AN ABSTRACT OF THE THESIS OF

Larry A. Compton for the degree of Doctor of Philosophy in

Biochemistry and Biophysics presented on July 29, 1987.

Title: Circular Dichroism Spectroscopy and the Conformational Analysis of Globular Proteins in the Native and Denatured State

Abstract approved: Redacted for Privacy

W. Curtis Johnson

A simple matrix method for the analysis of protein circular dichroism (CD) spectra is presented and used to compute the secondary structure of globular proteins in their native and denatured states. The method uses singular value decomposition and generalized inverse algorithms, and a basis set of CD spectra for proteins whose secondary structures are known from x-ray crystallography. Inverse CD functions for five major secondary structures are constructed, and the fraction of each structure present in a protein is computed by taking the dot product of the inverse CD function with its CD spectrum. The method is tested by computating secondary structure for proteins in the basis set. Furthermore, the secondary structure of dihydrofolate reductase from T4 bacteriophage, a protein whose x-ray structure is not known, is predicted. CD analysis in conjunction with amino acid sequence information indicates that the T4 enzyme contains both parallel and antiparallel segments of β-sheet. This β-sheet mixture is found in several other proteins, and may represent a new subclass of tertiary

folding. Remaining sections are devoted to analyzing the CD of proteins at extreme pH or temperature to determine the structural changes associated with denaturation. Studies of pH denaturation in myoglobin and concanavalin-A indicate that α-helices are affected by acid but not base; whereas, β-sheet is rearranged by base but is unaffected by acid. Thermal denaturation of ribonuclease and papain results in the denaturation of α-helix, without disturbing β-sheet. Analysis of papain indicates that its two separate domains are acting independently during denaturation.

Circular Dichroism Spectroscopy and the Conformational Analysis of Globular Proteins in the Native and Denatured State

by

Larry A. Compton

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Completed July 29, 1987

Commencement June 1988

APPROVED:

Redacted for Privacy

Professor of Biochemistry and Biophysics in charge of major

Redacted for Privacy

Chairman of Department of Biochemistry and Biophysics

Redacted for Privacy

Dean of Graduate School

Date thesis is presented July 29, 1987

Typed by Barbara Hanson for Larry A. Compton

ACKNOWLEDGMENTS

This thesis is dedicated to my beloved wife

DEBRA K. COMPTON

.. who has endured.

Manuscripts:

1) This work was supported by PHS grant GM-21479 from the Institute of General Medical Sciences.

2) We thank Gerald W. Lasser of this department for his skillful help in the isolation and purification of T4 dihydrofolate reductase. This research was supported by PHS grant GM-21479 from the Institute of General Medical Sciences (WCJ), and PHS grant AI-15145 from the National Institute of Allergy and Infectious Diseases (CKM).

3) We thank Arazdordi Toumadje and Jim Scott of this laboratory for performing the SDS polyacrylamide gel electrophoresis. This research was supported by PHS Grant GM-21479 from the Institute of General Medical Sciences.

PREFACE

The following thesis consists of three manuscripts written in publishable form. The first manuscript has been published in Analytical Biochemistry, 155, (1986), pp. 155-167. The second manuscript has been accepted for publication in the Journal of Biological Chemistry. In the second manuscript, the Christopher Mathews laboratory was responsible for the isolation and purification of the T4 dihydrofolate reductase. The W. Curtis Johnson laboratory was responsible for the CD analysis of the enzyme. The third manuscript is currently being submitted for editorial review to Biochemistry. Since the manuscript format is being used in this thesis, the manuscripts have undergone only the minimum changes necessary to conform to thesis requirements.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CIRCULAR DICHROISM SPECTROSCOPY AND THE CONFORMATIONAL ANALYSIS OF GLOBULAR PROTEINS IN THE NATIVE AND DENATURED STATE

INTRODUCTION

Circular dichroism (CD) is an optical phenomenon caused by the unequal interaction of circularly polarized light with chiral molecules. In 1812 Biot was the first to describe the theoretical laws of chirality on optical phenomena. Some of the first empirical evidence for physical chirality came with the separation of a racemic mixture of sodium ammonium tartrate into right- and left-handed crystals by Pasteur in the 1840s. The CD for solutions of chiral molecules was demonstrated by Cotton in 1896. However, the invention of the Bunsen burner, with its easily workable monochromatic sodium flame, brought to an end this early interest in optical chirality as a function of wavelength. It wasn't until the 1950s, with the discovery that helical polypeptides give rise to optical chirality that there was a resurgence of interest in the CD of polymers. Since this time all of the major secondary structures found within globular proteins and polypeptides have been shown to display characteristic CD spectra in the far ultraviolet region (Yang et al., 1986). These include the Pauling-Corey α-helix, β-sheet (both antiparallel and parallel), and several types of β-turns.

Since each of the commonly recognized secondary structures found within globular proteins are represented by their own unique CD spectra below 260 nm, during the last thirty years many attempts have

been made to elucidate the folding patterns of proteins in solution using their CD spectra. Theoretically, the CD spectrum of any globular protein should consist of a linear combination of known CD spectra for each secondary structure which reflects their structural makeup. For this reason, the original techniques for analyzing the structure of proteins using circular dichroism spectroscopy involved curve-fitting analysis of the experimental CD spectrum of interest and a given set of CD spectra for synthetic polypeptides in each of the know secondary structures (Greenfield and Fasman, 1969; Brahms and Brahms, 1980). However, it has become increasingly more obvious that the CD spectra of synthetic polypeptides in the different secondary structural conformations may not truly represent the CD of similar structures in globular proteins. Helical regions within globular proteins are normally composed of small, constrained elements unlike that of the extended α-helix of polypeptides. β-sheets may contain varying amounts of β-turn and can be twisted in a variety of positions. In addition, other spectral characteristics of globular proteins (i.e. aromatic side chains, disulfide interactions) may significantly alter the shape and intensity of the protein CD spectrum. Therefore, many investigators have begun using advanced mathematical techniques which do not depend upon reference spectra of synthetic polypeptides, but rather correlate the CD spectra for a set of basis proteins with known secondary structure with the experimental CD for the protein of interest (Saxena and Wetlaufer, 1971; Chen et al., 1972; Chen et al., 1974; Chang et al., 1978; Bolotina et al., 1979).

In our laboratory we have developed such a method for the analysis

of protein CD spectra using matrix algorithms that do not require standard matrix inversion techniques. The method automatically considers all parameters that contribute to the basis CD spectra, whether they are explicitly considered or not. Limiting the degrees of freedom in the analysis to the information content in the CD of the basis set proteins avoids the instability normally associated with large basis sets. The first chapter in this report contains a detailed description of the original method and several improvements which increase the usability and practicality of the method. Over the past several years, our method has been used quite extensively to predict the secondary structure of globular proteins in neutral salts, whose secondary structures are or were not as yet defined by x-ray crystallography (Manavalan et al., 1984; Manavalan et al., 1986; Hennessey et al., 1987; Ausio et al., 1987; Eckhardt et al., 1987). The validity of the method is supported by the result that the structures of several of the proteins studied after analysis of the CD have been elucidated by x-ray analysis, and give crystallographic structures that are in excellent agreement with the CD work (Manavalan et al., 1984; Manavalan et al., 1986). The second chapter of this thesis deals with using our method to determine the secondary structure of the dihydrofolate reductase (DHFR) from T4 bacteriophage, a globular protein whose secondary structure is, as of yet, undetermined by x-ray crystallography. The T4 DHFR protein is of interest because it is one of several enzymes involved in folate biosynthesis and is a target site for antimicrobial and anticancer agents such as trimethoprim and methotrexate. Results indicate that the T4 protein contains both antiparallel and parallel β-sheet

structures. This is the case for several other well-characterized proteins, including thymidylate synthase, which binds a derivative of tetrahydrofolate, the product of DHFR metabolism. The work suggests a subclass of tertiary folding which contains both α+β (antiparallel β-sheet) and α/β (parallel β-sheet) structural elements.

In addition to the determination of secondary structure for the native state of proteins, CD spectroscopy has also been used to monitor the structural changes in globular proteins due to ligand binding (Manavalan et al., 1985), quaternary interactions (Hennessey et al., 1982), and most importantly, the structural transitions involved in denaturation. Although it is beyond the scope of this introduction to discuss all of the relevant work in denaturation involving CD spectroscopy, a literature review of the most important results are given in the third manuscript of this thesis.

Early attempts to investigate the various modes of denaturation in globular proteins (i.e., organic salts, pH, temperature, and surfacants) used CD spectra terminated on or before 190 nm, and in many cases quantitative analysis was not even attempted, the authors relying on inspection of spectral transitions to qualify particular conclusions. However, modern vacuum CD spectrometers which normally record spectra to 178 nm or below, yield spectra with the information content necessary for analysis to give secondary structure. In conjunction with analytical methods designed to elucidate the structure corresponding to a protein CD spectra of interest, this allows a look at both the type and amount of secondary structural transitions accompanying physical or chemical denaturation. Several investigations of this type have indicated that denaturation is

neither an "all or nothing" phenomenon, nor one which always leads to a more disordered state. Indeed, some types of chemical or pH denaturation suggest that the denatured molecule contains more ordered secondary structure than the original native protein.

Although the preliminary investigations using CD spectroscopy to follow the course of denaturation have brought to light several interesting findings, these reports normally involved using either a variety of denaturants on one particular protein or interest, or conversely, they studied the effect of one denaturing agent on a variety of model globular proteins. Both of these approaches share the same failing, the fact that generalizations concerning the overall structural transitions cannot be made unless both types of experiments are combined. In the final section of this thesis we discuss CD experiments designed to study the effect of pH and thermal denaturation on a variety of well-characterized globular proteins. Quantitative results for CD spectra corresponding to acid, alkaline, and thermally denatured proteins are compared to their native structures and one another in order to draw general conclusions concerning effect of pH and temperature on the stability of the major secondary structural elements found within globular proteins.

ANALYSIS OF PROTEIN CD SPECTRA FOR SECONDARY STRUCTURE USING A SIMPLE MATRIX MULTIPLICATION

Larry A. Compton and W. Curtis Johnson, Jr.

Department of Biochemistry and Biophysics

Oregon State University

Corvallis, Oregon 97331 USA

ABSTRACT

Inverse CD spectra are presented for each of the five major secondary structures of proteins: α-helix, antiparallel and parallel β-sheet, β-turn, and other (random) structures. The fraction of the each secondary structure in a protein is predicted by forming the dot product of the corresponding inverse CD spectrum, expressed as a vector, with the CD spectrum of the protein digitized in the same way. We show how this method is based on the construction of the generalized inverse from the singular value decomposition of a set of CD spectra corresponding to proteins whose secondary structures are known from X-ray crystallography. These inverse spectra compute secondary structure directly from protein CD spectra without resorting to least squares fitting and standard matrix inversion techniques. In addition, spectra corresponding to the individual secondary structures, analogous to the CD spectra of synthetic polypeptides, are generated from the five most significant CD eigenvectors.

INTRODUCTION

Circular dichroism (CD) spectroscopy has been shown to be quite sensitive to the secondary structure of globular proteins. For the past several decades many attempts have been made to correlate the CD spectrum of a globular protein with its x-ray structure, all finding various degrees of success. One technique for estimating the secondary structure in globular proteins has been curve-fitting analysis of a protein CD spectrum assuming a linear relationship between the spectrum of interest and a set of CD spectra corresponding to synthetic polypeptides in theoretically known conformations (Greenfield and Fasman, 1969; Brahms and Brahms, 1980). The original methods often considered only the α-helical, β-sheet, and random (other) conformations (Greenfield and Fasman, 1969). Improved accuracy was achieved by the introduction of one or several β-turns, and the extention of spectra to 165 nm (Brahms and Brahms, 1980).

However, the validity of such methods have been in doubt for some time. Synthetic polypeptides do not resemble real proteins whose structural elements may differ substantially from ideal models. For instance, the CD spectrum of deprotonated poly-L-lysine at pH 11 has been chosen as a model for the helical protein structure (Greenfield and Fasman, 1969). However, the CD spectra of other polypeptides in the helical conformation show considerable variation in the intensity and position of major CD bands. In addition, x-ray crystallographic analysis of globular proteins has indicated that helical regions within biopolymers occur as small, constrained segments compared to the flexible, extended structure of synthetic polypeptides. That this

influences the far UV protein CD spectrum is supported by the finding that the intensity of the helical CD spectrum is chainlength dependent (Goodman et al., 1969; Barela and Darnall, 1974). CD spectra representing the β-sheet have an unknown length and may contain a significant amount of β-turn. Random coil CD spectra may well not represent the amides in proteins that are not in the classified structures. Comparison of the various CD spectra measured for each type of structure reveals notable differences and brings forth questions concerning the actual structural conformations represented by these synthetic spectra. Furthermore, workers should consider the known optical activity of disulfide bonds and aromatic amino acids that may contribute to protein CD spectra in the ultraviolet region (Woody, 1978). On the other hand, it is clear that attempting to account for all of these factors in a standard least-squares analysis would involve many parameters and that any estimation of secondary structure would be unstable due to experimental error (Provencher and Glockner, 1981).

An alternative approach to this problem based on the construction of basis CD spectra derived from a set of CD spectra of proteins with known secondary structures has been used by many investigators. Saxena & Wetlaufer (1971) used three protein CD spectra to calculate the basis CD spectra for α-helix, β-sheet, and "random" coil. Chen et al., generated the three major CD basis spectra using five (Chen et al., 1972) and eight (Chen et al., 1974) proteins. Chang et al., (1978) analyzed 15 proteins and derived the CD basis spectra for α-helix, β-sheet, a general β-turn, and "random" structures. Bolotina et al. (1979) included β-turns and separated parallel and

antiparallel-β-sheet in their construction of reference spectra from six proteins. However, all of the spectra used in the above investigations were terminated on or before 190 nm and statistical analysis of globular protein CD spectra in the range of 190-240 nm has been shown to correlate only to the amount of α-helical structure (Barela and Darnall, 1974; Siegel et al., 1980; Manavalan and Johnson, 1985). Thus it is not surprising that CD data restricted to wavelengths longer than 190 nm cannot accurately estimate the β-sheet and "random" structure of the proteins they represent. In addition, x-ray analysis has shown that secondary structure depends on tertiary folding in the sense that the secondary structure in a protein may be distorted from its ideal conformation (Richardson, 1981).

Recently, CD spectra of proteins with known secondary structure have been used for analysis without reference to the CD spectra for classified structures (Provencher and Glockner, 1981; Hennessey and Johnson, 1981). This report describes analysis of protein CD spectra using modern matrix techniques that automatically consider all parameters, whether they are recognized or not, and avoids the instability normally associated with large amounts of noisy data. Previously (Hennessey and Johnson, 1981) we have described the use of the singular-value decomposition (SVD) theorem on a set of CD spectra for globular proteins over the range of 178-260 nm to produce a set of orthogonal CD eigenvectors, their corresponding singular values, and a coefficient matrix which fits each protein CD spectrum as a linear combination of CD eigenvectors. In our original work, the decomposed matrices were used in a standard least-squares procedure to calculate the secondary structure corresponding to a given protein CD spectrum.

However, this required the use of matrix diagonalization subroutines to compute the inverse of our CD eigenvectors. These complex computer programs are known for their instability, and are not considered to be good algorithms for computer work. The next two sections of this paper describe how the matrices resulting from SVD can be used to construct inverse CD spectra based on the generalized (Moore-Penrose) inverse theory which will generate structure directly from a CD spectrum. This method is based on the orthogonal and unitary properties of SVD and does not involve the use of standard diagonalization or inversion techniques. The inverse CD spectra derived from our protein basis set represent the reciprocal linear solution for obtaining the five major x-ray structures (α-helix, antiparallel β-sheet, parallel β-sheet, β-turn, and "other" structures) from a digitalized CD spectrum.

In addition, secondary structure spectra analogous to the CD spectra of synthetic polypeptides are produced using the SVD and generalized inverse methods on the x-ray structure matrix of our protein basis set. These spectra will be used to investigate the correlation between CD signal and secondary structure.

Readers who are not concerned with the theory can go directly to the section entitled "Analyzing Protein CD Spectra". They will find that simply dotting vectors that we provide for each secondary structure into their digitized CD spectra will predict the corresponding structure for their protein.

METHODS OF ANALYSIS

A variety of methods have been employed for the analysis of the number and composition of the independent components contained in a protein CD spectrum. The most popular approach has been the use of standard least-squares analysis to solve several linear equations simultaneously. The CD spectra measured for a number of proteins can be tabulated as a matrix of CD signals with the proteins labeling the columns and the various wavelengths labeling the rows. Table I gives the spectra for the sixteen proteins of known secondary structure from our earlier work digitized every 2.0 nm, as these spectra were not tabulated originally (Hennessey and Johnson, 1981). This matrix of protein CD data will be given the symbol $\underset{\sim}{C}$. Then, in terms of the original methods which employed synthetic polypeptides of presumably a single secondary structure as basis CD spectra, the linear equations in matrix notation are given by $\underset{\sim}{C} = \underset{\sim}{B}\underset{\sim}{F}$. Here $\underset{\sim}{B}$ is the set of polypeptide CD spectra for each secondary structure expressed as column vectors in analogy to $\underset{\sim}{C}$, and $\underset{\sim}{F}$ is the matrix which gives the fraction of each secondary structure in the protein CD data. The least-squares solution for this equation is $\underset{\sim}{F} = (\underset{\sim}{B}^T\underset{\sim}{B})^{-1}\underset{\sim}{B}^T\underset{\sim}{C}$, where the superscript T indicates the transpose of the matrix where the columns are simply written as the rows. In this treatment the CD spectra for each secondary structure, $\underset{\sim}{B}$, are known and fit to the measured spectra, $\underset{\sim}{C}$, to find the fraction of each secondary structure, the unknown $\underset{\sim}{F}$.

