DISSERTATION

AROMATIC CIRCULAR DICHROISM IN GLOBULAR PROTEINS. APPLICATIONS TO PROTEIN STRUCTURE AND FOLDING.

Submitted by
Irina B. Grishina
Department of Biochemistry
& Molecular Biology

In partial fulfillment of the requirements
for the Degree of Doctor of Philosophy
Colorado State University
Fort Collins, Colorado
Fall, 1994

COLORADO STATE UNIVERSITY

July 11, 1994

WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER OUR SUPERVISION BY IRINA B. GRISHINA ENTITLED AROMATIC CIRCULAR DICHROISM IN GLOBULAR PROTEINS. APPLICATIONS TO PROTEIN STRUCTURE AND FOLDING BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

Committee on Graduate Work

Adviser

Department Head

ABSTRACT OF DISSERTATION

AROMATIC CIRCULAR DICHROISM IN GLOBULAR PROTEINS.

APPLICATIONS TO PROTEIN STRUCTURE AND FOLDING.

The exciton couplet approach was applied to estimate the circular dichroism (CD) of Trp side-chains in proteins. Calculations were performed by the origin-independent version of the matrix method, either for the indole B_b transition only or for the six lowest energy indole transitions.

The dependence of the CD of a Trp pair upon its distance and geometry has been analyzed. It was predicted that mixing with far-uv transitions are as important in determining the CD intensity of the near-uv transitions as the coupling among near-uv transition. The effects of varying exposure of Trp chromophores and nearby charges on Trp CD have been examined.

A survey of a large number of proteins from the Protein Data Bank reveals a number of cases where readily detectable exciton couplets are predicted to result from the exciton coupling of Trp B_b bands. The predicted CD spectra are generally couplets, often dominated by the contributions of the closest pair, but sometimes exhibit three distinct maxima. This CD depends on the distance and relative orientation of Trp pairs and thus reflects the spatial arrangement of Trp residues in the protein. It was shown that Trp side chains can make significant contributions to the CD of proteins in the far

ultraviolet. The distance dependence of exciton splitting, rotational and couplet strengths of Trp pairs show general agreement with theoretical predictions.

In several cases, changes in protein Trp CD can be attributed to a specific Trp pair and explained as a definite change in its conformation. Applications of the exciton couplet approach are discussed for various crystal forms of hen lysozyme, turkey and human lysozyme. Trp62 in hen and turkey lysozymes was found to be sensitive to the perturbations of the protein surface due to binding of substrate, antibodies and intermolecular contacts in the crystal. Conformational changes of Trp62 are predicted to have a strong effect on the overall Trp CD of lysozyme.

Predicted Trp CD is compared with experimental results for various lysozymes, α-chymotrypsin and chymotrypsinogen A, concanavalin, dihydrofolate reductase and ribonuclease from *Bacillus intermedius* 7P (binase). The calculated near-uv CD for hen lysozyme matches the experimental amplitude. Correlation of conformational changes in proteins with Trp CD is shown for α-chymotrypsin and chymotrypsinogen A. We found that the exciton couplet approach might be useful in relating Trp CD and changes in protein structure upon local mutations and conformational changes involved in enzyme activation.

Small globular proteins are usually composed of a single structural domain and undergo cooperative denaturation. We have demonstrated that a protein with a single structural domain, binase, and a protein with multiple structural domains, porcine pepsin,

contain fewer cooperative regions (energetic domains) under the conditions optimal for their functional activity. The study was performed by combining a CD analysis of the structural changes in the proteins during thermal denaturation and under various solvent conditions with thermodynamic properties observed by scanning microcalorimetry. Estimates of secondary structure were obtained from CD spectra, taking side-chain CD into account. It was found that neither of the proteins show any changes in secondary structure or local environment of aromatic amino acids upon separation of the energetic domains. The structural regions in binase corresponding to energetic domains were identified. It was shown that binase is converted from a single cooperative system into two separate energetic domains when ion pairs are disrupted, whereas the size of cooperative units in pepsin decrease as the electrostatic repulsion between regions in the molecule increases.

Irina Borisovna Grishina
Biochemistry & Molecular Biology Department
Colorado State University
Fort Collins, CO 80523
Fall, 1994

ACKNOWLEDGMENTS

I wish to express my deep gratitude to the two scientific communities where this work was performed, for scientific and moral support. I would like to acknowledge Prof. V. Vol'kenshtein, whose tea-parties I will long remember; Dr. A. Makarov, Prof. Yu. Morozov, Dr. N. Esipova, Dr. I. Bolotina, Dr. V. Chekhov, Dr. N Bajulina and Dr. I. Protasevich in the Engel'hardt Institute of Molecular Biology of the Russian Academy of Sciences; and Narasimha Sreerama, Andrey Volosov, Randy DeBey and A-Young Moon Woody in the Department of Biochemistry and Molecular Biology at Colorado State University where this work was successfully completed. I deeply appreciate the time spent under guidance of my advisor, Robert W. Woody, his timely advice, encouragement and broad scientific and cultural knowledge. Together with those mentioned, I would like to thank students, staff and faculty encountered during my predoctoral wandering in both departments for scientific discussions, moral and scientific support and for providing a wonderful atmosphere for my personal and scientific development. I would also like to thank my dear parents Maya and Boris Grishin for seeding the passion for science and discovery in my mind, and for all they have done for me; and a dear friend, who kept my spirits high surrounding me with warmth and care in most busy times.

CURRICULUM VITAE

February 29, 1964 - born - - Moscow, Russia.

1987 - M.S., Subdepartment of Biophysics, Department of Physics, Moscow State University, Moscow, Russia.

1987-1991 - Ph.D. graduate student at the Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia.

1994 - Ph.D, Department of Biochemistry and Molecular Biology, Colorado State University, Fort Collins, Colorado

FIELDS OF STUDY

Major Field : Circular Dichroism of Proteins and Polypeptides.

Experimental Circular Dichroism and Scanning Microcalorimetry of Globular Proteins. Professor Vladimir Mikhailovich Vol'kenstein, Dr. A.A. Makarov and Dr. N.G. Esipova (Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia).

Deconvolution Analysis of Protein Circular Dichroism. Professor Yuri V. Morozov and Dr. I. A. Bolotina (Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia);
Professor Robert W. Woody (CSU).

Theoretical Circular Dichroism of Globular Proteins.
Professor Robert W. Woody (CSU).

TABLE OF CONTENTS

Chapter I. Theory of Circular Dichroism

1. Historical Overview of Theories of Optical Activity.

a. Phenomenological Description. Fresnel's Formalism.

Circularly polarized light can be considered as the sum of two beams of linearly polarized light. The directions of the corresponding fields, electric and magnetic, in those linearly polarized beams are perpendicular to each other and their phases are shifted by $\pi/2$ (Fig. I-1a). If we evaluate the total electric field at each successive point in time from 1 to 4, separated by a quarter of a period from each other (Fig. I-1a), we obtain a circular rotation of the electric field vector in time. In this case, the rotation is counterclockwise as viewed by an observer looking toward the light source, which is left circularly polarized light (Fig. I-1b). The phenomenon of circular dichroism (CD) is a manifestation of the differential absorption of left and right circularly polarized light. This phenomenon is observed in media composed of molecules that are chiral, which means that they are not superimposable upon their mirror images.

Fresnel (1825) was the first to ascribe the rotation of plane-polarized light to the different velocities of transmission, or to the different refractive indices, of the two circularly polarized beams whose amplitudes add to form plane-polarized light. Considering $N_{L,R}$ as complex indices of refraction for left and right circularly polarized light:

$$N_{L,R} = n_{L,R} + ik_{L,R} , \qquad \text{(I-1)}$$

then the real part of their difference

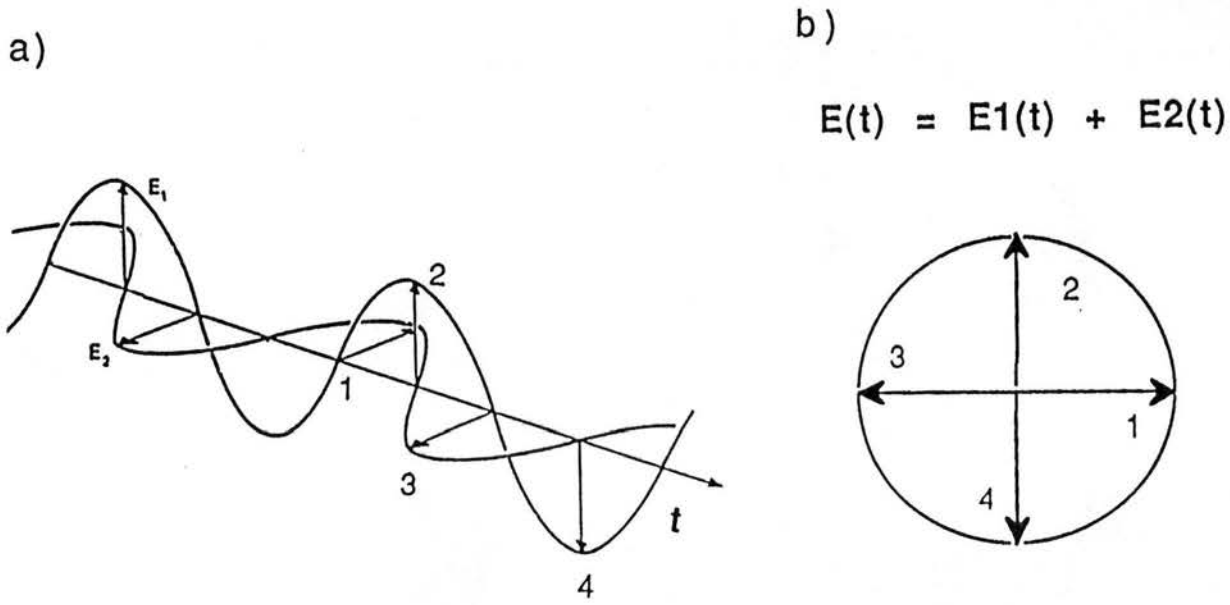


Figure I-1. Circular polarized light as sum of two linear polarized beams: a) propagation of the electric field vectors of the two linear beams; b) schematic projection of the sum of the electric field vectors on the plane perpendicular to the direction of propagation.

$$\Delta N = n_L - n_R + i\,(k_L - k_R) \qquad \text{(I-2)}$$

defines the phase retardation, while the imaginary part is the difference of the absorption indices, and interaction of the matter with light could be described by the value of the rotatory power Φ :

$$\Phi = (\pi\nu/c)\,(N_L - N_R) = (\pi\nu/c)\,[(n_L - n_R) + i\,(k_L - k_R)], \qquad \text{(I-3)}$$

where ν is the frequency of the light and c is the speed of light *in vacuo*.

When the light is absorbed by the sample, not only the relative phases of the left and right circularly polarized components will change but also the magnitudes of the electric field vectors of the two polarizations. The emerging electric field vector, rather than tracing a circle, will trace an ellipse (Fig.I-2), which arises from the sum of the two out-of-phase circularly polarized components of different amplitude. The imaginary part of Eq.I-3 is the ellipticity (the ratio of the minor to the major axis) per unit of length due to differential absorption of the left and right circularly polarized light:

$$\theta = (\pi\nu/c)\;(k_L - k_R). \qquad \text{(I-4)}$$

The absorption indices $k_{L,R}$ can be related to the extinction coefficientals $\varepsilon_{L,R}$ through the Beer-Lambert law. $k_{L,R}$ is defined through the relationship :

$$I_{L,R} = I^0_{L,R} \exp\,(-4\pi k_{L,R}\, l\, /\lambda), \qquad \text{(I-5)}$$

where $I^0_{L,R}$ is the intensity of the incident light beam, $I_{L,R}$ is the intensity of the transmitted beam, and l is the optical path length in cm. The absorbance A (λ) is defined as

$$A\,(\lambda)_{L,R} \equiv \log\,(I^0_{L,R}/I_{L,R}) = \varepsilon(\lambda)_{L,R}\, C\, l = \log\,(|\mathbf{E}^0_{L,R}|^2/|\mathbf{E}(\lambda)_{L,R}|^2, \qquad \text{(I-6)}$$

where C is the concentration of the sample in an optically inert solvent in mol/l; $\varepsilon(\lambda)_L$ and $\varepsilon(\lambda)_R$ are the molar extinction coefficients for the two polarizations; $\mathbf{E}^0$ and $\mathbf{E}$ are the

magnitudes of the electric field vectors for the entering and emerging light; and the subscripts L and R denote the left and right circularly polarized light, respectively.

Then $$\varepsilon(\lambda)_{L,R} = 4\pi k_{L,R} / 2.303\lambda C. \qquad \text{(I-7)}$$

The circular dichroism is proportional to the ellipticity, defined as the arctangent of the ratio of the minor axis ($|\mathbf{E}(\lambda)_R| - |\mathbf{E}(\lambda)_L|$) to the major axis ($|\mathbf{E}(\lambda)_R| + |\mathbf{E}(\lambda)_L|$). Considering that $\ln 10 = 2.303$, we can define the ellipticity θ from Equation I-5 as

$$\theta(\lambda) = \tan^{-1}\{ \frac{\exp[-2.303A(\lambda)_R / 2] - \exp[-2.303A(\lambda)_L / 2]}{\exp[-2.303A(\lambda)_R / 2] + \exp[-2.303A(\lambda)_L / 2]} \}. \qquad \text{(I-8)}$$

Because the CD is usually much smaller than the absorbance, the exponentials and the arctangent can be expanded in Taylor's series :

$$\theta(\lambda) \approx 2.303 [A(\lambda)_L - A(\lambda)_R] / 4 \quad \text{(radians)}. \qquad \text{(I-9)}$$

If we present the above value in degrees and in terms of molar units, as is done by convention, then we obtain for the molar elipticity $[\theta(\lambda)]$:

$$[\theta(\lambda)] = (100/lC)(180/\pi)\theta(\lambda) = 3298 [A(\lambda)_L - A(\lambda)_R] / lC$$

$$= 3298 (\varepsilon_L - \varepsilon_R) = 3298 \Delta\varepsilon \qquad \text{(deg cm}^2\text{ / dmol)}. \qquad \text{(I-10)}$$

What one measures experimentally is

$$\Delta A = A(\lambda)_L - A(\lambda)_R . \qquad \text{(I-11)}$$

Another way of describing the phenomenon of optical activity is through the optical rotation, which is defined as

$$\phi(\lambda) = [n(\lambda)_L - n(\lambda)_R]\pi l/\lambda \qquad \text{(rad)} \qquad \text{(I-12)}$$

or $$[\phi(\lambda)] \equiv [m(\lambda)] = (100/Cl)(180/\pi)\phi(\lambda) \quad \text{(deg cm}^2\text{ / dmol)}, \qquad \text{(I-13)}$$

where $\phi(\lambda)$ is in fact the angle between the long axis of the ellipse and the direction of polarization of the incident plane-polarized light (Fig.I-2). The optical rotation $\phi(\lambda)$ and

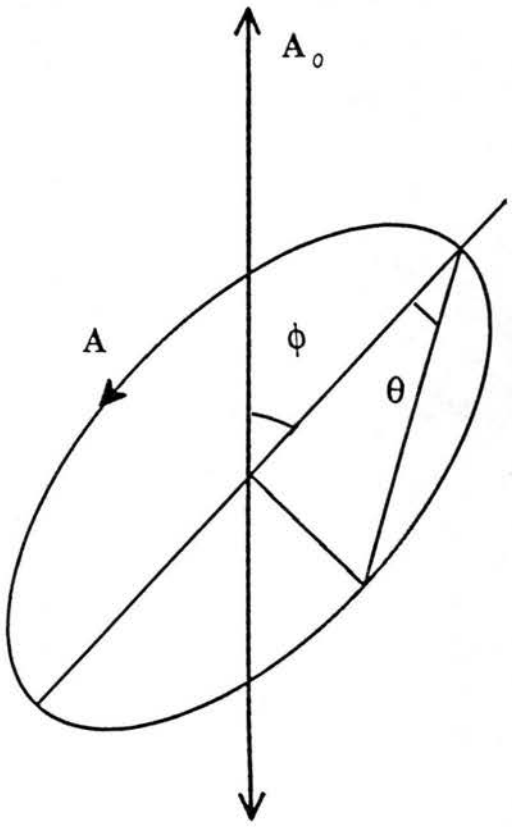


Figure I-2. The definitions of the rotational strength ϕ and the ellipticity θ. The light is approaching the observer. The vertical line labeled $\mathbf{A_0}$ represents the plane polarized light incident on the sample. The transmitted light **A** is elliptically polarized.

the ellipticity θ (λ) are interconvertible through the Kronig-Kramers transformation (Moffitt & Moscowitz, 1959).

b. Rosenfeld's Formalism. Rotational Strength.

Rosenfeld (1928) treated the interaction of light with optically active molecular media by first-order perturbation theory and arrived at the expression

$$[\phi\ (\nu)] = \quad (48N_A\nu^2/\hbar c)\ \ \sum_a R_{0a}/(\nu^2_{a\,0} - \nu^2), \qquad \text{(I-14)}$$

which links together the experimentally measurable molar rotation of the polarized light ($[\phi]$), the frequency (ν_{a0}) of the transition $0 \rightarrow a$ from the ground state (0) to the excited state (a), the frequency (ν) of the light for which the optical rotation is measured, and a parameter describing a molecular characteristic of the system interacting with the light beam, which is called the rotational strength (R), and is defined for each electronic transition of the molecule, $0 \rightarrow a$.

At a molecular level, optical activity manifests itself in electromagnetically induced charge displacements that have both linear and circular character. An optically active transition will have a characteristic vector describing the linear charge displacement in the system during the transition from the ground state (0) to the excited state (a), an electric dipole transition moment ($\boldsymbol{\mu}_{oa}$). Simultaneously, an optically active transition will be described by a vector characterizing the circular charge displacement, the magnetic dipole transition moment ($\mathbf{m}_{oa}$). In the Rosenfeld representation (1928), the rotational strength was defined as the imaginary part of the scalar product of the electric and magnetic dipole transition moments

$$R_{0a} = \text{Im}\ [\ \boldsymbol{\mu}_{0a} \bullet \mathbf{m}_{a0}] = -i\ \boldsymbol{\mu}_{0a} \bullet \mathbf{m}_{a0} \qquad \text{(I-15)}$$

In quantum mechanical representation the electric and magnetic dipole transition moments are :

$$\mu_{0a} = \langle 0 \mid \mu \mid a \rangle = \int \phi_0 \, \mu \phi_a d\tau \qquad \text{(I-16)}$$

and

$$\mathbf{m}_{a0} = \langle a \mid \mathbf{m} \mid 0 \rangle = \int \phi_a \, \mathbf{m} \phi_0 \, d\tau \qquad \text{(I-17)}$$

The operators μ and $\mathbf{m}$ are defined as :

$$\mu = e \sum_i \mathbf{r}_i \qquad \text{(I-18)}$$

and

$$\mathbf{m} = (e/2mc) \sum_i \mathbf{r}_i \times \mathbf{p}_i \qquad \text{(I-19)}$$

The wavefunctions in these equations represent the ground $| 0 \rangle$ and excited $| a \rangle$ states of the molecule. The integrals are taken over all space.

It can be seen from Equation I-15 that in order to have a nonvanishing rotatory strength, both the electric and magnetic transition moments must be nonzero. Molecules with a center of symmetry cannot have transitions that possess simultaneously nonvanishing electric and magnetic transition moments. Hence they cannot exhibit optical activity. Molecules with a plane of symmetry may have both μ and $\mathbf{m}$ nonzero, but these must be perpendicular to each other.

A serious problem in Rosenfeld's formulation is the necessity to know the accurate wavefunctions of the ground and excited states to evaluate the transition dipole moments. Obtaining accurate wavefunctions is possible so far only for very simple molecules. However the necessity to use the exact wavefunctions can be overcome by incorporating some experimental data in the calculations, such as experimentally determined directions and magnitudes for the electric transition dipole moments of electrically allowed transitions.

c. Determination of Transition Parameters from Experimental Bands.

The dipole strength of a transition D_{0a} , which in quantum mechanical reasoning defines the probability for this transition to occur, is determined from the area under the absorption band of the $0 \rightarrow a$ transition (Mulliken, 1939)

$$D_{0a} = \mu_{oa} \bullet \mu_{oa} = |\mu_{oa}|^2 = (6909hc/8\pi^3 N_A) \int_0^\infty \epsilon(\lambda) d\lambda/\lambda =$$

$$= 9.180 \times 10^{-39} \int_0^\infty \epsilon(\lambda) d\lambda/\lambda. \qquad \text{(I-20)}$$

Analogously, the rotational strength is determined from the area under the CD band (Moffitt & Moscowitz, 1959)

$$R_{0a} = (hc/48\pi^2 N_A) \int_0^\infty [\theta(\lambda)_{oa}] d\lambda/\lambda = 0.696 \times 10^{-42} \int_0^\infty [\theta(\lambda)_{oa}] d\lambda/\lambda. \qquad \text{(I-21)}$$

If the absorption band could be satisfactorily resolved, the amplitude of the dipole transition moment can be determined from Equation I-20

$$|\mu_{oa}| = D_{0a}^{1/2}. \qquad \text{(I-22)}$$

Analogously, from Equations I-21 and I-15, $|\mathbf{m}_{a0}|$ could be obtained, but the angle θ between $\mathbf{m}_{a0}$ and μ_{oa} must be known, as in the general case:

$$|\mathbf{m}_{a0}| = R_{0a} / |\mu_{oa}| \cos\theta = R_{0a} / D_{0a}^{1/2} \cos\theta. \qquad \text{(I-23)}$$

d. Shape of CD Bands.

It is most common for calculations of CD to assume that the shape of the CD bands is a Gaussian function of frequency or wavelength. This approximation frequently gives good agreement with the observed bands if the transitions considered are allowed. This is the case for indole, for which most of the transitions that I considered in my

calculations are allowed. The exception is the L_b transition, whose contribution to the overall CD spectrum is small. If we assume that the bands are Gaussian, then the absorption due to the transition $0 \rightarrow a$ can be described as

$$\epsilon(\lambda)_{0a} = \epsilon^0_{0a} \exp [-\{ (\lambda - \lambda_a) / \Delta_a \}^2], \qquad \text{(I-24)}$$

and for the CD band

$$[\theta(\lambda)_{0a}] = [\theta^0_{0a}] \exp [-\{ (\lambda - \lambda_a^0) / \Delta_a^0 \}^2]. \qquad \text{(I-25)}$$

where Δ_a and Δ_a^0 are the band halfwidths for the absorption and CD, respectively, which are the distances from the center of the band to the point on the bands where $\epsilon(\lambda)_{0a}$ or $[\theta(\lambda)_{0a}]$ falls to e^{-1} of its peak value, ϵ^0_{0a} or $[\theta^0_{0a}]$, respectively. If the transition is allowed, which can be assumed if $\epsilon_{0a} > 1000$, $\Delta_a = \Delta_a^0$ and $\lambda_a = \lambda_a^0$ (Moscowitz, 1965). In this case the absorption and CD bands of the transition are identically shaped and differ at any wavelength by a constant factor

$$[\theta(\lambda)_{0a}] / \epsilon(\lambda)_{0a} = (3298)\ 4R_{0a} / D_{0a}\ , \qquad \text{(I-26)}$$

where the factor $4R_{0a} / D_{0a}$ is the anisotropy factor of the transition (Condon, 1937). The assumption of the Gaussian shape for the absorption and CD bands allows us to obtain a simple expression for determination of the dipole and rotational strength of the transition. By substituting Eq. I-25 in Eq.I-21, and assuming $\lambda_a >> \Delta_a$ and $\lambda_a^0 >> \Delta_a^0$, we obtain (Moscowitz, 1960) :

$$D_{0a} \approx 9.180 \times 10^{-39} \sqrt{\pi \epsilon^0_{0a}}\ \Delta_a / \lambda_a \qquad \text{(I-27)}$$

and

$$R_{0a} \approx 0.696 \times 10^{-42} \sqrt{\pi [\theta^0_{0a}]}\ \Delta_a^0 / \lambda_a^0\ . \qquad \text{(I-28)}$$

e. One-electron Theory of Optical Activity.

The interaction of the electromagnetic field of light with the delocalized π electrons of the aromatic rings of Trp, Tyr and Phe side-chains generate considerable in-plane electric dipole transition moments connecting the ground states of those systems to a set of excited states, but the symmetry requires magnetic dipole transition moments in those systems to be perpendicular to the planes of the rings. Nevertheless, these aromatic chromophores do exhibit experimentally detectable optical activity in peptides and proteins. Several theories have been proposed to explain the mechanism that generates a nonzero rotational strength in these apparently symmetrical systems.

Condon, Altar and Eyring (1937) developed the so-called one-electron theory of optical activity. This theory describes the optical activity which arises from the mixing of zero-order states of differing local symmetry under the influence of the static electric field of the rest of the molecule. The mixing of states permits magnetically allowed and electrically forbidden transitions such as $n\pi^*$ to acquire electric dipole strength and become optically active. Conversely, electrically allowed but magnetically forbidden transitions such as $\pi\pi^*$ acquire magnetic dipole character through such mixing.

f. Exciton-Coupling Mechanisms of Optical Activity.

Kirkwood (1937) in his so-called polarizability theory rederived, using quantum mechanical formalism, the classical Born-Oseen-Kuhn coupled oscillator model (Kuhn, 1933). This theory explains the optical activity based on the treatment of two electric dipole transition moments as classical dipoles, coupled to produce a magnetic moment through the virtual rotation of one dipole about the other (Fig.I-3). Thus a circular

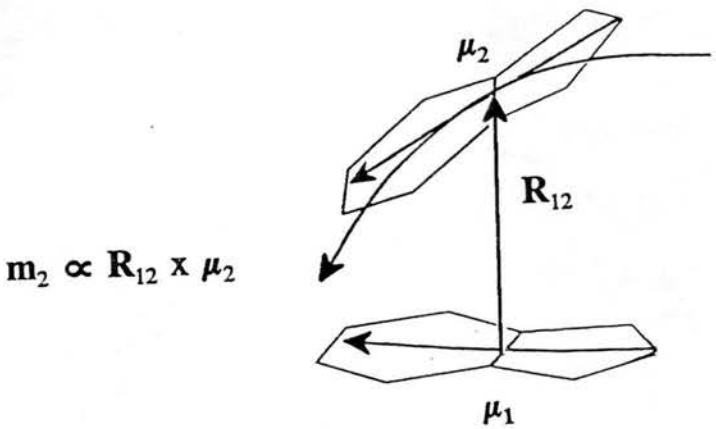


Figure I-3. Dipole-dipole model of exciton coupling interaction. The electric dipole transition moments μ_1 and μ_2 of the B_b transitions of two Trp are shown. $\mathbf{R}_{12}$ is the center-to-center distance vector connecting the middle of the CE2-CD2 bonds of the two Trp. The magnetic dipole transition moment of the coupled system $\mathbf{m}_2 = \mathbf{R}_{12} \times \mu_2$ results from the linear movement of charge within Trp2, defined by the dipole moment μ_2, which appears as a circular motion of charge relative to the center of Trp1.

motion of charge is achieved, giving rise to the magnetic moment. The coupled oscillator theory has been applied and further developed to describe optical activity in various systems (Moffitt, 1956; Schellman & Oriel; 1962; Tinoco, 1962; Harada & Nakanishi, 1983). For degenerate transitions, the rotational strength is given as (Moffit, 1956) :

$$R = \pm (\pi / 2\lambda_1) [\mathbf{R}_{12} \bullet \mu_2 \times \mu_1]. \qquad \text{(I-29)}$$

Schellman & Oriel (1962) introduced a mechanism for the coupling of the magnetically allowed transition of one chromophore with an electrically allowed transition of another. This is known as the "μ-**m**" mechanism.

The coupled oscillator model works for non-degenerate transitions as well, although the mixing between transitions won't be as complete as in the degenerate case due to differences in energy. In Kirkwood's theory for non-degenerate transitions, the rotational strength is calculated as

$$R = \{ - 2 V_{12}\nu_1 \nu_2 / \hbar c (\nu^2_1 - \nu^2_2)\}[\mathbf{R}_{12} \bullet \mu_2 \times \mu_1] \qquad \text{(I-30)}$$

The interaction potential in Kirkwood's theory, V_{12} , is derived in the dipole-dipole approximation. In matrix representation,

$$V_{12} = \mu_1 \bullet \mathbf{T}_{12} \bullet \mu_2 , \qquad \text{(I-31)}$$

where $\mathbf{T}_{12}$ is the dipole-dipole interaction tensor given by

$$\mathbf{T}_{12} = (1/R_{12}^3) (\mathbf{1} - 3\mathbf{R}_{12}\mathbf{R}_{12} / R_{12}^2), \qquad \text{(I-32)}$$

where $$\mathbf{1} = \mathbf{ii} + \mathbf{jj} + \mathbf{zz} \qquad \text{(I-33)}$$

is the unit tensor.

The coupled oscillator mechanism is the most important mechanism for the aromatic side-chain optical activity, and my calculations are based on this mechanism.

g. Generalization of the Theory of Optical Activity by Tinoco (1962).

Tinoco (1962) first developed a chiroptical formalism for polymers and protein-related structures by utilizing first-order perturbation theory to mix the wavefuctions for the various groups in a polymer and thereby obtain the wavefunctions for the interacting system. The final expression of Tinoco's model emphasizes all the important mechanisms for the generation of the optical activity described above. As in Kirkwood's theory, Tinoco suggested a division of the molecular system into N groups, between which electron exchange could be neglected. Then, the Hamiltonian of the system can be written as

$$\mathcal{H} = \sum_{i}^{N} H_i + V, \qquad \text{(I-34)}$$

$$\text{where } V = \sum_{i}^{N} \sum_{j>i}^{N} V_{is,jt} \qquad \text{(I-35)}$$

$$\text{and } V_{is,jt} = e^2 \sum_{s}^{\text{electrons}} \left[\sum_{t}^{\text{electrons}} 1/r_{is,jt} - \sum_{t}^{\text{nuclei}} Z_{jt} / r_{is,jt} \right]. \qquad \text{(I-36)}$$

Here H_i is the kinetic energy of the electrons in group i and that part of the potential energy which involves only the other electrons and nuclei of the same group. The potential energy V is the Coulombic energy of interaction of each electron s in group i with all electrons and nuclei t in all other groups j. The distance between particles *is* and *jt* is $\mathbf{r}_{is,jt}$. In the subsequent derivation, V was treated as a perturbation.

The group wave functions ψ_{i0}^{0}, ψ_{ia}^{0} ... in Tinoco's calculations are the eigenvectors of the Hamilonian H_i with eigenvalues E_0, E_a .... where 0 represents a ground state, and *a, b* ... singly excited states. As the electron exchange between the

groups is neglected, the zero-order wave functions of the system, Ψ_0^0, Ψ_A^0 ..., can be expressed as simple products of group wave functions. The zeroth order ground-state wave-function for the molecule Ψ_0^0 is :

$$\Psi_0^0 = \Pi_i \psi_{i0}^0 \tag{I-37}$$

The singly excited-state wave function in which the j-th group is excited to state *a*, Ψ_A^0, is :

$$\Psi_A^0 = (\psi_{ja}^0/\psi_{j0}^0)\, \Pi_i \psi_{i0}^0. \tag{I-38}$$

The doubly excited-state wave function in which the j-th group is excited to the *a*-th state and group k is excited to state b is :

$$\Psi_B^0 = (\psi_{ja}^0\psi_{kb}^0/\psi_{j0}^0\psi_{k0}^0)\, \Pi_i \psi_{i0}^0 \tag{I-39}$$

Using pertubation theory, the wave function for the first-order ground state would be

$$\Psi_0' = \Psi_0^0 - \sum_i \sum_{j\neq i} \sum_a V_{i0a,j00}\psi_{ja}^0 / h\nu_a - \sum_i \sum_{j>i} \sum_a \sum_b V_{i0a,j0b}\psi_{ia,jb} / h\,(\nu_a+\nu_b) \tag{I-40}$$

The wave function for the first-order nondegenerate excited state is :

$$\Psi_A' = \Psi_{ia}^0 + \sum_{j\neq i} V_{i0a,j00}\psi_0^0/h\nu_a - \sum_{j\neq i} \sum_{b\neq a} V_{i0a,j0b}\psi_{jb}^0/h\,(\nu_b - \nu_a) - \sum_{b\neq a} V_{iab,j00}\psi_{jb}^0/h\,(\nu_b - \nu_a), \tag{I-41}$$

where $\nu_a = (E_a - E_0)/ h$, (I-42)

$$V_{i0a,j00} = \int \psi_{i0}^*\psi_{j0}^*\, V\, \psi_{ia}\psi_{j0}\, d\tau\,, \tag{I-43}$$

$$V_{i0a,j0b} = \int \psi_{i0}^*\psi_{j0}^*\, V\, \psi_{ia}\psi_{jb}\, d\tau\,, \tag{I-44}$$

$$V_{iab,j00} = \int \psi_{ia}^*\psi_{j0}^*\, V\, \psi_{ib}\psi_{j0}\, d\tau\,. \tag{I-45}$$

These potential terms are the electrostatic interaction between the transition moments for transitions $0\rightarrow a$ or $0\rightarrow b$ in group i and the static dipole moment j00 or the transition

moment j0*b* in group j. The wave functions involving excited states of three or more different groups were omitted in the first-order wave functions, as their contribution to the rotational strength were expected to be small.

Using the first-order wave functions, Tinoco obtained the electric and magnetic dipole transition moments for the transition 0→A of the molecule, and the expression for the rotational strength R_A, which was used since then in numerous calculations for α- and 3_{10}-helices (Tinoco & Woody, 1967), for β-structure (Woody,1969) and for poly-L-Tyr (Chen & Woody, 1971).

$$R_{0A} = \sum_i [\ \mathrm{Im}\ \{\mu_{i0a} \bullet m_{ia0}\} \tag{I-46a}$$

$$-2\sum_{j\neq i}\sum_{b\neq a} \mathrm{Im}\ \{\ V_{i0a,j0b}\ (\mu_{i0a} \bullet \mathbf{m}_{jb0}\nu_a + \mu_{j0b} \bullet \mathbf{m}_{ia0}\nu_b\ )\}\ /\ h\ (\nu_b^2 - \nu_a^2\ ) \tag{I-46b}$$

$$-\sum_{j\neq i}\sum_{b\neq a} \mathrm{Im}\ \{\ V_{iab,j00}\ (\mu_{i0a} \bullet \mathbf{m}_{jb0} + \mu_{j0b} \bullet \mathbf{m}_{ia0})\}\ /\ h\ (\nu_b - \nu_a\ ) \tag{I-46c}$$

$$-\sum_{j\neq i}\sum_{b\neq a} \mathrm{Im}\ \{\ V_{i0b,j00}\ (\mu_{i0a} \bullet \mathbf{m}_{ib0} + \mu_{iab} \bullet \mathbf{m}_{ia0})\}\ /\ h\ \nu_b \tag{I-46d}$$

$$-\ (2\pi/c)\ \sum_{j\neq i}\sum_{b\neq a}\ V_{i0a,j0b}\ \nu_a\ \nu_b\ (\mathbf{R}_j - \mathbf{R}_i)\ (\mu_{j0b} \times \mu_{i0a}\ )\ /\ h\ (\nu_b^2 - \nu_a^2\ )], \tag{I-46e}$$

where μ_{i0a} and $\mathbf{m}_{i0a}$ are the electric and magnetic dipole transition moments associated with the transition from the ground state 0 to an excited state a in the group i; $\mathbf{R}_i$ is the position vector of the optical center of group i relative to a space-fixed coordinate system common to all groups in the molecule.

Each term in Equation I-46 has an explicit physical meaning. Term (a) is the intrinsic optical activity of the chromophoric group, which vanishes if the chromophore has one or more elements of reflection symmetry. Term (b) describes the mixing of electric dipole transition moments on one group with magnetic dipole transition moments on other groups. This is the μ-m term discussed by Schellman (1968). Terms (c) and (d)

arise from the mixing of excited states within individual chromophores due to the static field of the rest of the molecule. These are the one-electron contributions of Condon et al. (1937). Term (e) results from mixing of electric dipole transition moments in different chromophores, and corresponds to the coupled-oscillator mechanism emphasized by Kirkwood (1937). Physically, this term corresponds to the fact that a linear displacement of charge μ_{j0b} in group j leads to a circular displacement of charge about the center of group i and hence leads to a magnetic dipole transition moment.

An additional term was present in Tinoco's formulation (1962) that involved the change in electric dipole moment upon excitation. This term was an artifact of using the dipole-length formulation rather than the origin-independent dipole velocity formulation of the rotational strength (discussed in Chapter I.2.b). The coupled-oscillator mechanism is reflected in Equation 46 in terms (b) and (e) .

h. Dynamic methods.

A different approach to calculating the optical activity has been formulated by Applequist (1979) based on the earlier works of De Voe (1964, 1965). This is a classical all-order dynamical polarizability approach formulated in terms of induced dipole moments, oscillators and polarizabilities associated with the various groups into which the molecule is subdivided. A drawback with the De Voe/Applequist formalism is that the calculations must be done for each wavelength for which the CD is to be calculated, whereas in the exciton (or matrix) method described above, one only has to find the solution of the system once (diagonalize the matrix), from which all rotational strengths and dipole moments can be calculated.

2. Calculation of Exciton Coupling in Trp Chromophores.

a. Matrix Method.

A more convenient formalism of calculating both the coupled-oscillator and one-electron optical activity has been elaborated by Schellman and coworkers (Bayley et al., 1969) on the basis of a matrix approach, instead of first-order perturbation theory. This method emphasizes the same mechanisms of optical activity as in Tinoco's (1962) interpretation, although calculation of the eigenvectors, which are the wave functions of the system, becomes more convenient, and the interactions of all orders are included, rather than being limited to first order. In the matrix method, the molecule is divided into isolated groups between which electron exchange is neglected. Each of the groups has a characteristic set of electronic transitions. Hamiltonian matrix **V** is set up that describes the mixing of excited states of the various groups. In general for a system of N Trp, calculations of CD considering a single transition in each chromophore would require an N x N matrix, whereas if m transitions are considered in each chromophore an Nm x Nm matrix will have to be set up. The matrix below (Eq.I-47) is an example for a system of two chromophores with two transitions considered in each :

$$\mathbf{V} = \begin{vmatrix} \begin{pmatrix} E_{1a} & V_{1ab,200} \\ \cdots & E_{1b} \end{pmatrix} & \begin{pmatrix} V_{10a,20a} & V_{10a,20b} \\ V_{10b,20a} & V_{10b,20b} \end{pmatrix} \\ \begin{pmatrix} \cdots & \cdots \\ \cdots & \cdots \end{pmatrix} & \begin{pmatrix} E_{2a} & V_{2ab,100} \\ \cdots & E_{2b} \end{pmatrix} \end{vmatrix} \qquad \text{(I-47)}$$

The matrix is symmetrical, so only elements above the diagonal are shown.The diagonal elements of the matrix **V** represent the transition energies of the transitions in the isolated groups as obtained from experimental data. The off-diagonal matrix elements in the upper right and lower left corners of the matrix represent the mixing of excited states within different groups through coupled oscillator interactions. The off-diagonal elements in the upper left and lower right corners of the **V** matrix represent the mixing of transitions within a single group under the influence of the time-average field of the rest of the molecule, which give rise to CD through the one-electron mechanism (Condon et al., 1937). The information required to calculate the elements of this matrix is obtained from experiment, where available, or from molecular orbital calculations on models for the individual groups (Section I.2.c).

When this matrix is diagonalized, the eigenvalues of **V** are the energies of the composite transitions of the system that correspond to the individual exciton bands. The eigenvectors contain the mixing coefficients that describe the contribution of the various localized excited states to the excited state of the composite system of the molecule. These eigenvectors can be combined with row or column matrices containing the electric and magnetic dipole transition moments for the localized transitions to calculate the ∇ and $\mathbf{r} \times \nabla$ matrix elements that are the dipole velocity forms of the electric and magnetic dipole transition moments for the various excited states of the molecule. From these, the rotational strength can be calculated.

b. Origin-Independent Version of the Matrix Method.

The rotational strength as defined in Equation I-15, through the electric (Eq.I-18) and magnetic dipole transition moments (Eq.I-19), is a function of the distance operator **r**, and thus the magnetic moment operator depends on the choice of origin. The rotational strength defined as in Equation I-15 also obeys the sum rule (Condon et al., 1937) :

$$\sum_a R_{0a} = 0 , \qquad \text{(I-48)}$$

which states the fact that the net rotational strength of all the transitions in the molecule is zero. The origin dependence of the rotational strength can be circumvented by representing the electric dipole transition moment through the gradient operator ∇ instead of the distance operator **r**, where the form of the ∇ operator is by definition :

$$\nabla = \mathbf{i}\partial /\partial x + \mathbf{j}\,\partial /\partial y + \mathbf{k}\,\partial /\partial z, \qquad \text{(I-49)}$$

where **i**, **j** and **k** are the unit vectors in the Cartesian coordinate system. A modification of the matrix method (Bayley et al., 1969) allowing calculation of the rotational strength through ∇ and **r** x ∇ has been introduced by Goux & Hooker (1980), to ensure origin independence of the method.

Considering the quantum mechanical definition of the momentum through the gradient operator

$$\mathbf{p} = -i\hbar \nabla , \qquad \text{(I-50)}$$

for a system of localized charges we obtain for the *magnetic moment operator* :

$$\mathbf{m} = (2mc)^{-1} \sum q_s \mathbf{r}_s \times \mathbf{p}_s = (-i\hbar /2mc) \sum q_s \mathbf{r}_s \times \nabla_s \qquad \text{(I-51)}$$

or, assuming all $q_s = e$: $\qquad \sum_s r_s \times \nabla_s = (i\, 2mc / \hbar e)\, \mathbf{m} \qquad \text{(I-52)}$

Here m and e have their usual meaning, and $\hbar = h/2\pi$, where h is Planck's constant. The *electric dipole moment operator* μ can be defined through the gradient operator ∇ by applying Heisenberg's equation of motion to the distance operator :

$$i\hbar\,(\partial \mathbf{r} / \partial t) = [\mathbf{r}, H] \qquad \text{(I-53)}$$

$$(\partial \mathbf{r} / \partial t) = \mathbf{p}/m = (-i/\hbar)\,[\mathbf{r}, H] \qquad \text{(I-54)}$$

$$\nabla = (m / \hbar^2)\,[\mathbf{r}, H] \qquad \text{(I-55)}$$

$$(0 \mid \nabla \mid a) = (m / \hbar^2)\,(E_a - E_0)\,(0 \mid \mathbf{r} \mid a). \qquad \text{(I-56)}$$

Then for the dipole moment operator we obtain

$$\mu = e\mathbf{r} = \{e\,\hbar^2 / m\,(E_a - E_0)\}\,\nabla = \{e\,\hbar / 2\pi m\,(\upsilon_a - \upsilon_0)\}\,\nabla\,. \qquad \text{(I-57)}$$

Through Equations I-51 and I-57, the definition of the rotational strength becomes origin independent. Caution should be exercised in using the gradient operator form for rotational strength as: 1) this substitution is strictly valid only for exact wave functions, and 2) it might lead to a violation (Harris, 1969) of the sum rule for the rotational strength (Eq.I-48). In practice, one obtains the wave functions as exactly as possible, and obedience of the resulting rotational strength values to the sum rule is tested for each calculation.

