). These areas were identified as β-turn regions. Rose (1978) hypothesized that "turns occur at those sites in the polypeptide chain where the hydrophobicity is at a local minimum". Kuntz (1972), in a paper on protein folding, observed that the residues located at β-turns were relatively polar in nature. Areas adjacent to the active site were predicted as β-sheet and it was also noted that these areas were relatively hydrophobic. These observations may be of importance in defining the interaction between enzyme and substrate. Raap et al. (1983) analyzed the circular dichroism spectra for peptides containing the labile Phe-Met bond of κ-casein (in dilute sodium dodecyl sulfate solutions) and found that these peptides had β-structure forming potential. Similar results were obtained by the authors when they analyzed the peptide sequence using various secondary structure predictive methods (i.e. Lim (1974) and Chou and Fasman (1978b)). The results obtained in the present study appear to support the hypothesis put forth by Jenkins et al. (1976) and by Raap et al. (1983) that κ-casein may interact with the active site region of the enzyme through β-sheet-β-sheet interactions during the milk-clotting process.

The results obtained from the diagonal plot and the secondary structure prediction methods revealed that various structural features were common to chymosin, pepsin, penicillopepsin and Mucor miehei proteinase which belong to the aspartyl proteinases. Visser (1981), in

a review on proteolytic enzymes, stated that the overall conformation of one proteinase molecule may be very similar to that of other proteinases from the same class.

C. CIRCULAR DICHROISM SPECTRA

1. Far-UV spectra (190 to 240 nm)

The CD spectra of various aspartyl proteinases as a function of pH (pH 5.0 to 8.0) were measured over the wavelength range 190 to 240 nm and are represented as follows: chymosin (Figure 16); pepsin (Figure 17); Mucor miehei proteinase (Figure 18); Mucor pusillus proteinase (Figure 19); Endothia parasitica proteinase (Figure 20); Aspergillus saitoi proteinase (Figure 21) and penicillopepsin (Figure 22).

Chymosin showed three distinctive patterns in the CD spectra over the pH range (Figure 16). The magnitude of the peak at 193 nm of the pattern exhibited at low pH values (pH 5.0 and 5.3) was lower in comparison to those spectra at pH values above 5.3. At pH 5.8, 6.3 and 7.0 a shoulder at 215 nm also became evident. At pH 8.0 a blue shift in the maxima at 193 nm occurred although peak magnitude was similar to those spectra recorded at pH values greater than 5.3.

The change in CD spectra seen as a function of pH may be due to the associative properties of chymosin. Baldwin and Wake (1959) using sedimentation studies demonstrated that chymosin in solution was susceptible to polymerization; the extent of polymerization was dependent on both ionic strength and the concentration of the enzyme in solution. Dimers were found to form at high concentrations of enzyme and low ionic
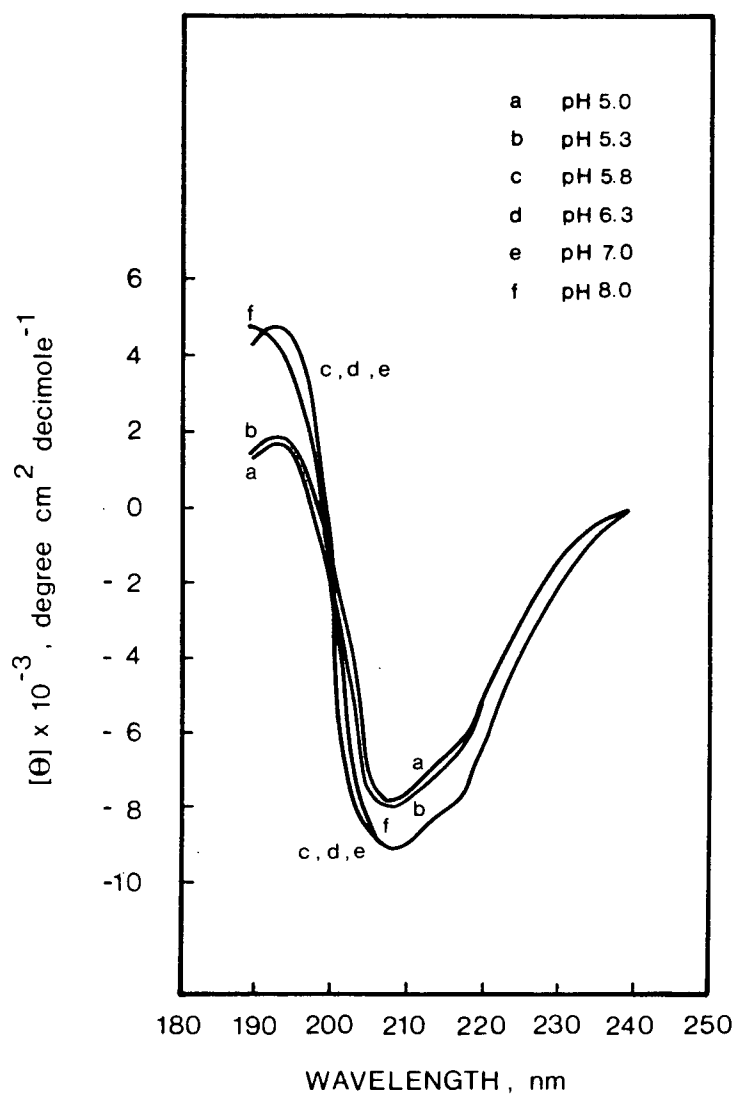
Figure 16. The effect of pH on the far-UV CD spectra of chymosin.
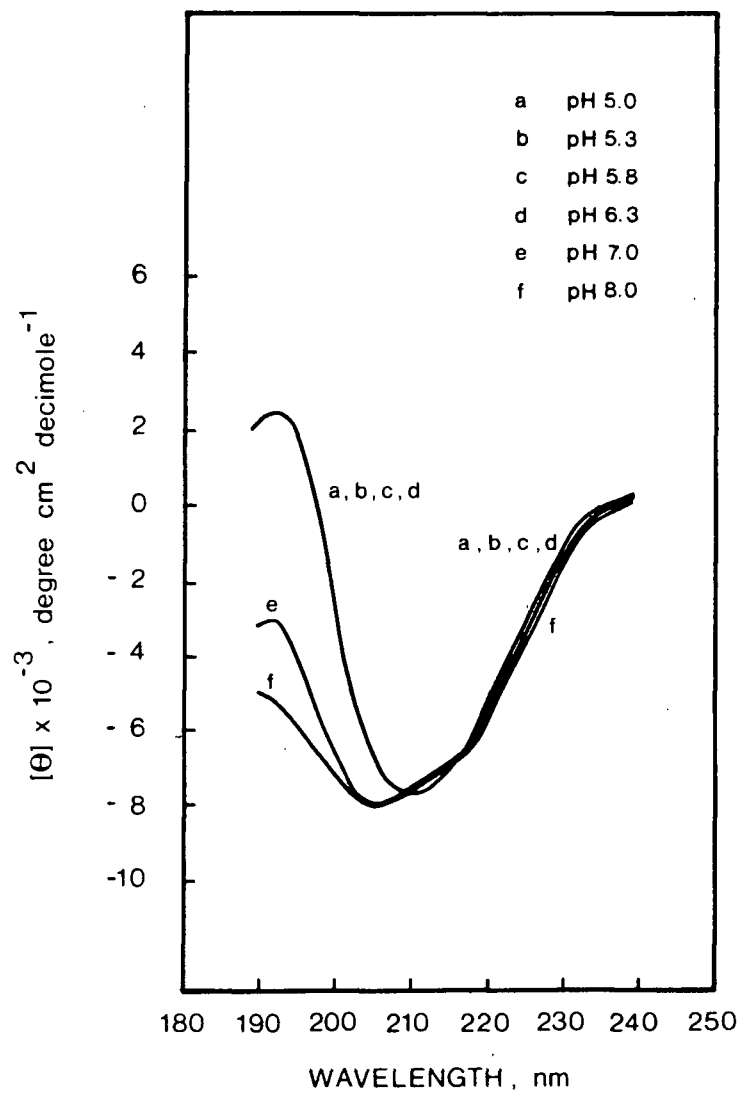
Figure 17. The effect of pH on the far-UV CD spectra of pepsin.

a   pH 5.0
b   pH 5.3
c   pH 5.8
d   pH 6.3
e   pH 7.0
f   pH 8.0
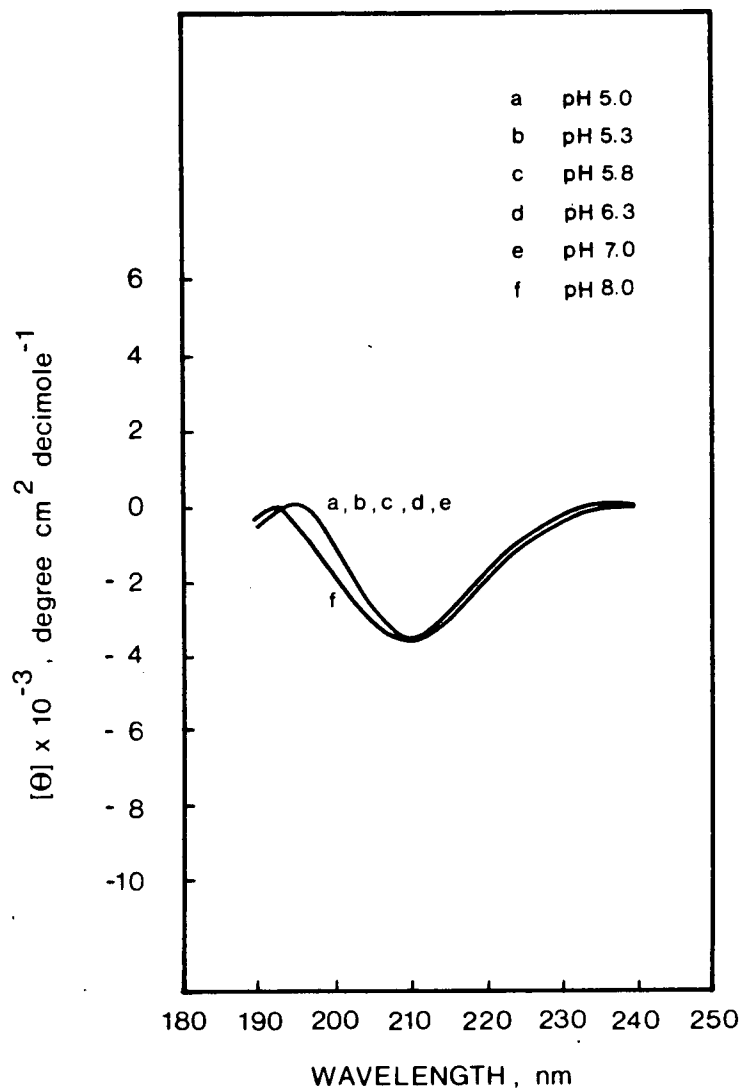
Figure 18. The effect of pH on the far-UV CD spectra of <u>Mucor</u> <u>miehei</u> proteinase.
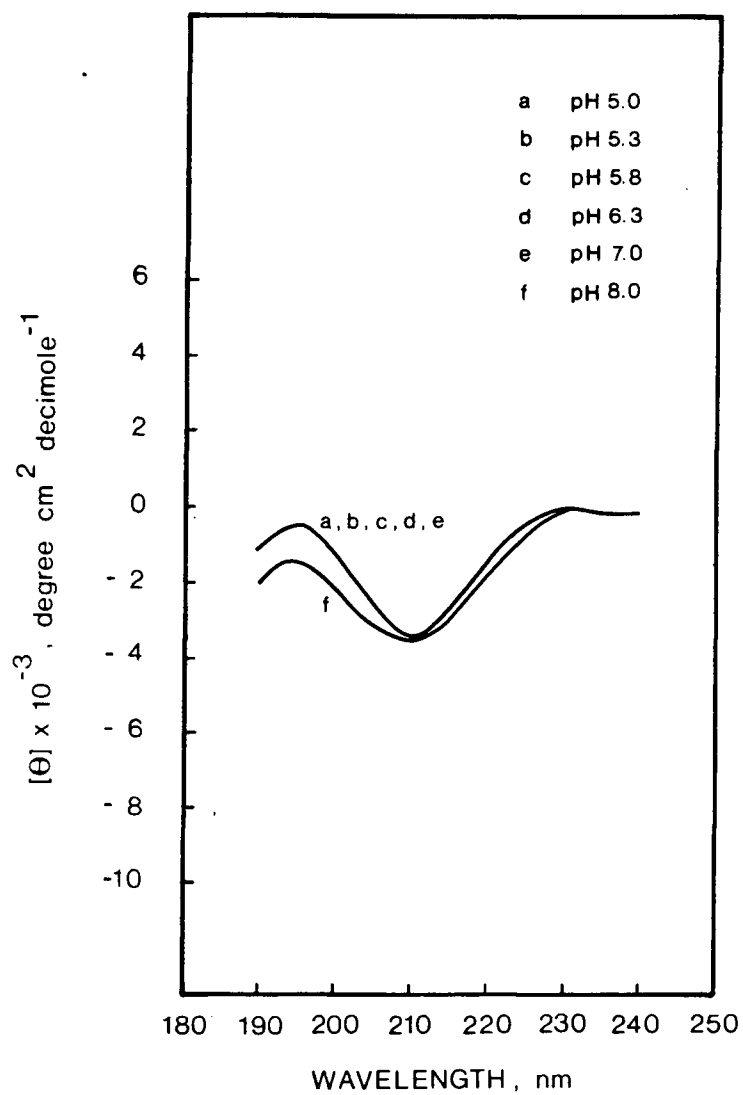
Figure 19. The effect of pH on the far-UV CD spectra of <u>Mucor pusillus</u> proteinase.
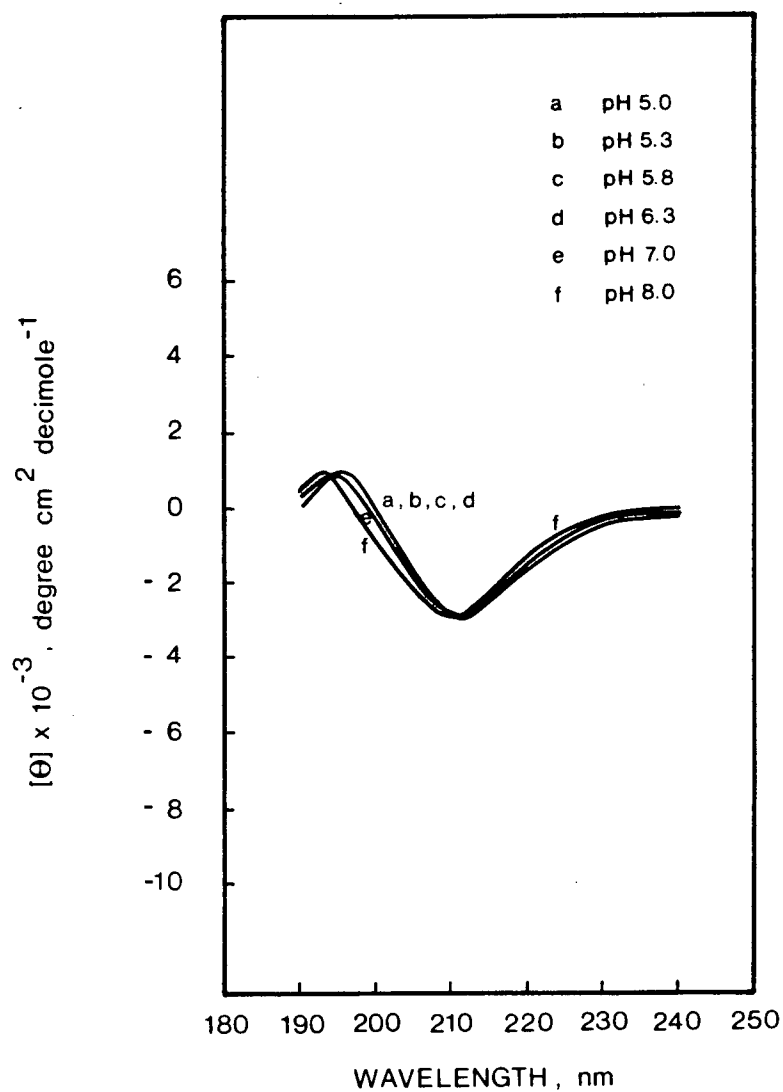
Figure 20. The effect of pH on the far-UV CD spectra of <u>Endothia parasitica</u> proteinase.
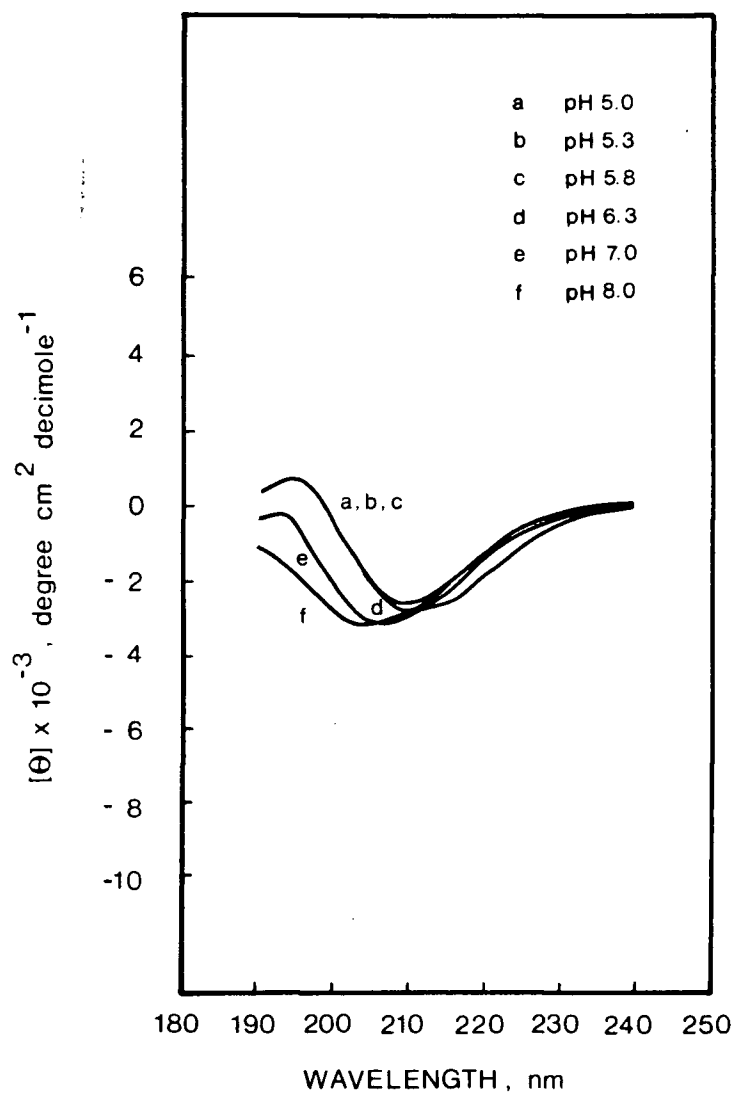
Figure 21. The effect of pH on the far-UV CD spectra of <u>Aspergillus</u> <u>saitoi</u> proteinase.
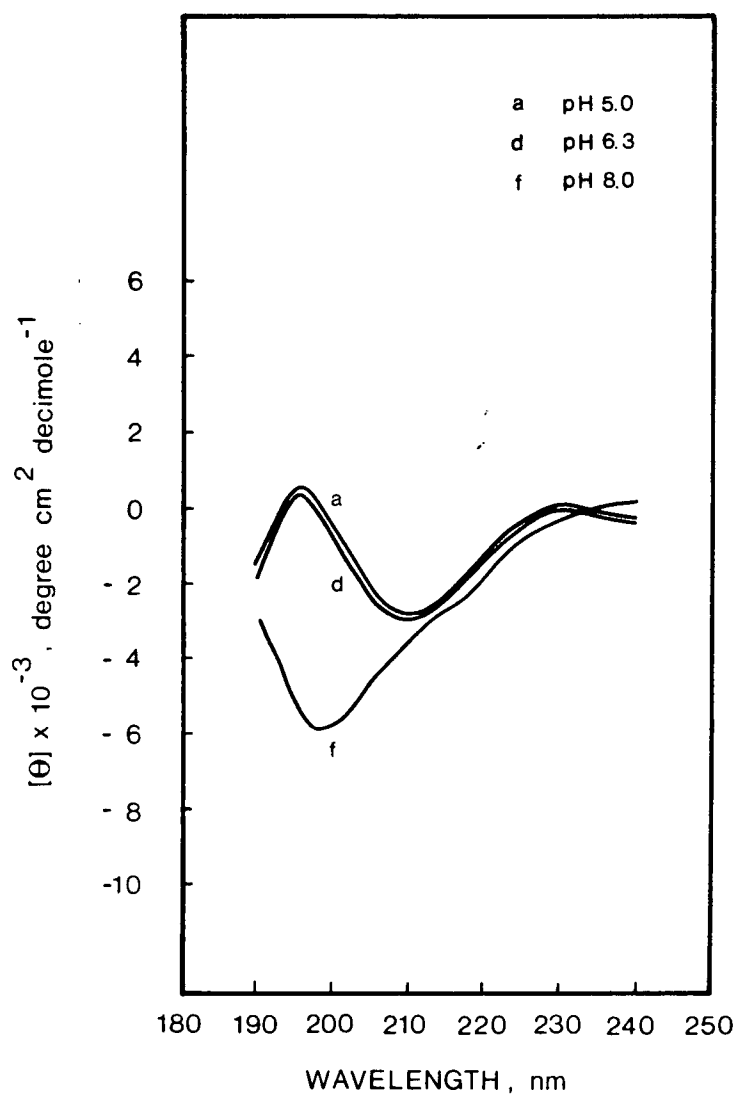
Figure 22. The effect of pH on the far-UV CD spectra of penicillo-pepsin.

strength, however, at low chymosin concentrations only the monomer was evident. Djurtoft et al. (1963) in a similar study showed that the sedimentation coefficient for chymosin increased with increasing concentration; increases in sedimentation were highly noticeable at chymosin concentrations greater than 3 mg/mL. In the present study, chymosin concentrations of less than 1.0 mg/mL were generally used at an ionic strength of 0.01.

The CD spectra seen at pH 5.0 and pH 5.3 might also be associated with a conformational change which may occur as chymosin approaches its isoelectric point at pH 4.6. The appearance of a precipitate was noted at pH 5.0 and 5.3. The precipitates were removed from the protein solutions prior to CD measurement. The CD spectrum of chymosin seen at pH 8.0 may be the result of the protein unfolding at alkaline pH values.

Pepsin showed virtually no change in the spectra from pH 5.0 to pH 6.3 (Figure 17). However, at pH values greater than 6.3 a drastic change in spectra occurred. Ahmad and McPhie (1978) working with the diazoacetylglycine ethyl ester derivative of swine pepsin (to prevent autolysis) observed that the CD spectrum of the proteinase underwent a substantial conformational change at pH 7.0. In the present study the change in conformation as reflected in CD spectra at pH values greater than 6.3 may be due to protein unfolding as a result of breakage of hydrogen bonds involving carboxyl groups (Edelhoch, 1958a and 1958b).

Nakayama et al. (1983) examined the electron spin resonance (ESR) spectra of 4-(3-diazo-2-oxopropylidene)-2,2,6,6-tetramethylpiperidine-1-oxyl labelled porcine pepsin over a pH range of 1.8 to 12.0. Between

pH 1.8 and 6.2 the ESR spectra remained unchanged and were representative of native porcine pepsin. At pH values greater than 7.4 the shape of the ESR spectra changed to a broad triplet which was assumed to represent denatured porcine pepsin. ESR spectra at pH values between 6.2 and 7.1 were a composite of the native and denatured spectra. The authors postulated that the change in conformation seen at alkaline pH values may be associated with the inactivation of pepsin reported in the literature.

Mucor miehei proteinase (Figure 18) showed very little change in the CD spectra over the pH range, although a slight change in the CD spectrum was noted at pH 8.0. Ottesen and Rickert (1970a) and Alais and Lagrange (1972) reported the loss of enzymatic activity for Mucor miehei proteinase at pH values near 8.0. The CD spectra of Mucor pusillus proteinase (Figure 19) also remained virtually unchanged for pH values 5.0 to 7.0. A decrease in the amplitude of the peak in the 190 to 200 nm range became evident at pH 8.0. Arima et al. (1970) reported that the milk-clotting activity of the enzyme decreased markedly at pH values above 7.0.

Endothia parasitica proteinase (Figure 20) showed very little change in the CD spectra over the pH range although a slight blue shift in the peak maximum in the 190 to 195 nm range was evident at pH values of 7.0 and 8.0. This slight change in the CD spectra at pH 7.0 and pH 8.0 may be associated with a loss of enzymatic activity reported in the literature. Whitaker (1970), in stability tests with the proteinase, reported that activity was rapidly lost above pH 6.5 while loss of activity was almost instantaneous above pH 8.0.

The Aspergillus saitoi proteinase CD spectra (Figure 21) remained relatively unchanged from pH 5.0 to pH 6.3, however, at pH values greater than 6.3 a breakdown in the 190 to 200 nm region was evident. Ichishima and Yoshida (1967) studied the optical rotatory properties of Aspergillus saitoi proteinase and found that the ORD spectra did not change at pH values between 2.7 and 5.7, but noted a breakdown in the ORD spectra on exposure to alkali conditions (i.e. pH values greater than 7.0).

The CD spectra of penicillopepsin at pH 5.0 and pH 6.3 (Figure 22) were virtually identical, however, the CD spectrum at pH 8.0 no longer showed a peak in the 190 to 200 nm region. This change in CD spectrum at pH 8.0 was very similar to that seen for pepsin at pH 7.0 and pH 8.0. Sodek and Hofmann (1970b) examined the effect of pH on the stability of penicillopepsin and found that at pH values greater than 6.6 there was a rapid drop in activity, similar to that reported for pepsin by Ahmad and McPhie (1978).

Based on the examination of the CD spectra for the various aspartyl proteinases some general observations could be made. The CD spectra were characterized by a maximum in the 190 to 200 nm range as well as a minimum around 210 nm. With the exception of chymosin, the CD spectra for the proteinases remained relatively constant in the lower pH range; the CD spectra were found to change at or above the neutral pH range depending on the proteinase examined.

2. Secondary structure determination (far-UV)

In order to quantitate the CD spectra of the various proteinases

as a function of pH, the method of Provencher and Glöckner (1981) was used to analyze the CD data for the determination of secondary structure. The results of the secondary structure determination are presented in Table 6. As would be expected, changes observed in the CD spectra tended to be reflected by changes in the amounts of secondary structure determined. The aspartyl proteinases were generally characterized by large amounts of β-sheet. The X-ray diffraction analysis of various aspartyl proteinases conducted by several researchers (Andreeva et al., 1976; Jenkins et al., 1976; Subramanian et al., 1976; James and Sielecki, 1983) has shown that a large portion of these molecules consist of parallel and anti-parallel β-sheet structures. The X-ray results together with the CD results of the present study would tend to indicate that as a class the aspartyl proteinases have similar secondary structures.

Of the aspartyl proteinases examined, chymosin contained a relatively high proportion of ∝-helix, a feature which readily distinguished it from the other aspartyl proteinases. At the low pH values (pH 5.0 and pH 5.3) the amount of ∝-helix was lower than that seen at pH values greater than 5.3 while the opposite was true for the β-sheet fractions.