An alternative method is to choose proteins with known secondary structure so that $\underset{\sim}{F}$ is known. These known structures are then fit to

TABLE I. Digitized Protein CD Spectra from 178-260 nm at 2 nm Intervals (Δε Units)

ALPHA CHYMOTRYPSIN														
-0.43	0.11	0.60	1.02	1.18	1.11	0.72	0.00	-0.75	-1.51	-2.10	-2.51	-2.66	-2.61	-2.38
-2.13	-1.89	-1.67	-1.51	-1.35	-1.25	-1.19	-1.14	-1.16	-1.21	-1.27	-1.22	-0.91	-0.58	-0.40
-0.26	-0.18	-0.11	-0.06	-0.04	-0.04	-0.04	-0.04	-0.04	-0.04	-0.04	-0.04			
CYTOCHROME C														
1.42	2.00	2.31	2.30	2.32	2.46	2.81	3.29	3.72	3.68	2.56	1.19	-0.16	-1.40	-2.30
-2.82	-2.91	-2.98	-3.19	-3.38	-3.58	-3.71	-3.78	-3.59	-3.28	-2.83	-2.16	-1.55	-1.07	-0.64
-0.38	-0.20	-0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00			
ELASTASE														
-0.99	-0.50	0.19	0.55	0.55	0.19	-0.65	-1.40	-2.19	-2.82	-3.20	-2.95	-2.64	-2.49	-2.37
-2.20	-2.00	-1.76	-1.54	-1.32	-1.12	-0.93	-0.76	-0.60	-0.47	-0.34	-0.28	-0.24	-0.22	-0.20
-0.19	-0.15	-0.09	-0.07	-0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00			
HEMOGLOBIN														
2.85	3.45	3.95	4.80	6.00	8.15	10.70	13.05	14.10	13.75	11.00	5.90	1.50	-2.30	-5.05
-6.15	-6.60	-6.55	-6.55	-6.75	-7.00	-7.15	-7.20	-7.00	-6.50	-5.85	-4.60	-3.60	-2.70	-1.80
-1.15	-0.65	-0.35	-0.15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00			
LACTATE DEHYDROGENASE														
-0.25	0.52	1.48	2.84	4.59	6.35	8.21	8.99	8.69	7.61	5.22	2.70	0.24	-1.21	-2.49
-3.27	-3.55	-3.62	-3.69	-3.85	-4.03	-4.16	-4.22	-4.15	-3.85	-3.36	-2.72	-2.13	-1.61	-1.17
-0.79	-0.52	-0.34	-0.19	-0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00			
LYSOZYME														
-0.20	0.40	1.22	2.30	3.40	4.20	4.90	5.18	5.05	4.00	2.10	-0.10	-2.25	-3.70	-4.40
-4.57	-4.18	-3.89	-3.66	-3.56	-3.50	-3.45	-3.25	-3.03	-2.88	-2.69	-2.43	-2.09	-1.59	-1.14
-0.71	-0.50	-0.32	-0.20	-0.16	-0.13	-0.09	-0.03	0.00	0.00	0.00	0.00			
MYOGLOBIN														
4.65	5.90	7.15	9.60	12.40	14.65	16.45	17.40	16.45	13.75	9.00	3.60	-0.60	-4.00	-6.40
-7.40	-7.55	-7.25	-7.35	-7.60	-7.80	-8.10	-8.30	-8.05	-7.55	-6.55	-4.80	-3.75	-2.75	-1.70
-1.00	-0.55	-0.35	-0.20	-0.15	0.00	0.00	0.00	0.00	0.00	0.00	0.00			
PAPAIN														
1.48	2.24	2.76	3.45	3.92	4.03	3.50	2.07	0.35	-1.07	-1.71	-1.91	-1.97	-2.28	-2.74
-3.06	-2.98	-2.84	-2.69	-2.66	-2.69	-2.71	-2.73	-2.67	-2.47	-2.13	-1.72	-1.31	-0.96	-0.69
-0.46	-0.29	-0.14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00			

Table I continued

RIBONUCLEASE-A														
0.30	0.59	1.00	1.30	1.39	1.49	1.79	2.00	2.00	1.72	0.80	-0.22	-1.34	-2.20	-2.81
-3.18	-3.21	-3.14	-3.09	-3.02	-2.87	-2.68	-2.39	-2.04	-1.63	-1.22	-0.82	-0.50	-0.25	-0.02
0.05	0.07	0.09	0.07	0.06	0.06	0.05	0.05	0.04	0.04	0.03	0.00			
SUBTILISIN BPN'														
-0.55	0.03	0.79	1.79	2.84	3.89	4.85	5.07	4.54	3.31	1.61	0.07	-0.78	-1.41	-1.90
-2.21	-2.39	-2.51	-2.61	-2.70	-2.72	-2.70	-2.60	-2.43	-2.19	-1.86	-1.52	-1.21	-0.92	-0.67
-0.47	-0.30	-0.20	-0.13	-0.04	-0.02	0.00	0.00	0.00	0.00	0.00	0.00			
FLAVODOXIN														
-1.20	-1.00	-0.10	0.60	1.40	2.20	3.00	3.80	4.20	4.00	3.20	2.10	0.90	-0.60	-1.60
-2.10	-2.40	-2.60	-2.80	-2.80	-2.90	-3.00	-2.90	-2.70	-2.40	-2.00	-1.60	-1.30	-1.10	-0.80
-0.60	-0.50	-0.40	-0.30	-0.20	-0.10	0.00	0.00	0.00	0.00	0.00	0.00			
GLYCERAL.-3-P-DEHDROGENASE														
-0.30	0.20	0.80	1.50	2.30	3.40	4.10	4.40	4.10	3.30	1.90	0.60	-0.80	-1.60	-2.20
-2.50	-2.80	-2.80	-2.90	-3.00	-3.10	-3.20	-3.10	-3.00	-2.80	-2.40	-2.00	-1.60	-1.30	-1.00
-0.70	-0.50	-0.40	-0.30	-0.20	-0.10	0.00	0.00	0.00	0.00	0.00	0.00			
PREALBUMIN														
-3.20	-2.70	-1.90	-1.10	-0.10	1.00	1.90	2.50	2.90	2.90	2.30	1.30	0.30	-0.40	-1.00
-1.30	-1.50	-1.60	-1.70	-1.80	-1.70	-1.60	-1.40	-1.20	-1.00	-0.70	-0.50	-0.30	-0.20	-0.10
-0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00			
SUBTILISIN NOVO														
-0.40	0.40	1.00	1.60	2.50	3.80	5.10	6.00	6.30	5.80	4.60	3.10	1.70	0.20	-1.00
-1.80	-2.50	-2.90	-3.00	-3.20	-3.30	-3.40	-3.30	-3.20	-2.90	-2.60	-2.20	-1.90	-1.60	-1.20
-0.80	-0.50	-0.30	-0.20	-0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00			
TRIOSE PHOSPHATE ISOMERASE														
1.20	1.80	3.00	4.30	6.10	7.60	9.10	10.30	10.50	9.40	6.60	4.00	1.10	-1.20	-3.00
-4.00	-4.40	-4.50	-4.70	-5.00	-5.20	-5.20	-5.10	-4.80	-4.40	-3.90	-3.20	-2.50	-1.90	-1.30
-0.90	-0.60	-0.30	-0.20	-0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00			
POLY GLUTAMIC ACID														
10.30	12.50	17.10	21.60	26.20	29.70	30.60	29.70	23.10	14.70	7.50	-0.50	-5.50	-10.20	-12.70
-12.50	-12.10	-11.90	-11.80	-12.00	-12.30	-13.00	-13.30	-12.40	-10.60	-8.80	-7.20	-5.50	-4.10	-2.50
-1.30	-0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00			

the measured CD spectra $\underset{\sim}{C}$ by the unknown inverse matrix $\underset{\sim}{X}$, so that $\underset{\sim}{F}=\underset{\sim}{X}\underset{\sim}{C}$. The problem then is to derive the linear solutions, $\underset{\sim}{X}$, to a large set of protein CD data. The standard least-squares solution to this inverse linear equation is $\underset{\sim}{X}=\underset{\sim}{F}\underset{\sim}{C}^{T}(\underset{\sim}{C}\underset{\sim}{C}^{T})^{-1}$. The inverse matrix can then be used to find the amount of secondary structure, $\underset{\sim}{F}'$, corresponding to the CD of other proteins, $\underset{\sim}{C}'$. Thus the unknown $\underset{\sim}{F}'$ is found from the measured $\underset{\sim}{C}'$ from $\underset{\sim}{F}'=\underset{\sim}{F}\underset{\sim}{C}^{T}(\underset{\sim}{C}\underset{\sim}{C}^{T})^{-1}\underset{\sim}{C}'$. However, when the number of protein CD spectra in $\underset{\sim}{C}$ is larger than the number of independent variables in the data, this method can be unstable.

Provencher and Glockner (1981) used Ridge Regression (Draper and Van Nostrand, 1979) to avoid instability when the number of CD spectra for proteins with known structure exceeds the number independent variables. Our laboratory attacked this problem (Hennessey and Johnson, 1981) by using the SVD theorem (Noble and Daniel, 1977). This theorem states that any m x n matrix, $\underset{\sim}{A}$, can be decomposed into the product of three matrices: 1) an m x m unitary column matrix of orthogonal eigenvectors, $\underset{\sim}{U}$, for $\underset{\sim}{A}\underset{\sim}{A}^{T}$; 2) a m x n matrix, $\underset{\sim}{S}$, with singular values on the main diagonal and zeros elsewhere; and 3) a n x n unitary row matrix of coefficients, $\underset{\sim}{V}^{T}$, whose entries fit the eigenvector matrix to the original data set. The columns of $\underset{\sim}{V}$ are the eigenvectors of $\underset{\sim}{A}^{T}\underset{\sim}{A}$ and the singular values are the positive square roots of the eigenvalues common to both $\underset{\sim}{A}\underset{\sim}{A}^{T}$ and $\underset{\sim}{A}^{T}\underset{\sim}{A}$. Since $\underset{\sim}{U}$ and $\underset{\sim}{V}^{T}$ are unitary, $\underset{\sim}{U}\underset{\sim}{U}^{T}=\underset{\sim}{I}$ and $\underset{\sim}{V}^{T}\underset{\sim}{V}=\underset{\sim}{I}$ where $\underset{\sim}{I}$ is the unit matrix. Although theoretically the $\underset{\sim}{U}$ matrix is m x m, in practice any eigenvector for which m > n will correspond to a singular value of 0, and is not produced by many of the currently available SVD subroutines. Therefore, the $\underset{\sim}{U}$ matrix as computed is m x n and $\underset{\sim}{S}$ is a square n x n

matrix.

In our application of this technique we start with a 42 x 16 data matrix, $\underset{\sim}{C}$, containing the CD spectra of 15 globular proteins and one polypeptide in the range of 178-260 nm at 2.0 nm intervals (Table I). This data matrix is decomposed using the SVD method into a matrix of 16 CD eigenvectors, a diagonal matrix of singular values, and a 16 x 16 matrix of coefficients that fit the CD eigenvectors to their corresponding CD spectra, such that $\underset{\sim}{C} = \underset{\sim\sim}{U}\underset{\sim\sim}{S}\underset{\sim}{V}^T$. Thus each of the protein CD spectra in $\underset{\sim}{C}$ can be expressed as a linear combination of eigenvectors weighted by their corresponding singular value $\underset{\sim\sim}{US}$. Subroutines for carrying out SVD are available at most computer centers, and a version may be found in Forsythe et al. (1977).

We see that the relative importance of each CD eigenvector as a component of the original CD spectrum is proportional to its corresponding singular value. The singular values in this case are: 116.8, 24.4, 14.9, 6.7, 4.0, 2.7, 1.4, 1.1, 1.0, 0.9, 0.6, 0.5, 0.3, 0.3, 0.3, 0.2. For perfect data the number of nonzero singular values indicates the number of independent components within the experimental data set. However, our real CD spectra contain signal noise and random error so no singular values are identically zero. However, the information content of the CD spectra comprising $\underset{\sim}{C}$ is limited, so it is reasonable to expect that only a few orthogonal CD eigenvectors (corresponding to the largest singular values) would be needed to reconstruct our original CD spectra with significance above the noise and random error of the data.

Hennessey & Johnson (1981) demonstrated that the five most significant CD eigenvectors from the decomposition give

reconstructions of the original CD spectra which are within two standard deviations of the measured values and provide for significant fine structure. Additional CD eigenvectors offer little improvement. Since only five CD eigenvectors will reconstruct the original CD spectra of our data set within experimental error, we can extract no more than five independent types of information from a given CD spectrum.

In our original application of the SVD method, the matrix of CD eigenvectors from our basis set proteins, $\underset{\sim}{U}$, and their corresponding singular values, $\underset{\sim}{S}$, were used in a standard least-squares approach analogous to the techniques used in previous studies. In this method the coefficient matrix, $\underset{\sim}{V'}^T$, of the experimental CD spectrum, $\underset{\sim}{C'}$, was determined by solving the linear equation, $\underset{\sim}{C'}=\underset{\sim\sim}{USV'}^T$ The standard least-squares solution was given by $\underset{\sim}{V'}^T=[(\underset{\sim\sim}{US})^T\underset{\sim\sim}{US}]^{-1}(\underset{\sim\sim}{US})^T\underset{\sim}{C'}$. The calculated coefficients, $\underset{\sim}{V'}^T$, could be used to derive the secondary structure of the globular protein of interest using the relationship $\underset{\sim}{F'}\underset{\sim}{V'}=\underset{\sim\sim}{FV}$ where the matrix, $\underset{\sim}{F}$, represents the x-ray structure matrix of our original CD data set, and $\underset{\sim}{V}$, their corresponding coefficients. The unknown secondary structures are then given by $\underset{\sim}{F'}=\underset{\sim\sim}{FVV'}^T$. Stability was achieved by using only the five most significant singular values and CD eigenvectors. This produces a rank reduced $\underset{\sim}{U}_5$ eigenvector matrix and a square diagonal matrix $\underset{\sim}{S}_5$ which describe our original CD data in terms of the five major components. However, this method is dependent upon the same types of diagonalization and linear inversion subroutines as the methods mentioned earlier. This is a potentially inaccurate procedure. Now we discuss the use of a matrix technique, the generalized (Moore-Penrose) inverse theorem which was

specifically designed to avoid this problem.

The Generalized Inverse Matrix

The three product matrices from the decomposition of our CD data matrix share important mathematical properties which can be used in conjunction with standard matrix multiplication techniques to produce the least-squares solution to a set of linear equations, without the use of matrix diagonalization or inversion subroutines. Since the $\underset{\sim}{U}$ and $\underset{\sim}{V}^T$ eigenvector matrices are orthogonal and unitary, $\underset{\sim}{U}\underset{\sim}{U}^T=\underset{\sim}{U}^T\underset{\sim}{U}=\underset{\sim}{I}$ and $\underset{\sim}{V}\underset{\sim}{V}^T=\underset{\sim}{V}^T\underset{\sim}{V}=\underset{\sim}{I}$. In addition, the singular value matrix is diagonal and it is easy to define a reciprocal matrix, $\underset{\sim}{S}^+$, such that $\underset{\sim}{S}\underset{\sim}{S}^+=\underset{\sim}{S}^+\underset{\sim}{S}=\underset{\sim}{I}$. Here $\underset{\sim}{S}^+$, called the pseudoinverse of $\underset{\sim}{S}$, has nonzero entries only on the diagonal and they are the reciprocals of their corresponding singular values.

These properties of SVD provide an easy way to produce the inverse CD spectra matrix, $\underset{\sim}{X}$, which solves the set of equations $\underset{\sim}{F}=\underset{\sim}{X}\underset{\sim}{C}=\underset{\sim}{X}(\underset{\sim}{U}\underset{\sim}{S}\underset{\sim}{V}^T)$. Multiplying by $\underset{\sim}{V}$, $\underset{\sim}{S}^+$, and $\underset{\sim}{U}^T$, respectively, gives $\underset{\sim}{X}=\underset{\sim}{F}(\underset{\sim}{V}\underset{\sim}{S}^+\underset{\sim}{U}^T)$. This inverse simply involves taking transposes (columns become rows) and a pseudoinverse (diagonal elements are replaced by their reciprocal). It is the solution to $\underset{\sim}{X}\underset{\sim}{C}=\underset{\sim}{F}$. The construction $\underset{\sim}{V}\underset{\sim}{S}^+\underset{\sim}{U}^T$ is called a generalized or Moore-Penrose inverse. Once obtained, $\underset{\sim}{X}$ can be used to estimate the secondary structure of other proteins from their CD spectra $\underset{\sim}{C}'$ using $\underset{\sim}{X}\underset{\sim}{C}'=\underset{\sim}{F}'$. Note that $\underset{\sim}{X}$ must be constructed using only the five most significant singular values to eliminate the effect of noisy data that leads to instability.

Figure 1 shows the set of inverse CD spectra corresponding to

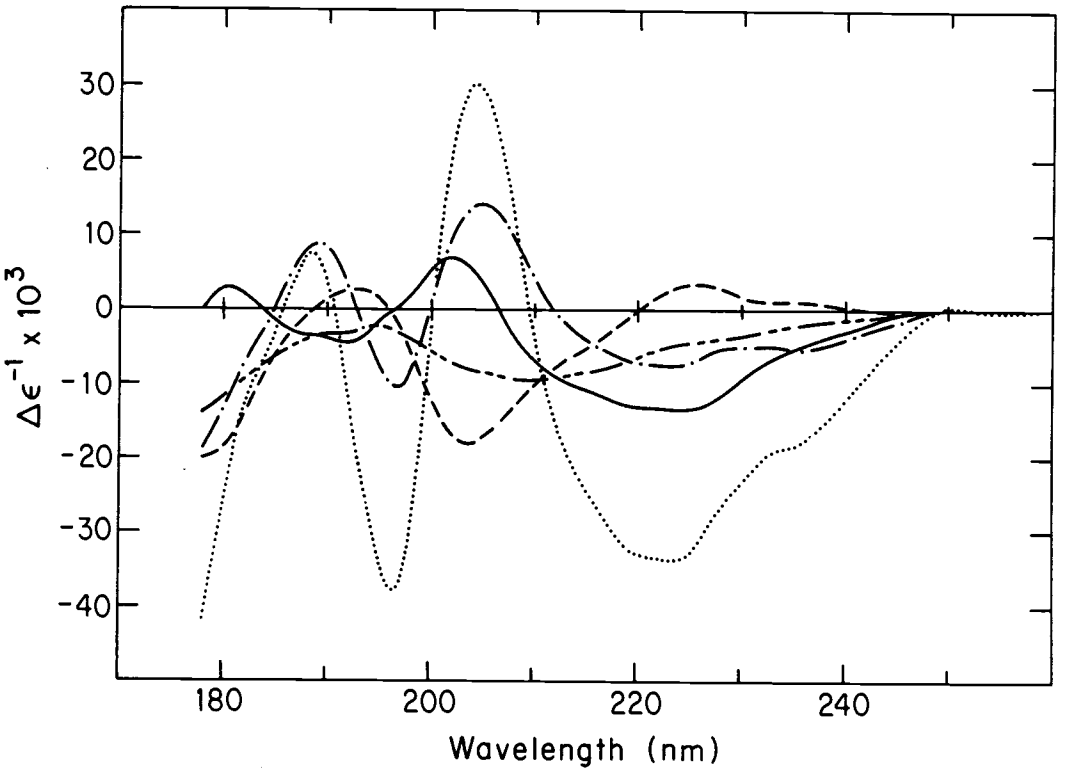

Figure 1. Generalized inverse CD spectra for five major secondary structures from 178-260 nm: α-helix (———), antiparallel β-sheet (----), parallel β-sheet (-•-), β-turn (— -- —), other (random) structure (••••).

secondary structure that are constructed using the generalized inverse method on the five most significant CD eigenvectors of our singular value decomposition. The units of these functions are $\Delta\varepsilon^{-1}$. The inverse CD spectra have a number of interesting features. First, they undergo oscillation from maxima to minima in a synchronous and systematic fashion. Inspection indicates that, in general, five major regions of high intensity alternate with nodal zones. Second, since the intensity of any particular region within an inverse CD basis spectrum is proportional to the amount of structural information available in a given CD spectrum, regions of high intensity reflect a large amount of structural knowledge within this area of the corresponding CD spectrum. In general, high intensity maxima and minima occur at 178-180, 188-190, 194-196, 206-208, and 222-224 nm for our inverse CD spectra. It is more than coincidence that these wavelengths are also the points of high intensity for many protein CD spectra. Finally, the number of regions of high intensity is the same as the number of CD eigenvectors used to construct them. A discussion of the generalized inverse matrix and a subroutine for rank reducing $\underset{\sim}{V}\underset{\sim}{S}^{T}\underset{\sim}{U}^{T}$ are available in Forsythe et al. (1977).

Secondary Structure Spectra

The SVD method can also be applied to the decomposition of our x-ray structure matrix, $\underset{\sim}{F}$, in order to produce spectra corresponding to secondary structures. For the X-ray structure matrix, $\underset{\sim}{F}$, used in this analysis, refer to Hennessey & Johnson (1981). These secondary structure spectra are the calculated equivalent of the CD spectra

obtained for synthetic polypeptides except that they optimize the fit in the least squares sense. They are the solutions to the linear equation used by early workers, $\underset{\sim}{C}=\underset{\sim\sim}{BF}$, where $\underset{\sim}{B}$ is now the unknown. Using the SVD and generalized inverse methods, the least-squares solution is given by $\underset{\sim}{B}=\underset{\sim}{C}(\underset{\sim}{V}_F\underset{\sim}{S}^{+}_F\underset{\sim}{U}^{T}_F)$, where $\underset{\sim}{F}=\underset{\sim}{U}_F\underset{\sim}{S}_F\underset{\sim}{V}^{T}_F$. In order to produce secondary structure basis spectra in which the noise and random error of our original CD spectra have been eliminated, CD spectra reconstructed from the five most significant CD eigenvectors are substituted for the data matrix $\underset{\sim}{C}$.

Figure 2 gives the secondary structure basis spectra constructed from our original CD spectra. We see that they are qualitatively similar to the CD spectra of synthetic polypeptides in similar conformations. Statistically, these secondary structure spectra represent the multiplicative inverse to the generalized inverse matrix. For perfectly correlated CD data, where each spectrum in $\underset{\sim}{B}$ represents only a single secondary structure, $\underset{\sim\sim}{XB} = \underset{\sim}{I}$, since $\underset{\sim\sim}{XC} = \underset{\sim}{F}$ and $\underset{\sim}{C} = \underset{\sim\sim}{BF}$. This is the case by definition for methods which exclusively employ polypeptide CD spectra as structural references.