In this origin-independent formalism (Goux & Hooker, 1980a), instead of the rotational strength defined by Eq.I-15, the transition linear ($\mathbf{p}_{0a}$) and angular momentum ($\mathbf{L}_{a0}$) are utilized to calculate the chiral strength C_{0a}, which is analogous to the rotational strength. The chiral strength is :

$$C_{0a} = (\alpha / 3)\,\mathrm{Re}\,[\,\mathbf{p}_{0a} \times \mathbf{L}_{a0}\,], \qquad \text{(I-58)}$$

where α is the fine structure constant ($\alpha \equiv e^2/\hbar c$), and transition linear and angular momenta are determined from the electric (I-57) and magnetic dipole transition moments (I-51). The results of the calculations can be compared with experimental CD spectra by calculating molar ellipticities :

$$[\theta] = 18 \times 10^{-7} \ (Ne^2/\sqrt{\pi}mc^2) \ (\lambda_{0a}{}^2 / \ \Delta_{0a}) \exp\{-[(\lambda-\lambda_{0a})/\Delta_{0a}]\}^2 \ C_{0a} \qquad \text{(I-59)}$$

where λ_{0a} is the characteristic wavelength in nm. The value of the numerical factors at the beginning of the term is 1.7234×10^4. The chiral strength can be converted to the rotatory strength :

$$R_{0a} = (3he^2 / 8\pi^2 mc) \lambda_{0a} C_{0a} \qquad \text{(I-60)}$$

The approach used in this, as in other works (Manning & Woody, 1989; Sreerama et al,, 1991; Woody, 1994) incorporates the direct calculation of the rotational strength from ∇ and $\mathbf{r}\times\nabla$:

$$R_{0a} = (e^2\hbar^3 / 2mc \ E_{oa}) \ \langle\psi_o|\nabla|\psi_a\rangle \bullet \langle\psi_a \ |\mathbf{r}\times\nabla|\psi_0\rangle \qquad \text{(I-61)}$$

where E_{oa} is the energy of the transition $0 \rightarrow a$.

The matrix method (Bayley, 1969) in its origin-independent version (Goux & Hooker, 1980a) has been applied to calculations of optical activity for side-chain chromophores in proteins. Calculations of the near-uv CD have been performed for ribonuclease S (Goux & Hooker, 1980a), considering transitions in Tyr, Phe, His and disulfide bonds, and hen egg lysozyme (Goux & Hooker, 1980b), which has six Trp chromophores. In both cases, only partial agreement in position and intensities of experimental bands has been achieved, although important data have been gathered regarding possible interactions of various chromophore transitions in globular proteins.

c. Parameters in the Matrix Method.

The application of either Tinoco's method or the matrix method to the calculation of optical activity requires information about the excited state energies, and both the magnitude and direction of the magnetic and electric dipole transition moments in all the chromophores considered in the molecule.

(1) Assignment of Transitions in the Indole Spectrum.

Platt (1949) applied group theory and symmetry rules to a number of π-electron systems (aromatic chromophores), which can be extended to indole. Platt's theoretical approach resulted in prediction of various electronic transitions in indole which have been confirmed experimentally (Yamamoto & Tanaka, 1972). Platt's (1949) model treats indole as a perturbed cyclodecapentaene, an aromatic system with 10 π electrons donated from carbon atoms. The perturbations are made by the covalent bond between the CD2 and CE2 atoms and the replacement of two carbons in the ring with a nitrogen atom. The model predicts four excited states for the indole chromophore: two low-energy excited states which were named L_b and L_a, and two high-energy excited states named B_b and B_a. In Platt's model of the indole electronic excited states, the B states are fully allowed. In unperturbed cyclodecapentaene transitions to the L states are forbidden, but they become weakly allowed in indole due to the perturbations.

Yamamoto & Tanaka (1972) have obtained absorption spectra of Trp model compound (3-indolylacetic acid) in water, and identified four electronic transitions between 180 nm and 300 nm (Table I-1). The observed transitions are assigned, in order of increasing energy, to the L_b, L_a, B_b and B_a excited states according to Platt's

Table I-1. Comparison of Theory and Experiment for Indole Transitions.

Transition	Experiment			Calculated[d]		
	λ (nm)[a,b,c]	f[a]	θ (deg)	λ (nm)	f	θ (deg)
L_b	282	0.01	49[a], 58[a] 45[b], 60[c]	284	0.007	80
L_a	269	0.11	-42[a], -30[a]	259	0.107	-56
B_b	214	0.68		213	0.807	15
B_a	200			200	0.117	-44
V				193	0.187	-18
VI				190	0.399	-80

(a) Yamamoto & Tanaka (1972); (b) Philips & Levy (1986a); (c) Philips & Levy (1986b); (d) Woody (1994), parameters of Nishimoto & Forster (1965, 1966).

definition (Table I-1 and I-2). The L_b and L_a transitions of indole are readily observable in the near uv spectra of many model compounds, and there is little doubt about their assignment (Strickland, 1974; Kahn, 1979; Callis, 1991). Auer (1973) reported far-uv absorption and CD spectra of a number of indole derivatives in water and hydrophobic solvents. Auer also performed deconvolution analysis of the absorption and CD spectra, and identified five to six Gaussian components in each spectrum. The band occuring at highest energy in these spectra, between 190 nm and 198 nm, was attributed to the B_a indole transition. This band is of high intensity and lacks vibrational structure, which is characteristic of strongly allowed transitions. The other Gaussian bands resolved in the far uv are most probably vibronic components of the B_b transition, although Auer suggested another C transition in that region which has not been confirmed.

(2) Transition Dipole Moments.

The energies of the transitions and magnitudes of the electric dipole transition moments can be obtained from experiments on model compounds (Yamamoto & Tanaka, 1972; Auer, 1973). The oscillator strengths estimated for aqueous solution of L-Trp (Woody, 1994) are reported in Table I-1. Electric dipole transition moment directions have commonly been taken from linear dichroism measurements on single crystals or stretched films. Recently, the L_b transition moment direction has been determined for indole (Philips & Levy, 1986a) and for tryptamine (Philips & Levy, 1986b) in the gas phase by rotationally resolved absorption spectroscopy on supercooled jets. Given the differences in the environment and molecular structure, the polarization for the transitions in crystalline 3-indolylacetic acid ($+49^\circ$ with respect to the long axis) and Gly-Trp

Table I-2. Energy, dipole and magnetic moments for the electronic transitions in indole.

Transition	E (eV)	λ (nm)	/μ/ (D)	∇_x (Å^{-1})[a]	∇_y (Å^{-1})	∇_z (Å^{-1})	m_z (BM)[b]
L_b	4.3966	189.5	0.774	0.028321	0.088570	0.0	-0.0633
L_a	4.6091	193.4	2.508	0.208254	-0.237381	0.0	0.1114
B_b	5.5105	225.0	5.702	0.832213	0.210753	0.0	-0.4533
B_a	6.1902	200.3	2.231	0.279420	-0.253608	0.0	-0.0758
V	6.4118	269.0	2.773	-0.459844	0.156315	0.0	-0.5786
VI	6.5414	282.0	4.006	-0.129107	0.704400	0.0	-0.2384

[a] The long diagonal of the indole lies on the x axis; the y axis is drawn through the CE2-CD2 bond; and the z axis originates in the middle of the CE2-CD2 perpendicular to the plane of the indole ring.

[b] m_z were calculated with respect to the center of the indole ring, defined as the middle of the CE2-CD2 bond. The units of m are Bohr magnetons (0.9224×10^{-21} erg/Gauss). All m_x and m_y equal to zero.

(+58°) agree reasonably well with those for indole (+45°) and tryptamine (+60°) in the gas phase (Table I-1).

The magnitude of the magnetic dipole transition moment and the directions of both magnetic and electric dipole transition moments are difficult or impossible to obtain from experiment. Even in those cases where the electric dipole transition moment direction is determined by symmetry or from reliable experimental data, calculation of the Coulomb interactions between transition charge densities requires additional information, unless the chromophores are sufficiently far apart that the point dipole approximation can be safely used. Generally, a distributed dipole or monopole approximation (Tinoco, 1962) is used to approximate the transition or stationary state charge densities.

(3) Monopole Approximation.

Tinoco proposed a method to calculate the interaction potential that needs the group wavefunctions. The interaction potential $V_{i0a,j0b}$ was approximated by the electrostatic interaction between point monopoles, the idea of which was first introduced by London (1942) and extended by Haugh and Hirschfelder (1955). Practically the transition charge density is approximated by point charges. These point charges are chosen so that they reproduce the electric dipole transition moment. In the case of an electrically forbidden transition, for which the total dipole transition moment equals zero, the point charges would reproduce the lowest non-vanishing moment of the transition charge density. Transition monopoles are defined by

$$q_{it0a} = e \int \psi_{i0}^* \psi_{ia} d\tau_t. \qquad \text{(I-62)}$$

The integration is carried out over the region t in which the sign of the expression under the integral is the same throughout.

$$q_{i0a} = \sum_t q_{it0a} = 0 \qquad \text{(I-63)}$$

because ψ_{i0} and ψ_{ia} are orthogonal to each other. The position of the monopole is given by

$$\mathbf{r}_{it} = \left(\int \psi_{i0}^* \mathbf{r} \psi_{ia} d\tau_t \right) / \left(\int \psi_{i0}^* \psi_{ia} d\tau_t \right). \qquad \text{(I-64)}$$

The V matrix elements are then calculated using Coulombs law:

$$V_{i0a;job} = \sum_{i,s} \sum_{j,t} q_{is0a}\, q_{it0b} / \; r_{is,jt} \qquad \text{(I-65)}$$

where $r_{is,jt} = \left| \mathbf{r}_{jt} - \mathbf{r}_{is} \right|$ is the distance between the monopoles *is* and *jt*, and q_{is0a} is the charge of the monopole *s* associated with the transition $0 \rightarrow a$ in chromophore *i*.

(4) Parameters from Molecular Orbital Calculations.

The transition monopoles and energies of the unperturbed transitions (L_a, L_b, B_a, B_b , transition V at 193.4 nm, and transition VI 189.5 nm) used in our calculations of CD were determined from π-MO (molecular orbital) calculations of the Pariser-Parr-Pople type (Murrell and Harget, 1972) with Nishimoto-Forster (1965, 1966) parameters for π-electrons. The geometry of the indole has been taken from the crystal structure of 3-indolylacetic acid (Karle et al., 1964).

Transition monopoles and ground-state charges for the six lowest-energy $\pi\pi^*$ transitions were evaluated from the molecular orbitals obtained after deorthogonalization (Shillady et al., 1971).The magnitudes of the monopoles are calculated as

$$\rho_a = \sum_i A_i \rho_{ia} = \sqrt{2} \sum_i A_i C_{0a} C_{ua} \; . \qquad \text{(I-66)}$$

Here ρ_a is the monopole charge in units of e, the electronic charge; A_i is the coefficient of configuration i, corresponding to the excitation $0 \rightarrow u$, in the CI expansion; C_{0a} and C_{ua} are the MO coefficients of the π orbital on atom a in occupied orbital 0, and u, respective. In the monopole approximation that I used, the monopole charges for nine monopoles were calculated and placed at the positions of the carbons and the nitrogen of the indole ring (Table I-2).

In our calculations, the transition monopole charges for indole transitions calculated from the π-electron wave functions were multiplied by factors specific for each transition but lying in a narrow interval from 0.745 to 0.824. This scaling has been performed to reproduce the experimental oscillator strengths, available for the L_b, L_a and B_b transitions. Monopole charges for transitions connecting the excited states were scaled by a factor that is the average of the scaling factors for the transitions from the ground state, equal to 0.793. The parameters used are shown in Table I-2 and I-3.

The calculated values for the transition energies were correlated, whenever possible, with experimental observations for Trp and Trp derivatives in various solvents (Auer, 1973) and with those observed CD bands in globular proteins that are most likely due to aromatic contributions (Woody, 1994). In fact π-MO calculations on indole give good agreement with experiment (Woody, 1987; 1994).

The first two transitions (L_b, L_a) were placed at the energies reported by Yamamoto and Tanaka (1972), and their transition dipole velocity magnitudes were adjusted to agree with experimental oscillator strengths, but the transition dipole velocity directions were taken from the calculated transition dipole moments.

The B_b transition was placed at 5.51 eV (225 nm) and the transition dipole velocity magnitude was adjusted to reproduce the observed oscillator strength (Yamamoto & Tanaka, 1972). For the higher energy transitions, transition dipole velocity magnitudes were adjusted to give agreement with the theoretical oscillator strengths calculated as the geometric mean of the dipole length and dipole velocity oscillator strengths (Hansen, 1967).

Both $\pi\pi^*$ MO calculations and CNDO/S (Complete Neglect of Differential Overlap, Spectroscopic version) calculations predict several electrically allowed transitions of considerable intensity below the B_a transition. For these higher energy transitions, which were predicted at 6.54 eV (189.5 nm) and 6.19 eV (193.4 nm), no experimental data are available at present. Including these two transitions with energies higher than B_a in our calculations has a basis in experimental observations of absorbance and CD of indole and its derivatives (Auer, 1973) and in the results of estimations of aromatic contributions in globular proteins through spectral deconvolution (Bolotina & Lugauskas, 1985; Chapter III). In both cases, intense bands centered around 200-205 nm were observed.

Table I-3. Ground state and transition monopoles for the electronic transitions in indole.

	Indole ground state atomic charges*								
	q_{CG}	q_{CD1}	q_{CD2}	q_{NE1}	q_{CE2}	q_{CE3}	q_{CZ2}	q_{CZ3}	q_{CH2}
	-0.459287	-0.350480	-0.237033	1.701983	-0.303044	0.014631	-0.219699	-0.112991	-0.034060
Mixing of transitions	Transition monopoles*								
0 - L_b	-0.003951	0.160270	0.183745	-0.365614	0.280513	0.030737	-0.308335	-0.044775	0.067411
L_b- L_a	0.264759	-0.358272	-0.137726	0.222249	-0.119752	0.212724	0.022072	0.286639	-0.392691
L_b- B_b	0.858572	0.395888	-0.807709	-0.143556	-1.714903	1.404722	1.350768	-1.421786	0.078094
L_b- B_a	0.326882	-0.423523	-0.238982	0.321822	-0.185441	0.355689	0.215905	0.168955	-0.541306
L_b- V	-0.070788	-0.111187	0.706933	0.070650	0.221404	-0.724581	-0.460342	-0.135039	0.502949
L_b- VI	0.098164	0.534948	0.367310	0.139047	-0.325504	-0.335135	-0.235916	-0.803089	0.560174
0 - L_a	-0.637774	0.763618	0.442475	0.075043	-0.373892	-1.010600	0.924834	0.593960	-0.777662
L_a- B_b	0.332474	-0.010800	0.070341	0.489321	-0.736465	-0.053813	0.186646	-0.435879	0.158176
L_a- B_a	-0.853288	-0.312000	0.302752	-0.287371	-0.311630	0.504626	0.709791	-0.075879	0.322999
L_a- V	- 0.275571	0.204139	0.098027	-0.113877	-0.128177	0.005284	-0.372902	-0.006382	0.038316
L_a- VI	-0.461135	-0.644743	-0.030278	0.063927	-0.017519	0.239755	0.475055	-0.029716	0.404655

* units of ground state charges and transition monopoles are 10^{-10} esu.

Table I-3 (Continued). Ground state and transition monopoles for the electronic transitions in indole.

Mixing of transitions	Transition monopoles								
	q_{CG}	q_{CD1}	q_{CD2}	q_{NE1}	q_{CE2}	q_{CE3}	q_{CZ2}	q_{CZ3}	q_{CH2}
0 - B_b	-0.054692	0.904626	0.456463	0.338668	-0.274699	0.416004	-0.476148	-0.613996	-0.696226
B_b - B_a	-0.243893	-0.247500	0.359036	-0.206650	0.001372	0.084601	-0.093968	0.029995	0.317007
B_b - V	-0.673134	-0.375125	0.485153	0.073315	0.360795	-0.109628	0.044853	0.075748	0.118022
B_b - VI	0.062635	-0.506552	0.295251	0.133926	-0.487415	0.121788	0.011130	0.075236	0.294003
0 - B_a	-0.048207	-0.149456	-0.706761	0.439226	1.017310	0.468824	-0.385815	-1.073380	0.438260
B_b - V	-0.231687	0.702509	-0.548159	0.038009	0.003203	0.087978	0.196893	-0.231897	-0.016849
B_b - VI	-0.266891	0.410312	0.338840	-0.181501	-0.299429	-0.135928	0.002658	-0.078307	0.210246
0 - V	-0.660496	-0.149602	-0.059869	0.589353	-0.631161	0.586731	-0.529822	0.619001	0.235864
V - VI	-0.600746	-0.771867	-0.061936	0.210443	-0.054913	0.351468	0.237252	0.422398	0.267902
0 - VI	0.617767	0.021256	0.105287	-0.488358	-0.746207	0.635455	-0.452660	0.544312	-0.236853

Chapter II. Prediction of Tryptophan Circular Dichroism in Globular Proteins.

1. Introduction.

A number of proteins have been found to have positive CD bands near 230 nm (Woody, 1987). Such positive bands are generally assumed to result from Tyr or Trp side chains (Auer, 1973; Shiraki, 1969) because peptide contributions exhibit negative CD in this region in all conformations, except the rarely occurring poly-(Pro)II (Woody, 1992).

We applied the exciton approach (Davydov, 1949; Moffitt, 1956; Tinoco, 1962; Harada and Nakanishi, 1983), based on interactions between the transition dipole moments of identical chromophores, to calculate the CD of Trp side chains interacting in globular proteins.

2. Methods.

Calculations were performed by the origin-independent version (Goux & Hooker, 1980a) of the matrix method (Bayley et al., 1969). In this method, the molecule is divided into groups between which electron exchange is neglected, each of which has a characteristic set of electronic transitions. The elements of the energy matrix are calculated by a monopole approximation in which the transition charge density is approximated by point charges that reproduce the electric dipole transition moment. The transition monopoles and the transition moment directions for the six lowest energy indole transitions were obtained (Woody, 1987; 1994) from MO calculations of the Pariser-Parr-Pople type (Murrell & Harget, 1972) with Nishimoto & Forster (1965,

1966) parameters. The energies for the six transitions (L_b at 4.40 eV, 282 nm; L_a at 4.61 eV, 269 nm; B_b at 5.51 eV, 225 nm; B_a at 6.19 eV, 200.3 nm; transition V at 6.41 eV or 193.4 nm; and transition VI at 6.54 eV or 189.5 nm) and transition dipole moment magnitudes were based upon experimental values (Cosani et al., 1968; Yamamoto & Tanaka, 1972), where available, as described in Chapter 1. A matrix is constructed so that its diagonal elements are the energies of the transitions in the isolated groups, and the off-diagonal elements represent the mixing of excited states within different groups (coupled-oscillator interaction) or within a single group (one-electron interaction). After diagonalization, the eigenvalues of this matrix are the energies of the composite transitions of the system, while the eigenvectors contain the mixing coefficients that describe the contribution of the various localized excited states to the excited states of the composite system. From these eigenvectors, the matrix elements for the ∇ and $\mathbf{r}\times\nabla$ operators can be calculated, corresponding to the dipole velocity forms of the electric and the magnetic dipole transition moments, respectively. From these, the rotational strengths are calculated.

Couplet strength was calculated as (Bayley, 1973) :

$$S \equiv \Delta\epsilon_+ - \Delta\epsilon_- = 3.92\ \lambda_i\ (\Delta\lambda)R\ /\ \Delta_i^2 \qquad (II-1)$$

where subscripts (+) and (-) indicate the long-wavelength and the short-wavelength exciton bands, respectively; λ_i is the wavelength at the absorbance maximum of the transition, which is 225 nm for the B_b; Δ_i is the half bandwidth ; $\Delta\lambda$ is the theoretical splitting between the long-wavelength (+) and the short-wavelength (-) exciton bands;

and R is the mean value of the rotational strengths for the long-wavelength band, R_1, and the short-wavelength band, R_2 : $R = (R_1 - R_2)/2$.

For the CD calculations, a Gaussian shape of the CD bands was assumed. Half bandwidths were calculated from the empirical equation (Brown et al., 1971) :

$$\Delta_i = k\lambda_i^{1.5} \qquad \text{(II-2)}$$

The constant k was chosen so that the B_b band at 225 nm has a half bandwidth of 9.1nm.

Crystallographic coordinates for the Trp residues in globular proteins were obtained from the Protein Data Bank (PDB version 1990). Superposition of Trp residues in chymotrypsin and chymotrypsinogen structures was performed using INSIGHTII software (Biosym, Inc., San Diego, CA) on a Silicon Graphics Iris workstation.

3. Calculations of Circular Dichroism of Tryptophan Chromophores.

a. Exciton Coupling of B_b Transitions in Trp Pairs.

The exciton interaction (Davydov, 1949) between two degenerate excited states in indole results in a CD spectrum that is the sum of two components of opposite sign (Figure II-1), separated by an energy gap. The exciton couplet strength (Bayley, 1973) for the B_b band has been calculated for 150 Trp pairs from 118 proteins. Figure II-2 shows the calculated couplet strength as a function of the center-to-center distance of the Trp pairs. Each point corresponds to a Trp pair. The CD of a Trp pair depends not only on their distance but also on their relative orientation. Thus at a given distance between two Trp, the CD intensity ranges from a maximal negative to a maximal positive value, depending on the relative orientations of the Trp residues.

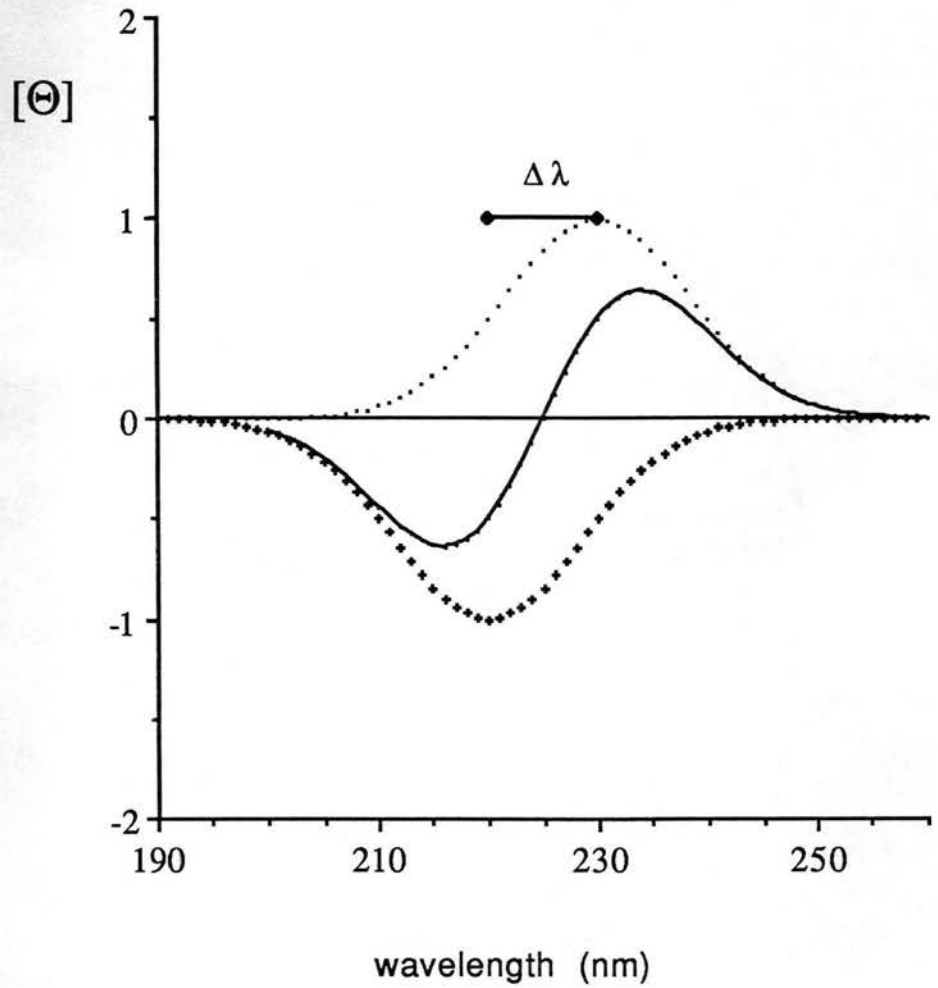


Figure II-1. Exciton couplet CD (-) as the sum of positive (•) and negative (+) Cotton effects.

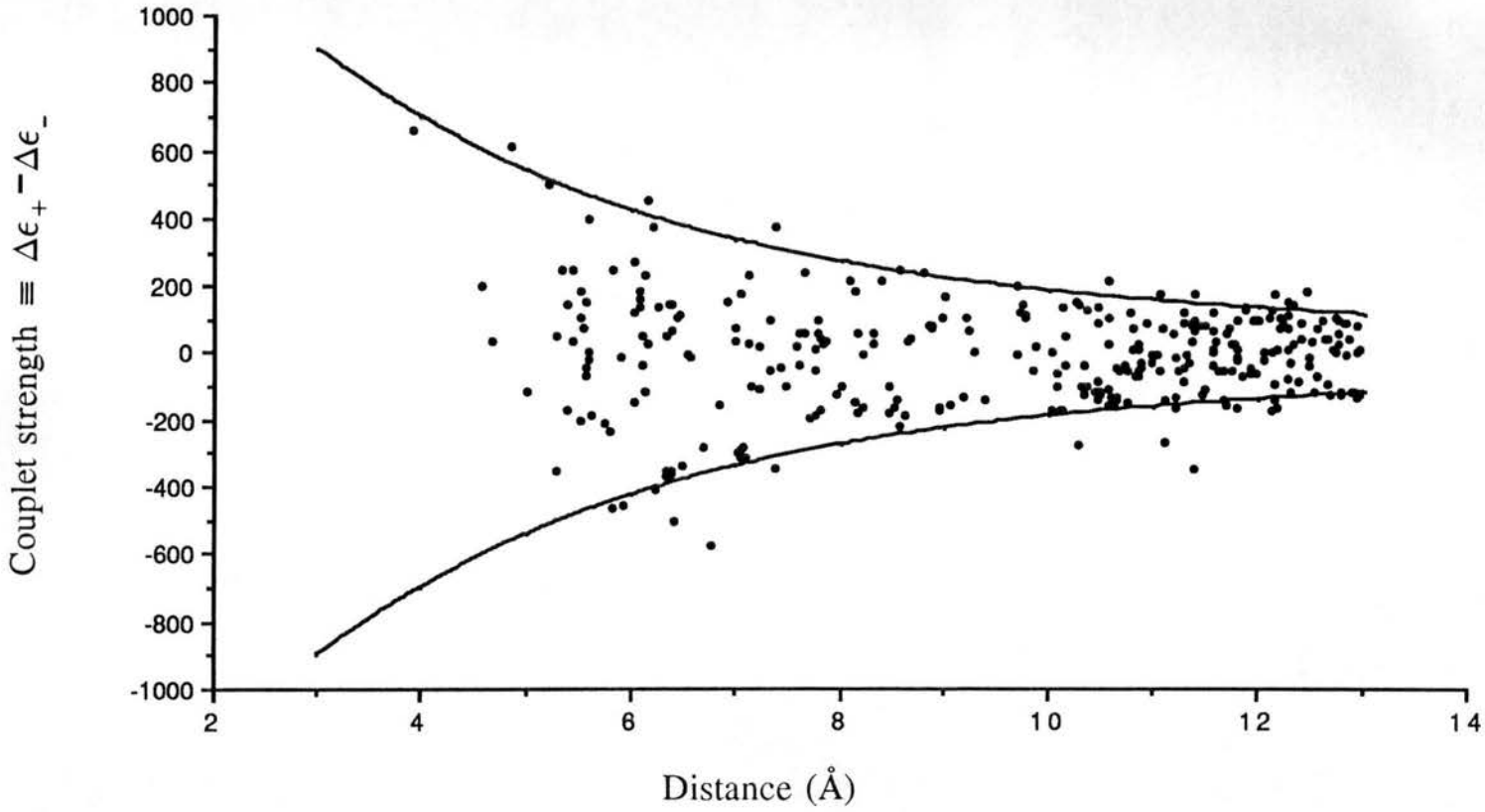


Figure II-2. Dependence of couplet strength of Trp pairs in globular proteins on the center-to-center distance between Trp (•). Distance dependence for two Trp in parallel planes with their center-to-center vector normal to both planes and one ring rotated 45° relative to the other is shown with continuous lines.

The theoretical maximal CD (or maximal couplet strength), in the dipole-dipole approximation, results when the two Trp rings are in parallel planes, with their center-to-center vector normal to both planes and one ring rotated by 45° relative to the other (Fig.II-3a). This conformation of a pair of Trp will be referred to as the P45 geometry. The distribution of couplet strengths calculated for Trp pairs observed in protein crystal structures has been found to fit within the limits for the P45 geometry, given as the outer pair of curves in Figure II-3.

The distance dependences of the rotational strength and the exciton splitting were examined separately. The distribution of calculated rotational strengths and exciton splittings as functions of distance fall within the range expected for the stacked geometry, as shown in Figures II-4 and II-5, respectively. The Trp geometry that corresponds to maximal rotational strength is shown in Figure II-3b. The two Trp are in stacked in parallel planes and one ring is rotated 90° relative to other (Fig.II-3b).

Maximal exciton splitting is predicted in dipole-dipole approximation when the two Trp are in parallel planes with their long axes parallel, i.e., at 0° rotation. As can be seen (Fig.II-5) the dipole-dipole approximation underestimates the maximal splitting at short distances (below 12 Å) compared with calculations in the monopole approximation (Chapter I-2; Chapter II.2), whereas estimation of maximal rotational strength (Fig.II-4) and, especially, maximal couplet strength (Fig.II-2) define correctly the upper limits of the statistical distribution. The reason is that the conformations of Trp pairs which give maximal splitting (Fig.II-5) also have the rotational strength close to zero, thus their contribution to the couplet strength (Eq.II-1) is small.

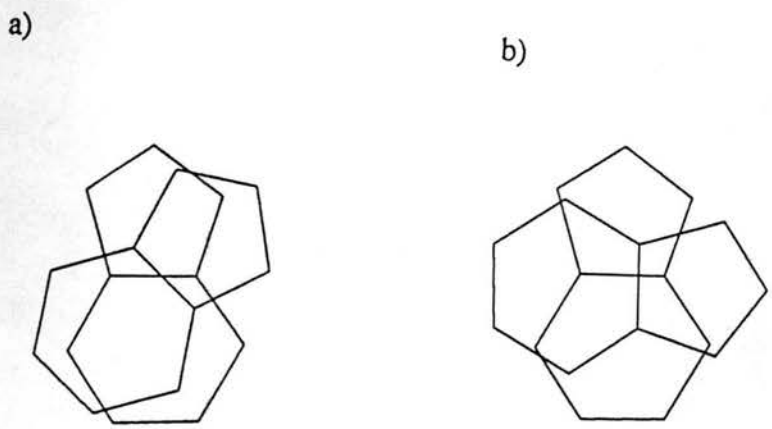


Figure II-3. Idealized conformations of Trp pairs: a) two Trp in parallel planes with their center-to-center vector normal to both planes and one ring rotated 45° relative to the other (P45 geometry, see text); b) two Trp in parallel planes with their center-to-center vector normal to both planes and one ring rotated 90° relative to the other.

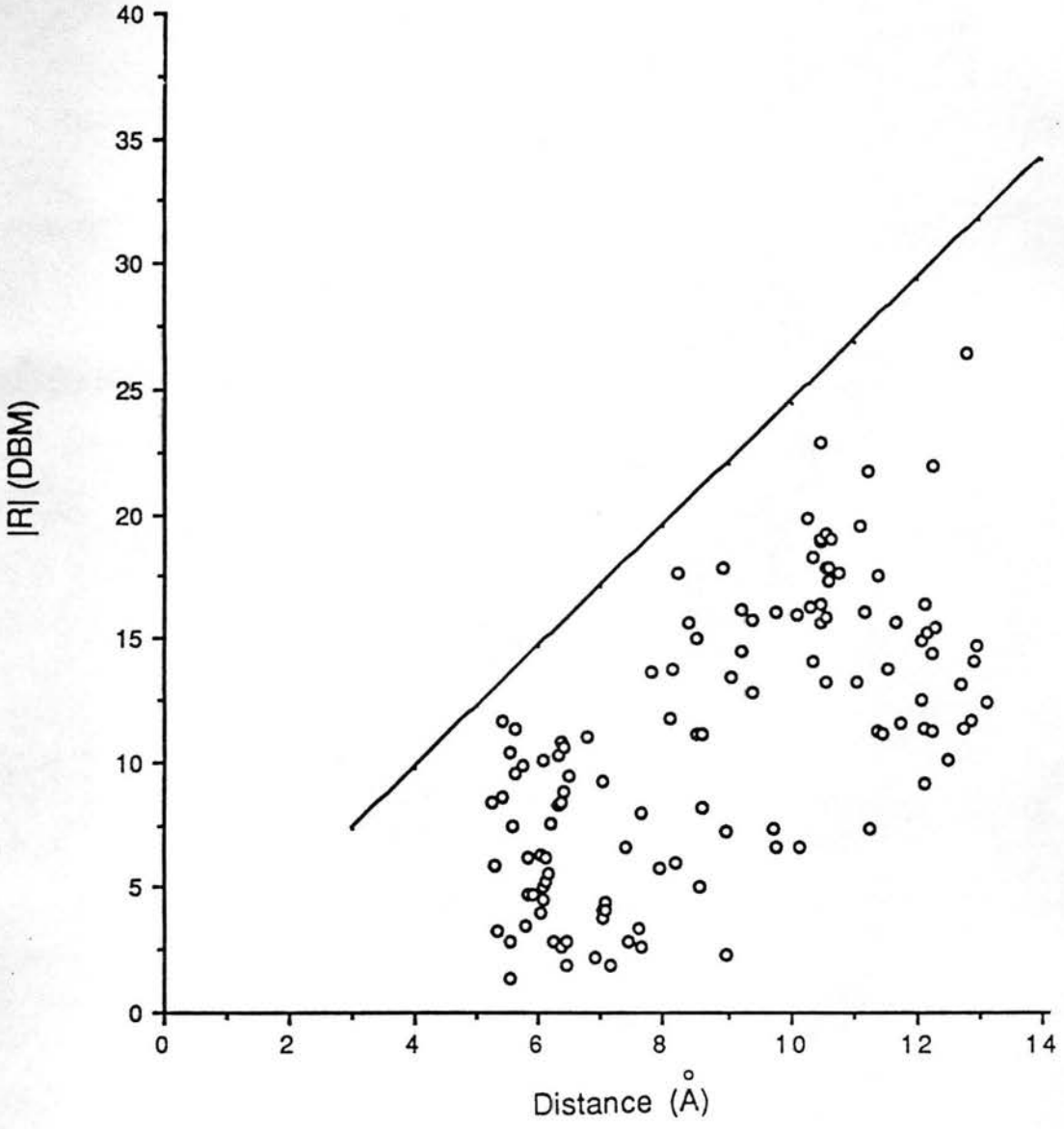


Figure II-4. Distance dependence of rotational strength of Trp pairs in globular proteins (o). Maximal rotational strength in dipole-dipole approximation corresponding to the orientation of Trps shown in Figure II-3b .

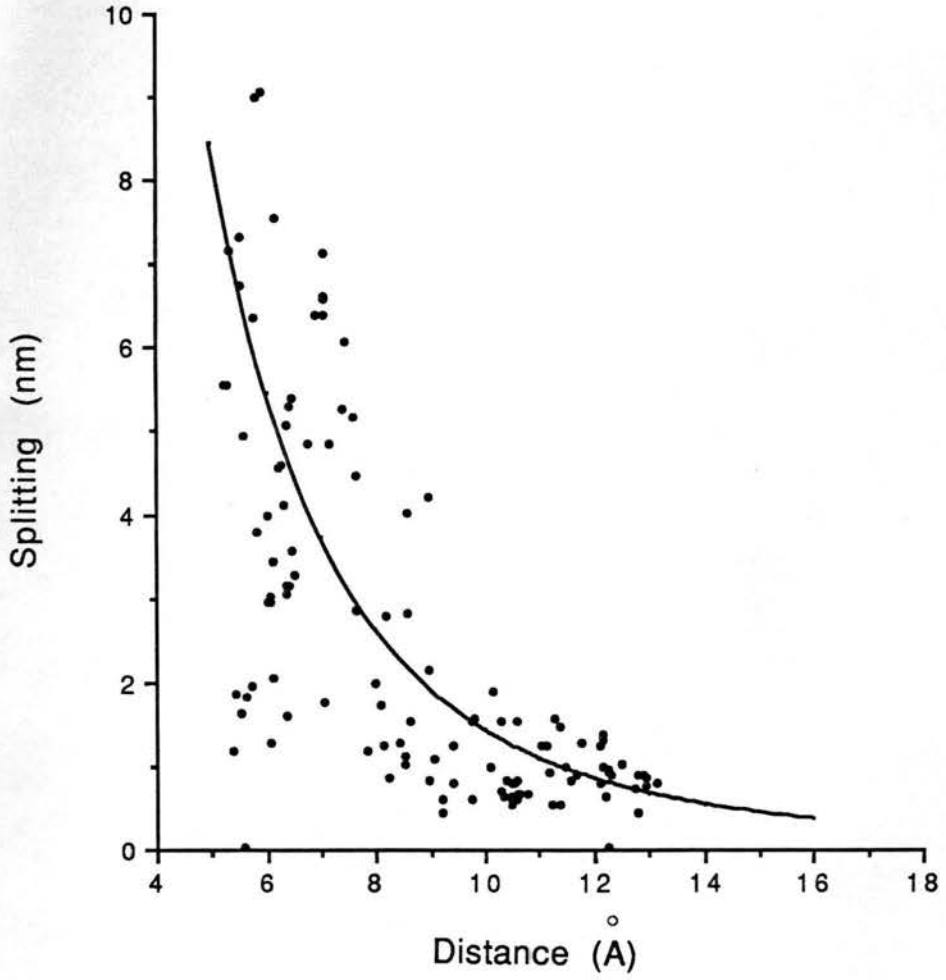


Figure II-5. Distance dependence of exciton splitting of Trp pairs in globular proteins (•). Exciton splitting corresponding to the parallel stacked geometry with 0° rotation (solid line). See text.

b. Analysis of CD of a Trp Pair. Directions for New Approaches.

As it appears that the predicted CD of Trp pairs in globular proteins might reach significant amplitudes in both the near and far uv, it seems reasonable to ask what kind of information we can obtain by applying our theoretical analysis to experimental data, and on the other hand what kind of characteristics of the globular protein should we account for to make our predictions more realistic. This chapter deals with four questions that come to mind: the effect of varying hydrophobicity of the Trp environment within the protein, or, in other words, the effect of Trp exposure to the solvent; the effect of nearby uncompensated charges or ion pairs; correlation between the changes in amplitude of near- and far-uv transitions; and dependence of the optical activity of a Trp pair upon its conformation. These questions show directions for application of theoretical approaches in the analysis of aromatic contributions in the CD of globular proteins.

(1). Accounting for Trp Exposure to the Solvent.

Based on the experimental data on absorption and CD of Trp model compounds in solvents with varying hydrophobicity, the B_b transition occurs at 219 nm in H_2O and around 223 - 225 nm in hydrophobic solvents (Cosani et al, 1968; Auer, 1972). The latter results correlate with observations of anomalous positive bands in CD spectra of some globular proteins around 225-230 nm (Woody, 1994), which suggest that the Trp B_b band is near 225 nm for Trp side chains buried in a protein. If we take different energies for the indole transitions but maintain the degeneracy and use the same transition monopoles, ground state charges and the value for the intrinsic magnetic moment (Table I-3), no qualitative changes in Trp spectra are predicted (Fig.II-6): only a shift in the

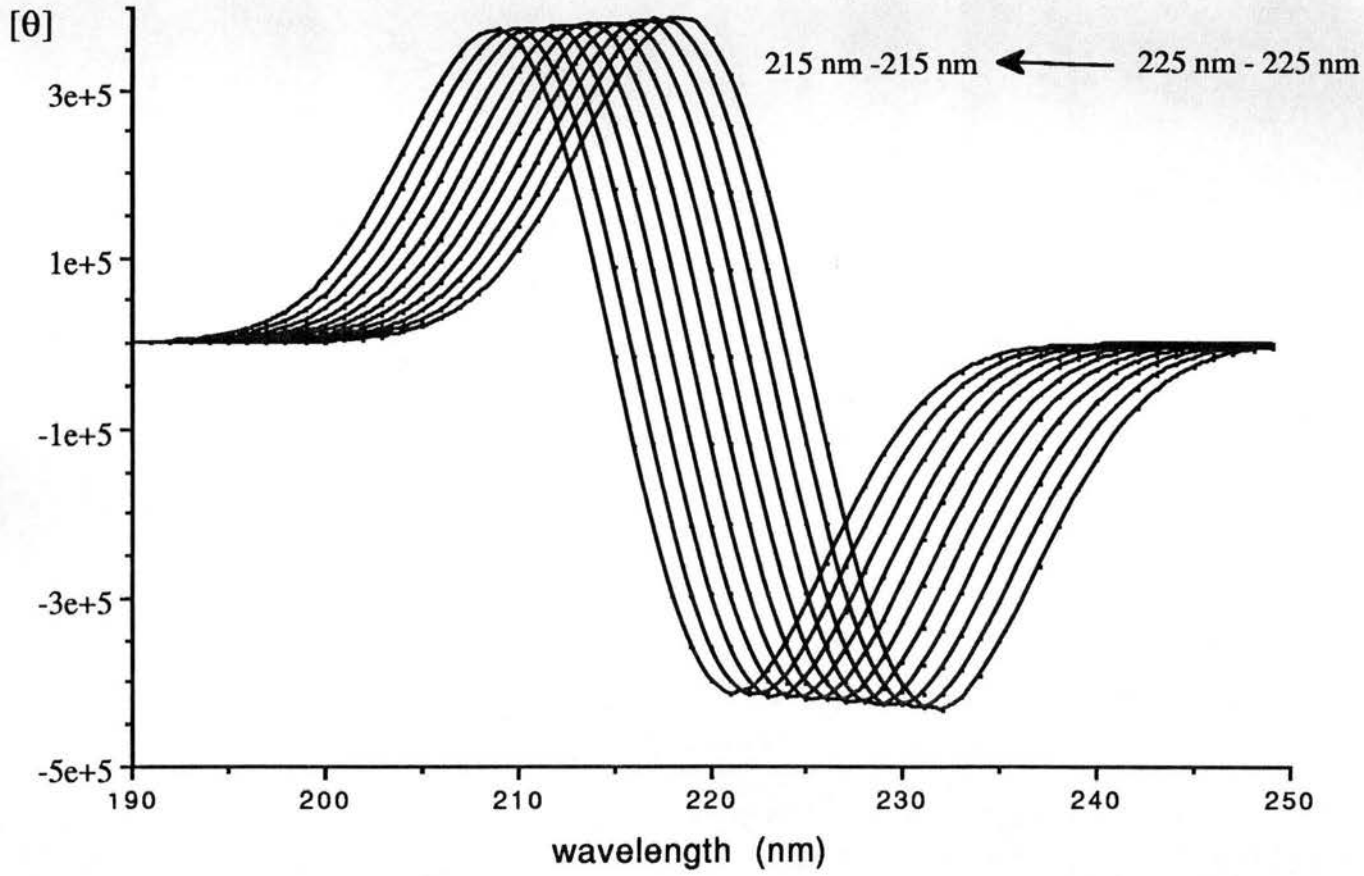


Figure II-6. Calculated CD of Trp pairs of identical Trp with B_b transitions energies from 5.51 eV (225 nm) to 5.77 eV (215 nm).

direction of the energy change is observed, but the amplitude of the couplet is essentially unchanged.

I estimated the effect of having Trp with analogous transitions distributed through a set of energies. Figure II-7 shows the changes in the exciton couplet CD for a system in which one Trp is buried (Trp B_b at 225 nm) and the exposure of the other gradually increases, corresponding to a change in the position of the B_b band from 225 to 215 nm. As the energy of the B_b transition increases, the couplet experiences a blue shift and decreases in amplitude. In fact, the more dissimilar the interacting transition energies are, the lower the optical activity of the composite transitions (Fig.II-8). This result is intuitively clear and theoretically obvious, but it is interesting to see the effect of varying exposure of chromophores on the prediction of CD for a particular protein and this is performed for lysozyme.