At pH values of 6.3 and lower, the secondary structure fractions of pepsin compared favourably to those obtained by Ahmad and McPhie (1978) who calculated 0.00, 0.58 and 0.42 for ∝-helix, β-sheet and unordered fractions respectively for pepsin at pH 4.5 using the method of White (1976). Rao and Dunn (1981) using the CD spectrum of porcine pepsin at pH 5.5 determined that the secondary structure consisted of

Table 6. Secondary structure determination from CD spectra for various proteinases.

Protein			Secondary structure fraction			
			α-helix	β-sheet	β-turn	random
Aspartyl proteinases						
chymosin	pH	5.0	0.09	0.44	0.19	0.28
		5.3	0.15	0.43	0.17	0.25
		5.8	0.24	0.37	0.14	0.25
		6.3	0.23	0.35	0.15	0.27
		7.0	0.22	0.37	0.16	0.25
		8.0	0.22	0.36	0.16	0.26
pepsin	pH	5.0	0.12	0.59	0.14	0.15
		5.3	0.13	0.60	0.13	0.14
		5.8	0.13	0.58	0.14	0.15
		6.3	0.12	0.58	0.15	0.15
		7.0	0.12	0.48	0.16	0.24
		8.0	0.12	0.44	0.15	0.29
M. miehei proteinase	pH	5.0	0.01	0.61	0.22	0.16
		5.3	0.01	0.62	0.21	0.16
		5.8	0.01	0.60	0.22	0.17
		6.3	0.02	0.60	0.21	0.17
		7.0	0.03	0.60	0.20	0.17
		8.0	0.03	0.58	0.22	0.17
M. pusillus proteinase	pH	5.0	0.01	0.62	0.20	0.17
		5.3	0.00	0.61	0.22	0.17
		5.8	0.00	0.61	0.21	0.18
		6.3	0.01	0.61	0.20	0.18
		7.0	0.03	0.60	0.19	0.18
		8.0	0.03	0.58	0.20	0.19
E. parasitica proteinase	pH	5.0	0.00	0.62	0.23	0.15
		5.3	0.00	0.63	0.22	0.15
		5.8	0.00	0.62	0.21	0.17
		6.3	0.03	0.58	0.22	0.17
		7.0	0.02	0.59	0.22	0.17
		8.0	0.03	0.57	0.19	0.21

Table 6. (continued)

Protein			Secondary structure fraction			
			α-helix	β-sheet	β-turn	random
A. <u>saitoi</u> proteinase	pH	5.0	0.00	0.62	0.23	0.15
		5.3	0.00	0.62	0.23	0.15
		5.8	0.00	0.63	0.22	0.15
		6.3	0.00	0.63	0.21	0.16
		7.0	0.00	0.55	0.24	0.21
		8.0	0.03	0.51	0.20	0.26
penicillopepsin	pH	5.0	0.00	0.70	0.23	0.07
		6.3	0.00	0.69	0.22	0.09
		8.0	0.01	0.43	0.22	0.34
Non-aspartyl proteinases						
papain	pH	6.3	0.14	0.36	0.22	0.28
α-chymotrypsin	pH	6.3	0.03	0.50	0.29	0.18
trypsin	pH	6.3	0.06	0.47	0.28	0.19

0.11 α-helix, 0.39 β-sheet and 0.50 random structure. The content of β-sheet determined by these authors was low compared to the results of the present study and might have been due to the least squares method used for the analysis. This method for the determination of secondary structure fractions used reference protein spectra which were generated from only five proteins. In the present study the proportion of β-sheet decreased at pH 7.0 and pH 8.0 while the fraction of unordered structure increased. This loss of organized structure would tend to suggest that the molecule was unfolding. Edelhoch (1957) found that as pepsin was exposed to alkaline pH conditions, a change in viscosity occurred which was attributed to the unfolding of the protein to a linear polyelectrolyte.

Secondary structure determination from the CD spectra of Endothia parasitica proteinase over the pH range indicated a β-sheet content of approximately 0.6 with very little α-helix content. Jenkins et al. (1976) using three different methods for the analysis of the CD spectrum for E. parasitica proteinase at pH 4.48, found that the amount of β-sheet ranged from 0.34 to 0.59 while the proportion of α-helix ranged from 0.03 to 0.08 depending on the algorithm used. As the pH was increased from 5.0 to 8.0 slight decreases in β-sheet with corresponding slight increases in both α-helix and unordered fractions were observed. Similar trends in the secondary structure fractions over the pH range were noted for both Mucor miehei proteinase and Mucor pusillus proteinase. Secondary structure data revealed that these two proteinases were almost identical in structural nature. Etoh et al. (1979) studied the

physical-chemical and immunochemical properties of Mucor miehei protein-ase and Mucor pusillus proteinase and found that a high structural similarity exists between the two proteinases.

Secondary structure determination of Aspergillus saitoi proteinase indicated a β-sheet fraction of approximately 0.6 and no α-helix at pH values of 6.3 and lower. Ichishima and Yoshida (1966a) on the basis of ORD and infrared data concluded that the proteinase from Aspergillus saitoi was void of helical conformation and that the molecule existed in the anti-parallel β-structure. Aspergillus saitoi proteinase, as with the other microbial proteinases, also exhibited a decrease in the β-sheet fraction with a corresponding increase in the unordered fraction. Ichishima and Yoshida (1967) found that at pH values above 7.0 the rotatory dispersion curve resembled that of an unordered coil found in denatured proteins.

Penicillopepsin was determined to have a high β-sheet content of approximately 0.7 at pH 5.0 and pH 6.3, analogous to the other aspartyl proteinases from microbial sources. This result compares favourably to that obtained by James and Sielecki (1983) who calculated a β-sheet content of 0.66 based on results obtained from X-ray diffraction analy-sis of penicillopepsin. In the present study the fraction of β-sheet decreased markedly from approximately 0.7 to 0.4 at pH 8.0; the propor-tion of unordered fraction greatly increased from approximately 0.1 at pH 5.0 and pH 6.3 to 0.3 at pH 8.0. These changes would suggest that the molecule was unfolded as a result of exposure to alkaline pH condi-tions. The structural behaviour exhibited by penicillopepsin is thus

very similar to that displayed by pepsin. Mains et al. (1971) found that many of the physical-chemical properties of penicillopepsin and pepsin were very similar.

The secondary structure data for the non-aspartyl proteinases papain, α-chymotrypsin and trypsin at pH 6.3 are also presented in Table 6. The fact that the three non-aspartyl proteinases had relatively high contents of β-sheet may aid in explaining the reported milk-clotting ability of these enzymes (Ernstrom, 1974).

On the basis of the results obtained for the aspartyl proteinases it would appear that an increase in pH to the neutral-alkaline range results in a transition of the secondary structure from β-sheet to unordered fraction, which may be indicative of denaturation. The extent of transition, however, was proteinase dependent. Chymosin, Mucor miehei proteinase, Mucor pusillus proteinase and Endothia parasitica proteinase showed relatively strong resistance to structural change with an increase in pH, whereas pepsin, Aspergillus saitoi proteinase and penicillopepsin were markedly affected by the pH change.

3. Near-UV spectra (240 to 320 nm)

The near-UV (240 to 320 nm) spectra for chymosin, pepsin, Mucor miehei proteinase, Mucor pusillus proteinase, Endothia parasitica proteinase and Aspergillus saitoi proteinase are presented in Figures 23, 24, 25, 26, 27 and 28, respectively. In contrast to the far-UV CD spectra which is indicative of secondary structure, the near-UV CD spectra reflects changes in the tertiary structure (Strickland, 1974).
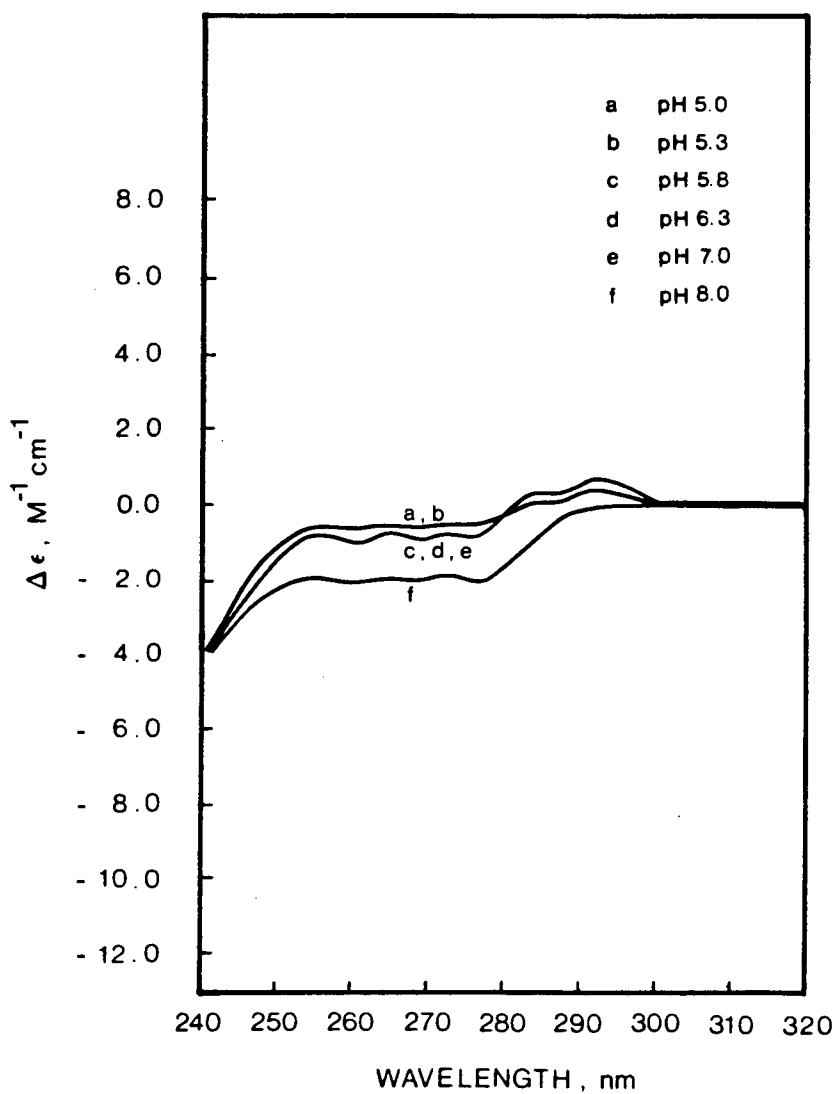
Figure 23. The effect of pH on the near-UV CD spectra of chymosin.
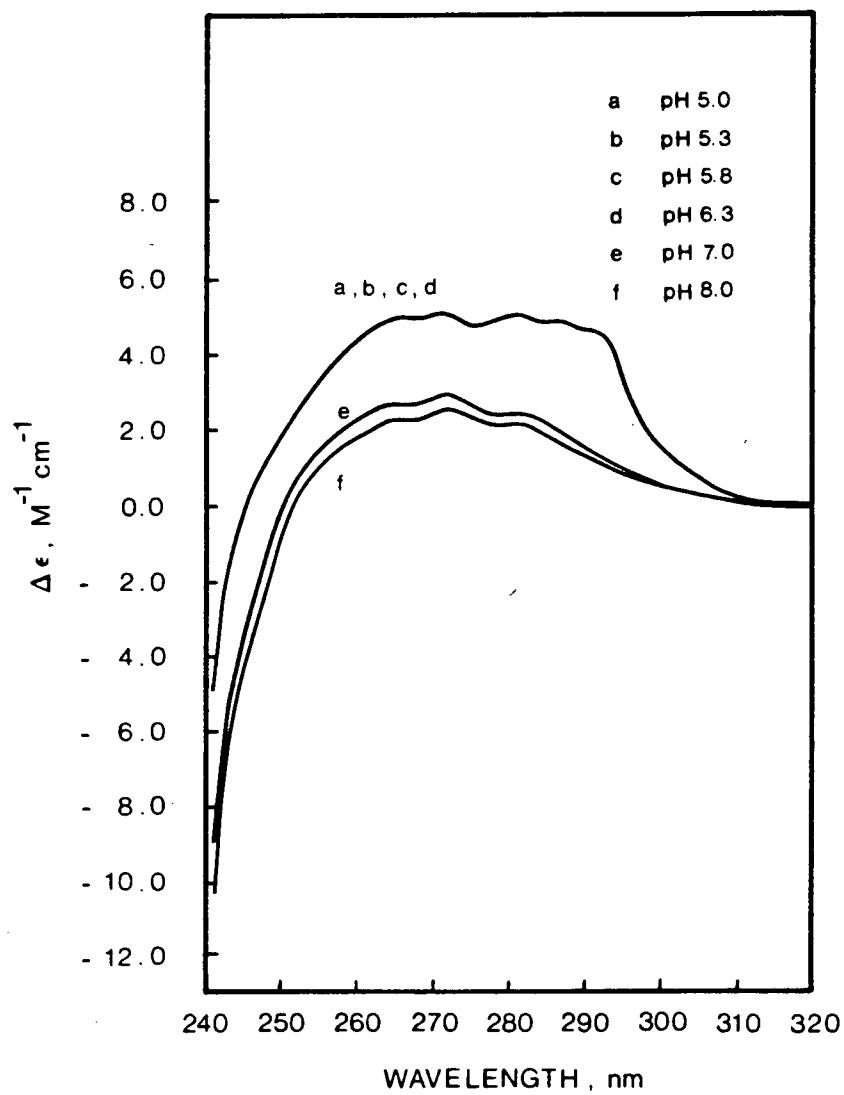
Figure 24. The effect of pH on the near-UV CD spectra of pepsin.
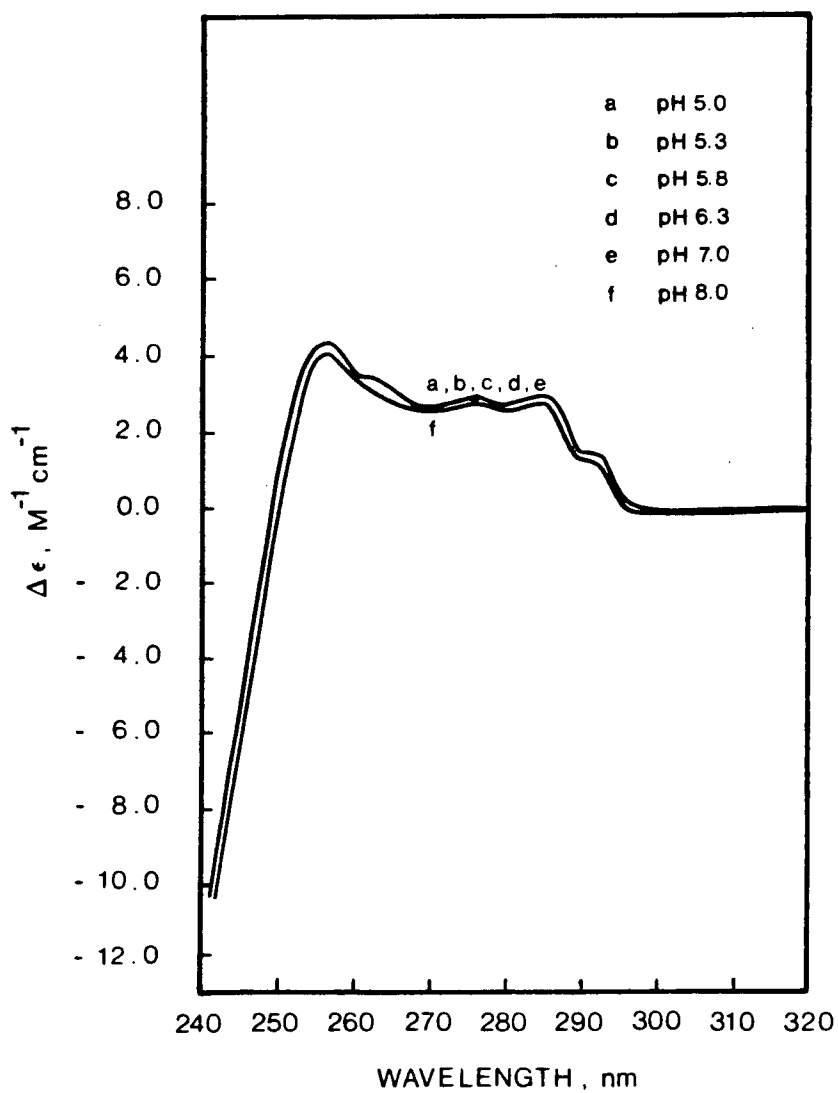
Figure 25. The effect of pH on the near-UV CD spectra of <u>Mucor</u> <u>miehei</u> proteinase.
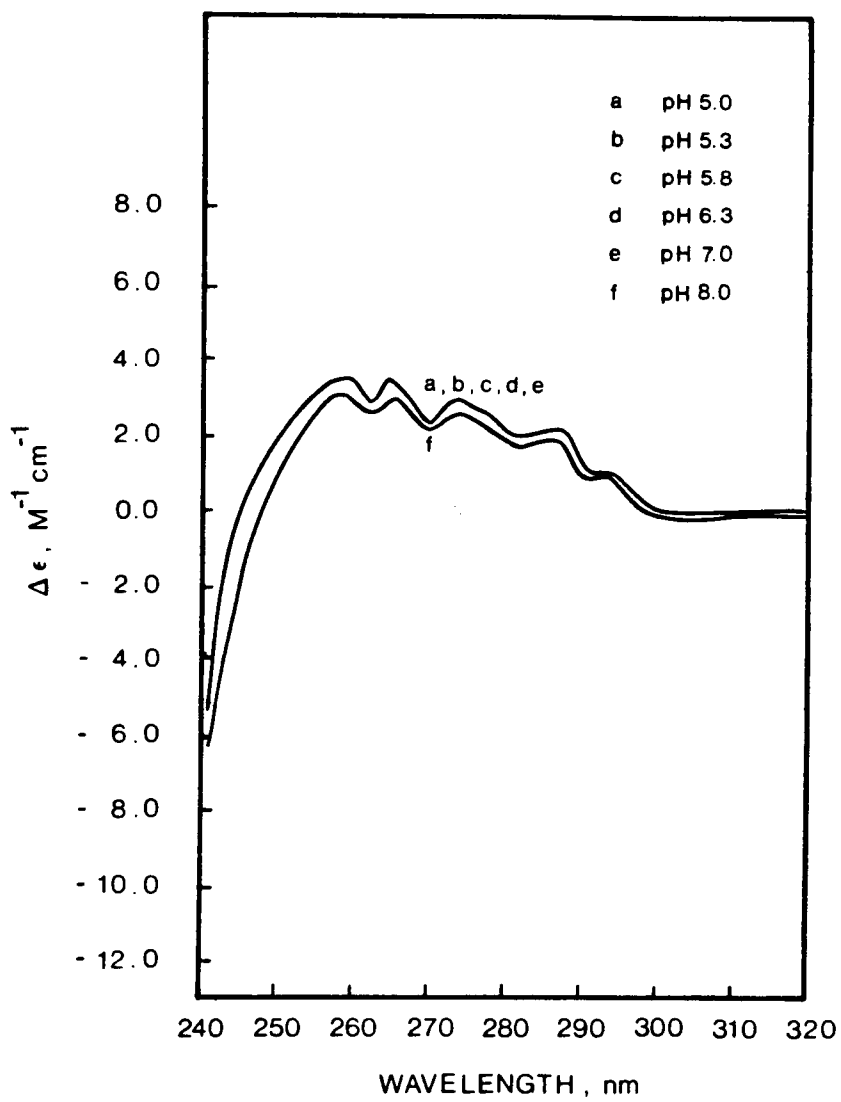
Figure 26. The effect of pH on the near-UV CD spectra of <u>Mucor</u> <u>pusillus</u> proteinase.
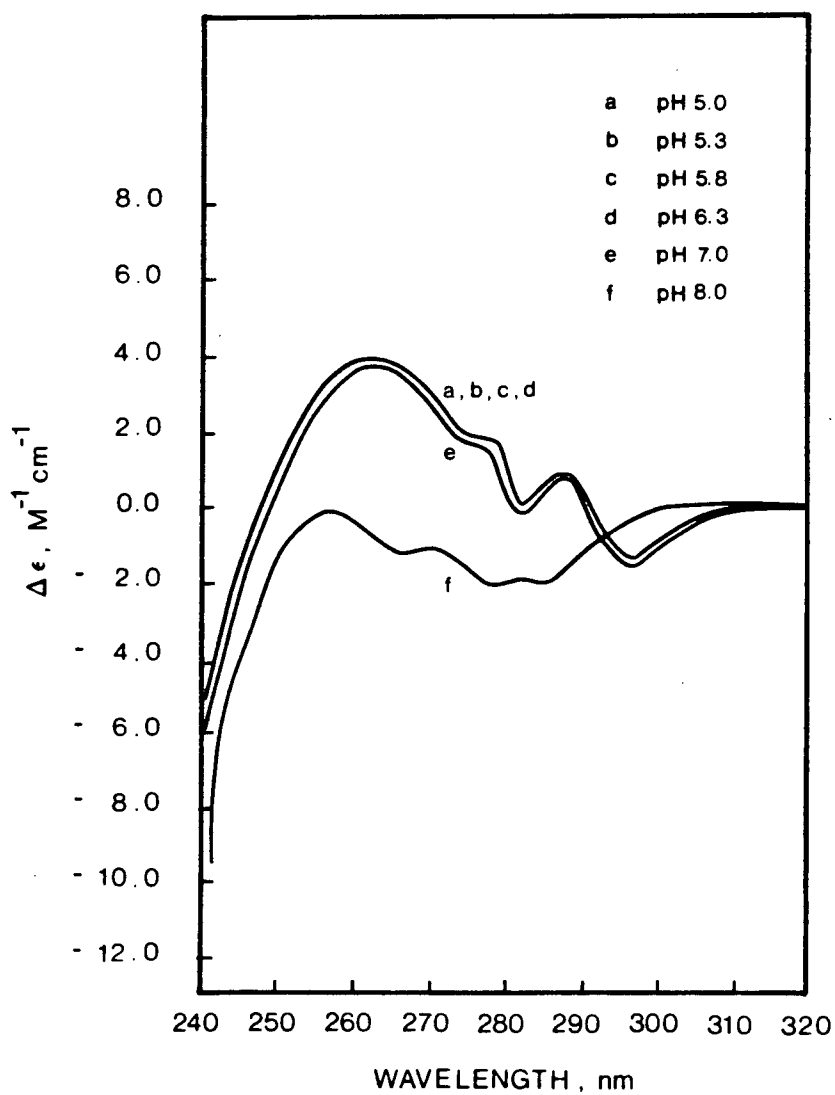
Figure 27. The effect of pH on the near-UV CD spectra of _Endothia_ _para-sitica_ proteinase.
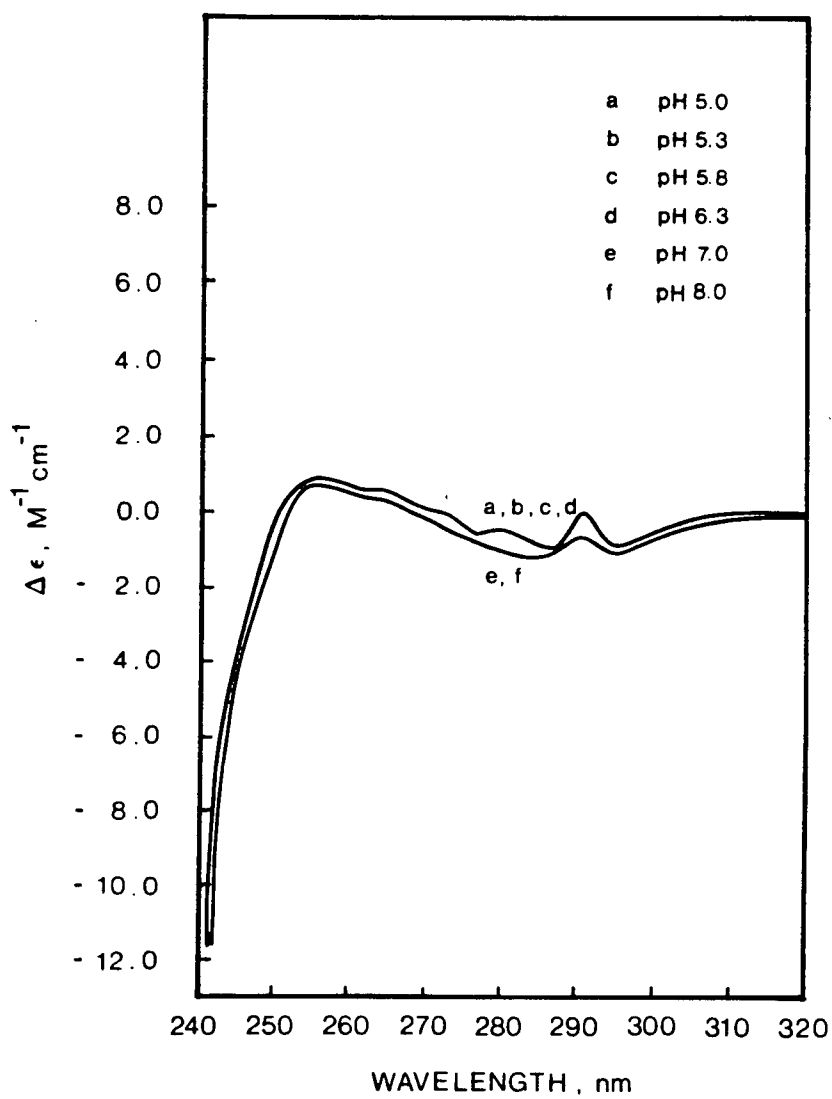
Figure 28. The effect of pH on the near-UV CD spectra of Aspergillus
saitoi proteinase.

In the near-UV region, the aromatic rings of tyrosine, tryptophan and phenylalanine give rise to CD bands through interactions with the amino acid moiety or with nearby groups in the protein. In addition, the disulfide chromophore of cystinyl residues may contribute to the near-UV spectra (Strickland, 1974; Heindl et al., 1980). In general, the changes observed in the CD spectra from the 190 to 240 nm range were reflected in the CD spectra from the near-UV range.

The relative intensity of the near-UV spectra for chymosin (Figure 23) was low (in the range of 0 to 2 $M^{-1}cm^{-1}$) which may result from the relatively large number of aromatic residues found in the molecule. Chymosin contains approximately 12 percent aromatic amino acids. Proteins containing large numbers of aromatic side chains may not have very large CD bands as a result of cancellations by positive and negative contributions (Strickland, 1974). Since intense CD bands occur when aromatic groups are in close proximity to one another it is also possible that the aromatic amino acids found in the tertiary structure of chymosin are not close to one another. Due to the low intensity of the CD spectra especially at low pH values (pH 5.0 and pH 5.3), the identification of the fine structure for the various aromatic groups was difficult. Chymosin also contains three disulfide linkages (Foltmann et al., 1977), however the shape and intensities of disulfide CD bands have not been well characterized (Strickland, 1974); therefore, no attempts were made to identify these structures. Identification of the disulfide fine structure was not attempted for any of the other aspartyl proteinases examined.

Tentative identification of phenylalanine fine structure for chymosin was made at 261 nm and 269 nm while the O+850 cm^{-1} ^1L_b and the O-O cm^{-1} ^1L_b bands of tryptophan were assigned to peaks centered at 285 nm and 292 nm, respectively. The O+800 cm^{-1} band of tyrosine was identified at 276 nm, however, the O-O cm^{-1} band was not detected and may have been obscured by the tryptophan fine structure. No change in the pattern of fine structure was noted until the pH was increased to 8.0 where the loss of tryptophan fine structure (bands at 285 nm and 292 nm) became apparent. This loss of fine structure may indicate a slight perturbation of the tertiary structure resulting from the motility of the tryptophan side chains. In studies with model compounds it has been shown that an increase in the motility of the aromatic side chains tends to decrease the CD intensity (Strickland, 1974). Although no change in CD fine structure was apparent at pH 5.0 and pH 5.3, the decreased intensity of the spectra may indicate a change in tertiary structure. The intensity of the near-UV spectra is affected by the rigidity of the protein, interactions of aromatic rings with its surroundings and the number of aromatic residues (Strickland, 1974).