The construction of secondary structure spectra from protein CD spectra, and the relationship between these spectra and the generalized inverse CD matrix, provide for a novel method to investigate the correlation between the x-ray structure of a globular protein and its CD spectrum. Normally, studies of the correlation between CD signal and secondary structure have involved the use of Pearson product-moment correlation coefficients (Baker and Isenberg, 1976). In the Pearson method, the correlation coefficients $\underset{\sim}{r}_i$, are given by

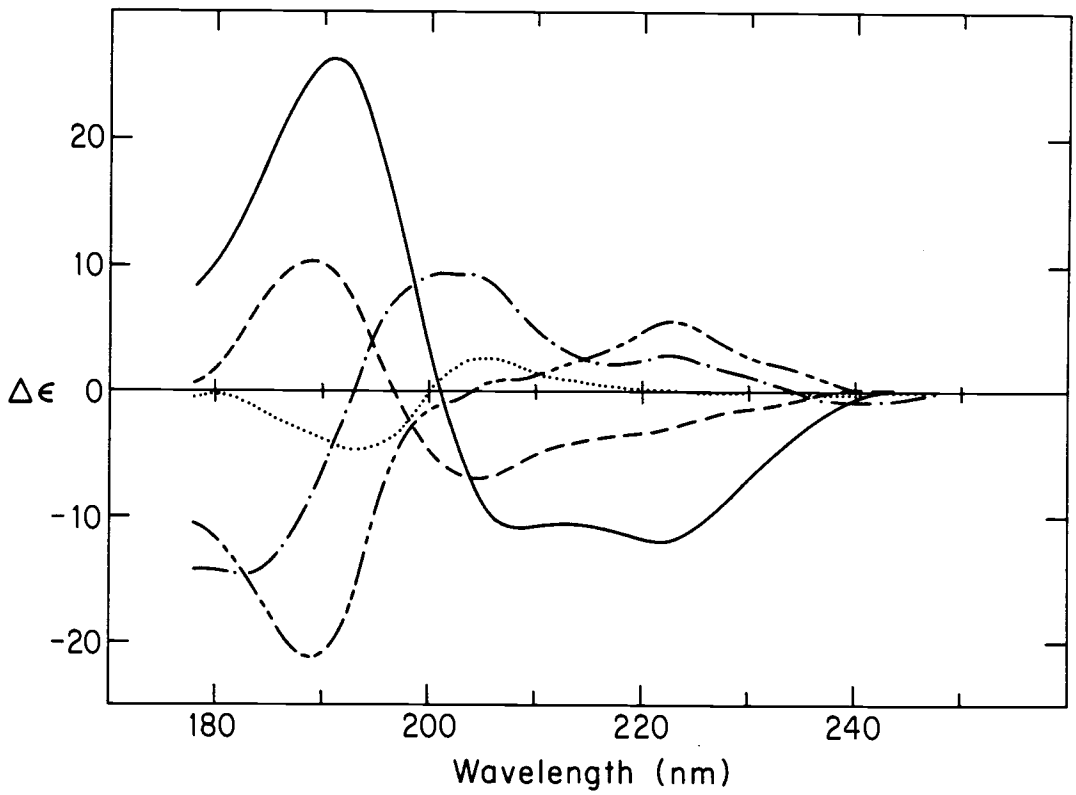

Figure 2. Secondary structure basis spectra for five major secondary structures from 178-260 nm: α-helix (———), antiparallel β-sheet (----), parallel β-sheet (-•-), β-turn (— -- —), other (random) structure (••••).

$$r_i = \frac{N\sum_i x_i y_i - \sum_i x_i \sum_i y_i}{((N\sum_i x_i^2 - (\sum_i x_i)^2)^{1/2} \cdot (N\sum_i y_i^2 - (\sum_i y_i)^2)^{1/2}}$$

where the summations range from i=1 to N, the number of measurements. The quantities x_i and y_i are the observed and computed fractions of secondary structure, respectively. A correlation coefficient of +1 indicates a successful prediction, a value of 0 equals a random assignment, and a coefficient of -1 means inverse agreement between observed and computed results. To date, most methods for the analysis of CD spectra have used this summation. The Pearson equation, however, is based upon the result of a more fundamental concept. For simple linear equations, the correlation coefficient, r_i, is calculated as the product of the observed mathematical function and its inverse, $r_i=(m_i m_i')^{1/2}$ where m_i and m_i' are the slopes of the original and inverse linear equations. In our application, where $\underset{\sim}{\underset{\sim}{X}}\underset{\sim}{C} = \underset{\sim}{F}$, and $\underset{\sim}{C} = \underset{\sim}{\underset{\sim}{B}}\underset{\sim}{S}$, the product-moment correlation coefficients r_i, correspond to the square roots of the diagonal elements of $\underset{\sim}{\underset{\sim}{X}}\underset{\sim}{\underset{\sim}{B}}$, the matrix which gives the secondary structure corresponding to each secondary structure spectrum.

Table II summarizes the calculation of correlation coefficients using both the summation and product-moment methods. The summation method yields very good results for α-helical, anti-parallel β-sheet, and random structures, and fair results for the remaining structures. The product-moment method yields similar results for helical structures, but indicates a somewhat lower correlation between the other secondary structures. In general, both methods yield

TABLE II. Correlation Coefficients[1]

	H	A	P	T	O
r_{sum}[1]	0.98	0.74	0.82	0.97	0.86
r_{pm}[2]	1.00	0.66	0.75	0.81	0.78

[1] H = α-helix; A = antiparallel β-sheet; P = parallel β-sheet; T = β-turns; O = other structure.

[2] Summation method used.

[3] Product-moment matrix method used.

coefficients which are much more favorable than previously documented (Provencher and Glockner, 1981; Chang et al., 1978; Hennessey and Johnson, 1981).

Analyzing Protein CD Spectra

The generalized inverse provides a particularly simple method for analyzing the CD spectrum of a protein for secondary structure. Each secondary structure corresponds to a vector, which is a particular inverse CD spectrum. The fraction of the secondary structure corresponding to each inverse spectrum is found by taking the dot product of the vector with the CD spectrum of the protein digitized in the same way. The inverse CD spectra depend on the wavelength range of the data. We provide the inverses for the ranges 178-260 nm, 180-260 nm, 182-260 nm, and 184-260 nm in Tables III-VI, respectively. Data truncated above 184 nm should not be used for analyses of protein CD spectra for secondary structure. Clearly, the further the CD spectra extend into the vacuum UV the more reliable the analysis. There is no constraint on the sum of structures and the fraction of a structure is permitted to be negative. The absence of large, negative fractions of structure and sums of structure between 0.9 and 1.1 are indicative of a successful analysis. If a protein has structure that is outside of the vector space provided by the sixteen proteins that are used to construct the inverse spectra, then the analysis can not be successful.

Of course the inverses for the five secondary structures can be combined into a row matrix, and a number of protein CD spectra can be

TABLE III. Generalized Inverses for Various Secondary Structures[4] 178-260 nm at 2 nm Intervals ($\Delta\varepsilon^{-1}x10^3$).

H:	-0.04	3.01	2.20	-0.24	-2.69	-3.46	-3.64	-4.94	-3.07	-0.26	1.23	5.81	7.05	5.24	1.28
	-3.68	-7.19	-9.04	-10.28	-11.28	-12.56	-13.04	-13.21	-13.52	-13.05	-11.70	-9.33	-7.25	-5.95	-4.83
	-3.78	-2.91	-1.98	-0.97	-0.50	-0.04	0.15	0.02	-0.04	-0.04	-0.03	-0.01			
A:	-19.90	-19.09	-14.41	-8.70	-3.85	-0.85	1.39	2.54	2.50	-0.20	-5.60	-12.13	-17.61	-17.98	-16.01
	-13.17	-9.93	-7.64	-5.94	-4.25	-2.07	-0.27	1.87	3.20	3.24	3.00	1.59	1.00	1.17	0.96
	0.68	0.40	0.04	-0.28	-0.27	-0.40	-0.30	-0.08	0.02	0.02	0.00	-0.05			
P:	-18.87	-13.41	-7.02	-2.74	3.23	8.30	9.13	3.16	-4.25	-10.33	-9.95	1.58	9.80	14.15	13.66
	9.50	3.89	-0.59	-3.04	-4.80	-6.21	-7.29	-7.25	-7.63	-6.64	-4.90	-5.30	-4.64	-5.15	-5.39
	-4.73	-3.82	-2.84	-1.99	-1.20	-0.26	0.17	-0.04	-0.12	-0.12	-0.09	0.01			
T:	-13.86	-11.69	-9.23	-7.08	-5.54	-4.27	-3.18	-3.43	-2.38	-2.31	-4.42	-5.26	-7.45	-8.14	-8.74
	-9.58	-9.46	-9.06	-8.66	-8.10	-7.37	-6.43	-5.11	-4.48	-4.27	-3.79	-3.57	-2.93	-2.23	-1.89
	-1.57	-1.33	-1.08	-0.73	-0.46	-0.28	-0.11	-0.04	-0.01	-0.01	-0.02	-0.04			
O:	-41.99	-25.26	-12.23	-6.95	0.58	7.71	5.60	-12.74	-28.45	-38.45	-34.12	0.11	22.31	31.28	26.54
	11.68	-4.19	-15.73	-22.15	-26.65	-30.98	-33.52	-33.28	-34.30	-31.51	-25.98	-23.96	-19.62	-18.83	-17.97
	-15.21	-12.20	-8.80	-5.53	-3.29	-0.69	0.54	-0.07	-0.31	-0.31	-0.24	-0.03			

[4]Abbreviations as in Table II

TABLE IV. Generalized Inverses for Various Secondary Structures[5] 180-260 nm at 2 nm Intervals ($\Delta\varepsilon^{-1}x10^{3}$).

H:	2.51	1.86	-0.40	-2.64	-3.23	-3.34	-4.77	-3.08	-0.47	0.97	5.72	7.09	5.41	1.49	-3.51
	-7.10	-9.02	-10.29	-11.31	-12.61	-13.10	-13.25	-13.58	-13.11	-11.75	-9.42	-7.34	-6.05	-4.96	-3.89
	-3.00	-2.06	-1.03	-0.54	-0.05	0.15	0.02	-0.04	-0.04	-0.03	-0.01				
A:	-28.67	-21.50	-13.26	-5.48	-0.07	3.63	4.26	3.09	-1.31	-7.90	-13.39	-18.44	-17.73	-15.32	-12.78
	-10.19	-8.47	-6.98	-5.33	-2.94	-0.94	1.76	3.25	3.40	3.37	1.34	0.56	0.50	0.04	-0.22
	-0.40	-0.68	-0.89	-0.65	-0.58	-0.34	-0.10	0.00	0.00	-0.01	-0.05				
P:	-23.80	-14.71	-7.56	1.78	9.66	12.13	5.31	-3.72	-11.95	-12.92	0.08	9.04	14.86	14.95	10.36
	3.91	-1.27	-4.00	-5.84	-7.12	-8.04	-7.45	-7.75	-6.73	-4.76	-5.87	-5.42	-6.17	-6.69	-5.96
	-4.91	-3.80	-2.77	-1.70	-0.50	0.12	-0.08	-0.15	-0.15	-0.11	0.00				
T:	-19.27	-14.73	-10.52	-6.55	-3.30	-1.10	-1.94	-2.04	-3.51	-6.54	-6.26	-7.87	-7.58	-7.81	-8.96
	-9.47	-9.62	-9.44	-8.95	-8.11	-7.05	-5.32	-4.58	-4.29	-3.60	-3.89	-3.38	-2.88	-2.74	-2.39
	-2.05	-1.71	-1.25	-0.79	-0.43	-0.14	-0.06	-0.03	-0.03	-0.03	-0.04				
O:	-46.39	-28.11	-17.11	-2.90	9.84	11.18	-8.60	-27.19	-41.25	-39.72	-2.92	20.38	32.18	28.59	12.97
	-4.43	-17.24	-24.13	-28.73	-32.72	-34.89	-33.49	-34.37	-31.58	-25.68	-25.09	-21.18	-20.85	-20.55	-17.64
	-14.38	-10.73	-7.10	-4.28	-1.17	0.43	-0.15	-0.36	-0.36	-0.29	-0.05				

[5]Abbreviations as in Table II

TABLE V. Generalized Inverses for Various Secondary Structures[6] 182-260 nm at 2 nm Intervals ($\Delta\epsilon^{-1}x10^3$).

H:	3.08	0.45	-2.21	-3.18	-3.60	-5.01	-3.27	-0.47	1.19	5.93	7.33	5.45	1.45	-3.51	-7.04
	-8.92	-10.19	-11.22	-12.57	-13.11	-13.36	-13.69	-13.19	-11.83	-9.42	-7.30	-5.99	-4.85	-3.78	-2.90
	-1.96	-0.94	-0.49	-0.03	0.16	0.02	-0.04	-0.04	-0.03	-0.01					
A:	-38.67	-24.85	-10.54	0.54	8.40	8.32	5.51	-2.09	-12.03	-16.57	-21.58	-17.89	-14.05	-11.99	-10.41
	-9.37	-7.98	-6.28	-3.31	-0.75	3.10	4.50	4.15	4.09	0.70	-0.69	-1.09	-2.13	-2.26	-2.22
	-2.40	-2.29	-1.53	-1.02	-0.51	-0.21	-0.06	-0.06	-0.07	-0.08					
P:	-28.68	-17.09	-2.52	10.03	16.09	8.56	-1.69	-12.42	-16.31	-2.75	6.14	14.66	16.05	11.03	3.83
	-1.75	-4.46	-6.25	-7.10	-7.62	-6.21	-6.84	-6.52	-4.79	-6.93	-6.97	-7.87	-8.83	-7.87	-6.59
	-5.38	-4.00	-2.46	-0.89	-0.04	-0.19	-0.21	-0.21	-0.17	-0.03					
T:	-29.15	-19.94	-10.10	-1.83	3.79	1.94	-0.18	-4.89	-10.74	-9.08	-10.16	-7.11	-6.07	-7.71	-9.29
	-10.23	-10.26	-9.81	-8.55	-7.11	-4.43	-3.77	-3.81	-3.07	-4.57	-4.57	-4.40	-4.78	-4.30	-3.72
	-3.26	-2.51	-1.58	-0.81	-0.27	-0.15	-0.09	-0.09	-0.08	-0.06					
O:	-58.35	-37.42	-11.51	11.55	20.56	-1.05	-22.82	-42.82	-47.62	-9.23	14.28	32.06	31.36	14.83	-4.34
	-18.18	-25.11	-29.65	-32.76	-34.11	-30.92	-32.46	-31.08	-25.59	-27.35	-24.51	-24.57	-25.25	-21.86	-18.06
	-14.19	-9.81	-5.96	-2.03	0.08	-0.38	-0.51	-0.51	-0.41	-0.10					

[6]Abbreviations as in Table II

TABLE VI. Generalized Inverses for Various Secondary Structures[7] 184-260 nm at 2 nm Intervals ($\Delta\varepsilon^{-1}x10^{3}$).

H:	2.57	-0.91	-2.84	-4.07	-5.60	-3.75	-0.55	1.59	6.35	7.87	5.57	1.27	-3.72	-7.09	-8.83
	-10.13	-11.18	-12.61	-13.22	-13.66	-13.97	-13.38	-11.93	-9.16	-6.93	-5.58	-4.36	-3.36	-2.55	-1.65
	-0.70	-0.35	0.06	0.20	0.05	-0.02	-0.02	-0.02	0.00						
A:	-47.74	-24.70	-4.05	12.11	13.38	10.34	-0.72	-15.45	-20.26	-26.87	-19.01	-12.42	-10.28	-10.20	-10.33
	-8.71	-6.74	-2.96	0.47	6.32	7.34	5.92	4.82	-1.88	-4.36	-5.11	-7.03	-6.47	-5.81	-5.58
	-4.86	-3.00	-1.94	-0.92	-0.46	-0.21	-0.21	-0.19	-0.15						
P:	-36.41	-14.13	6.92	20.22	13.53	2.55	-11.71	-19.87	-6.44	1.28	13.72	17.74	12.92	4.37	-2.28
	-4.76	-6.34	-6.58	-6.46	-3.43	-4.53	-5.25	-4.53	-9.70	-10.69	-11.84	-13.54	-11.84	-9.94	-8.31
	-6.35	-3.80	-1.75	-0.43	-0.43	-0.36	-0.36	-0.29	-0.09						
T:	-42.73	-22.86	-4.27	9.71	8.32	4.64	-4.81	-15.85	-13.52	-15.38	-7.52	-3.53	-5.27	-8.67	-11.09
	-11.07	-10.46	-8.41	-6.15	-1.43	-1.13	-2.15	-2.23	-7.15	-8.29	-8.60	-9.93	-8.78	-7.56	-6.65
	-5.26	-3.19	-1.77	-0.67	-0.41	-0.24	-0.24	-0.22	-0.13						
O:	-76.16	-34.87	5.12	28.58	8.74	-14.31	-41.20	-54.55	-16.53	4.51	30.05	34.56	18.43	-3.34	-19.24
	-25.69	-29.79	-31.66	-31.71	-25.28	-27.78	-28.51	-25.08	-32.84	-31.91	-32.46	-34.60	-29.76	-24.75	-20.04
	-14.50	-8.64	-3.74	-0.70	-0.87	-0.81	-0.81	-0.67	-0.23						

[7]Abbreviations as in Table II

combined as a column matrix. Then matrix multiplication will give the fractions of each secondary structure with each protein corresponding to a column.

As an example of how to carry out a computation, let us consider the fraction of α-helix corresponding to the CD of papain measured to 178 nm. Simply form the dot product of the inverse vector for H from Table III and the CD spectrum for papain from Table I:

$$(-0.00004,\ 0.00301,\ \ldots,\ -0.01028,\ \ldots)$$
$$\cdot\ (1.48,\ 2.24,\ \ldots,\ -2.69,\ \ldots)$$
$$= -0.00006 + 0.00674 + \ldots + 0.02765 + \ldots$$
$$= 0.27$$

The fractions for other secondary structures are calculated analogously to give the results for papain shown in Table VII. It is not even necessary to use a computer to do this computation.

Mathematically this method is identical to the original method presented in Hennessey and Johnson (1981). Thus the agreement between predicted and X-ray structures will be no better or worse than in that paper. The value of the generalized inverse presented here is the simplicity of computation.

The original paper (Hennessey and Johnson, 1981) presented computed structures for all sixteen proteins using the complete basis set. These were normalized to a total of 1.0 using a multiplicative factor. We also presented structures computed from a basis set with the protein in question removed. In Table VII we present the results for papain, lactate dehydrogenase, α-chymotrypsin, and hemoglobin that

TABLE VII. Analysis of Protein Structure[8]

	H	A	P	T	O	Total
Papain						
X-ray	0.28	0.09	0.00	0.14	0.49	
Cutoff 178 nm	0.27	0.09	0.05	0.15	0.50	1.06
Cutoff 180 nm	0.27	0.10	0.05	0.15	0.51	1.07
Cutoff 182 nm	0.27	0.10	0.05	0.15	0.52	1.10
Cutoff 184 nm	0.27	0.09	0.05	0.15	0.50	1.06
Papain eliminated[9]	0.26	0.10	0.08	0.15	0.49	1.08
Lactate dehydrogenase						
X-ray	0.41	0.06	0.11	0.11	0.31	
Cutoff 178 nm	0.40	0.08	0.13	0.12	0.28	1.01
Cutoff 180 nm	0.40	0.07	0.13	0.12	0.28	1.01
Cutoff 182 nm	0.40	0.07	0.13	0.13	0.29	1.03
Cutoff 184 nm	0.40	0.06	0.13	0.13	0.28	1.01
Lactate dehydrogenase eliminated[9]	0.40	0.08	0.14	0.12	0.27	1.00
α-chymotrypsin						
X-ray	0.10	0.34	0.00	0.20	0.36	
Cutoff 178 nm	0.13	0.22	0.01	0.19	0.34	0.89
Cutoff 180 nm	0.13	0.21	0.01	0.19	0.34	0.88
Cutoff 182 nm	0.13	0.22	0.01	0.20	0.35	0.91
Cutoff 184 nm	0.13	0.22	0.02	0.21	0.36	0.94
α-chymotrypsin eliminated[9]	0.13	0.19	0.01	0.19	0.33	0.86
Hemoglobin						
X-ray	0.75	0.00	0.00	0.14	0.11	
178 nm	0.76	(-0.02)	0.02	0.16	0.16	1.10
180 nm	0.76	(-0.01)	0.02	0.16	0.17	1.10
182 nm	0.76	0.00	0.03	0.16	0.19	1.13
184 nm	0.76	0.03	0.04	0.17	0.22	1.21
Hemoglobin eliminated[9]	0.78	(-0.04)	0.05	0.18	0.27	1.27

[8]Abbreviations as in Table II.

[9]Basis set from fifteen protein CD spectra with protein being analyzed eliminated.

are unnormalized and compare these with X-ray data.