Hen egg-white lysozyme (HEWL) is a relatively small (129 a.a.) and well-studied protein. It contains six Trp residues (Trp28, Trp62, Trp63, Trp108, Trp111 and Trp123), three of which lie in the active site cleft and are involved in substrate binding (Goux & Hooker, 1980a). First I calculated the optical activity for mixing of degenerate B_b (225 nm) transitions in all six Trp for the tetragonal lysozyme crystal structure (2lyz, Diamond, 1974). Most of the lysozyme Trp residues are in relatively close proximity to one another (6 - 8 Å), and multiple chromophore interactions are important in the resulting Trp CD (Chapter II.4.b). Visual analysis of the HEWL structure reveals that Trp28 and Trp111 are buried, Trp123 lies close to the surface, and Trp62, Trp63 and Trp108 are partially exposed in the active site cleft. Based on this analysis, I assigned

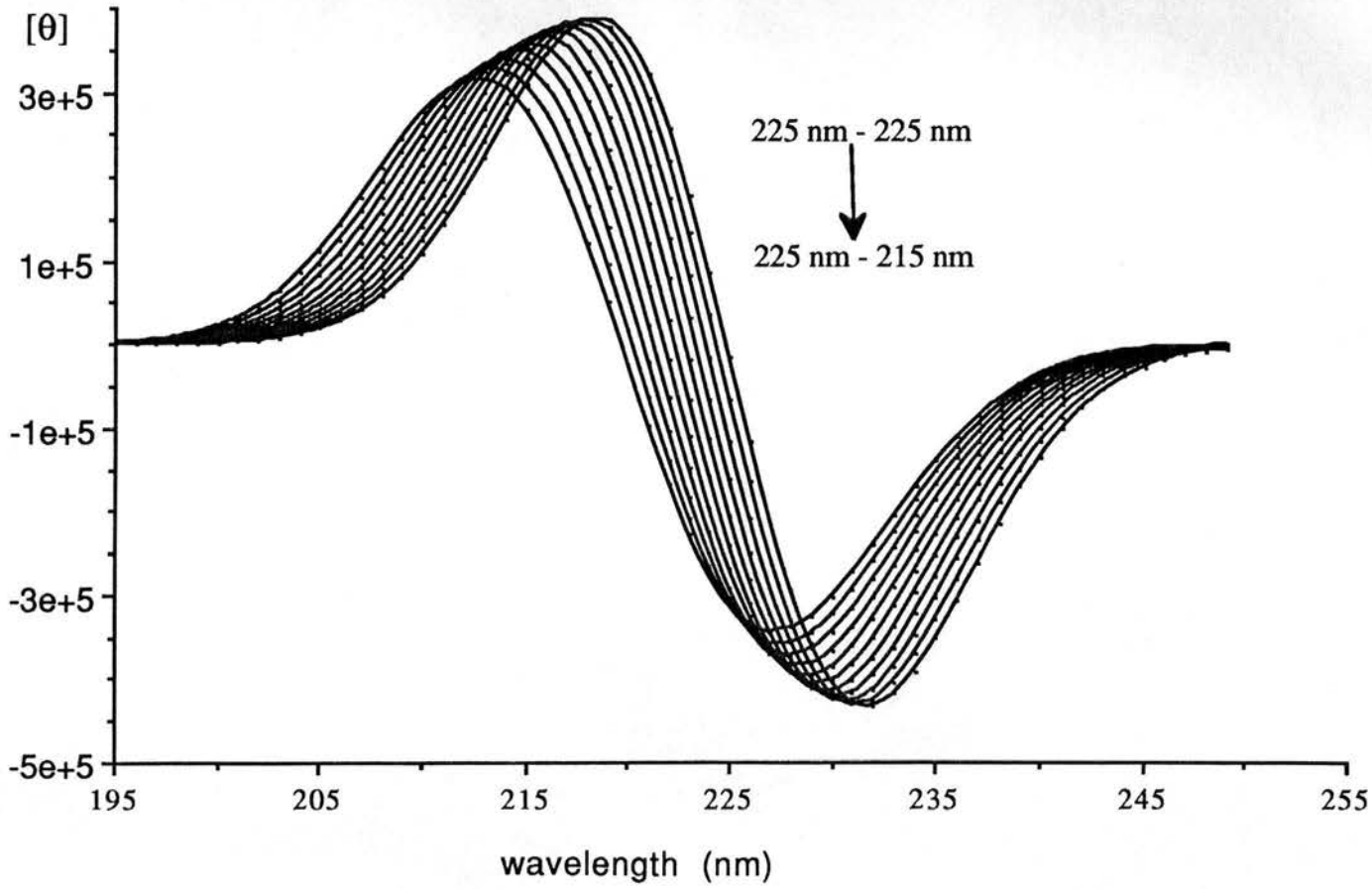


Figure II-7. Calculated CD of Trp pairs between a Trp, for which the B_b transition is at 5.51 eV (225 nm) and a Trp for which the B_b transition is varied from 5.51 eV to 5.77 eV (215 nm).

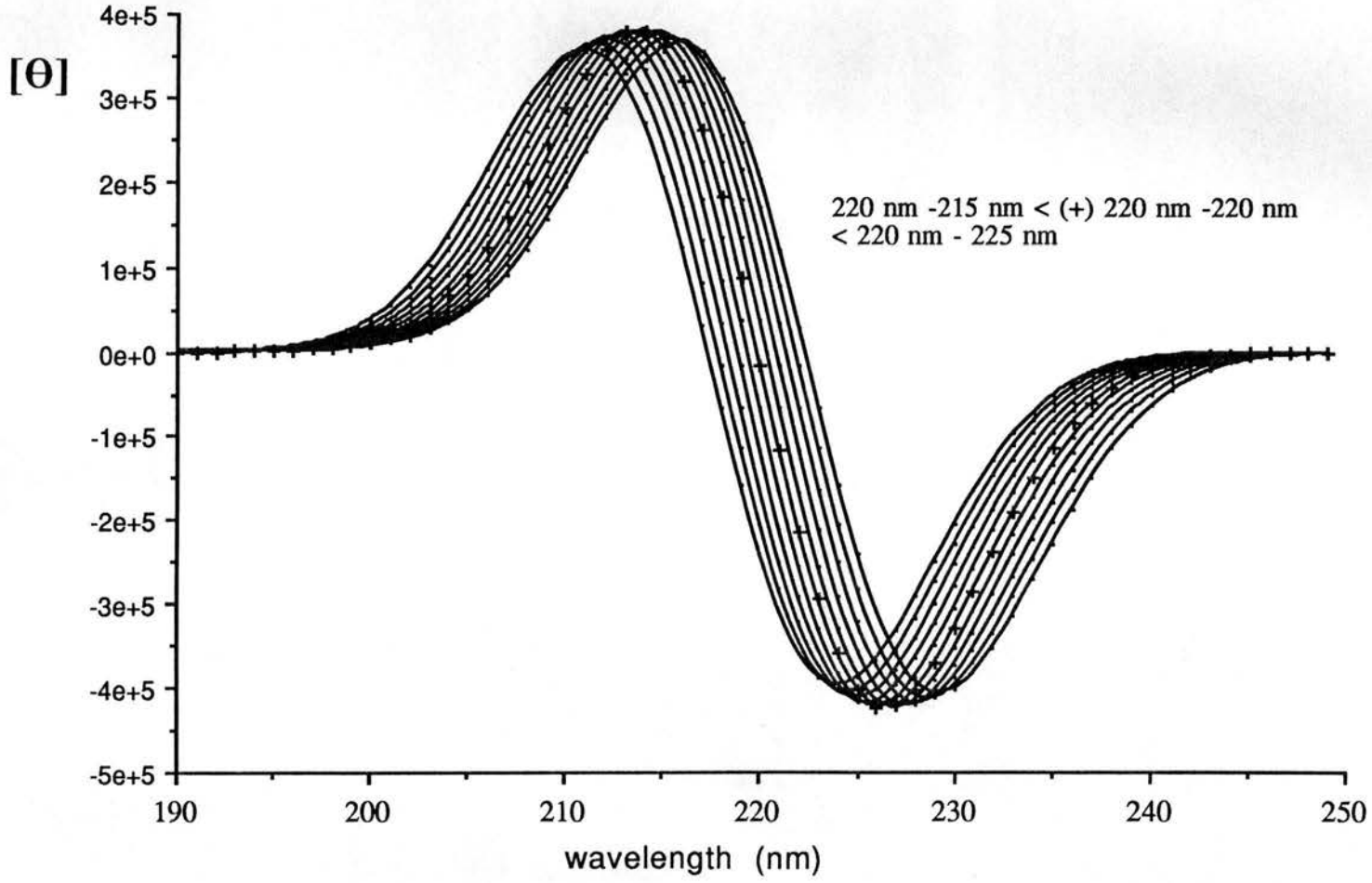


Figure II-8. Calculated CD of Trp pairs between a Trp for which the B_b transition is at 5.63 eV (220 nm) and a Trp for which the B_b transition varied from 5.51 eV (225 nm) to 5.77 eV (215 nm)

the B_b indole transitions of Trp123 at 219 nm, Trp62 at 220 nm, Trp63 at 221 nm, Trp108 at 223 nm, Trp111 at 224 nm, and Trp28 at 225 nm. Based on these estimates, another prediction of the Trp CD was made (Fig.II-9). An analysis of local molecular structure and the resulting dispersion of the transition energies is required to achieve more accurate predictions. However, our simple analysis gives the right idea of the effect of varying Trp exposure on CD. The whole calculated spectrum experiences a 3 nm blue shift. This is understandable since some Trp are considered partially exposed and the energies of electronic transitions increase in an environment of lower polarizability. Inside the protein hydrophobic core the transition energies are lowered. In the core the polarizability of the environment is higher and the excited states of the electronic transitions, which are more polar, are favored and thus their energy level is lowered. The amplitude of the positive couplet decreases by about 35% (Fig.II-9) and the short-wavelength positive band is almost abolished.

I found that most Trp-containing proteins give a CD couplet in the B_b region (Chapter II.4.a). The predicted Trp B_b CD spectrum of HEWL is unusual in exhibiting this third short-wavelength band. Interestingly, this band in HEWL is predicted to vanish upon binding to monoclonal antibodies as discussed in Chapter II.4.b. The existence of the third maximum means that the spectrum is composed of two couplets of opposite sign, one shifted relative to the other. It is not clear what geometrical properties of the group of Trps are required to give a triplet rather than a couplet. Thus I decided to distinguish contributions of individual Trp to the short-wavelength band in a homogenous hydrophobic medium (all B_b transitions at 225 nm), and while exposure to solvent is

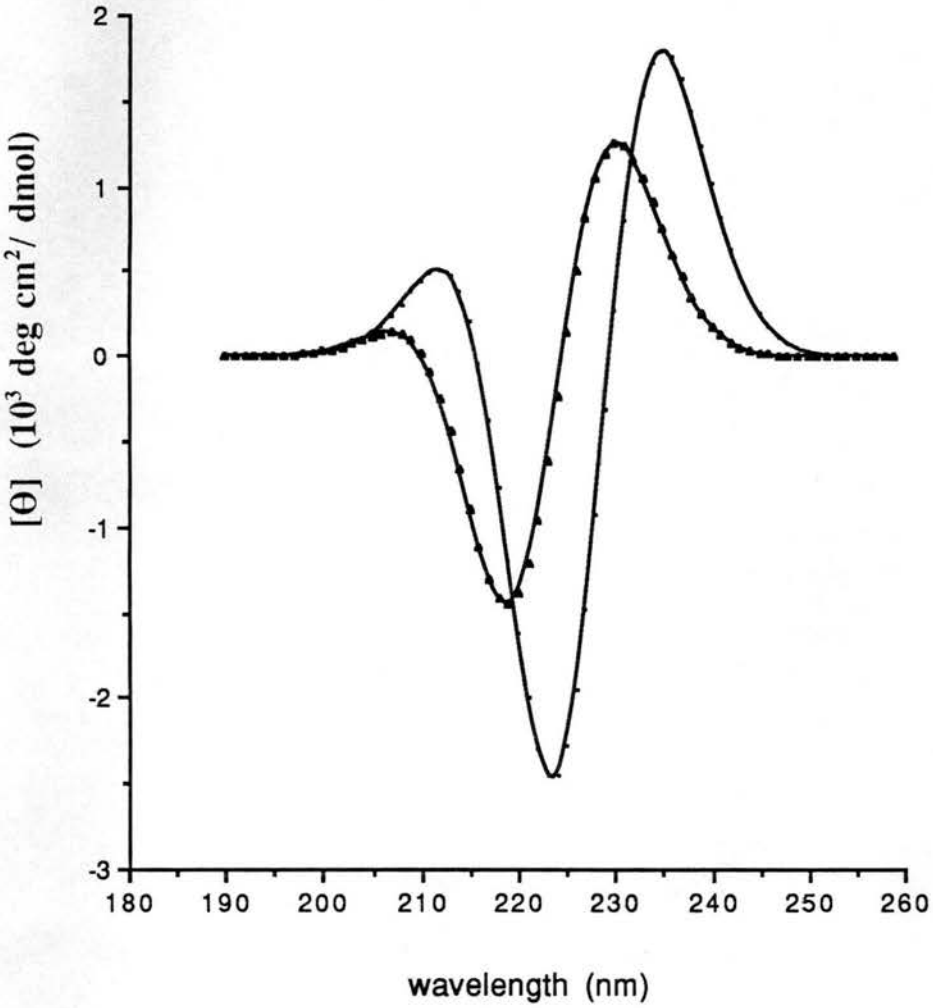


Figure II-9. Calculated B_b Trp CD of HEWL accounting for variable exposure of Trp to the solvent (▲). Comparison with predicted CD, when B_b transitions in all Trp are at 5.51 eV (225 nm) (-).

considered as discussed. I estimated the "partial" CD for an individual Trp as the difference between the total Trp CD and the Trp CD of the five other Trp for both degenerate and nondegenerate cases. I found that the short-wavelength band decreases in the nondegenerate case due to the shift of the partial contributions of Trp111 and Trp123 (Fig.II-10). The maxima of B_b transitions in Trp111 and Trp123 shifted to 220 nm and 217 nm, respectively. The quantitative estimates obtained should be taken with caution, since the contributions of individual Trps to the CD of the group are not additive. However, we can conclude that Trp111 and Trp123 contribute positive CD around 212 nm. Figure II-11 shows the difference spectrum of HEWL considering and not considering Trp exposure, and the differences in partial CD of Trp111 and Trp123 within those two models. The other HEWL Trps have the opposite effect on the CD in that region.

(2) Effect of nearby charge on the CD of Trp pair.

It is reasonable to assume that nearby charges would affect the optical activity of Trp pairs in a protein by altering the direction and/or magnitude of the electric dipole moments in the ground and excited states, as well as the transition moments, of the chromophores. We decided to start by only considering the contribution of the charge to the ground state charge distribution, which will modify the one-electron mixing of the transitions. In these calculations, I considered all six low-energy transitions of indole as specified in Methods (Chapter II.2).

I took a pair of Trp in P45 conformation separated by 5 Å and calculated the CD of that system with an additional uncompensated charge or an ion pair placed nearby.

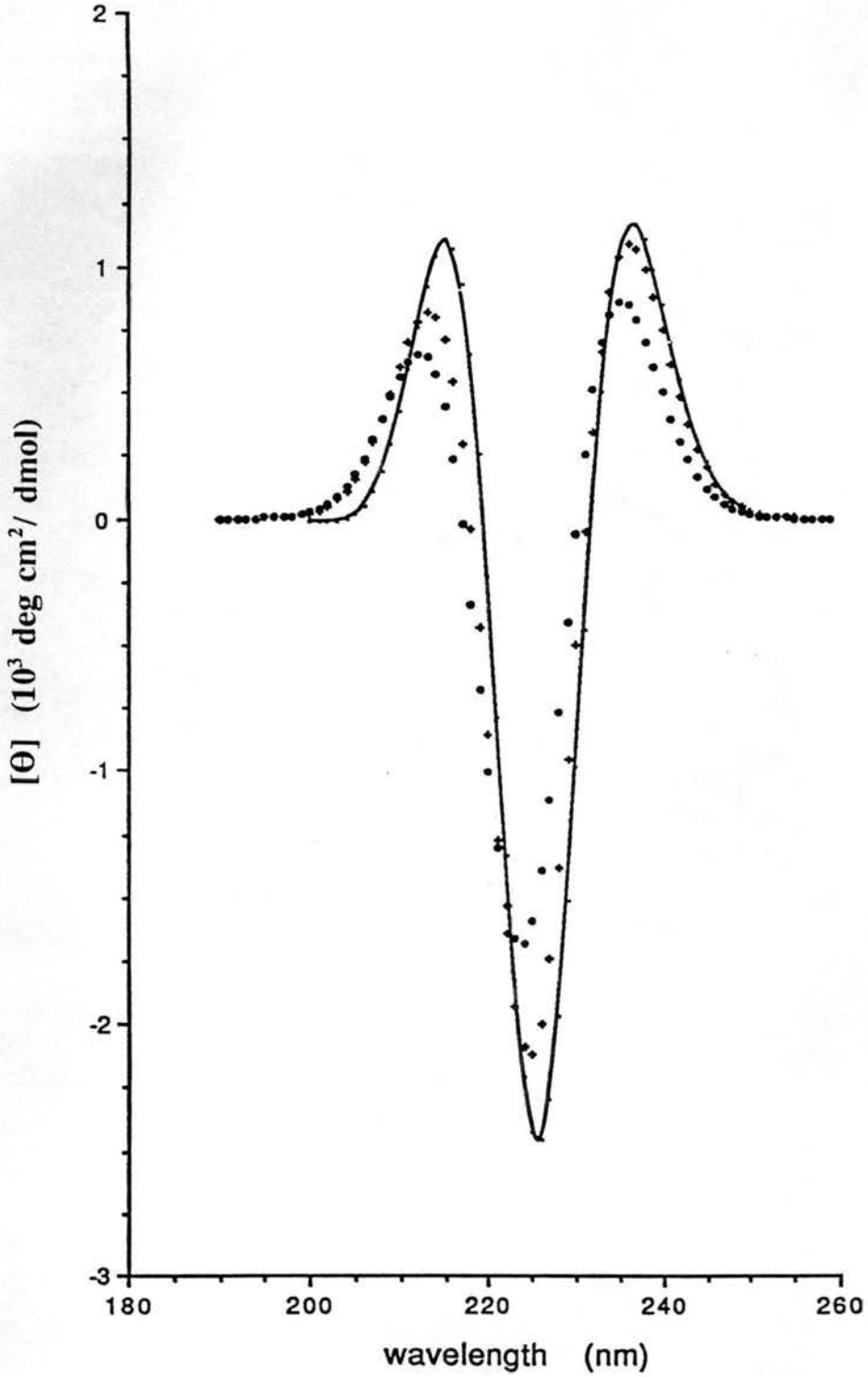


Figure II-11. Difference spectrum of HEWL accounting for variable exposure of Trp to the solvent (-), and difference spectra of partial contributions of Trp111 (o) and Trp123 (+) between the degenerate and nondegenerate cases (see text).

One of the Trp lies in the xy plane with its long axis along (x). The centers of Trp are defined as before, in the middle of the CE2-CD2 bond. The location of charges were chosen as follows (Fig.II-12): in the yz plane between the indole rings at 5 Å from their centers (0; 4.33; 2.5), defined as position P1; above both rings, shifted by 5 Å along the Z axis from P1, defined as position P2 (0; 4.33; 7.5); reflected through the xz plane from position P1, defined as P3 (0; -4.33; 2.5); and shifted along the z axis from P3, defined as P4 (0; -4.33; 7.5). The magnitude of the charges were taken to be -1, -0.5, +0.5 and +1 (in units of e = 4.803207×10^{-10}esu). Ion pairs were constructed by placing an additional charge of opposite sign at 4 Å from the first one along the y axis : (0; 8.33; 2.5) at P1, (0; 8.33; 7.5) at P2, (0; -8.33; 2.5) at P3, and (0; -8.33; 7.5) at P4.

The CD calculations **in the far uv** led to the following results :

1) the effect of a half-charge is lower in magnitude, but has the same qualitative effect on CD as a full charge, the results for which are shown in Figures II-13, II-14 and II-15;

2) the effect of the counterion in the ion pair was less than 15 % for all geometries considered and in most cases negligible, as the effect of the external charge on the dipole moment of the ring decreases rapidly with distance;

3) the effect of placing single charges at the positions considered resulted in a set of far-uv CD spectra, the relations of which can be summarized as follows (Fig.II-13): $P1(1) \approx P3(1) \approx P4(-1) \sim P2(-1)$ result in similar spectra that we will call type A; $P1(-1) \approx P3(-1) \approx P4(1) \approx P2(+1)$ result in another type of spectrum that we call type B. The effect of placing the charge at position P2 results in some qualitatively different features

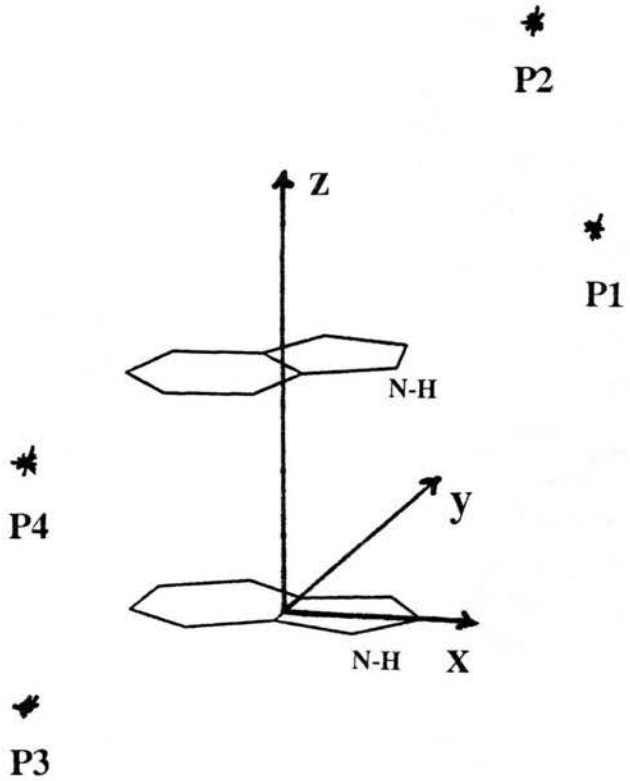


Figure II-12. Schematic location of the charges near the Trp pair (see text).

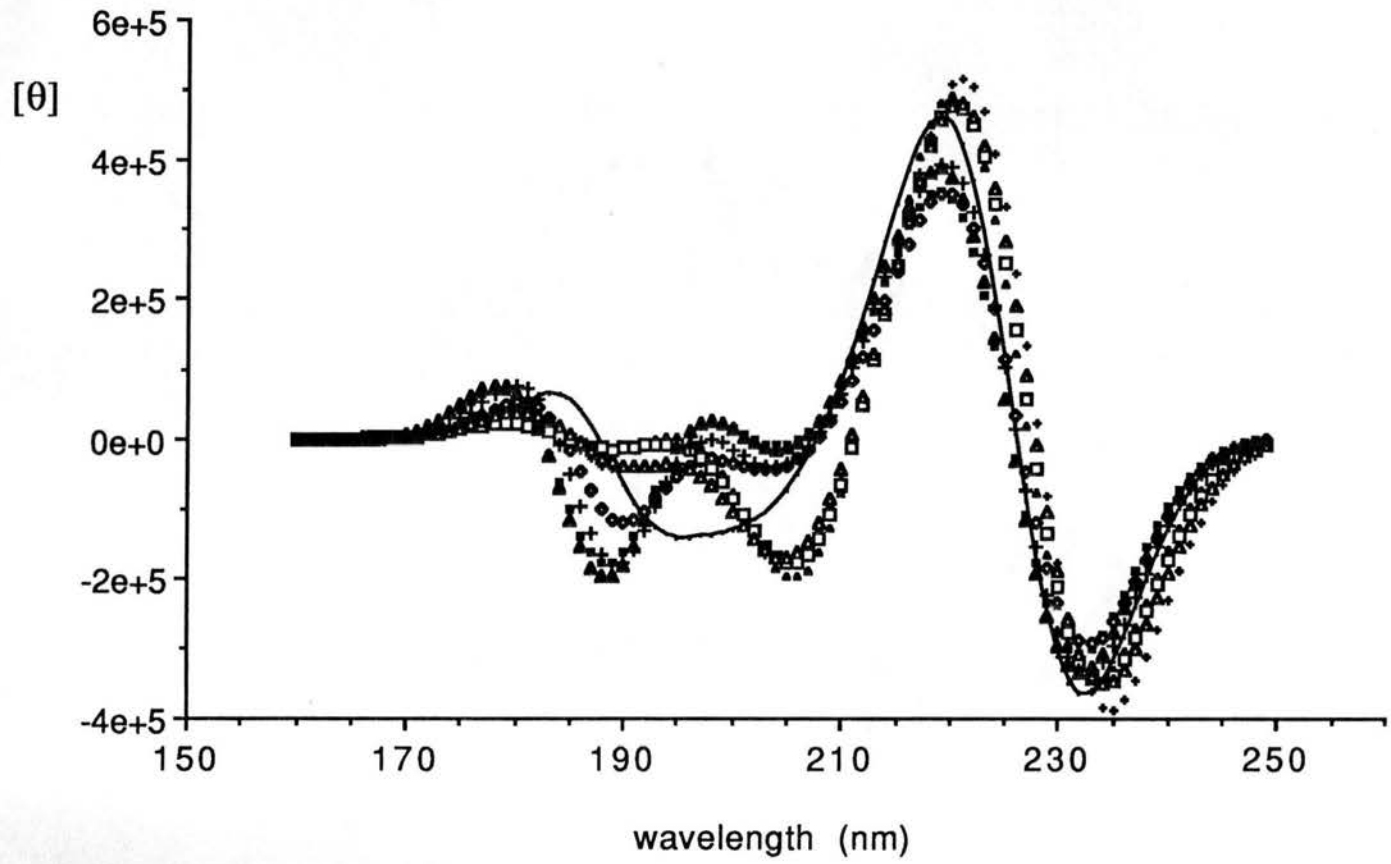


Figure II-13. Effect of the nearby charges on the calculated far-uv CD of a Trp pair (position . charge) : no charges (-); P1.1 (■); P1.-1 (▴); P2.1(+); P2.-1 (◇); P3.1 (▲); P3.-1 (□); P4.1 (△); P4.-1 (+).

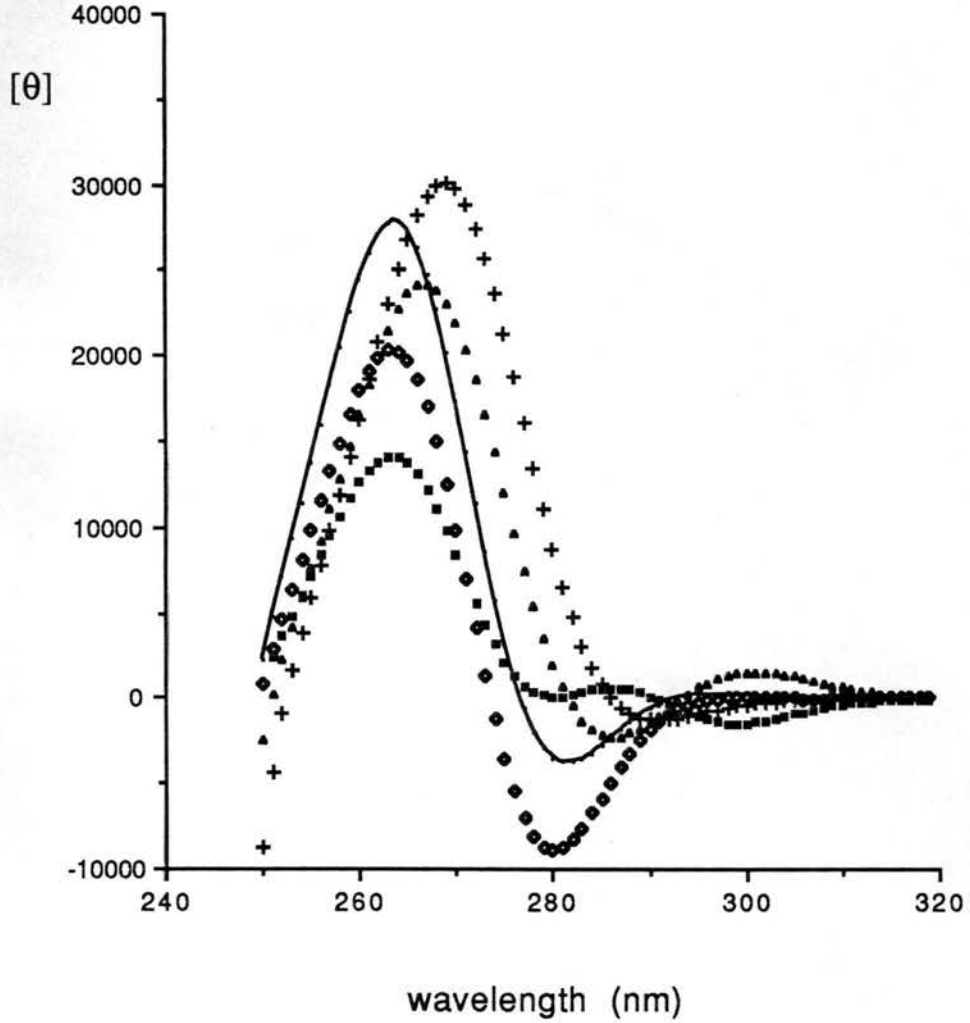


Figure II-14. Effect of the nearby charges on the calculated near-uv CD of a Trp pair (position . charge): no charges (-); P1.1 (■); P1.-1 (▲); P2.1 (+); P2.-1 (◇).

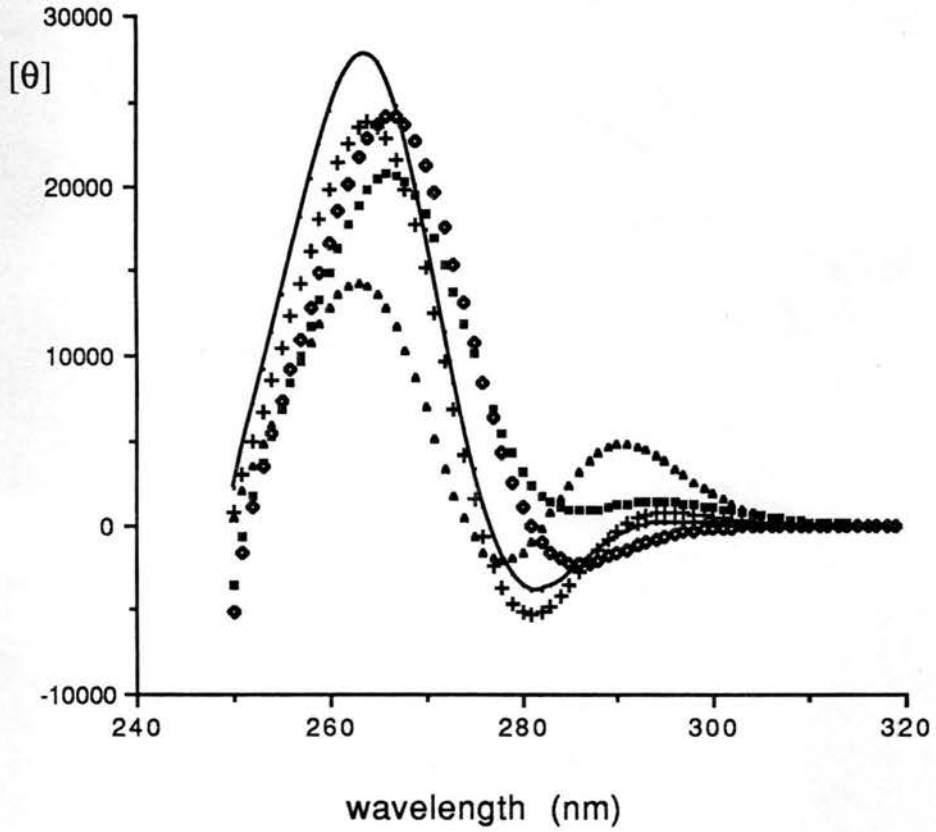


Figure II-14. Effect of the nearby charges on the calculated near-uv CD of a Trp pair (position . charge): no charges (-); P3.1 (▲); P3.-1 (■); P4.1 (◇); P4.-1 (+).

of the spectra. Placing the charge at any of the other geometries considered, results in a spectrum which falls in either A or B type depending on the sign of the charge;

4) placing the charge at any location results in a significant deviation from the spectra of the isolated Trp pair. In all cases: a) both extrema of the B_b couplet are red shifted by 2 nm; b) the positive band around 185 nm is either abolished or blue shifted by 3- 4 nm; c) the broad negative band around 195-205 nm in the CD of an isolated pair is resolved into two bands;

5) the cases resulting in type A spectra: a) show larger positive amplitudes around 180 nm; b) the short-wavelength part of the 195-205 nm band is enhanced and blue-shifted, and its longer-wavelength part is decreased or abolished; c) the magnitude of the B_b band is decreased;

6) the type B spectra: a) show weak CD around 180 nm; b) the long-wavelength part of the 195 - 205 nm band is enhanced and red-shifted, and the short-wavelength part of it is abolished; c)the magnitude of the B_b band is increased;

7) the spectra corresponding to the geometry P2: a) show extreme values at the B_b band: the lowest intensity with P2(-1) and the highest with P2(1); b) the geometry P2(-1) also shows unusually low amplitude at 190 nm.

The effect of nearby charges can be also seen in **the near uv** (Fig.II-14; II-15). As can be seen, in all cases when charges were added, the short-wavelength extremum shifts to the red, but the position of the L_a maximum varies. The similarities in the spectra obtained in the near uv correlate with the division of cases into A and B types in the far uv, although the relative deviation between the spectra in the near uv are larger

than in the far uv. Thus, in fact, it might be promising to follow the effect of a nearby charge in some experimental system in the near-uv CD. It is easier to analyze the spectra predicted in the near uv if we concentrate on the CD intensities in two intervals: at 260 - 270 nm, where we assume that most of the contribution is from the L_a band at 269 nm, and at 280-290 nm where the contribution of the L_b transition at 282 nm is dominant. The P2(1) geometry shows maximal amplitude in the L_a region, and the P1(1) $\approx$ P3(1) shows the lowest, which is less than 50 % of the highest value. In the L_b region, all geometries other then P3(-1) exhibit negative CD. The largest negative band in this region is obtained for geometry P2(-1). Thus, an uncompensated charge in a globular protein near a Trp pair might influence the transitions in the near uv in such a manner that its location or at least the quadrant of its location in the coordinate system associated with one of the Trp could be determined.

In these calculations the charges placed in the plane perpendicular to the indole ring and containing the CE2-CD2 bond were shown to affect mostly the amplitude and position of the B_a transition. The B_a transition dipole moment is closer to parallel to that plane than is that of the B_b transition (Table I-1). When the charges were placed in the xz plane (data not shown), which is close to parallel to the direction of the B_b transition moment, their presence affected mostly the B_b region of the spectrum.

(3) Observable effect of mixing of indole transitions.

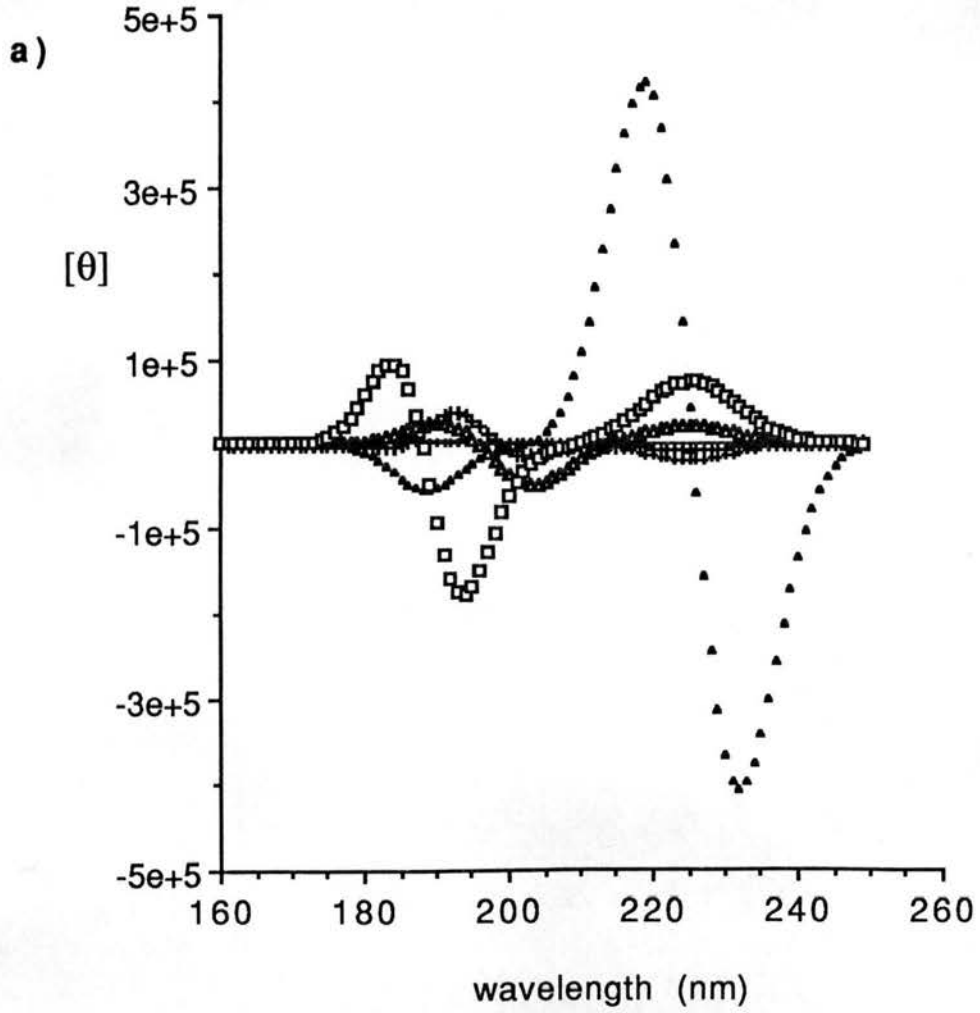
Suppose we observe the near-uv CD of a particular protein and, under certain conditions, the optical activity has vanished or some other distinctive changes have occurred. For an experimentalist this would indicate that something has happened around

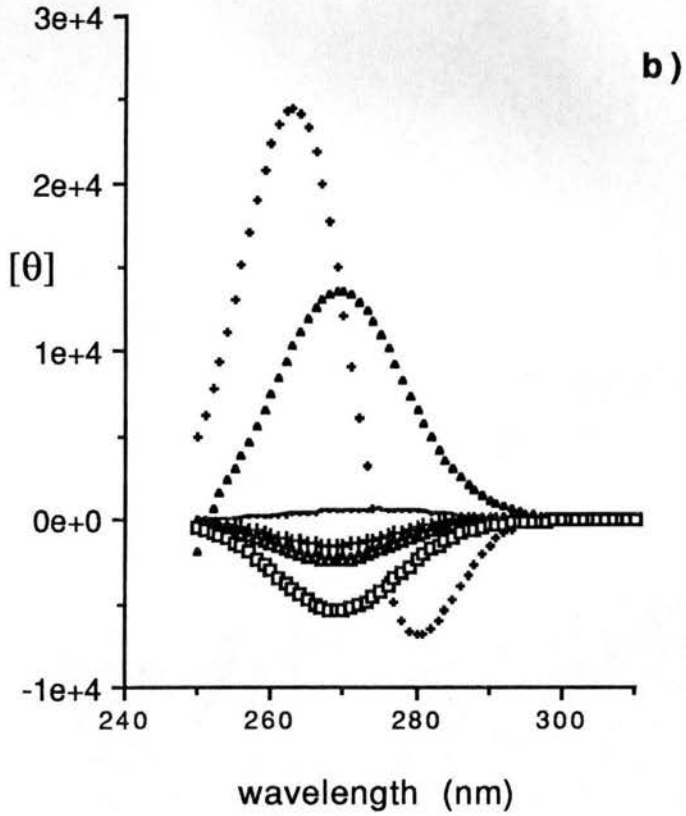
one or more aromatic chromophores. Aromatic chromophores, Trp in particular, exhibit CD both in the near and far uv (Table I-1 and I-2). Should we make any corrections in the analysis of the far-uv spectra of that protein under the same conditions ? The changes in aromatic CD might have occurred in the far uv as well, but they are difficult to distinguish from the overwhelming peptide CD in that region. Can we quantitatively correlate aromatic CD changes in the near and far uv ? In the following section I explore this question.

I considered the interaction between six indole transitions, two of which lie in the near uv, the L_b at 282 nm and the L_a at 269 nm, and the others in the far uv : B_b at 225 nm, B_a at 200.3 nm, V at 193.4 nm and VI at 189.5 nm. Other than by the energy of excitation, the transitions differ by the redistribution of charge upon excitation. The latter is characterized by electric dipole and magnetic dipole transition moments (Table I-2), which are accounted for by transition monopoles (Table I-3). I performed two sets of theoretical experiments to reveal the effect of some of the transitions on the intensity of the others.

In the first experiment, I calculated the CD of a pair of Trp in P45 geometry, including all six transitions for one of the indoles, denoted as W1, and one transition at a time for the other, denoted as W2. The calculated spectra in the far (Fig.II-16a) and near uv (Fig.II-16b) are shown. The first apparent result is not surprising: each transition from W2 exhibits maximal CD within the wavelength interval where it is mixing with the identical transition from W1. This interaction results in a characteristic couplet band.

Figure II-16. Contribution of individual transitions to the CD of indole in the far uv (a) and in the near uv (b) considering all 6 transitions of W1 and one of the transitions of W2, namely: L_b (-); L_a (+); B_b (▲); B_a (△); V (+); VI (□).


a)
5e+5
3e+5
1e+5
-1e+5
-3e+5
-5e+5
[θ]
160
180
200
220
240
260
wavelength (nm)


b)
3e+4
2e+4
1e+4
0e+0
-1e+4
[θ]
240
260
280
300
wavelength (nm)

As we examine the effect of mixing of various individual W2 transitions with W1, we see the following:

L_b - the magnitude of the L_b transition throughout the region is too small to give any significant effect.

L_a - mixing of the W2 L_a transition results in a strong asymmetric couplet centered at its wavelength. The asymmetry results from mixing with the larger magnitude B_b band. The B_b (5.51 eV, 225 nm) transition is relatively distant from the L_a (4.61 eV, 269 nm) on the energy scale.

B_b - as can be seen from Figure II-16b, introduction of any W2 far-uv transition results in optical activity in the near uv, the largest effect being from the B_b transition. In fact the B_b transition contributes as much to the optical activity in the L_a region as the L_a transition itself. The band produced in the near uv is asymmetric, which indicates mixing with both the L_a and L_b transitions. B_b also contributes to the optical activity at 190 nm, through mixing with the B_a, V and VI transitions.

B_a - contributes to a low intensity band in the B_b region and to the near uv as well (Fig. II-16b). The sign of the contribution in the near uv is opposite to that of B_b and about 12% of its amplitude (Fig. II-16b).

V - contributes to the near uv, in the B_b region and at transition VI. The couplet at transition V is asymmetric due to mixing with the B_b and VI transitions. The sign of the contribution in the near uv is opposite to that of B_b and reaches about 18 % of its amplitude (Fig. II-16b).

VI - is second in effectiveness (after the B_b transition) among the far-uv transitions in contributing to the near uv, despite being the most distant on the energy scale. The sign of the contribution in the near uv is opposite to that of B_b and reaches about 40 % of the amplitude of the latter (Fig. II-16b). The sizable effect of this transition is due to its large electric dipole transition moment (4 D). The couplet at the wavelength of transition VI is skewed towards longer wavelengths, primarily due to mixing with the B_b transition.