The near-UV CD spectra of pepsin is presented in Figure 24. Unlike the chymosin CD spectra which was mainly negative, the CD spectra of pepsin was positive. Tentative identification of the phenylalanine fine structure was made at 266 nm and 271 nm, the O+800 cm^{-1} band of tyrosine at 281 nm and the O+850 cm^{-1} ^1L_b and the O-O cm^{-1} ^1L_b bands of tryptophan at 287 nm and 292 nm, respectively. Again, as with chymosin the O-O cm^{-1} band of tyrosine was absent. The appearance of a

shoulder in the 295 to 305 nm region may indicate the presence of the ^1L_a band of tryptophan, however, the CD band due to disulfides may also be contributing to the region. Pepsin contains three disulfide bonds (Moravek and Kostka, 1974). No change in the spectra with an increase in pH was observed until pH 7.0 when a general decrease in CD intensity was noted. As with chymosin, a decrease in intensity was accompanied by a loss of tryptophan fine structure which may indicate a change in tertiary structure.

The near-UV spectra of Mucor miehei proteinase (Figure 25) and Mucor pusillus proteinase (Figure 26) were relatively similar. This similarity in tertiary structure supports the work of Etoh et al. (1979), who concluded that the two proteinases shared a "morphological identity". In comparison to chymosin and pepsin, the resolution of some fine structures for M. miehei and M. pusillus proteinases was improved, especially for phenylalanine. With Mucor miehei proteinase the phenyl-alanine fine structure was identified at 257 nm and 263 nm, with the 0+800 cm^{-1} band of tyrosine at 276 nm and the 0+850 cm^{-1} ^1L_b band and 0-0 cm^{-1} ^1L_b band of tryptophan at 285 nm and 292 nm, respectively. As the pH was increased, very little change occurred in the near-UV spectra other than the disappearance of the phenylalanine peak at 263 nm. Only minor changes accompanied an increase in pH which would sug-gest that the tertiary structure of Mucor miehei proteinase is relative-ly stable to pH change. A similar stability to pH was also seen in the far-UV spectra for this proteinase.

The near-UV spectra of Mucor pusillus proteinase (Figure 26) indicated that the phenylalanine fine structure was characterized by

bands at 259 nm and 265 nm, the 0+800 cm^{-1} band of tyrosine at 274 nm and the 0+850 cm^{-1} ^1L_b band and the 0-0 cm^{-1} ^1L_b band of tryptophan at 287 nm and 293 nm, respectively. Some evidence of the 0-0 cm^{-1} band of tyrosine was seen at 277 nm, although the band was very weak. As with the near-UV spectra of M. miehei proteinase, both CD intensity and band shape were virtually unaffected until pH 8.0. At pH 8.0 a noticeable decrease in CD intensity, especially with regards to tyrosine and tryptophan fine structure, became apparent with the disappearance of the 0-0 cm^{-1} band of tyrosine. The disappearance of this band may be the result of the ionization of the tyrosine side chain. Like M. miehei proteinase, the tertiary structure of M. pusillus proteinase appeared to be relatively resistant to pH change.

The near-UV CD spectra of Endothia parasitica proteinase (Figure 27) was highly resolved. Similar highly resolved CD spectra were also obtained by Jenkins et al. (1976). In the present study tryptophan fine structure predominated the CD spectra with the 0+850 cm^{-1} ^1L_b band located at 288 nm and the 0-0 cm^{-1} ^1L_b band located at 297 nm. In addition to the ^1L_b bands, the ^1L_a band was tentatively identified at 262 nm. The presence of this ^1L_a band may have obscured by the presence of the characteristic double band of phenylalanine which is found in this wavelength region. No change in CD intensity or band position was noted until pH values greater than 6.3 were achieved. At pH 7.0 a general decrease in the intensity of the spectra occurred with no change in band position. At pH 8.0 a completely different spectra from those at the other pH values was observed with the appearance of

two new bands at 280 nm and 285 nm. The results obtained would suggest that unlike the other microbial proteinases (M. miehei proteinase and M. pusillus proteinase), the tertiary structure of Endothia parasitica proteinase was more susceptible to conformational change as pH was increased.

The near-UV CD spectra of Aspergillus saitoi proteinase (Figure 28) showed some similarity to the spectra exhibited by Endothia parasitica proteinase with respect to tryptophan fine structure. Tryptophan fine structure was identified as negative peaks at 287 nm (0+850 cm^{-1} $^{1}L_b$ band) and at 294 nm (0-0 cm^{-1} $^{1}L_b$ band) and a positive peak at 257 nm ($^{1}L_a$ band). The presence of the $^{1}L_a$ band of tryptophan made it difficult to distinguish the phenylalanine fine structure, although tentative identification was made at 267 nm and 272 nm. The 0+800 cm^{-1} band of tyrosine was located at 277 nm. Increasing the pH to above 6.3 resulted in a general loss of CD intensity with the broadening of some bands (e.g. 0+850 cm^{-1} $^{1}L_b$ band of tryptophan) with a consequent loss of resolution. The loss of both intensity and resolution may indicate a change in the tertiary structure.

From the results of the near-UV CD spectra study it was concluded that as the pH was increased, changes in the tertiary structure occurred, the extent and degree of which was proteinase dependent. A similar type of conclusion was made when the far-UV region was examined. The fact that changes in the near-UV CD spectral range were observed for all aspartyl proteinases examined may indicate the importance of the aromatic groups in maintaining the structural stability of

the native protein. Structural stability in the aspartyl proteinases may be a function of hydrophobicity due to the inherent relationship between aromatic amino acids and hydrophobicity (Bigelow, 1967). The importance of hydrophobic forces to the structural stability of the aspartyl proteinases may be reinforced from the results of X-ray crystallography for some of these enzymes. These X-ray results have indicated that a significant portion of the aspartyl proteinases consist of hydrophobic, parallel and anti-parallel β-sheet structures (Tang, 1979; Visser, 1981).

D. BIGELOW AVERAGE HYDROPHOBICITY

Hydrophobic forces may play an important role in the structural stability of a protein and also in the interaction between enzyme and substrate. On the basis of hydrolytic studies using synthetic peptide substrates, Tang (1963) postulated that there was a site on the pepsin molecule that binds the hydrocarbon side-chains of the amino acids of the substrate through hydrophobic bonding which was essential for enzymatic action. Hydrophobic forces may be involved in the enzymatic (or primary) phase of the milk-clotting process. Green and Marshall (1977) working with chemically modified casein micelles found that the neutralization of the negative charge on the micelles by a positively charged additive resulted in an increased affinity of chymosin for the micelles.

The average hydrophobicities ($H\Phi_{AVG}$) for the various proteinases using the method of Bigelow (1967) are presented in Table 7. A comparison of the average hydrophobicities of the non-aspartyl proteinases

Table 7. Bigelow average hydrophobicity values ($H\Phi_{AVG}$) obtained for various proteinases.

Protein	$H\Phi_{AVG}$ cal res^{-1}
Aspartyl proteinases	
chymosin	1120
pepsin	1063
M. miehei proteinase	1109
M. pusillus proteinase	1041
E. parasitica proteinase	955
A. saitoi proteinase	973
penicillopepsin	933
Non-aspartyl proteinases	
papain	1159
∝-chymotrypsin	1030
trypsin	1034

(papain, α-chymotrypsin and trypsin) to those of the aspartyl protein-ases showed no distinct difference as evidenced by the overlap of some values between the two groups. Within the group of aspartyl proteinases examined, the $H\Phi_{AVG}$ values ranged from 933 cal res^{-1} for penicillopep-sin to the highest value of 1120 cal res^{-1} for chymosin. This rela-tively large range is contrary to the findings of Bigelow (1967) who reported that proteins related to each other generally have similar hydrophobicity values; (e.g. for various hemoglobins and myoglobins, the $H\Phi_{AVG}$ values ranged from 1060 to 1180 cal res^{-1}). A narrower range of hydrophobicity values would have been expected for aspartyl proteinases due to the sequence homology displayed by various aspartyl proteinases examined.

The use of average hydrophobicity values as a means for classifi-cation may have some limitation if one assumes that function is deter-mined by tertiary structure. Average hydrophobicity involves the summa-tion of the contribution of all amino acids in the protein (Bigelow, 1967) and therefore would not be expected to give an indication of the hydrophobicity on the surface of the native protein structure.

E. HYDROPHOBICITY USING FLUORESCENT PROBES

In addition to an empirical calculation of Bigelow hydrophobicity, the hydrophobicity of the various proteinases was measured using fluorescent probes. Various fluorescent probes have been used to study aspects of the structure and interactions of protein in relation to protein hydrophobicity. In the present study the fluorescent probes

cis-parinaric acid (CPA) and 1-anilino-8-naphthalene sulfonate (ANS) were used. CPA is a conjugated polyene fatty acid and would bind aliphatic regions on the protein surface (Sklar et al., 1977), whereas ANS being aromatic in nature would bind aromatic regions on the protein surface (Figure 29). The support for the use of cis-parinaric acid as a hydrophobic probe was provided by Townsend (1982) who found a significant correlation (P<0.01) between the polarity of a series of solvents (as described by Snyder (1978)) and the quantum yield of cis-parinaric acid in these solvents (as described by Sklar et al. (1977)). The use of ANS as a hydrophobic probe has been well documented in the literature (Stryer, 1965; Weber and Daniel, 1966; Stryer, 1968) where it has been demonstrated that the quantum yield of fluorescence and the wavelength of maximum emission of ANS bound to protein is dependent on the polarity of the binding region.

The results using the two fluorescent probes (CPA and ANS) for the hydrophobicity determination of the various proteinases are presented in Table 8. No distinct difference was observed between aspartyl proteinases and non-aspartyl proteinases as indicated by the overlap of the hydrophobicity values between the two groups using both probes. These results were therefore similar to the results obtained using the Bigelow average hydrophobicity calculation. Of the aspartyl proteinases, chymosin showed high values for both CPA and ANS in comparison to the other proteinases, which would indicate that chymosin had a relatively aliphatic and aromatic surface area. Endothia parasitica proteinase, on the other hand, had a high CPA value but a low ANS value, indicating a

CH$_3$ cis trans trans cis (CH$_2$)$_7$COO$^-$
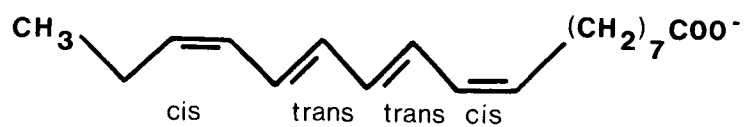
cis-parinaric acid (CPA)
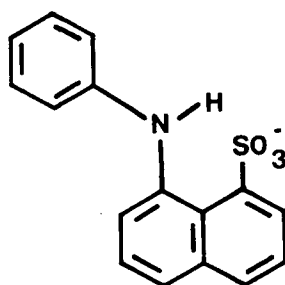
1-anilino-8-naphthalene sulfonate (ANS)

Figure 29. Structures of cis-parinaric acid and 1-anilino-8-naphthalene sulfonate.

Table 8. Hydrophobicity values obtained for various proteinases using fluorescent probes.

Protein	cis-parinaric acid (CPA)	1-anilino-8- naphthalene sulfonate (ANS)
Aspartyl proteinases		
chymosin	96.0[1]	48.0[1]
pepsin	6.0	1.0
M. miehei proteinase	21.0	2.0
M. pusillus proteinase	3.0	7.0
E. parasitica proteinase	113.0	7.0
A. saitoi proteinase	73.0	6.0
penicillopepsin	23.0	3.0
Non-aspartyl proteinases		
papain	19.0	12.0
α-chymotrypsin	9.0	5.0
trypsin	6.0	12.0

[1]Values represent the mean of duplicate determinations.

surface area which was relatively aliphatic in nature with little aroma-tic characteristic. Pepsin had low values for both CPA and ANS indi-cating that the surface area of pepsin was neither very aliphatic nor aromatic in nature. These low values may be related to the fact that pepsin contains a large amount of charged groups, especially aspartic acid, which would restrict the binding of hydrophobic probes. The non-aspartyl proteinases (papain, α-chymotrypsin and trypsin) generally had low CPA and ANS values which may also indicate a highly charged surface.

The surface hydrophobicity characteristics of the aspartyl proteinases were quite varied and were likely a function of the folding of the protein molecule. The folding of the protein into the tertiary structure will ultimately decide the surface characteristics of the protein by determining which residues in the primary amino acid sequence will be exposed and which residues will be buried. Further elucidation of the exact nature of the surface characteristics of the proteinases may require high resolution X-ray diffraction analysis. The similari-ties in secondary structures as previously demonstrated for the aspartyl proteinases were not shown to be reflected in surface hydrophobicity characteristics.

F. CHARGE RATIOS

In further attempts to distinguish the aspartyl proteinases from the non-aspartyl proteinases, various charge ratios were calculated on the basis of amino acid composition data (Table 9). No distinct differ-ences between aspartyl and non-aspartyl proteinases were seen when the

Table 9. Charge ratios of various proteinases.

Protein	$\dfrac{\text{Total charge}[1]}{\text{Total AA}[2]}$	$\dfrac{\text{Acidic}[3]}{\text{Total AA}}$	$\dfrac{\text{Basic}[4]}{\text{Total AA}}$	$\dfrac{\text{Acidic}}{\text{Basic}}$
Aspartyl proteinases				
chymosin	0.156	0.096	0.062	1.55
pepsin	0.141	0.128	0.012	10.67
M. miehei proteinase	0.157	0.106	0.052	2.04
M. pusillus proteinase	0.292	0.228	0.064	3.56
E. parasitica proteinase	0.167	0.118	0.049	2.41
A. saitoi proteinase	0.249	0.197	0.052	3.80
penicillopepsin	0.108	0.084	0.025	3.39
Non-aspartyl proteinases				
papain	0.184	0.071	0.113	0.63
∝-chymotrypsin	0.139	0.057	0.082	0.70
trypsin	0.265	0.161	0.085	1.89

[1] Total charge = Acidic + Basic residues.

[2] Total AA = Total number of amino acids.

[3] Acidic = Asp, Glu residues.

[4] Basic = Lys, His, Arg residues.

ratios involving total charged groups and acidic groups were examined, however, some differences were noted when the ratio involving basic amino acids and the ratio of acidic to basic amino acids were examined. The proportion of basic amino acids was relatively low for the aspartyl proteinases in comparison to the non-aspartyl proteinases. On the other hand, the proportion of acidic amino acids in the aspartyl proteinases was generally much higher than the proportion of the basic amino acids, resulting in ratios greater than one.

Examination of the acidic to basic ratio indicated that pepsin had a substantially higher ratio than the other aspartyl proteinases. This excess of acidic to basic groups may partially explain the conformational change seen in the CD spectra for pepsin at pH values greater than 6.3. Lowenstein (1974), working with chemically modified pepsin, postulated that at pH values greater than 6.0 an electrostatic expansion of the negatively charged polypeptide may occur which could lead to denaturation. Chymosin, Mucor miehei proteinase, Mucor pusillus proteinase and Endothia parasitica proteinase which had low acidic to basic ratios were relatively stable to conformational change with increasing pH as detected by CD spectral analysis (both far-UV and near-UV). However, the relationship between a low acidic to basic amino acid ratio and conformation stability could not be extended to penicillopepsin and Aspergillus saitoi proteinase. The ability to link charge ratios to a functional property such as conformational stability may suffer from the same type of limitation that was experienced with the use of average hydrophobicity as a means for classification. Both parameters (average

hydrophobicity and charge ratio) describe the total protein regardless of the location of specific residues in the topography of the native protein, which is undoubtedly important in determining the function of the protein. In addition, the charge ratio of a proteinase may not reflect the charge density at a specific pH since charge density is affected by the various pK values of the acidic and basic side chains.

G. ZETA POTENTIAL

An indication of a protein's net charge may be obtained from the zeta potential or the electrical potential at the surface of a particle coated with protein. The zeta potential of the various proteinases was measured over the pH range of 5.0 to 8.0 and the results are presented in Figure 30.

The net charge of a protein at a specific pH is dependent on the degree of ionization of the amino acid side chains. Negative charges result from carboxyl groups whereas ε-amino groups, imidazole, indole and guanidyl groups contribute to the positive charges. Sulfhydryl and phenolic groups may also contribute to negative charges at high pH values (Townsend, 1982).

The aspartyl proteinases studied showed a net negative charge over the pH range examined, with the degree of negativity increasing with increasing pH. This increase in negativity can be attributed to the dissociation of the carboxylic side chains coupled with the deprotonation of the amino groups. The net negative charge observed for the proteinases would be expected since the aspartyl proteinases have
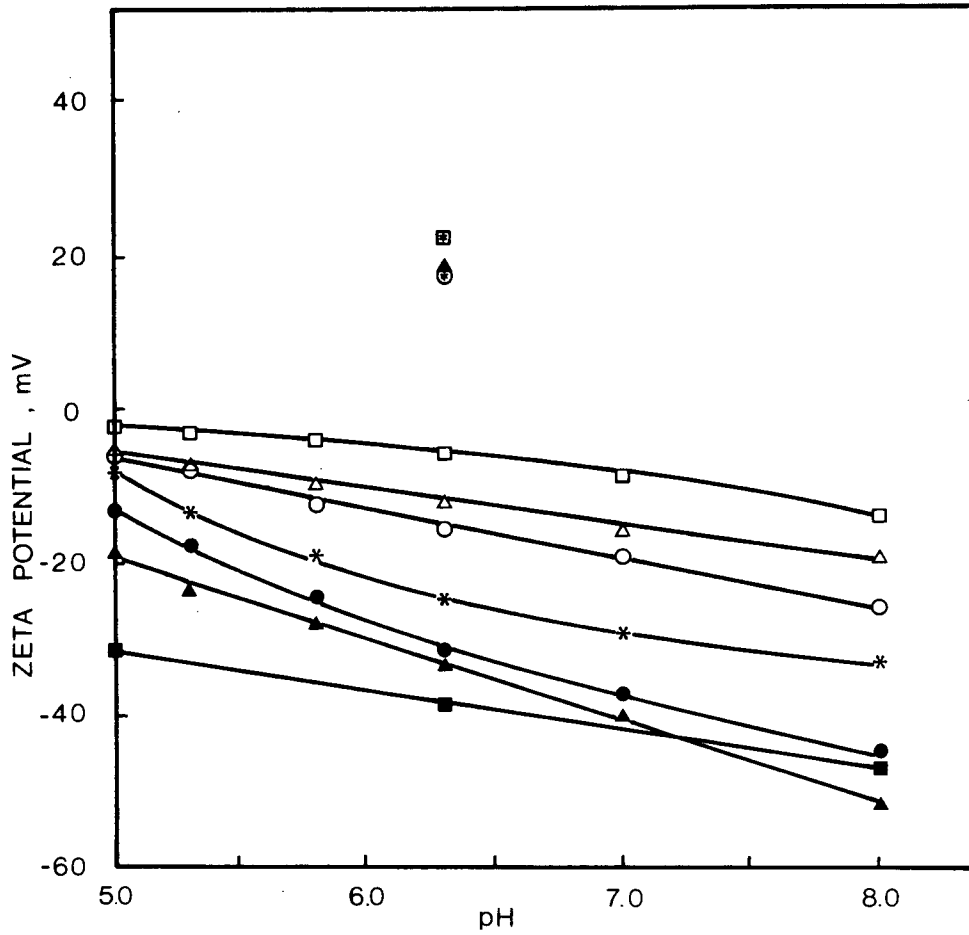
Figure 30. The effect of pH on the zeta potential of various protein-
ases. Chymosin △ , pepsin ▲ , M. miehei proteinase ● , M.
pusillus proteinase ○ , E. parasitica proteinase ∗ , A. sai-
toi proteinase □ , penicillopepsin ■ , papain ⊞ , α-chymo-
trypsin ▲ and trypsin ⊗ .

isoelectric points below pH 5.0. Attempts were made to extrapolate the zeta potential curves of the aspartyl proteinases using linear regression analysis in order to determine the isoelectric points. Results were generally unsuccessful when compared to literature values. Abramson et al. (1964) demonstrated that extrapolation of the zeta potential versus pH curve to zero zeta potential may yield erroneous isoelectric points because of a curvilinear relationship between zeta potential and pH.

With the exception of Aspergillus saitoi proteinase, the other aspartyl proteinases had lower (i.e. more negative) zeta potentials than did chymosin over the pH range. The large negative zeta potential values obtained for pepsin are undoubtedly related to the low isoelectric point of this proteinase. Fruton (1970) reported that the isoelectric point of porcine pepsin is below 1.0 since even at this pH the molecule is still migrating as an anion. The high negative charge of pepsin is also consistent with the calculated charge ratio of acidic to basic groups (Table 9) since pepsin has a definite predominance of acidic groups over basic groups. Curves of higher zeta potential values (i.e. less negative) might be expected for the other aspartyl proteinases which have lower acidic to basic amino acid ratios, and would likely have higher isoelectric points as compared to pepsin. Higher isoelectric points have been reported in the literature for the other aspartyl proteinases (e.g. Mucor miehei proteinase with an isoelectric point of pI = 4.2 (Ottesen and Rickert, 1970b)).

Chymosin which had a nearly equal proportion of acidic to basic groups (Table 9) also had high (i.e. less negative) zeta potential

values. The zeta potential curves of _Mucor miehei_ proteinase, _Mucor pusillus_ proteinase and _Endothia parasitica_ proteinase were intermediate to chymosin and pepsin as might be expected if one assumes that charge ratio provides an indication of zeta potential.

Aspergillus saitoi proteinase which had a greater proportion of acidic to basic groups (Table 9), had higher zeta potential values (i.e. less negative) than those of chymosin. The zeta potential values of penicillopepsin on the other hand were lower than expected based on charge ratios. The availability of certain acidic or basic groups may be affected by unfolding of the protein molecule during zeta potential measurement. Proteins that are adsorbed onto a carrier surface (e.g. 3,3'-dimethylbiphenyl) may not have the zeta potential values expected for the same proteins dissolved in the absence of a carrier. Abramson et al. (1964) stated that changes in protein conformation may occur upon adsorption to a carrier thus affecting the exposure of charged groups.

H. ACCESSIBLE SURFACE AREA

Lee and Richards (1971) introduced the term "accessible surface area" to describe the proportion of the protein surface which could form contacts with water. Since the native structure of a protein exists only in the presence of water (Bernal et al., 1938), the determination of accessible surface area may yield information pertinent to the theories of protein structure and protein-protein interaction (Kauzmann, 1959).

Janin (1976), on the basis of the work by Chothia (1975), formulated an equation (Eq. 23) which relates the molecular weight of a

protein to its accessible surface area (A_s). The accessible surface areas obtained for the various proteinases are presented in Table 10. The aspartyl proteinases were found to have similar accessible surface areas in the range of 11000 to 13000 $Å^2$, whereas the non-aspartyl proteinases had A_s values in the range of 9000 to 10000 $Å^2$. These results would be expected due to the nature of the relationship between accessible surface area and molecular weight, and the fact that the aspartyl proteinases are characterized by having relatively similar molecular weights (Table 10). As a class, the aspartyl proteinases also have similar tertiary structures (Jenkins et al., 1976), thus it follows that molecules with similar size and shape would have similar A_s values.

Accessible surface area may help to explain protein function. However, the non-aspartyl proteinases which are also known to exhibit milk-clotting ability (Ernstrom, 1974), were found to have lower A_s values than the aspartyl proteinases (milk-clotting enzymes). It is therefore apparent that other parameters in addition to A_s must be contributing to enzymatic activity (e.g. milk-clotting).

I. MILK-CLOTTING TO PROTEOLYTIC ACTIVITY RATIO

The milk-clotting to proteolytic activity ratio has been used in attempts to identify proteinases which are suitable as milk-clotting agents in the cheese-making process (Puhan and Irvine, 1973; de Koning et al., 1978). Proteinases which are successfully used in cheese-making have a high milk-clotting to proteolytic activity ratio (Visser, 1981).

Table 10. Accessible surface areas (A_s) of various proteinases.

Protein	M^1	A_s (Å^2)
Aspartyl proteinases		
chymosin	30700	11300
pepsin	35000	12300
M. miehei proteinase	38000	13000
M. pusillus proteinase	30600	11200
E. parasitica proteinase	37500	12900
A. saitoi proteinase	34500	12200
penicillopepsin	32000	11600
Non-aspartyl proteinases		
papain	20900	8700
α-chymotrypsin	21600	8900
trypsin	24000	9600

^{1}M = Molecular weight.

Both milk-clotting and proteolytic activities are calculated on the basis of proteinase weight, but due to impurities in some of the proteinase preparations, it was felt that the use of the milk-clotting to proteolytic activity ratio would reduce the effects of non-proteolytic components within the preparations. The milk-clotting to proteolytic activity ratios obtained for the various proteinases are presented in Table 11. Of the enzymes examined, chymosin exhibited the highest milk-clotting to proteolytic activity ratios over the pH range. Ernstrom (1974) also reported that chymosin had the highest milk-clotting to proteolytic activity ratio among proteolytic enzymes indicating that chymosin had the highest specific milk-clotting activity. The microbial proteinases from _Mucor_ _miehei_, _Mucor_ _pusillus_ and _Endothia_ _parasitica_ generally showed ratios that were intermediate to those of chymosin and porcine pepsin. _Aspergillus_ _saitoi_ proteinase, penicillopepsin and the non-aspartyl proteinases (papain, α-chymotrypsin and trypsin) all showed low ratios.

According to Huang and Dooley (1976), the majority of cheese manufacturing operations in the United States utilize the proteinases from _Mucor_ _miehei_, _Mucor_ _pusillus_ and _Endothia_ _parasitica_; a similar situation may exist in Canada. Arima et al. (1970), in a comparison of the milk-clotting to proteolytic activity ratios of various proteinases, found that the ratio obtained for _Endothia_ _parasitica_ proteinase was lower than that of _Mucor_ _pusillus_ var. Lindt proteinase, which in turn was lower than that obtained for chymosin. The authors observed that low ratios were obtained for proteinases such as trypsin, papain and

Table 11. Milk-clotting to proteolytic activity ratios for various proteinases.

Protein	Milk-clotting to proteolytic activity ratio[1] $(\times 10^{-2})$					
	pH					
	5.0	5.3	5.8	6.3	7.0	8.0
Aspartyl proteinases						
chymosin	159.61	132.81	97.06	95.75	13.64	N.D.[2]
pepsin	74.43	57.89	53.12	47.43	N.D.	N.D.
M. miehei proteinase	114.31	105.96	97.58	71.05	6.93	N.D.
M. pusillus proteinase	127.30	90.01	86.04	63.37	8.32	N.D.
E. parasitica proteinase	82.11	62.52	62.49	60.36	N.D.	N.D.
A. saitoi proteinase	1.45	0.93	0.58	N.D.	N.D.	N.D.
penicillopepsin	22.95	–[3]	–	0.75	–	N.D.
Non-aspartyl proteinases						
papain	–	–	–	2.17	–	–
⍺-chymotrypsin	–	–	–	3.07	–	–
trypsin	–	–	–	0.02	–	–

[1]Ratios were calculated using the mean values of duplicate determinations for both milk-clotting and proteolytic activity.

[2]N.D. = Not detected.

[3]Not determined.

Aspergillus saitoi proteinase and were generally the result of low milk-clotting ability to high general proteolytic activity.