We used the basis set derived from all sixteen proteins for cutoffs of 178, 180, 182, and 184 nm. Results for 178, 180, and 182 nm are similar for all three proteins, demonstrating the stability of the method. Results for the 184 nm cutoff tend to be poorer, showing that the information content of this truncated data is marginal. We also present structures computed from a basis set of fifteen proteins that does not contain the protein in question.

DISCUSSION

CD spectra of globular proteins can be viewed as defining a vector space whose dimensions are defined by the asymmetric characteristics of the protein that contribute to the CD signal. The space contains various secondary structures plus possible contributions from the distortions introduced in secondary structures by tertiary structure, and from aromatic and other absorbing sidechains. Application of SVD to our basis set of 16 protein CD spectra provides us with a set of 16 orthogonal CD eigenvectors and their associated singular values. These CD eigenvectors describe all the independent components within a protein CD spectrum. That is, they define the vector space of our basis set.

SVD provides a powerful tool for the analysis of a set of related vectors. The usefulness of this technique to the evaluation of CD spectra is evident in the fact that 1) it is not necessary for the data matrix be square, symmetric, or even nonsingular; 2) the necessity of using standard matrix inversion techniques is eliminated and 3) the unitary characteristics of the decomposed matrix, $\underset{\sim\sim}{USV}^T$, allows for the accurate and economical calculation of basis vectors which contain only the most significant components. This effectively removes both random low frequency error and high frequency noise.

Having truncated our data set at five CD eigenvectors, we constructed inverse CD spectra which are the structural solutions to the problem of obtaining secondary structure from CD spectra. That these inverse spectra contain enough information to solve for the five major structures is supported by the observation that good

reconstructions of our x-ray structure matrix are obtained when these limited inverse functions are used to evaluate our original CD spectra (Hennessey and Johnson, 1981). This is true even when the protein in question is eliminated from the basis set.

The concept of using an inverse matrix as the simultaneous solution to a set of related linear equations is not new. For some time x-ray crystallographers have used reciprocal lattice energy functions to calculate the electron density of an x-ray diffraction pattern. Many of the principles of quantum mechanics and molecular spectroscopy depend on wavefunctions that are orthogonal and unitary, the same properties inherent to SVD and its generalized inverse. In 1976 Baker and Isenberg (1976) theorized that there existed reciprocal CD functions which could be used to evaluate the secondary structure of a globular protein from CD spectrum. They demonstrated that if a CD spectrum, Ψ, is composed of a linear combination of reference spectra, b_i, where, f_i, is the relative amount of each reference spectra, then $\underset{\sim}{\Psi}=\sum f_i b_i$. Now, defining a set of reciprocal vectors, ϕ_i, such that $\int \phi_i b_i=1$ and $\int \phi_i b_j=0$, then $\phi_i \Psi=f_i$.

On close inspection, we see that the reciprocal CD function, ϕ_i, is nothing more than an integral approach using synthetic polypeptides which is analogous to the generalized inverse CD spectrum. Although Baker & Isenberg developed an excellent formalism, they were still forced to rely on synthetic polypeptide CD spectra in the calculation of inverse integrals. The generalized inverse CD spectra constructed using the SVD of our data set of globular protein CD spectra represent a matrix interpretation of the Baker-Isenberg reciprocal wavefunction.

REFERENCES

Baker, C.C., & Isenberg, I. (1976) Biochemistry 15, 629-634.

Barela, T.D., & Darnall, D.W. (1974) Biochemistry 13, 1694-1700.

Bolotina, A., Chekov, V.O., & Lugauskas, V. (1979) Int. J. Quantum Chemistry, 26, 819-824.

Brahms, S., & Brahms, J. (1980) J. Mol. Biol. 138, 149-178.

Chang, X.T., Wu, C.S.C., & Yang, J.T. (1978) Anal. Biochem. 91, 13-31.

Chen, Y.H., Yang, J.T., & Martinez, H.M. (1972) Biochemistry 11, 4120-4131.

Chen, Y., Yang, J.T., & Chau, K.H. (1974) Biochemistry 13, 3350-3359.

Draper, N.R., & Van Nostrand, R.C. (1979) Technometrics, 21, 451-467.

Forsythe, G.E., Malcolm, M.A., & Moler, C.B. (1977) Computer Methods for Mathematical Computations, pp. 192-239, Prentiss-Hall, Englewood Cliffs, New Jersey.

Goodman, M., Verdini, A.S., Toniolo, C., Philips, W.D., & Bovey, F.A. (1969) Proc. Natl. Acad. Sci. USA 64, 444-450.

Greenfield, N., & Fasman, G.D. (1969) Biochemistry 8, 4108-4116.

Hennessey, J.P., Jr., & Johnson, W.C., Jr. (1981) Biochemistry 20, 1085-1094.

Manavalan, P., & Johnson, W.C., Jr. (1985) Pro. Intl. Sym. Biomol. Struc. Intr., Suppl. J. Biosci., 8, 141-149.

Noble, B., & Daniel, J.W. (1977) Applied Linear Algebra, 2nd Edition, pp. 323-330, Prentiss-Hall, Englewood Cliffs, New Jersey.

Provencher, S.W., & Glockner, J. (1981) Biochemistry 20, 33-37.

Richardson, J.S. (1981) Adv. Protein Chemistry 34, 167-339.

Saxena, V.P., & Wetlaufer, D.B. (1971) Proc. Nat. Acad. Sci. USA 68, 969-972.

Siegel, J.B., Steinmetz, W.E., & Long, G.L. (1980) Anal. Biochem. 104, 160-167.

Woody, R.W. (1978) Biopolymers 17, 1451-1467.

Young, H.D. (1962) Statistical Treatment of Experimental Data, pp. 126-132, McGraw-Hill, New York.

THE CONFORMATION OF T4 BACTERIOPHAGE DIHYDROFOLATE REDUCTASE FROM CIRCULAR DICHROISM

Larry A. Compton, Christopher K. Mathews, and W. Curtis Johnson, Jr.

Department of Biochemistry and Biophysics

Oregon State University

Corvallis, Oregon 97331

ABSTRACT

The secondary and tertiary structure of T4 bacteriophage dihydrofolate reductase is investigated by vacuum ultraviolet CD spectroscopy and probability analysis of the primary amino acid sequence. The far ultraviolet CD spectrum of the enzyme in the range of 260-178 nm is analyzed by the generalized inverse and variable selection methods developed by our laboratory. Variable selection yields an average content of 26% α-helix, 21% antiparallel β-sheet, 10% parallel β-sheet, 20% β-turns, and 32% "other" structures within the T4 protein. The characteristic peaks of the CD spectrum indicate that the enzyme has a lot of antiparallel β-sheet, which is typical of the α+β tertiary class of globular proteins. The secondary structure of the protein is also analyzed by using four statistical methods on the amino acid sequence. Although the secondary structures predicted by each individual statistical method vary to a considerable extent, the fractions of each structure jointly predicted by a majority of the methods are in excellent agreement with our CD analysis. The alternating arrangement for some segments of α-helix and β-sheet predicted from primary structure to be within the enzyme is characteristic of proteins containing parallel β-sheet. This supports our conclusion that the protein contains both parallel and antiparallel β-sheet structures, but finding both types of β-sheet also means that the protein may have the variation on α/β tertiary structure recently found in EcoRI endonuclease and thymidylate synthase. These observations, in conjunction with other physical properties of the T4 reductase, suggest that the enzyme perhaps shares

an evolution in common with the dihydrofolate reductases derived from type I R-plasmids rather than with the host-cell protein.

INTRODUCTION

Dihydrofolate reductase (DHFR) catalyzes the NADPH-dependent reduction of 7,8-dihydrofolate to 5,6,7,8-tetrahydrofolate, a cofactor required in nucleotide and amino acid metabolism. Because of its role as a target site for important chemotherapeutic drugs, DHFRs from a variety of sources have been studied from the viewpoint of protein structure and function (Freisheim and Matthews, 1984). Several folate analogs, such as methotrexate and trimethoprim, which inhibit DHFR and result in cell death, bind tightly to the catalytic region of the molecule (Bolin et al., 1982; Hood et al., 1979). However, the benefits of such anticancer and antimicrobial agents have been limited due to the emergence of drug-resistant cell lines. Resistance can result from any one of three mechanisms. First, overproduction of the enzyme, either by increased translation, transcription, or amplification of the DHFR gene, results in drug resistance in certain cell lines (Sheldon and Brenner, 1976; Alt et al., 1978). Second, the effects of folate analogs can be overcome by mutational events which alter the affinity of the enzyme for the inhibitor (Baccanari et al., 1981). Third, some trimethoprim resistance factors are, in fact, plasmid-encoded DHFRs which are characterized by unusual biophysical properties (Fling and Richards, 1983). This has promoted interest in the origin and evolutionary role of these plasmid-derived enzymes.

T4 bacteriophage DHFR was the first enzyme of its kind to be purified to homogeneity (Erickson and Mathews, 1971). For the past several years, the Mathews laboratory has studied the molecular properties of the T4-encoded enzyme. The native enzyme has a

molecular weight of 44,500, as indicated by sedimentation velocity and gel filtration. The subunit molecular weight is approximately 23,000 as shown by SDS gel electrophoresis and aminopterin titration (Purohit et al., 1981). Therefore, we have concluded that the enzyme is a dimer, with each subunit containing one ligand binding site. It is of interest that the DHFR specified by the Tn7 transposon of type I R-plasmids grown in E. coli are also dimers of a relatively low subunit molecular weight (Novak et al., 1983). The T4 enzyme also shares a number of other properties in common with the DHFR specified by transposon Tn7; these include a moderate resistance to trimethoprim and a limited ability to substitute NADH for NADPH as electron donor (Fling and Richards, 1983; Novak et al., 1983).

Purohit and Mathews have also determined the nucleotide sequence of the structural gene which encodes the T4 DHFR protein, using recombinant DNA techniques (Purohit and Mathews, 1984). The cloned fragment contains a 193-residue polypeptide coding site with a calculated molecular weight of 22,053 daltons. This is in agreement with our reported subunit molecular weight. Sequential amino acid analysis has determined that the first twenty residues of the N-terminus of the protein are identical to those deduced from the nucleotide sequence. The amino acid sequence shows partial homology with other dihydrofolate reductases, most notably those residues involved in ligand binding and catalysis.

The resolution of the overall secondary structure of the enzyme molecule would be of great value in the determination of the mechanism of catalysis, drug resistance, and evolutionary origin of the T4 protein. Therefore, in this report we discuss studies designed to

elucidate the secondary and tertiary structure of T4 DHFR. The overall secondary structure of the native enzyme in solution is determined by analysis of the vacuum CD spectrum in the range of 178-260 nm. In addition, we also empirically predict the secondary structure of the protein from the nucleotide-derived amino acid sequence. We use four independent prediction methods, and combine the results into a joint prediction of the α-helical and β-sheet regions along the polypeptide chain. The advantage of using two separate techniques is that the more reliable CD analysis of secondary structure can be interpreted in terms of the two-dimensional array of structural regions along the polypeptide as shown by empirical prediction.

METHODS AND MATERIALS

Materials

T4 DHFR was purified according to the procedure of Purohit et al. (1981). The enzyme was given a final purification on a hydroxylapatite column (1 x 5 cm) using a 200-ml linear gradient from 0.01 M to 0.10 M $KHPO_4$, pH 6.8, and stored at 0°C until use. The enzyme has a specific activity identical to that previously reported for homogeneous T4 DHFR and shows one band on sodium dodecyl sulfate polyacrylamide gel electrophoresis (Erickson and Mathews, 1971; Purohit et al., 1981). Prior to use, the enzyme was dialyzed for 48 hrs with 2 changes of one liter 0.10 M $KHPO_4$, pH 6.8, at 4°C. Guanidine hydrochloride was purchased from Schwarz/Mann.

Methods

CD measurements. CD spectra were taken on a McPherson vacuum UV spectrophotometer modified for CD as described elsewhere (Johnson, 1971). Measurements were made using a 50 μm pathlength cell thermostatted at 20 ± 1°C. The instrument was calibrated by using (+)-10-camphorsulfonic acid, $\Delta\varepsilon$ = +2.37 M^{-1} cm^{-1} at 290.5 nm and -4.95 at 192.5 nm (Chen and Yang, 1977). Data were collected by a microcomputer at 0.5 nm intervals using a 0.1 sec time constant. The pen recorder was set at a 20-sec time constant and a 1 nm/min scanning speed throughout the experiments. Absorbance measurements were made on a Cary 15 spectrophotometer. Protein concentrations were

determined by using the guanidine hydrochloride method (Mulvey et al., 1974; Elwell and Schellman, 1977). The extinction coefficient for T4 DHFR at 190 nm was calculated to be 10,600 ± 80 M^{-1} cm^{-1} on an amino acid basis. The protein concentration varied from 1.0-2.0 mg/ml. Digitized data for both sample and baseline were smoothed with 9-point and 17-point Savitsky-Golay smoothing functions, respectively, before subtraction of the baseline (Savitzky and Golay, 1964). The resulting sample spectrum was then resmoothed using 11 points. The procedure was repeated three times on separate occasions from separate isolations of T4 protein. The three CD spectra were averaged and analyzed for secondary structure by using 42 data points at 2 nm intervals between 260-178 nm using the method described below.

CD data analysis. The CD spectrum of T4 DHFR was analyzed for structure by using a combination of the generalized inverse and variation selection methods developed in our laboratory. A detailed review of the generalized inverse method is available elsewhere (Compton and Johnson, 1986). Basically, this method is dependent upon the linear relationship between a CD spectrum and its secondary structure, which in matrix notation is given by $\underset{\sim}{X}\underset{\sim}{C} = \underset{\sim}{F}$, where $\underset{\sim}{C}$ is a set of CD basis spectra, $\underset{\sim}{F}$ their corresponding secondary structures, and $\underset{\sim}{X}$ the matrix of vectors which relate $\underset{\sim}{C}$ and $\underset{\sim}{F}$. In our method $\underset{\sim}{X}$ is determined by inverting $\underset{\sim}{C}$ such that $\underset{\sim}{X} = \underset{\sim}{F}\underset{\sim}{C}^{-1}$. To do so, the matrix $\underset{\sim}{C}$ is decomposed into the product of three matrices $\underset{\sim}{C} = \underset{\sim}{U}\underset{\sim}{S}\underset{\sim}{V}^T$, by using the singular value decomposition theorem (Johnson, 1985). The $\underset{\sim}{U}$ and $\underset{\sim}{V}$ matrices are orthogonal and unitary, and $\underset{\sim}{S}$ is a matrix with singular values on the main diagonal. Therefore, $\underset{\sim}{C}^{-1} = \underset{\sim}{V}\underset{\sim}{S}^+\underset{\sim}{U}^T$, where $\underset{\sim}{S}^+$ is a matrix with entries on the main diagonal, which are the reciprocals of

their corresponding singular values. Once formulated, $\underset{\sim}{X}$ can be used to calculate the secondary structure of an unknown CD spectrum of interest by using $\underset{\sim}{F} = \underset{\sim}{X}\underset{\sim}{C}$.

Variable selection. Any given set of reference proteins whose basic secondary structures are known from X-ray crystallography may contain extraneous CD contributions that are not contained in the protein being analyzed. This may be due to several reasons, e.g., aromatic sidechain absorption, tertiary folding, etc. Therefore, we would like to process our CD spectrum with a basis set of reference proteins which contain the same variables as in our experimental spectrum. The variable selection method consists of removing one or more reference proteins from the basis set and recalculating the secondary structure. However, since we do not know beforehand which proteins should be removed, the following method was applied in order to choose the optimal basis set for our experimental data analysis. If there exists n proteins in a given basis set and r proteins are removed, then there are n!r!(n-r)! possible subsets to choose from. Secondary structural analysis was performed on all possible combinations of r = 2 using the 22 protein basis set of Hennessey et al. (1987). The protein whose removal leads to the most combinations giving an improvement in both secondary structural analysis and curve-fit is removed from the basis set and the procedure is repeated on n-1 proteins until the following criteria are met: 1) the sum of secondary structures is in the range of 0.90-1.10; 2) no negative values greater than -0.05 are obtained for any structure; 3) the reconstructed spectrum is within the noise level of the reference protein basis set. This is equivalent to the sum of squared residuals

being less than 2.0; 4) the basis set is the largest set giving reasonable results which still contains all reference CD spectra similar to that of the experimental protein.

Secondary structure from amino acid sequence. The amino acid sequence of T4 DHFR used in the analysis is the 193 residue polypeptide encoded by the 579-nucleotide open reading frame taken from the sequence of Purohit et al. (1984). The prediction methods of Burgess et al. (1974), Chou & Fasman (1978), Garnier et al. (1978), and Lim (1974) were used to estimate the amounts and spatial arrangement of α-helices and β-sheets within the T4 enzyme. The results were combined in a joint prediction using the following rules: 1) The secondary structure was predicted by using the four individual methods. 2) A residue was assigned to a particular conformation if it was predicted by a majority of the methods. To be considered α-helix, a minimum of four consecutive residues must be in the α-helical conformation; to be considered β-strand, a minimum of three consecutive residues must be predicted as β-sheet. 3) If any region of sequence is predicted as 50% α-helical and 50% β-sheet, the hydrophobicity/hydrophilicity profile curves of Cid et al. (1982) were used to determine which structure was most probable. 4) Any segment predicted by two of the four methods and which extends a jointly predicted region was incorporated as long as rules 1)-3) were not violated.

RESULTS AND DISCUSSION

Secondary structure analysis of CD data

The CD spectrum of T4 DHFR from 178-260 nm is given in Figure 3. The spectrum is the average of three measurements. The reproducibility of individual $\Delta\varepsilon$'s for our measurements is represented by standard deviation error bars at the maximum and minimum of the spectrum. The general shape of the spectrum did not change between individual runs. The spectrum is characterized by three minima at 180, 212, and 228 nm, and by a maximum at 191 nm. The negative minimum at 212 nm is characterized by an intensity approximately twice that of the minimum at 228 nm. The baseline is crossed twice, once at 200 nm from negative to positive and again at 184 nm from positive to negative.

The secondary structure of T4 DHFR as calculated by the generalized inverse method is given in Table VIII. Analysis of the CD spectrum in Figure 3 with the entire 22 protein basis set of Hennessey et al. (1987) yields a structural content of 23% α-helix, 31% antiparallel β-sheet, 9% parallel β-sheet, 23% β-turns, and 37% other structure, giving a total of 1.23. Since we do not constrain the method in any fashion, the total of structures need not add to one. This is due, for the most part, to CD contributions within the basis CD set which are not contained within the experimental CD spectrum. Removal of extraneous basis proteins can greatly improve the structural prediction. When we applied the variable selection method, we obtained ten subsets in which the total of secondary structures

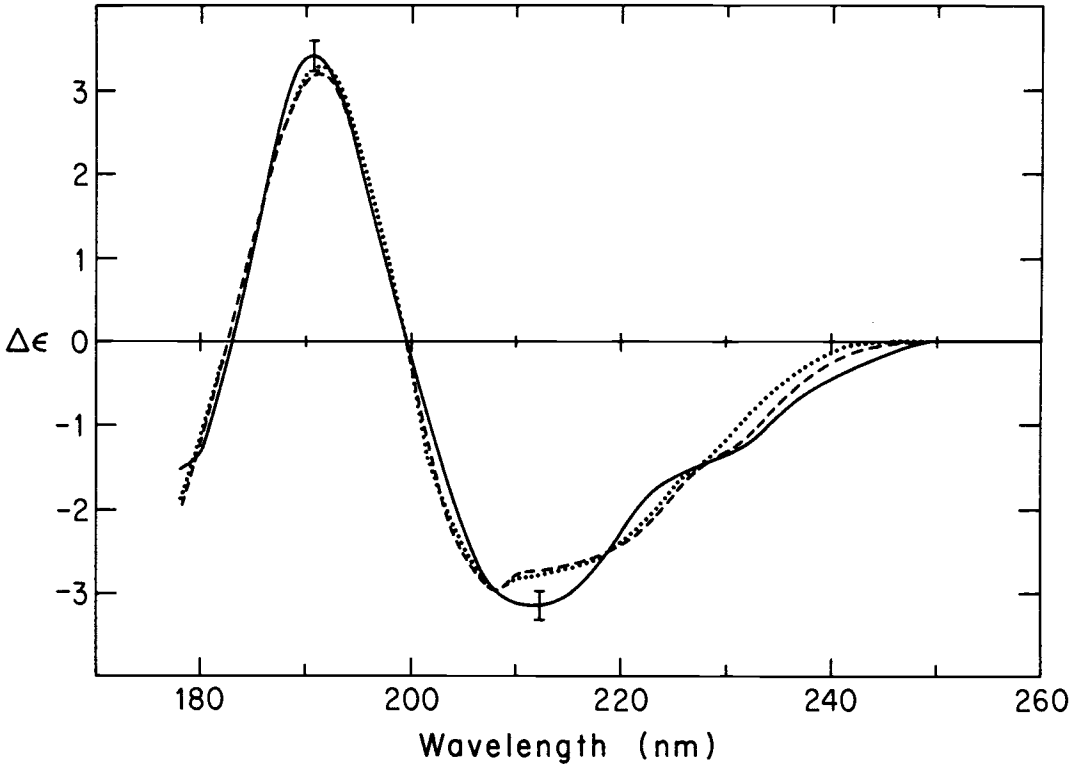

Figure 3. CD spectra of T4 dihydrofolate reductase: (——) original data, (••••) 22 protein reconstructed spectrum, (----) variable selection.