Another version of this experiment is to calculate the optical activity of two indoles with an identical but limited number of transitions, and record the changes upon adding each additional transition to the system (Fig. II-17). I started with calculation of the near-uv CD due to the L_a transition and L_b transition separately and together (Fig. II-17b), and then added the other transitions one by one:

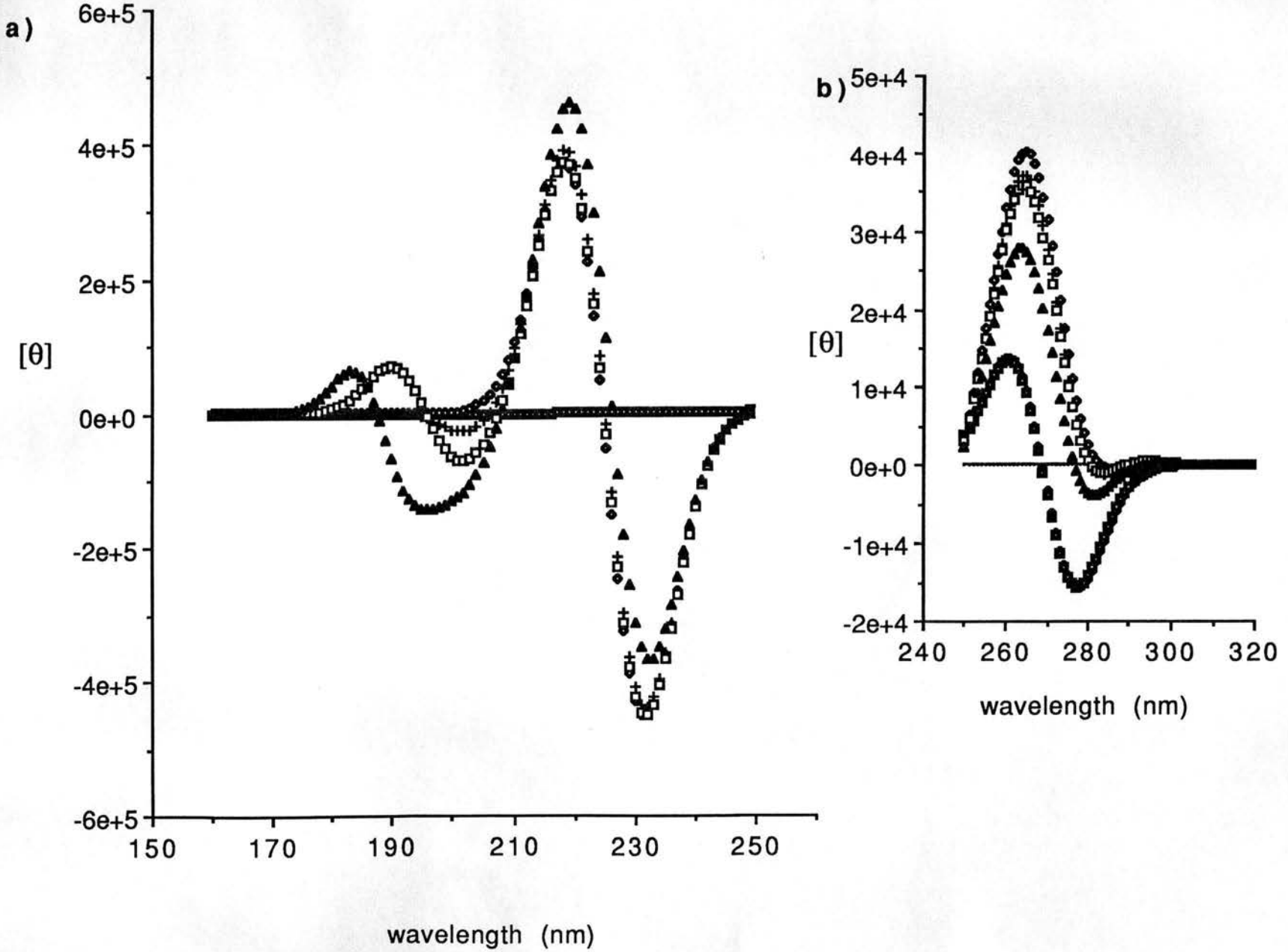
L_b - The transition is of low intensity;

L_a - The transition is of sufficient intensity to be observed in an average protein. A characteristic couplet band appears due to the interaction of L_a transitions in W1 and W2. The band does not extend below 240 nm.

2 (L_a + L_b) - Almost no difference from L_a alone.

3 (...+ B_b) - Characteristic couplet at the wavelength of the B_b transition. Significant effect on the near uv. A broad positive band of high intensity dominates the couplet produced by the interaction of L_a of W1 and L_a of W2.

Figure II-17. Effect of the higher energy transitions in indole on the intensity of the lower energy transitions in the far uv (a) and in the near uv (b). Calculations are performed considering the same number and type of transitions in both Trp, namely : L_b (-); L_a (+); L_b+L_a (•); $L_b+L_a+B_b$ (▲); $L_b+L_a+B_b+B_a$ (△); $L_b+L_a+B_b+B_a+V$ (+); $L_b+L_a+B_b+B_a+V+VI$ (□).


a)
6e+5
4e+5
2e+5
0e+0
-2e+5
-4e+5
-6e+5
[θ]
150
170
190
210
230
250
wavelength (nm)
b)
5e+4
4e+4
3e+4
2e+4
1e+4
0e+0
-1e+4
-2e+4
[θ]
240
260
280
300
320
wavelength (nm)

4 (...+ B_a) - Almost no effect on B_b. A single negative band around 200 nm appears. Positive band in the near uv slightly decreases. Marginal negative band at L_b appears.

5(...+ V) - No effect in B_b region. Negative couplet appears at location of transition V ($\sim$ 195 nm), overwhelming the negative CD at B_a.

6(...+ VI) - The positive lobe of the transition V couplet shifts to the location of transition VI, significantly broadening the negative band, as if the contribution of V is added to the others; affects the magnitude of the B_b couplet ($\sim$ 12 %); decreases the positive contribution in the near uv; and increases the negative band at L_b.

Overall, from these two sets of predicted spectra, one clearly sees the large effect of the B_b transition on the near uv, resulting in a significant change in magnitude and shape of the spectrum of the near-uv transitions. Next in order of importance is transition VI, located at the highest energy limit considered in our calculations. The effect of transition VI is opposite in sign to that of B_b. Although we cannot ignore the effect of the distance, on the energy scale, between various transitions on their mixing, it appears that the magnitude of the dipole transition moment is the most important parameter. This conclusion is applicable both to the mixing of various transitions and to the mixing between the same transitions in different Trp placed at various energies to account for solvent exposure (Chapter II.3.a.(1)).

Thus in fact, since transitions in the far uv appear through mixing with the near-uv, a correlation between the optical activity in the near and far uv does exist as I have shown for a pair of Trp chromophores in P45 geometry. Other conformations must be

analyzed to draw quantitative conclusions, which is performed in the next section. In fact, one reason for performing this work was my guess that mixing between the far- and near-uv transitions should depend on the geometry of Trp in the system.

(4) Relations Between Conformation and CD of a Trp Pair.

The optical activity of a Trp pair due to exciton coupling depends on the magnitudes and relative orientations of the electric dipole transition moments of the two Trp, μ_{1a} and μ_{2a}, and the vector defining the center-to-center distance ($\mathbf{R}_{12}$). The absolute values of the optical activity at the maxima of an exciton couplet band for a particular transition are proportional to the parameter defined as the couplet strength (Eq.II-1; Bayley, 1973), and the shape of each band corresponding to an individual transition was assumed to be Gaussian, with the half bandwidth calculated from Mason's equation (Eq. II-2). For the couplet strength calculated from Equation II-1

$$S \equiv \Delta\epsilon_+ - \Delta\epsilon_- \propto (\Delta\lambda)R \,, \qquad \text{(II-3)}$$

where R is the rotational strength of the transition and $\Delta\lambda$ the splitting of the exciton components.

In the framework of our model, for any conformation of a pair of Trp, the optical activity can be calculated if the positions and orientations of the two aromatic rings are known. Practically, if calculations are made for proteins, the atomic coordinates for the Trp side-chain atoms are taken from Brookhaven Protein Data Bank. In case a definite model conformation of Trp pairs is to be considered, atomic coordinates for the indole can be used for one of the Trp, while the coordinates of the other indole are obtained by a transformation of the coordinate system. The question I want to discuss in this section

is whether the conformation or conformational changes of the chromophores could be elucidated from the known CD.

We can roughly approximate the dependence of CD on the geometry of the two Trp using the dipole-dipole approximation. In this approximation, the rotational strength of the exciton couplet is:

$$R = (\pi/2\lambda_{0a})\ \mathbf{R}_{12}\ [\ \mu_{1a} \times \mu_{2a}\ ], \qquad \text{(II- 4)}$$

where the subscript "a" specifies the excited state. It is useful to analyze the case when the two Trp are stacked in parallel planes with the center-to-center vector ($\mathbf{R}_{12}$) perpendicular to both planes (Fig.II-18a,c). If we consider two identical tryptophans, so that $|\mu_{1a}| = |\mu_{2a}|$,

$$R = (\pi/2\lambda_{0a})\ |\mathbf{R}_{12}|\ |\mu_{1a}|^2 \sin\theta \qquad \text{(II-5)}$$

where θ is the angle between the vectors μ_{1a} and μ_{2a}, and in our case this angle is in the plane parallel to xy. On the other hand, the splitting of the exciton bands, $\Delta\lambda$, is proportional to the energy of interaction of the dipole transition moments, which in the dipole-dipole approximation is

$$\Delta\lambda \propto V = (1/R^3)\ \{\ \mu_{1a} \bullet \mu_{2a} - 3(\mu_{1a} \bullet \mathbf{R}_{12})(\mu_{2a} \bullet R_{12})/R^2\}. \qquad \text{(II-6)}$$

Considering that for the Trp stacked in parallel planes $\mu_{1a} \bullet \mathbf{R}_{12} = \mu_{2a} \bullet \mathbf{R}_{12} = 0$ and assuming $|\mu_{1a}| = |\mu_{2a}|$, we obtain :

$$\Delta\lambda \propto (|\mu_{1a}|^2/R^3)\ \cos\theta \qquad \text{(II-7)}$$

$$\text{Thus } S \equiv \Delta\epsilon_+ - \Delta\epsilon_- \propto \Delta\lambda R \propto \sin\theta\cos\theta \propto \sin 2\theta \qquad \text{(II-8)}$$

Thus, the intensity of the optical activity for two Trp in a planar stacked geometry is a periodic function of θ or, in other words, rotation in the plane perpendicular to the Z

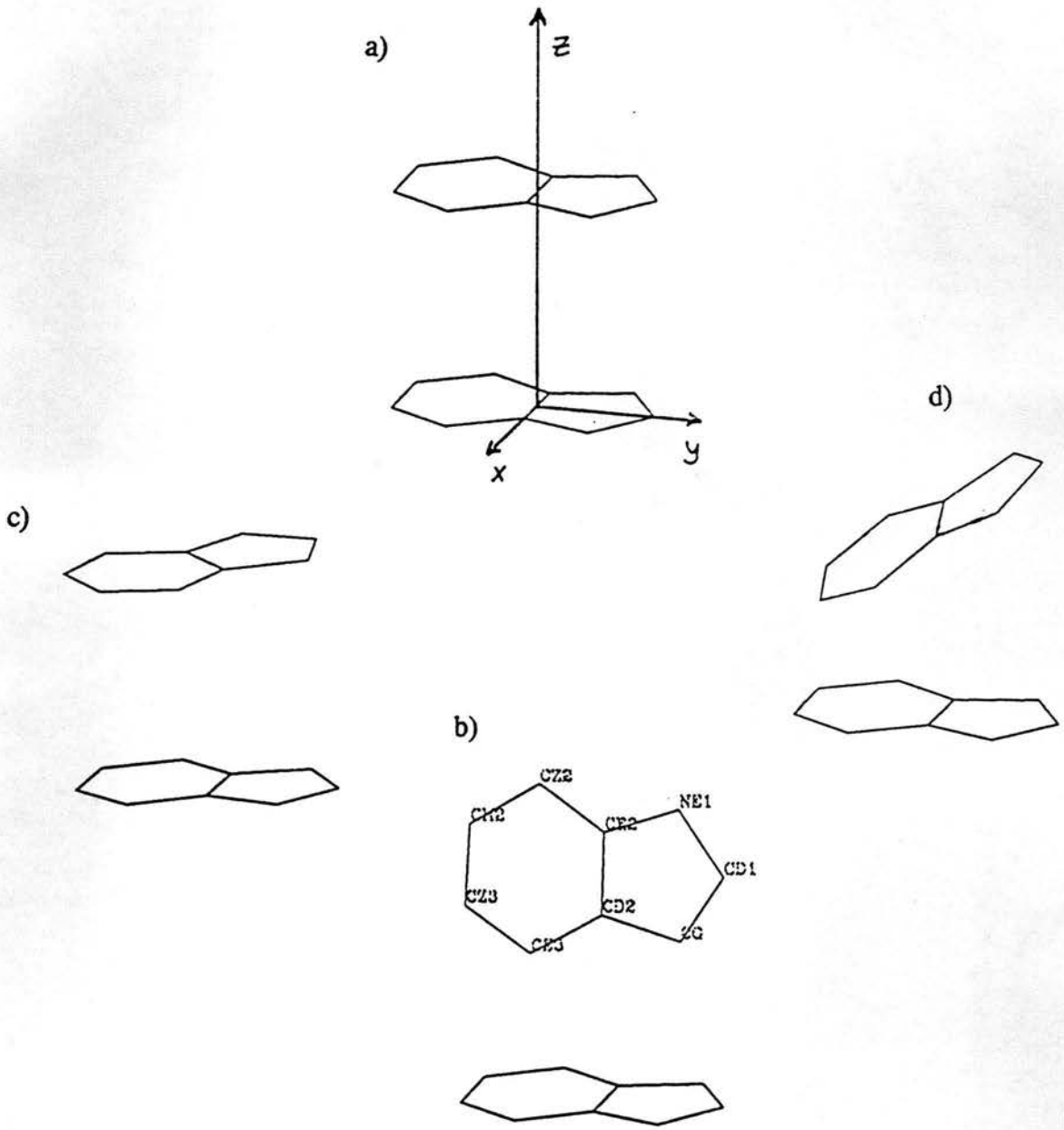


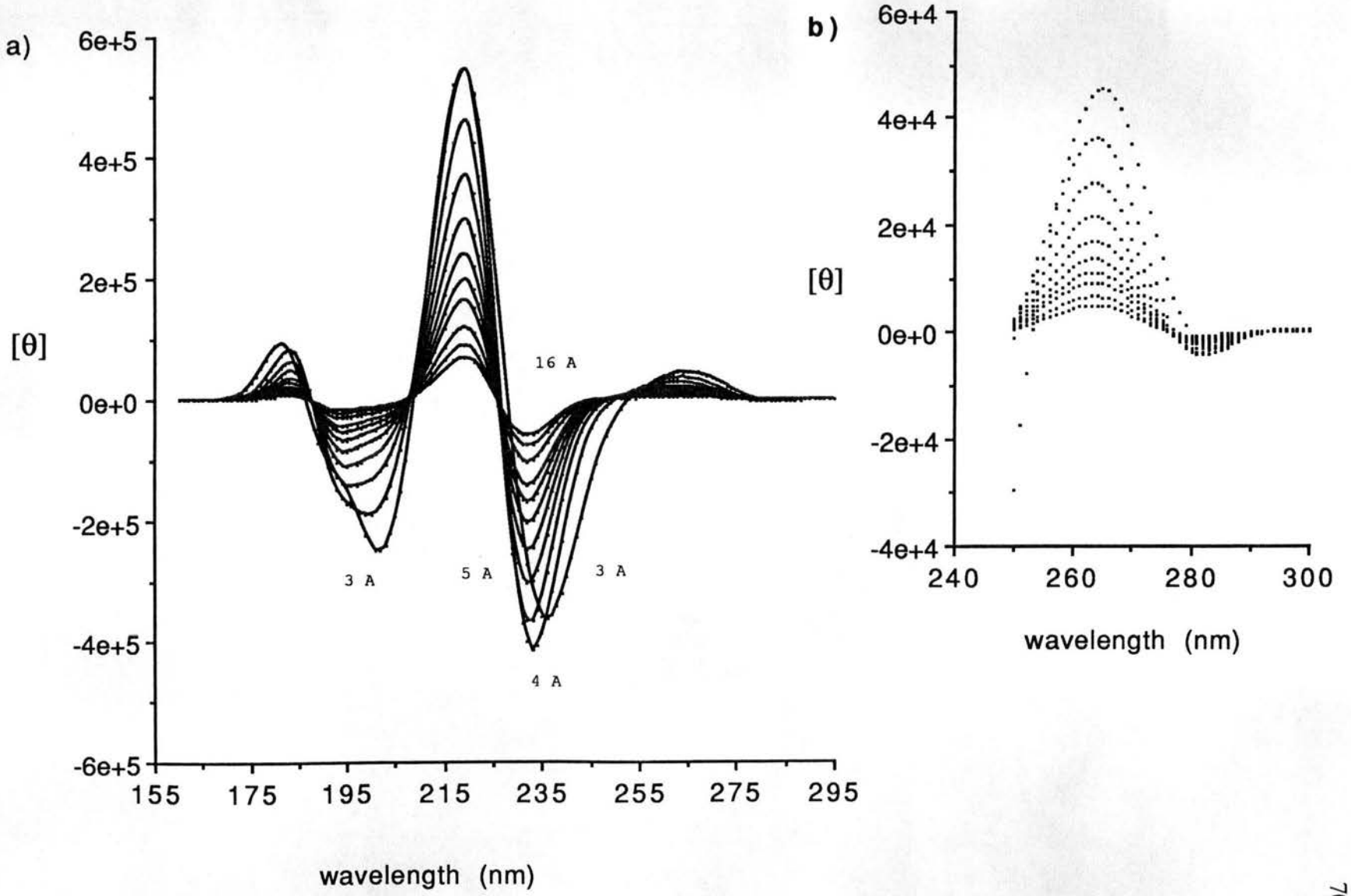
Figure II-18. Conformations of a Trp pair : parallel stacking (a); W2 rotated around y axis (b); W2 rotated around z axis for 45^0, defined as P45 conformation (c); W2 rotated around x axis (d).

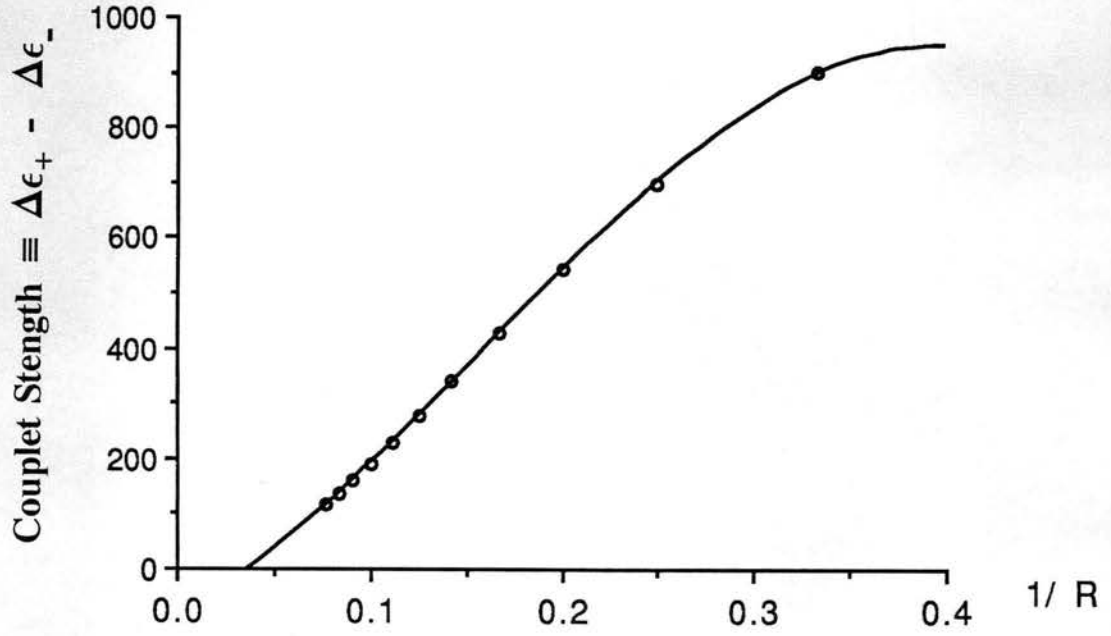
axis as in Figure II-18c. The periodicity of $\sin 2\theta$ is π, and the maximum values are reached at $\theta = 45° \pm k\pi$, where k is an integer. This geometry was defined earlier as P45 and used in previous examples.

The first step I took was to estimate the distance dependence of the CD intensity in the B_b region of the spectra at a given orientation. I considered the six lowest energy Trp transitions defined in Methods. The calculated CD spectra are shown (Fig.II-19) for the geometry P45 at various center-to center distances from 3 Å to 16 Å. These spectra give an idea of the positions of the composite bands and the dependence of their shape and intensity on the distance between the Trp. It can be seen that the intensity of the spectra at each wavelength decreases monotonically with an increase in distance, other than for the unrealistically short distance, 3Å, where crossovers with other spectra are observed in the negative bands around 200 nm and 235 nm. I obtained an analytical expression for the distance dependence of the CD intensity by performing a set of functional approximations shown in Figure II-20. As can be seen, polynomial approximations of the third and higher order provide adequate fitting accuracy, and thus could be used to estimate the change in the CD of Trp pair with changes in distance between the indole rings.

For further analysis, I considered the distance between the Trp fixed, and varied their relative orientation. I started with the geometry of a Trp pair shown in Figure II-18a. The two Trp are stacked in parallel planes with the center-to-center vector perpendicular to both planes. I have chosen the coordinate system with the origin at the center of one of the Trp (W1), which is fixed in the xy plane with its long axis along the

Figure II-19. CD of a Trp in P45 conformation (Fig. II-17c) from 3Å to 10Å with one Å step, and at 12Å, 14Å and 16Å between the centers of Trp rings in the region 160 - 300 nm (a); in the near uv at distances 3Å to 10Å, 12Å and 14Å (b). Center of indole ring is defined as the middle of CE2-CD2 bond. The CD intensity decreases monotonically as distance increases, other than for 3Å distance as pointed out in the figure.


a)
6e+5
4e+5
2e+5
0e+0
-2e+5
-4e+5
-6e+5
[θ]
16 A
3 A
5 A
3 A
4 A
155
175
195
215
235
255
275
295
wavelength (nm)
b)
6e+4
4e+4
2e+4
0e+0
-2e+4
-4e+4
[θ]
240
260
280
300
wavelength (nm)



y = - 119.68 + 3170.8x R^2 = 0.995

y = - 205.35 + 4260.1x - 2766.8x^2 R^2 = 0.999

y = - 84.858 + 1950.1x + 1.0160e+4x^2 - 2.1426e+4x^3 R^2 = 1.000

Figure II-20. Dependence of the couplet strength of the Trp pair in P45 geometry upon the center-to-center distance between Trp and interpolation of the dependence with 3rd order polynomial (equations shown for 1st, 2nd and 3rd order polynomials).

y axis. This system of coordinates is very convenient to analyze the dependence of the CD intensity upon the relative orientation of the Trp rings, which can be performed by rotation of W2. The direction of the largest electric dipole transition moment in indole (B_b) almost coincides with its long axis. Thus rotation of W2 around the y axis won't significantly change the direction of μ_{2Bb} (Fig. II-18b). On the contrary, rotation of W2 around the z axis (Fig. II-18c) and around the x axis (Fig. II-18d) does change the direction of that transition dipole moment.

I performed calculations of the exciton couplet CD for a pair of Trp starting from the parallel stacked geometry (Fig.II-21a) and rotating the W2 indole around the z axis (Fig.II-18c) and around the x axis (Fig.II-18d). I chose to follow the changes in CD of the pair at 232 nm, which corresponds to the longer wavelength part of the B_b couplet, at the L_a maximum at 263 nm and at the L_b band at 283 nm. The bands in the near uv could be compared to the experimental CD of a protein. The B_b transition is the most intense in the indole spectrum, so it is likely to be elucidated from the far uv CD of model compounds (Auer, 1973). The B_b transition is also the one that can be, in some cases, estimated from the protein far-uv CD (Chapter III). The choice of the band at 232 nm over the one at 219 nm in the B_b region was made to minimize overlap with the $n\pi^*$ peptide transitions ($\sim$220 nm), so the obtained estimates will be easier to apply for direct comparison with protein CD. Following the CD intensity of the transitions in both the near and far uv upon relative rotation of the indole rings will also provide us with data to estimate the correlation between those transitions.

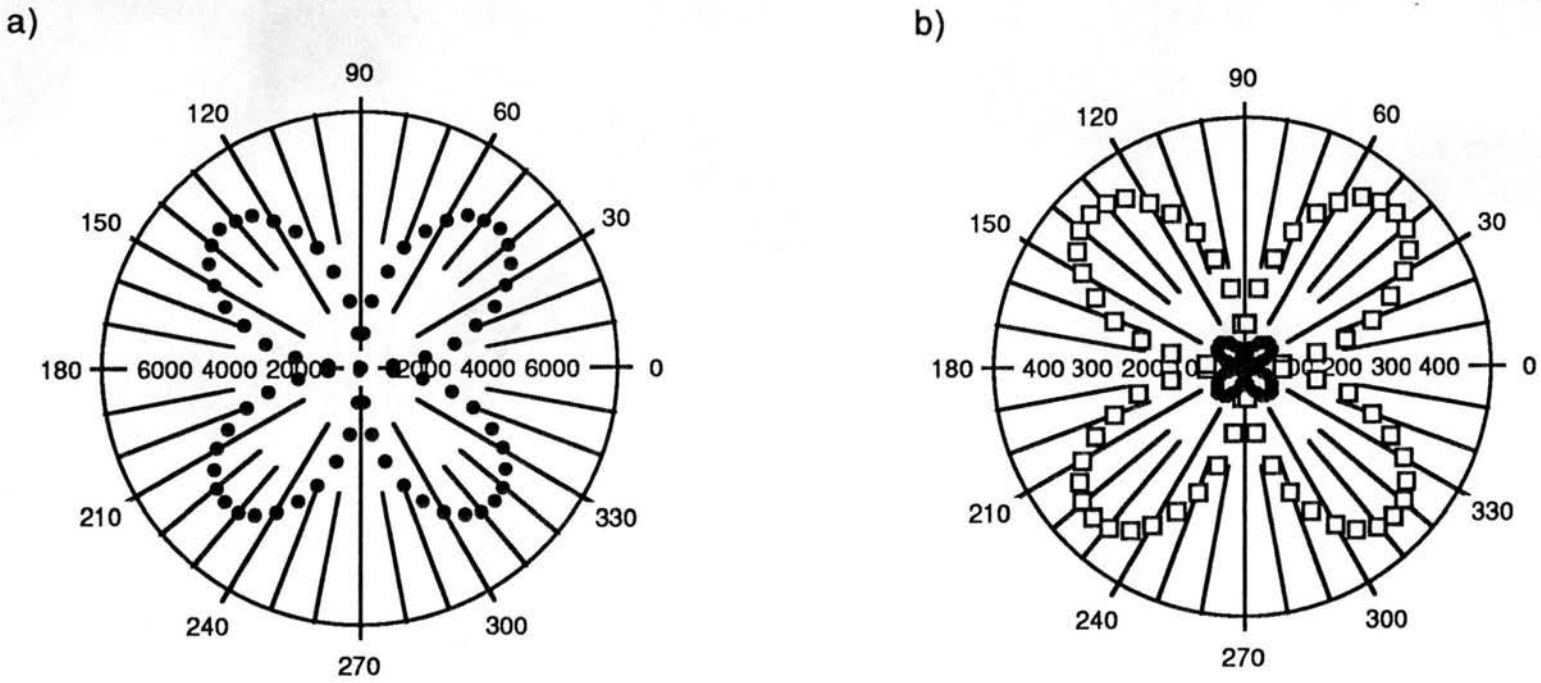


Figure II-21. Dependence of the CD intensity of a Trp pair at 6 Å upon rotation of W2 around the z axis : a) in the far uv at 232 nm (•); b) in the near uv at 263 nm (□) and 283 nm (◆). CD intensity is shown in [Θ] (deg cm^2 / dmol) per 100 amino acids.

The results for the rotation around the z axis are shown in Figure II-21. As can be seen, the magnitude of the CD at each band is a function with $\pi/2$ periodicity, and the same CD magnitudes occur twice during each period. The sign of the CD is positive in the 1st and 3rd quadrants and negative in the 2nd and 4th quadrants. The CD intensities at bands in the near and far uv in this experiment look synchronous, reflecting the same preferences for the relative orientation of the corresponding dipole transition moments. The results are presented in $[\theta]$ per 100 amino acids to make the value of CD intensity comparable to the experimental values for proteins calculated on per residue basis. The intensities for the B_b band at 232 nm are the largest. Intensities for the L_a band are an order of magnitude smaller, and about two orders smaller for the L_b band. The observed periodicity of the CD intensity and its maximum as a function of the relative orientation of the indole rings agrees with the dipole-dipole approximation, even though actual calculations are performed in the monopole approximation, which is certainly more appropriate at the range of distances between Trp considered.

The periodicity for the intensity of CD upon indole rotation around the x axis (Fig.II-18d) is also π (Fig.II-22) and the function changes sign every half period. The amplitudes at the L_a band are out of phase with the L_b and B_b bands, which results from the fact that L_b and B_b bands make a positive angle with the y axis, whereas L_a makes a negative angle. The maximum CD intensities, which are determined by the relative angle between the electric dipole transition moments in each case is, of course, the same and occurs at $\pm 45°$. As can be already seen from Figure II-21 and Figure II-22, the problem of identifying the conformation of a Trp pair from its CD has multiple solutions

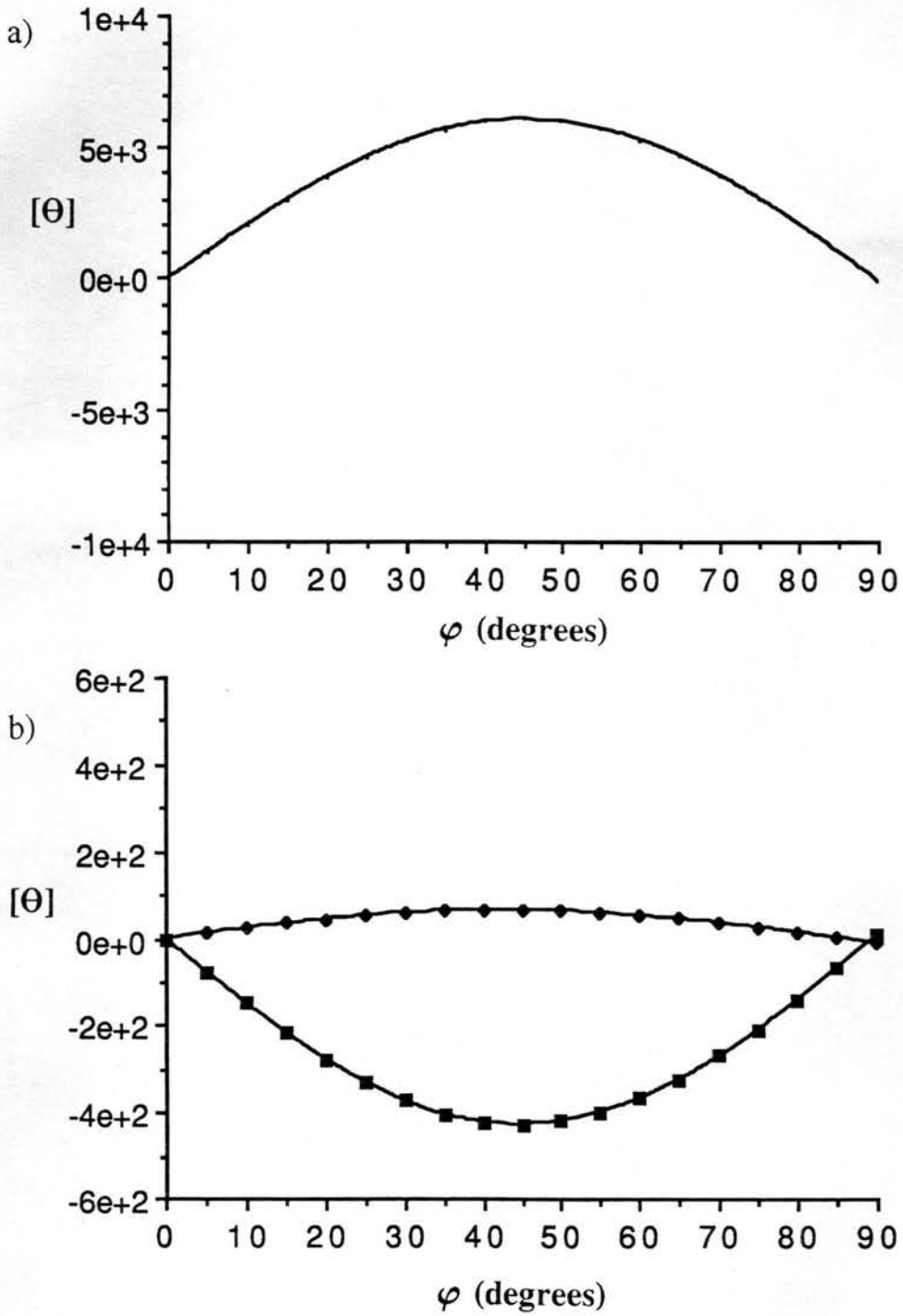


Figure II-21. Dependence of the CD intensity of a Trp pair at 6 Å upon rotation of W2 around the x axis : a) in the far uv at 232 nm (•); b) in the near uv at 263 nm (■) and 283 nm (♦). CD intensity is shown in [Θ] (deg cm^2 / dmol) per 100 amino acids.

due to periodicity of the function. Eight solutions could be obtained by rotation around the z axis only and four by rotation around the x axis only. In fact, for any value between the maximal and minimal amplitude possible at a particular distance for any rotation around one of the mentioned axes, four solutions by rotation around the other axis could be found. Two of these solutions can be eliminated if we can estimate the sign of the first derivative $\partial[\theta]/\partial\varphi$, where φ is a small variation of the angle in the area. The ambiguity in estimates of the relative conformation of Trp pair from CD might be narrowed if some structural data could be incorporated, for instance, if we are interested in local conformational changes in proteins in solution due to substrate binding or other changes in external variables. If the crystal structure for the particular protein is known, then the area of conformational freedom of the particular Trp pair could be limited significantly by considering the rigidity of the backbone and restrictions from the neighboring side-chains as well. Thus, with careful consideration of structural data, CD might be applied to estimate the local conformational changes. If, instead of, or together with, the x-ray coordinates, some modeling of the local structure around the Trp pair in a protein could be performed, then CD could be used to discriminate between several theoretical solutions. In any case, our predictions make a bridge between spectroscopic properties of a system of chromophores and quantitative analysis of changes in their relative conformation in the protein structure.

A sufficiently detailed scan of the CD of a Trp pair (in $[\theta]$ per 100 a.a) through its conformational space is shown in Figures II-23, II-24, II-25. These tables present CD amplitudes at the negative (Fig. II-23) and positive (Fig. II-24) extrema in the region

of the B_b couplet (215-235 nm) in the sector from 0 to -45° for the rotation around the z axis and from 0 to -90° for the rotation around the x axis. Information on both maxima of the couplet are included because the negative band in the 219-229 interval is measurable over almost all conformational space, other than near R_x=-90°, whereas the positive band in that region occurs over a broader interval 214 - 250 nm. The bands with maxima at longer wavelength than 235 nm are left out, to simplify Figure II-24. Their conformational characteristics correspond to the blank lower left corner. The positions of bands occurring below 221 nm are given in parentheses (R_z =0; R_x=-50 to -90°). The CD values are calculated at 5° steps in each direction, which seems a reasonable accuracy for relating the CD with corresponding conformational changes. Figure II-25 provides analogous information for CD in the near-uv region, which reflects mostly the L_a transition. Analysis of the tables can be systematized by dividing them into sectors, corresponding to changes in CD amplitudes of 10^3 deg cm^2 / dmol per 100 amino acids in the far uv and 50 to 25 deg cm^2 / dmol per 100 amino acids in the near uv. The predicted CD values are shown in units per 100 amino acids to ease the comparison with Trp CD in globular proteins, where experimental CD is usually calculated on per residue basis. For both L_a and B_b bands, the CD generally increases while R_z changes from 0 to -45°, although the profile of CD changes as a function of the two angles is complex (Fig. II-23, II-24, II-25) and differs for the B_b and L_a bands. CD changes at each band depend both on the initial conformation and about which axis the rotation is performed.

These data (Fig. II-20, II-23, II-24, II-25) can be used to provide rough estimates of the conformational changes in Trp pairs in model compounds, and, perhaps, in

Rotation around x axis \ Rotation around z axis	0°	-5°	-10°	-15°	-20°	-25°	-30°	-35°	-40°	-45°
0°	0	-1294	-2557	-3739	-4796	-5722	-6457	-7004	-7333	-7436
-5°	-138	-1362	-2601	-3737	-4764	-5660	-6368	-6881	-7193	-7287
-10°	-289	-1405	-2608	-3725	-4710	-5565	-6246	-6736	-7020	-7094
-15°	-457	-1441	-2614	-3696	-4647	-5465	-6097	-6551	-6813	-6866
-20°	-649	-1459	-2599	-3639	-4558	-5330	-5922	-6344	-6570	-6599
-25°	-854	-1465	-2565	-3568	-4442	-5171	-5718	-6095	-6290	-6290
-30°	-1081	-1461	-2516	-3471	-4299	-4980	-5484	-5820	-5970	-5947
-35°	-1319	-1448	-2441	-3346	-4134	-4762	-5222	-5508	-5619	-5569
-40°	-1560	-1438	-2334	-3206	-3924	-4519	-4919	-5160	-5233	-5152
-45°	-1797	-1440	-2215	-3044	-3708	-4223	-4573	-4768	-4806	-4694
-50°	-2024	-1487	-2084	-2830	-3443	-3903	-4193	-4339	-4339	-4206
-55°	-2224	-1572	-1936	-2589	-3152	-3539	-3772	-3868	-3833	-3683
-60°	-2375	-1671	-1789	-2335	-2811	-3149	-3322	-3365	-3298	-3134
-65°	-2460	-1761	-1643	-2060	-2442	-2712	-2830	-2831	-2733	-2557
-70°	-2458	-1787	-1510	-1763	-2054	-2251	-2315	-2280	-2156	-1967
-75°	-2354	-1739	-1388	-1484	-1668	-1771	-1783	-1708	-1567	-1379
-80°	-2135	-1607	-1257	-1217	-1280	-1304	-1248	-1139	-981	-794
-85°	-1800	-1374	-1083	-974	-932	-864	-751	-599	-433	-291
-90°	-1081	1174	1379	1603	1811	1996	2138	2246	2324	2382

Figure II-23. CD of Trp pair at 6 Å distance and various orientations of indole rings. Amplitude at the negative minimum in the interval 219-229 nm. CD intensity is shown in [Θ] (deg cm^2/ dmol) per 100 amino acids.

Rotation around z axis

Rotation around x axis	0°	-5°	-10°	-15°	-20°	-25°	-30°	-35°	-40°	-45°
0°	-	1034	2048	2999	3851	4611	5216	5682	5972	6086
-5°	-	939	1951	2878	3726	4482	5084	5533	5825	5943
-10°	-	818	1805	2741	3573	4314	4917	5368	5649	5762
-15°	-	683	1664	2584	3409	4142	4721	5162	5448	5555
-20°	-	518	1502	2396	3220	3937	4506	4942	5212	5314
-25°	-	357	1319	2206	3003	3710	4262	4680	4945	5038
-30°	-	195	1109	1998	2773	3454	3997	4402	4644	4732
-35°	-	62	886	1756	2535	3182	3715	4097	4323	4403
-40°	-		660	1509	2262	2909	3402	3767	3976	4038
-45°	-		447	1260	1989	2599	3058	3403	3598	3641
-50°	307 (214)		248	987	1686	2280	2708	3017	3186	3223
-55°	393 (215)		98	735	1389	1930	2334	2604	2749	2773
-60°	477 (215)		10	501	1077	1586	1947	2180	2295	2304
-65	545 (216)			290	793	1232	1543	1739	1827	1814
-70	588 (216)			129	527	886	1146	1301	1351	1318
-75	586 (216)			32	295	569	756	860	879	824
-80	526 (216)				115	281	396	438	416	342
-85	407 (215)				12	65	89	69	19	
-90		1174	1379	1603	1811	1996	2138	2246	2324	2382

Figure II-24. CD of Trp pair at 6 Å distance and various orientations of indole rings. Amplitude at the positive maximum in the interval 214-235 nm. CD intensity is shown in [Θ] (deg cm^2/ dmol) per 100 amino acids.

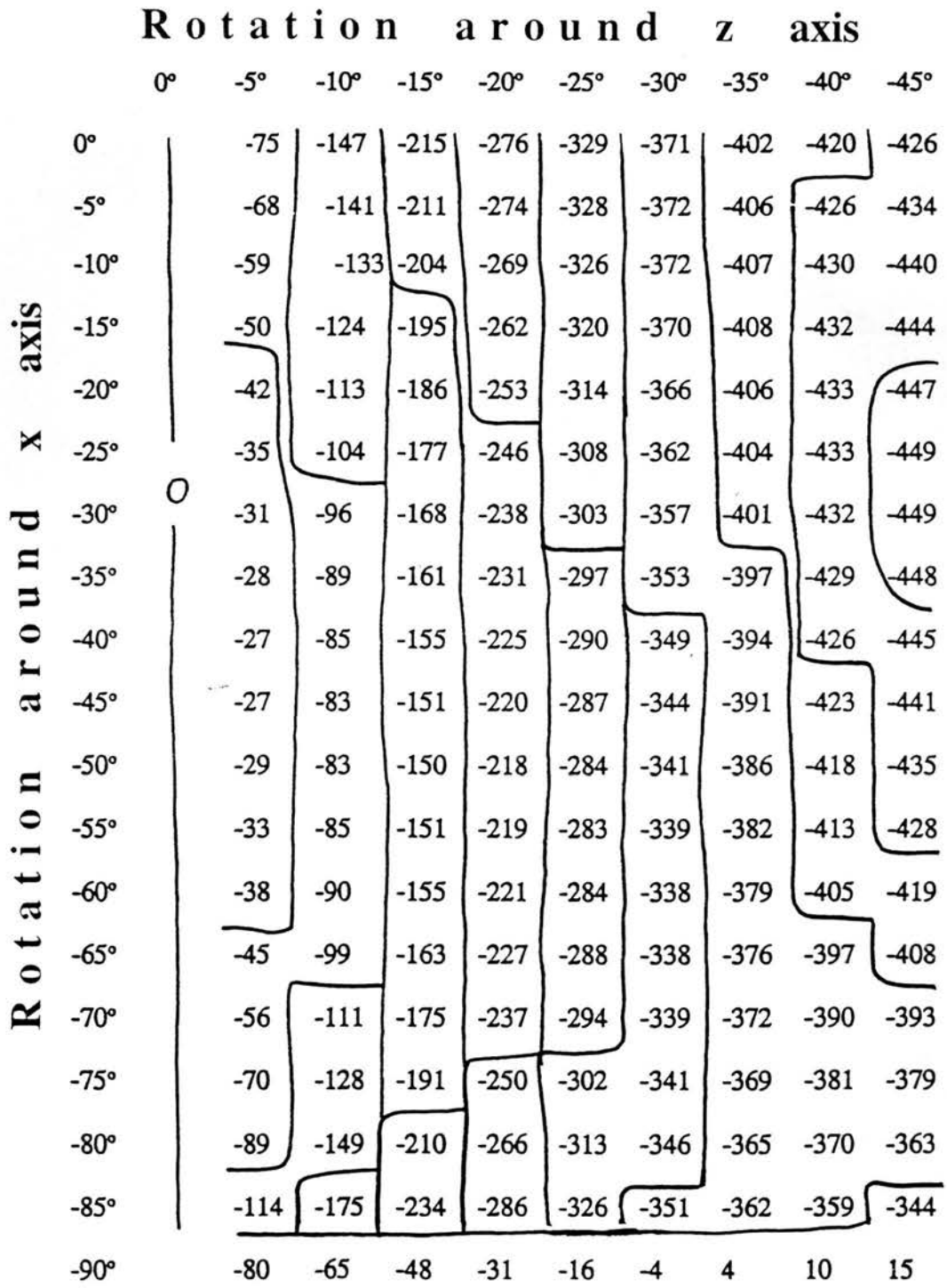


Figure II-25. CD of Trp pair at 6 Å distance and various orientations of indole rings. Amplitude at the negative minimum in the interval 250-270 nm. CD intensity is shown in [Θ] (deg cm^2/ dmol) per 100 amino acids.

proteins as well. Practically, one should compare the experimental CD intensity in several intervals both in the near and far uv with the data in the tables and find the appropriate sector. Data for other bands from my calculations are available as well, such as those around 180 nm, 190 nm and 200 nm.

4. Trp CD of Globular Proteins.

a. Survey of the Protein Data Bank.