Examination of the aspartyl proteinases showed that as pH was increased from 5.0 to 8.0, the milk-clotting to proteolytic activity ratio decreased indicating loss of enzymatic activity. The decrease in enzymatic activity for the aspartyl proteinases could be partially explained by a change in the far-UV CD spectra as reflected by changes in the protein secondary structure fractions. At pH 8.0 no enzymatic activity was detected for any of the aspartyl proteinases, which corresponded to changes in the far-UV CD spectra noted at this pH. However, not all of the decrease in enzymatic activity could be attributed to changes in secondary structure since virtually no change in secondary structure was seen at pH values of 6.3 or lower for most of the aspartyl proteinases (except chymosin). It was interesting to note that the proteinases (pepsin, Aspergillus saitoi proteinase and penicillopepsin) which showed relatively large changes in secondary structure at pH values greater than 6.3 (Table 6), showed greater losses of enzymatic activity (Table 11) than those aspartyl proteinases which showed only small changes in secondary structure.

The relationship between the secondary structure of a protein and its function has been documented in the literature (Ptitsyn and Finkelstein, 1983). It is undoubtedly the folding of the secondary structure fractions into the tertiary structure that governs the intimate contact between enzyme and substrate which eventually results in catalysis. From the present study it is apparent that changes in

secondary structure may affect enzymatic activity. A loss of enzymatic activity was not always associated with a change in secondary structure in the present study, however, subtle changes in the secondary structure may have occurred, but were not detected by the method of analysis.

The relationship seen between the far-UV CD spectra and the milk-clotting to proteolytic activity ratio was similar to that observed between the near-UV CD spectra and the milk-clotting to proteolytic activity ratio. The largest reductions in enzymatic activity for the various aspartyl proteinases generally occurred at pH values greater than 6.3. It was also at these pH values that changes in the near-UV CD fine structure were noted, the extent of these changes being proteinase dependent. Losses in enzymatic activity in the pH range of 5.0 to 6.3 were not detected when the near-UV CD spectra of the aspartyl protein-ases were examined. Minor changes in the tertiary structure may have occurred in the pH range of 5.0 to 6.3, but were not detected.

Cheeseman (1969) examined the loss of milk-clotting activity by measuring changes in the UV absorption spectra obtained for chymosin treated with urea and subjected to different pH values. Results indi-cated that loss of activity occurred with changes in absorption spectra and that the rate of change increased as both urea concentration and pH value increased. However, it was also found that 50 percent inactiva-tion was obtained with changes varying from 27 to 94 percent of the total spectral change, which may suggest that there is only a partial correlation between the degree of unfolding and the loss of activity. Factors other than the availability of the chromophore groups as a

result of the unfolding of the molecule must influence the loss of enzyme activity.

Kay and Valler (1981) noted that the aspartyl proteinases were susceptible to alkaline denaturation due to high contents of acidic amino acid residues. Ma (1979) studied the chemical modification of carboxyl groups in porcine pepsin and stated that decreases in enzymatic activity were due either to conformational changes in the protein molecule or to changes in the charge distribution; both changes could effect the binding between enzyme and substrate. In the present study, pepsin which had a high predominance of acidic to basic residues (Table 9) showed no detectable enzymatic activity at pH values greater than 6.3. Chymosin, on the other hand, which had a lower ratio of acidic to basic residues (Table 9) showed substantially higher enzymatic activity at pH values greater than 6.3 indicating greater enzymatic stability. Mucor miehei proteinase, Mucor pusillus proteinase and Endothia parasitica proteinase which had charge ratios intermediate to chymosin and pepsin showed milk-clotting to proteolytic activity ratios which were intermediate to chymosin and pepsin. However, penicillopepsin and Aspergillus saitoi proteinase which also had intermediate acidic to basic amino acid ratios showed no detectable enzymatic activity at pH values greater than 6.3, therefore, charge ratios cannot totally explain enzymatic stability to alkaline conditions. In addition to the charge ratio, the specific arrangement of the charged groups in the protein's tertiary structure must be important for determining enzymatic stability. Futhermore, it is undoubtedly the charge and conformation of the residues in the active

site which have a major influence on the enzyme-substrate interaction and subsequent catalysis.

Most work concerning the active site region of the aspartyl proteinases has been carried out on pepsin. If it is assumed that all aspartyl proteinases react in a similar manner, then some of the hypotheses proposed for the pH dependency of pepsin catalyzed reactions may be generalized to the aspartyl proteinases as a group. Clement (1973) reported that the pK_a values of the catalytically essential aspartic acid residues of pepsin, Asp_{32} and Asp_{215}, were 1.5 and 4.5, respectively. The difference in pK_a values was attributed to the existence of a hydrogen bond between the two carboxylic acid groups. Antonov (1976), also working with pepsin, proposed that the decrease in catalytic activity with an increase in pH may be due to the loss of a proton from the hydrogen bonded essential aspartic acid groups. Once this proton has been eliminated due to the dissociation of the carboxylic acid group of Asp_{215}, the orientation of the carboxyl groups may change. Hsu et al. (1977) discussed the possibility of this type of mechanism for the active site aspartic acid residues in penicillopepsin. A similar situation could occur with other aspartyl proteinases leading to the loss of catalytic activity with increasing pH.

Raymond and Bricas (1979) were able to determine the pK values of the groups essential for the reaction between chymosin and synthetic peptide substrates; pK values of 3.3 and 5.7 were obtained on the basis of a plot of log k_{cat}/K_m vs pH. The authors postulated that the catalytic activity of chymosin was dependent on the ionization of two

carboxyl groups, similar to the reaction mechanism proposed for pepsin (Fruton, 1970). Visser et al. (1980) studied the hydrolysis of tryptic fragments of bovine κ-casein by chymosin and found that the apparent pK_1 and pK_2 values of catalytically important groups on the enzyme-substrate complexes were in the region of 4.0 to 4.2 and 6.5 to 6.7, respectively. It was assumed that these pK values represent the pK values of the active site aspartic acid residues although the values for the free enzyme may be different due to substrate binding.

The differences in loss of enzymatic activity as a function of pH exhibited by the various aspartyl proteinases examined in the present study may be partially due to the differences in the pK_a values of the active site aspartic acid residues. The pK_a values of the active site aspartic acid residues may be affected by the residues adjacent to them in the active site, and by the characteristics of the substrate. Kitson and Knowles (1971) observed that by chemically modifying arginine residues (by phenylglyoxal) in pepsin and thereby changing the microenvironment, it was possible to change the pK_a value of one of the active site aspartic acid residues (Asp_{32}).

Raap et al. (1983) examined various peptide substrates for chymosin and stated that in order to better understand the specificity and the kinetic mechanism of enzyme action both the conformation of the active site of the enzyme and the conformation of the substrate molecule are important. It is possible that the decrease in enzymatic activity with increasing pH observed in the present study for the various aspartyl proteinases may have also been due to changes in the substrate(s). No attempts were made to measure any substrate conformation.

J. PRINCIPAL COMPONENT ANALYSIS

Data collected for the various proteinases examined in the present study was subjected to principal component analysis (PCA) in attempts to identify parameters important for the classification of these protein-ases. The sixteen variables used in the PCA were previously described (see Materials and Methods). Milk-clotting to proteolytic activity ratio was excluded in this analysis.

When the sixteen original variables for each of the proteinase samples were entered in the principal component analysis program, four factors were returned which met the criteria of their eigenvalues exceeding 1.0 (Table 12). Aishima (1979a; 1979b; 1979c) used the same criteria for selecting principal components. The four factors or prin-cipal components obtained accounted for more than 85 percent of the total variance (Table 12), where the total variance is the sum of the individual variances for each of the original variables. Each factor represents a combination of the original variables. The factor loadings for each factor give an indication of the importance of the original variables to that factor. The factor loadings for the four factors are presented in Table 13.

Examination of the factor loadings indicated that factor 1 was primarily concerned with circular dichroism spectral data. Variables contributing to factor 2 included molar ellipticity values at wave-lengths 193, 198, 200 and 202 nm. It is interesting to note that these wavelengths corresponded to a characteristic wavelength for each of the major secondary structure fractions. The CD band at 193 nm is one of the bands characteristic of α-helix, the 198 nm band characteristic of

Table 12. The variance explained and the cumulative proportion of total variance accounted for by each factor derived from the principal component analysis.

Factor	Variance explained (eigenvalue)	Cumulative proportion of total variance
1	8.113	0.5071
2	2.793	0.6816
3	1.867	0.7984
4	1.280	0.8783
5	0.749	0.9251
6	0.562	0.9603
7	0.380	0.9840
8	0.145	0.9931
9	0.061	0.9969
10	0.028	0.9987
11	0.011	0.9994
12	0.006	0.9997
13	0.002	0.9998
14	0.001	0.9999
15	0.001	0.9999
16	0.000	1.0000

Table 13. Factor loadings for the factors whose eigenvalues exceed 1.0.

Original variable	Factor loadings			
	Factor 1	Factor 2	Factor 3	Factor 4
213 nm	0.978	0.000	0.000	0.000
202 nm	0.974	0.000	0.000	0.000
210 nm	0.965	0.000	0.000	0.000
224 nm	0.954	0.000	0.000	0.000
225 nm	0.945	0.000	0.000	0.000
190 nm	-0.905	0.000	0.000	0.000
193 nm	-0.879	0.359	0.000	0.000
$H\Phi_{AVG}$	-0.719	0.000	0.000	0.394
ANS	-0.626	0.000	-0.495	-0.259
198 nm	0.000	0.971	0.000	0.000
200 nm	0.000	0.967	0.000	0.000
202 nm	0.537	0.811	0.000	0.000
Acidic/Basic	0.000	0.000	0.870	0.000
Zeta potential	0.000	0.000	-0.692	0.000
A_s	0.000	0.000	0.000	0.794
CPA	0.000	0.000	-0.470	-0.780

β-sheet, the 200 nm band characteristic of the unordered fraction and the 202 nm band is characteristic of the β-turn fraction. Factor 3 was characterized by ANS hydrophobicity, acidic to basic amino acid ratio, zeta potential and CPA hydrophobicity. Variables contributing to factor 4 included Bigelow average hydrophobicity, ANS hydrophobicity, accessible surface area, as well as CPA hydrophobicity.

Two-dimensional plots involving the first three factors are presented in Figures 31 and 32. Alphabetic codes used to identify the proteinases are given in the figure legends. The plot of factor 2 versus factor 1 (Figure 31) indicated that the proteinases from microbial sources (i.e. M. miehei, M. pusillus, E. parasitica and A. saitoi proteinases and penicillopepsin) were closely associated with one another, while pepsin at pH values 5.0 to 6.3 formed one group, and chymosin at pH values 5.8, 6.3 and 7.0 formed another group. The non-aspartyl proteinases (papain, α-chymotrypsin and trypsin) were not associated with any group. Chymosin at pH 5.0, 5.3 and 8.0, as well as pepsin at pH 7.0 and 8.0, were not associated with their respective groups which could be explained by noting that these samples showed different CD spectral patterns than those from the previously mentioned pH values. Since factor 1 and factor 2 were almost exclusively concerned with CD spectral data, any change in spectral characteristics of a sample within a group may result in the exclusion of that sample from that group.

The far-UV spectra of the aspartyl proteinases examined generally changed at pH values greater than 6.3. This change in CD spectra was
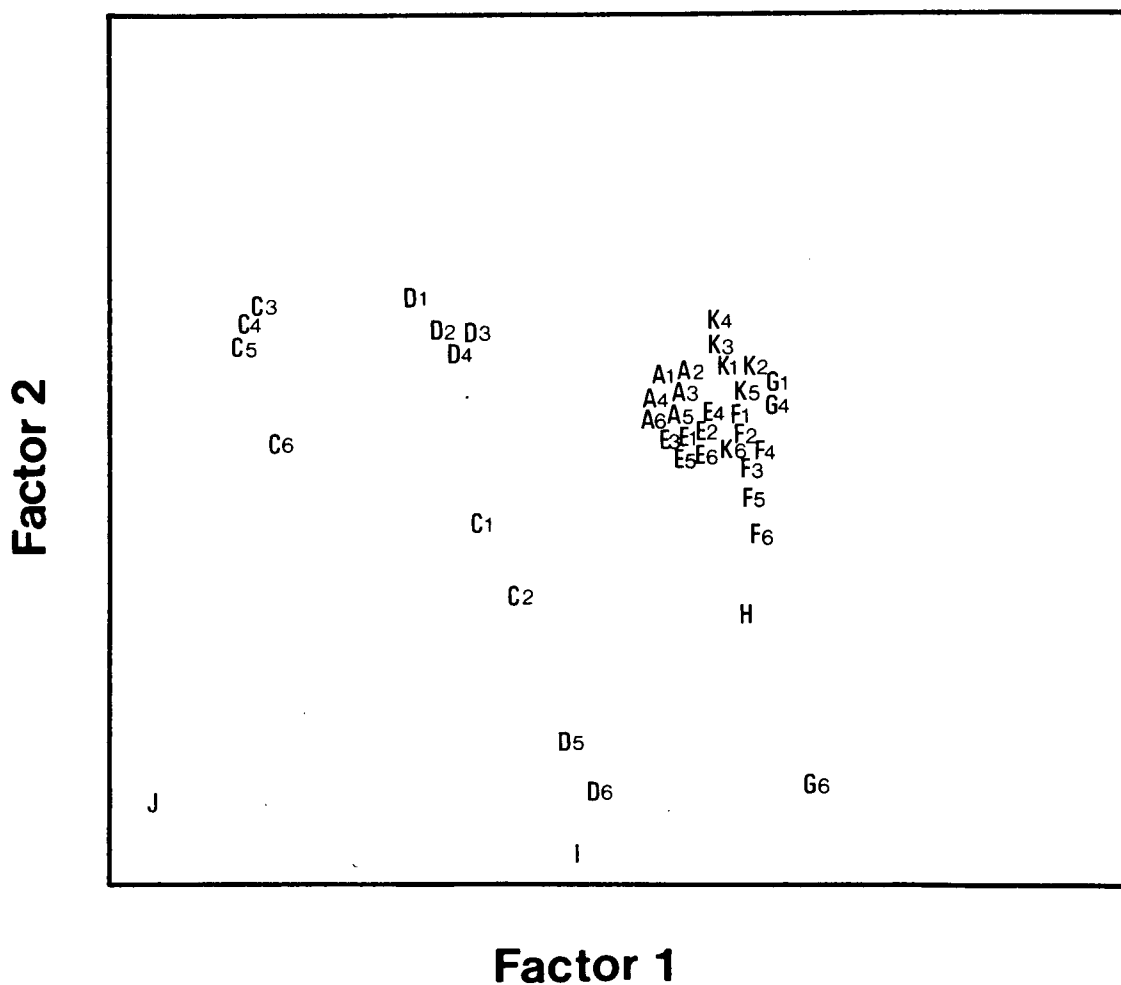
Figure 31. The plot of factor 2 vs factor 1 obtained from the principal component analysis of the various structural and intrinsic properties of the proteinases. A - Mucor miehei proteinase, K - Endothia parasitica proteinase, C - chymosin, D - pepsin, E - Mucor pusillus proteinase, F - Aspergillus saitoi proteinase, G - penicillopepsin, H - trypsin, I - α-chymotrypsin, and J - papain. The numbers 1,2,3,4,5,6 represent pH 5.0, 5.3, 5.8, 6.3, 7.0, 8.0 respectively.
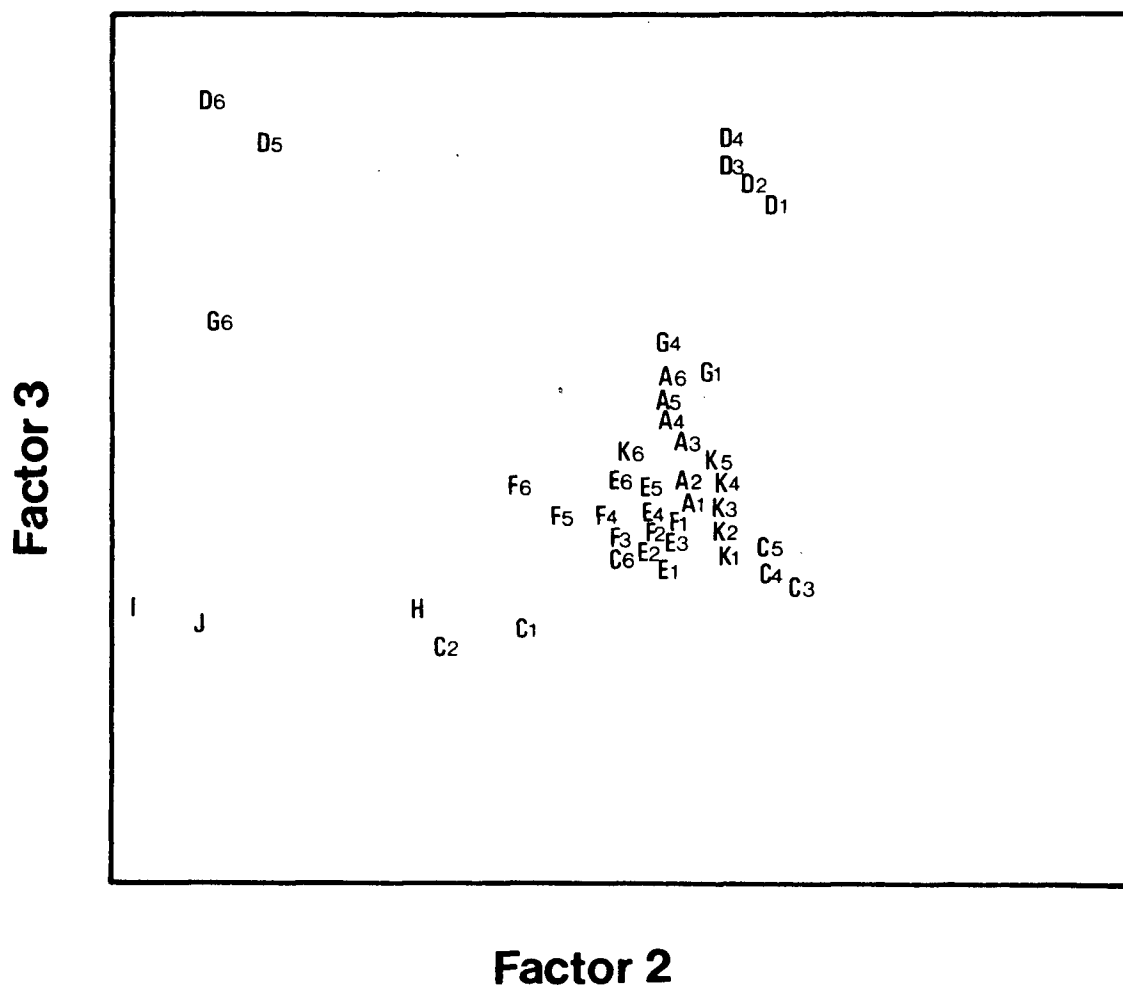
Figure 32. The plot of factor 3 vs factor 2 obtained from the principal component analysis of the various structural and intrinsic properties of the proteinases. A - <u>Mucor miehei</u> proteinase, K - <u>Endothia parasitica</u> proteinase, C - chymosin, D - pepsin, E - <u>Mucor pusillus</u> proteinase, F - <u>Aspergillus saitoi</u> proteinase, G - penicillopepsin, H - trypsin, I - α-chymotrypsin, and J - papain. The numbers 1,2,3,4,5,6 represent pH 5.0, 5.3, 5.8, 6.3, 7.0, 8.0 respectively.

reflected in a downward shift of the points representing these samples in the plot. The degree of downward shift was related to the degree of change in the CD spectra (i.e. the greater the change in the CD spectra, the larger the downward shift). Changes in the vertical direction (Y-axis) of the plot would be controlled by factor 2 while changes in the horizontal direction (X-axis) would be controlled by factor 1. The shifts or changes observed were in the vertical direction which would tend to imply that changes occurring in the short wavelength far-UV CD range were more critical than those seen at the long wavelength far-UV CD range. If one examines the reference spectra for the four secondary structure fractions (α-helix, β-sheet, β-turn and unordered) as proposed by Chang et al. (1978) (Figure 33) it is apparent that the wavelengths comprising factor 2 show the largest differences in molar ellipticity values among the secondary structures. These wavelengths may therefore have some ability to discriminate among the four secondary structures.

Unlike the plot of factor 2 versus factor 1, in which both factors were concerned with CD spectral data, the plot of factor 3 versus factor 2 (Figure 32) was a plot of properties associated with structure (zeta potential, acidic to basic amino acid ratio, ANS and CPA hydrophobicity) versus CD spectral properties (molar ellipticity values at wavelengths 193, 198, 200 and 202 nm). The plot of factor 3 versus factor 2 (Figure 32) again showed that the microbial proteinases were closely associated with one another. Chymosin as a group, however, was much more closely associated with the microbial proteinases than in the plot of factor 2 versus factor 1 (Figure 31). Penicillopepsin at pH 5.0 and pH 6.3 formed a group intermediate to the microbial enzymes and a group formed
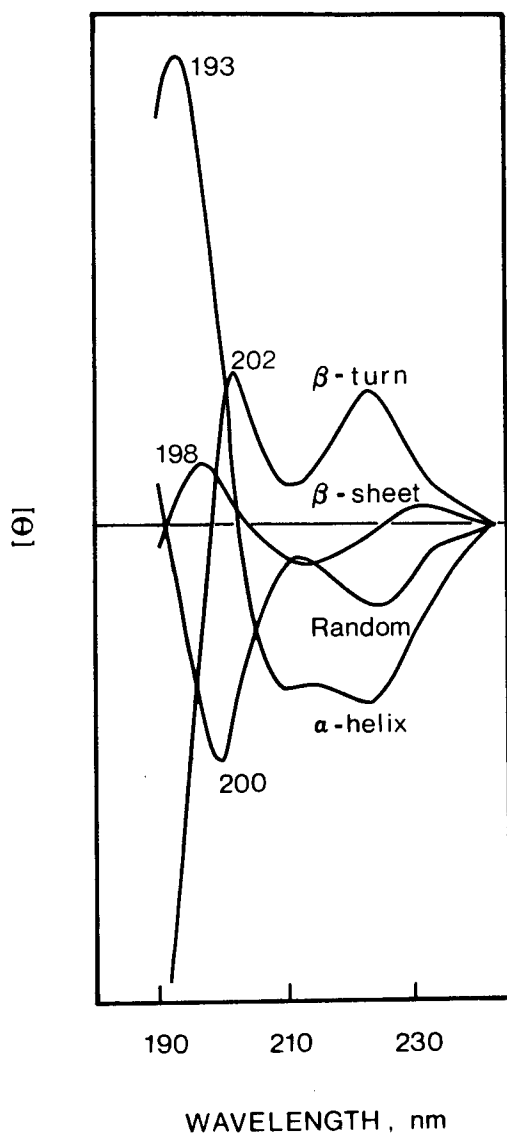
Figure 33. Reference spectra for α-helix, β-sheet, β-turn and random coil fractions. (Adapted from Chang et al., 1978.)

from pepsin at pH 5.0, pH 5.3, pH 5.8 and pH 6.3. Again papain, α-chymotrypsin and trypsin were not associated with any group.

In Figure 32 it was interesting to note that as the pH of the aspartyl proteinases samples was increased, the points representing those samples tended to shift horizontally to the left-hand side of the plot. Since factor 2 represents CD spectral data and is plotted on the abscissa, any change in CD spectral data would be expected to cause changes along that axis. The samples corresponding to the points which had been shifted horizontally also had lower milk-clotting to proteolytic activity ratios than those points on the right-hand side of the plot.

In attempts to identify the variables that contribute to the milk-clotting to proteolytic activity ratio, the milk-clotting to proteolytic activity ratio which acted as the dependent variable was regressed on the principal components that were derived from the sixteen original variables. Regression of the dependent variable on the sixteen original variables using standard multiple linear regression analysis (both forward and backward stepwise) was unsuccessful due largely to the intercorrelation found among the independent variables. Lukovits (1983), in a study on the quantitative structure-activity relationships (QSAR) employing independent quantum chemical indices, also pointed out the problem of intercorrelations of various measured indices used in a QSAR analysis. Due to these intercorrelations interpretation of the regression equations between activity and the chemical indices of molecules was reported to be difficult. The author used principal

component analysis to reduce the number of indices, without loss of information, and also to decompose the indices into mutually independent components which could then be regressed against the dependent variable.

In the present study principal components were entered into the regression model on the basis of their correlation with the dependent variable (Table 14); the component having the largest coefficient was entered first. Multivariate linear regression indicated that three components were necessary to describe the dependent variable (Table 14). The regression model containing the three components was highly significant (P<0.01) and included components four, seven and two, the components being entered in order of correlation to the dependent variable. Component four represented zeta potential, Bigelow average hydrophobicity, acidic to basic amino acid ratio, accessible surface area and CPA hydrophobicity; component seven represented the molar ellipticity value at 190 nm, zeta potential, Bigelow average hydrophobicity, acidic to basic amino acid ratio and ANS hydrophobicity; and component two represented molar ellipticity values at wavelengths 193, 198, 200 and 202 nm. Analysis of the components indicated that the properties and the combination of the properties measured for the proteinases were important in partially describing the milk-clotting to proteolytic activity ratio.

Since a significant relationship (P<0.01) was established between the dependent variable and various combinations of properties determined for the proteinases, this may imply that proteinases with similar combinations have similar enzymatic activities. Although the relationship was significant, some of the variability observed for the dependent

Table 14. Regression analysis for the milk-clotting to proteolytic activity ratio on the principal components computed from various structural and intrinsic properties of the proteinases.

Number of components (p)	Number of cases (N)	Index of components entering	F-value Regression model	R^2
1	42	4[a]	5.58**	.1224
2	42	7[b]	6.06**	.2370
3	42	2[c]	6.83**	.3503

** Significant at $P < 0.01$.

[a]Represents zeta potential, Bigelow average hydrophobicity, acidic/basic amino acid ratio, accessible surface area and CPA hydrophobicity.

[b]Represents the molar ellipticity value at 190 nm, zeta potential, Bigelow average hydrophobicity, acidic/basic amino acid ratio and ANS hydrophobicity.

[c]Represents molar ellipticity values at wavelengths 193, 198, 200 and 202 nm.

variable was not accounted for by the independent variables (principal components) which would indicate that other properties not examined in the present study were critical for describing enzymatic activity. For example, the proteinase from Aspergillus saitoi had measured characteristics generally similar to those of the microbial proteinases examined and was grouped close to these proteinases, but had a very much lower milk-clotting to proteolytic activity ratio.

In attempts to improve the regression model it was decided to exclude the data from Aspergillus saitoi proteinase since this proteinase was the least pure of the enzymes examined. The resulting regression model was also significant (P<0.01) and was described by the same three components (i.e. components, 2, 4 and 7) as previously discussed, however, the order in which the components were entered into the equation was different. The coefficient of determination was increased from .35 to .43 (Table 15).

It may be possible to examine the two-dimensional principal component plots and on the basis of factor loadings and standardized scores, hypothesize how a manipulation of the original variables may result in a protein with similar structural combination characteristics having similar functional properties. For example, in the plot of factor 3 versus factor 2 (Figure 32) no change in factor 2 (CD spectral characteristics) would be necessary to move pepsin at pH values 5.0 to 6.3 closer to chymosin since the respective groups are situated above one another in the plot. However, a change in factor 3 (representing acidic to basic amino acid ratio, zeta potential, ANS and CPA hydrophobicity) would be necessary to move pepsin closer to chymosin in

Table 15. Regression analysis for the milk-clotting to proteolytic activity ratio on the principal components computed for various structural and intrinsic properties of the proteinases in the absence of <u>Aspergillus</u> <u>saitoi</u> proteinase.