TABLE VIII. Secondary Structure Analysis of CD Data for T4 Dihydrofolate Reductase[10]

Method	H	A	P	T	O	Tot	SSR
unconstrained	0.23	0.31	0.09	0.23	0.37	1.23	2.266
variable selection	0.26	0.21	0.10	0.20	0.32	1.09	1.882

[10] H = α-helix, A = antiparallel β-sheet, P = parallel β-sheet, T = β-turns, O = other, Tot = total, SSR = sum of squared residuals for 42 points.

fell between 0.90 and 1.10, and the sum of squared residuals was less than 2.0. All of these subsets contained 12 basis CD spectra. The average of the predicted structural values is 26% α-helix, 21% antiparallel β-sheet, 10% parallel β-sheet, 20% β-turns, and 32% other structure. Surprisingly, the final structural selection indicates that the protein contains not only an abundance of antiparallel β-sheet, but also a substantial amount of parallel β-strands. This is not the usual case for globular proteins of low monomeric molecular weight, since most common examples of similar proteins exhibit an almost exclusive dominance of either one or the other structure.

Reconstructions of the CD spectrum for T4 DHFR from the generalized inverse and variable selection methods are also given in Figure 3. Both the entire 22 protein basis set and results of variable selection with the best 12 proteins give reasonably good reconstructions of the original data. However, the measured CD spectrum shows a broad 208-216 peak, whereas the calculated spectral curves show a sloping region between 210 and 220 nm, and a sharp bump at 208 nm. The generalized inverse method does not even reproduce the shoulder at 220 nm. This fit occurs because we have no protein in the basis set with a CD similar to T4 DHFR, so the calculated curve ends up looking similar to the α+β proteins in the basis set--i.e., lysozyme and ribonuclease. The unique characteristics of the T4 CD spectrum and their structural implications are discussed below.

Secondary structure from amino acid sequence

In addition to CD analysis, the secondary structural array of the

193-residue polypeptide chain encoded by our T4 DHFR clone (Purohit and Mathews, 1984) was analyzed using the empirical prediction methods as discussed in the methods section. Table IX summarizes the total of each structure calculated by these methods. Figure 4 indicates that all of the methods generally predict α-helix and β-strand in similar regions along the sequence. However, there exists quite a variation in the length and number of structural elements indicated by each individual method, and therefore a difference in the total amount of each structure indicated in Table IX. The α-helical content varies between 23% and 40%, with Lim (1974) giving the largest prediction. The β-strand differs from 25% with the Lim method to 59% with the Burgess (1974) method. Clearly, any one method is not exact, but we can take advantage of the fact that all the methods predict similar regions of structure by combining the results into a joint prediction as outlined earlier. The location of α-helices and β-strands along the protein chain is given in Figure 4. Included is the location of β-turns as predicted by the Chou and Fasman method (1978). The secondary structural totals for joint prediction are also given in Table IX. The contribution of ordered segments to the CD spectrum is dependent upon the amide, not the residue, so for the purposes of comparison the fraction of secondary structures in the joint prediction is also given in terms of amides. It is obvious from the tabular data that the joint prediction method yields excellent agreement with our CD analysis. In addition, we were pleased to note that the Chou and Fasman prediction (1978) fits β-turns neatly into the spaces between structural elements, even though these pieces of information were derived separately. This is a good indication that

TABLE IX. Amino Acid Sequence Analysis for T4 Dihydrofolate Reductase[11]

Method	α-helix	β-sheet (A + P)	Other (T + O)
Burgess et al.	0.20	0.55	0.25
Chou & Fasman	0.30	0.41	0.29
Garnier et al.	0.36	0.42	0.22
Lim	0.34	0.28	0.38
Joint method (per residue)	0.27	0.35	0.38
Joint method (per amide)	0.23	0.29	0.48

[11]Symbols as Table VIII.

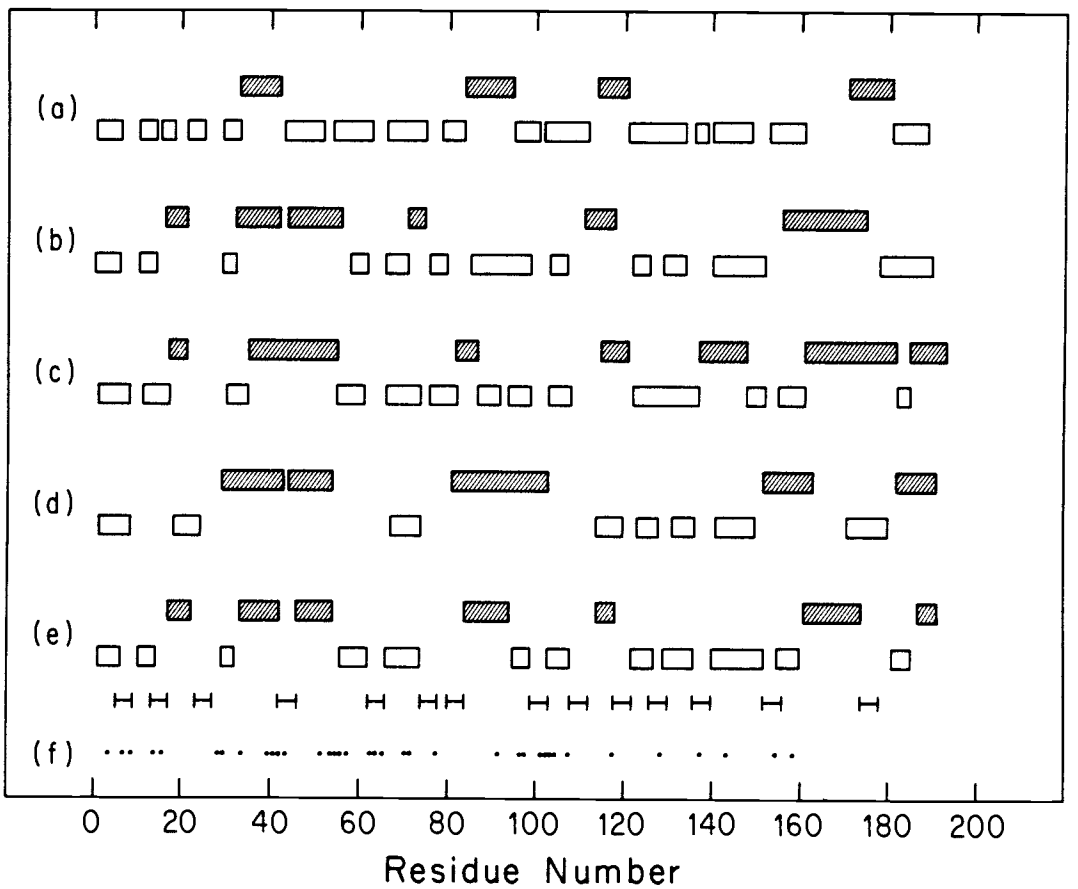

Figure 4. Prediction of secondary structure from amino acid sequence with α-helix, β-sheet, β-turn: (a) Burgess et al., (b) Chou and Fasman, (c) Garnier et al., (d) Lim, (e) joint prediction. Part (f) shows homologies of T4 DHFR with DHFR from the type I R-plasmid.

the alignment of structural elements is correct.

The conformation of T4 DHFR

The overall shape of the CD spectrum for T4 DHFR, and most notably the relationship between the two negative minima in the 200-260 nm range provide a valuable insight into the tertiary folding of the enzyme. Previously, our laboratory has shown that the major features of a CD spectrum share a distinct relationship with the tertiary folding classes described by Levitt and Chothia (1976). In their scheme an α+β protein consists of two domains, one containing α-helix and the other containing antiparallel β-sheet with β-turns reversing the chain direction. An α/β protein mixes α-helix and β-sheet structure in a single domain, and typically has a lot of parallel β-sheet as the chain fits into the β structure in one direction, but returns to the other side of the protein as α-helix. Our CD measurements (Manavalan and Johnson, 1983) on α/β proteins lactate dehydrogenase, triose phosphate isomerase, and other proteins characterized by alternating parallel β-sheet and α-helical structural elements all exhibit a negative minimum in the 220-228 nm range which is more intense than the accompanying minimum at 208-212 nm. These proteins are also characterized by a crossover at a wavelength longer than 200 nm, a maximum at ~192 nm, and a crossover at ~180 nm. On the other hand, α+β proteins such as ribonuclease and lysozyme, which contain predominantly segregated segments of antiparallel β-strands and α-helical elements, show a strong minimum in the 208-212 nm range with a weak shoulder at 224-228 nm. They also are characterized by a

crossover at 200 nm, a maximum at ~192 nm, and a crossover at wavelengths shorter than 180 nm.

Although the CD of T4 DHFR shares many of the characteristics found in the spectra of α+β proteins, there exist some features that are normally associated with proteins which contain a substantial amount of parallel β-sheet. These features include a minor minimum in the 224-228 nm range which is redshifted almost to 230 nm, a broad rather than sharp 208-212 nm peak, and a negative minimum at 180 nm which is normally absent in purely α+β proteins. The results of analyzing the CD through variable selection show (Table VIII) that although T4 DHFR contains a large amount of antiparallel sheet, it also contains significant amounts of parallel β-sheet. This suggests that the T4 protein may be representative of a subclass of tertiary folding which contains both α+β and α/β elements, as recently found for EcoRI endonuclease (McClarin et al., 1986) and thymidylate synthase (Hardy et al., 1987).

EcoRI endonuclease, which recognizes and cleaves specific sequences within double-stranded DNA molecules, was analyzed in this laboratory by using methods similar to those used on T4 DHFR (Manavalan et al., 1984). These studies indicate that, although the CD spectrum of the enzyme is characteristic of an α+β protein, the sequence prediction data are consistent with a protein which contains a substantial amount of parallel β-sheet. Recently, X-ray crystallography has shown that the tertiary structure of the protein is a mixture of both antiparallel and parallel β-structure. The N-terminal and C-terminal ends of the molecule are dominated by an α+β motif, while the central portion of the molecule forms a classic

α-helix/parallel β-sheet nucleotide binding site (McClarin et al., 1986). Further, X-ray analysis has shown that thymidylate synthase from L. casei contains the same type of antiparallel/parallel/ antiparallel folding pattern in the domains responsible for substrate binding and catalysis (Hardy et al., 1987). The CD spectrum is similar to EcoRI endonuclease, and the analysis in good agreement with the X-ray results (Manavalan et al., 1986).

The arrangement of secondary structure within the T4 protein as shown by analysis of amino acid sequence also supports the conclusion that the enzyme may contain parallel β-sheet folding patterns. Inspection of Figure 4 indicates that α-helical and β-sheet regions sometimes alternate along the polypeptide chain, indicative of the α/β tertiary folding.

The central portion of the T4 DHFR sequence contains the only cysteinyl, residue number 72, which in analogy to other DHFRs is probably responsible for the sulfhydryl reduction of dihydrofolate, and the majority of β-sheet residues involved in enzyme catalysis which are conserved among DHFRs (Purohit and Mathews, 1984). It also contains a group of closely spaced homologies with the type I R-plasmid DHFR not found in other DHFRs. The N-terminal (and less notably the C-terminal) region of the T4 protein contains large or multiple α-helices. These are also the known sites of cofactor substrate binding in many DHFRs. It seems quite possible that the conformation of the T4 enzyme is quite similar to that described for EcoRI endonuclease and thymidylate synthase.

All in all, these examples lend credence to the conclusion that the T4 DHFR protein is neither exclusively an α+β nor an α/β protein.

We suggest that T4 DHFR and the other examples cited here represent a subclass of tertiary folding which is a cross between the two well-defined classes of Levitt and Chothia (1976). This subclass is characterized by a folding motif in which the central portion of the molecule is dominated by a classic α/β structure (parallel β-sheet and α-helices), and the terminal ends of the molecule (or domain) are characterized by α+β folding (antiparallel β-sheet and α-helices). Whether or not this is indeed the structure of T4 DHFR will be confirmed by X-ray crystallography, which is now in progress for the bacteriophage protein.

The evolution of T4 DHFR

As discussed above, the CD spectrum of T4 DHFR exhibits many of the major components associated with a protein belonging to the α+β class of tertiary folding (Manavalan and Johnson, 1983). These include a stronger minimum at 212 than at 228 nm, and the positive to negative crossover at 200 nm. However, the dihydrofolate reductase encoded by the E. coli genome, the major host cell for T4 bacteriophage, has been shown both by CD spectroscopy and X-ray crystallography to contain exclusively the α/β type of tertiary folding (Bolin et al., 1982; Greenfield et al., 1972). Indeed, the CD spectrum of E. coli genome DHFR is distinctly different from T4 protein, showing the weak 208 nm and the strong 224 nm minima of a classic α-helical/β-strand protein (Greenfield et al., 1972). This indicates that the bacteriophage protein is not recently diverged from the E. coli gene.

In contrast, the molecular properties of the T4 and type I R-plasmid DHFRs show many similarities, so we wished to investigate the possible structural relationship between the two proteins. Therefore, Figure 4 also contains a comparison of the amino acid homologies between the two enzymes and how they correspond to the structural analysis by empirical prediction. Surprisingly, the relationship is not random, with almost all of the conserved residues being those predicted as β-sheet and only a few falling within the α-helical regions of the molecule. This suggests that the β-sheet component of the two proteins has been conserved, and that the two proteins perhaps share a common evolutionary origin. It is also of interest that the 40-residue C-terminus of the T4 protein, which is not present in the type I R-plasmid DHFR, is predicted as being almost exclusively α-helical with only one segment of β-sheet.

REFERENCES

Alt, F.W., Kellems, J.R., & Schimke, R.T. (1978) J. Biol. Chem. 253, 1357-1370.

Baccanari, D.P., Stone, D., & Kupyer, L. (1981) J. Biol. Chem. 256, 1738-1747.

Bolin, J.T., Filman, D.J., Matthews, D.A., Hamlin, R.C., & Kraut, J. (1982) J. Biol. Chem. 257, 13650-13662.

Burgess, A.W., Ponnuswamy, P.K., & Scheraga, H.A. (1974) Isr. J. Chem. 12, 239-286.

Chen, G.C., & Yang, J.T. (1977) Anal. Lett. 10, 1195-1207.

Chou, P.Y., & Fasman, G.D. (1978) Adv. Enzymol. Relat. Areas Mol. Biol. 47, 45-148.

Cid, H., Bunster, M., Arriagada, E., & Campos, M.(1982) FEBS Lett. 150, 247-254.

Compton, L.A., & Johnson, W.C., Jr. (1986) Anal. Biochem. 155, 155-167.

Elwell, M.L., & Schellman, J.A. (1977) Biochim. Biophys. Acta 494, 367-383.

Erickson, J.S., & Mathews, C.K. (1971) Biochem. Biophys. Res. Commun. 43, 1164-1170.

Fling, M.E., & Richards, C. (1983) Nucleic Acids Res. 11, 5147-5158.

Freisheim, J.H., & Matthews, D.A. (1984) in Folate Antagonists as Therapeutic Agents (Sirotnak, F.M., Burchall, J.J., Ensminger, W.B., & Montgomery, J.A., Eds.), Vol. 1, Academic Press, New York, pp. 69-131.

Garnier, J., Osguthorpe, D.J., & Robson, B. (1978) J. Mol. Biol. 120, 97-120.

Greenfield, N.J., Williams, M.N., Poe, M., & Hoogsteen, K. (1972) Biochemistry 11, 4706-4711.

Hardy, L.W., Finer-Moore, J.S., Montfort, W.R., Jones, M.O., Santi, D.V., & Stroud, R.W. (1987) Science 235, 448-455.

Hennessey, J.P., Jr., Manavalan, P., Johnson, W.C., Jr., Malencik, D.A., Anderson, S.R., Schimerlik, M.I., & Shalitin, Y. (1987) Biopolymers 26, in press.

Hood, K., Bayley, P.M., & Roberts, G.C.K. (1979) Biochem. J. 177, 425-432.

Johnson, W.C., Jr. (1971) Rev. Sci. Instrum. 42, 1283-1286.

Johnson, W.C., Jr. (1985) in Physical Optics of Dynamic Phenomena and Processes in Macromolecular Systems (Sedlacek, B., Ed.), pp. 492-506.

Levitt, M., & Chothia, C. (1976) Nature (London) 261, 552-557.

Lim, V.I. (1974) J. Mol. Biol. 88, 873-894.

Manavalan, P., & Johnson, W.C., Jr. (1983) Nature 305, 831-832.

Manavalan, P., Johnson, W.C., Jr., & Modrich, P. (1984) J. Biol. Chem. 259, 11666-11667.

Manavalan, P., Mittelstaedt, D.M., Schimerlik, M.I., & Johnson, W.C., Jr. (1986) Biochemistry 25, 6650-6655.

McClarin, J.A., Frederick, C.A., Wang, B., Greene, P., Boyer, H.W., Crable, J., & Rosenberg, J.M. (1986) Science 234, 1526-1541.

Mulvey, R.S., Gualtieri, R.J., & Beychok, S. (1974) Biochemistry 13, 782-787.

Novak, P., Stone, D., & Burchall, J.J. (1983) J. Biol. Chem. 258, 10956-10959.

Purohit, S., Bestwick, R.K., Lasser, G.W., Rogers, C.M., & Mathews, C.K. (1981) J. Biol. Chem. 256, 9121-9125.

Purohit, S., & Mathews, C.K. (1984) J. Biol. Chem. 259, 6261-6266.

Savitzky, A., & Golay, A.E. (1964) Anal. Chem. 36, 1627-1639.

Sheldon, R., & Brenner, S. (1976) Molec. Gen. Genet. 147, 91-97.

CONFORMATIONAL ANALYSIS OF PH AND THERMAL DENATURATION IN GLOBULAR PROTEINS BY USING CIRCULAR DICHROISM SPECTROSCOPY

Larry A. Compton and W. Curtis Johnson, Jr.

Department of Biochemistry and Biophysics

Oregon State University

Corvallis, Oregon 97331

ABSTRACT

Thermal and pH denaturation in globular proteins is monitored by circular dichroism spectra measured into the far uv region to 178 nm. The spectra are analyzed for secondary structure to show the change in conformation upon denaturation. Myoglobin, an α-helical protein, denatures in acid but not in base. In contrast, the β-sheet protein, concanavalan-A denatures in base but not in acid. The β-sheet structure in the globular proteins ribonuclease and papain appears quite stable to high temperature, but the α-helical secondary structure tends to denature as the temperature is raised. The segregated α-helical and β-sheet domains of papain behave in an independent fashion. In all cases, denaturation is not simply from a native to a disordered state, but rather to other structural arrangements that often involve β-sheet and β-turn.

INTRODUCTION

Since the discovery that globular proteins contain highly ordered and sophisticated elements of secondary structure (i.e. α-helices, & β-sheet) the stability of such structures and their importance in the biological function of the protein molecule has been the focus of intensive research. Proteins retain biological function only within a limited range of pH, ionic strength, and temperature. Not unexpectedly, it has been shown that the structural elements within many globular proteins undergo changes when their physical or chemical environment is altered. The term "denaturation" has been coined to describe the process by which the ordered, native state of a globular protein is transformed into a nonfunctional, presumably less ordered arrangement (Scheraga, 1961).