As can be seen from Figure II-2 Trp pairs within 13 Å of each other and in a favorable conformation might give a couplet strength ($\Delta\epsilon_+$-$\Delta\epsilon_-$) of about 200, which corresponds to a band with 165,000 deg cm^2/dmol at the maximum, calculated per Trp chromophore. The average CD intensity for an $\alpha+\beta$ protein is around 10^6 deg cm^2/dmol. Thus the CD of a pair of Trp even at 13 Å, in favorable conformation, might provide 15-20 % of the amplitude of protein CD.

I calculated the magnitude of B_b Trp CD for proteins for which coordinates were available in the Brookhaven Protein Data Bank (PDB) and which contain pairs of Trp pairs residues within 13 Å. All Trp in each protein were included. Predicted CD values at all extrema are reported in Table II-1 in 10^3 deg cm^2/dmol, where dmol refer to the amino acid residues in the protein, i.e., the calculated Trp CD is divided by N, the number of amino acid residues in the protein. Table II-1 also provides the amplitude of the experimental CD, where available, to estimate the relative contribution of the predicted Trp CD to the experimental CD of the protein. Values in parenthesis in Table II-1 show data per number of Trp chromophores in the protein, which is more appropriate to reveal groups of Trp in conformations most favorable for CD.

Proteins in Table II-1 are arranged according to functional classes and related structures. As can be seen, the calculated exciton Trp CD in globular proteins is usually a couplet band. For some proteins three extrema are predicted, such as in case of hen egg-white lysozyme, monomeric bovine chymotrypsinogen A and modified human carbonic anhydrase. Initially Trp CD was calculated for all 118 proteins for which Trp pairs appear in Figure II-2. Then groups of related proteins for which significant CD was predicted (more than $1.5x10^3$ deg $cm^2/dmol$), were extended and are discussed in this chapter. As can be seen from Table II-1, a number of proteins show significant B_b Trp CD, which reaches about 10-30 % of the amplitude of experimental far uv spectrum. This is the case in several enzyme classes such as lysozymes and chymotrypsins. Significant CD was predicted for concanavalin A, dihydrofolate reductase, high potential iron potential protein and several others. Trp CD in evolutionarily related proteins, for different crystal forms and protein modifications of lysozymes, trypsins, chymotrypsins and others is described in the present chapter.

The information on absolute and relative Trp contribution to the far uv CD in globular proteins presented in Table II-1 can be applied: 1) to compose protein basis sets for estimation of protein secondary structure by deconvolution approaches (Chapter III.2), and 2) to estimate the tertiary structure changes that correlate with Trp conformations in cases where Trp CD is significant.

Table II-1. Calculated Trp B_b contribution to the proetein CD.

FILE	PROTEIN	CALCULATED CD[a]			EXPERIMENTAL CD[b]
		EXTREMUM1[c]	EXTREMUM2	EXTREMUM3	
	CALCIUM BINDING PROTEIN				
1ALC	α-LACTALBUMIN, BABOON MILK	.870 (35.7)	-.974 (-39.9)		(baboon) -9.60
	TRANSFERASE				
1RHD	RHODANESE, BOVINE LIVER	1.017 (42.6)	-1.114 (-47.8)		(229 nm) -4.64
	HYDROLASE				
1CHG	CHYMOTRYPSINOGEN A	-1.032 (-31.6)	2.192 (67.1)	-.916 (28.1)	(230 nm) -1.80 (210 nm) -8.00
1CHO	α-CHYMOTRYPSIN COMPLEX WITH TURKEY OVOMUCOID THIRD DOMAIN	3.346 (124.2)	-3.797 (-141.0)		(200 nm) -12.00
6CHA	α-CHYMOTRYPSIN A WITH PHENYLETHANE BORONIC ACID, COW	3.211 (96.7)	-3.592 (-108.2)		(229 nm) -4.64
7GCH	γ CHYMOTRYPSIN + N-ACETYL-L--LEUCYL-L-PHENYLALANYL TRIFLUORO-METHYL KETONE AT THE ACTIVE SITE, BOVINE	2.543 (76.6)	-2.832 (-85.3)		
2TGD	TRYPSINOGEN, DIISOPROPYL-PHOSPHORYL INHIBITED, BOVINE PANCREAS	1.150 (65.8)	-1.291 (-73.9)		(210 nm) -4.40
2TGP	-//- COMPLEX WITH PANCREATIC TRYPSIN INHIBITOR, BOVINE PANCREAS	.937 (67.2)	-1.051 (-75.4)		
2TGT	-//-, BOVINE PANCREAS	1.155 (66.1)	-1.296 (-74.2)		
2TPI	-//- -PANCREATIC TRYPSIN INHIBITOR-ILE-VAL COMPLEX BOVINE PANCREAS	1.031 (74.0)	-1.156 (-82.9)		(kalinogen inhibitor) (227 nm) -5.50
2TRM	TRYPSIN (ASP 102->ASN) WITH BENZAMIDINE, RAT	1.135 (63.3)	-1.273 (-71.0)		
1TPP	β-TRYPSIN WITH P-AMIDINO--PHENYL-PYRUVATE, BOVINE PANCREAS	1.000 (55.8)	-1.122 (-62.3)		
2PTN	TRYPSIN BOVINE PANCREAS	.964 (53.7)	-1.082 (-60.3)		(208 nm) -4.22

3PTB	β-TRYPSIN BENZAMIDINE INHIBITED BOVINE PANCREAS	.244 (13.6)	-1.059 (-59.0)	(235 nm) -.25
3EST	NATIVE ELASTASE, PORCINE	-.819 (-28.1)	.920 (31.5)	
1HNE	HUMAN NEUTROPHIL ELASTASE INHIBITED	-.047 (-3.5)	.053 (3.9)	
1TON	TONIN RAT HYDROLASE	1.067 (62.7)	-1.193 (- 70.1)	
2PAD	PAPAIN PAPAYA	2.015 (85.4)	-2.257 (-95.7)	-9.00
9PAP	PAPAIN CYS-25 OXIDIZED, PAPAYA	.236 (10.0)	-.265 (-11.2)	
1PPD	2-HYDROXYETHYLTHIOPAPAIN PAPAYA	-.361 (-15.3)	.233 (17.2)	
2RSP	ROUS SARCOMA VIRUS PROTEASE	2.325 (144.2)	-2.602 (-161.3)	
1PSG	PEPSINOGEN PORCINE	-1.007 (-74.5)	1.128 (83.5)	(212 nm) -8.00
5PEP	PEPSIN, PIG	-.953 (-62.1)	1.068 (69.6)	
2PKA	KALLIKREIN A, PORCINE PANCREAS	2.492 (82.6)	-2.800 (-92.8)	(203 nm) -9.25
1TEC	THERMITASE + EGLIN-C	-.0014 (-.081)	.0012 (.07)	
3RP2	PROTEINASE, RAT	.583 (43.5)	-.653 (-48.8)	
7LYZ	LYSOZYME, TRICLINIC CRYSTAL HEN EGG WHITE	-2.708 (-58.2)	2.222 (47.8)	-7.20
2HFL	LYSOZYME (HEW) + IG*G1 FAB HYHEL-5	-.456 (-16.8)	.521 (19.2)	
	a. LYSOZYME	-2.028 (-43.6)	2.275 (48.9)	
	b. FAB	-.136 (-8.4)	.165 (10.2)	
3HFM	LYSOZYME (HEW) + IG*G1 FAB HYHEL-10	.548 (21.8)	-.767 (-30.6)	
	a. LYSOZYME	-3.125 (-67.2)	3.472 (69.3)	
	b. FAB	.528 (32.8)	-.591 (-36.7)	
	LECTIN			
1CN1	CONCANAVALIN A JACK BEAN	-3.005 (-178.0)	3.293 (195.1)	-7.80

	ELECTRON TRANSPORT			
1CYC	FERROCYTOCHROME C	-.871 (-44.9)	.975 (50.2)	-7.40
2CCY	CYTOCHROME C	4.448 (189.8)	-4.882 (-208.3)	-11.30
3CYT	CYTOCHROME C, TUNA HEART	1.126 (58.6)	-1.262 (-65.6)	-12.50
451C	CYTOCHROME C551(REDUCED)	-.724 (-29.7)	.810 (33.2)	
4FD1	FERREDOXIN	-.116 (-6.1)	.130 (6.9)	-6.35
	ELECTRON TRANSFER			
1FX1	FLAVODOXIN	-.637 (-47.1)	.713 (52.8)	species dependant
3FXN	FLAVODOXIN (OXIDIZED)	.868 (39.9)	-.972 (-47.7)	
1HIP	HIGH POTENTIAL IRON PROTEIN	8.272 (234.4)	-9.220 (-261.2)	
	PEPTIDE ANTIBIOTIC			
1GMA	GRAMICIDIN A	-22.313 (-94.8)	24.998 (106.2)	
	OXIDOREDUCTASE			
1GP1	GLUTATHIONE PEROXIDASE	.576 (57.0)	-.646 (-64.0)	
3GRS	GLUTATHIONE REDUCTASE, OXIDIZED	-.015 (-2.3)	.017 (2.7)	(bovine) -6.05
1FCB	FLAVOCYTOCHROME B2 YEAST	-.0083 (-1.2)	.0093 (1.4)	
2CYP	CYTOCHROME C PEROXIDASE	-1.805 (-75.8)	2.022 (84.9)	(bovine) -8.60
3CPP	CYTOCHROME P450CAM - REDUCED	-2.294 (-189.9)	2.554 (214.5)	
2DHF	DIHYDROFOLATE REDUCTASE, INHIBITED	.180 (11.2)	-.201 (-12.5)	
5DFR	APO-DIHYDROFOLATE REDUCTASE (E.coli)	-4.069 (-129.4)	4.480 (142.5)	
2LDB	L-LACTATE DEHYDROGENASE	-.017 (-1.8)	.020 (2.1)	
2LDX	APO-LACTATE DEHYDROGENASE	-.596 (-32.9)	.669 (36.9)	
8LDH	M4 APO-LACTATE DEHYDROGENASE	.445 (24.5)	-.498 (-27.4)	
	OXYGEN TRANSPORT and STORAGE			
1HDS	HEMOGLOBIN (DEER)	.267 (25.5)	-.299 (-28.5)	
2HHB	HEMOGLOBIN (DEOXY), HUMAN	.310 (29.7)	-.347 (-33.2)	
1THB	HEMOGLOBIN (T STATE), HUMAN	.175 (20.1)	-.196 (-22.5)	-32.00
2LHB	HEMOGLOBIN V, SEA LAMPREY	-.098 (-7.3)	.109 (8.1)	-16.00

2MHB	HEMOGLOBIN (HORSE)	.235 (22.4)	-.263 (-22.6)		
1LH3	LEGHEMOGLOBIN, YELLOW LUPIN	2.322 (118.4)	-2.572 (-131.2)		
1MBS	MYOGLOBIN, COMMON SEAL	.762 (58.3)	-.852 (-65.2)		
1PMB	MYOGLOBIN, PORCINE	.413 (31.6)	-.463 (-35.4)		
4MBA	MYOGLOBIN, SEA HARE	-1.826 (-134.2)	2.043 (150.2)		
4MBN	MYOGLOBIN, SPERM WHALE	.400 (30.6)	-.428 (-32.7)		-25.70
1HMQ	HEMERYTHRIN SIPUNCULID WORM	-1.693 (-63.7)	1.894 (71.3)		
2MHR	MYOHEMERYTHRIN SIPUNCULAN WORM RETRACTOR MUSCLE	-.025 (-.98)	.028 (1.1)		
	PHOTOSYNTHETIC REACTION CENTER				
1PRC	RHODOPSEUDOMONAS VIRIDIS	-.806 (-22.9)	.706 (20.0)		
	LYASE				
2CAB	CARBONIC ANHYDRASE B, HUMAN	1.092 (45.7)	-1.228 (-53.4)		
3CA2	-//- INHIBITED, HUMAN	-.451 (16.7)	.790 (29.2)	-.229 (8.5)	
6CTS	CITRATE SYNTHASE/COENZYME A	.194 (9.3)	-.232 (-11.2)		
	T-CELL SURFACE GLYCOPROTEIN				
2CD4	CD4(1-182) HUMAN EXPRESSED IN CHINESE HAMSTER OVARY CELLS	-.513 (-31.1)	.577 (35.0)		
	PERIPLASMIC BINDING PROTEIN				
2GBP	D-GALACT./D-GLUC. BINDING PROTEIN	-1.300 (-80.7)	1.458 (90.1)		
2LBP	LEU-BINDING PROTEIN, MOUSE	.446 (38.6)	-.501 (-43.3)		
	EYE LENS PROTEIN				
2GCR	γ IVA-CRYSTALLIN, BOVINE LENS	-.867 (-37.5)	.971 (42.0)		(dogfish) (217 nm) -4.18
	LIGASE				
2GLS	GLUTAMINE SYNTHETASE	-.020 (-9.4)	.023 (10.8)		(216 nm) -7.10
2TS1	TYROSYL-tRNA SYNTHETASE	-.259 (-18.1)	.290 (20.3)		
	TRANSPORT IN SERUM				
2PAB	PREALBUMIN HUMAN	1.348 (85.6)	-1.509 (-95.8)		(215 nm) -4.70

	GENE REGULATORY PROTEIN		
3WRP	TRP APOREPRESSOR, E. COLI	.389 (21.0)	-.436 (-23.5)
	HISTOCOMPATIBILITY ANTIGEN		
3HLA	HUMAN CLASS I HISTOCOMPATIBILITY ANTIGEN A2.1 LYMPHOBLASTOID CELL	-3.043 (124.8)	3.406 (-139.6)
	HORMONE		
4RLX	RELAXIN, PIG OVARY	1.027 (26.7)	-1.156 (-30.0)
	IMMUNOGLOBULIN		
1MCW	IMMUNOGLOBULIN HETEROLOGOUS LIGHT CHAIN DIMER HUMAN	.820 (59.0)	-.973 (-70.1)
1PFC	P/FC(PRIME) FRAGMENT OF AN IG*G1 GUINEA PIG SERUM	1.037 (58.6)	-1.161 (-65.6)
2MCP	IMMUNOGLOBULIN MC/PC603 FAB-PHOSPHOCHOLINE COMPLEX MOUSE	-.752 (-55.4)	.842 (62.0)
2RHE	BENCE-JONES PROT., HUMAN MYELOMA	-.116 (-6.6)	.130 (7.4)
3FAB	λ IMMUNOGLOBULIN FAB HUMAN	.278 (17.0)	-.312 (-19.1)
3MCG	IMMUNOGLOBULIN LAMBDA LIGHT CHAIN DIMER, HUMAN	.120 (8.5)	-.135 (-9.7)
	VIRUS PROTEINS		
1BMV	BEAN POD MOTTLE VIRUS	-.311 (-11.9)	.351 (13.4)
2HMG	INFLUENZA VIRUS HEMAGGLUTININ	-.355 (-19.8)	.398 (22.2)
2MEV	CARDIO PICORNAVIRUS COAT PROTEIN	.441 (21.6)	-.491 (-24.1)
2PLV	POLIOVIRUS HUMAN	.372 (25.2)	-.422 (-28.6)
2RR1	RHINOVIRUS COAT PROTEIN	.126 (18.2)	-.141 (-20.4)
4RHV	RHINOVIRUS 14 COAT PROTEIN	.257 (18.3)	-.287 (-20.4)
4SBV	VIRUS COAT PROTEIN	.464 (30.3)	-.521 (-34.0)

a. CD is shown in 10^3 deg cm^2/dmol. The first number in each column shows CD at the maxima per residue of the protein, second number in parenthesis shows CD per Trp chromophore in a protein.

b. Experimental protein CD in 10^3 deg cm^2/dmol is measured at 222 nm, if not specified otherwise.

c. Extrema are numbered in order of increasing wavelength.

b. Hen egg-white lysozyme.

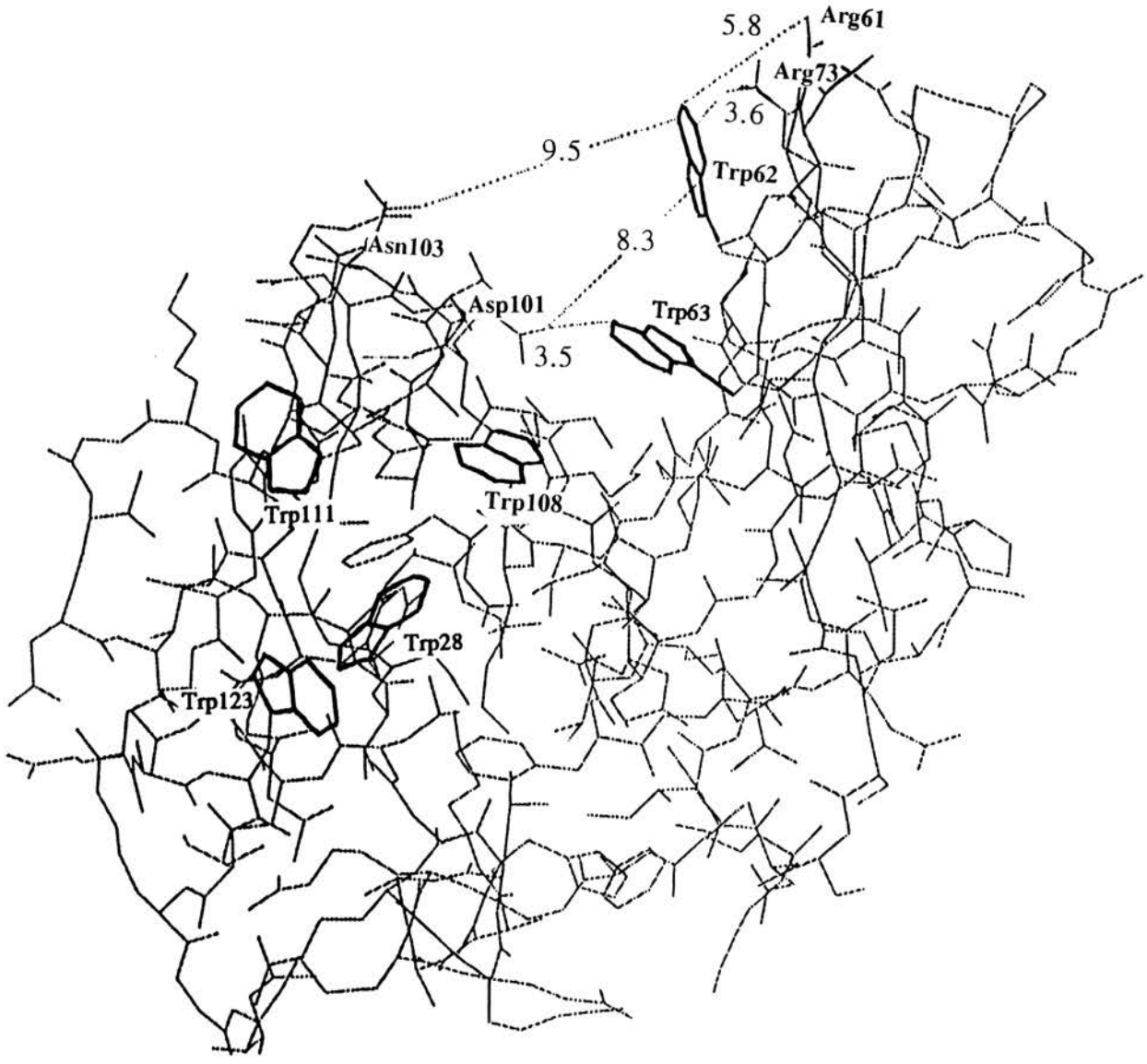
(1) Tetragonal and Triclinic Lysozyme.

Hen egg-white lysozyme (HEWL) is a relatively small (129 a.a.) and well-studied protein. It contains six Trp residues, Trp28, Trp62, Trp63, Trp108, Trp111 and Trp123, three of which, Trp62, Trp63 and Trp108, lie in the active site cleft (Fig.II-26) and are involved in substrate binding (Goux & Hooker, 1980a).

HEWL is known to crystallize in a variety of crystal forms (Phillips, 1966; Moult et al., 1976; Kurachi et al., 1976). The most common are the tetragonal and triclinic crystal forms, both containing one enzyme molecule per asymmetric unit (Diamond, 1974; Moult et al., 1976). Fig.II-27 shows the comparison of extrema amplitudes for the B_b Trp CD of tetragonal (2lyz), triclinic (2lzt) and monoclinic crystal forms (1lym), as well as crystals grown under pressure (2lym, 3lym) and iodine-inactivated HEWL (8lyz). As can be seen, tetragonal and triclinic lysozyme do not exhibit any considerable difference either in position or in the amplitude of the CD bands.

The exciton coupling and one-electron mixing for six indole transitions in Trp chromophores have been calculated for the L_b, L_a, B_b, B_a , V and VI indole transitions in tetragonal and triclinic lysozyme (Fig.II-28a,b). Most of the lysozyme Trp residues lie near at least one other Trp (6 - 8 Å), as can be seen from Figure II-26, and multiple chromophore interactions are important in the resulting Trp CD. Fig.II-28a shows the predicted CD for tetragonal and triclinic lysozyme, considering six indole transitions, in comparison with B_b CD. As can be seen, accounting for six transitions allows us to

Figure II-26. Conformations of Trp chromophores (bold) in hen lysozyme. Width of the cleft is shown (distance between Trp62 and Asn103). Distances between Asp101, Arg61, Arg73 and Trp62 and Trp63 are shown.


5.8
Arg61
Arg73
3.6
9.5
Trp62
Asn103
8.3
Asp101
Trp63
3.5
Trp111
Trp108
Trp28
Trp123

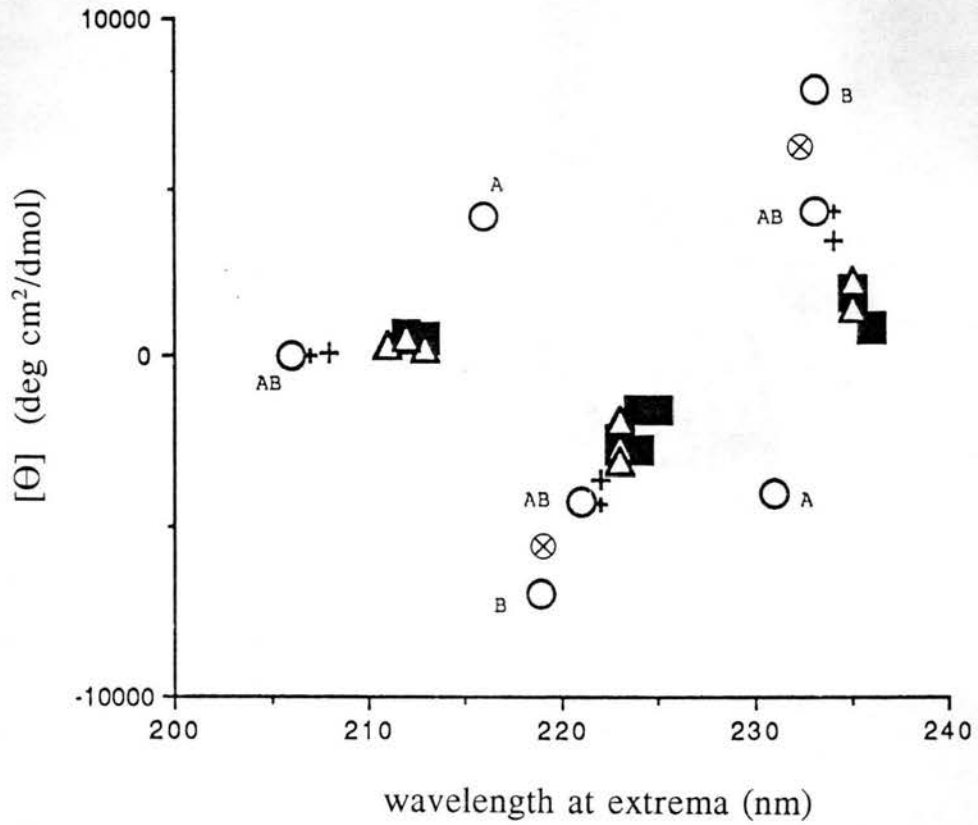
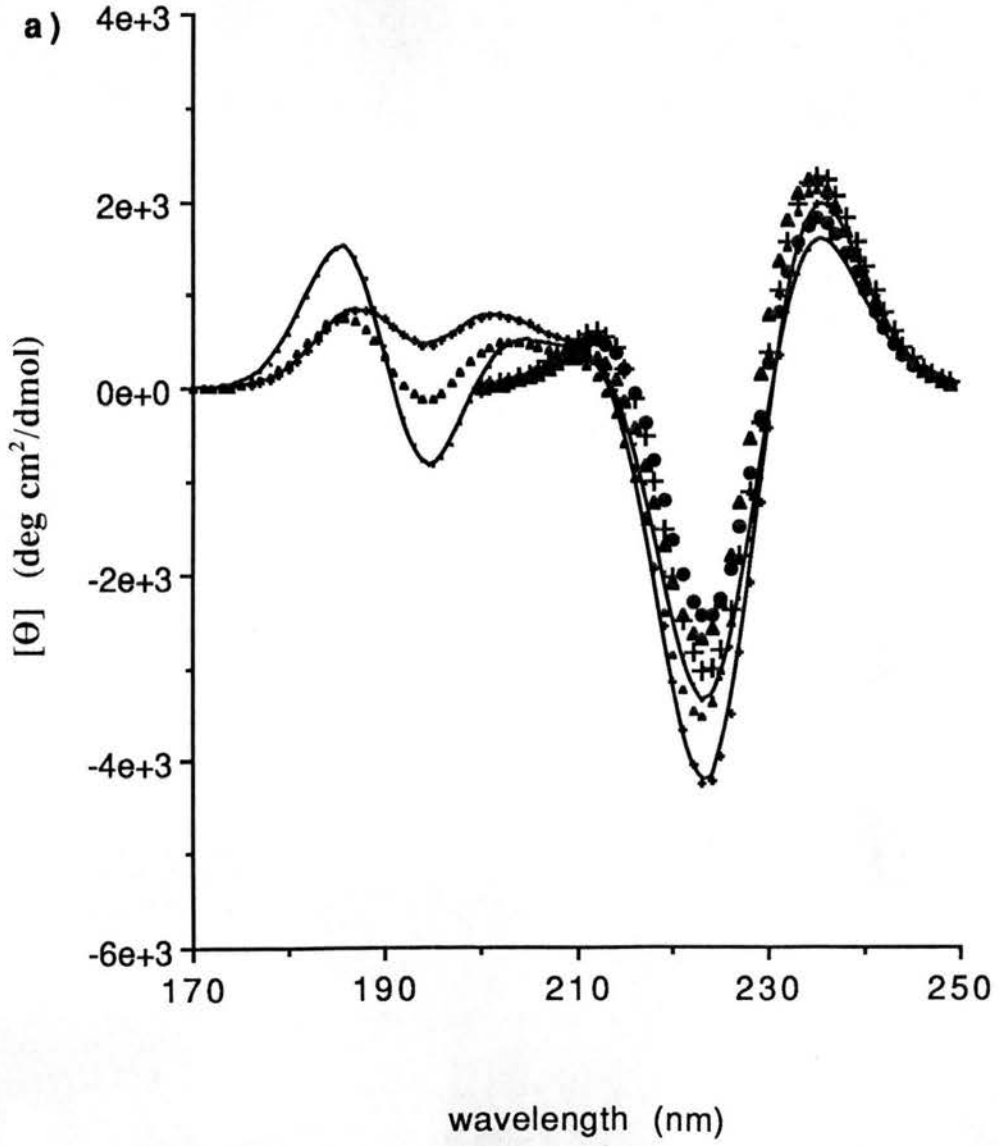
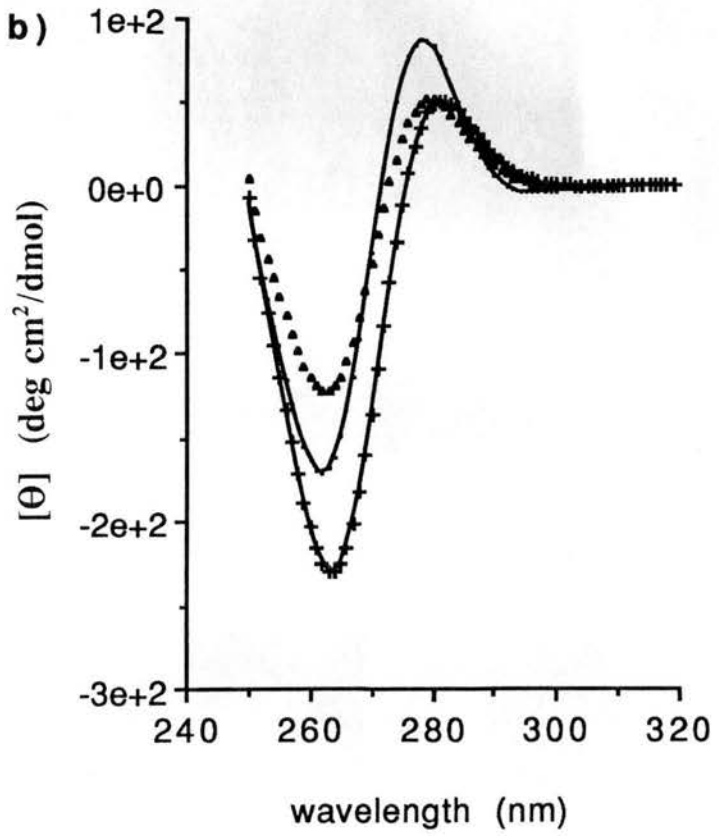


Figure II-27. Calculated B_b Trp CD of HEWL obtained at different refinements and crystal forms: tetragonal, 1lyz, 2lyz, 3lyz, 4lyz, 5lyz, 6lyz (■); triclinic, 7lyz, 1lzt, 2lzt (▵); monoclinic dimer (AB)(◯) and subunits A and B (◯); tetragonal grown under 1 atm and 1000 atm (+); iodine inactivated (⊗).

Figure II-28. Calculated Trp CD of tetragonal and triclinic HEWL in the far uv (a) and near uv (b) considering all six transitions : tetragonal (-); triclinic (▲); triclinic +substrate (+); and considering B_b transition only : tetragonal (•); triclinic (▲); triclinic +substrate (+).


a)
4e+3
2e+3
0e+0
-2e+3
-4e+3
-6e+3
[θ] (deg cm²/dmol)
170
190
210
230
250
wavelength (nm)


b)
1e+2
0e+0
-1e+2
-2e+2
-3e+2
[θ] (deg cm²/dmol)
240
260
280
300
320
wavelength (nm)

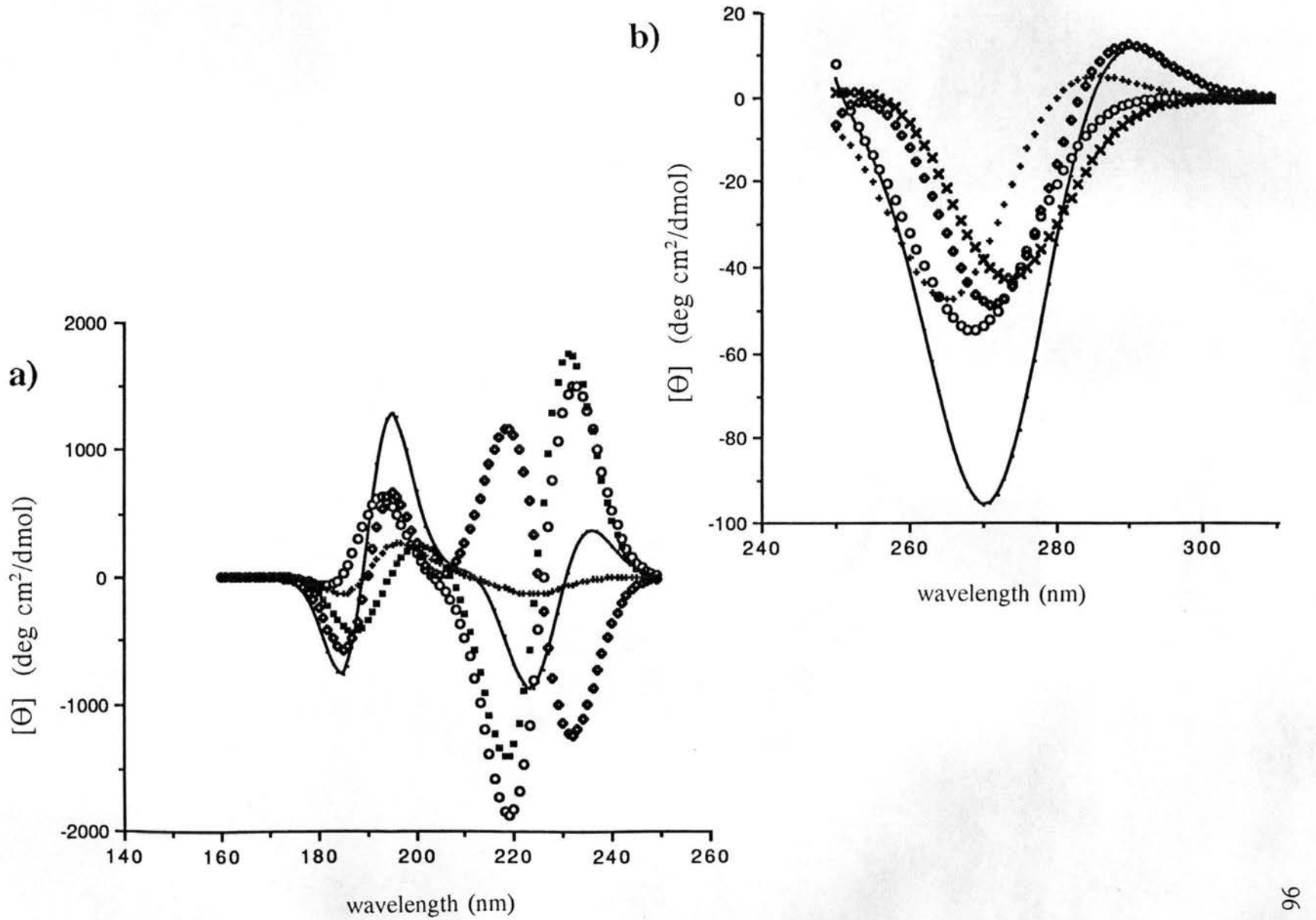
discriminate between tetragonal and triclinic HEWL forms (Fig.II-28a). The Trp B_b CD of triclinic HEWL does not show any significant changes upon binding the substrate. Considerable differences can be seen in the higher energy transitions (Fig.II-28a) between tetragonal (2lyz), triclinic HEWL (2lzt), and triclinic HEWL with substrate (1lzt).

In the near uv (Fig.II-28b), the triclinic HEWL with and without substrate show opposite amplitude changes at the L_a band with respect to tetragonal HEWL. The CD at the L_b band exhibited by triclinic HEWL shows an amplitude half as large as that of the tetragonal form (Fig.II-28b). Comparison of calculated lysozyme Trp CD with experimental spectra in the far and near uv is discussed in Chapter III.3.e.

Analysis of CD exhibited by corresponding pairs in triclinic and tetragonal lysozyme revealed that the most significant CD changes occur in pairs containing Trp28 or Trp62 (Fig.II-29). I analyzed the differences in distances and torsional angles within Trp pairs in triclinic (2lzt) and tetragonal (2lyz) lysozyme, and concluded that torsional distortions rather than changes in distances correlate with observed changes in CD. This conclusion is not suprising since we are looking at very localized conformational changes that do not affect most of the molecule.

The differences between tetragonal and triclinic forms in the B_a region become more profound when substrate is bound to the triclinic enzyme (Fig.II-28a,b). I analyzed the difference CD between Trp pairs with Trp28 and Trp62 in tetragonal (2lyz) and triclinic with substrate HEWL (1lzt), considering six indole transitions. Figure II-29a shows the difference CD of tetragonal and triclinic lysozyme in the far uv, and the difference CD of particular Trp pairs that provide the major contribution. The difference

Figure II-29. Calculated differential CD between triclinic and tetragonal HEWL (-) in the far (a) and near uv (b), and the major contributions of individual Trp pairs to the differential CD : Trp28-Trp108 (o); Trp28-Trp123 (■); Trp62-Trp63 (◇) and Trp62-Trp108 (+).


a)
b)
[Θ] (deg cm²/dmol)
wavelength (nm)
2000
1000
0
-1000
-2000
140
160
180
200
220
240
260
20
0
-20
-40
-60
-80
-100
240
260
280
300

spectrum in the B_b region is dominated by the contributions of the pairs involving Trp28: Trp28-Trp108 and Trp28-Trp123. The sign of the couplet exhibited by Trp62-Trp63 in the B_b region is opposite to that of the difference spectrum and the contribution of the other major pairs, whereas in the B_a region contribution of all major pairs coincide in sign. As result, the amplitude of the difference spectrum in the B_a region is even larger than that in B_b (Figure II-29a). In the region below 195 nm the difference CD is determined by the Trp62-Trp63 pair in the 180-190 nm region and by this pair and Trp28-Trp108 in the 190-200 nm region. In the near uv, differences in the L_a band result from additive contributions of all the mentioned pairs, although the changes in the L_b band characteristic for triclinic lysozyme appear as a result of the Trp62-Trp63 interaction only. Comparison of the tetragonal and triclinic crystal forms of HEWL has shown no significant deviation in the backbone structure, though slight differences in conformation of various surface residues have been described (Moult et al., 1976; Kurachi et al., 1976).

Tetragonal crystals are less compact, containing one dalton of the protein per 2.04Å^3, compared to 1.81Å^3 in triclinic (Hogle et al., 1981) and this might allow higher mobility for some surface residues. Upon binding of the substrate, Trp62 moves towards it (Kurachi et al., 1976). This might lead to the differences in CD obtained in our calculations. Fig.II-30 shows the superposition of Trp62-Trp63 pairs in tetragonal and triclinic crystal forms. Torsional changes can be observed at Trp62.

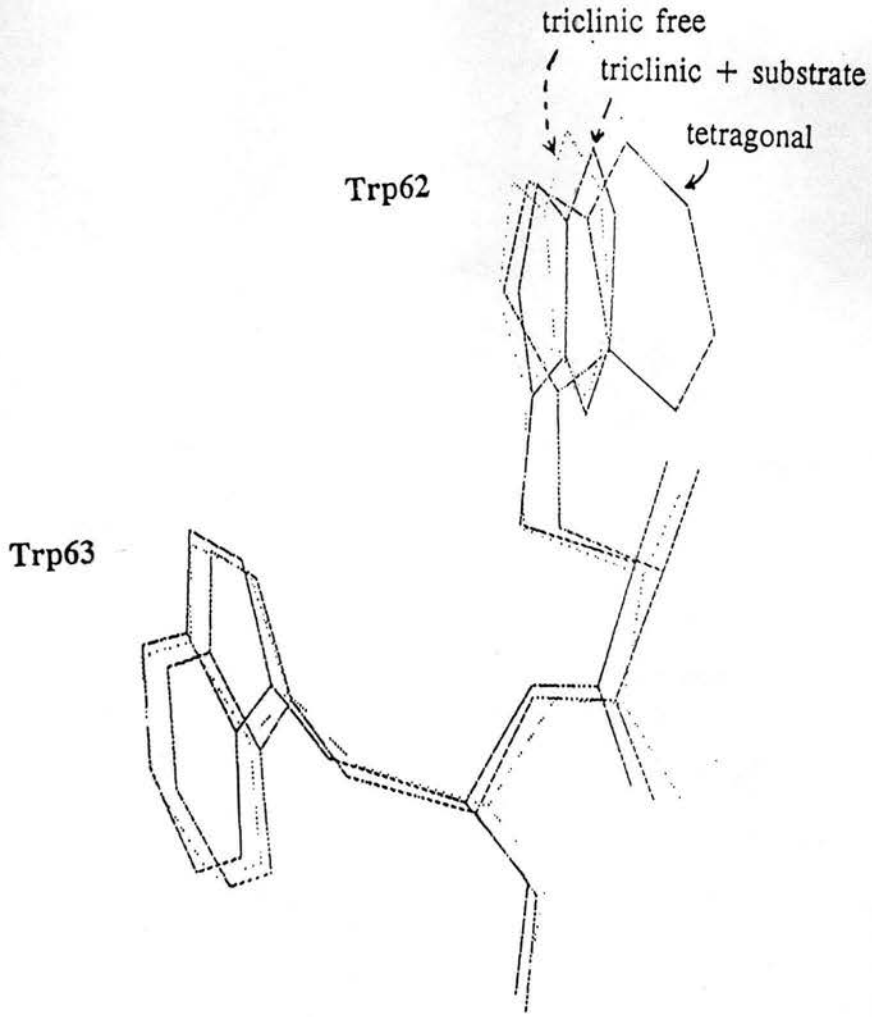


Figure II-30. Superposition of Trp62-Trp63 pair in tetragonal and triclinic HEWL.

(2) CD of HEWL-Fab complexes.

Figure II-31 shows the calculated Trp CD for the B_b bands of free tetragonal lysozyme in comparison with lysozyme complexes with Fabs of HyHEL-5 (h5) and HyHEL-10 (h10) monoclonal antibodies. The CD spectra based on two refinements for tetragonal lysozyme are shown to give an idea of the systematic error of our approach due to the limited accuracy of the x-ray data. The h5 and h10 spectra exhibit couplet CD bands of opposite sign. The h5 complex spectrum is of the same sign as for free HEWL. None of the complexes shows positive features around 212 nm characteristic of free HEWL.

Figure II-32 shows the Trp CD of h5 and h10 complexes after the contribution of the corresponding Fab fragment has been subtracted. Calculations of the Trp CD for free lysozyme and lysozyme complexes with two monoclonal antibodies have shown that formation of protein complexes is readily distinguishable by Trp CD (Fig.II-31), although the difference between h5 and h10 complexes spectra is largely attributable to the Trp pairs in the Fab component. We could not analyze the Trp CD of free Fab fragments since their crystal structure was not available, although the original authors believe that no significant conformational changes occur in Fabs upon association with lysozyme, relying on the overall similarity of h5 and h10 structures with other Fabs and the similar ways in which they associate (Padlan, et al., 1989). The secondary structure of Fabs are essentially similar and consist of four β-sheets. Four out of nine Trp in Fabs, Trp103(106), Trp154(157), Trp34(35) and Trp161(163), where positions in h10 are shown in parenthesis, lie in regions with high (RMS <0.60 Å) structural homology

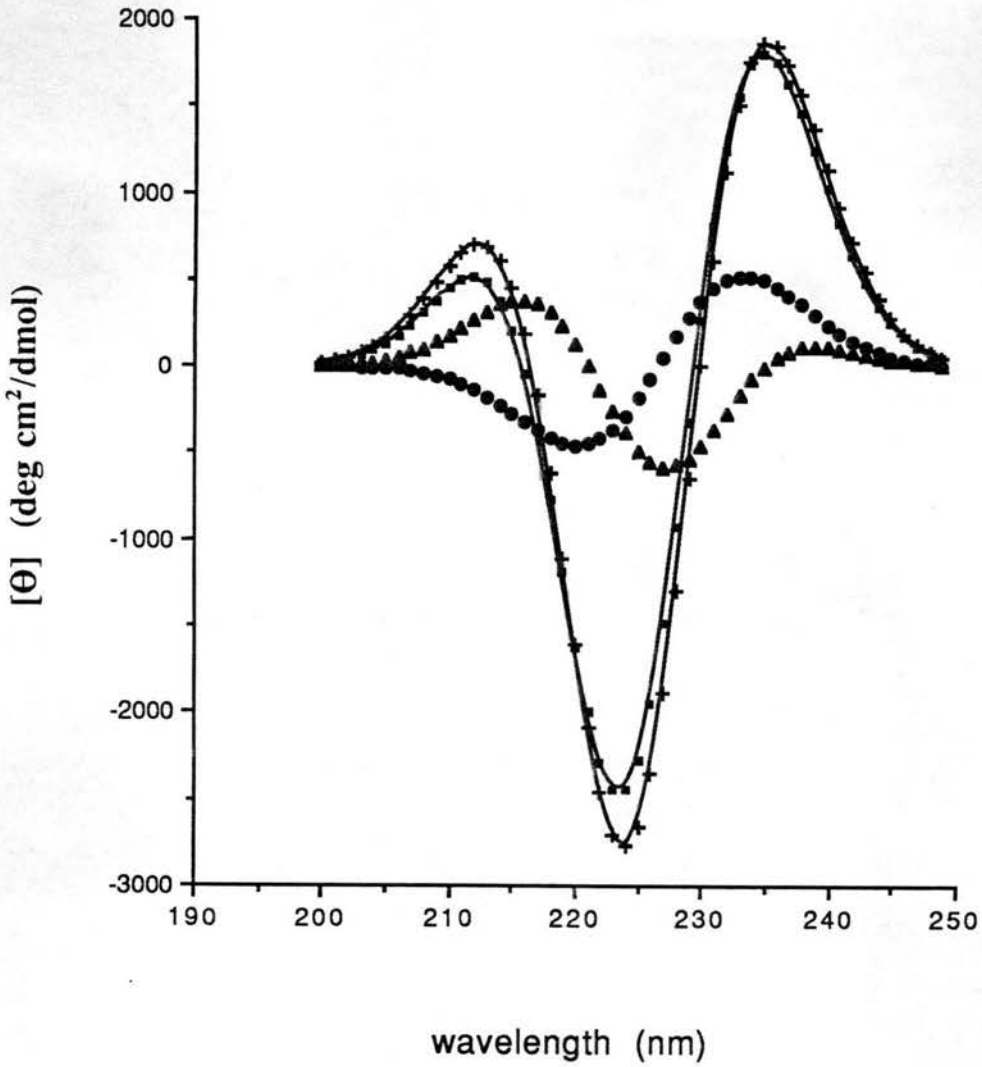


Figure II-31. B_b Trp CD of free tetragonal lysozyme, 2lyz (■) and 6lyz (+), and lysozyme-Fab complexes, h5 (•) and h10 (▲). The free HEWL spectra are shown in [Θ] per 129 lysozyme residues. The CD of the complexes are shown in [Θ] per 129+425=554 residues for the h5 complex, and per 129+429=558 residues for the h10 complex.