Number of components (p)	Number of cases (N)	Index of components entering	F-value Regression model	R^2
1	36	7^a	7.30**	0.1768
2	36	2^b	7.71**	0.3186
3	36	4^c	8.14**	0.4329

** Significant at $P<0.01$.

[a] Represents molar ellipticity value at 190 nm, zeta potential, Bigelow average hydrophobicity, acidic/basic amino acid ratio and ANS hydrophobicity.

[b] Represents molar ellipticity values at wavelengths 193, 198, 200 and 202 nm.

[c] Represents zeta potential, Bigelow average hydrophobicity, acidic/basic amino acid ratio, accessible surface area and CPA hydrophobicity.

Figure 32. Since chymosin had a low acidic to basic amino acid ratio, high zeta potential values and high ANS and CPA hydrophobicity values as compared to pepsin, hypothetically altering those values for pepsin to more closely resemble those of chymosin would undoubtedly move pepsin closer to chymosin. A similar approach could be used for the other proteinases in the plot. Although a relationship between milk-clotting to proteolytic activity ratio and structural and intrinsic parameter combinations has been established, whether or not such changes could result in enzymatic properties similar to those of a desired enzyme (e.g. chymosin) is speculative and must await further experimental evidence (e.g. chemical modification studies).

CONCLUSIONS

The major objectives of the present study were: to examine various structural and intrinsic properties of some aspartyl proteinases as well as some non-aspartyl proteinases; to classify these proteinases using principal component analysis; and to identify parameters important for the milk-clotting to proteolytic activity ratio. Using the diagonal plot method, it was found that pepsin and chymosin had the highest degree of primary sequence homology when compared to the other pairs of aspartyl proteinases. It was also found that the active site regions were highly homologous between the aspartyl proteinases examined. The secondary structure prediction methods of Chou and Fasman (1978b) and Cid et al. (1982) indicated that chymosin, pepsin, penicillopepsin and Mucor miehei proteinase had a high proportion of β-sheet and that the active site aspartic acid residues were located in β-turn regions. Examination of the far-UV CD spectra of the aspartyl proteinases indicated that changes in the spectra occurred in the neutral to alkaline pH range, the extent of change being proteinase dependent. Secondary structure determination from far-UV CD spectral data demonstrated that the aspartyl proteinases had a high proportion of β-sheet. The proportion of β-sheet generally decreased at pH values greater than 6.3. Results obtained from the near-UV CD spectra of the aspartyl proteinases indicated a change in spectra in the neutral to alkaline pH range and may implicate the importance of aromatic groups to tertiary structure stability.

Bigelow average hydrophobicity showed no clear distinction between aspartyl and non-aspartyl proteinases. Of the aspartyl proteinases, chymosin had the highest hydrophobicity value. The determination of hydrophobicity using fluorescent probes (CPA and ANS) again indicated no clear distinction between aspartyl and non-aspartyl proteinases; chymosin showed relatively high values for both probes. Charge ratios indicated that the aspartyl proteinases generally had low proportions of basic amino acids and correspondingly high acidic to basic amino acid ratios as compared to the non-aspartyl proteinases. For all aspartyl proteinases examined, the zeta potential became more negative with increasing pH. Calculation of the accessible surface area indicated that the aspartyl proteinases had similar accessible surface areas that were higher than those of the non-aspartyl proteinases examined.

The milk-clotting to proteolytic activity ratio of the aspartyl proteinases decreased with increasing pH. Changes in the far- and near-UV CD spectra were partially correlated to changes in the milk-clotting to proteolytic activity ratio. Of the aspartyl proteinases, chymosin showed the highest milk-clotting to proteolytic activity ratio, whereas penicillopepsin, Aspergillus saitoi proteinase and the non-aspartyl proteinases had low ratios.

Principal component analysis of various structural and intrinsic properties of aspartyl and non-aspartyl proteinases indicated that aspartyl proteinases formed distinct groups; the non-aspartyl proteinases were not associated with any of these groups. Regression of the milk-clotting to proteolytic activity ratio on the principal components

indicated that three components were necessary to describe the dependent variable. By examining the parameters that contribute to the principal components it is possible to conclude that a high milk-clotting to proteolytic activity ratio is partially dependent on relatively high hydrophobicity, β-sheet structure and relatively low charge.

REFERENCES CITED

Abramson, H. A., Moyer, L. S. and Gorin, M. H. 1964. Electrophoresis of Proteins. Hafner Publishing Co., Inc., New York, NY.

Ahmad, F. and McPhie, P. 1978. The denaturation of covalently inhibited swine pepsin. Int. J. Peptide Protein Res. 12: 155.

Aishima, T. 1979a. Evaluation and discrimination of soy sauce by computer analysis of volatile profiles. Agric. Biol. Chem. 43:1711.

Aishima, T. 1979b. Classification of soy sauce on principal components in gas chromatography profiles. Agric. Biol. Chem. 43: 1905.

Aishima, T. 1979c. Objective evaluation of soy sauce by statistical analysis of G.C. profiles. Agric. Biol. Chem. 43: 1935.

Aishima, T., Nagasawa, M. and Fukushima, D. 1979. Differentiation of the aroma quality of soy sauce by statistical evaluation of gas chromatographic profiles. J. Food Sci. 44: 1723.

Alais, C. and Lagrange, A. 1972. Etude biochimique d'une protease coagulante produite par Mucor miehei. 1. Activite coagulante et activite proteolytique. Le Lait 52: 407.

Andreeva, N. S., Gustchina, A. E., Fedorov, A. A., Shutzkever, N. E. and Volnova, T. V. 1976. X-ray crystallographic studies of pepsin. In: Acid Proteases: Structure, Function and Biology. Tang, J. (Ed.). p. 23. Adv. Exp. Med. Biol. Vol. 95. Plenum Press, New York, NY.

Andren, A. and de Koning, P. J. 1982. Changes in immunologically and enzymatically active centres of some milk clotting enzymes. In: Uses of Enzymes in Food Technology. Dupuy, P. (Ed.). p. 275. Technique et Documentation, Lavoisier, Paris.

Anfinsen, C. B. 1973. Principles that govern the folding of protein chains. Science 181: 223.

Anfinsen, C. B., Haber, E., Sela, M. and White, F. H. Jr. 1961. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. Proc. Natl. Acad. Sci. U.S.A. 47: 1309.

Antonov, V. K. 1976. New data on pepsin mechanism and specificity. In: Acid Proteases: Structure, Function and Biology. Tang, J. (Ed.). p. 179. Adv. Exp. Med. Biol. Vol. 95. Plenum Press, New York, NY.

Arima, K., Iwasaki, S. and Tamura, G. 1967. Milk clotting enzyme from microorganisms. Part I. Screening test and the identification of the potent fungus. Agric. Biol. Chem. 31: 540.

Arima, K., Yu, J., and Iwasaki, S. 1970. Milk-clotting enzyme from Mucor pusillus var. Lindt. In: Methods in Enzymology. Vol. XIX. Perlmann, G. E. and Lorand. L. (Eds.). p. 446. Academic Press, New York, NY.

Arnon, R. 1970. Papain. In: Methods in Enzymology. Vol. XIX. Perlmann, G.E. and Lorand, L. (Eds.). p. 226. Academic Press, New York, NY.

Baker, C. C. and Isenberg, I. 1976. On the analysis of circular dichroic spectra of proteins. Biochemistry 15: 629.

Baldwin, R. L. and Wake, R. G. 1959. Boundary spreading in the ultracentrifuge of an associating system (rennin). Abstr. Am. Chem. Soc. 136: 35c.

Bamford, C. H., Elliott, A. and Hanby, W. E. 1956. Synthetic Polypeptides. Academic Press, New York, NY.

Bang-Jensen, V., Foltmann, B. and Rombauts, W. 1964. On the proteolytic specificity of rennin. Compt. Rend. Trav. Lab. Carlsberg 34: 326.

Bannister, W. H. and Bannister J. V. 1974. A study of three-component fitting of protein circular dichroism spectra. Int. J. Biochem. 5: 679.

Barela, T. D. and Darnall, D. W. 1974. Practical aspects of calculating protein secondary structure from circular dichroism spectra. Biochemistry 13: 1694.

Barry, C. D., Friedman, J., Chou, P. Y., Fasman, G. D., Finkelstein, A. V., Lim, V. I., Ptitsyn, O. B., Kabat, E. A., Wu, T. T., Levitt, M., Robson, B. and Nagano, K. 1974. Comparison of predicted and experimentally determined secondary structure of adenyl kinase. Nature 250: 140.

Bech, A.-M. and Foltmann, B. 1981. Partial primary structure of Mucor miehei protease. Neth. Milk Dairy J. 35: 275.

Berger, A. and Schechter, I. 1970. Mapping the active site of papain with the aid of peptide substrates and inhibitors. Phil. Trans. Roy. Soc. Ser. B. 257: 249.

Bernal, J. D., Fankuchen, I. and Perutz, M. 1938. An X-ray study of chymotrypsin and haemoglobin. Nature 141: 523.

Berridge, N. J. 1945. The purification and crystallization of rennin. Biochem. J. 39: 179.

Berridge, N. J. 1952. Some observations on the determination of the activity of rennet. Analyst 77: 57.

Beychok, S. 1966. Circular dichroism of biological macromolecules. Science 154: 1288.

Beynon, R. J. 1982. The diagonal plot for sequence comparisons. Biochemistry Microcomputer Group 7: 11.

Bigelow, C. C. 1967. On the average hydrophobicity of proteins and the relation between it and protein structure. J. Theoret. Biol. 16: 187.

Blout, E. R. 1971. Polypeptide and proteins. In: Fundamental Aspects and Recent Developments in Optical Rotatory Dispersion and Circular Dichroism. Ciardelli, F. and Salvadori, P. (Eds.). p. 352. Heyden and Sons, Ltd., New York, NY.

Blumenfeld, O. O., Leonis, J. and Perlmann, G. E. 1960. The effect of guanidine hydrochloride on crystalline pepsin. J. Biol. Chem. 235: 379.

Bolotina, I. A., Chekhov. V. O. and Lugauskas, V. Y. 1979. Determination of the secondary structures of protein by circular dichroism spectra. Calculation of the protein basic circular dichroism spectra for anti-parallel and parallel β-structures and β-bends. Int. J. Quantum Chem. 16: 819.

Bolotina, I. A., Chekhov, V. O., Lugauskas, V. Y., Finkelstein, A. V. and Ptitsyn, O. B. 1980. Determination of the secondary structure of proteins from circular dichroism spectra. II. Consideration of the contribution of β-bends. Mol. Biol. 14: 709.

Box, G. E. P. 1957. Evolutionary operation: a method for increasing industrial productivity. Appl. Statist. 6: 3.

Brahms, S. and Brahms, J. G. 1979. Comparison of protein secondary structure in solution and in crystals. J. Chim. Phys. 76: 841.

Brahms, S. and Brahms, J. G. 1980. Determination of protein secondary structure in solution by vacuum ultraviolet circular dichroism. J. Mol. Biol. 138: 149.

Brahms, S., Brahms, J. G., Spach, G. and Brack, A. 1977. Identification of β, β-turns and unordered conformations in polypeptide chains by vacuum ultraviolet circular dichroism. Proc. Natl. Acad. Sci. U.S.A. 74: 3208.

Bunn, C. W., Moews, P. C. and Baumber, M. E. 1971. Crystallography of calf rennin (chymosin). Proc. R. Soc. London B178: 245.

Burgess, A. W. and Scheraga, H. A. 1975. Assessment of some problems associated with prediction of the three-dimensional structure of a protein from its amino acid sequence. Proc. Natl. Acad. Sci. U.S.A. 72: 1221.

Bush, C. A., Ralapati, S. and Duben, A. 1981. Computer-controlled vacuum ultraviolet circular dichroism spectrometer with Fourier digital data smoothing. Anal. Chem. 53: 1140.

Chang, C. T., Wu, C.-S. C. and Yang, J. T. 1978. Circular dichroic analysis of protein conformation: inclusion of the β-turns. Anal. Biochem. 91: 13.

Cheeseman, G. C. 1969. Effects of some protein modifying agents on the properties of rennin. J. Dairy Res. 36: 299.

Chen, G. C. and Yang, J. T. 1977. Two-point calibration of circular dichrometer with d-10-camphorsulfonic acid. Anal. Letters 10: 1195.

Chen, Y.-H. and Yang, J. T. 1971. A new approach to the calculation of secondary structures of globular proteins by optical rotatory dispersion and circular dichroism. Biochem. Biophys. Res. Commun. 44: 1285.

Chen, Y.-H., Yang, J. T. and Chau, K. H. 1974. Determination of the helix and β form of proteins in aqueous solution by circular dichroism. Biochemistry 13: 3350.

Chen, Y.-H., Yang, J. T. and Martinez, H. M. 1972. Determination of the secondary structures of proteins by circular dichroism and optical rotatory dispersion. Biochemistry 11: 4120.

Chothia, C. 1975. Structural invariants in protein folding. Nature 254: 304.

Chou, P. Y. and Fasman, G. D. 1977. β-turns in proteins. J. Mol. Biol. 115: 135.

Chou, P. Y. and Fasman, G. D. 1978a. Empirical predictions of protein conformation. Ann. Rev. Biochem. 47: 251.

Chou, P. Y. and Fasman, G. D. 1978b. Prediction of the secondary structure of proteins from their amino acid sequence. Adv. Enzymol. 47: 45.

Cid, H., Bunster, M., Arriagada, E. and Campos, M. 1982. Prediction of secondary structure of proteins by means of hydrophobicity profiles. FEBS Letters 150: 247.

Clement, G. E. 1973. Catalytic activity of pepsin. Prog. Bioorg. Chem. 2: 177.

Crawford, J. L., Lipscomb, W. N. and Schellman, C. G. 1973. The reverse turn as a polypeptide conformation in globular proteins. Proc. Natl. Acad. Sci. U.S.A. 70: 538.

Cunningham, L. W. Jr., Tietze, F., Green, N. M. and Neurath, H. 1953. Molecular-kinetic properties of trypsin and related proteins. Discussion Faraday Soc. 13: 58.

Dalgleish, D. G. 1982. The enzymatic coagulation of milk. In: Developments in Dairy Chemistry-1. Fox, P. F. (Ed.). p. 157. Applied Science Publishers, New York, NY.

Daultry, S. 1976. Principal Components Analysis. Geo Abstracts Ltd., University of East Anglia, Norwich.

Deker, C. M., Bovey, F. A., Carver, J. P. and Blout, E. R. 1970. Nuclear magnetic resonance evidence for cis-peptide bonds in proline oligomers. J. Am. Chem. Soc. 92: 6191.

de Koning, P. J. 1968. Studies on rennin and the genetic variants of casein. Neth. Milk Dairy J. 22: 121.

de Koning, P. J., van Rooijen, P. J. and Visser, S. 1978. Application of a synthetic hexapeptide as a standard substrate for the determination of the activity of chymosin. Neth. Milk Dairy J. 32: 232.

Deming, S. N. and Parker, L. R. 1978. A review of simplex optimization in analytical chemistry. Crit. Rev. Anal. Chem. 7: 187.

Derde, M. P. and Massart, D. L. 1982. Extraction of information from large data sets by pattern recognition. Z. Anal. Chem. 313: 484.

Deschamps (1840). J. Pharm. Sci. Acces. 26: 412. Cited in: Foltmann, B. 1966. A review of prorennin and rennin. Compt. Rend. Trav. Lab. Carlsberg 35: 143.

Djurtoft, R., Foltmann, B. and Johansen, A. 1963. Studies on rennin. VIII: On the molecular weight of prorennin and rennin. Compt. Rend. Trav. Lab. Carlsberg 34: 287.

Dols, T. J. and Armbrecht, B. H. 1976. Simplex optimization as a step in method development. J. Assoc. Offic. Anal. Chem. 59: 1204.

Donovan, J. 1969. Ultraviolet absorption. In: Physical Principles and Techniques of Protein Chemistry. Part A. Leach, S. (Ed.). p. 101. Academic Press, New York, NY.

Drenth, J. 1981. The three-dimensional structure of proteolytic enzymes. Neth. Milk Dairy J. 35: 197.

Edelhoch, H. 1957. The denaturation of pepsin. I. Macromolecular changes. J. Am. Chem. Soc. 79: 6100.

Edelhoch, H. 1958a. The denaturation of pepsin. II. Hydrogen ion equilibria of native and denatured pepsin. J. Am. Chem. Soc. 80: 6640.

Edelhoch, H. 1958b. The denaturation of pepsin. III. The effects of various protein denaturants on the kinetics of pepsin inactivation. J. Am. Chem. Soc. 80: 6648.

Ernstrom, C. A. 1974. Milk clotting enzymes and cheese chemistry. In: Fundamentals of Dairy Chemistry, 2nd ed. Webb, B. H., Johnson, A. H. and Alford, J. A. (Eds.). AVI Publishing Co., Westport, CT.

Etoh, Y., Shoun, H., Beppu, T. and Arima, K. 1979. Physicochemical and immunochemical studies on similarities of acid proteases Mucor pusillus rennin and Mucor miehei rennin. Agric. Biol. Chem. 43: 209.

Fasman, G. D. 1980. Prediction of protein conformation from the primary structure. Ann. N.Y. Acad. Sci. 348: 147.

Foltmann, B. 1959. On the crystallization, stability and proteolytic activity of rennin. Acta Chem. Scand. 13: 1927.

Foltmann, B. 1966. A review on prorennin and rennin. Compt. Rend. Trav. Lab. Carlsberg 35: 143.

Foltmann, B. 1971. The biochemistry of prorennin (prochymosin) and rennin (chymosin). In: Milk Proteins: Chemistry and Molecular Biology. Vol 2. McKenzie, H. A. (Ed.). p. 217. Academic Press, New York, NY.

Foltmann, B. 1981. Mammalian milk-clotting proteases: structure, function, evolution and development. Neth. Milk Dairy J. 35: 223.

Foltmann, B. and Pedersen, V. B. 1976. Comparison of the primary structures of acidic proteinases and their zymogens. In: Acid Proteases: Structure, Function and Biology. Tang, J. (Ed.). p. 3. Adv. Exp. Med. Biol. Vol. 95. Plenum Press, New York, NY.

Foltmann, B., Pedersen, V. B., Jacobsen, H., Kaufmann, D. and Wybrandt, G. 1977. The complete amino acid sequence of prochymosin. Proc. Natl. Acad. Sci. U.S.A. 74: 2321.

Foss, J. G. 1963. Absorption, dispersion, circular dichroism, and rotary dispersion. J. Chem. Education 40: 592.

Fox, P. F. 1969. Milk-clotting and proteolytic activities of rennet and of bovine pepsin and porcine pepsin. J. Dairy Res. 36: 427.

Fox, P. F. 1981. Proteinases in dairy technology. Neth. Milk Dairy J. 35: 233.

Frane, J., Jennrich, R. and Sampson, P. 1981. Factor analysis. BMDP Statistical Software. Department of Biomathematics, University of California, Los Angeles, University of California Press.

Frank, I. E. and Kowalski, B. R. 1982. Chemometrics. Anal. Chem. 54: 232R.

Fruton, J. S. 1970. Pepsin. In: The Enzymes. Vol. 3. Boyer, P. D. (Ed.). p. 119. Academic Press, New York, NY.

Fukumoto, J., Tsuru, D. and Yamamoto, T. 1967. Studies on mold protease. Part I. Purification, crystallization and some enzymatic properties of acid protease of Rhizopus chinensis. Agric. Biol. Chem. 31: 710.

Fujii, S. and Nakai, S. 1980. Optimization of data transformations for linearization. Can. Inst. Food Sci. Technol. J. 13: 188.

Gabeloteau, C. and Desnuelle, P. 1960. Sur l'activation du trypsinogène de boeuf par une proteinase cristallisee d'Aspergillus saitoi. Biochim. Biophys. Acta 42: 230.

Garnier, J., Osguthorpe, D. J. and Robson, B. 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J. Mol. Biol. 120: 97.

Garnier, J., Salesse, R., Rerat, B., Rerat, C. and Blake, C. 1976. Comparison of X-ray data to estimated secondary structures from amino acid sequence and circular dichroism of human prealbumin. J. Chim. Phys. 73: 1018.

Gayle, J. B. and Bennett, H. D. 1978. Consequences of model departures on the resolution of multicomponent spectra by multiple regression and linear programming. Anal. Chem. 50: 2085.

Gibbs, A. J. and McIntyre, G. A. 1970. The diagram, a method for comparing sequences. Its use with amino acid and nucleotide sequences. Eur. J. Biochem. 16: 1.

Gower, J. C. 1982. Data analysis: multivariate or univariate and other difficulties. In: Food Research and Data Analysis. Martens, H. and Russwarm, H. Jr. (Eds.). p. 39. Applied Science Publishers, New York, NY.

Green, M. L. 1972. Assessment of swine, bovine and chicken pepsins as rennet substitutes for cheddar cheese-making. J. Dairy Res. 39: 261.

Green, M. L. and Marshall, R. J. 1977. The acceleration by cationic materials of the coagulation of casein micelles by rennet. J. Dairy Sci. 44: 521.

Greenfield, N., Davidson, B. and Fasman, G. D. 1967. The use of computed optical rotatory dispersion curves for the evaluation of protein conformation. Biochemistry 6: 1630.

Greenfield, N. and Fasman, G. D. 1969. Computed circular dichroism spectra for the evaluation of protein conformation. Biochemistry 8: 4108.

Grosse, R., Malur, J., Meiske, W. and Repke, K. R. H. 1974. Statistical behaviour and suitability of protein-derived circular dichroic-basis spectra for the determination of globular protein conformation. Biochim. Biophys. Acta 359: 33.

Grouke, M. J. and Gibbs, J. H. 1967. Transition from random coil to ∝-helix induced by sodium dodecyl sulfate. Biopolymers 5: 586.

Guzzo, A. V. 1965. The influence of amino acid sequence on protein structure. Biophys. J. 5: 809.

Hamed, M. M., Robinson, R. M. and Mattice, W. L. 1983. Helix formation upon acidification of protein-dodecyl sulfate complexes. Biochim. Biophys. Acta. 743: 260.

Hartley, B. S. 1960. Proteolytic enzymes. Ann. Rev. Biochem. 29: 45.

Hayakawa, S. and Nakai, S. 1983. Relationships of hydrophobicity and net charge to the solubility of milk proteins. Paper No. D101, presented at 78th Annual Meeting of the American Dairy Science Association, University of Wisconsin, Madison, June 26.

Hayashi, K., Fukushima, D. and Mogi, K. 1967. Physicochemical properties and amino acid composition of alkaline proteinase from Aspergillus sojae. Agric. Biol. Chem. 31: 1171.

Heindl, M.-C., Fermandjian, S. and Keil, B. 1980. Circular dichroism comparative studies of two bacterial collagenases and thermolysin. Biochim. Biophys. Acta 624: 51.

Hennessey, J. P. Jr. and Johnson, W. C. Jr. 1981. Information content in the circular dichroism of proteins. Biochemistry 20: 1085.

Hennessey, J. P. Jr. and Johnson, W. C. Jr. 1982. Experimental errors and their effect on analyzing circular dichroism spectra of proteins. Anal. Biochem. 125: 177.

Hennessey, J. P. Jr., Johnson, W. C. Jr., Bahler, C. and Wood, H. G. 1982. Subunit interactions of transcarboxylase as studied by circular cichroism. Biochemistry 21: 642.

Hill, R. D. 1969. Synthetic peptide and ester substrates for rennin. J. Dairy Res. 36: 409.

Hill, R. L. 1965. Hydrolysis of proteins. Adv. Protein Chem. 20: 37.

Hoffman, D. L. and Young, F. W. 1982. Quantitative analysis of qualitative data: applications in food research. In: Food Research and Data Analysis. Martens, H. and Russwarm, H. Jr. (Eds.). p. 63. Applied Science Publishers, New York, NY.

Hofmann, T. 1974. Structure, function and evolution of acid proteases. In: Food Related Enzymes. Whitaker, J. R. (Ed.). p. 146. Adv. Chem. Ser. 136. American Chemical Society, Washington, D.C.

Hsu, I.-N., Delbaere, L. T. J., James, M. N. G. and Hofmann, T. 1977. Penicillopepsin from Penicillium janthinellum crystal structure at 2.8Å and sequence homology with porcine pepsin. Nature 266: 140.

Huang, H. T. and Dooley, J. G. 1976. Enhancement of cheese flavours with microbial esterases. Biotech. Bioeng. 18: 909.

Ichishima, E. 1970. Purification and mode of assay for acid proteinase of Aspergillus saitoi. In: Methods in Enzymology. Vol. XIX. Perlmann, G. E. and Lorand, L. (Eds.). Academic Press, New York, NY.

Ichishima, E. and Yoshida, F. 1965. Molecular weight and amino acid composition of the acid proteinase of Aspergillus saitoi. Biochim. Biophys. Acta 110: 155.

Ichishima, E. and Yoshida, F. 1966a. Conformation of Aspergillopeptidase A in aqueous solution. Biochim. Biophys. Acta 128: 130.

Ichishima, E. and Yoshida, F. 1966b. N- and C- terminal residues in the acid proteinase of Aspergillus saitoi. J. Biochem. 59: 183.

Ichishima, E. and Yoshida, F. 1967. Conformation of Aspergillopeptidase A in aqueous solution. Part II. Ultraviolet optical rotatory dispersion of Aspergillopeptidase A. Biochim. Biophys. Acta 147: 341.

IUPAC-IUB. 1970. Commission on biochemical nomenclature - abbreviations and symbols for the description of the conformation of polypeptide chains. Tentative rules 1969. Biochemistry 9: 3471.

Iwasaki, S., Tamura, G. and Arima, K. 1967a. Milk clotting enzyme from microorganisms. Part II. The enzyme production and the properties of the crude enzyme. Agric. Biol. Chem. 31: 546.

Iwasaki, S., Tamura, G. and Arima, K. 1967b. Milk clotting enzyme from microorganisms. Part III. The purification of the enzyme and its properties. Agric. Biol. Chem. 31: 1421.

James, M. N. G. and Sielecki, A. R. 1983. Structure and refinement of penicillopepsin at 1.8Å resolution. J. Mol. Biol. 163: 299.

Janin, J. 1976. Surface area of globular proteins. J. Mol. Biol. 105: 13.

Jenkins, J. A.,. Blundell, T. L., Tickle, I. J. and Ungaretti, L. 1975. The low resolution structure analysis of an acid proteinase from Endothia parsitica. J. Mol. Biol. 99: 583.

Jenkins, J. A., Tickle, I. J., Sewell, T., Ungaretti, L., Wollmer, A. and Blundell, T. L. 1976. X-ray analysis and circular dichroism of the acid protease from Endothia parasitica and chymosin. In: Acid Proteases: Structure, Function and Biology. Tang, J. (Ed.). p. 43. Adv. Exp. Med. Biol. Vol. 95. Plenum Press, New York, NY.

Jirgensons, B. 1969. Optical Rotatory Dispersions of Proteins and Other Macromolecules. Springer-Verlag, New York, NY.

Jollès, J., Alais, C. and Jollès, P. 1968. The tryptic peptide with the rennin-sensitive linkage of cow's κ-casein. Biochim. Biophys. Acta. 168: 591.

Jones, D. D. 1975. Amino acid properties and side-chain orientation in proteins: a cross correlation approach. J. Theoret. Biol. 50: 167.

Kabat, E. A. and Wu, T. T. 1973a. The influence of nearest neighbouring amino acid residues on aspects of secondary structure of proteins. Attempt to locate ∝-helices and β-sheets. Biopolymers 12: 751.