Originally it was generally accepted that the denaturation of a globular protein always proceeded from a highly ordered state to a more or less random configuration, and that the native and disordered states were the only ones available to the protein. For the most part transitions were observed indirectly using the measurement of some physical property other than loss of secondary or tertiary structure (e.g., intrinsic viscosity, solubility, or loss of biological function), and a completed transition was taken to represent a large if not total loss of secondary structure. A concise review of early techniques used to characterize protein denaturation is available elsewhere (Tanford, 1968). However, this "all-or-none" character of denaturation has recently been challenged by the use of modern biophysical methods which monitor the actual configurational

transitions of the polypeptide chain (Anfinsen and Scheraga, 1975). These include nuclear magnetic resonance (NMR) (Jardetzky et al., 1971), circular dichroism (CD) (Saito and Wada, 1983; Henkens et al., 1982; Takeda et al., 1981; Labhardt, 1982; Koser, 1983; Komiyama et al., 1984), and fluorescence spectroscopy (Henkens et al., 1982). All of these studies indicate that denaturation in a variety of globular proteins is either multiphasic or leads to "denatured" molecules which contain a high degree of residual structure.

One of the foremost methods currently in use to evaluate the secondary structure of native and denatured proteins is far-ultraviolet CD spectroscopy, which has been shown to be quite sensitive to the secondary structure of globular proteins (Chen et al., 1974; Saxena and Wetlaufer, 1971; Chang et al., 1978; Provencher and Glockner, 1981; Bolotina and Lugauskas, 1986; Hennessey and Johnson, 1981; Compton and Johnson, 1986). In the last two decades several CD studies of polypeptides and globular proteins in solution have given results consistent with the hypothesis that denaturation may not always proceed from a highly ordered to a totally disorganized state. Polypeptides such as poly-L-glutamic acid or poly-L-lysine, which in the native state are considered to be in an unordered configuration, will adopt highly organized secondary structures under conditions of extreme pH and temperature that would normally tend to "denature" globular proteins (Davidson and Fasman, 1967; Itoh et al., 1976). Denaturation using sodium dodecyl sulfate (SDS) on a variety of globular proteins involves an increase in the α-helical content of the protein molecule at the expense of disordered structure (Mattice et al., 1976; Su and Jirgensons, 1977; Jirgensons, 1981; Mori and

Jirgensons, 1981); subsequent dialysis of such solutions leads to an increase in β-sheet content (Visser and Blout, 1971). The addition of halogenated alcohols (i.e. 2-chloroethanol and trifluoroethanol) to β-sheet rich proteins such as concanavalin-A or the exotoxins from S. aureus results in the conversion of β-sheet to α-helix (McCubbin et al., 1971; Bailey et al., 1982). All of this evidence suggests that protein denaturation is much more complex than originally postulated.

In this manuscript we report experiments designed to determine the secondary structural changes associated with pH and thermal denaturation in globular proteins by using CD spectroscopy. CD spectra for several globular proteins obtained in acid, alkali, and neutral solution were analyzed for five major secondary structures using methods developed previously in this laboratory. The CD spectra for pH-denatured myoglobin, concanavalin-A, and papain were compared to the structure of the native protein and to one another, in order to determine the primary conformational transitions and any residual structure remaining within the denatured molecule.

In addition, high temperature CD spectra were measured for papain and ribonuclease. Ribonuclease is a globular protein which does not undergo any substantial change in secondary structure when the pH is lowered or raised. However, the temperature at which the protein "melts" is drastically affected by alkaline or acidic pH. CD spectra of thermally denatured ribonuclease in low and high pH solution were analyzed in order to determine the effect of pH on thermal denaturation.

MATERIALS AND METHODS

Materials

Myoglobin (sperm whale) was purchased from CalBioChem, ribonuclease-A (bovine pancreas) from Beohringer Mannheim, and Concanavalin-A (jack bean) and papain (papaya) from Sigma Corp. In addition, all experiments on papain were repeated with mercuripapain to ensure autolysis was not affecting the experimental outcome. All proteins were found to be at least 95% pure as monitored by polyacrylamide gel electrophoresis and used without further purification. Ultra-Pure grade guanidine hydrochloride was obtained from Schwarz-Mann. All other chemicals were reagent grade or better.

Methods

Protein Solutions. 0.01 M phosphate buffer, pH 6.8, was made by combining equimolar solutions of 0.01 M dibasic and monobasic sodium salts until the desired pH was obtained. Acid solutions used were 0.01 M $HClO_4$, pH 2.0; and 0.03 M H_3PO_4, pH 2.0. The phosphoric acid solution represents changing the pH without changing the anion used in neutral solution, but a higher ionic strength was necessary in order to obtain pH 2.0. In order ensure that ionic strength was not affecting the CD spectra, CD spectra for papain in 0.01 M HCl and 0.01 M oxalic acid, both pH 2.0, were measured to 190 nm and gave the same results as phosphoric acid (data not shown). However, high optical density prohibits using these acids below 190 nm. Perchloric acid

solution offers a transparent buffer and a different anionic composition at the same ionic strength as neutral pH sollution. Alkali pH solution was 0.01 M NaOH, pH 12.0. Prior to use all protein samples were dialyzed for 48 hours using two changes of one liter of the appropriate solvent.

CD Measurements. All CD spectra were measured on a McPherson vacuum UV spectrophotometer modified for CD as described elsewhere (Johnson, 1971), using a 50 or 100 μm pathlength cell thermostatted at the appropriate temperature. The instrument was calibrated using (+)-10-camphorsulfonic acid, $\Delta\varepsilon$ = +2.37 M^{-1} cm^{-1} at 290.5 nm and -4.9 at 192.5 nm (Chen and Yang, 1977). Data were collected between 260-178 nm by a microcomputer at 0.5 nm intervals using a 0.1 sec time constant. The pen recorder was set at a 10 or 20 sec time constant and a 1 or 2 nm/min scanning speed throughout the experiments. Protein concentrations were determined using the guanidine hydrochloride method (Edelhoch, 1967; Elwell and Schellman, 1977). Extinction coefficients at 280 nm used in this method for tyrosine, tryptophan and cystine in the acid and neutral range were taken from Mulvey et al. (1974). Extinction coefficients corrected for alkali solution were taken from Beaven and Holiday (1952). Normal absorption was then measured over the range of 260-178 nm on a Cary 15 spectrophotometer in order to ensure that the absorbance did not exceed 1.0. Furthermore, we determined extinction coefficients at 190 nm for each protein in each solvent (Table X) for ease in monitoring optical concentrations. In general, protein sample concentrations were 1.0-2.0 mg/ml. Digitized CD data for both sample and baseline were smoothed with 9-point and 17-point Savitsky-Golay smoothing

TABLE X. Extinction Coefficients at 190 nm ($\underline{M}^{-1}cm^{-1}$)

pH	6.8	2.0 (H_3PO_4)	2.0 ($HClO_4$)	12.0
Myoglobin	9,200	11,600	10,700	8,500
Concanavalin-A	9,900	10,800	10,700	9,100
Papain	10,100	11,000	10,900	9.600

functions, respectively, before subtraction of the baseline (Savitzky and Golay, 1964). The resulting sample spectrum was then resmoothed using 11 points. The procedure was repeated 3 times on separate samples for each experiment. CD spectra were averaged and analyzed for secondary structure using 42 data points at 2 nm intervals between 260-178 nm by the method described below.

CD Data Analysis. The CD spectra of all protein samples were analyzed for structure using the generalized inverse method of Compton and Johnson (1986). Briefly, this method uses a set of reference spectra which are derived from CD spectra for globular proteins whose secondary structures are well characterized by x-ray crystallography. In this report we have analyzed all the CD spectra of the various proteins for five major secondary structures; α-helix (H), parallel (P) and antiparallel (A) β-sheet, β-turns (T), and other configurations (O). Since all the proteins used in this study contain only antiparallel β-sheet, the term β-sheet will be used to describe results for both the parallel and antiparallel structures.

Variable Selection. Any given set of reference proteins whose secondary structures are known from x-ray crystallography may contain contributions to the CD spectra which are not represented in the protein being analyzed. The variable selection method consists of removing from the basis set one or more reference proteins that contain contributions which are extraneous to the analysis. However, since we do not know a priori which proteins should be removed, the variable selection method was applied on all possible subsets until the following criteria were met: 1) The sum of secondary structures, which is not constrained, is between 0.9 and 1.10, 2) no negative

values greater than -0.05 are obtained for any particular secondary structure, 3) the reconstructed CD spectrum for the protein is within experimental error. (This is equivalent to the sum of squared residuals for the 42 points being less than 2.0.), and 4) the basis set is the largest subset consistent with rules 1-3 which still contains all reference CD spectra similar to the experimental CD spectrum. All subsets meeting the criteria are then averaged.

Since we are comparing analyses closely in this work, we strengthened rule 1 so that all sums are 1.0. This is not unreasonable since concentrations, the main experimental source for error in the sum, are internally consistent for each protein. A more detailed discussion of the use of variable selection is available elsewhere (Hennessey et al., 1987).

RESULTS

Denaturation by pH

Myoglobin. Myoglobin, an oxygen binding protein found in mammals, was the first globular protein to be analyzed for structure using x-ray crystallography. It has been shown to be comprised of predominately α-helix and is representative of the all-α class of tertiary folding (Richardson, 1981). The CD spectra for sperm whale myoglobin in neutral, acidic, and alkaline pH are given in Figure 5. In addition, the low pH form is measured in both phosphoric acid and perchloric acid (pH 2.0) as discussed in Materials and Methods. The first interesting feature of note is that the CD spectrum of myoglobin in 0.01 M NaOH, pH 12.0, is similar if not identical to the CD spectrum of the native protein. That the alkaline solution has little or no effect on the shape of the spectrum is direct evidence that the molecule has not undergone a conformational change. The spectrum in neutral or alkaline solution is characterized by intense minima at 210 and 222 nm with a positive maximum at about 192 nm. This is typical of an α-helical secondary structure. The conclusion is that alkaline pH solution (pH 12.0) does not affect α-helix (at least in myoglobin).

The low pH forms of myoglobin display CD spectra distinctly different from the native form, and indeed differ from one another. The spectrum in phosphoric acid is characterized by a single major minimum at approximately 200 nm, and a significant negative shoulder between 210 and 220 nm. The CD spectrum of myoglobin in phosphoric acid is somewhat similar to the CD spectra of a variety of charged

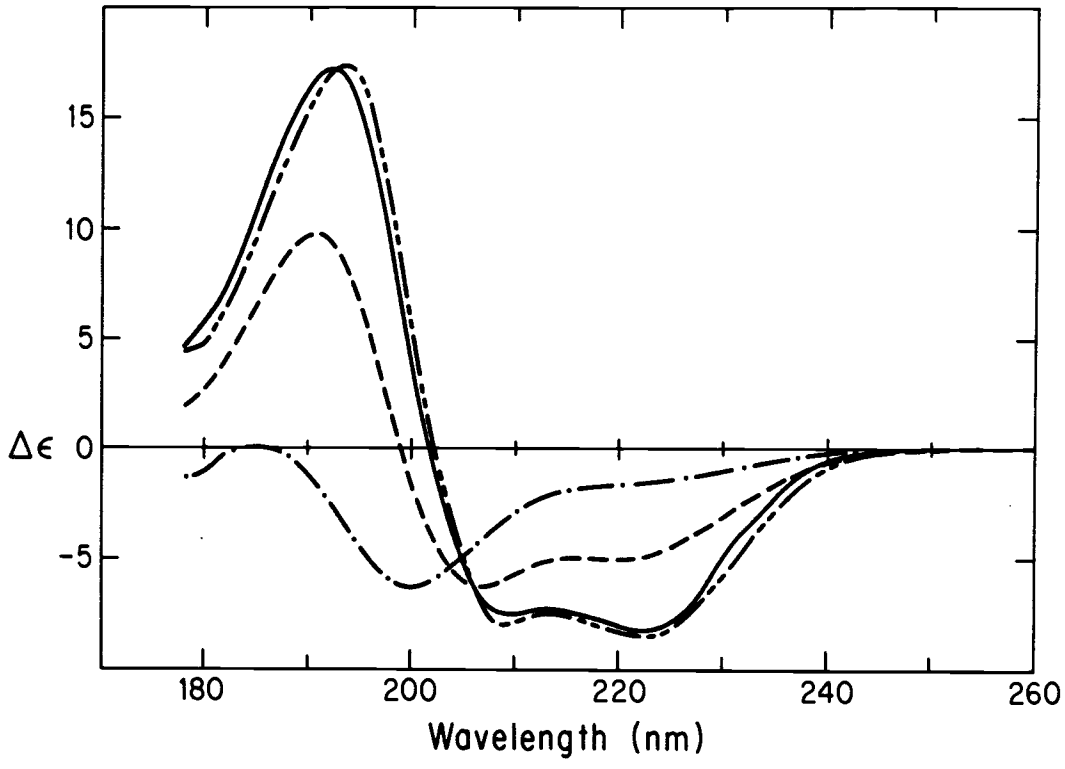

Figure 5. CD spectra of pH-denatured myoglobin
(———) pH 6.8, (— -- —) pH 12.0
(—·—) pH 2.0 (H_3PO_4)
(-----) pH 2.0 ($HClO_4$)

polypeptides, which are believed to be in the unordered conformation (Greenfield and Fasman, 1969; Johnson and Tinoco, 1972; Brahms et al., 1977). This suggests the denatured molecule contains a substantial amount of disordered structure. However, these reference spectra are characterized by an extremely intense minimum at 190-200 nm (about twice the spectral intensity of the denatured myoglobin spectrum) and a positive (rather than negative) maximum at 210-220 nm. Several investigations have questioned the interpretation that charged polypeptides in aqueous solution are truly representative of a completely disordered conformation (Tiffany and Krimm, 1969; Tiffany, 1975). Their studies show that synthetic polypeptides in high concentrations of inorganic salts, which are normally considered "denaturing" agents for globular proteins, produce CD spectra which contain significant negative shoulders at 210-220 nm identical to that for acid denatured myoglobin. Our interpretation is the reverse, however, since other work indicates that globular proteins with CD spectra similar to myoglobin in phosphoric acid at pH 2.0 (i.e., with a positive shoulder in the 210-220 nm region) still contain a significant amount of secondary structure.

A number of globular proteins in their native state display CD spectra that are similar to myoglobin in phosphoric acid. In particular, Ausio et al. (1987) have measured the CD for the nuclear sperm-specific protein from Spisula solidissma. Analysis of the CD spectrum for secondary structure indicates very little α-helix, some β-sheet, and a substantial amount of β-turn and "other" structure. The analysis is reasonable for a globular protein. Apomucin from the porcine submaxillary gland (Eckhardt et al., 1987) also has a CD

spectrum with a shape similar to myoglobin in phosphoric acid. A Chou and Fasman (1978) analysis of primary structure predicts primarily β-turn, which confirms the analysis of the CD spectrum which gives 40% β-turn and 42% disordered structure. The CD predicts little α-helix or β-sheet structure. This suggests that proteins which display CD spectra of this shape contain a substantial amount of β-turn structure. While we would not go so far as to say that the CD for charged polypeptides with the positive band at 220 nm corresponds to a truly disordered structure, we do note that the CD attributed to "other" structure reconstructed from our protein basis set (Compton and Johnson, 1986) is similar to the CD of the charged polypeptides.

The CD spectrum of myoglobin in 0.01 M perchloric acid, pH 2.0, is representative of a protein CD spectrum containing a much higher percentage of ordered structure. The spectrum is characterized by a maximum at approximately 190 nm, a major minimum at 208 nm, and a broad negative shoulder at 220 nm. This spectrum is indicative of the CD spectra corresponding to a class of proteins containing significant amounts of both α-helical and β-sheet structures, such as lysozyme and ribonuclease (Manavalan and Johnson, 1983).

In order to more quantitatively analyze the effect of acid pH on the conformation of myoglobin, the secondary structures corresponding to the CD spectra of the native and denatured forms were calculated using the generalized inverse and variable selection methods described earlier. The results are given in Table XI. The structure of myoglobin is most denatured in 0.03 M phosphoric acid. Analysis indicates that almost all of the α-helix has been destroyed, giving rise to an increase in β-turns and "other" structures, with little

TABLE XI. Secondary Structure Analysis of pH Denaturation[12]

	H	A	P	T	O
Myoglobin					
X-Ray	0.78	0.00	0.00	0.12	0.10
pH 6.8	0.75	(-0.03)	0.03	0.10	0.12
pH 2.0 (H_3PO_4)	0.19	0.05	0.01	0.41	0.34
pH 2.0 ($HClO_4$)	0.45	0.12	0.01	0.26	0.16
pH 12.0	0.82	(-0.05)	0.02	0.11	0.05
Concanavalin-A					
X-Ray	0.02	0.41	0.00	0.15	0.42
pH 6.8	0.12	0.19	0.13	0.11	0.45
pH 2.0 (H_3PO_4)	0.12	0.22	0.13	0.14	0.39
pH 2.0 ($HClO_4$)	0.11	0.20	0.15	0.13	0.41
pH 12.0	0.11	0.16	(-0.02)	0.33	0.40
Papain					
X-Ray	0.28	0.09	0.00	0.14	0.49
pH 6.8	0.27	0.14	0.02	0.12	0.45
pH 2.0 (H_3PO_4)	0.17	0.23	0.03	0.23	0.34
pH 2.0 ($HClO_4$)	0.17	0.27	0.08	0.14	0.34
pH 12.0	0.23	0.12	0.02	0.20	0.43

[12] H = α-helix; A = antiparallel β-sheet; P = parallel β-sheet; T = β-turns; O = other structures.

change in the β-sheet categories. This is similar to the undenatured proteins with this type of CD, since they also show a great deal of β-turn and "other" structure, and some antiparallel β-sheet. This suggests that the denaturation of myoglobin in low pH phosphoric acid leads to a protein molecule characterized by disordered regions which are folded by the inclusion of β-turns.

The denaturation of myoglobin in 0.01 M perchloric acid at pH 2.0 is quite distinctly different from that observed in phosphoric acid. Analysis indicates that the denatured molecule contains less α-helix than the native form, as is the case with phosphoric acid. However, there is no subsequent increase in the amount of "other" structures. Instead, the low pH molecule is characterized by a mixture of α-helix, β-sheet, and β-turns. Since the amount of increase in the β-sheet and β-turn categories is equivalent to the loss of helical structure, and since the native molecule contains no β-sheet, this is direct evidence that the primary conformational transition in low pH perchloric acid is one of α-helix to β-sheet and β-turn.

Concanavalin-A. Concanavalin-A, a lectin from legumes, is representative of the all β-sheet tertiary class of globular proteilns (Richardson, 1981). X-ray diffraction indicates that it contains only β-sheet, β-turns, and "other" structural components. The CD spectra of concanavalin-A in neutral, acid, and alkaline solutions are given in Figure 6. The primary observation is that although the CD spectrum in 0.01 M NaOH, pH 12, is drastically altered from the native form, the spectrum of the protein in either low pH solution is not significantly different from that of the native molecule at neutral pH. The CD spectrum of concanavalin-A in 0.01 M NaOH, pH 12.0, is

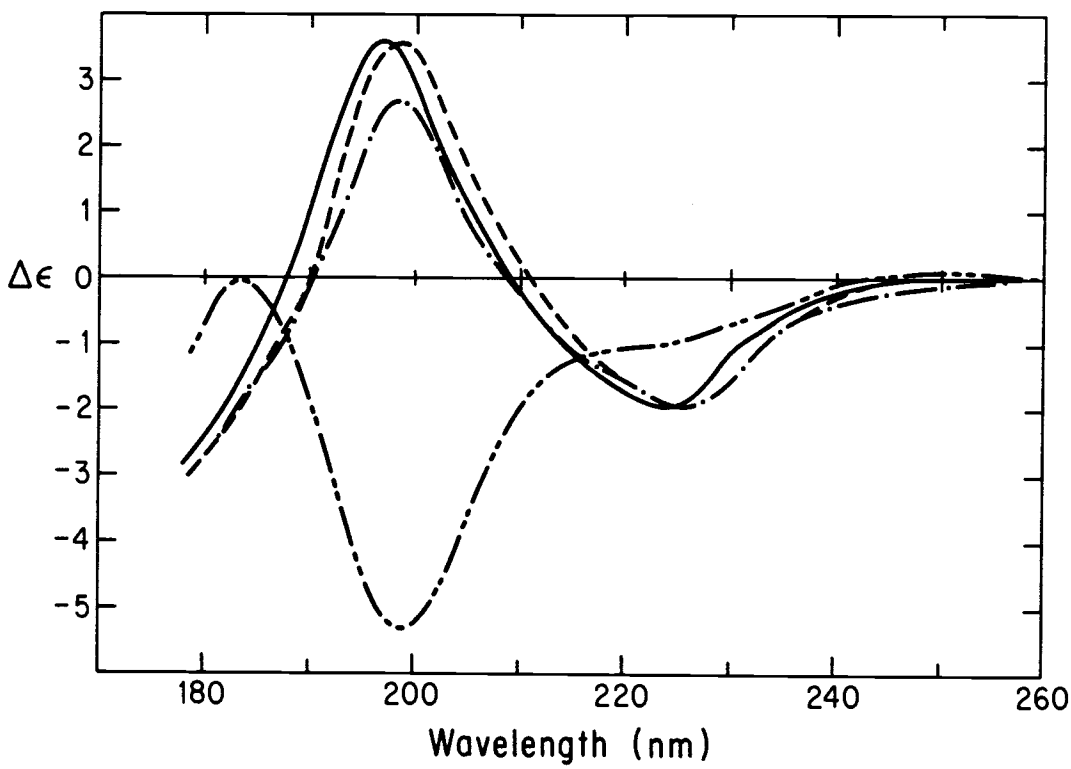

Figure 6. CD spectra of pH-denatured concanavalin-A
(——) pH 6.8, (— -- —) pH 12.0
(—·—) pH 2.0 (H_3PO_4)
(-----) pH 2.0 ($HClO_4$)

characterized by a single major minimum at 200 nm and a substantial negative shoulder at 210-220 nm. It is quite similar to the denatured spectrum of myoglobin in phosphoric acid. The qualitative conclusion is that the structure of the β-sheet protein is affected by alkaline solution, but is not altered by acidic pH. This indicates that the primary transition in 0.01 M NaOH, pH 12.0, is a loss of β-sheet opposite to the effect seen in myoglobin, a predominately α-helical protein that is stable in base. This conclusion is not supported by quantitative analysis (Table XI) which indicates little or no change within the struacture of the protein.