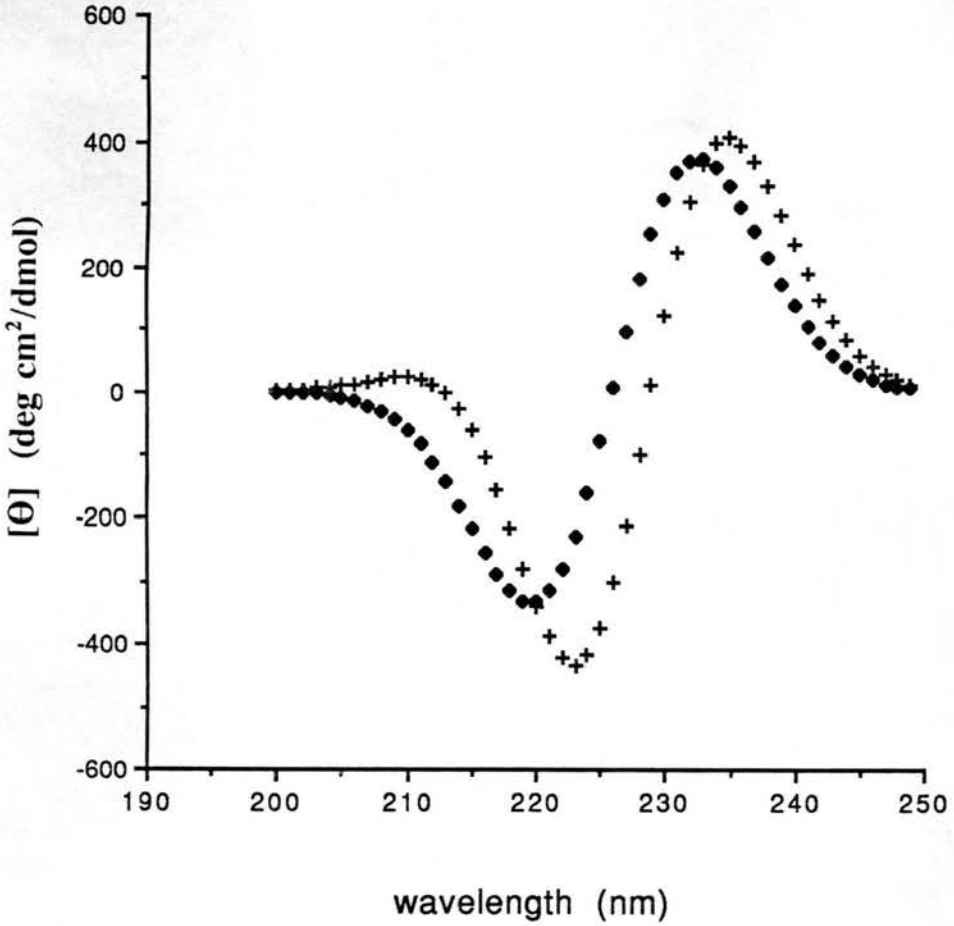


Figure II-32. B_b Trp CD of the lysozyme-Fab complexes, h5 (•) and h10 (+), with Fabs Trp contributions subtracted (see text). CD spectra are calculated per number of residues in each lysozyme-Fab complex, 554 in h5 and 558 in h10.

(Table II-2). As can be seen (Fig.II-31), Fabs h5 and h10 complexes with HEWL have different characteristic aromatic CD although their peptide CD should be very similar due to homologous secondary structure. Tyr residues are also abundant in Fabs, which will be considered in further studies.

After the Fab contribution has been subtracted, the resulting CD reflects contributions of the lysozyme Trp (Figure II-33) and the intermolecular CD due to Trp pairs, one member of which belongs to lysozyme and the other to Fab (Fig.II-34). In the complex with h5 the lysozyme spectrum is essentially the same as that of native enzyme and the intermolecular CD is small. In h10 the amplitude of the lysozyme spectrum has increased, corresponding to a significant intermolecular CD. The differences in intermolecular CD might reflect the geometry of the complexes. The epitopes for h5 and h10 are located at different sites of lysozyme, which to a certain extent determines the orientation of Fab Trps relative to lysozyme Trps. Other sources of difference in intermolecular CD might be the conformational changes in lysozyme and differences in structure and sequence between Fabs.

The most characteristic feature in the Trp CD of free lysozyme (both tetragonal and triclinic) is an extra positive band at 211-212 nm (Fig.II-33), which vanishes upon binding of Fabs. Our analysis of the partial Trp contribution to the exciton bands (Table II-3) shows that the dominance of the Trp28-Trp108 pair in the highest energy exciton band is a common feature in all studied lysozyme structures. The comparison of rotational strengths for this exciton band (around 221 nm) for different lysozyme structures shows distinctive differences between bound and free lysozyme (Fig.II-35).

Table II-2. Analysis of Trp homology between Fab fragments of HyHEL-5 and HyHEL-10 monoclonal antibodies.

	Frame limit\ Trp content			
HyHEL-5	H11-\Trp33,Trp36\-H50*		H102-\Trp103\-H109	HTrp154 -14
HyHEL-10	H11-\Trp34,Trp36\-H50		H102-\Trp106\-H109	HTrp157 -14
RMS	1.03		0.51	0.40
HyHEL-5	LTrp34 ±2	LTrp161 +4,-6	LTrp90 +4	
HyHEL-10	LTrp35±2	LTrp163 +4,-6	LTrp94 -4	
RMS	0.45	0.81	1.26	

*H for residues in Fab heavy chain; L for residues in Fab light chain.

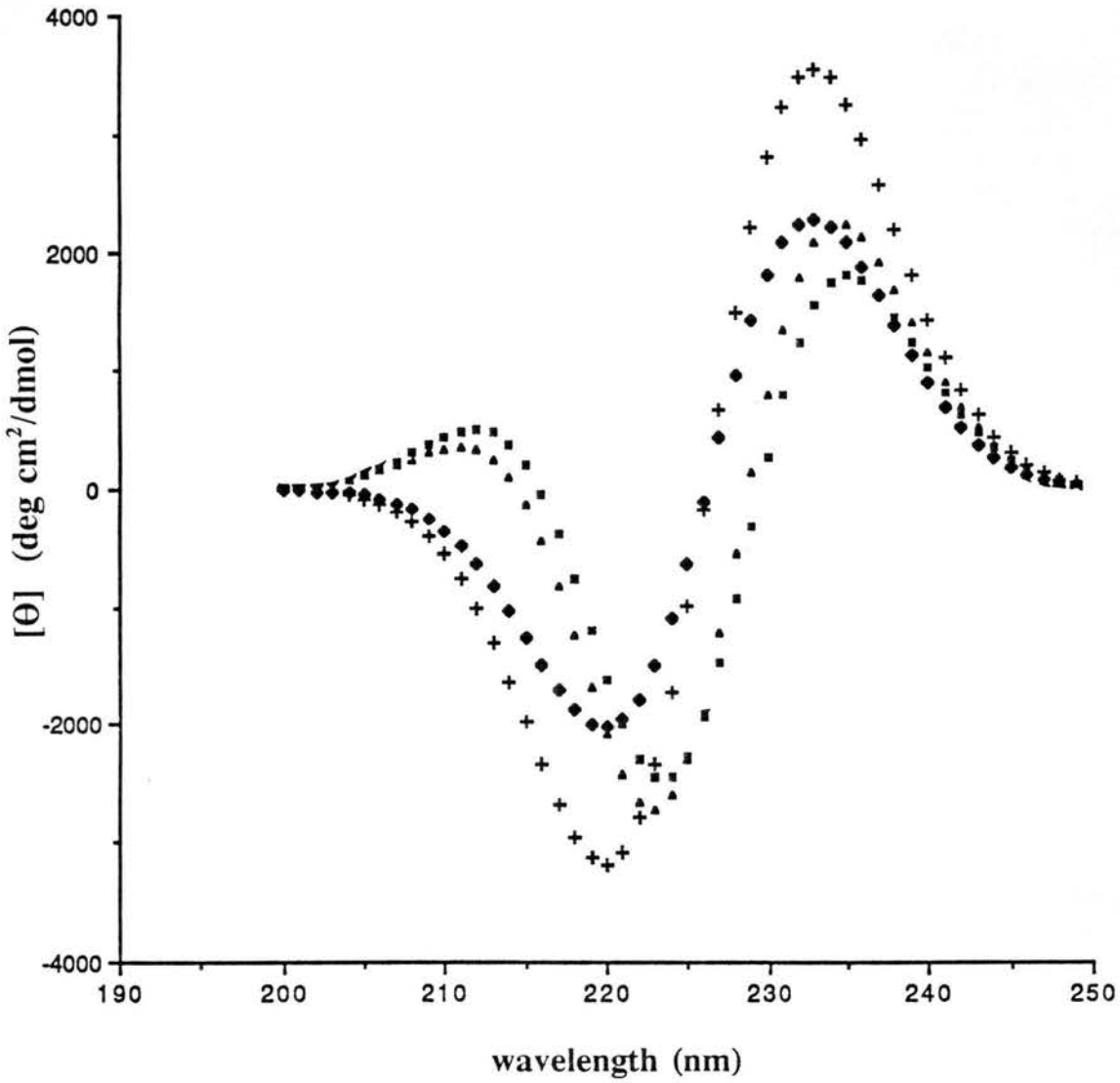


Figure II-33. B_b Trp CD of HEWL: free tetragonal, 2lyz (■); free triclinic (▲); in h5 complex (◆); in h10 complex (+). CD spectra are calculated per number of residues in HEWL (129) both for the free protein and bound in complexes.

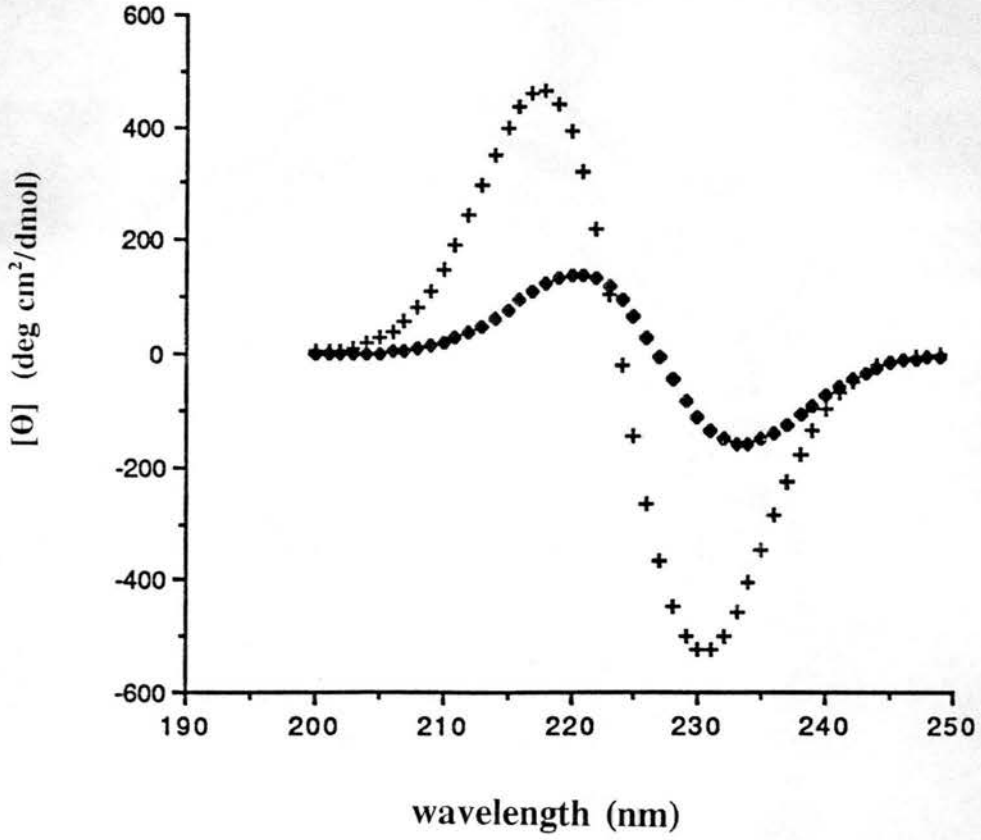


Figure II-34. Intermolecular B_b Trp CD in h5 complex (◆) and h10 complex (+). CD spectra are calculated per number of residues in each lysozyme-Fab complex, 554 in h5 and 558 in h10.

Table.II-3 Partial contributions of individual Trp in various lysozyme structures to the eigenvectors of the B_b exciton bands.

Band number*		Trp:	28	62	63	108	111	123
	a. tetragonal		.43	.02	.06	.46	.05	.00
1	b. triclinic		.31	.10	.17	.36	.05	.00
	c. in h5		.31	.10	.18	.37	.04	.00
	d. in h10		.43	.00	.03	.46	.08	.00
	a.		.06	.49	.44	.02	.00	.00
2	b.		.16	.41	.32	.10	.01	.01
	c.		.16	.40	.32	.09	.09	.03
	d.		.00	.14	.14	.02	.13	.56
	a.		.00	.00	.00	.03	.30	.66
3	b.		.00	.00	.00	.04	.20	.75
	c.		.00	.01	.00	.05	.29	.64
	d.		.04	.61	.11	.01	.07	.16
	a.		.03	.00	.00	.00	.65	.32
4	b.		.04	.00	.00	.00	.73	.23
	c.		.03	.00	.00	.00	.64	.33
	d.		.04	.24	.59	.01	.12	.00
	a.		.02	.49	.49	.00	.00	.00
5	b.		.13	.42	.38	.06	.00	.01
	c.		.19	.39	.33	.08	.00	.00
	d.		.03	.01	.13	.00	.58	.24
	a.		.47	.00	.02	.50	.00	.01
6	b.		.36	.07	.12	.44	.00	.01
	c.		.31	.09	.17	.41	.00	.01
	d.		.46	.00	.01	.50	.01	.02

* Exciton bands are numbered from higher to lower energies, which correspond to wavelengths in Figure II-35.

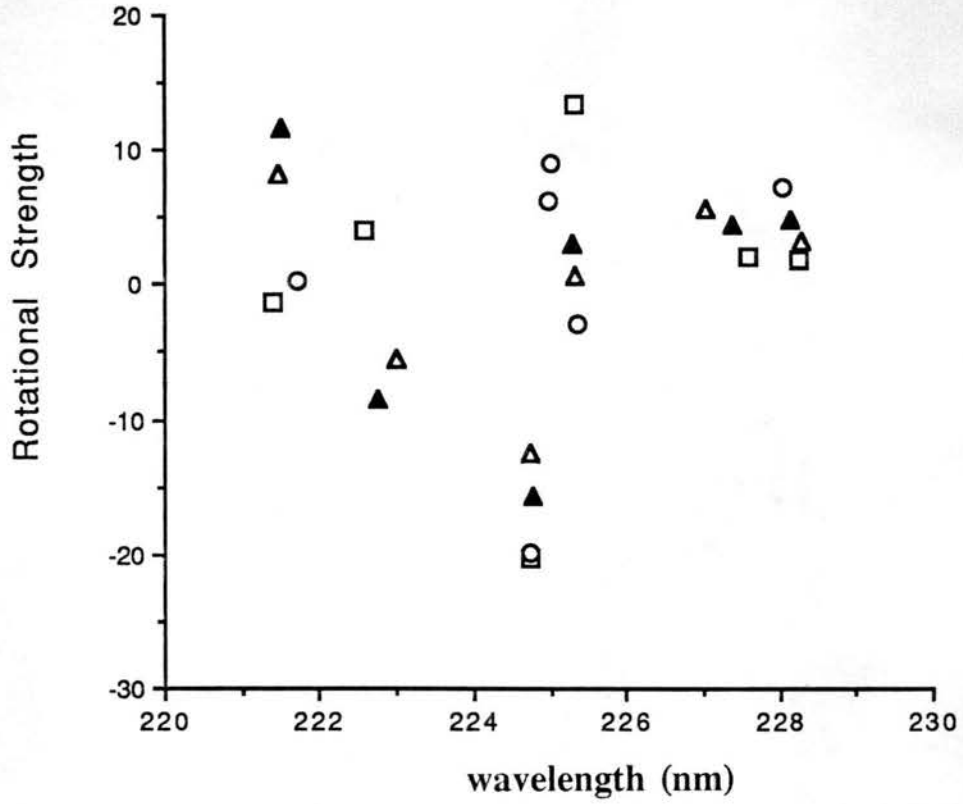


Figure II-35. Positions of B_b Trp exciton bands in calculated CD of HEWL: free tetragonal, 2lyz (■); free triclinic, 2lzt (▲); in h5 complex (□); in h10 complex (o).

The CD of the Trp28-Trp108 pair changes in the same direction in h5 and h10 complexes, which might explain the decrease of positive CD at 211 nm (Fig.II-35, Fig.II-36). The superposition of the Trp residues (α carbons) in tetragonal lysozyme and in the complexes (Fig.II-37) does reveal some torsional changes at Trp28 and Trp108, which are more profound in the h10 complex (Table II-4). The difference between the orientation of lysozyme Trps in h5 and h10 complexes is largest for Trp63 (58°) and is least significant for Trp108 and Trp123.

Estimation of the CD of separate Trp pairs in HEWL has shown that they might reach amplitudes comparable to that of the whole lysozyme Trp spectrum. The CD of particular Trp pairs are different in h5 and h10 complexes. Our results show that conformational changes in lysozyme in the h10 complex are reflected at Trp62 and Trp108 (Fig.II-36, Fig.II-38d). The CD contribution of Trp63 is mainly conserved upon binding to h10 but changes significantly in the h5 complex (Fig.II-36, Fig.II-38b,c). In the h5 complex, changes in lysozyme CD occur mostly at the Trp62-Trp63 and Trp63-Trp108 pairs, which largely cancel each other in the total lysozyme spectrum. We might assume that though conformational changes in the h5 complex might occur in all the three Trps, the most likely to account for the predicted CD change is a change in conformation of Trp63. This coincides with large conformational changes reported at Trp63, which undergoes a 180 degree rotation around the C_β-C_γ bond (Fig.II-37; Sheriff et al., 1987). In the h5 complex, the Fab binds far from the active site cleft. Thus rotation of the Trp63 ring gives evidence of long-range conformational changes. In h10, significant changes occur at the Trp62-Trp108 pair, which coincides in sign with the total

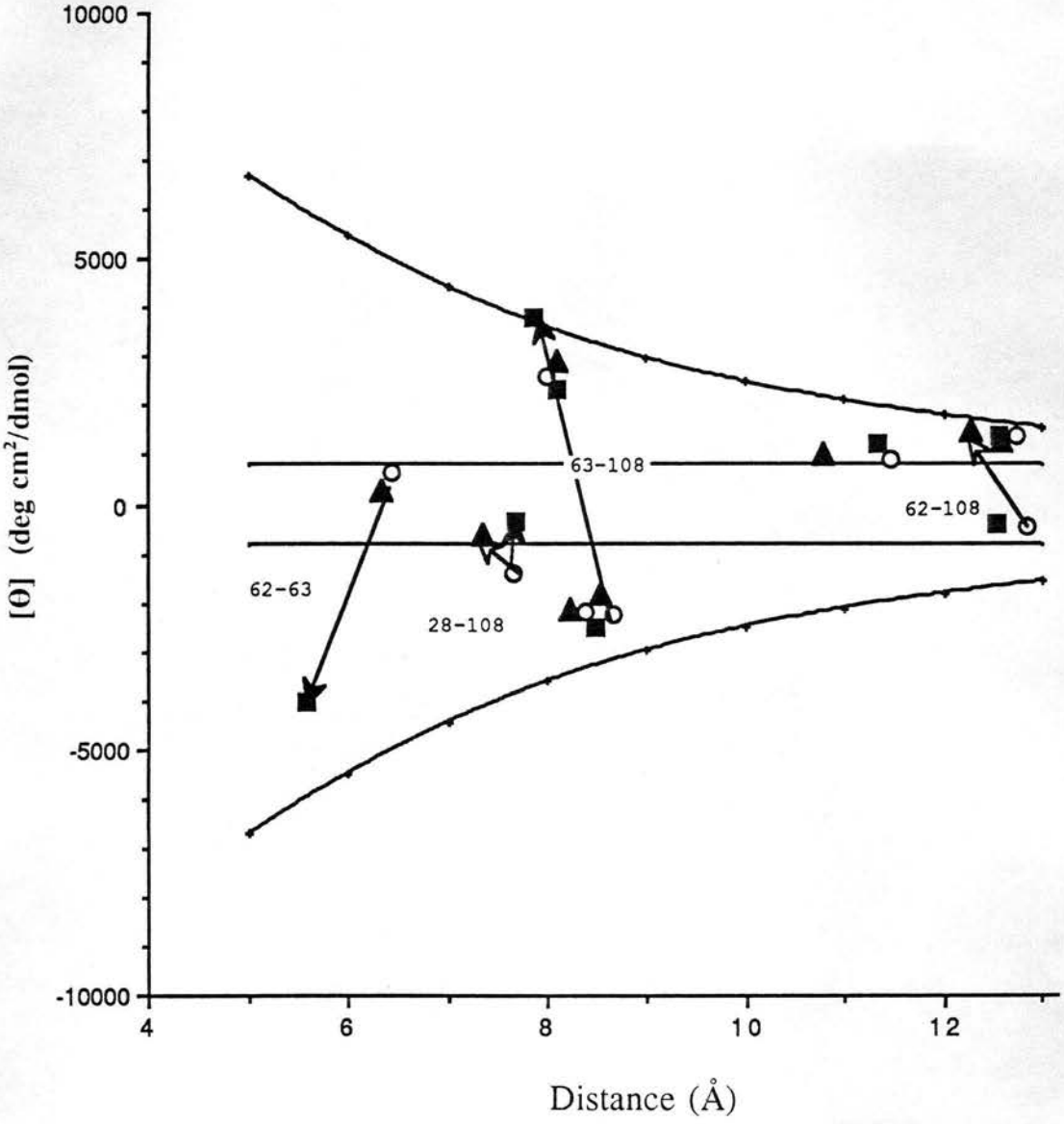


Figure II-36. B_b CD of Trp pairs at 231 nm in HEWL: free tetragonal, 2lyz (o); in h5 complex (■); in h10 complex (▲). Changes in CD for the individual pairs between free lysozyme and in h5 (-▸) and h10 (-▷) complexes are indicated with arrows, and the numbers of Trps in those pairs are shown.

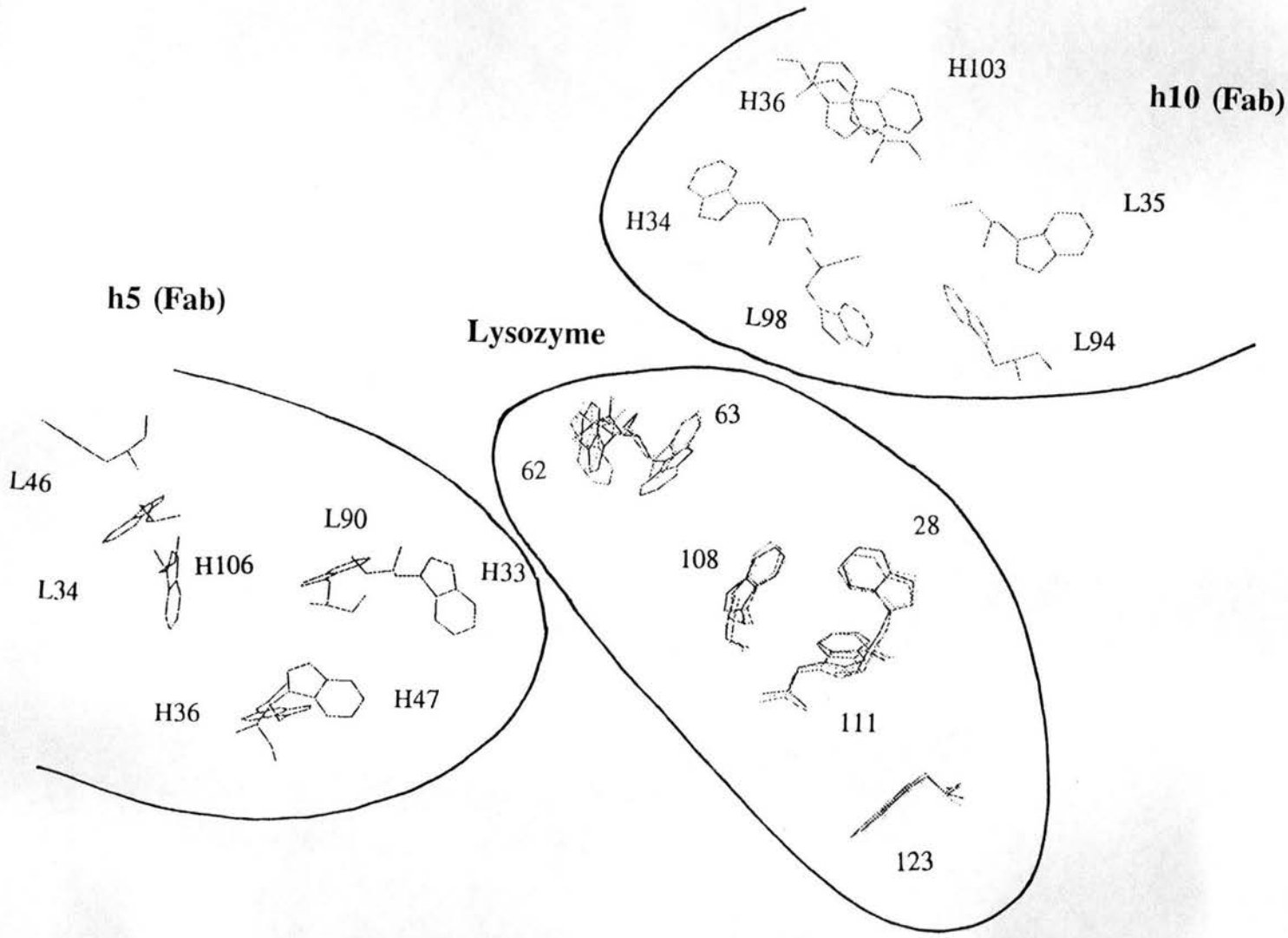


Figure II-37. Trp residues in HEWL, h5 and h10 lysozyme-Fab complexes

Table II-4. Conformational changes in lysozyme Trp in h5 and h10 complexes.

	Change in dihedral angle*					
	CD1-CZ3			CE3-CZ2		
Trp	free/h5	free/h10	h5/h10	free/h5	free/h10	h5/h10
Y28	4.2	5.5	5.1	5.1	**-24.4**	**23.0**
Y62	-3.1	1.2	**-21.2**	0.0	4.6	5.1
Y63	**62.3**	1.8	**58.1**	-8.3	-3.8	-4.3
Y108	-5.5	-3.9	3.2	-8.8	**17.0**	2.0
Y111	-2.8	3.5	1.3	**27.9**	**15.0**	**11.3**
Y123	-5.7	-1.7	4.4	5.7	**13.7**	5.8

* Lysozyme structures were superimposed through C_α atoms of corresponding Trp residues. The measured angle is the angle between the projections of CD1->CZ3 (CE3->CZ2) vectors in corresponding Trps on the plane perpendicular to the line connecting CD1 (CE3) atoms in corresponding Trps.

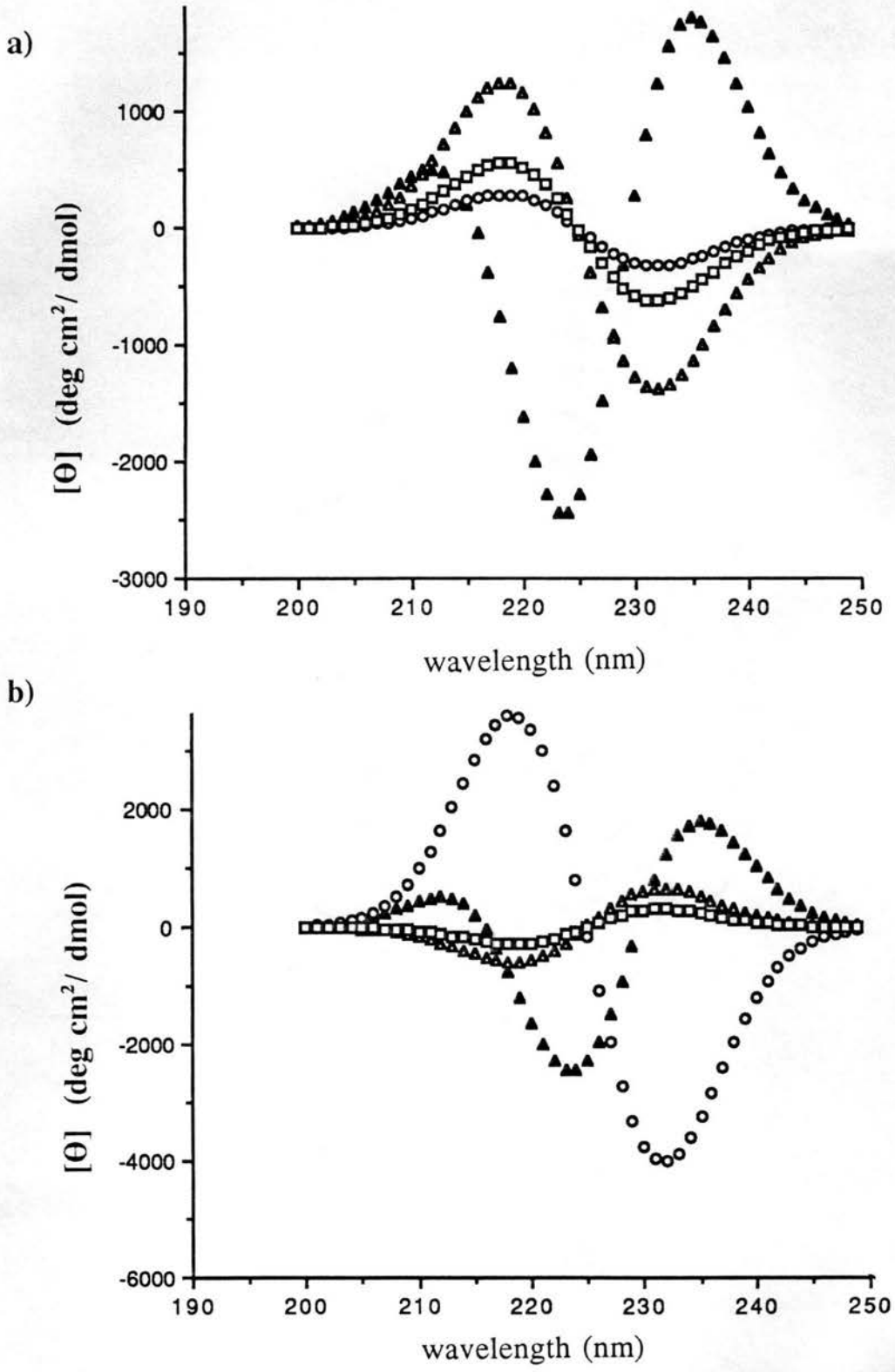


Figure II-38 B_b CD of lysozyme Trp pairs (△) in free lysozyme (▲), h5 (o) and h10 (□) complexes: a) Trp28-Trp108; b) Trp62-Trp63.

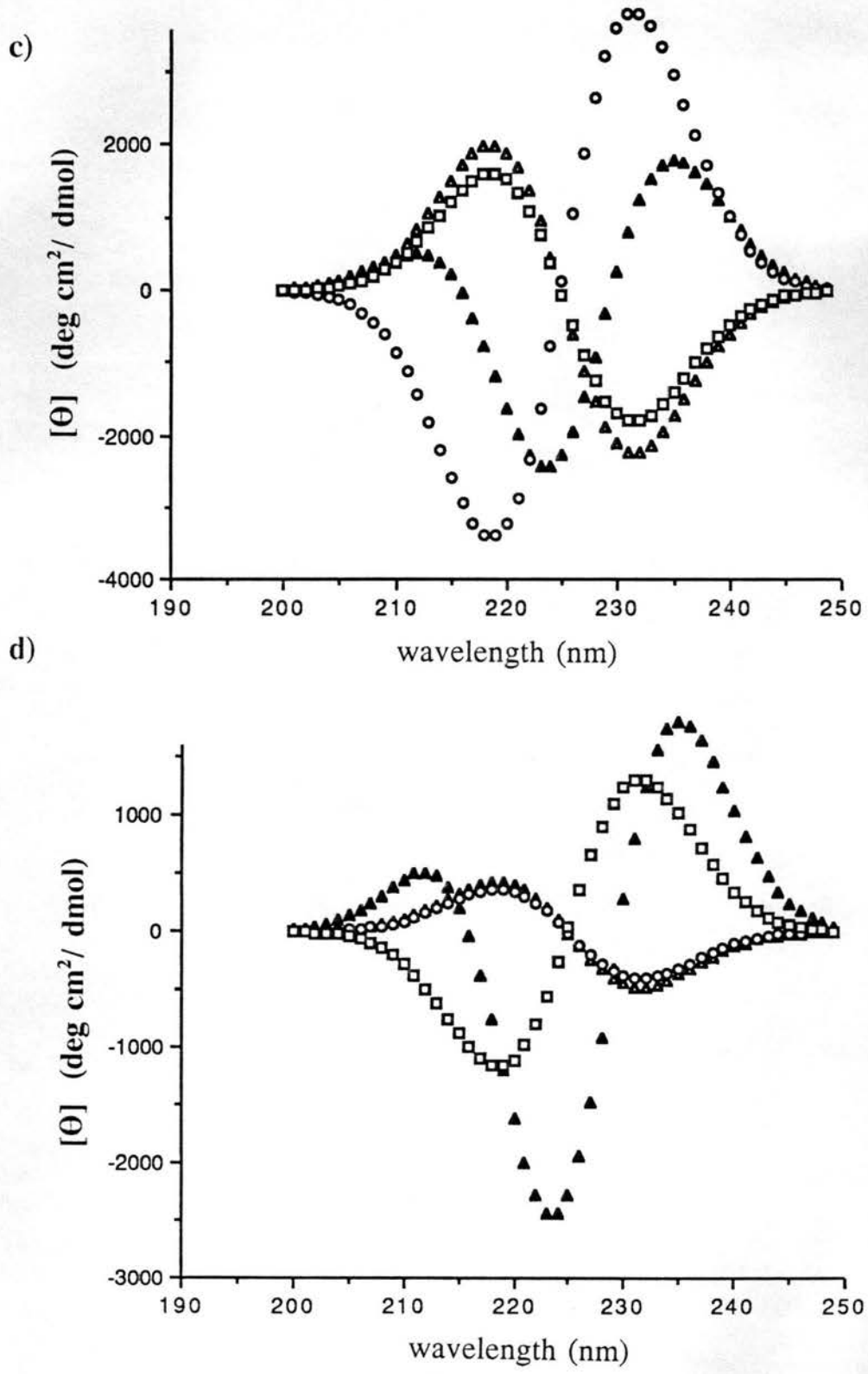


Figure II-38 (continued). B_b CD of lysozyme Trp pairs (△) in free lysozyme (▲), h5 (o) and h10 (□) complexes: c) Trp63-Trp108; d) Trp62-Trp108.

lysozyme Trp spectrum (Fig.II-38d). This pair might be responsible for the most of the amplitude increase (1.6 times) in lysozyme in the h10 complex. Rotation of the Trp62 indole ring by 150 degrees around the C_{β}-C_{γ} bond has been reported by the original authors (Fig.II-37 and Padlan et al., 1989) as the largest and unique conformational change among lysozyme side-chain residues upon binding to Fab h10. The CD of individual Trp pairs reaches considerable values and might be experimentally detectable. Both Trp62 and Trp63 lie inside the active site cleft and are involved in substrate binding (Phillips, 1966). Thus CD estimation and detection of conformational changes in this region might be useful in monitoring the enzyme function.

Figure II-37 shows the superposition of Trp residues in free tetragonal lysozyme and lysozyme in h5 and h10 complexes. Three Fab Trp residues are close to the lysozyme surface in the h5 complex. These are TrpH47, which belongs to the heavy chain framework, TrpL90 and TrpH33. The epitope of h10 crosses the active site cleft and includes Trp63. In h10, several Fab Tyr are in contact with the lysozyme surface and three Trp are close to it: TrpH98, TrpL94 and TrpH34. Seven out of nine Trp residues have similar positions in the two Fabs (Table II-2).

Figure II-39 shows partial contributions of lysozyme Trps to the interprotein difference spectrum between h5 and h10 complexes. Table II-5 describes interprotein Trp pairs that contribute 30 % and more to intermolecular CD. Differences between CD of h5 and h10 complexes result from several Trp-Trp interactions (Table II-5). For instance, TrpH47 in h5 interacts with Trp62 and Trp108. TrpL46 in h5 contributes to the CD of Trp28. In all cases (Table II-5), differences in CD between interprotein Trp pairs are

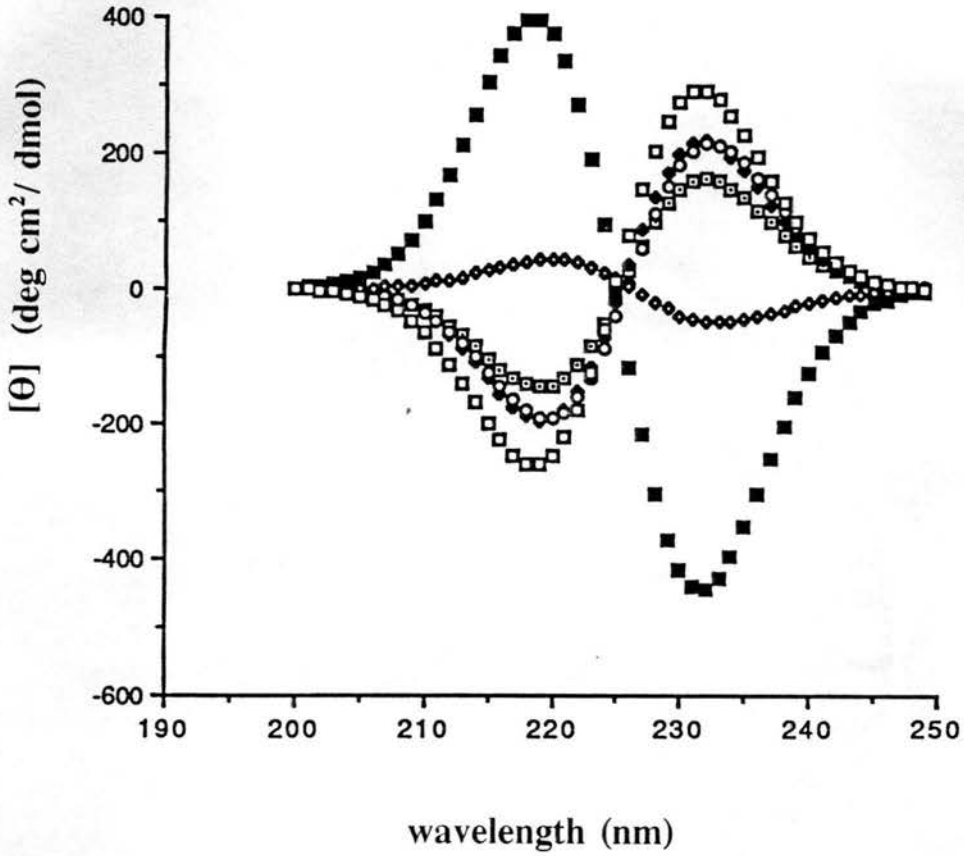


Figure II-39. Partial contribution of lysozyme Trps to the interprotein difference CD spectrum between h5 and h10 complexes: Trp28 (⊡); Trp62 (♦); Trp63 (o); Trp108 (◇); Trp111(■); Trp123 (□). See text.

Table II-5. Conformational and spectroscopic properties of Trp pairs in h5 and h10 complexes.

Pair	CD at 219 nm	Dihedral angles*		Distance (Å)
		CD1-CH2	CE3-CZ2	
(h5) Y108-H33	196	142.5	-59.0	19.6
(h10) Y108-H34	90	90.2	60.0	21.8
(h5) Y108-H47	76	-177.2	-8.4	25.1
(h0) Y108-H98	-18	-24.7	-96.3	15.1
(h5) Y62-H47	-130	-147.9	-176.7	22.5
(h10) Y62-H98	159	-74.0	145.7	16.3
(h5) Y62-L90	171	119.3	-178.1	17.7
(h10) Y62-L94	-42	145.8	-12.7	22.3
(h5) Y63-H33	67	-76.7	-44.4	17.7
(h10) Y63-H34	168	78.2	-103.7	14.7
(h5) Y111-L46	9	160.3	-89.4	45.7
(h10) Y111-L94	-163	-114.5	-116.5	20.3
(h5) Y111-H47	-26	-110.2	-57.4	30.9
(h10) Y111-H98	-135	34.2	42.4	19.2
(h5) Y28-L46	23	107.4	-0.5	46.2
(h10) Y28-L94	79	-47.6	-79.3	17.9
(h5) Y28-H33	-12	-156.4	-32.8	25.4
(h10) Y28-H34	-80	-94.7	-145.4	24.0

* The measured angle is the angle between the projections of CD1->CH2 (CE3->CZ2) vectors in corresponding Trps on the plane perpendicular to the line connecting CD1 (CE3) atoms in corresponding Trps.

caused by large differences in distance, 10 Å and more. The most important conserved TrpH33 in h5 (TrpH34 in h10) shows the largest exciton coupling with lysozyme Trp28, Trp63 and Trp108. The indole ring of Trp63 is nearly perpendicular to that of TrpH33 (TrpH34) in both complexes (Table II-5), despite conformational changes in lysozyme structure (Fig. II-37; Table II-4). Conformational changes in TrpH33 (TrpH34)-Trp28 and TrpH33 (TrpH34)-Trp108 pairs occur both at lysozyme and Fab Trps. TrpL90 in h5 (TrpL94 in h10) contributes considerably to the intermolecular CD of Trp62. CD spectra of TrpL90 (TrpL94)-Trp62 pairs are of opposite sign in the two complexes.

Trp111 contributes the strongest negative couplet to the difference interprotein CD between h10 and h5 complexes (Figure II-39) which is largely due to the strong negative contributions in the h10 complex from Trp111-TrpL94 and Trp111-TrpH98 pairs. The main interactions of Trp111 in h5 complex, with TrpL46 and TrpH47, are weak (Table II-5). The sign of the Trp111 difference spectrum is opposite to that of the total difference between interprotein CD of the complexes (Fig.II-34, Table II-5).

(3) Crystals Under pressure.