Kabat, E. A. and Wu, T. T. 1973b. The influence of nearest neighbouring amino acids on the conformation of the middle amino acid in predicted and experimental determination of β-sheets in concanavalin A. Proc. Natl. Acad. Sci. U.S.A. 70: 1473.

Kabsch, W. and Sander, C. 1983. How good are predictions of protein secondary structure? FEBS Letters 155: 179.

Kanehisa, M. I. and Tsong, T. Y. 1980. Local hydrophobicity stabilizes secondary structures in proteins. Biopolymers 19: 1617.

Katchalski, E. and Sela, M. 1958. Synthesis and chemical properties of poly-∝-amino acids. Adv. Protein Chem. 13: 243.

Kato, A. and Nakai, S. 1980. Hydrophobicity determined by a fluorescent probe method and its correlation with surface properties of proteins. Biochim. Biophys. Acta 624: 13.

Kauzmann, W. 1959. Some factors in the interpretation of protein denaturation. Adv. Protein Chem. 14: 1.

Kay, J. and Valler, M. J. 1981. The aspartate proteinases used as rennet substitutes. Neth. Milk Dairy J. 35: 281.

Kitson, T. M. and Knowles, J. R. 1971. The effect of arginine modification on the pH dependence of pepsin activity. FEBS Letters 16: 337.

Kotelchuck, D. and Scheraga, H. A. 1969. The influence of short-range interactions on protein conformation. II. A model for predicting the ∝-helical regions of proteins. Proc. Natl. Acad. Sci. U.S.A. 62: 14.

Kowalski, B. R. 1980. Chemometrics. Anal. Chem. Rev. 52: 112.

Kubota, Y., Takahashi, S., Nishikawa, K. and Ooi, T. 1981. Homology in protein sequences expressed by correlation coefficients. J. Theoret. Biol. 91: 347.

Kuntz, I. D. 1972. Protein folding. J. Am. Chem. Soc. 94: 4009.

Kwan, K. K. H., Nakai, S. and Skura, B. J. 1983. Comparison of four methods for determining protease activity in milk. J. Food Sci. 48: 1419.

Kwan, W.-O. and Kowalski, B. R. 1980. Correlation of objective chemical measurements and subjective sensory evaluations. Wines of Vitis vinifera variety "Pinot Noir" from France and the United States. Anal. Chim. Acta 122: 215.

Kyte, J. and Doolittle, R. F. 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157: 105.

Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227: 680.

Lee, B. and Richards, F. M. 1971. The interpretation of protein structures: estimation of static accessibility. J. Mol. Biol. 55: 379.

Leggett, D. J. 1977. Numerical analysis of multicomponent spectra. Anal. Chem. 49: 276.

Lehninger, A. L. 1970. Biochemistry. Worth Publishers, Inc. New York, NY.

Lehninger, A. L. 1982. Principles of Biochemistry. Worth Publishers, Inc. New York, NY.

Levitt, M. and Chothia, C. 1976. Structural patterns in globular proteins. Nature 261: 552.

Levitt, M. and Greer, J. 1977. Automatic identification of secondary structure in globular proteins. J. Mol. Biol. 114: 181.

Lewis, P. N., Momany, F. A. and Scheraga, H. A. 1973. Chain reversals in proteins. Biochim. Biophys. Acta 303: 211.

Li, L.-K. and Spector, A. 1969. The circular dichroism of β-poly-L-lysine. J. Am. Chem. Soc. 91: 220.

Li-Chan, E., Helbig, N., Holbek, E., Chau, S. and Nakai, S. 1979. Covalent attachment of lysine to wheat gluten for nutritional improvement. J. Agric. Food Chem. 27: 877.

Lim, V. I. 1974. Algorithms for prediction of α-helical and β-structural regions in globular proteins. J. Mol. Biol. 88: 873.

Linderstrom-Lang, K. U. and Schellman, J. A. 1959. Protein structure and enzyme activity. In: The Enzymes. Vol. 1, 2nd ed. Boyer, P. D. (Ed.). p. 443. Academic Press, New York, NY.

Lloyd, D. A. 1969. Partial nucleotide sequences for three unique T-1 RNase fragments of TMV-RNA. Ph.D. Thesis. University of California, Berkeley.

Lowenstein, H. 1974. The use of ethylene maleic anhydride for the preparation of a water-soluble polyanionic derivative of pepsin. Preparation and properties. Acta. Chem. Scand. B. 28: 1098.

Lukovits, I. 1983. Quantitative structure-activity relationships employing independent quantum chemical indices. J. Med. Chem. 26: 1104.

Ma, C.-Y. 1979. Chemical modification of carboxyl groups in porcine pepsin. Ph.D. Thesis. University of British Columbia, Vancouver.

Mackinlay, A. G. and Wake, R. G. 1971. κ-casein and its attack by rennin (chymosin). In: Milk Proteins: Chemistry and Molecular Biology. Vol. II. McKenzie, H. A. (Ed.). p. 175. Academic Press, New York, NY.

Madison, V. and Schellman, J. 1970. Location of proline derivatives in conformational space. II. Theoretical optical activity. Biopolymers 9: 569.

Madison, V. and Schellman, J. 1972. Optical activity of polypeptides and proteins. Biopolymers 11: 1041.

Magar, M. E. 1968. On the analysis of the optical rotatory dispersion of proteins. Biochemistry 7: 617.

Mains, G., Takahashi, M., Sodek, J. and Hofmann, T. 1971. The specificity of penicillopepsin. Can. J. Biochem. 49: 1134.

Martens, H. 1982. Understanding food research data. In: Food Research and Data Analysis. Martens, H. and Russwarm, H. Jr. (Eds.). p. 5. Applied Science Publishers, New York, NY.

Martens, H. and Russwarm, H. Jr. (Eds.). 1982. Food Research and Data Analysis. Applied Science Publishers, New York, NY.

Mattice, W. L. and Harrison, W. H. 1976. The importance of coulombic interactions for the induction of β-structure in lysine oligomers by sodium dodecyl sulfate. Biopolymers 15: 559.

McCord, R. W., Blakeney, E. W. and Mattice, W. L. 1977. Conformational changes induced in ionized poly(L-arginine) and poly(L-histidine) by sodium dodecyl sulfate. Biopolymers 16: 1319.

Mickelson, R. and Ernstrom, C. A. 1963. Factors affecting the stability of rennin. J. Dairy Sci. 46: 613.

Moews, P. C. and Bunn, C. W. 1970. An X-ray crystallographic study of the rennin-like enzyme of Endothia parasitica. J. Mol. Biol. 54: 395.

Moffitt, W. 1956. Optical rotatory dispersion of helical polymers. J. Chem. Phys. 25: 467.

Moffitt, W. and Moscowitz, A. 1959. Optical activity in absorbing media. J. Chem. Phys. 30: 648.

Moffitt, W. and Yang, J. T. 1956. The optical rotatory dispersion of simple polypeptides. I. Proc. Natl. Acad. Sci. U.S.A. 42: 596.

Moravek, L. and Kostka, V. 1974. Complete amino acid sequence of hog pepsin. FEBS Letters 43: 207.

Morgan, S. L. and Deming, S. N. 1974. Simplex optimization of analytical methods. Anal. Chem. 46: 1170.

Myer, Y. P. 1970. A new method for the conformational analysis of proteins and polypeptides from circular dichroism spectra. Res. Commun. Chem. Pathol. Pharmacol. 1: 607.

Nagano, K. 1973. Logical analysis of the mechanism of protein folding. I. Prediction of helices, loops and β-structures from primary structure. J. Mol. Biol. 75: 401.

Nakai, S. 1982. Comparison of optimization techniques for application to food product and process development. J. Food Sci. 47: 144.

Nakai, S. 1983. Structure-function relationships of food proteins with an emphasis on the importance of protein hydrophobicity. J. Agric. Food Chem. 31: 676.

Nakano, M. and Yang, J. T. 1981. Helical conformation of myoglobin and its cyanogen bromide-treated fragments in sodium dodecyl sulfate solution. Arch. Biochem. Biophys. 207: 69.

Nakayama, S., Nagashima, Y., Hoshino, M., Moriyama, A., Takahashi, K., Watanabe, T. and Yoshida, M. 1983. Structural study on the active site of porcine pepsin and Rhizopus chinensis acid protease. Spin labeling with diazoketone reagents. J. Biochem. 93: 1297.

Nelder, J. A. and Mead, R. 1965. A simplex method for function minimization. Computer J. 7: 308.

Nozaki, T. and Tanford, C. 1971. The solubility of amino acids and two glycine peptides in aqueous ethanol and dioxane solutions. Establishment of a hydrophobicity scale. J. Biol. Chem. 246: 2211.

Okabayashi, H., Isemura, T. and Sakakibara, S. 1968. Steric structure of L-proline oligopeptides. II. Far-ultraviolet absorption spectra and optical rotations of L-proline oligopeptides. Biopolymers 6: 323.

Ottesen, M. and Rickert, W. 1970a. The isolation and partial characterization of an acid protease produced by Mucor miehei. Compt. Rend. Trav. Lab. Carlsberg 37: 301.

Ottesen, M. and Rickert, W. 1970b. The acid protease of Mucor miehei. In: Methods in Enzymology. Vol. XIX. Perlmann, G. E. and Lorand, L. (Eds.). p. 459. Academic Press, New York, NY.

Palau, J., Argos, P. and Puigdomenech, P. 1982. Protein secondary structure. Studies on the limits of prediction. Int. J. Peptide Protein Res. 19: 394.

Pauling, L. and Corey, R. B. 1951. Configurations of polypeptide chains with favored orientations around single bonds: two new pleated sheets. Proc. Natl. Acad. Sci. U.S.A. 37: 295.

Pauling, L., Corey, R. B. and Branson, H. R. 1951. The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain. Proc. Natl. Acad. Sci. U.S.A. 37: 205.

Perlmann, G. E. 1959. Effect of solvents and of temperature on the optical rotatory properties of pepsin. Proc. Natl. Acad. Sci. U.S.A. 45: 915.

Pham, A.-M. 1981. Prediction program of secondary structure from sequence of proteins according to the method of Chou and Fasman. M.Sc. Thesis. University of British Columbia, Vancouver.

Ponnuswamy, P. K., Prabhahakaran, M. and Manavalan, P. 1980. Hydrophobic packing and spatial arrangement of amino acid residues in globular proteins. Biochim. Biophys. Acta 623: 301.

Powers, J. C., Harley, A. D. and Myers, D. V. 1976. Subsite specificity of porcine pepsin. In: Acid Proteases: Structure, Function and Biology. Tang, J. (Ed.). p. 141. Adv. Exp. Med. Biol. Vol. 95. Plenum Press, New York, NY.

Provencher, S. W. 1980. CONTIN - a portable Fortran IV program for the regularized solution of linear algebraic and linear integral equations of the first kind with options for peak constraints and linear equality and inequality constraints. European Molecular Biology Laboratory, Heidelberg, Federal Republic of Germany.

Provencher, S. W. and Glöckner, J. 1981. Estimation of globular protein secondary structure from circular dichroism. Biochemistry 20: 33.

Ptitsyn, O. B. and Finkelstein, A. V. 1970. Connection between the secondary and primary structures of globular proteins. Biofizika 15: 757.

Ptitsyn, O. B. and Finkelstein, A. V. 1980. Similarities of protein topologies: evolutionary divergence, functional convergence or principles of folding? Q. Rev. Biophys. 13: 339.

Ptitsyn, O. B. and Finkelstein, A. V. 1983. Theory of protein secondary structure and algorithm of its prediction. Biopolymers 22: 15.

Puhan, Z. and Irvine, D. M. 1973. Proteolysis by proteases of Bacillus subtilis used to make Canadian cheddar cheese. J. Dairy Sci. 56: 317.

Raap, J., Kerling, K. E. T., Vreeman, H. J. and Visser, S. 1983. Peptide substrates for chymosin (rennin): conformational studies of κ-casein and some κ-casein-related oligopeptides by circular dichroism and secondary structure prediction. Arch. Biochem. Biophys. 221: 117.

Rao, S. P. and Dunn, B. M. 1981. Conformation of pepsin and pepsinogen: some aspects of the role of tyrosine residues and the 1-44 segment of pepsinogen on conformational stability. Arch. Biochem. Biophys. 207: 135.

Raymond, M. N. and Bricas, E. 1979. New chromophoric peptide substrates for chymosin (rennin). J. Dairy Sci. 62: 1719.

Raymond, M. N., Bricas, E. and Mercier, J. C. 1973. Synthesis of new oligopeptide substrates of chymosin (rennin) and kinetic parameters of their hydrolyses. Neth. Milk Dairy J. 27: 298.

Richardson, J. S. 1977. β-sheet topology and the relatedness of proteins. Nature 268: 495.

Richardson, J. S. 1981. The anatomy and taxonomy of protein structure. Adv. Protein Chem. 34: 167.

Rickert, W. S. 1970. The degradation of the β-chain of oxidized insulin by Mucor miehei protease. Compt. Rend. Trav. Lab. Carlsberg 38: 1.

Rickert, W. S. and Elliott, J. E. 1973. Acid proteases from species of Mucor: molecular weight of Mucor miehei protease from amino acid analysis data. Can. J. Biochem. 51: 1638.

Rickert, W. S. and McBride-Warren, P. A. 1974. Structural and functional determinants of Mucor miehei protease. III. Isolation and composition of the carbohydrate moiety. Biochim. Biophys. Acta 336: 437.

Rickert, W. S. and McBride-Warren, P. A. 1977. Structural and functional determinants of Mucor miehei protease. VI. Inactivation of the enzyme by diazoacetyl norleucine methyl ester, pepstatin and 1,2-epoxy-3-(p-nitrophenoxy) propane. Biochim. Biophys. Acta 480: 262.

Rose, G. D. 1978. Prediction of chain turns in globular proteins on a hydrophobic basis. Nature 272: 586.

Rose, G. D. and Roy, S. 1980. Hydrophobic basis of packing in globular proteins. Proc. Natl. Acad. Sci. U.S.A. 77: 4643.

Sakar, P. K. and Doty, P. 1966. The optical rotatory properties of the β-conformation in polypeptides and proteins. Proc. Natl. Acad. Sci. U.S.A. 55: 981.

Satake, I. and Yang, J. T. 1973. Effect of chain length and concentration of anionic surfactants on the conformational transitions of poly(L-ornithine) and poly(L-lysine) in aqueous solution. Biochem. Biophys. Res. Commun. 54: 930.

Saxena, V. P. and Wetlaufer, D. B. 1971. A new basis for interpreting the circular dichroic spectra of proteins. Proc. Natl. Acad. Sci. U.S.A. 68: 969.

Schattenkerk, C. and Kerling, K. E. T. 1973. Relation between structure and capacity to function as a rennin substrate. Neth. Milk Dairy J. 27: 286.

Schober, R., Heimburger, N. and Printz, J. 1960. Reinigung und Charakterisierung von Labpraparaten durch Saulenchromatographie an Carboxylmethyl-Cellulose und durch Papier-electrphorese. Milchwiss. 15: 506.

Schulz, G. E. and Schirmer, R. H. 1978. Principles of Protein Structures. Springer-Verlag, New York, NY.

Sears, D. W. and Beychok, S. 1973. Circular dichroism. In: Physical Principles and Techniques of Protein Chemistry. Part C. Leach, S. J. (Ed.). p. 445. Academic Press, New York, NY.

Siegel, J. B., Steinmetz, W. E. and Long, G. L. 1980. A computer-assisted model for estimating protein secondary structure from circular dichroic spectra: comparison of animal lactate dehydrogenases. Anal. Biochem. 104: 160.

Sklar, L. A., Hudson, B. S., Peterson, M. and Diamond, J. 1977. Conjugated polyene fatty acids as fluorescent probes: spectroscopic characterization. Biochemistry 16: 813.

Smith, E. L. and Kimmel, J. R. 1960. Papain. In: The Enzymes. Vol. 4. Boyer, P. D., Lardy, H. and Myrbäck, K. (Eds.). p. 133. Academic Press, New York, NY.

Smith, J. A. and Pease, L. G. 1980. Reverse turns in peptides and proteins. Crit. Rev. Biochem. 8: 400.

Sober, H. A. (Ed.). 1970. Handbook of Biochemistry. 2nd ed. Chemical Rubber Co., Cleveland, OH.

Sodek, J. and Hofmann, T. 1970a. Microbial acid proteinases. In: Methods in Enzymology. Vol. XIX. Perlmann, G. E. and Lorand, L. (Eds.). p. 391. Academic Press, New York, NY.

Sodek, J. and Hofmann, T. 1970b. Large-scale preparation and some properties of penicillopepsin, the acid proteinase of Penicillium janthinellum. Can. J. Biochem. 48: 425.

Sodek, J. and Hofmann, T. 1970c. Amino acid sequence around the active site aspartic acid in penicillopepsin. Can. J. Biochem. 48: 1014.

Spendley, W., Hext, G. R. and Himsworth, F. R. 1962. Sequential application of simplex designs in optimization and evolutionary operation. Technometrics 4: 441.

Sternberg, M. Z. 1972. Bond specificity, active site and milk clotting mechanism of the Mucor miehei protease. Biochim. Biophys. Acta 285: 383.

Strickland, E. H. 1974. Aromatic contributions to circular dichroism spectra of protein. Crit. Rev. Biochem. 11: 113.

Stryer, L. 1965. The interaction of a naphthalene dye with apomyoglobin and apohemoglobin. A fluorescent probe for nonpolar binding sites. J. Mol. Biol. 13: 482.

Stryer, L. 1968. Fluorescence spectroscopy of proteins. Science 162: 521.

Stuper, A. J., Brugger, W. E. and Jurs, P. C. 1979. Computer Assisted Studies of Chemical Structure and Biological Functions. Wiley, New York, NY.

Subramanian, E., Liu, M., Swan, I. D. A. and Davies, D. R. 1976. The Crystal structure of an acid protease from Rhizopus chinensis at 2.5Å resolution. In: Acid Proteases: Structure, Function and Biology. Tang, J. (Ed.). p. 33. Adv. Exp. Med. Biol. Vol. 95. Plenum Press, New York, NY.

Snyder, L. R. 1978. Classification of the solvent properties of common liquids. J. Chromatog. Sci. 16: 223.

Tam, J. J. and Whitaker, J. R. 1972. Rates and extents of hydrolysis of several caseins by pepsin, rennin, Endothia parasitica protease and Mucor pusillus protease. J. Dairy Sci. 55: 1522.

Tanford, C. 1962. Contribution of hydrophobic interactions to the stability of the globular conformation of proteins. J. Am. Chem. Soc. 84: 4240.

Tang, J. 1963. Specificity of pepsin and its dependence on a possible hydrophobic binding site. Nature 199: 1094.

Tang, J. (Ed.). 1976. Acid Proteases: Structure, Function and Biology. Plenum Press, New York, NY.

Tang, J. 1979. Evolution in the structure and function of carboxyl proteases. Mol. Cell. Biochem. 26: 93.

Tang, J., Sepulveda, P., Marciniszyn, J. Jr., Chen, K. C. S., Huang, W.-Y., Tao, N., Liu, D. and Lanier, J. P. 1973. Amino acid sequence of porcine pepsin. Proc. Natl. Acad. Sci. U.S.A. 70: 3437.

Townend, R., Kumosinki, T. F. and Timasheff, S. N. 1967. The circular dichroism of variants of β-lactoglobulin. J. Biol. Chem. 242: 4538.

Townsend, A.-A. E. 1982. Relationship between physicochemical proper-
    ties of proteins and their foaming characteristics. Ph.D. Thesis.
    University of British Columbia, Vancouver.

Trout, G. E. and Fruton, J. S. 1969. The side-chain specificity of
    pepsin. Biochemistry 8: 4231.

UBC MATRIX. 1979. A guide to solving matrix problems. Bird, C.
    (Ed.). Computing Centre, University of British Columbia, Vancouver.

UBC PLOT. 1981. The UBC plot subroutines and programs. Mair, S. G.
    (Ed.). Computing Centre, University of British Columbia, Vancouver.

van de Voort, F. R., Ma, C.-Y. and Nakai, S. 1979. Molecular weight
    distribution of interacting proteins calculated by multiple regres-
    sion analysis from sedimentation equilibrium data. Arch. Biochem.
    Biophys. 195: 596.

Venkatachalam, C. M. 1968. Stereochemical criteria for polypeptides
    and proteins. V. Conformation of a system of three linked peptide
    units. Biopolymers 6: 1425.

Visser, S. 1981. Proteolytic enzymes and their action on milk pro-
    teins. A review. Neth. Milk Dairy J. 35: 65.

Visser, S., van Rooijen, P. J., Schattenkerk, C. and Kerling, K. E. T.
    1976. Peptide substrates for chymosin (rennin). Kinetic studies
    with peptides of different chain length including parts of the
    sequence 101-112 of bovine κ-casein. Biochim. Biophys. Acta 438:
    265.

Visser, S., van Rooijen, P. J., Schattenkerk, C. and Kerling, K. E. T.
    1977. Peptide substrates for chymosin (rennin). Kinetic studies
    with bovine κ-casein (103-108) hexapeptide analogues. Biochim.
    Biophys. Acta 481: 171.

Visser, S., van Rooijen, P. J. and Slangen, C. J. 1980. Peptide
    substrates for chymosin (rennin). Isolation and substrate behaviour
    of two tryptic fragments of bovine κ-casein. Eur. J. Biochem. 108:
    415.

Voynick, I. M. and Fruton, J. S. 1971. The comparative specificity of
    acid proteinases. Proc. Natl. Acad. Sci. U.S.A. 68: 257.

Walsh, K. A. and Neurath, H. 1964. Trypsinogen and chymotrypsinogen as
    homologous proteins. Proc. Natl. Acad. Sci. U.S.A. 52: 884.

Weber, G. and Daniel, E. 1966. Cooperative effects in binding by
    bovine serum albumin. II. The binding of 1-anilino-8-naphthalene
    sulfonate. Polarization of the ligand fluorescence and quenching of
    the protein fluorescence. Biochemistry 5: 1900.

Weber, G. and Young, L. B. 1964. Fragmentation of bovine serum albumin
    by pepsin. J. Biol. Chem. 239: 1415.

Whitaker, J. R. 1970. Protease of Endothia parasitica. In: Methods in Enzymology. Vol. XIX. Perlmann, G. E. and Lorand, L. (Eds.). p. 436. Academic Press, New York, NY.

White, F. H. Jr. 1976. Studies on secondary structure in chicken egg-white lysozyme after reductive cleavage of disulfide bonds. Biochemistry 15: 2906.

Williams, D. C., Whitaker, J. R. and Caldwell, P. V. 1972. Hydrolysis of peptide bonds of the oxidized β-chain of insulin by Endothia parasitica protease. Arch. Biochem. Biophys. 149: 52.

Wold, S., Albano, C., Dunn, W. J., Esbensen, K., Hellberg, S., Johansson, E. and Sjöström, M. 1982. Pattern recognition: finding and using regularities in multi-variate data. In: Food Research and Data Analysis. Martens, H. and Russwarm, H. Jr. (Eds.). p. 147. Applied Science Publishers, New York, NY.

Woody, R. W. and Tinoco, I. Jr. 1967. Optical rotation of oriented helixes. III. Calculation of the rotatory dispersion and circular dichroism of the alpha- and 3_{10}-helix. J. Chem. Phys. 46: 4927.

Yang, J. T. 1956. On the phenomenological treatments of optical rotatory dispersion of polypeptides and proteins. Proc. Natl. Acad. Sci. U.S.A. 42: 438.

Yoshida, F. 1956. Studies on the proteolytic enzymes of black Aspergilli. Part I. Investigation of strains producing superior proteinase yields in black Aspergillus and the crystallization of proteolytic enzyme from Aspergillus saitoi. Bull. Agric. Chem. Soc. Japan 20: 252.

Yoshida, F. and Nagasawa, M. 1956a. Studies on the proteolytic enzymes of black Aspergilli. Part II. Several properties of the crystalline proteolytic enzyme obtained from Aspergillus saitoi. Bull. Agric. Chem. Soc. Japan 20: 257.

Yoshida, F. and Nagasawa, M. 1956b. Studies on the proteolytic enzyme of black Aspergilli. Part III. The specificity of crystalline Aspergillus saitoi proteinase on synthetic substrates. Bull. Agric. Chem. Soc. Japan 20: 262.

Yu, J., Ozawa, H. and Arima, K. 1970. (Cited in: Arima, K., Yu, J. and Iwasaki, S. 1970.) Milk-clotting enzyme from Mucor pussillus var. Lindt. In: Methods in Enzymology. Vol. XIX. Perlmann, G. E. and Lorand, L. (Eds.). p. 446. Academic Press, New York, NY.

Zar, J. H. 1974. Biostatistical Analysis. Prentice-Hall, Inc., Englewood Cliffs, NJ.

Zimmerman, S. S. and Scheraga, H. A. 1977. Local interactions in bends of proteins. Proc. Natl. Acad. Sci. U.S.A. 74: 4126.