Papain. Papainase, a cysteine protease from papaya, is a low molecular weight protein belonging to the α+β tertiary class of globular proteins (Richardson, 1981). It is composed of two distinct domains, one comprised exclusively of α-helix, the other being composed primarily of β-sheet. The two domains are separated by a deep cleft in the molecule, which constitutes the active site. The CD spectra of papain in neutral, low, and high pH solutions are given in Figure 7. The spectrum at neutral pH is characterized by the typical 208 and 220 nm minima of α-helical proteins. However, the 220 nm peak is lower in intensity and has a broad sloping character indicative of proteins containing β-sheet. The conformational analysis of CD for secondary structure within the native and denatured molecule is given in Table XI.

The CD spectra corresponding to the low pH forms of papain in phosphoric and perchloric acid are typical of molecules containing less α-helix than the native enzyme. The phosphoric acid spectrum has a broad negative minimum at 200 nm and a sloping shoulder at

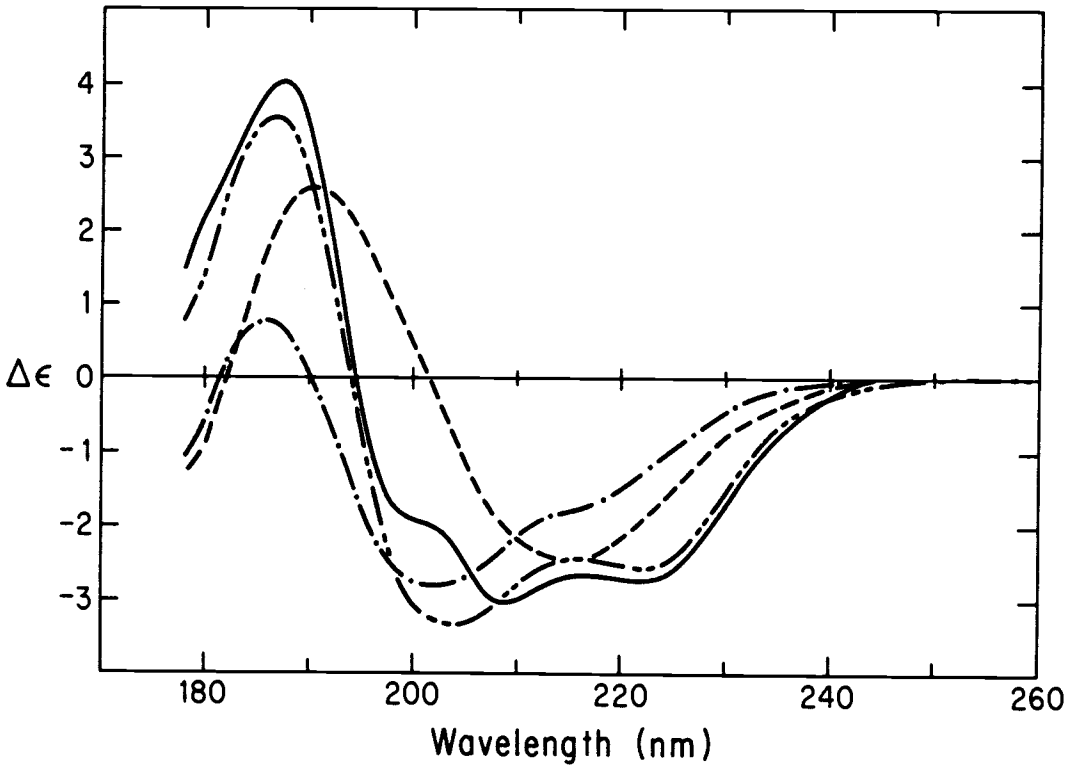

Figure 7. CD spectra of pH-denatured papain
(———) pH 6.8, (— -- —) pH 12.0
(—•—) pH 2.0 (H_3PO_4)
(-----) pH 2.0 ($HClO_4$)

approximately 210-220 nm. This is similar to the effect seen in myoglobin, only of less intensity. Quantitative analysis indicates that a loss of α-helix occurs with a concurrent gain in β-turns, and some β-sheet. However, there is no increase in the "other" structures as seen with myoglobin in phosphoric acid at pH 2.0. This may be real or merely due to the fact that the spectral changes observed in papain are of a lower magnitude than those for myoglobin, and therefore more suspectible to errors in analysis. Inspection of the spectrum with its 200 nm minimum is in agreement with a transition of α-helix to β-turn and "other" structures as seen with myoglobin in the same solution. Since the α-helical and β-sheet components of the papainase molecule are spatially segregated, the primary transition seen in phosphoric acid is consistent with a loss of the helical domain, without loss of structure in the β-sheet domain.

The perchloric acid CD spectrum of papain is quite different from the phosphoric acid form. It is characterized by a single broad minimum at 215 nm, a maximum which is red shifted to 195 nm, and a negative minimum at 178 nm. The spectrum is very similar in both shape and intensity to the typical β-sheet spectrum of synthetic polypeptides and the CD spectrum of some globular proteins comprised of predominately β-sheet structure. Analysis for secondary structure indicates that an α-helical loss has occurred, similar in magnitude to the transition seen in phosphoric acid; however, the largest gain of secondary structure has occurred in the β-sheet category. This suggests that the structural transition for papain in perchloric acid is one of α-helix to β-sheet and β-turn, just as that observed in myoglobin. Since the loss of α-helical structure is equivalent to

that seen in phosphoric acid, is reasonable to assume that the same region of the molecule has been affected. This then would support the conclusion that both of the domains within the denatured protein are comprised primarily of β-sheet structures in acidic solution.

The CD spectrum of papain in 0.01 M NaOH, pH 12.0, offers a third distinct form of the enzyme. The spectrum is characterized by a major negative minimum at 203 nm and another at 222 nm. In addition, the negative to positive crossover at 195 nm, and the positive maximum at 188 nm are very similar to those for the native protein. The two distinct minima and the positive shoulder at 178 nm in the far UV are all characteristic of the CD spectra of α-helical polypeptides, although the spectrum is blue-shifted by superposition of the CD spectrum for the disordered conformation. In addition, the isodichroic point for the CD spectra of native and alkaline denatured protein at 208 nm indicates that the α-helical content has not changed (Greenfield and Fasman, 1969). Taken in conjunction with the results for concanavalin-A in base, this suggests the primary transition for papain in alkaline solution may be one of β-sheet to β-turn and "other" structure. If our hypothesis is correct, then high pH affects only the β-sheet domain of the papain molecule. This is in contrast to the affect of acid pH, which rearranges only the α-helical portion of the molecule without disturbance of the β-sheet regions. However, once again this conclusion is not supported by quantitative analysis which indicates only minor changes within the protein molecule.

Thermal Denaturation

Papain. In order to investigate the effect of thermal denaturation on the secondary structure of globular proteins, the high temperature CD spectrum of papain can be compared to the already characterized high and low pH forms of the enzyme. The CD spectrum of papain in neutral solution at 80°C is given in Figure 8. It is immediately evident that the spectrum is very similar if not identical to that obtained for papain in 0.03 M phosphoric acid, pH 2.0. Quantitative analysis (Table XII) indicates that the primary structural transition is one of α-helix to antiparallel β-sheet and β-turn, the same result as in acid pH.

Additional evidence that acid pH and thermal denaturation of papain is differently affecting only specific regions (domains) of the molecule, without rearranging the structure of the entire protein is evidenced by the CD spectrum of papain in 6 M guanidine hydrochloride. Although optical density prohibits measuring the CD spectrum below 200 nm inspection of the CD in Figure 8 indicates that the molecule contains substantially less ordered structure than any of the pH or thermally denatured states of papain. This is consistent with the hypothesis that 6 M guanidine hydrochloride is rearranging both the β-sheet and α-helical regions (domains) of the protein.

Ribonuclease. Ribonuclease-A is a globular protein belonging to the same class of tertiary folding as papain. It is comprised of a mixture of α-helices and β-sheets, which are segregated into separate regions in the overall folding of the protein molecule. The CD spectrum for ribonuclease in neutral solution is given in Figure 9.

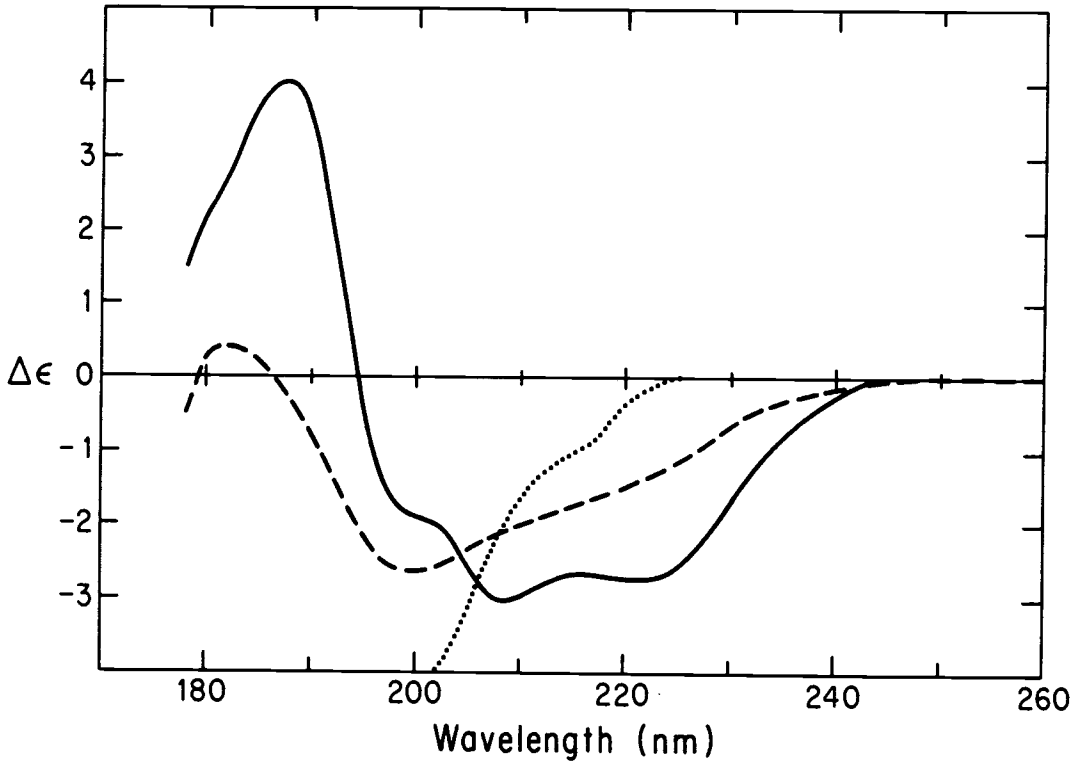

Figure 8. CD spectra of thermal denaturation in papain
(——) pH 6.8, 20°C
(----) pH 6.8, 80°C
(••••) 6M Guanidine HCl

TABLE XII. Secondary Structure Analysis of Thermal Denaturation[13]

	H	A	P	T	O
Papain					
X-Ray	0.28	0.09	0.00	0.14	0.49
pH 6.8, 20°C	0.27	0.14	0.02	0.12	0.45
pH 6.8, 80°C	0.17	0.21	0.01	0.21	0.40
Ribonuclease					
X-Ray	0.24	0.33	0.00	0.14	0.29
pH 6.8, 20°C	0.32	0.23	0.02	0.17	0.26
pH 6.8, 80°C	0.17	0.21	(-0.02)	0.27	0.35
pH 2.0, 50°C[14]	0.15	0.21	(-0.01)	0.29	0.35
pH 12.0, 50°C	0.13	0.26	(-0.01)	0.27	0.34

[13] Abbreviations as in Table XI.

[14] 0.01 M $HClO_4$

It is characterized by a major minimum at 208 nm, and a broad, sloping 220 nm shoulder typical of α+β proteins. However, unlike papain, the ribonuclease molecule is resistant to pH denaturation since the CD spectrum of the protein in acid or alkaline pH solution is identical to that of the native enzyme (data not shown). That ribonuclease does not undergo pH denaturation has been known for some time, but the structure of the molecule is destabilized by acid or alkaline solution since the thermal transition of the enzyme is achieved at substantially lower temperatures (Tanford, 1968). This effect can be seen in Figure 9, where it is evident that the high and low pH forms of the enzyme achieve thermal transition at a temperature 30 degrees C lower than the protein in neutral solution. All of the CD spectra corresponding to "melted" protein are quite similar. This suggests that, unlike pH denaturation in other globular proteins, there is no structural difference in the thermal transitions occurring in acid or alkaline buffers. Quantitative analysis (Table XII) indicates that at any given pH the α-helical portion of the ribonuclease molecule is partially destroyed by increasing the temperature, the primary structural transition being from α-helix to β-turn and "other" structure. This is in agreement with the result for papain at neutral pH and high temperature. The results for both papain and ribonuclease offer direct evidence that only the α-helical portions of globular proteins are affected by high temperature, leaving any β-sheet regions intact.

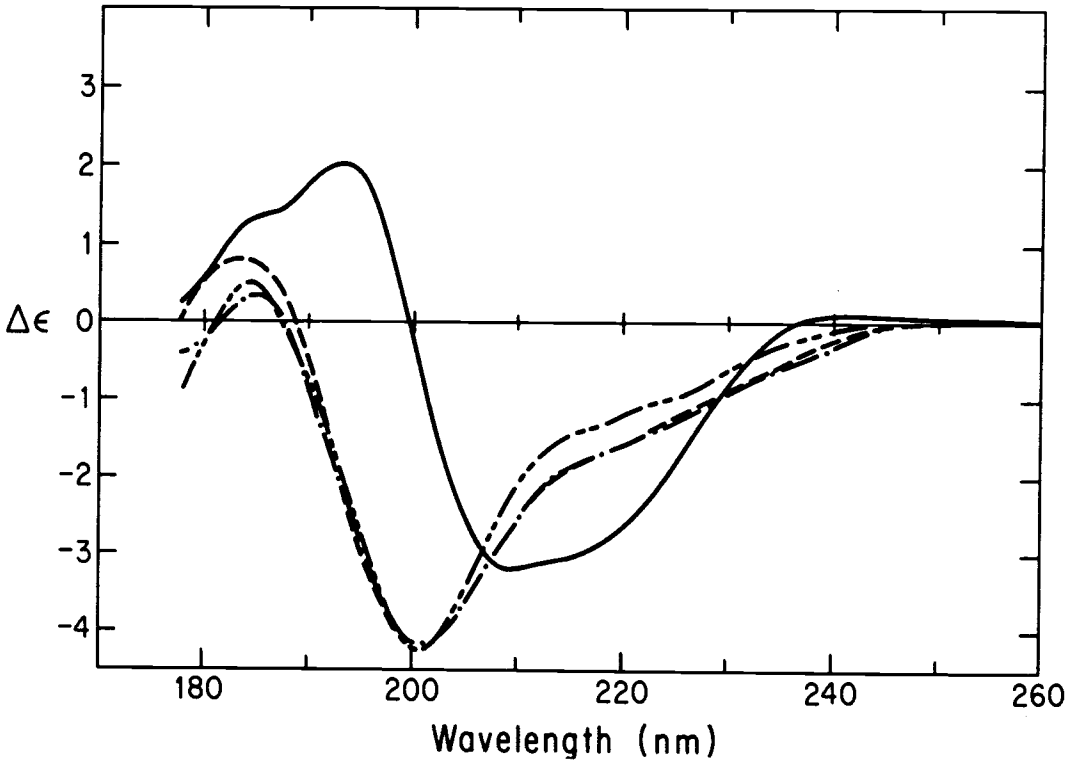

Figure 9. CD spectra of thermal denaturation in ribonuclease-A
(———) pH 6.8, 20°C
(----) pH 6.8, 80°C
(—·—) pH 2.0, 50°C ($HClO_4$)
(— -- —) pH 12.0, 50°C

DISCUSSION

The CD studies on globular proteins given in this report offer insight into the stability of the major secondary structures. The acid and alkaline pH experiments on myoglobin and concanavalin-A indicate that there exists a fundamental difference in the structural stability and weak bonding forces between the α-helix and β-sheet structures. This is evident in the result that α-helical structures are stable in base but undergo rearrangement in acid while β-sheets are stable in acid but are affected by alkaline solutions. This indicates that the Corey-Pauling structure of globular proteins are not only differentiated by torsional angles, dipole-dipole interactions, and other general topological constraints, but by denaturational stability that result from these interactions as well. The reason for this difference in conformational stability is not known; however, some data are available which offer a plausible theory for this effect. Tanford (1968) suggests that the instability of myoglobin to acid pH is due to the exposure of six histidyl residues normally buried within the hydrophobic interior of the molecule. These buried groups become unstable as the pH is lowered below the pK of the groups and the residues become protonated (and henced charged). This would also explain the stability toward alkaline pH, since these residues would be unaffected by high pH. However, this does not explain the reverse effect as seen in β-sheet proteins. A more general explanation along this same line of logic is that α-helix and β-sheet structures are preferentially stabilized by amino acids of negative and positive charge, respectively. Our explanation is

supported by such studies as the probability analysis of amino acids in different conformations used in the prediction methods of Chou and Fasman (1978). Their study indicates that negatively charged amino acids such as glutamic and aspartic acid are known to occur in helical structure with a higher probability than positively charged residues. β-sheet does not incorporate hydrophilic residues to any great extent. However, when such residues are incorporated, positive charges are preferred over negative. The proteins used in our study also show a distribution of charged amino acids which reflects this differential incorporation of charged residues. The helical regions of myoglobin contain 92% of the glutamic acid and 66% of the aspartic acid present within the entire molecule, but only 75% and 60% of the arginine and lysine, respectively. The β-sheets of concanavalin-A incorporate 50% of the arginine and 42% of the l;ysine, but only 42% and 35% of the glutamic and aspartic acid residues, respectively. Papain contains both α-helices and β-sheets. The α-helical regions incorporates 3 of 6 glutamic acid and 2 of 7 aspartic acid residues; whereas, the β-sheets contain only one glutamic acid residue and no aspartic acid. The β-sheet regions also contain 3 of the 11 arginine residues, the α-helical regions only one. Both regions incorporate 2 lysine residues. This means that the helical regions contain 5 negative charges but only 3 positive charges at neutral pH. Conversely, the β-sheet regions incorporate 5 positive charges and only 1 negatively charged residue.