Hydrostatic pressure of 1000 atm (Kundrot & Richards, 1987; 1988) is predicted to result in CD differences only in the B_b region, while the B_a and higher energy transitions match closely (Fig.II-40a). The short-wavelength positive band, which is at 211-212 nm in tetragonal and triclinic lysozyme, almost disappears in crystal structures grown under pressure (Fig.II-40b). The Trp B_b CD of the HEWL grown under 1000 atm was predicted to be about 1000 deg cm^2/dmol smaller than for the crystals grown at 1atm. The largest deviation in the couplet strength among individual Trp pairs was found

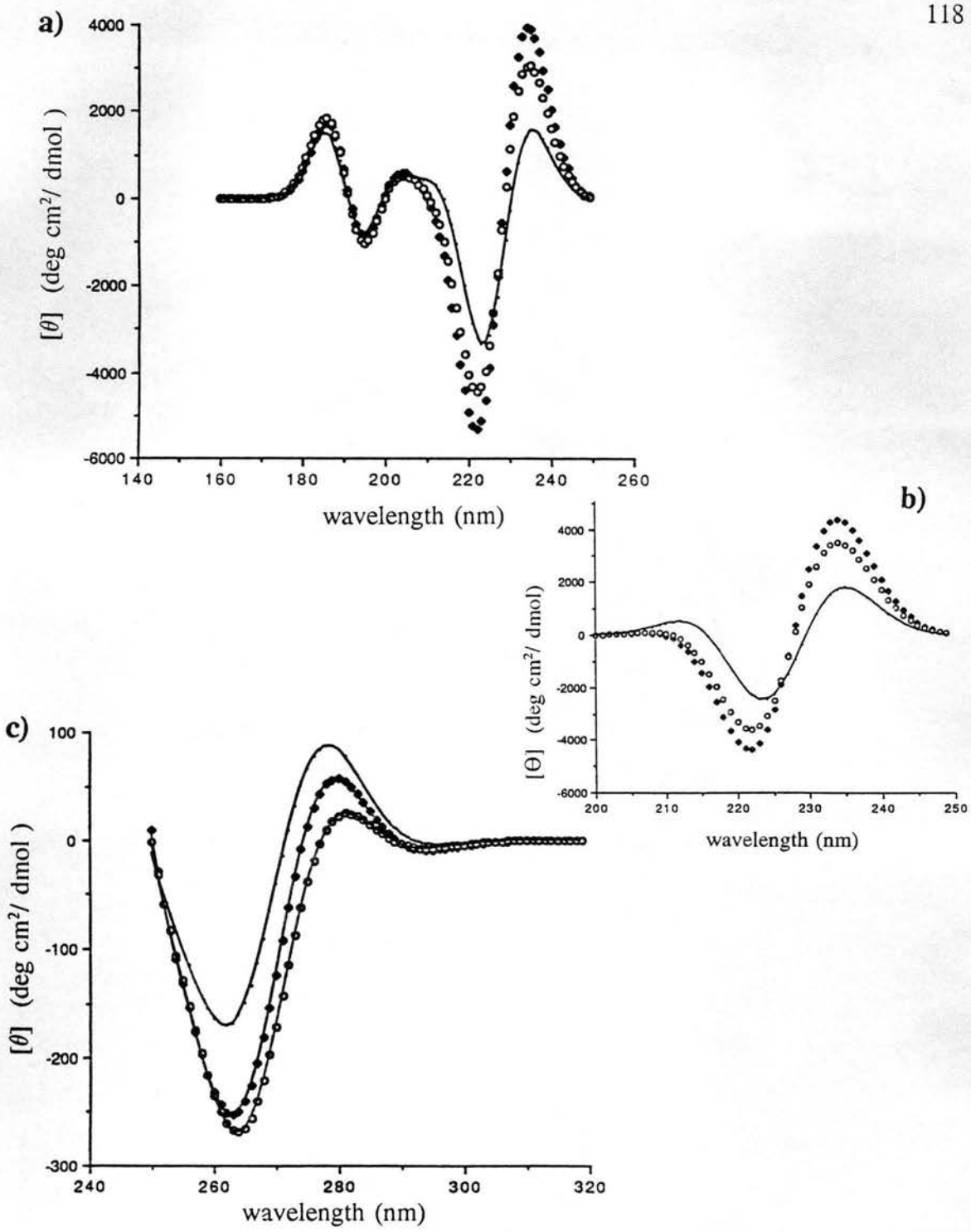


Figure II-40. Trp CD calculated for crystals of HEWL grown under pressure of : 1atm (◆); 1000 atm (o); atmospheric pressure, 2lyz (-). Spectra in the far uv calculated considering all six transitions (a) and B_b transition only (b), and in the near uv considering all six transitions (c).

for the Trp28-Trp108 pair. Figure II-41 does reveal some torsional differences at Trp28. The discrepancy between the CD of the crystals grown under 1 atm (Kundrot & Richards, 1988) and that reported originally for tetragonal lysozyme (Diamond, 1975) might reflect the difference in ionic conditions, which were 1M NaCl (Phillips et al., 1967) compared to 1.4M NaCl (Kundrot & Richards, 1988).

(4) Inactivated Lysozyme.

HEWL can be chemically modified by the addition of iodine. Iodine promotes covalent linkage between CD1 of Tyr108 and OE2 of Glu35. Thus Tyr108 is present in iodine-inactivated HEWL in an acylated form.

In our calculations the indole-ester formed at Trp108 (Beddell et al., 1975) was considered as an indole chromophore. The reason for this consideration is that the ester bond was found to lie approximately in the same plane as the indole ring (dihedral angle NE1Trp108-CD1Trp108-OE2Glu35-CDGlu35 equals 12° and CD1Trp108-OE2Glu35-CDGlu35-OE1Glu35 equals 13.6°). This makes the case similar to that of O-acetyl-Tyr, for which the absorption spectrum is closer to that of benzene than to the parent Tyr (Simpson et al., 1963), since the hydroxyl oxygen π-electrons are attracted towards the acetyl group, which practically abolishes their interaction with the aromatic ring. Figure II-42a shows the predicted Trp CD of iodine-inactivated lysozyme in the far-uv region when six transitions are taken into consideration and Figure II-42b shows the B_b Trp CD. The Trp CD of iodine-inactivated HEWL shows considerable deviation from the tetragonal form both in far- and near-uv regions (Fig.II-42). The B_b Trp CD of iodine-inactivated enzyme does not exhibit the positive band around 210 nm as in regular

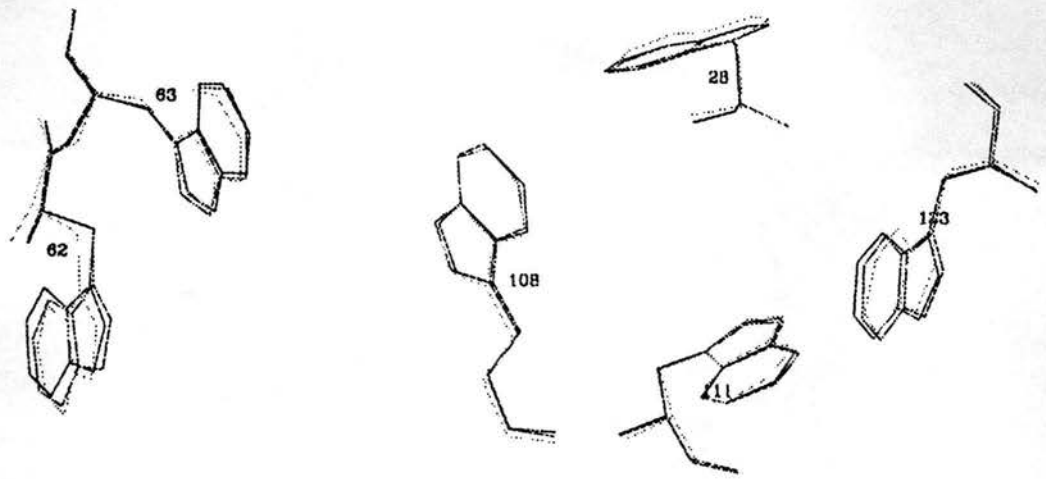


Figure II-41. Conformations of Trp residues in original tetragonal HEWL, 2lyz (dark line); crystals grown under 1 atm (dark line between tetragonal and 1000 atm crystal conformation); and crystalls grown under 1000 atm.

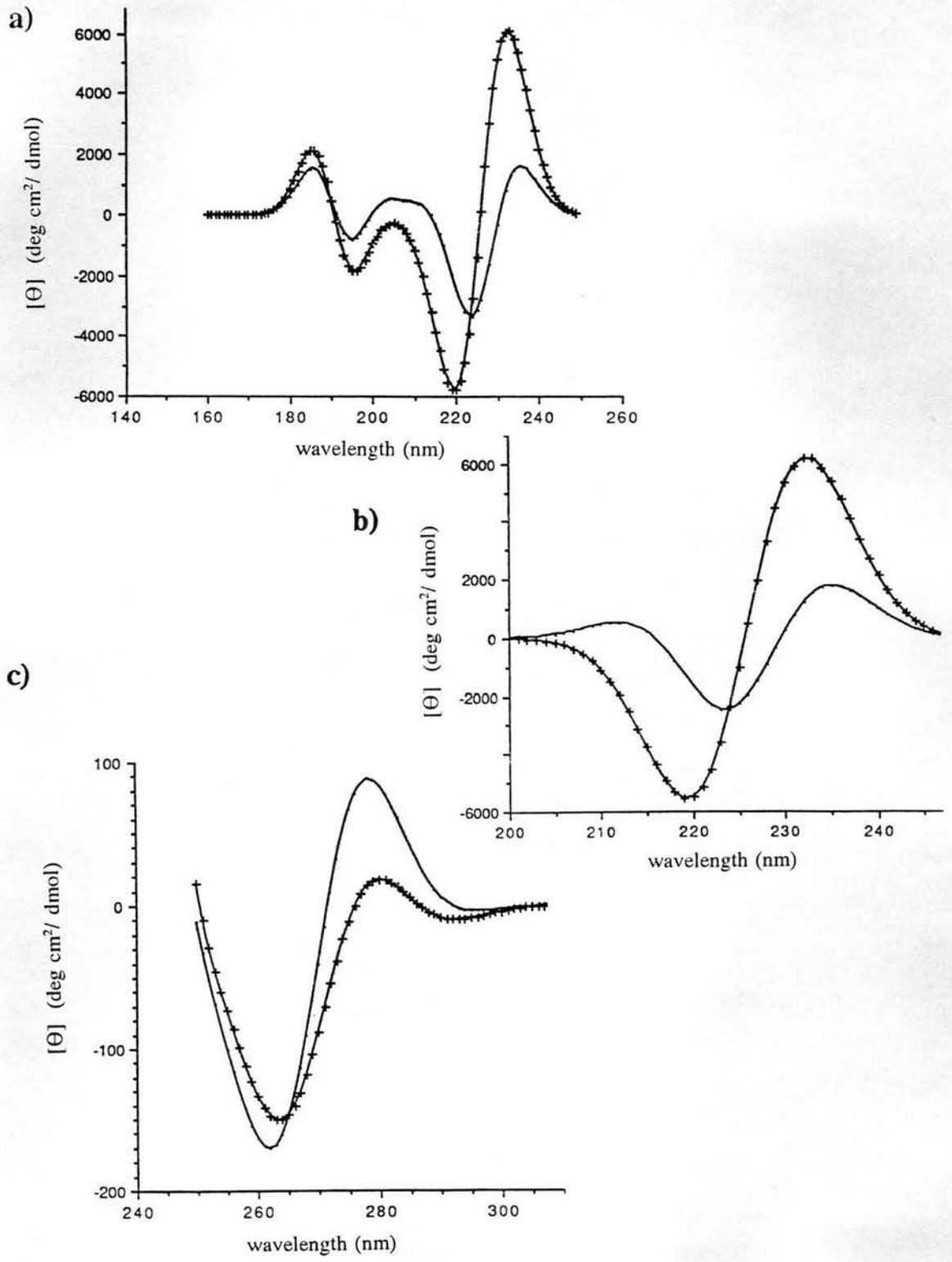


Figure II-42. Trp CD of iodine-inactivated HEWL (+) compared with tetragonal lysozyme (-). Spectra in the far uv calculated considering all six transitions (a) and B_b transition only (b), and in the near uv considering all six transitions (c).

tetragonal crystals (Fig.II-42b). Its positive couplet reaches an amplitude about three times higher and it is blue-shifted about 5 nm. This shift might also be responsible for the increase of the negative band in the B_a region (190-200 nm). Analysis of the couplet strength values reveals that changes in B_b Trp CD occur only in pairs involving Trp108 (Table II-6). It is reasonable to assume that differences in CD calculated accounting for six transitions (Fig.II-42a) occur in the same pairs. Analysis of the difference CD revealed that the major contribution to differences in the B_b region results from the Trp62-Trp108 (12.8 Å in native and 12.6 Å in inactivated enzyme), Trp28-Trp108 (7.7Å and 8.2Å) and Trp63-Trp108 CD (8.7Å and 8.5 Å), but differences in the B_a and higher energy transitions arise mainly from the Trp63-Trp108 pair (Fig.II-43).

In the near-uv CD of the inactivated enzyme the L_b transition is the weakest among the crystal forms studied. The L_a band is close to the tetragonal form in amplitude, although it is red-shifted by 2 nm (Fig.II-42c). Differences between native and inactivated lysozyme in the L_b band are most probably due to Trp62-Trp108 and Trp63-Trp108 pairs (Fig.II-44). The blue shift in the L_a band in native HEWL (Fig.II-42) most probably results from the overlap with a large positive contribution to the L_b band. The near-uv CD of other Trp pairs does not change significatly.

I compared Trp conformations in native and inactivated HEWL by superimposing the Trp C_α atoms. The only difference was localized at Trp108 (Fig.II-45). According to the analysis of the iodine-inactivated enzyme by the original authors (Beddell et al., 1975), Trp108 forms a covalent bond through CD1 to OE2 of the Glu35 carboxyl group. That explains our finding that all the changes in Trp CD occur at pairs involving Trp108.

Table.II-6. Trp pairs in native and iodine-inactivated tetragonal HEWL. Trp pairs where significant changes are observed are shown in bold.

Trp pair	Couplet strength		Angle difference*
	native HEWL	inactivated	
28-62	22	23	
28-63	7	7	
28-108	**114**	**53**	**5°**
28-111	164	166	
28-123	106	107	
62-63	51	51	
62-108	**36**	**65**	**39°**
62-111	22	22	
62-123	4	4	
63-108	**169**	**27**	**10°**
63-111	76	75	
63-123	5	5	
108-111	**198**	**214**	**4°**
108-123	**47**	**32**	**5°**
111-123	71	70	

* Difference in the angle between B_b transition dipole moments of Trps in a particular pair is calculated between native (tetragonal) HEWL and iodine-inactivated HEWL.

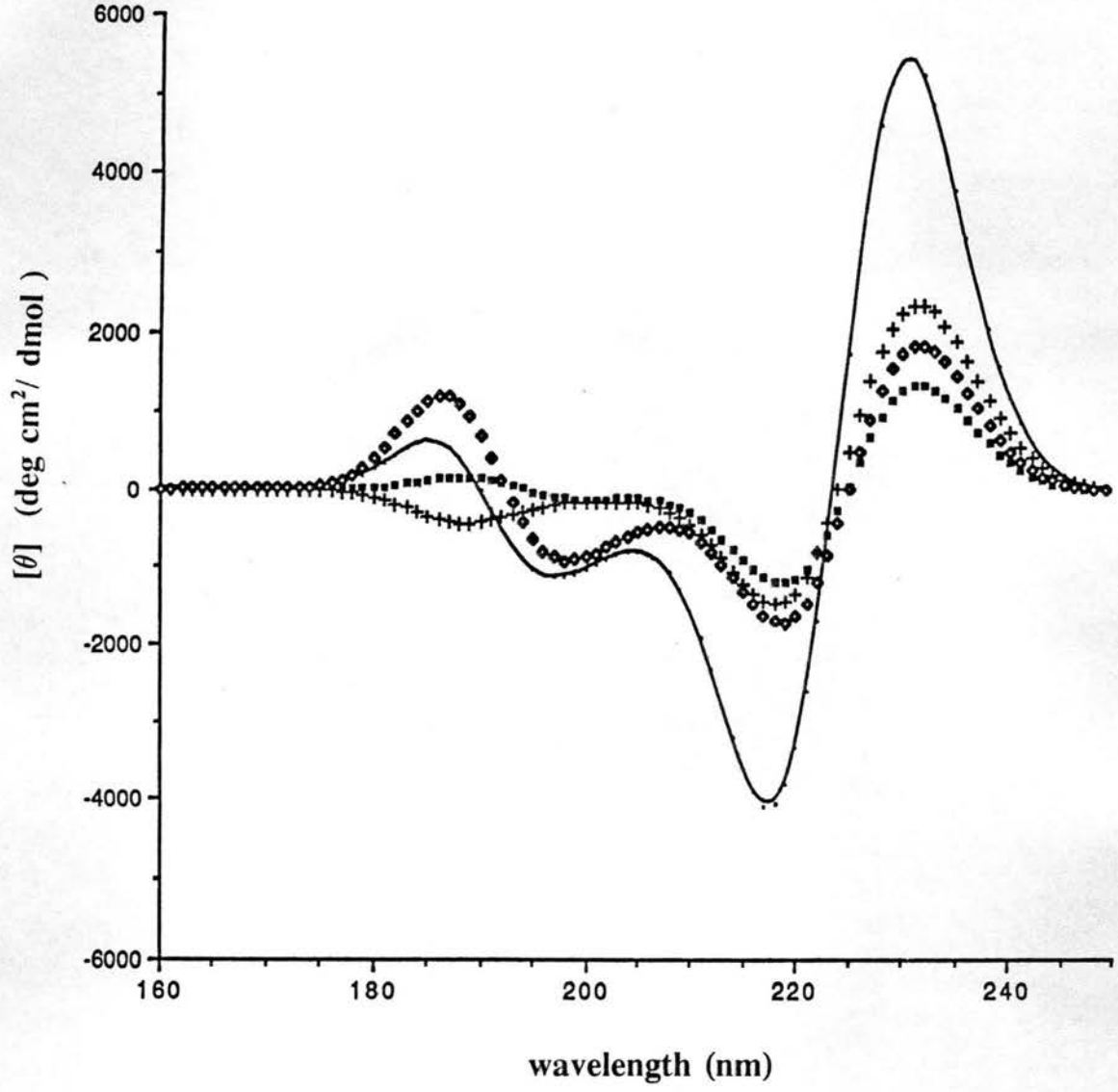


Figure II-43. Differential CD between native and inactivated HEWL (-), and difference in CD of the major pairs: Trp28-Trp108 (+); Trp62-Trp108 (■); Trp63-Trp108 (◇); calculated in the far uv considering all six transitions.

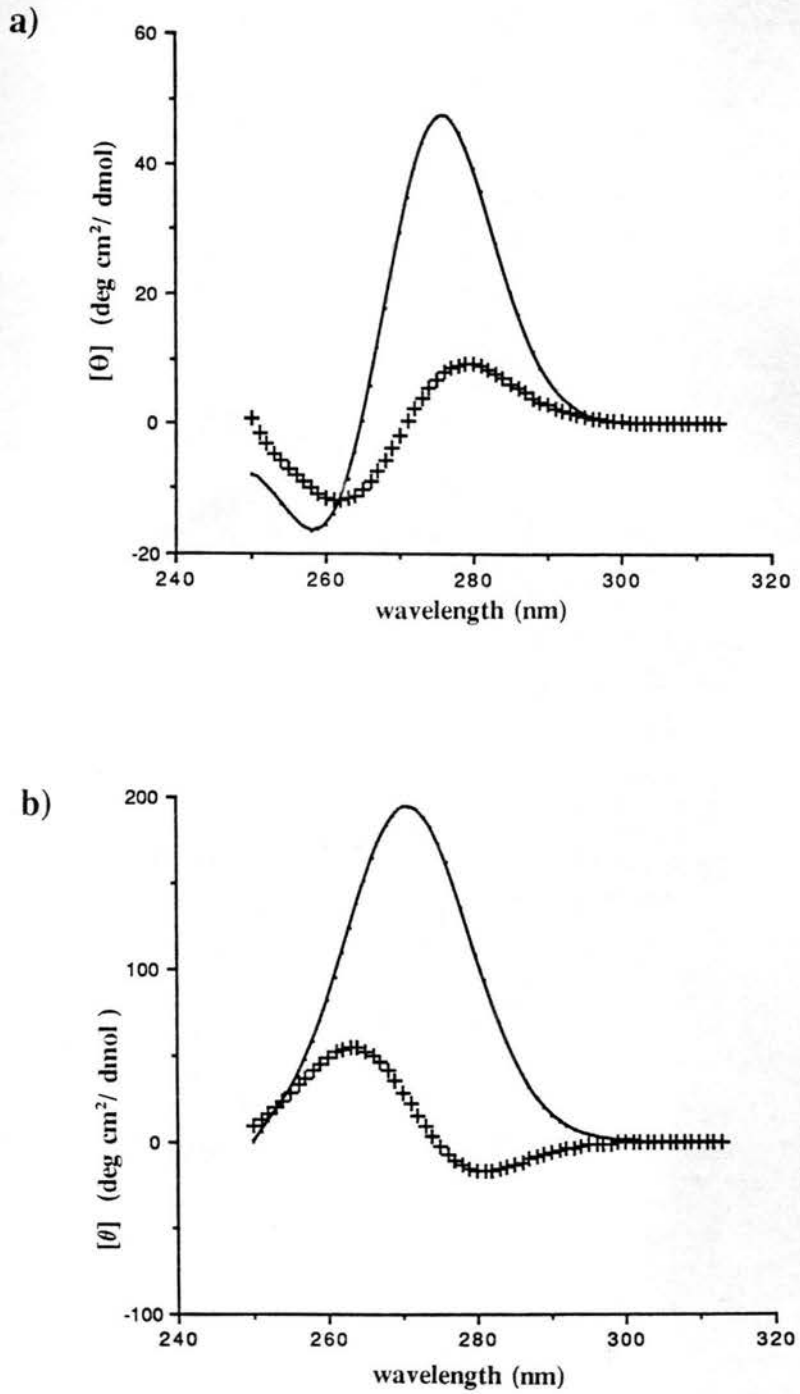


Figure II-44. CD of Trp pairs, Trp63-Trp108 (a) and Trp62-Trp108 (b), in native (-) and iodine inactivated (+) HEWL in the near uv, which contribute most of the difference between the native and iodine inactivated HEWL CD around 280nm.

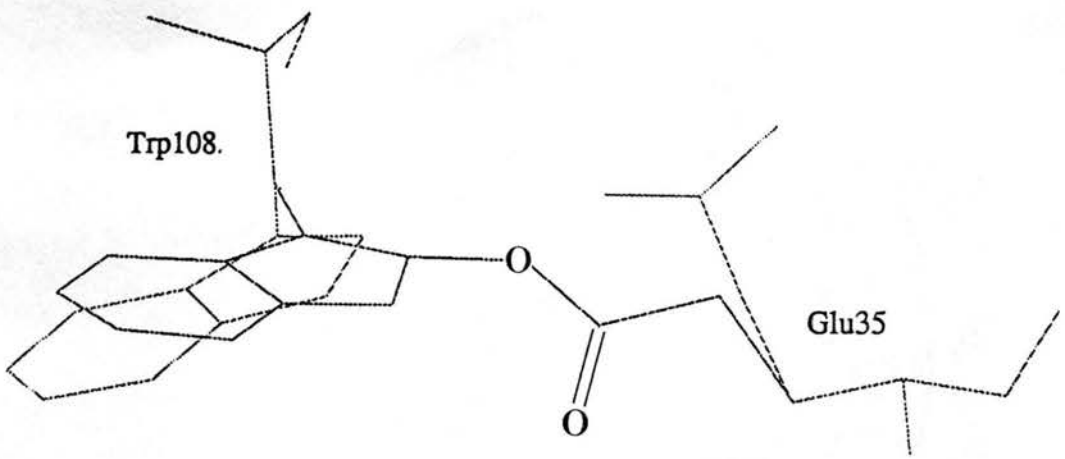


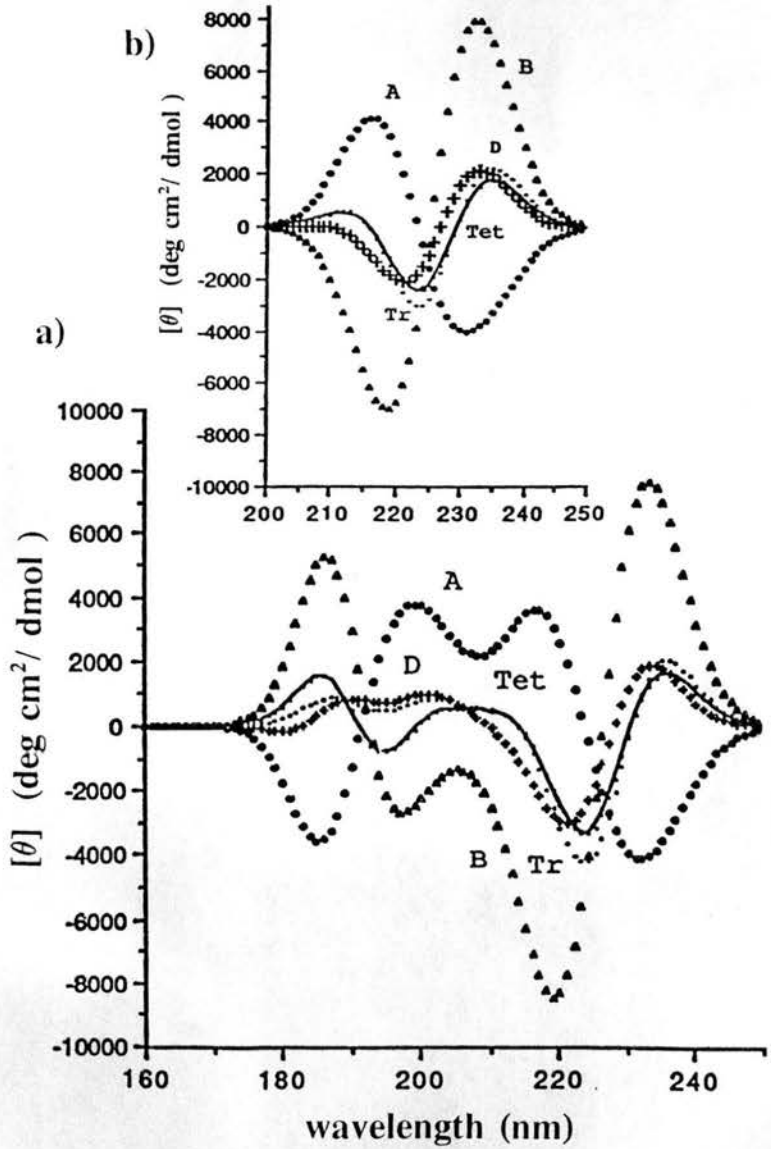
Figure II-45. Conformation of Trp108 and Glu35 in native HEWL (side chains separated) and iodine-inanctivated HEWL (side chains covalently bonded).

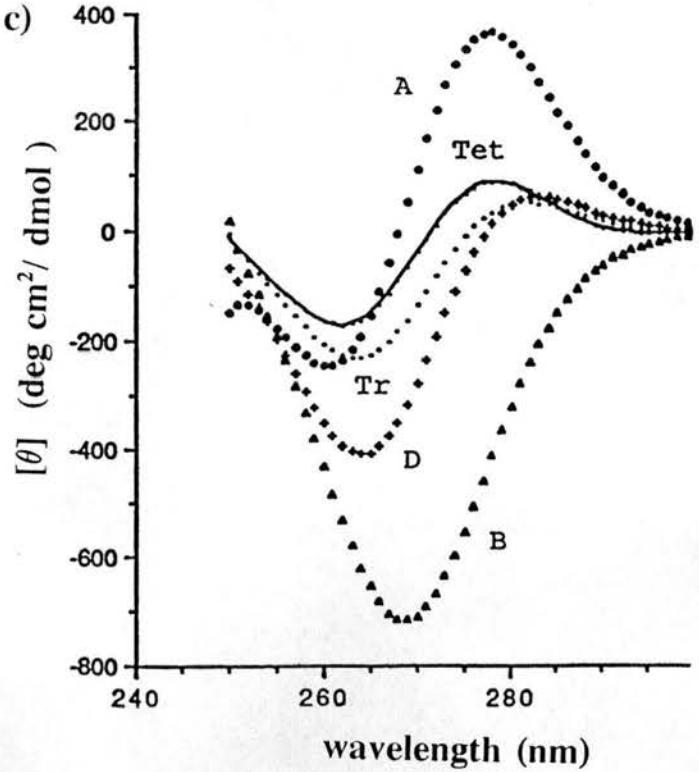
The lack of changes in Trp CD of the Trp108-Trp123 pair might be attributed to its relatively large distance (15Å), though this is not the case for Trp108-Trp111. Comparison of the angles between the directions of transition dipole moments in those Trp pairs (Table II-6) reveal that changes in CD of particular pairs are related to the torsionally induced movement of Trp rings rather than linear movement. As can be seen, considerable changes in CD occur when the torsion is more the 10°.

(5) Monoclinic Lysozyme.

Monoclinic and triclinic crystal forms of HEWL are similar in compactness (Hogle et al., 1981). Monoclinic crystals contain two lysozyme molecules per asymmetric unit. The B_b Trp CD of monoclinic dimer is close to that of tetragonal lysozyme and lysozyme crystals grown under pressure (Fig.II-27). The short-wavelength positive band in the CD of monoclinic dimer is blue-shifted relative to the CD predicted for tetragonal and triclinic crystal structures. Monoclinic dimer CD in B_b region is close to that of tetragonal form, and it nearly coincides with that of triclinic lysozyme with substrate at higher energies (below 205 nm). This suggests that the average conformations of the Trp residues in lysozyme molecules in the monoclinic crystal are close to those in the triclinic form with bound mono-N-acetylglucosamine. It is interesting that the predicted Trp CD of the lysozyme dimer in the monoclinic crystal shows all the basic features of native lysozyme CD, such as the main positive couplet and the small positive band near 210 nm (Fig.II-46), although this type of spectrum has been calculated for monomer molecules in the case of tetragonal and triclinic forms. The CD of monomeric molecules A and B differ substantially from each other and from tetragonal and triclinic HEWL (Fig.II-46).

Figure II-46. Calculated Trp CD of monoclinic HEWL : monoclinic dimer (D, +); subunit A (•); subunit B (▲). Tetragonal HEWL (Tet, -). Triclinic HEWL with substrate (Tr, .). Spectra in the far uv calculated considering all six transitions (a) and B_b transition only (b), and in the near uv considering all six transitions (c).


a)
b)
[θ] (deg cm²/ dmol)
wavelength (nm)
A
B
D
Tet
Tr


c)
[θ] (deg cm²/ dmol)
wavelength (nm)
A
Tet
Tr
D
B

The CD of monoclinic subunits reach values about four times higher than those predicted for tetragonal lysozyme. The two subunits exhibit B_b CD of opposite sign in four out of six exciton bands and four- to five-fold differences in amplitude in two other cases (Fig.II-47). The sign of the CD bands of the B subunit (considering either B_b only or all six transitions) coincide with those of the dimer and the tetragonal form. The signs of the CD bands exhibited by molecule A are opposite (Fig.II-46, II-47). The CD of the dimer in the monoclinic crystal form can be considered as a sum of the CD of the two subunits. In the near uv, the A subunit shows distinct bands for L_a and L_b transitions, forming a positive couplet (Fig.II-46c). In the case of the B subunit, a strong negative L_a contribution dominates in the near-uv region. We compared values of B_b CD, distances and angles between transition dipole moments directions among individual Trp pairs in A and B subunits and in the tetragonal crystal form. Pairs showing considerable differences in CD are presented in Table II-7. As can be seen, corresponding pairs in A and B molecules exhibit CD of opposite sign in most cases. The CD and conformations of Trp residues in the B subunit seem to be closer to the tetragonal form than in the A subunit. Analyzing the contributions of the six lysozyme Trps to each of the exciton bands, we found that differences in the CD of A and B subunits correlate with the contributions of Trp62-Trp63 and Trp28-Trp108 pairs to the particular exciton bands (Fig.II-48). The Trp62-Trp63 pair largely contributes to the highest energy exciton band in the A subunit and it is the Trp28-Trp108 pair that contributes mostly to this band in the B subunit. This finding might be explained by selective alterations of the Trp residue conformations in the monoclinic monomers. We might assume that a conformation more

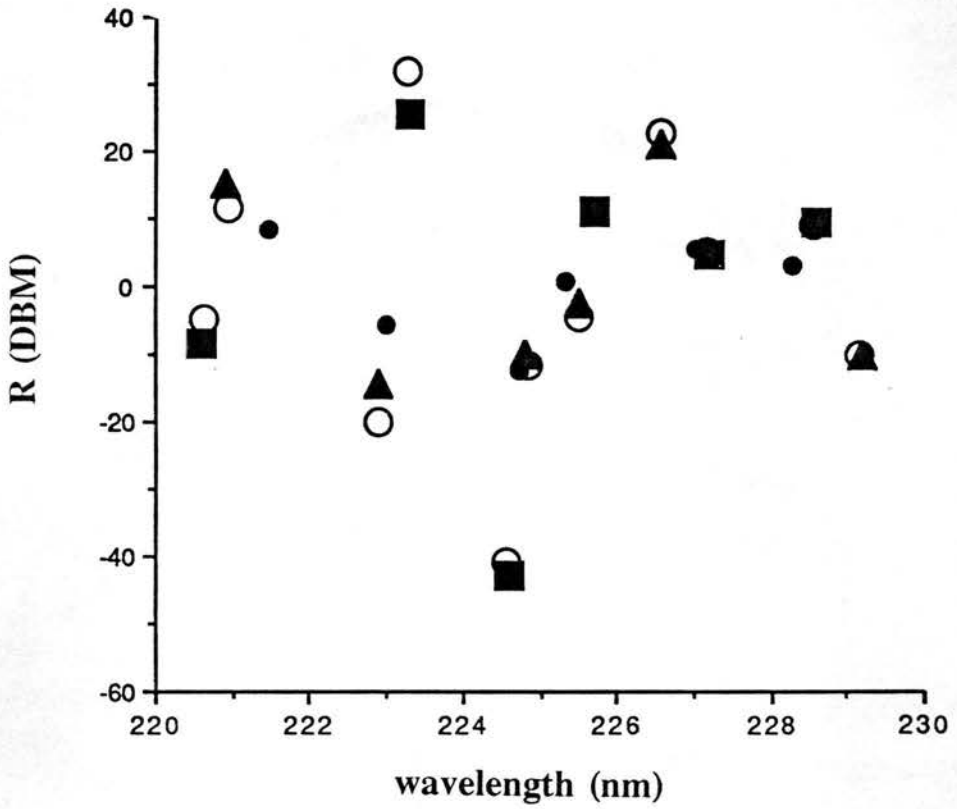


Figure II-47. Position and rotational strength of B_b exciton bands in Trp CD of monoclinic HEWL: dimer (O); molecule A (▲); molecule B (■); compared with tetragonal HEWL (•).

Table II-7. Parameters of Trp pairs in A and B subunits of monoclinic hen lysozyme, showing major CD differences.

Molecule	Trp pair	CD at 231 nm	Distance, Å	Angle*
A		-913	17.1	105.1
B	28-62	26	20.4	77.0
tetragonal		296	19.5	72.7
A		2130	8.5	69.5
B	28-108	-259	7.3	102.5
tetragonal		-1369	7.7	101.9
A		1597	12.0	40.7
B	28-123	196	13.1	68.1
tetragonal		1399	12.7	64.2
A		-8579	4.5	35.8
B	62-63	3380	7.7	87.0
tetragonal		653	6.4	47.6
A		1131	8.9	36.3
B	62-108	-30	15.5	33.8
tetragonal		-471	12.8	61.7

* Angle between the normals to indole planes.

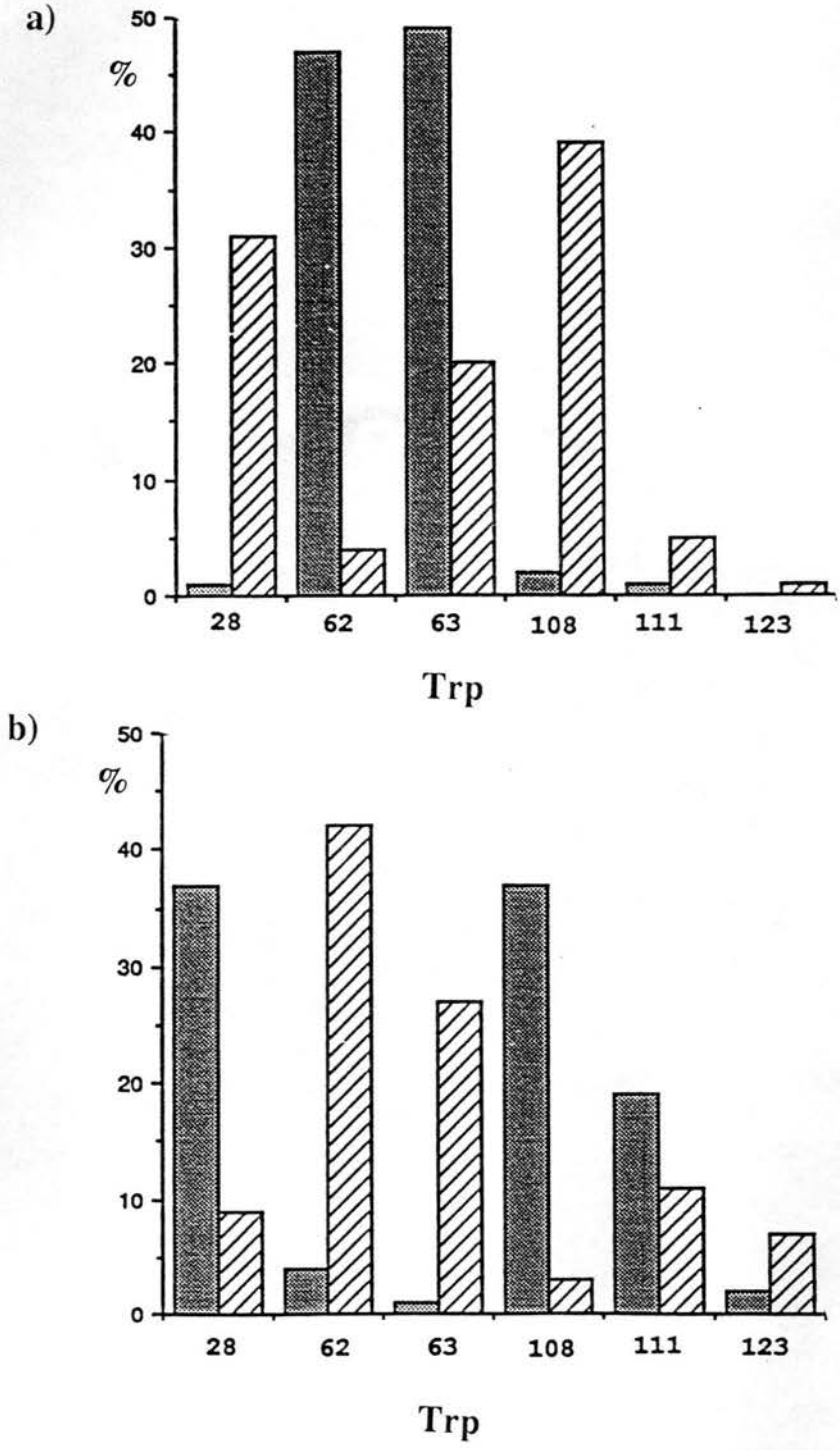


Figure II-48. Partial contributions of individual Trp in A (dotted column) and B (striped column) molecules of monoclinic HEWL crystal to the eigenvectors of the 1st (a) and second (b) B_b exciton bands.

favorable for CD exists in the A subunit for the Trp62-Trp63 pair and in the B subunit for the Trp28-Trp108 pair. We found that the difference in distance between Trp62 and Trp63 in the A and B molecules is 3.2 Å and 1 Å for the Trp28-Trp108 pair. Conformational changes in subunit A (compared to the tetragonal form) occur at Trp62 and Trp28, and in subunit B, torsions are observed at Trp62, Trp63 and Trp123 (Fig.II-49). It has been found (Kundrot & Richards, 1987) that HEWL can be considered as being composed of two domains with different rigidity, the more flexible comprised of residues 44-80, which might explain the differences in deformations of A and B molecules by their orientation in the crystal and different areas of intermolecular contacts. The most important might be the movement of the C-terminal segment 125-129 and the N-terminal top of the helix 5-15 of one molecule A towards the active site cleft of another A molecule related by crystallographic symmetry, approaching the segments 62-64 and 101-109, which include Trp62, Trp63 and Trp108.

c. Hen and Turkey Lysozyme.

Hen and turkey lysozymes have the same Trp side chains at conserved positions (Fig.II-50). The alignment of C_α atoms for all Trp in hen and turkey gives a very low RMS deviation (0.29). Nevertheless, the predicted Trp CD for turkey lysozyme is about two times higher and does not exhibit a positive band around 211-212 nm characteristic for free HEWL (Fig.II-51). The comparison of Trp conformations in hen and turkey lysozymes (Fig.II-50) revealed a significant difference only at one residue, Trp62, which is twisted by about 90° in turkey lysozyme relative to hen lysozyme. Comparison with the Trp62 conformation in the h10 complex revealed rotation around the same bond by

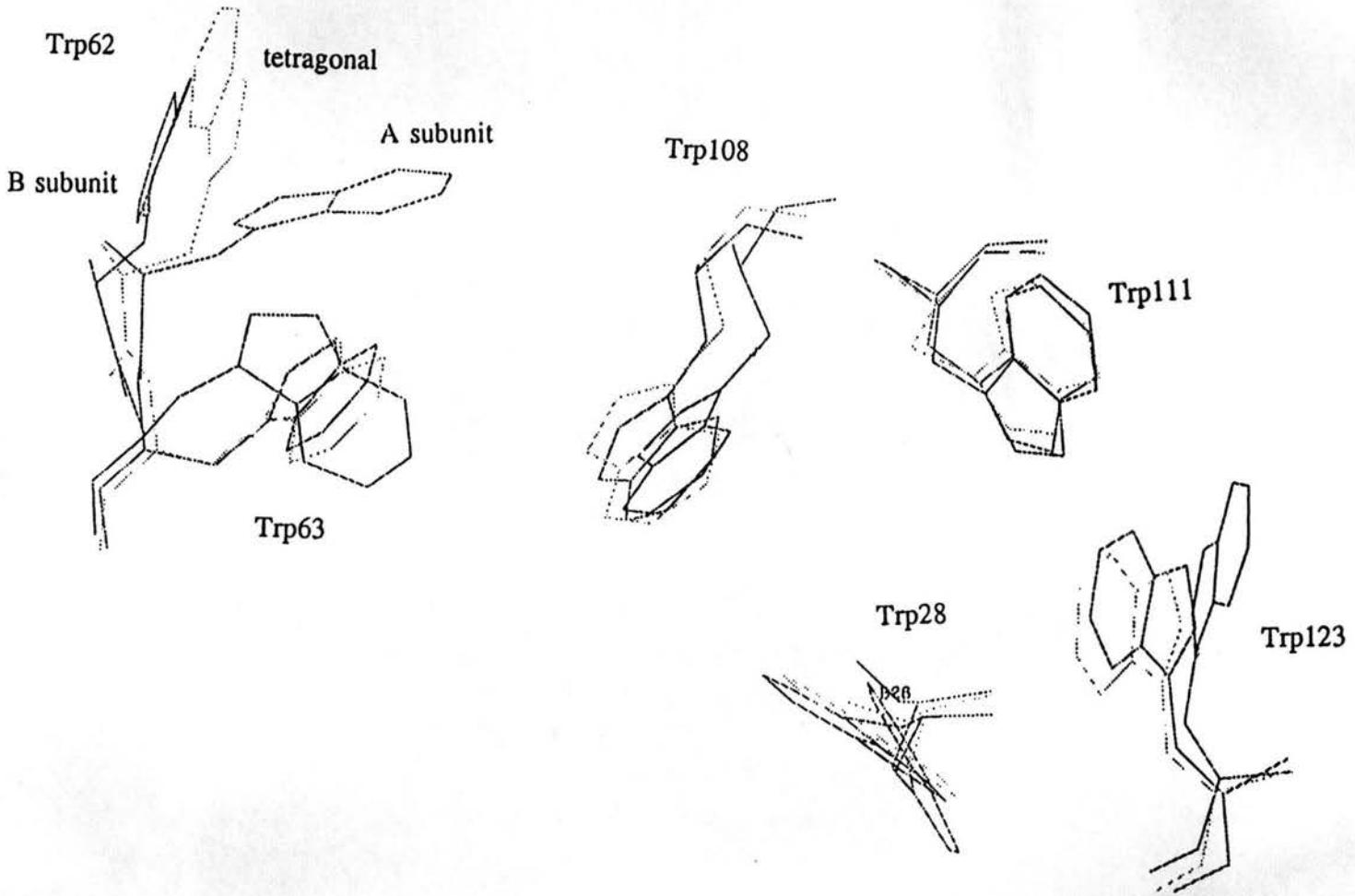


Figure II-49. Conformations of Trp residues in A and B molecules of monoclinic HEWL crystal.