APPENDIX

Appendix 1. Listing of a Fortran IV version of the computer program used by Siegel et al. (1978) to determine secondary structure fractions from CD spectral data.

```
1     C    *****************************************************************
2     C    *                                                               *
3     C    * THIS PROGRAM CALCULATES PERCENTAGE HELIX ACCORDING TO THE METHOD *
4     C    * OF SIEGEL,STEINMETZ AND LONG.1980. ANAL.BIOCHEM.104:160-167    *
5     C    *                                                               *
6     C    *****************************************************************
7     C
8     C    VARIABLES:
9     C      CORCTN : REFERENCE ELLIPTICITY CORRESPONDING TO THE C TERM
10    C      FRACTN : FRACTION OF HELIX AT SPECIFIC WAVELENGTH
11    C      HELIX  : FRACTION OF HELIX DETERMINED
12    C      N      : NUMBER OF WAVELENGTHS EXAMINED
13    C      PROTN  : PROTEIN EXAMINED
14    C      REFTHT : REFERENCE ELLIPTICITY CORRESPONDING TO THE HELIX
15    C      SAMPLE : MEASURED ELLIPTICITY OF SAMPLE PROTEIN
16    C      SHEET  : FRACTION OF SHEET DETERMINED
17    C      SSQDC  : VARIANCE CORRESPONDING TO THE C TERM
18    C      SSQTHC : VARIANCE CORRESPONDING TO THE HELIX AND C TERM
19    C      SSQTHT : STANDARD DEVIATION OF RESIDUAL ELLIPTICITY
20    C      STHETA : VARIANCE CORRESPONDING TO THE HELIX REFERENCE
21    C
22    C      VALUES FOR CORCTN. REFTHT, SSQDC, SSQTHC, SSQTHT,
23    C             AND STHETA ARE OBTAINED FROM THE PAPER BY
24    C             SIEGEL ET AL.(1980)
25    C
26    C      ** DATA ENTRY **
27    C       - NAME OF PROTEIN IS ENTERED FIRST FORMAT 80A1
28    C       - NUMBER OF WAVELENGTHS FORMAT 6X,I6
29    C       - MEASUED ELLIPTICITIES ARE ENTERED AT WAVELENGTHS:
30    C         210, 213, 216, 218. 220. 222. 224. 227. 229. 231.
31    C         234. 237. AND 240.
32    C       - MEASURED ELLIPTICITY ON ONE LINE FORMAT 6X,E10.4
33    C       - THE NEXT LINE CONTAINS VALUES FOR CORCTN. REFTHT,
34    C         SSQDC. SSQTHC. SSQTHT, AND STHETA FORMAT AS PER LINE 54
35    C
36         INTEGER N,I
37         INTEGER PROTN(80)
38         REAL SAMPLE,CORCTN,REFTHT,FRACTN,SSQDC,STHETA,SSQTHT,SSQTHC,HELIX,
39        1SUMONE,SUMTWO,FACTOR,CORFAC
40         DIMENSION SAMPLE(50),CORCTN(50),REFTHT(50),FRACTN(50),SSQDC(50),
41        1SSQTHT(50),STHETA(50),FACTOR(50),CORFAC(50),SSQTHC(50)
42         SUMONE=0
43         SUMTWO=0
44         I=1
45         READ(5,1)PROTN
46    1    FORMAT(80A1)
47         READ(5,10) N
48    10   FORMAT(6X,I6)
49    20   IF(I.GT.N) GO TO 50
50         READ(5,30) SAMPLE(I)
51    30   FORMAT(6X,E10.4)
52         READ(5,40) CORCTN(I),REFTHT(I),SSQDC(I),SSQTHT(I),STHETA(I),SSQTHC
53        1(I)
54    40   FORMAT(6X,E10.4,2X,E10.4,2X,E10.4,2X,E10.4,2X,E10.4,2X,E10.4)
55         FRACTN(I)=(SAMPLE(I)-CORCTN(I))/REFTHT(I)
56         FACTOR(I)=(SSQDC(I)+SSQTHT(I)**2)/(REFTHT(I)**2)+((SAMPLE(I)-
57        1CORCTN(I))**2)*STHETA(I)/(REFTHT(I)**4)
58         CORFAC(I)=FACTOR(I)+2*(SAMPLE(I)-CORCTN(I))*SSQTHC(I)/(REFTHT(I
```

```
59            1)**3)
60            SUMONE=SUMONE+(FRACTN(I)/CORFAC(I))
61            SUMTWO=SUMTWO+(1.0/CORFAC(I))
62            I=I+1
63            GO TO 20
64      50    CONTINUE
65            HELIX=(SUMONE/SUMTWO)*100
66            SHEET=(-0.729*HELIX)+0.583
67            WRITE(6,55) PROTN
68      55    FORMAT('1',80A1)
69            WRITE(6,60) HELIX
70      60    FORMAT('O','THE PERCENTAGE OF HELIX IS',F6.2)
71            WRITE(6,70) SHEET
72      70    FORMAT('2','THE PERCENTAGE OF SHEET IS',F6.2)
73            STOP
74            END
```

Appendix 2. Listing of a Fortran IV computer program similar to that used by Chang et al. (1978) to determine the secondary structure fractions from CD spectral data.

```
1      C    ******************************************************************
2      C    *                                                                *
3      C    *     THIS PROGRAM CALCULATES THE SECONDARY STRUCTURE OF PROTEINS *
4      C    *     USING THE METHOD OF CHANG ET AL.(1978) ANAL.BIOCHEM 91:13   *
5      C    *                                                                *
6      C    ******************************************************************
7      C
8      C    THIS PROGRAM UTILIZES THE SUBROUTINE DBEST WHICH IS AN ITERATIVE
9      C    ALGORITHM WHICH IS SUBJECT TO A NUMBER OF LINEAR EQUALITY
10     C    CONSTRAINTS. THIS PROGRAM IS DOCUMENTED IN THE UBC MATRIX MANUAL.
11     C    CONSTRAINTS ARE ENTERED AS THE FIRST SET OF DATA. TO HAVE THE
12     C    SECONDARY STRUCTURE FRACTIONS TOTAL ONE, 1.0'S SHOULD CONSTITUTE
13     C    THE FIRST ROW OF THE REFERENCE ELLIPTICITY MATRIX. A VALUE OF
14     C    ONE SHOULD ALSO BE THE FIRST VALUE OF THE OBSERVED ELLIPTICITY
15     C    DATA VECTOR.
16     C
17     C
18     C    VARIABLES:
19     C       DA    : IS A TWO DIMENSIONAL ARRAY CONTAINING THE REFERENCE
20     C               ELLIPTICITIES FOR THE FOUR SECONDARY STRUCTURE FRACTIONS.
21     C       DB    : IS A VECTOR CONTAINING THE OBSERVED ELLIPTICITIES
22     C       DETA  : IS A CONVERGENCE CRITERION
23     C       DRES  : IS A VECTOR CONTAINING THE RESIDUES
24     C       DX    : IS A VECTOR CONTAING THE SOLUTION
25     C       NHRS  : IS THE NUMBER OF VECTORS CONTAINING THE OBSERVED
26     C               ELLIPTICITIES
27     C
28          REAL*8 DA,DB,DETA,DTOL,DX,DRES
29          DIMENSION DA(100,20),DB(100,1),DX(100,1),DRES(100,1)
30     C
31     C    READ IN AND WRITE OUT THE REFERENCE ELLIPTICITIES
32     C    THE REFERENCE ELLIPTICITIES ARE OBTAINED FROM THE OUTPUT
33     C    OF PROGRAM YANG WHICH GENERATES THE REFERENCE ELLIPTICITIES
34     C
35          READ(5,10)((DA(I,J),J=1,4),I=1,52)
36     10   FORMAT(4F10.1)
37          WRITE(6,20)
38     20   FORMAT('1',' REFERENCE ELLIPTICITIES'/)
39          WRITE(6,30)
40     30   FORMAT('0',' HELIX     B-SHEET      B-TURN    RANDOM'//)
41          WRITE(6,40)((DA(I,J),J=1,4),I=1,52)
42     40   FORMAT(1X,4F10.1)
43          NRHS=1
44     C
45     C    READ IN OBSERVED ELLIPTICITIES 240-190NM, AT 1NM INTERVALS
46     C
47          READ(5,50)((DB(I,J),I=1,52),J=1,NRHS)
48     50   FORMAT(10F10.1)
49          WRITE(6,60)
50     60   FORMAT(///' OBSERVED ELLIPTICITIES 240-190NM, 1NM INTERVALS'/)
51          WRITE(6,70)((DB(I,J),I=1,52),J=1,NRHS)
52     70   FORMAT(1X,10F10.1)
53     C
54     C    CALL SUBROUTINE DBEST TO FIND THE BEST LEAST SQUARES SOLUTION
55     C
56          DTOL=1.D-8
57          DETA=5.D-7
58          CALL DBEST(DA,DB,NRHS,1,52,4,100,DETA,DTOL,DX,DRES)
```

```
59      C
60      C    WRITE OUT THE SOLUTION
61      C
62           WRITE(6,80)
63      80   FORMAT(///' SECONDARY STRUCTURE FRACTIONS'//'    HELIX       B-SHEET
64      1      B-TURN        RANDOM     '/)
65           WRITE(6,90)((DX(I,J),I=1,4),J=1,NRHS)
66      90   FORMAT(1X,4G12.4)
67           WRITE(6,100)((DRES(I,J),I=1,52),J=1,NRHS)
68      100  FORMAT(///' RESIDUALS'/(1X,10G12.4))
69           STOP
70           END
```

Appendix 3. Listing of a Fortran IV computer program to generate the
reference ellipticity values for the various secondary
structure fractions (∝-helix, β-sheet, β-turn and random
fractions).

```
1       C**************************************************************************
2       C*                                                                        *
3       C* THIS PROGRAM CORRECTS THE HELIX REFERENCE ELLIPITICITY USING           *
4       C* THE WAVELENGTH DEPENDENT FACTOR AND NUMBER OF RESIDUES PER HELIX.       *
5       C* FORMATS THE DATA FOR THE PROGRAM THAT CALCULATES THE SECONDARY          *
6       C* STRUCTURE USING THE METHOD OF CHANG ET AL.(1978) ANAL.BIOCHEM           *
7       C* 91:13 OR THE SIMPLEX-LEAST SQUARES METHOD                               *
8       C*                                                                        *
9       C**************************************************************************
10      C
11      C    VARIABLES
12      C      RES   : NUMBER OF AMINO ACID RESIDUES PER HELIX
13      C      WDF   : WAVELENGTH DEPENDENT FACTOR
14      C      HELIX : REFERENCE ELLIPITICITY- HELIX
15      C      BSHT  : REFERENCE ELLIPITICITY- BETA SHEET
16      C      BTURN : REFERENCE ELLIPITICITY- BETA TURN
17      C      RAND  : REFERENCE ELLIPITICITY- RANDOM COIL
18      C      ELLP  : OBSERVED ELLIPITICITIES AT WAVELENGTHS 240-190 NM
19      C
20      C    REFERENCE VALUES FOR HELIX, WAVELENGTH DEPENDENT FACTOR, BETA SHEET,
21      C    BETA TURN AND RANDOM COIL WERE OBTAINED FROM J.T.YANG (UNPUBLISHED
22      C    DATA) AND ARE CONTAINED IN FILE REFERENCE.
23      C
24      C    ** DATA ENTRY **
25      C     - NUMBER OF RESIDUES PER HELICAL REGION IS ENTERED FIRST, IF
26      C       UNKNOWN 10.4 SHOULD BE ENTERED. THIS IS AN AVERAGE VALUE
27      C       BASESD ON X-RAY DATA.
28      C     - REFERENCE DATA IS THEN ENTERED IN THE FOLLOWING ORDER: HELIX,
29      C       WAVELENGTH DEPENDENT FACTOR, BETA SHEET, BETA TURN AND
30      C       RANDOM COIL.
31      C     - OBSERVED ELLIPTICITIES ARE THE ENTERED AT 1NM INTERVALS 240
32      C       TO 190 NM (51 DATA POINTS).
33      C
34      C    * POSSIBLE CHANGES TO THE PROGRAM *
35      C     - IF THE METHOD OF CHANG ET AL.(1978) IS TO BE USED THEN THE
36      C       DO LOOP CONTAINING STATEMENT 60 SHOULD BE CHANGED TO I=1,52.
37      C       THE INITIAL ENTRY OF 1.0 INTO THE VARIOUS REFERENCE
38      C       ELLIPTICITIES ALLOWS FOR THE CONSTRAINT OF THE SUM OF THE
39      C       SECONDARY FRACTIONS TO TOTAL ONE TO BE MET.
40      C     - FOR THE SIMPLEX METHOD I=2,52 SHOULD BE USED
41      C
42      C    ** DATA OUTPUT **
43      C     - OUTPUT FROM THIS PROGRAM COULD BE USED IN EITHER THE METHOD OF
44      C       CHANG ET AL.(1978) OR THE SIMPLEX METHOD DEPENDING ON I= STATE-
45      C       MENT.
46      C
47            DIMENSION BSHT(100),BTURN(100),HELIX(100),WDF(100),RAND(100),ELLP(
48           1100)
49            READ(5,10) RES
50      10    FORMAT(F4.1)
51            HELIX(1)=1.0
52            READ(5,20)(HELIX(I),I=2,52)
53      20    FORMAT(10F10.0/10F10.0/10F10.0/10F10.0/10F10.0/F10.0)
54            WDF(1)=1.0
55            READ(5,30)(WDF(I),I=2,52)
56      30    FORMAT(10F6.1/10F6.1/10F6.1/10F6.1/10F6.1/F6.1)
57            DO 40 I=2,52
58            HELIX(I)=HELIX(I)*(1.0-(WDF(I)/RES))
```

```
59      40      CONTINUE
60              BSHT(1)=1.0
61              READ(5,20)(BSHT(I),I=2,52)
62              BTURN(1)=1.0
63              READ(5,20)(BTURN(I),I=2,52)
64              RAND(1)=1.0
65              READ(5,20)(RAND(I),I=2,52)
66              ELLP(1)=1.0
67              READ(5,20)(ELLP(I),I=2,52)
68              DO 60 I=1,52
69              WRITE(6,50) HELIX(I),BSHT(I),BTURN(I),RAND(I)
70      50      FORMAT(4F10.1)
71      60      CONTINUE
72              WRITE(6,70)(ELLP(I),I=1,52)
73      70      FORMAT(10F10.1)
74              STOP
75              END
```

```
1
2     -3510..-4630..-6000..-7640..-9560..-11800..-14200..-16200..-18900..-21500.;
3     -24000..-26700..-29000..-31200..-33200..-35500..-37000..-37500..-37400..-36900.;
4     -36300..-35700..-35300..-35000..-34800..-35000..-35300..-36000..-36800..-37300.;
5     -38000..-37500..-36000..-33200..-28800..-22600..-14500..-4530..7060..20000.;
6     36700..47600..61000..73200..83400..91000..95600..97000..95300..90700.;
7     83800.;
8     2.5,2.5,2.5,2.5,2.5,2.5,2.5,2.5,2.5,2.5;
9     2.5,2.5,2.5,2.5,2.5,2.5,2.5,2.5,2.5,2.5;
10    2.6,2.6,2.7,2.7,2.8,2.9,3.0,3.1,3.1,3.2;
11    3.3,3.4,3.5,3.6,3.7,4.1,4.8,9.4,1.5,1.2;
12    1.8,2.1,2.2,2.3,2.3,2.4,2.4,2.4,2.4,2.4;
13    2.4;
14    970..1232..1436..1541..1564..1578..1708..1876..2137..1313.;
15    2227..2167..2024..1551..1077..454..-596..-981..-1763..-2488..
16    -3208..-3858..-4657..-5257..-5814..-6240..-6486..-6492..-6324..-5707.;
17    -4921..-4504..-3981..-3201..-2278..-1196..69..957..2164..4882.;
18    8107..9584..10287..10015..8748..7491..4955..1584..-1611..-3476.;
19    -5804.;
20    1840..2174..2684..3342..3881..4344..5133..5424..6957..8120.;
21    10837..13319..15407..17257..18614..20488..21928..21350..20139..18728.;
22    17057..15279..13634..12041..10291..8951..7718..6973..6033..5156.;
23    6299..7301..8079..10641..13476.;16834..19434..21290..24258..22609.;
24    16881..14134..2490..-11077..-25634..-38681..-50781..-61324..-70284..-75556.;
25    -77435..
26    -1879..-2235..-2515..-2816..-2986..-3133..-3639..-4504..-5691..-6344.;
27    -8661..-10249..-11636..-12485..-13023..-13222..-12870..-12890..-12281..-11782.;
28    -11053..-10245..-9159..-8130..-7034..-6112..-5519..-5253..-5016..-5401.;
29    -6666..-7990..-8974..-11018..-13440..-16472..-20133..-23447..-25840..-32054.;
30    -37206..-37208..-35064..-31528..-26383..-21216..-15243..-8059..-285..5999.;
31    11985.;
32
33
34
35
36
37
```

This file contains the reference ellipticities for α-helix (lines 2-7),
β-sheet (lines 14-19), β-turn (lines 20-25) and random (lines 26-31).
Lines 8 to 13 contain the wavelength dependent factors. The reference
ellipticities as well as wavelength dependent factors were obtained from
Yang (unpublished data). The average number of residues per helix ($\bar{n}$)
is entered in line one, while the observed ellipticities (240-190 nm)
for the protein in question are entered into lines 32 to 37.

Appendix 4. Listing of a Fortran IV computer program which utilizes the simplex optimization algorithm of Morgan and Deming (1974) to determine the secondary structure fractions from CD spectral data.

```
 1      C    ********************************************************************
 2      C    *                                                                  *
 3      C    *     THIS PROGRAM UTILIZES THE MORGAN-DEMING SIMPLEX ALGORITHM     *
 4      C    *     TO CALCULATE THE SECONDARY STRUCTURE OF PROTEINS USING THE    *
 5      C    *     ELLIPTICITY DATA FROM C.D. DATA                               *
 6      C    *                                                                  *
 7      C    ********************************************************************
 8      C
 9      C
10      C    VARIABLES:
11      C       A       : RESPONSE VALUES FOR SIMPLEX USED IN THE TERMINATION
12      C                 PROCEDURE
13      C       BLOC    : BEST LOCATION IN THE SIMPLEX MATRIX, ROW NO.
14      C       CENT    : CENTROID VALUES
15      C       CP      : P VALUES OF STARTING SIMPLEX MATRIX
16      C       CQ      : Q VALUES OF STARTING SIMPLEX MATRIX
17      C       CR      : CONTRACTION OF R RESPONSES
18      C       CRP     : MASSIVE CONTRACTION OF R RESPONSES
19      C       CW      : CONTRACTION OF W RESPONSES
20      C       CWP     : MASSIVE CONTRACTION OF W RESPONSES
21      C       EOBS    : VECTOR OF OBSERVED ELLIPTICITIES
22      C       EXPAN   : EXPANSION RESPONSES
23      C       ISIMN   : SIMPLEX NUMBER
24      C       IVERT   : VERTEX NUMBER
25      C       LIMIT   : VECTOR OF LIMITS FOR EACH FACTOR, LOWER THEN UPPER
26      C       N       : NUMBER OF SECONDARY STRUCTURE FRACTIONS
27      C       NW      : NUMBER OF WAVELENGTHS EXAMINED
28      C       RC      : MATRIX OF REFERENCE ELLIPTICITIES, ROW REPRESENTS A
29      C                 WAVELENGTH
30      C       REFLEC  : REFELECTION RESPONSES
31      C       SD      : REPONSES FOR EACH SIMPLEX CONDITION
32      C       SS      : SIMPLEX MATRIX
33      C       WLOC    : WORST LOCATION IN SIMPLEX MATRIX, ROW NO.
34      C
35      C    ** DATA ENTRY **
36      C
37      C     - FIRST SET DATA ENTERED IS N AND NW, FORMAT 2I3
38      C     - LIMITS FOR EACH FACTOR ARE THEN ENTERED LOWER FIRST,
39      C       THEN UPPER, FORMAT 8F6.0
40      C     - OBSERVED ELLIPTICITIES ARE ENTERED 7 PER LINE FOR 7 LINES,
41      C       2 DATA FOR THE 8TH LINE, TOTAL OF 51 DATA POINTS FROM 240
42      C       TO 190 NM, 1 NM INTERVALS
43      C     - REFERENCE ELLIPTICITIES ARE ENTERED 4 PER LINE (REPRESENT THE
44      C       4 SECONDARY STRUCTURES) FOR 51 WAVELENGTHS 240 TO 190 NM
45      C       FORMAT 4E12.4. THESE ARE VALUES OBTAINED FROM THE REFERENCE
46      C       PROGRAM. REFERENCE PROGRAM CORRECTS FOR HELIX REFERENCE FOR
47      C       CHAIN LENGTH DEPENDENCY. REFERENCE ELLIPTICITY PROGRAM IS
48      C       RUN PRIOR TO THIS PROGRAM.
49      C
50           REAL LIMIT
51           INTEGER WLOC,BLOC
52           DIMENSION SS(100,100),CQ(100),CP(100),LIMIT(100),RC(100,100)
53           DIMENSION EOBS(100),SD(10),CENT(10),A(10)
54           DIMENSION REFLEC(10),CW(10),EXPAN(10),CR(10),CRP(10),CWP(10)
55
56           READ(5,10) N,NW
57      10   FORMAT(2I3)
58           L=N+1
```

```
59              M=N*2
60              ISIMN=1
61              IVERT=L
62              READ(5,20)(LIMIT(I),I=1,M)
63      20      FORMAT(8F6.0)
64              READ(5,25)(EOBS(I),I=1,NW)
65      25      FORMAT(7(F7.0,6F8.0,/),F7.0,F8.0)
66              WRITE(6,25)(EOBS(I),I=1,NW)
67              READ(5,27)((RC(I,J),J=1,N),I=1,NW)
68      27      FORMAT(4E12.4)
69              WRITE(6,27)((RC(I,J),J=1,N),I=1,NW)
70              WRITE(6,30) N
71      30      FORMAT(' NUMBER OF VARIABLES=',I3///)
72              P=(1/(N*SQRT(2.0)))*((N-1)+SQRT(N+1.0))
73              Q=(1/(N*SQRT(2.0)))*(SQRT(N+1.0)-1.0)
74              J=1
75              DO 60 I=1,N
76              WRITE(6,40) I
77      40      FORMAT(' VARIABLE',I3/)
78              WRITE(6,50) LIMIT(J),LIMIT(J+1)
79      50      FORMAT(' LOWER LIMIT=',F6.2,'   UPPER LIMIT=',F6.2//)
80              CP(I)=(LIMIT(J+1)-LIMIT(J))*P+LIMIT(J)
81              CQ(I)=(LIMIT(J+1)-LIMIT(J))*Q+LIMIT(J)
82              J=J+2
83      60      CONTINUE
84              J=1
85              DO 70 I=1,M,2
86              SS(1,J)=LIMIT(I)
87              J=J+1
88      70      CONTINUE
89              K=1
90              DO 90 I=1,N
91              DO 80 J=2,L
92              SS(J,I)=CQ(K)
93      80      CONTINUE
94              K=K+1
95      90      CONTINUE
96              J=1
97              I=1
98              DO 100 K=2,L
99              SS(K,J)=CP(I)
100             I=I+1
101             J=J+1
102     100     CONTINUE
103     C
104     C  CONSTRAINTING VALUES IN THE SIMPLEX   SUCH THAT
105     C  THE SUM OF THE FRACTIONS EQUALS ONE
106     C
107             DO 107 I=1,L
108             SUM=0.0
109             DO 103 J=1,N
110             SUM=SUM+SS(I,J)
111     103     CONTINUE
112             IF(SUM.EQ.0.0) GOTO 107
113             DO 105 J=1,N
114             SS(I,J)=SS(I,J)/SUM
115     105     CONTINUE
116     107     CONTINUE
```

```
117        120  WRITE(6,121)
118        121  FORMAT(125(1H*))
119             WRITE(6,125) ISIMN
120        125  FORMAT(/' SIMPLEX NUMBER',I3/1X,17(1H*)//)
121             WRITE(6,127)((SS(I,J),J=1,N),I=1,L)
122        127  FORMAT(4G12.4)
123
124        C
125        C   CALCULATION OF SIMPLEX RESPONSES
126        C
127
128             IF(ISIMN.NE.1) GOTO 153
129             DO 150 K=1,L
130             ALSD=O.O
131             DO 140 I=1,NW
132             CALC=O.O
133             DO 130 J=1,N
134             RESULT=RC(I,J)*SS(K,J)
135             CALC=CALC+RESULT
136        130  CONTINUE
137             ALSD=ALSD+(EOBS(I)-CALC)**2
138        140  CONTINUE
139             SD(K)=ALSD
140        150  CONTINUE
141        153  WRITE(6,155)
142        155  FORMAT(/' SIMPLEX RESPONSES'/)
143             WRITE(6,157)(SD(K),K=1,L)
144        157  FORMAT(6G12.4)
145
146        C
147        C   FIND THE WORST AND THE BEST RESPONSE
148        C
149
150             WORST=SD(1)
151             WLOC=1
152             DO 160 K=2,L
153             IF(WORST.GT.SD(K)) GOTO 160
154             WORST=SD(K)
155             WLOC=K
156        160  CONTINUE
157             BEST=SD(1)
158             BLOC=1
159             DO 170 K=2,L
160             IF(BEST.LT.SD(K)) GOTO 170
161             BEST=SD(K)
162             BLOC=K
163        170  CONTINUE
164
165        C
166        C   TERMINATION PROCEDURE
167        C
168
169             DO 180 I=1,N
170             A(I)=SD(I)
171        180  CONTINUE
172             LAST=N
173             NLAST=N-1
174             DO 200 J=1,NLAST
```

```
175            MLIM=LAST-1
176            DO 190 I=1,MLIM
177            IF(A(I).LT.A(I+1)) GOTO 190
178            TEMP=A(I)
179            A(I)=A(I+1)
180            A(I+1)=TEMP
181     190    CONTINUE
182            LAST=LAST-1
183     200    CONTINUE
184            CRIT=.05*A(1)
185            DIFF=ABS(A(N)-A(1))
186            IF(DIFF.LE.CRIT) GOTO 570
187
188     C
189     C   CALCULATE CENTROID
190     C
191
192     220    DO 240 J=1,N
193            CENT(J)=0.0
194            DO 230 I=1,L
195            IF(I.EQ.WLOC) GOTO 230
196            CENT(J)=CENT(J)+SS(I,J)
197     230    CONTINUE
198            CENT(J)=CENT(J)/(L-1)
199     240    CONTINUE
200            WRITE(6,245)
201     245    FORMAT(/' CENTROID VALUES'/)
202            WRITE(6,247)(CENT(J),J=1,N)
203     247    FORMAT(6G12.4)
204
205     C
206     C   CALCULATE AVERAGE
207     C
208
209            AVG=0.0
210            DO 250 I=1,L
211            IF(I.EQ.WLOC) GOTO 250
212            IF(I.EQ.BLOC) GOTO 250
213            AVG=AVG+SD(I)
214     250    CONTINUE
215            AVG=AVG/(L-2)
216            WRITE(6,255) AVG
217     255    FORMAT(/' AVERAGE RESPONSE',G12.4/)
218
219     C
220     C   CALCULATION OF THE REFLECTION POINT
221     C
222
223            IVERT=IVERT+1
224            WRITE(6,270) IVERT
225     270    FORMAT(/' VERTEX NUMBER',I3/)
226            WRITE(6,275)
227     275    FORMAT('     REFLECTION'/)
228            DO 280 I=1,N
229            REFLEC(I)=CENT(I)+1.0*(CENT(I)-SS(WLOC,I))
230     280    CONTINUE
231            CALL CHECK(REFLEC,LIMIT,M)
232            CALL CONST(REFLEC,N)
```

```
233           CALL LEAST(EOBS,RC,REFLEC,NW,N,TOTAL)
234           SD(L+1)=TOTAL
235           WRITE(6,290)(REFLEC(I),I=1,N),SD(L+1)
236     290   FORMAT(4G12.4,6X,'RESPONSE=',G12.4)
237           IF(SD(L+1).LT.SD(BLOC)) GOTO 340
238           IF(SD(L+1).LT.AVG) GOTO 370
239           IF(SD(L+1).LT.SD(WLOC)) GOTO 430
240
241     C
242     C   CONTRACTION OF W
243     C
244
245           IVERT=IVERT+1
246           WRITE(6,270) IVERT
247           WRITE(6,300)
248     300   FORMAT(/'    CONTRACTION OF W'/)
249           DO 310 I=1,N
250           CW(I)=CENT(I)-0.5*(CENT(I)-SS(WLOC,I))
251     310   CONTINUE
252           CALL CHECK(CW,LIMIT,M)
253           CALL CONST(CW,N)
254           CALL LEAST(EOBS,RC,CW,NW,N,TOTAL)
255           SD(L+2)=TOTAL
256           WRITE(6,290)(CW(I),I=1,N),SD(L+2)
257           IF(SD(L+2).GT.SD(WLOC)) GOTO 530
258           DO 320 I=1,N
259           SS(WLOC,I)=CW(I)
260     320   CONTINUE
261           SD(WLOC)=SD(L+2)
262           WRITE(6,330)
263     330   FORMAT(/' CONTRACTION-W REPLACES WORST'//)
264           ISIMN=ISIMN+1
265           GOTO 120
266
267     C
268     C   EXPANSION CALCULATION
269     C
270
271     340   IVERT=IVERT+1
272           WRITE(6,270) IVERT
273           WRITE(6,350)
274     350   FORMAT('    EXPANSION'/)
275           DO 360 I=1,N
276           EXPAN(I)=CENT(I)+2.0*(CENT(I)-SS(WLOC,I))
277     360   CONTINUE
278           CALL CHECK(EXPAN,LIMIT,M)
279           CALL CONST(EXPAN,N)
280           CALL LEAST(EOBS,RC,EXPAN,NW,N,TOTAL)
281           SD(L+2)=TOTAL
282           WRITE(6,290)(EXPAN(I),I=1,N),SD(L+2)
283           IF(SD(L+2).LT.SD(L+1)) GOTO 400
284
285     C
286     C   REFLECTION REPLACES WORST
287     C
288
289     370   DO 380 I=1,N
290           SS(WLOC,I)=REFLEC(I)
```

```
291        380   CONTINUE
292              SD(WLOC)=SD(L+1)
293              WRITE(6,390)
294        390   FORMAT(/' REFLECTION REPLACES WORST'//)
295              ISIMN=ISIMN+1
296              GOTO 120
297
298     C
299     C   EXPANSION REPLACES WORST
300     C
301
302        400   DO 410 I=1,N
303              SS(WLOC,I)=EXPAN(I)
304        410   CONTINUE
305              SD(WLOC)=SD(L+2)
306              WRITE(6,420)
307        420   FORMAT(/' EXPANSION REPLACES WORST'//)
308              ISIMN=ISIMN+1
309              GOTO 120
310
311     C
312     C   CONTRACTION OF R
313     C
314
315        430   IVERT=IVERT+1
316              WRITE(6,270) IVERT
317              WRITE(6,440)
318        440   FORMAT(/'    CONTRACTION OF R'/)
319              DO 450 I=1,N
320              CR(I)=CENT(I)+0.5*(CENT(I)-SS(WLOC,I))
321        450   CONTINUE
322              CALL CHECK(CR,LIMIT,M)
323              CALL CONST(CR,N)
324              CALL LEAST(EOBS,RC,CR,NW,N,TOTAL)
325              SD(L+2)=TOTAL
326              WRITE(6,290)(CR(I),I=1,N),SD(L+2)
327              IF(SD(L+2).GT.SD(L+1)) GOTO 480
328              DO 460 I=1,N
329              SS(WLOC,I)=CR(I)
330        460   CONTINUE
331              SD(WLOC)=SD(L+2)
332              WRITE(6,470)
333        470   FORMAT(/' CONTRACTION OF R REPLACES WORST'//)
334              ISIMN=ISIMN+1
335              GOTO 120
336
337     C
338     C   CALCULATION- MASSIVE CONTRACTION OF R
339     C
340
341        480   IVERT=IVERT+1
342              WRITE(6,270) IVERT
343              WRITE(6,490)
344        490   FORMAT(/'    MASSIVE CONTRACTION OF R'/)
345              DO 500 I=1,N
346              CRP(I)=CENT(I)+0.25*(CENT(I)-SS(WLOC,I))
347        500   CONTINUE
348              CALL CHECK(CRP,LIMIT,M)
```

```
349            CALL CONST(CRP,N)
350            CALL LEAST(EOBS,RC,CRP,NW,N,TOTAL)
351            SD(L+3)=TOTAL
352            WRITE(6,290)(CRP(I),I=1,N),SD(L+3)
353            IF(SD(L+3).GT.SD(L+1)) GOTO 370
354            DO 510 I=1,N
355            SS(WLOC,I)=CRP(I)
356      510   CONTINUE
357            SD(WLOC)=SD(L+3)
358            WRITE(6,520)
359      520   FORMAT(/' MASSIVE CONTRACTION OF R REPLACES WORST'//)
360            ISIMN=ISIMN+1
361            GOTO 120
362
363    C
364    C   MASSIVE CONTRACTION OF W
365    C
366
367      530   IVERT=IVERT+1
368            WRITE(6,270) IVERT
369            WRITE(6,535)
370      535   FORMAT(/'    MASSIVE CONTRACTION OF W '/)
371            DO 550 I=1,N
372            CWP(I)=CENT(I)-0.25*(CENT(I)-SS(WLOC,I))
373      550   CONTINUE
374            CALL CHECK(CWP,LIMIT,M)
375            CALL CONST(CW,N)
376            CALL LEAST(EOBS,RC,CWP,NW,N,TOTAL)
377            SD(WLOC)=TOTAL
378            DO 555 I=1,N
379            SS(WLOC,I)=CWP(I)
380      555   CONTINUE
381            WRITE(6,560)
382      560   FORMAT(' MASSIVE CONTRACTION OF W REPLACES WORST'//)
383            WRITE(6,290)(CWP(I),I=1,N),SD(WLOC)
384            ISIMN=ISIMN+1
385            GO TO 120
386      570   WRITE(6,580)
387      580   FORMAT('1',' THE BEST SOLUTION'//)
388            WRITE(6,590)
389      590   FORMAT('   HELIX    B-SHEET    B-TURN    RANDOM'//)
390            WRITE(6,600)(SS(BLOC,J),J=1,N)
391      600   FORMAT(6G12.4)
392            STOP
393            END
394
395    C
396    C   THIS SUBROUTINE CALCULATES THE LEAST SQUARES DIFFERENCE
397    C
398
399    C    VARIABLES:
400    C    SIMCO : SIMPLEX COEFFICIENTS
401    C    TOTAL : SUM OF SQUARES OF THE DIFFERENCE BETWEEN CALCULATED
402    C            AND OBSERVED VALUES
403    C
404            SUBROUTINE LEAST(EOBS,RC,SIMCO,NW,N,TOTAL)
405            DIMENSION EOBS(100),SIMCO(100),RC(100,100)
406            TOTAL=0.0
```

```
407              DO 750 I=1,NW
408              RESULT=0.0
409              DO 700 K=1,N
410              RESULT=RESULT+(RC(I,K)*SIMCO(K))
411        700   CONTINUE
412              TOTAL=TOTAL+(RESULT-EOBS(I))**2
413        750   CONTINUE
414              RETURN
415              END
416
417     C
418     C   THIS SUBROUTINE CHECKS FOR BOUNDARY VIOLATIONS - IF A FACTOR
419     C   IS LESS THAN THE LOWER LIMIT IT IS SET TO THE LOWER LIMIT. IF
420     C   IT EXCEEDS THE UPPER LIMIT IT IS SET TO THE UPPER LIMIT
421     C
422     C   VARIABLES:
423     C      FAC   : SIMPLEX COEFFICIENTS
424     C      LIMIT : LIMITS FOR THE FACTORS
425     C      NUM   : COUNTER
426     C      INDIC : INDICATOR OF BOUNDARY VIOLATION
427     C      IFLAG : VECTOR CONTAINING FACTORS EXCEEDING BOUNDARIES.
428     C
429
430              SUBROUTINE CHECK(FAC,LIMIT,M)
431              REAL LIMIT
432              DIMENSION FAC(10),LIMIT(20),IFLAG(10)
433              INDIC=0
434              K=1
435              J=1
436              DO 850 I=1,M,2
437              IF(FAC(K).LT.LIMIT(I)) GOTO 810
438              IF(FAC(K).GT.LIMIT(I+1)) GOTO 820
439              K=K+1
440              GOTO 850
441        810   FAC(K)=LIMIT(I)
442              IFLAG(J)=K
443              INDIC=1
444              K=K+1
445              J=J+1
446              NUM=J-1
447              GOTO 850
448        820   FAC(K)=LIMIT(I+1)
449              IFLAG(J)=K
450              INDIC=1
451              K=K+1
452              J=J+1
453              NUM=J-1
454        850   CONTINUE
455              IF(INDIC.EQ.0) GOTO 900
456              WRITE(6,860)
457        860   FORMAT(' BOUNDARY VIOLATION- THE FOLLOWING FACTORS EXCEEDED THEIR
458             1 LIMIT'/)
459              DO 870 J=1,NUM
460              WRITE(6,865) IFLAG(J)
461        865   FORMAT(1X,I3/)
462        870   CONTINUE
463        900   RETURN
464              END
```

```
465
466    C
467    C   THIS SUBROUTINE CONSTRAINTS THE FRACTIONS SUCH THAT
468    C   THE SUM OF THE FRACTIONS EQUALS ONE
469    C
470    C    VARIABLES:
471    C     FRAC  : SIMPLEX COEFFICIENTS
472    C     SUM   : TOTAL OF SIMPLEX COEFFICIENTS
473    C
474
475        SUBROUTINE CONST(FRAC,N)
476        DIMENSION FRAC(10)
477        SUM=0.0
478        DO 905 I=1,N
479        SUM=SUM+FRAC(I)
480    905 CONTINUE
481        IF(SUM.EQ.0.0) GOTO 920
482        DO 910 I=1,N
483        FRAC(I)=FRAC(I)/SUM
484    910 CONTINUE
485    920 RETURN
486        END
```