The thermal denaturation of papain and ribonuclease also indicates that there exists distinct differences among the structural stability of secondary structures within globular proteins. This is supported

by the observation that in the proteins studied only α-helix is affected within the temperature range investigated. This preference has also been noted for several other proteins (Ayala and Nieto, 1978; Komiyama et al., 1983). Our results for ribonuclease are in opposition to those presented by Labhardt (1982), in which the primary thermal transition observed was calculated to be loss of β-sheet. However, CD spectra in that report were terminated at about 210 nm and least-squares analysis was performed using polypeptide models. We believe that our results in conjunction with other reports demonstrate that only α-helix is responsive to thermal denaturation. Although the physical reason behind such a preference is unknown at this time, it is possible that differences in the forces giving rise to β-sheet and α-helix are responsible for the greater stability of β-sheet. The hydrogen bonds, and indeed all of the stabilizing interactions in a classic α-helix, are of short-range intramolecular type. The formation of hydrogen bonds along the helical axis gives rise to a rigid, tubular structure which lacks any substantial flexibility. Indeed, α-helices are often found on the hydrophilic exterior of many globular proteins and act as structural reinforcers. On the other hand, β-sheet structures are characterized by hydrogen bonds which are between adjacent strands, almost intermolecular in nature. In addition, β-sheets are usually found within the interior of a protein, and are closely associated by hydrophobic interactions. Salemme (1982) has suggested that these properties give rise to an intrinsic flexibility in the β-sheet that allows transient energy fluctuations to be accommodated by cooperative motion of the structure without breaking local hydrogen bonds. It is possible then, that thermal

agitation of the polypeptide chain leads to fluctations in the secondary structure which are strong enough to overcome the rigid short-range interactions which stabilize easily accessible exterior α-helices, but do not disturb the tertiary folding enough to overcome interstrand stabilization of the internalized β-sheets.

The difference in stability between α-helix and β-sheet toward the effects of pH and temperature are also evident in proteins which are a mixture of secondary structures, such as papain or ribonuclease. Since such proteins have topologically separated regions (or domains) of secondary structure, the difference in pH denaturation between the two structures can be used to investigate the separate stability, vis a vis, the separate functionality of such domains. For some time the theory of such spatially separate regions has been that they both fold independently (nucleate) and function as separate units (Wetlaufer, 1981). To date the only available confirmation of the hypothesis has been the evidence that the proteolytic fragments of some globular proteins retain their overall secondary structure and that these fragments can complement to form an active enzyme (Wetlaufer, 1981). The differential denaturation results given for papain in acid and alkaline solution offer additional evidence for the independence of domain-like regions within such globular proteins. It seems that not only can "domains" be defined as structural or physically separate regions, but as denaturationally independent units as well.

The structural transitions studied in this report also provide support to the hypothesis that denaturation is neither an "all-or-nothing" phenomenon nor that the process always preceeds from an ordered to a more disordered state. None of the experiments

indicate that a totally unfolded molecule is the result of acid, alkaline, or thermal denaturation in any of the proteins studied. Some of the structural changes are simple rearrangements of one type of secondary structure to another. Proteins in alkaline solution may be characterized by a β-sheet to β-turn and "other" transition. Phosphoric acid and high temperatures induce a change from α-helix to β-turns and "other" structures. This evidence indicates that even protein molecules which are highly "denatured" by pH or temperature are characterized by an arrangement of irregular or "disordered" segments interspersed with β-turns. The β-turns are perhaps necessary in order to allow the unordered segments of the polypeptide chain to fold back on themselves in a spatially conservative fashion. This suggests that even in the denatured state globular proteins can retain their overall globular nature, unlike synthetic polypeptides which are normally considered to be in an extended conformation.

The most interesting structural transition occurs in perchloric acid, where results indicate that α-helix has been replaced by β-sheet structures. This β-reconstructive denaturation effect has been noted in other proteins as well as those discussed in this study (Gratzer et al., 1967; Le Gaillard et al., 1976). Taken in conjunction with the results of Jirgensons (1981) that α-reconstructive denaturation of proteins occurs in SDS solution and that a β-sheet to α-helix transition is observed with halogenated alcohols (McCubbin et al., 1971; Bailey et al., 1982), it seems the process of protein denaturation is much more complex than originally thought.

REFERENCES

Anfinsen, C.B., & Scheraga, H.A. (1975) Adv. Protein Chem. 29, 205-300.

Ausio, J., Toumadje, A., McParland, R., Becker, R.R., Johnson, W.C., Jr., & van Holde, K.E. (1987) Biochemistry 26, 975-982.

Ayala, J.A., & Nieto, M. (1978) Biochem. J. 169, 371-380.

Bailey, C.J., Martin, S.R., & Bayley, P.M. (1982) Biochemical Society 203, 775-778.

Beaven, G.H., & Holiday, E.R. (1952) Ad. Pro. Chem. 7, 319-386.

Bolotina, I.A., & Lugauskas, V.Y. (1985) Mol. Biol. (USSR) 19, 1409-1421.

Brahms, S., Brahms, J., Spach, G., & Brack, A. (1977) Proc. Natl. Acad. Sci. USA 74, 3208-3212.

Chang, C.T., Wu, C-S.C., & Yang, J.T. (1978) Anal. Biochem. 93, 13-31.

Chen, G.C., & Yang, J.T. (1977) Anal. Lett. 10, 1195-1207.

Chen, Y.H., Yang, J.T., & Chau, K.H. (1974) Biochemistry 13, 3350-3359.

Chou, P.Y., & Fasman, G.D. (1978) Adv. Enzymol. Relat. Areas Mol. Biol. 47, 45-148.

Compton, L.A., & Johnson, W.C., Jr. (1986) Anal. Biochem. 155, 155-167.

Davidson, B., & Fasman, G.D. (1967) Biochemistry 6, 1616-1629.

Eckhardt, A.E., Timpte, C.S., Abernethy, J.L., Toumadje, A., Johnson, W.C., Jr., & Hill, R.L. (1987) J. Biol. Chem. in press.

Edelhoch, H. (1967) Biochemistry 6, 1948-1954.

Elwell, M.L., & Schellman, J.A. (1977) Biochim. Biophys. Acta 494, 367-383.

Gratzer, W.B., Beaven, G.H., Rattle, H.W.E., & Bradbury, E.M. (1968) European J. Biochem. 3, 276-283.

Greenfield, N., & Fasman, G.D. (1969) Biochemistry 8, 4108-4116.

Henkens, R.W., Kitchell, B.B., Lottich, S.C., Stein, P.J., & Williams, T.J. (1982) Biochemistry 21, 5918-5923.

Hennessey, J.P., Jr., & Johnson, W.C., Jr. (1981) Biochemistry 20, 1085-1094.

Hennessey, J.P., Jr., Manavalan, P., Johnson, W.C., Jr., Malencik, D.A., Anderson, S.R., Schimerlik, M.I., & Shalitin, Y. (1987) Biopolymers 26, 561-571.

Itoh, K., Foxman, B.M., & Fasman, G.D. (1976) Biopolymers 15, 419-455.

Jardetzky, O., Thielmann, H., & Arata, Y. (1971) CSHS Quant. Biol. 36, 257-261.

Jirgensons, B. (1981) Makromol. Chem. 2, 213-217.

Johnson, W.C., Jr., & Tinoco, I., Jr. (1972) J. Amer. Chem. Soc. 94, 4389-4390.

Johnson, W.C., Jr. (1971) Rev. Sci. Instrum. 42, 1283-1286.

Komiyama, T., Miwa, M., Yatabe, T., & Ikeda, H. (1984) J. Biochem. 95, 1569-1575.

Kosen, P.A., Creighton, T.E., & Blout, E.R. (1983) Biochemistry 22, 2433-2440.

Labhardt, A.M. (1982) J. Mol. Biol. 157, 331-355.

Le Gaillard, F., Aubert, J.P., Dautrevaux, M., & Loucheux-Lefebvre, M.H. (1976) FEBS Letters 64, 278-284.

Mattice, W.L., Riser, J.M., & Clark, D.S. (1976) Biochemistry 15, 4264-4272.

McCubbin, W.D., Oikawa, K., & Kay, C.M. (1971) Biochem. Biophys. Res. Commun. 43, 666-674.

Mori, E., & Jirgensons, B. (1981) Biochemistry 20, 1630-1634.

Mulvey, R.S., Gualtieri, R.J., & Beychok, S. (1974) Biochemistry 13, 782-787.

Provencher, S.W., & Glockner, J. (1981) Biochemistry 20, 33-37.

Richardson, J.S. (1981) Advances in Protein Chemistry 34, 167-339.

Saito, Y., & Wada, A. (1983) Biopolymers 22, 2105-2132.

Salemme, F.R. (1982) Nature 299, 754-756.

Savitzky, A., & Golay, A.E. (1964) Anal. Chem. 36, 1627-1639.

Saxena, A.P., & Wetlaufer, D.B. (1971) Proc. Nat. Acad. Sci. USA 68, 969-972.

Scheraga, H.A. (1961) in Protein Structure (Kaplan, N.O., & Scheraca, H.A., Eds.) pp 81-128, Academic Press, New York.

Su, Y-Y.T., & Jirgensons, B. (1981) Arch. Biochem. Biophys. 181, 137-146.

Takeda, K., Miura, M., & Takagi, T. (1981) J. Colloid Inter. Sci. 82, 38-44.

Tanford, C. (1968) Adv. Pro. Chem. 23, 122-275.

Tiffany, M.L. (1975) Physiol. Chem. Phys. 7, 191-207.

Tiffany, M.L., & Krimm, S. (1969) Biopolymers 8, 347-359.

Visser, L., & Blout, E.R. (1971) Biochemistry 10, 743-751.

Wetlaufer, D.B. (1981) Adv. Pro. Chem. 34, 61-92.

BIBLIOGRAPHY

Alt, F.W., Kellems, J.R., & Schimke, R.T. (1978) J. Biol. Chem. 253, 1357-1370.

Anfinsen, C.B., & Scheraga, H.A. (1975) Adv. Protein Chem. 29, 205-300.

Ausio, J., Toumadje, A., McParland, R., Becker, R.R., Johnson, W.C., Jr., & van Holde, K.E. (1987) Biochemistry 26, 975-982.

Ayala, J.A., & Nieto, M. (1978) Biochem. J. 169, 371-380.

Baccanari, D.P., Stone, D., & Kupyer, L. (1981) J. Biol. Chem. 256, 1738-1747.

Bailey, C.J., Martin, S.R., & Bayley, P.M. (1982) Biochemical Society 203, 775-778.

Baker, C.C., & Isenberg, I. (1976) Biochemistry 15, 629-634.

Barela, T.D., & Darnall, D.W. (1974) Biochemistry 13, 1694-1700.

Beaven, G.H., & Holiday, E.R. (1952) Ad. Pro. Chem. 7, 319-386.

Bolin, J.T., Filman, D.J., Matthews, D.A., Hamlin, R.C., & Kraut, J. (1982) J. Biol. Chem. 257, 13650-13662.

Bolotina, A., Chekov, V.O., & Lugauskas, V. (1979) Int. J. Quantum Chemistry, 26, 819-824.

Bolotina, I.A., & Lugauskas, V.Y. (1985) Mol. Biol. (USSR) 19, 1409-1421.

Brahms, S., Brahms, J., Spach, G., & Brack, A. (1977) Proc. Natl. Acad. Sci. USA 74, 3208-3212.

Brahms, S., & Brahms, J. (1980) J. Mol. Biol. 138, 149-178.

Burgess, A.W., Ponnuswamy, P.K., & Scheraga, H.A. (1974) Isr. J. Chem. 12, 239-286.

Chang, C.T., Wu, C-S.C., & Yang, J.T. (1978) Anal. Biochem. 93, 13-31.

Chang, X.T., Wu, C.S.C., & Yang, J.T. (1978) Anal. Biochem. 91, 13-31.

Chen, G.C., & Yang, J.T. (1977) Anal. Lett. 10, 1195-1207.

Chen, Y.H., Yang, J.T., & Martinez, H.M. (1972) Biochemistry 11, 4120-4131.

Chen, Y., Yang, J.T., & Chau, K.H. (1974) Biochemistry 13, 3350-3359.

Chen, Y.H., Yang, J.T., & Chau, K.H. (1974) Biochemistry 13, 3350-3359.

Chou, P.Y., & Fasman, G.D. (1978) Adv. Enzymol. Relat. Areas Mol. Biol. 47, 45-148.

Cid, H., Bunster, M., Arriagada, E., & Campos, M.(1982) FEBS Lett. 150, 247-254.

Compton, L.A., & Johnson, W.C., Jr. (1986) Anal. Biochem. 155, 155-167.

Davidson, B., & Fasman, G.D. (1967) Biochemistry 6, 1616-1629.

Draper, N.R., & Van Nostrand, R.C. (1979) Technometrics, 21, 451-467.

Eckhardt, A.E., Timpte, C.S., Abernethy, J.L., Toumadje, A., Johnson, W.C., Jr., & Hill, R.L. (1987) J. Biol. Chem. in press.

Edelhoch, H. (1967) Biochemistry 6, 1948-1954.

Elwell, M.L., & Schellman, J.A. (1977) Biochim. Biophys. Acta 494, 367-383.

Erickson, J.S., & Mathews, C.K. (1971) Biochem. Biophys. Res. Commun. 43, 1164-1170.

Fling, M.E., & Richards, C. (1983) Nucleic Acids Res. 11, 5147-5158.

Forsythe, G.E., Malcolm, M.A., & Moler, C.B. (1977) Computer Methods for Mathematical Computations, pp. 192-239, Prentiss-Hall, Englewood Cliffs, New Jersey.

Freisheim, J.H., & Matthews, D.A. (1984) in Folate Antagonists as Therapeutic Agents (Sirotnak, F.M., Burchall, J.J., Ensminger, W.B., & Montgomery, J.A., Eds.), Vol. 1, Academic Press, New York, pp. 69-131.

Garnier, J., Osguthorpe, D.J., & Robson, B. (1978) J. Mol. Biol. 120, 97-120.

Goodman, M., Verdini, A.S., Toniolo, C., Philips, W.D., & Bovey, F.A. (1969) Proc. Natl. Acad. Sci. USA 64, 444-450.

Gratzer, W.B., Beaven, G.H., Rattle, H.W.E., & Bradbury, E.M. (1968) European J. Biochem. 3, 276-283.

Greenfield, N., & Fasman, G.D. (1969) Biochemistry 8, 4108-4116.

Greenfield, N.J., Williams, M.N., Poe, M., & Hoogsteen, K. (1972) Biochemistry 11, 4706-4711.

Hardy, L.W., Finer-Moore, J.S., Montfort, W.R., Jones, M.O., Santi, D.V., & Stroud, R.W. (1987) Science 235, 448-455.

Henkens, R.W., Kitchell, B.B., Lottich, S.C., Stein, P.J., & Williams, T.J. (1982) Biochemistry 21, 5918-5923.

Hennessey, J.P., Jr., & Johnson, W.C., Jr. (1981) Biochemistry 20, 1085-1094.

Hennessey, J.P., Jr., Johnson, W.C., Jr., Bahler, C., & Wood, H.G. (1982) Biochemistry 21, 642-646.

Hennessey, J.P., Jr., Manavalan, P., Johnson, W.C., Jr., Malencik, D.A., Anderson, S.R., Schimerlik, M.I., & Shalitin, Y. (1987) Biopolymers 26, 561-571.

Hennessey, J.P., Jr., Manavalan, P., Johnson, W.C., Jr., Malencik, D.A., Anderson, S.R., Schimerlik, M.I., & Shalitin, Y. (1987) Biopolymers 26, in press.

Hood, K., Bayley, P.M., & Roberts, G.C.K. (1979) Biochem. J. 177, 425-432.

Itoh, K., Foxman, B.M., & Fasman, G.D. (1976) Biopolymers 15, 419-455.

Jardetzky, O., Thielmann, H., & Arata, Y. (1971) CSHS Quant. Biol. 36, 257-261.

Jirgensons, B. (1981) Makromol. Chem. 2, 213-217.

Johnson, W.C., Jr., & Tinoco, I., Jr. (1972) J. Amer. Chem. Soc. 94, 4389-4390.

Johnson, W.C., Jr. (1971) Rev. Sci. Instrum. 42, 1283-1286.

Johnson, W.C., Jr. (1985) in Physical Optics of Dynamic Phenomena and Processes in Macromolecular Systems (Sedlacek, B., Ed.), pp. 492-506.

Komiyama, T., Miwa, M., Yatabe, T., & Ikeda, H. (1984) J. Biochem. 95, 1569-1575.

Kosen, P.A., Creighton, T.E., & Blout, E.R. (1983) Biochemistry 22, 2433-2440.

Labhardt, A.M. (1982) J. Mol. Biol. 157, 331-355.

Le Gaillard, F., Aubert, J.P., Dautrevaux, M., & Loucheux-Lefebvre, M.H. (1976) FEBS Letters 64, 278-284.

Levitt, M., & Chothia, C. (1976) Nature (London) 261, 552-557.

Lim, V.I. (1974) J. Mol. Biol. 88, 873-894.

Manavalan, P., & Johnson, W.C., Jr. (1983) Nature 305, 831-832.

Manavalan, P., Johnson, W.C., Jr., & Modrich, P. (1984) J. Biol. Chem. 259, 11666-11667.

Manavalan, P., & Johnson, W.C., Jr. (1985) Pro. Intl. Sym. Biomol. Struc. Intr., Suppl. J. Biosci., 8, 141-149.

Manavalan, P., Taylor, P., & Johnson, W.C., Jr. (1985) Biochim. Biophys. Acta 829, 365-370.

Manavalan, P., Mittelstaedt, D.M., Schimerlik, M.I., & Johnson, W.C., Jr. (1986) Biochemistry 25, 6650-6655.

Mattice, W.L., Riser, J.M., & Clark, D.S. (1976) Biochemistry 15, 4264-4272.

McClarin, J.A., Frederick, C.A., Wang, B., Greene, P., Boyer, H.W., Crable, J., & Rosenberg, J.M. (1986) Science 234, 1526-1541.

McCubbin, W.D., Oikawa, K., & Kay, C.M. (1971) Biochem. Biophys. Res. Commun. 43, 666-674.

Mori, E., & Jirgensons, B. (1981) Biochemistry 20, 1630-1634.

Mulvey, R.S., Gualtieri, R.J., & Beychok, S. (1974) Biochemistry 13, 782-787.

Noble, B., & Daniel, J.W. (1977) Applied Linear Algebra, 2nd Edition, pp. 323-330, Prentiss-Hall, Englewood Cliffs, New Jersey.

Novak, P., Stone, D., & Burchall, J.J. (1983) J. Biol. Chem. 258, 10956-10959.

Provencher, S.W., & Glockner, J. (1981) Biochemistry 20, 33-37.

Purohit, S., Bestwick, R.K., Lasser, G.W., Rogers, C.M., & Mathews, C.K. (1981) J. Biol. Chem. 256, 9121-9125.

Purohit, S., & Mathews, C.K. (1984) J. Biol. Chem. 259, 6261-6266.

Richardson, J.S. (1981) Adv. Protein Chemistry 34, 167-339.

Saito, Y., & Wada, A. (1983) Biopolymers 22, 2105-2132.

Salemme, F.R. (1982) Nature 299, 754-756.

Savitzky, A., & Golay, A.E. (1964) Anal. Chem. 36, 1627-1639.

Saxena, A.P., & Wetlaufer, D.B. (1971) Proc. Nat. Acad. Sci. USA 68, 969-972.

Saxena, V.P., & Wetlaufer, D.B. (1971) Proc. Nat. Acad. Sci. USA 68, 969-972.

Scheraga, H.A. (1961) in Protein Structure (Kaplan, N.O., & Scheraca, H.A., Eds.) pp 81-128, Academic Press, New York.

Sheldon, R., & Brenner, S. (1976) Molec. Gen. Genet. 147, 91-97.

Siegel, J.B., Steinmetz, W.E., & Long, G.L. (1980) Anal. Biochem. 104, 160-167.

Su, Y-Y.T., & Jirgensons, B. (1981) Arch. Biochem. Biophys. 181, 137-146.

Takeda, K., Miura, M., & Takagi, T. (1981) J. Colloid Inter. Sci. 82, 38-44.

Tanford, C. (1968) Adv. Pro. Chem. 23, 122-275.

Tiffany, M.L. (1975) Physiol. Chem. Phys. 7, 191-207.

Tiffany, M.L., & Krimm, S. (1969) Biopolymers 8, 347-359.

Visser, L., & Blout, E.R. (1971) Biochemistry 10, 743-751.

Wetlaufer, D.B. (1981) Adv. Pro. Chem. 34, 61-92.

Woody, R.W. (1978) Biopolymers 17, 1451-1467.

Yang, J.T., Wu, C.S.C., & Martinez, H.M. (1986) Methods in Enzymology 130, 208-269.

Young, H.D. (1962) Statistical Treatment of Experimental Data, pp. 126-132, McGraw-Hill, New York.