Appendix 5. Listing of a Fortran IV computer program for the diagonal
plot method.

```
 1    C  ****************************************************************
 2    C  *                                                              *
 3    C  *   DIAGONAL PLOT PROGRAM USING THE METHOD OF BEYNON(1982)     *
 4    C  *           BIOCHEMISTRY MICROCOMPUTER GROUP 7:11              *
 5    C  *                                                              *
 6    C  ****************************************************************
 7    C
 8    C  VARIABLES:
 9    C     B     : TWO DIMENSIONAL ARRAY CONTAINING THE PRIMARY SEQUENCE
10    C             OF PROTEIN 1
11    C     C     : TWO DIMENSIONAL ARRAY CONTAINING THE PRIMARY SEQUENCE
12    C             OF PROTEIN 2
13    C     D     : VECTOR CONTAINING THE PRIMARY SEQUENCE OF PROTEIN 1
14    C     E     : VECTOR CONTAINING THE PRIMARY SEQUENCE OF PROTEIN 2
15    C     IMIN  : INTEGER VALUE FOR XMIN
16    C     IMAX  : INTEGER VALUE FOR XMAX
17    C     N     : NUMBER OF AMINO ACIDS TO BE COMPARED IN THE SEQUENCE
18    C     P     : NUMBER OF LINES OF DATA
19    C     XMIN  : START OF SEQUENCE TO BE COMPARED
20    C     XMAX  : END OF SEQUENCE TO BE COMPARED
21    C
22    C  ** DATA ENTRY **
23    C    - AMINO ACID SEQUENCE WAS DIGITIZED USING THE FOLLOWING CODE:
24    C      1(ALA), 2(ARG), 3(ASN), 4(ASP), 5(CYS), 6(GLN), 7(GLU), 8(GLY),
25    C      9(HIS), 10(ILE), 11(LEU), 12(LYS), 13(MET), 14(PHE), 15(PRO),
26    C      16(SER), 17(THR), 18(TRP),19(TYR), AND 20(VAL).
27    C    - 20 DATA POINTS ARE ENTERED PER LINE, FORMAT 20I2
28    C    - SUBROUTINES AXIS AND SYMBOL ARE DOCUMENTED IN THE UBC PLOT
29    C      MANUAL
30    C
31    C  * POSSIBLE CHANGES TO THE PROGRAM *
32    C    - XMIN,XMAX WILL HAVE TO BE CHANGED DEPENDING ON THE REGION
33    C      TO BE EXAMINED.
34    C    - NAMES IN THE 2 AXIS SUBROUTINES WILL HAVE TO BE CHANGED
35    C      DEPENDING ON THE PROTEINS TO BE COMPARED (I.E. PROTEIN 1
36    C      AND PROTEIN 2).
37    C    - THE NUMBERS DIRECTLY FOLLOWING THE NAMES WILL HAVE TO BE
38    C      CHANGED SINCE THESE NUMBERS CORRESPOND TO THE NUMBER OF
39    C      LETTERS IN THE PROTEIN NAME. N.B. BLANKS MUST BE INCLUDED
40    C      IN THE NUMBER COUNT. THE NEGATIVE SIGN SHOULD BE KEPT FOR
41    C      THE FIRST AXIS SUBROUTINE.
42    C    - INITIALLY FOR THE THE FIRST COMPARISONS THE XMIN IN THE AXIS
43    C      SUBROUTINES SHOULD BE REPLACED BY O.O (E.G. FOR COMPARISONS
44    C      OF RESIDUES 1 TO 85). FOR SUBSEQUENT COMPARISONS O.O SHOULD
45    C      BE REPLACED IN THE AXIS SUBROUTINES (E.G. RESIDUES 85 TO 170).
46    C      THIS IS DONE TO FACILITATE THE SYNCHRONIZING OF RESIDUE NUMBER
47    C      WITH PLOT NUMBER.
48    C    - INITIALLY JA AND KA SHOULD BE SET TO 1 AND O, RESPECTIVELY.
49    C      HOWEVER, FOR COMPARISONS OF RESIDUES LATER ON IN THE SEQUENCE
50    C      JA AND KA SHOULD BE RESET TO O AND -1, RESPECTIVELY (E.G. RES-
51    C      DUES 85 TO 170 JA=O, KA=-1 HOWEVER FOR RESIDUES 85 TO 170
52    C      JA=O, KA=-1). AGAIN THIS IS DONE TO SYNCHRONIZE RESIDUE
53    C      NUMBER TO PLOT NUMBER.
54    C
55    C
56          INTEGER B(500,500),C(500,500),D(500),E(500),P
57          READ(5,10) N,P
58     10   FORMAT(2I3)
```

```
59              READ(5,15) XMIN,XMAX
60        15    FORMAT(2F6.0)
61              READ(5,20) ((B(I,J),J=1,40),I=1,P)
62              READ(5,20) ((C(I,J),J=1,40),I=1,P)
63        20    FORMAT(40I2)
64              I=1
65              DO 30 K=1,P
66              DO 30 M=1,40
67              D(I)=B(K,M)
68              E(I)=C(K,M)
69              I=I+1
70        30    CONTINUE
71              CALL AXIS(0.0,0.0,'PROTEIN 1',-9,9.0,0.0,XMIN,10.0)
72              CALL AXIS(0.0,0.0,'PROTEIN 2',9,9.0,90.0,XMIN,10.0)
73              IMIN=IFIT(XMIN)
74              IMAX=IFIT(XMAX)
75              JA=0
76              DO 50 J=IMIN,IMAX
77              KA=-1
78              DO 40 K=IMIN,IMAX
79              KA=KA+1
80              IF(D(K).NE.E(J)) GOTO 40
81              X=(FLOAT(KA)-0.0)/10.0
82              Y=(FLOAT(JA)-0.0)/10.0
83              CALL SYMBOL(X,Y,0.07,11,0.0,-1)
84        40    CONTINUE
85              JA=JA+1
86        50    CONTINUE
87              CALL PLOTND
88              STOP
89              END
```

Appendix 6. Listing of a Fortran IV computer program for the prediction of secondary structure using the hydrophobicity profile method of Cid et al. (1982).

```
 1      C     *****************************************************
 2      C     *                                                   *
 3      C     *   SECONDARY STRUCTURE USING HYDROPHOBICITY SCALE  *
 4      C     *   OF CID ET AL.(1982) FEBS LETTERS 150:247        *
 5      C     *                                                   *
 6      C     *****************************************************
 7      C
 8      C     VARIABLES :
 9      C        A         : BUFFER VECTOR
10      C        C         : TWO DIMENSIONAL ARRAY CONTAINING THE PRIMARY AMINO
11      C                    ACID SEQUENCE OF THE PROTEIN
12      C        F         : VECTOR CONTAINING THE BULK HYDROPHOBICITY VALUES FOR
13      C                    THE PRIMARY SEQUENCE.
14      C        IMAX      : INTEGER VALUE OF THE LAST RESIDUE OF THE SEQUENCE
15      C        IMIN      : INTEGER VALUE OF THE FIRST RESIDUE OF THE SEQUENCE
16      C        XMAX      : LAST RESIDUE OF THE SEQUENCE TO BE EXAMINED
17      C        XMIN      : FIRST RESIDUE OF THE SEQUENCE TO BE EXAMINED
18      C
19      C     ** DATA ENTRY **
20      C        - SUBROUTINES AXIS AND SYMBOL ARE DOCUMENTED IN THE UBC PLOT
21      C          MANUAL.
22      C        - SINCE A MAXIMUM X AXIS OF 10 INCHES WAS USED ONLY A CERTAIN
23      C          PORTION OF ANY SEQUENCE CAN BE PLOTTED USUALLY CA. 100 RESIDUES.
24      C          TO DELIMIT THE RANGE OF RESIDUES TO BE EXAMINED XMIN AND XMAX
25      C          HAVE TO BE CHANGED (E.G. IF RESIDUES 1 TO 100 ARE TO BE PLOTTED
26      C          XMIN=O AND XMAX=100). FOR SUBSEQUENT COMPARISONS XMIN SHOULD BE
27      C          ST TO 100, 200, ETC. PLOTTING WILL START AT RESIDUE XMIN+1.
28      C        - PRIMARY AMINO ACID SEQUENCE IS ENTERED 16 DATA PER LINE 16I5
29      C        - AMINO ACID SEQUENCE IS DIGITIZED USING THE FOLLOWING CODE:
30      C          1(ALA), 2(ARG), 3(ASN), 4(ASP), 5(CYS), 6(GLN), 7(GLU), 8(GLY),
31      C          9(HIS), 10(ILE), 11(LEU), 12(LYS), 13(MET), 14(PHE), 15(PRO),
32      C          16(SER), 17(THR), 18(TRP), 19(TYR) AND 20(VAL).
33      C
34            INTEGER A,C
35            DIMENSION A(500),C(50,50),F(500)
36            READ(5,10) M,N
37       10   FORMAT(2I3)
38            READ(5,15) XMIN,XMAX
39       15   FORMAT(2F6.0)
40            READ(5,20) ((C(I,J),J=1,16),I=1,N)
41       20   FORMAT(16I5)
42            K=1
43            DO 30 I=1,N
44            DO 30 J=1,16
45            A(K)=C(I,J)
46            K=K+1
47       30   CONTINUE
48            CALL AXIS(0.0,0.0,'RESIDUE NUMBER',-14,7.0,0.0,XMIN,10.0)
49            CALL AXIS(0.0,0.0,'BULK HYDROPHOBICITY',19,6.0,90.0,10.0,1.0)
50            DO 40 K=1,M
51            IF(A(K).EQ.1) F(K)=12.28
52            IF(A(K).EQ.2) F(K)=11.49
53            IF(A(K).EQ.3) F(K)=11.00
54            IF(A(K).EQ.4) F(K)=10.97
55            IF(A(K).EQ.5) F(K)=14.93
56            IF(A(K).EQ.6) F(K)=11.28
57            IF(A(K).EQ.7) F(K)=11.19
58            IF(A(K).EQ.8) F(K)=12.01
```

```
59            IF(A(K).EQ.9) F(K)=12.84
60            IF(A(K).EQ.10) F(K)=14.77
61            IF(A(K).EQ.11) F(K)=14.10
62            IF(A(K).EQ.12) F(K)=10.80
63            IF(A(K).EQ.13) F(K)=14.33
64            IF(A(K).EQ.14) F(K)=13.43
65            IF(A(K).EQ.15) F(K)=11.19
66            IF(A(K).EQ.16) F(K)=11.26
67            IF(A(K).EQ.17) F(K)=11.65
68            IF(A(K).EQ.18) F(K)=12.95
69            IF(A(K).EQ.19) F(K)=13.29
70            IF(A(K).EQ.20) F(K)=15.07
71      40    CONTINUE
72            L=1
73            IMIN=IFIT(XMIN)
74            IMAX=IFIT(XMAX)
75            IMIN=IMIN+1
76            DO 50 I=301,369
77            Y=(F(I)-10.0)
78            X=(FLOAT(L)-0.0)/10.0
79            IF(L.EQ.1) CALL SYMBOL(X,Y,0.07,11,0.0,-1)
80            CALL SYMBOL(X,Y,0.07,11,0.0,-2)
81            L=L+1
82      50    CONTINUE
83            CALL PLOTND
84            STOP
85            END
```

Appendix 7. Listing of a Fortran IV computer program for the determination of mean residue ellipticity from CD spectral data.

```
 1     C   ******************************************************************
 2     C   *                                                                *
 3     C   *   THIS PROGRAM CALCULATES MOLECULAR ELLIPTICITY USING THE     *
 4     C   *            JASCO J-500A SPECTROPOLARIMETER                     *
 5     C   *                                                                *
 6     C   *   INPUT FROM THE TEXAS INSTRUMENTS DATA ACQUISITION UNIT       *
 7     C   *                                                                *
 8     C   ******************************************************************
 9     C
10     C
11     C   VARIABLES :
12     C       BASELN : BASELINE VOLTAGE READING
13     C       CDATA  : SAMPLE VOLTAGE CORRECTED FOR BASELINE
14     C       CONC   : CONCENTRATION OF THE SAMPLE IN GM/CM**3
15     C       DATA   : VOLTAGE READING FOR THE SAMPLE
16     C       DPBL   : NUMBER OF DATA READINGS FOR EACH BASELINE
17     C       ELLP   : ELLIPTICITY IN DEGREES
18     C       MOLELP : MOLECULAR ELLIPTICITY IN DEGREES CM**2/DECIMOLE
19     C       MOLWT  : MOLECULAR WEIGHT IN GM/MOLE
20     C       NUMBAS : NUMBER OF BASELINES USED
21     C       PATHLG : PATHLENGTH OF THE CUVETTE IN CM
22     C       PEAKHT : PEAK HEIGHT IN CM
23     C       SENSET : SENSITIVITY IN MILLIDEGREES/CM
24     C       WAVELG : WAVELENGTH IN NM
25         INTEGER DPBL,NUMBAS,MOLWT,I,J,K,N,Y
26         INTEGER PROTN(80)
27         REAL BASELN,DATA,CDATA,PEAKHT,ELLP,MOLELP,PATHLG
28         REAL CONC,SENSET,WAVELG
29         DIMENSION BASELN(11,10),DATA(11,10),MOLELP(200)
30         READ(5,10)NUMBAS
31     10  FORMAT(I2)
32         P=1
33     15  IF(P.GT.NUMBAS) GOTO 210
34         WRITE(6,20)
35     20  FORMAT('1','INFORMATION CONCERNING THE CD RUN :'//)
36         READ(5,30)SENSET
37     30  FORMAT(F6.2)
38         WRITE(6,40) SENSET
39     40  FORMAT('0','SENSITIVITY MILLIDEGREES/CM :',F6.3)
40         READ(5,50) PATHLG
41     50  FORMAT(F6.3)
42         WRITE(6,60) PATHLG
43     60  FORMAT('0','PATHLENGTH OF THE CUVETTE IN CM :',F6.3)
44         M=6
45         READ(5,70)((BASELN(I,J),J=1,10),I=1,M)
46     70  FORMAT(10F10.2)
47         READ(5,80)DPBL
48     80  FORMAT(I2)
49         Y=1
50     85  IF(Y.GT.DPBL) GOTO 200
51         WAVELG=240.0
52         READ(5,90) PROTN
53     90  FORMAT(80A1)
54         WRITE(6,100) PROTN
55    100  FORMAT('0','PROTEIN :',80A1)
56         READ(5,110) MOLWT
57    110  FORMAT(I6)
58         WRITE(6,120) MOLWT
```

```
59        120   FORMAT('0','MOLECULAR WEIGHT GM/MOLE :',I6)
60              READ(5,130) CONC
61        130   FORMAT(E9.3)
62              WRITE(6,140) CONC
63        140   FORMAT('0','CONCENTRATION GM/CM**3   :',E9.3)
64              WRITE(6,150)
65        150   FORMAT('0',4X,'DATA',8X,'BASELN',8X,'CDATA',9X,'PEAKHT',8X,
66              1'ELLP',7X,'MOLELP',8X,'WAVELG'///)
67              READ(5,70)((DATA(I,J),J=1,10),I=1,M)
68              K=1
69              DO 170 I=1,M
70              DO 160 J=1,10
71              CDATA=DATA(I,J)-BASELN(I,J)
72              PEAKHT=(CDATA+0.07)/2.42
73              PEAKHT=PEAKHT/10.0
74              ELLP=(SENSET*PEAKHT)/1000.0
75              MOLELP(K)=(ELLP*MOLWT)/(10.0*PATHLG*CONC)
76              WRITE(6,155)DATA(I,J),BASELN(I,J),CDATA,PEAKHT,ELLP,MOLELP(K),
77              1WAVELG
78        155   FORMAT('0',2X,F7.2,7X,F7.2,6X,F7.2,6X,F6.1,4X,E10.4,3X,E10.4,
79              17X,F5.1)
80              WAVELG=WAVELG-1.0
81              K=K+1
82        160   CONTINUE
83        170   CONTINUE
84              N=K-10
85              WRITE(6,180)
86        180   FORMAT('2','DATA INPUT FOR PROVENCHER PROGRAM'///)
87              WRITE(6,190) (MOLELP(K),K=1,N)
88        190   FORMAT('0',10F10.0)
89              Y=Y+1
90              GOTO 85
91        200   CONTINUE
92              P=P+1
93              GOTO 15
94        210   CONTINUE
95              STOP
96              END
```

Appendix 8. Listing of a Fortran IV computer program to calculate aver-
age hydrophobicity based on the algorithm of Bigelow
(1967).

```
 1      C  ************************************************
 2      C  *                                              *
 3      C  *        BIGELOW'S AVERAGE HYDROPHOBICITY       *
 4      C  *                                              *
 5      C  ************************************************
 6      C
 7      C  N = NUMBER OF AMINO ACIDS IN THE PROTEIN
 8      C  M = ARRAY OF THE AMINO ACIDS IN FORMAT 16I5(READING ACROSS)
 9      C  ND= NUMBER OF DATA CARDS
10      C  A = TEMPORARY ARRAY TO STORE THE AMINO ACID SEQUENCE
11      C  X = HYDROPHOBICITY VALUE ASSIGNED TO EACH AMINO ACID
12      C  AVG = BIGELOW'S AVERAGE HYDROPHOBICITY
13      C
14      C  THE NUMBER SYSTEM FOR THE AMINO ACIDS FOLLOWS THAT OF PHAM
15      C  UNIT 16 MUST BE ASSIGNED A TEMPORARY FILE INORDER TO RUN
16      C  THIS PROGRAM
17      C
18      C
19            REAL*8 X,TOTAL,AVG
20            INTEGER A,PROTN(80)
21            DIMENSION M(30,30),A(500)
22            READ(5,10) PROTN
23      10    FORMAT(80A1)
24            WRITE(6,15) PROTN
25      15    FORMAT('1',80A1)
26            READ(5,20) N,ND
27      20    FORMAT(2I3)
28            WRITE(6,25) N
29      25    FORMAT(' THE NUMBER OF AMINO ACIDS IN THE SEQUENCE =',I3)
30            READ(5,30) ((M(I,J),J=1,16),I=1,ND)
31      30    FORMAT(16I5)
32            REWIND 16
33            WRITE(16,40) ((M(I,J),J=1,16),I=1,ND)
34      40    FORMAT(I5)
35            REWIND 16
36            TOTAL=0.0
37            DO 50 K=1,N
38            READ(16,40) A(K)
39            IF(A(K).EQ.3.OR.A(K).EQ.4.OR.A(K).EQ.6.OR.A(K).EQ.7.OR.A(K).EQ.8
40           1.OR.A(K).EQ.9.OR.A(K).EQ.16) X=0.0
41            IF(A(K).EQ.18) X=3.00
42            IF(A(K).EQ.10) X=2.95
43            IF(A(K).EQ.19) X=2.85
44            IF(A(K).EQ.14) X=2.65
45            IF(A(K).EQ.15) X=2.60
46            IF(A(K).EQ.11) X=2.40
47            IF(A(K).EQ.20) X=1.70
48            IF(A(K).EQ.12) X=1.50
49            IF(A(K).EQ.13) X=1.30
50            IF(A(K).EQ.5)  X=1.00
51            IF(A(K).EQ.1)  X=0.75
52            IF(A(K).EQ.2)  X=0.75
53            IF(A(K).EQ.17) X=0.45
54            TOTAL=TOTAL+X
55      50    CONTINUE
56            AVG=(TOTAL*1000)/N
57            WRITE(6,60) AVG
58      60    FORMAT(' BIGELOW AVERAGE HYDROPHOBICITY', F12.4)
59            STOP
60            END
```