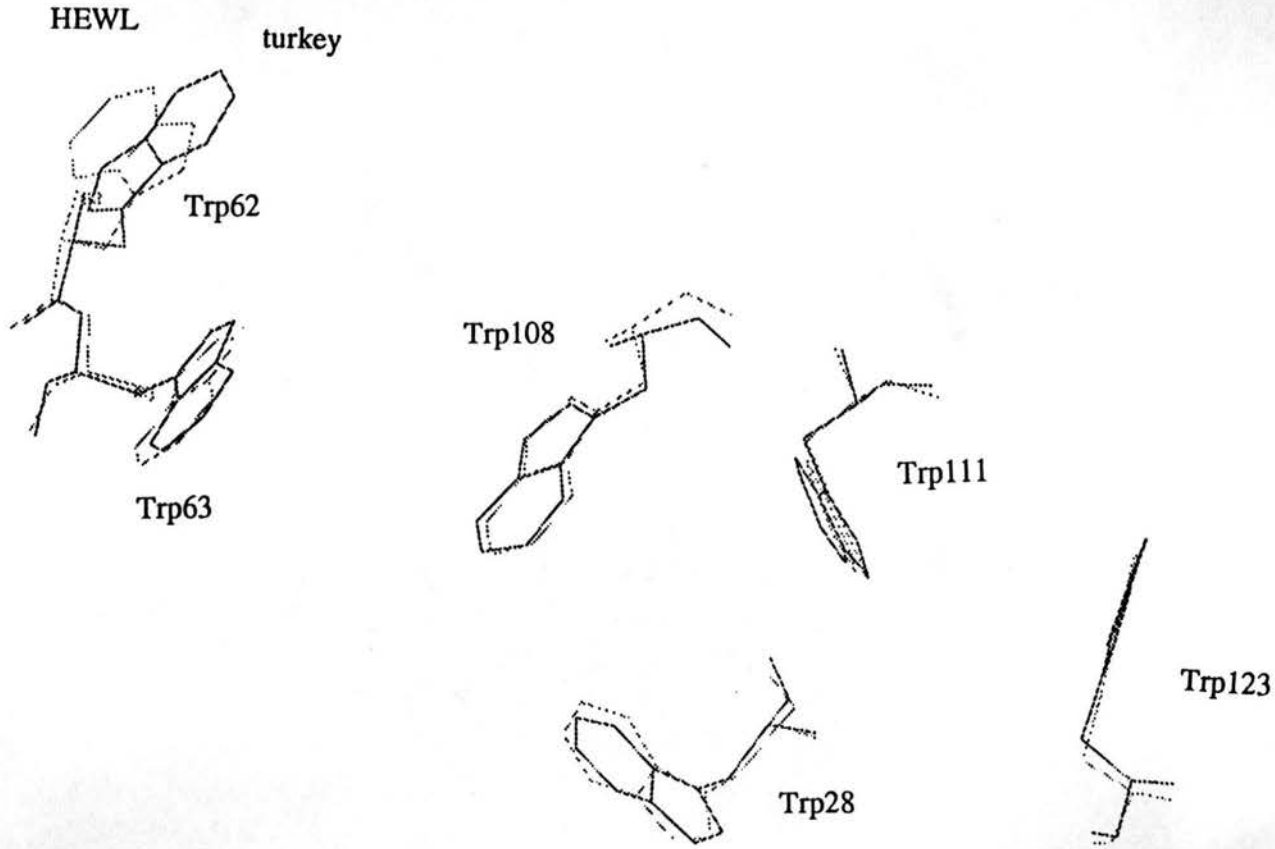


Figure II-50. Conformations of Trp residues in hen (light line)and turkey (dark line) lysozyme.

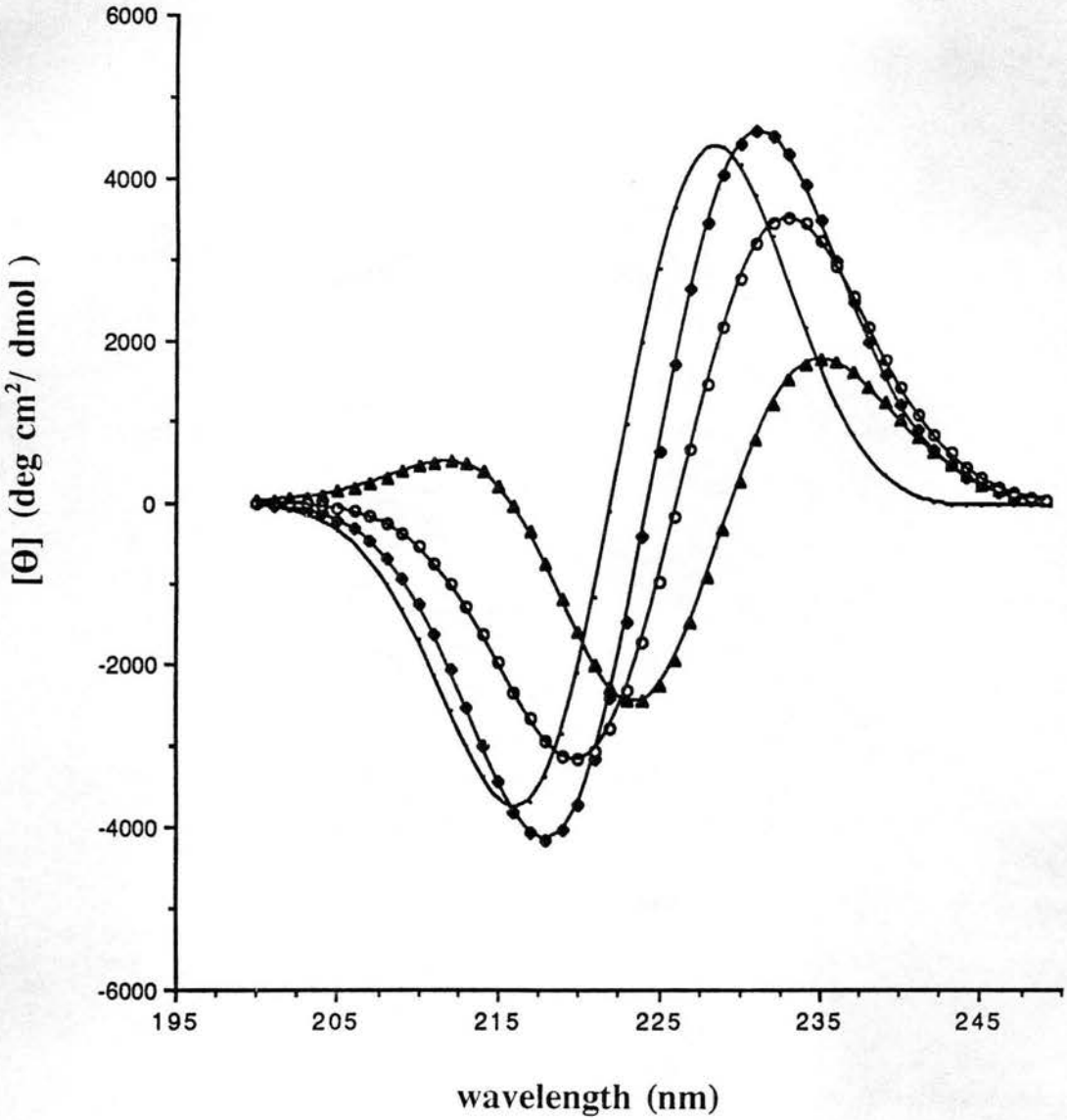


Figure II-51. B_b Trp CD in hen lysozyme (▲), turkey lysozyme (◆), and hen lysozyme bound to Fab HyHEL-10 (o). Difference spectrum between hen and turkey lysozyme B_b Trp CD (-).

an angle (χ_2) smaller than that for conformation in turkey lysozyme (Fig.II-52). The B_b CD of HEWL Trps in h10 complex is more intense than in free hen lysozyme, but less intense than in turkey lysozyme (Fig.II-51).

Analysis of the effects of sequence differences between hen and turkey (T) lysozymes on the conformation of Trp residues revealed that Arg73 (hen (Fig.II-53), which is close to Trp62 (3.6 Å), is replaced with Lys(T) with a distance of 5.3 Å from its amino group to Trp62(T). The Trp63 in HEWL (Fig.II-53) might also experience an effect of the negative charge on Asp101 (3.5 Å), which is replaced with Gly in turkey lysozyme. All the other mutations between hen and turkey lysozymes occur at no less than 10 Å from Trp residues. They are not likely to affect the conformations of Trp residues or the B_b transition moment directions, even though three of these mutations do involve charged residues (His).

The distance between Trp62 and the guanidino group on Arg73 in the h10 complex is intermediate between that for free hen and turkey lysozymes (5.0 Å), which correlates with the increase of the χ_2 angle of Trp62 and the increase of the predicted CD (Fig.II-51). We can see that the conformation of Trp62 is likely to be altered as result of the interactions with neighboring residues and that the flexibility of Trp62 is strongly contributing to the differences in CD of the lysozyme structures. As can be seen from Fig.II-53, Trp62 is located on the edge of the active-site cleft, bordered with positively charged Arg73 (3.6Å), Arg61 (5.8Å) and Lys47 (14Å) from the opposite site.

Figure II-54 and II-55 show the Trp CD of hen and turkey lysozyme, in the far

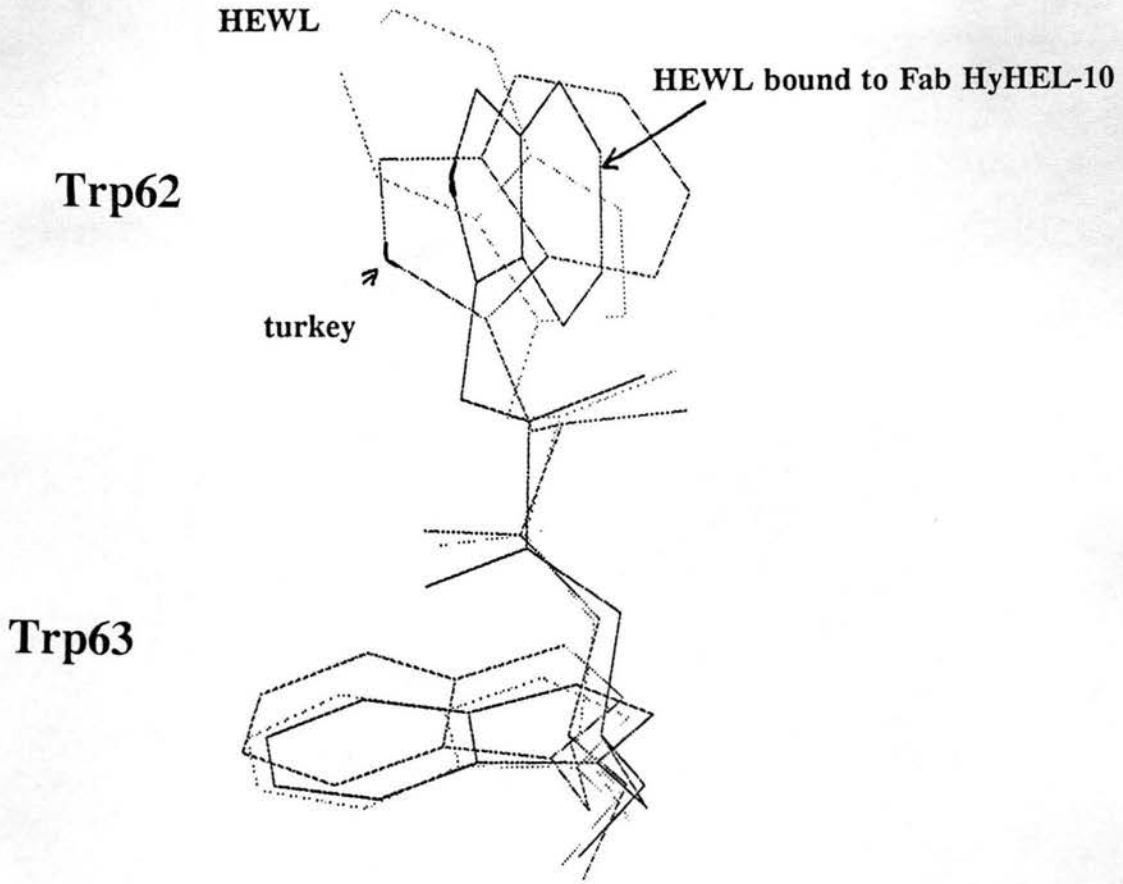
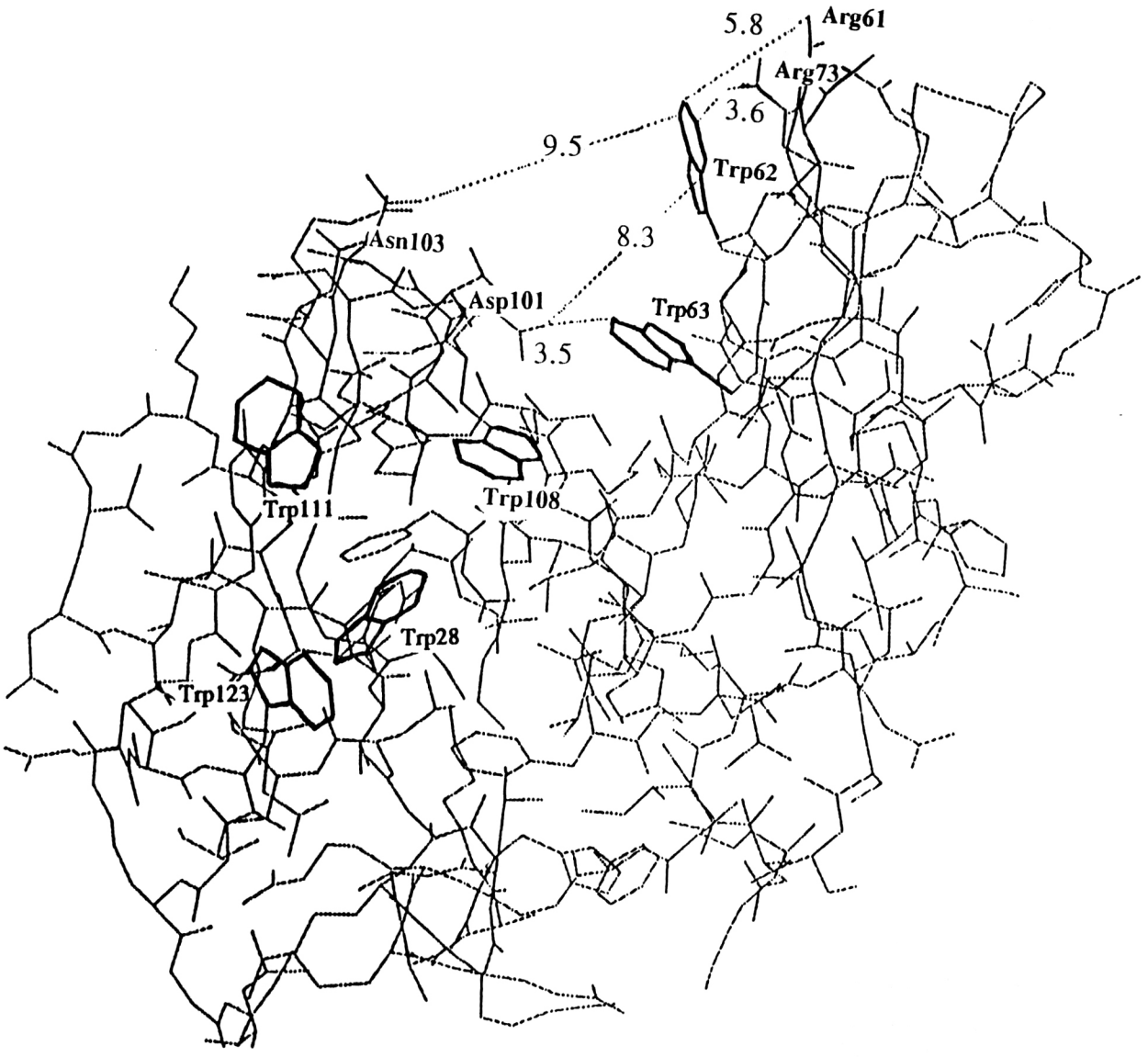


Figure II-52. Conformation of Trp62 and Trp63 in hen lysozyme, turkey lysozyme and hen lysozyme bound to Fab HyHEL-10.

Figure II-53. Conformations of Trp chromophores (bold) in hen lysozyme. Width of the cleft is shown (distance between Trp62 and Asn103). Distances between Asp101, Arg61, Arg73 and Trp62 and Trp63 are shown.


5.8
Arg61
Arg73
3.6
9.5
Trp62
8.3
Asn103
Asp101
Trp63
3.5
Trp111
Trp108
Trp28
Trp123

and near uv respectively, accounting for six transitions. These predicted spectra and the difference spectrum were used for comparison with experimental CD (Chapter III.4.e).

d. Human Lysozyme and Baboon Lactalbumin.

Human lysozyme and baboon lactalbumin have significant similarity to HEWL in backbone and side-chain conformations (Artymiuk and Blake, 1981; Acharya et al., 1989). The calculated B_b Trp CD of baboon lactalbumin differs significantly from that of HEWL (Fig.II-56). Trp CD of baboon lactalbumin is about two times smaller, and is represented by a negative couplet, similar to human lysozyme, while HEWL has a triplet band with a large positive couplet on the longer wavelength side. Lactalbumin contains only three Trp residues, whose relative conformations match satisfactorily those of Trp63, Trp108 and Trp123 in HEWL (Fig.II-57). Human lysozyme, relative to HEWL, lacks Trp62 and Trp123. Trp34 is present in human lysozyme, instead of Phe34 in HEWL (Fig.II-56). Both human lysozyme and baboon lactalbumin have CD similar to that of HEWL in the 200-215 nm region, where free HEWL exhibits a characteristic positive band. Comparison of predicted and experimental CD for human lysozyme is discussed in Chapter III.

e. T4 phage lysozyme and mutants.

The lysozyme of T4 bacteriophage has only three Trp residues and its predicted CD is similar to that of human lysozyme and baboon lactalbumin (Fig.II-56). The folding of T4 phage lysozyme is different from that of HEWL. Thus no close homology in the conformation of Trp residues was found, although locations of the three T4

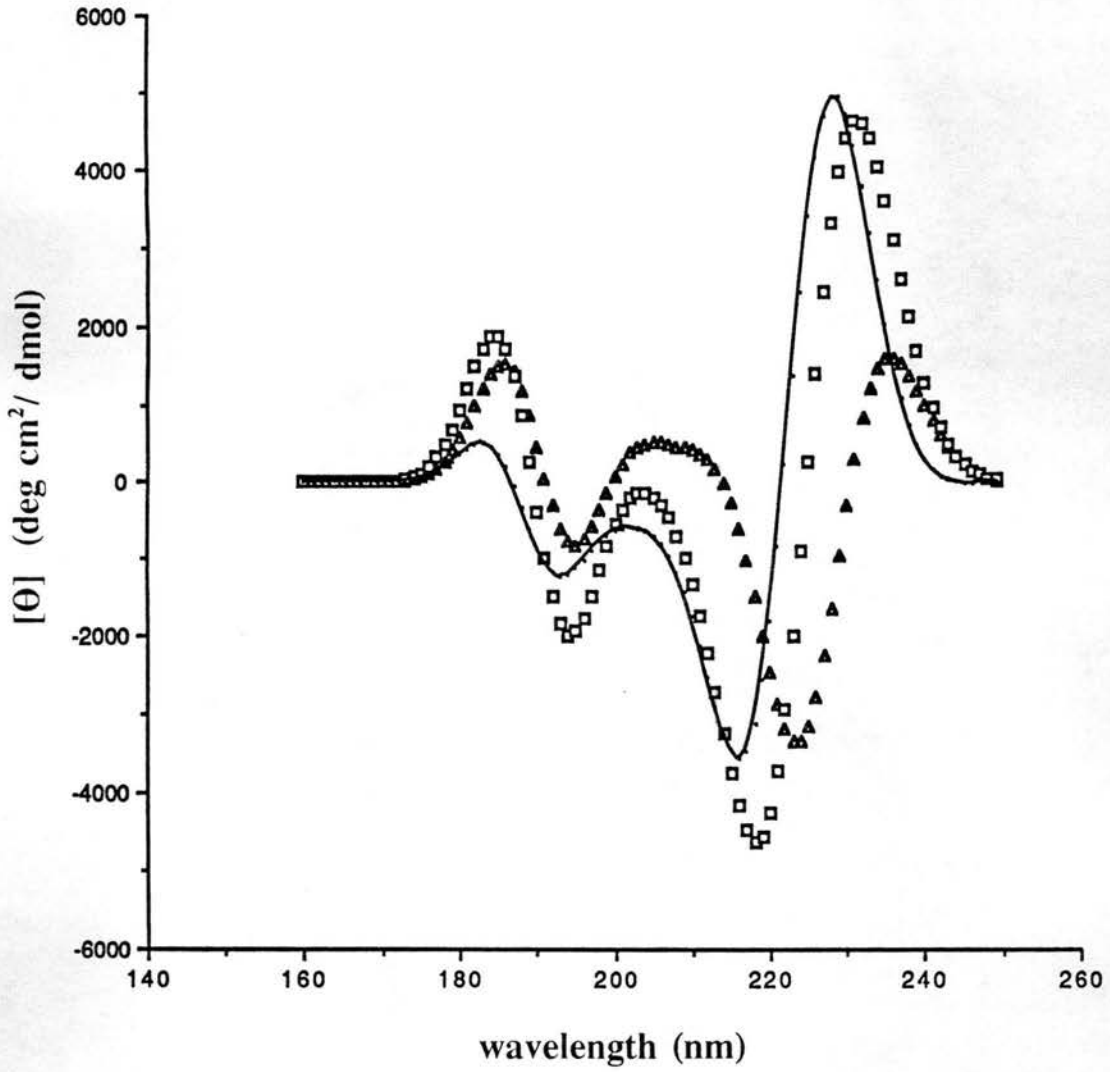


Figure II-54. Calculated far-uv Trp CD of hen lysozyme (▲) and turkey lysozyme (□) considering six indole transitions. Difference Trp CD spectrum between hen and turkey lysozyme (-).

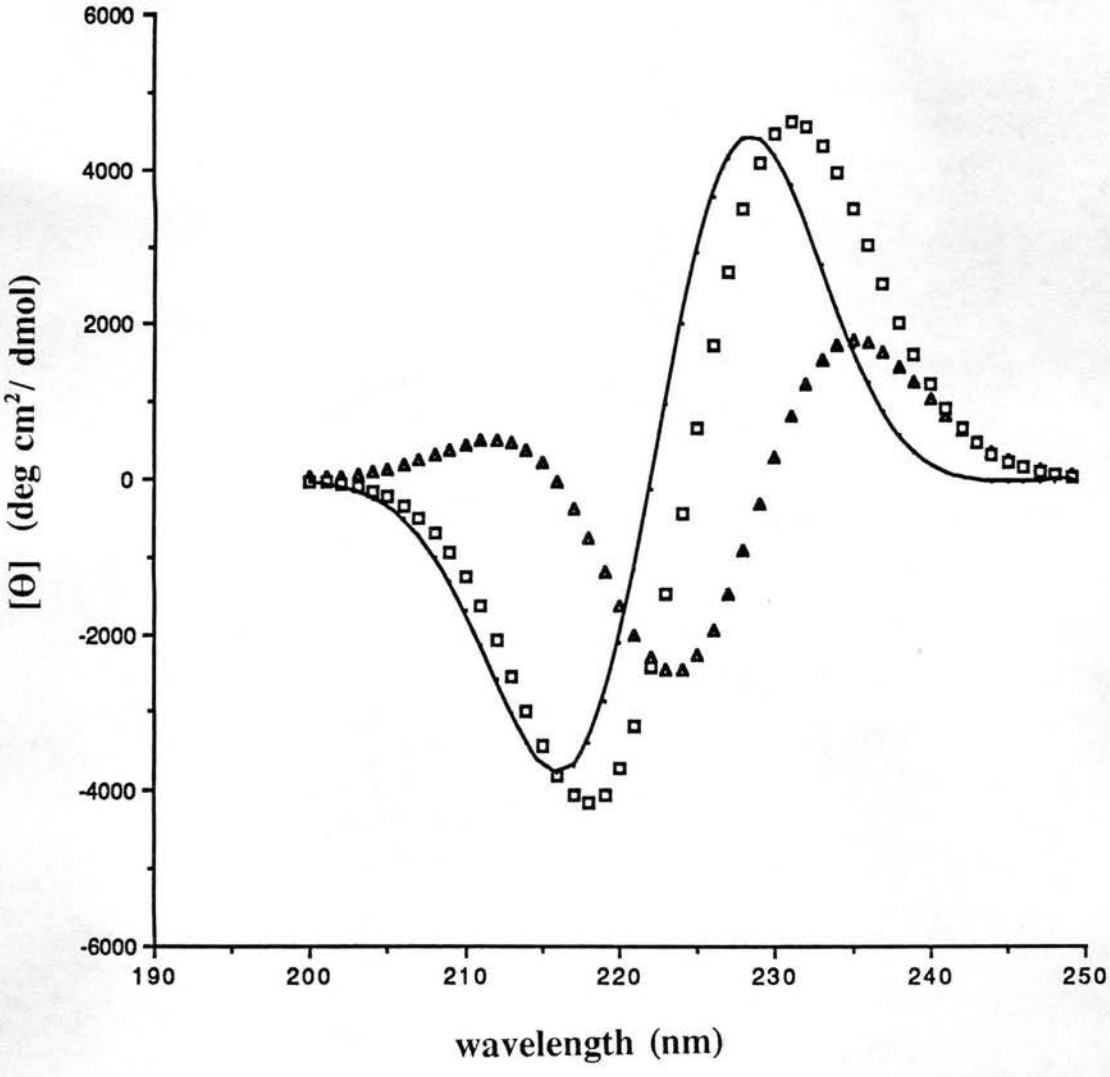


Figure II-55. Calculated near-uv Trp CD of hen lysozyme (▲) and turkey lysozyme (□) considering six indole transitions. Difference Trp CD spectrum between hen and turkey lysozyme (-).

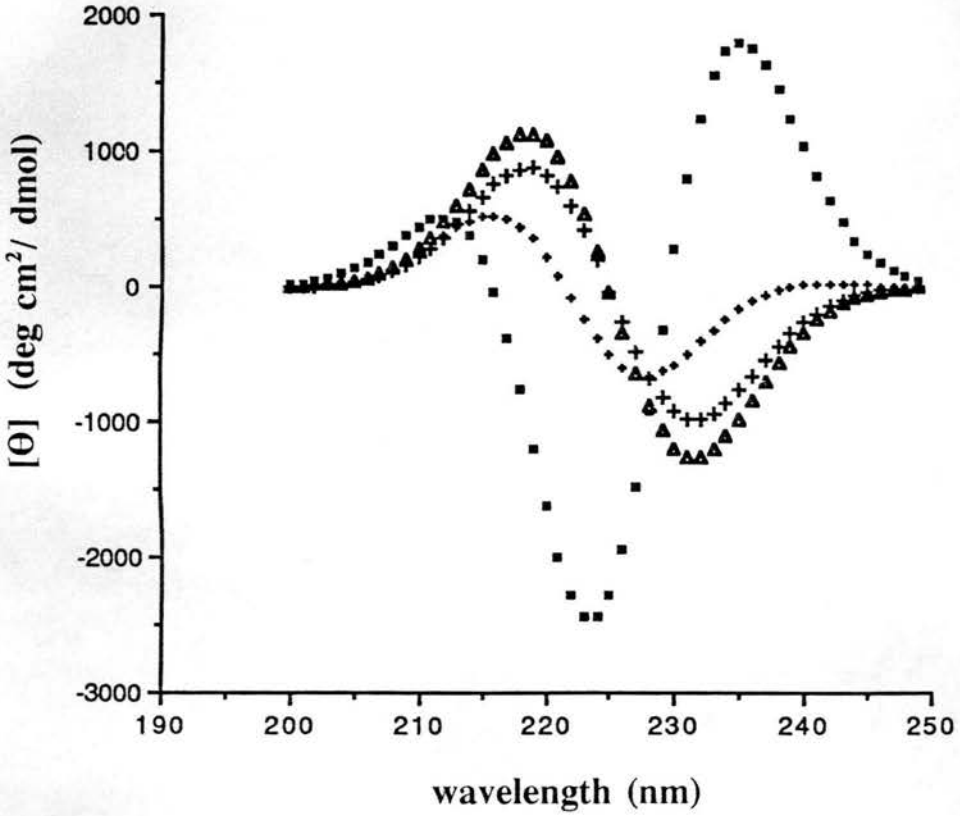


Figure II-56. Calculated B_b Trp CD of T4 phage lysozyme (▵); baboon lactalbumin (+); hen lysozyme (■) and human lysozyme (+).

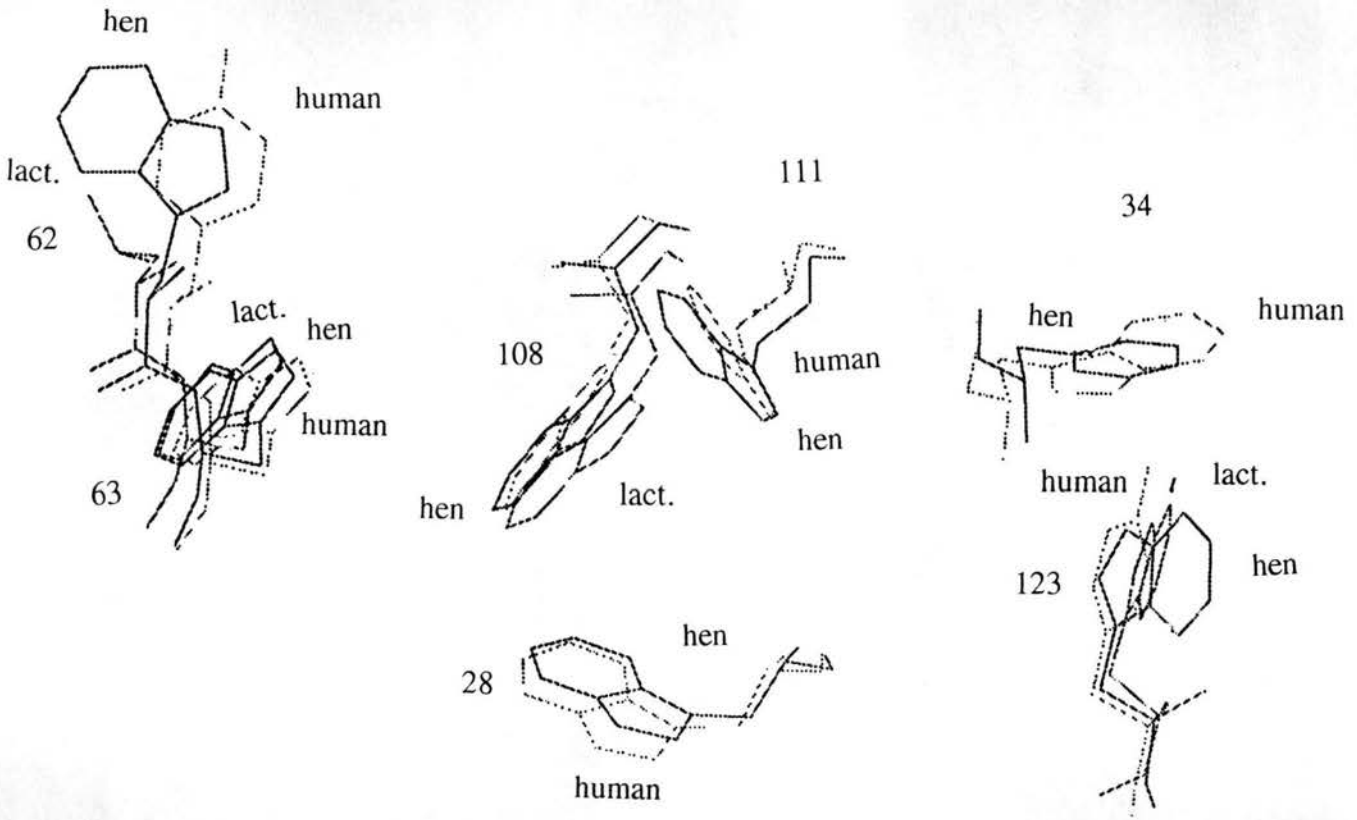


Figure II-57. Alignment of Trps and side chains at corresponding positions in hen lysozyme, human lysozyme and baboon lactalbumin.

lysozyme Trps in the structure resemble those of Trp63, Trp108 and Trp123 in hen lysozyme.

Matthews and coworkers (Matthews et al., 1987; Alber et al., 1988; Nicholson et al., 1988; Matsumura et al., 1988; Nicholson et al., 1989; Weather et al., 1989) have obtained a number of mutants of T4 lysozyme, which differ in thermal stability. As can be seen (Fig.II-58), the mutations have little effect on Trp conformations. Effects of the distribution of charges in the local surroundings of Trp residues might, in principle, affect the direction of the transition dipole moment and thus the CD. Most of the mutants show a tendency to increase the magnitude of Trp CD. The only significant change in CD was predicted for the I3Y mutant. This means that the substitution of Ile3 for Tyr has the most disruptive effect on the conformation of Trp residues in the bacteriophage lysozyme structure.

f. α- and γ-Chymotrypsin.

α- and γ-chymotrypsin are two sequential intermediates in chymotrypsin activation. Inactive chymotrypsinogen is produced by acinous cells of the pancreas and carried as such by pancreatic juice into the duodenum where it is activated by trypsin (Merck Index, 1968) by conversion into π-chymotrypsin, which in turn is degraded to δ-chymotrypsin in the presence of large amounts of trypsin, or is converted to α-chymotrypsin in the presence of small amounts of trypsin. After long standing α-chymotrypsin gives rise to β- and γ-chymotrypsin. Figure II-59 presents the amplitude of Trp B_b CD in α-chymotrypsin and γ-chymotrypsin.

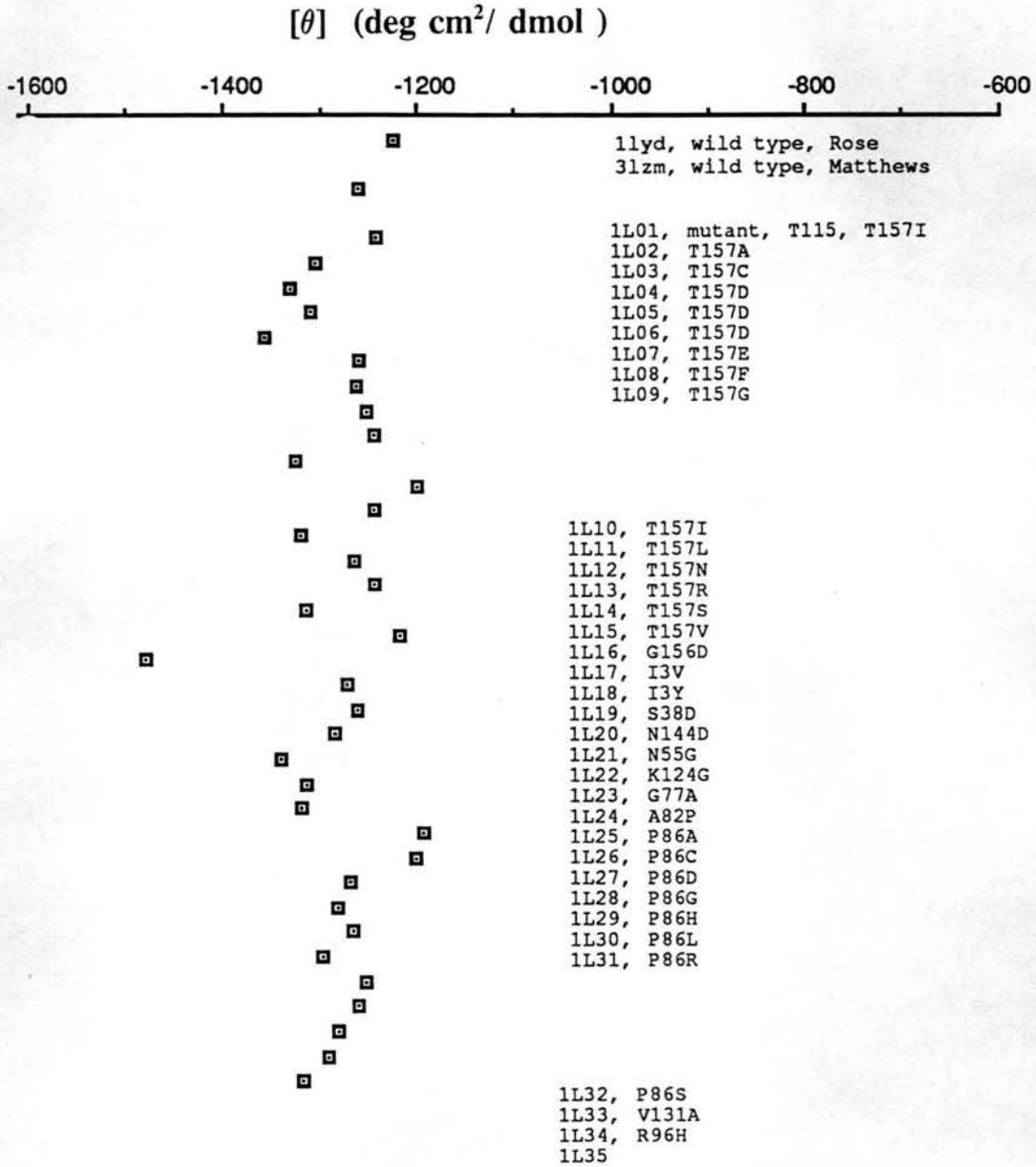
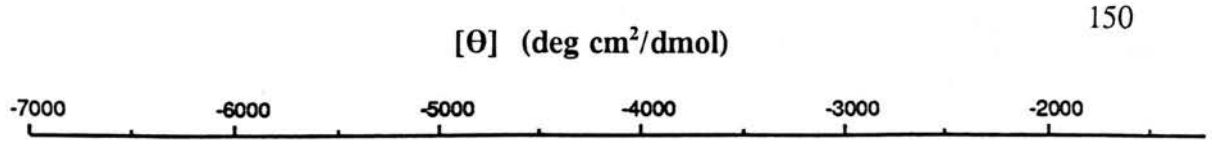
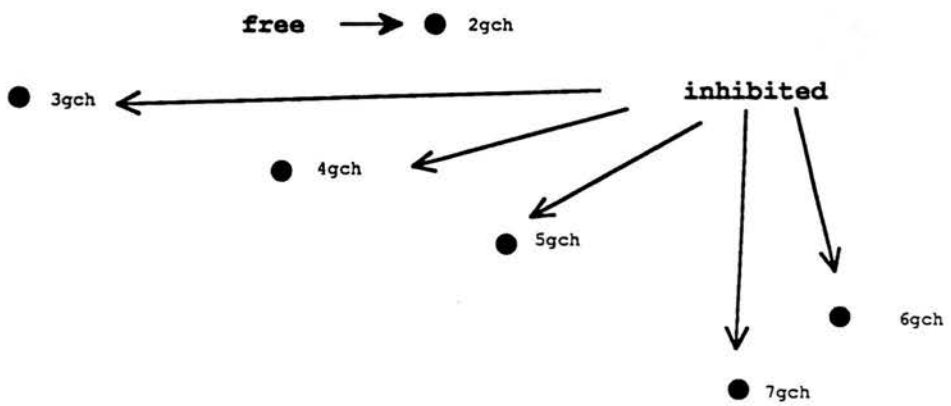


Figure II-58. Calculated B_b Trp CD of wild-type T4 bacteriophage lysozyme and mutants at 231 nm.

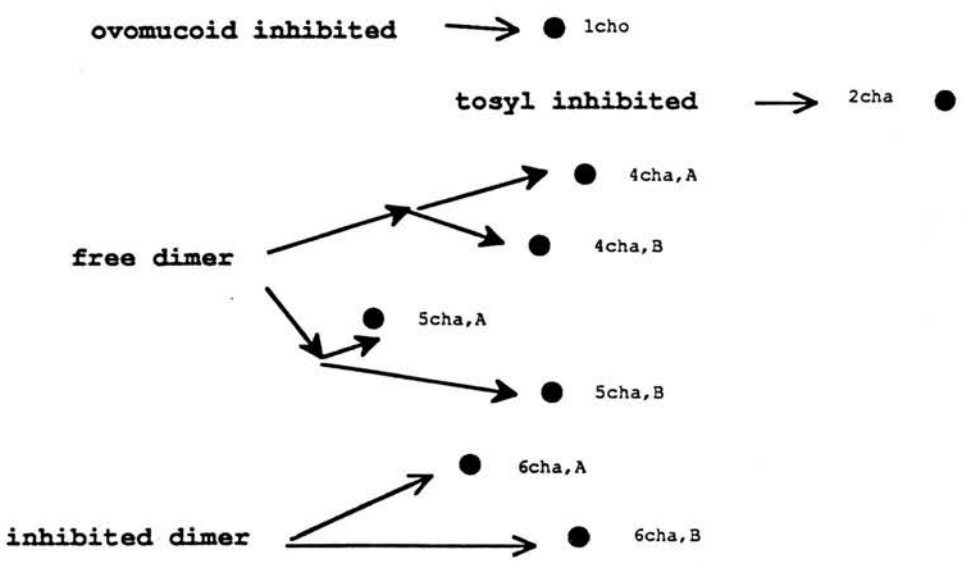
Figure II-59. Amplitude of B_b Trp CD at lond-wavelength extremum in bovine α- and γ-chymotrypsins. α-chymotrypsin crystal structures considered : 1cho, in complex with turkey ovomucoid third domain; 2cha, tosylated; 4cha, dimer refined by Tsukada & Blow (1983); 5cha, dimer refined by Tulinsky & Blewins (1987); 6cha, dimer, phenylethane boronic acid bound to each molecule. γ-chymotrypsin crystal structures considered : 2gch, native; 3gch, complex with trans-O-hydroxy-α-methyl cinnamate; 4gch, complex with P-diethylamino-O-hydroxy-α-methyl cinnamate; 5gch, photolysis product of 4gch; 6gch, complex with N-acetyl-phenylalanyl trifluoromethyl ketone bound at the active site; 7gch, complex with N-acetyl-L-leucyl-L-phenylalanyl trifluoromethyl ketone bound at the active site.



γ -chymotrypsin



α -chymotrypsin



The γ-chymotrypsin Trp CD is more variable and sensitive to crystallization conditions than that of α-chymotrypsin. The γ-chymotrypsin was crystallized as a monomer either in free form or with inhibitors. In fact, γ-chymotrypsin shows much weaker dimerization in solution than α-chymotrypsin, which dimerizes between pH 5.5 and 3.5 (Gorbunoff, 1971). α-chymotrypsin has two molecules in the asymmetric cell unit, which show slightly different Trp CD. The CD of one of the monomers is close to that of ovomucoid-inhibited enzyme. The tosyl inhibited α-chymotrypsin Trp CD shows a difference in amplitude of about 2000 degcm^2/dmole from the other α-chymotrypsin spectra, which signifies distortions of Trp residues. Analysis of crystal structures showed slight distortions in Trp conformations, although no significant conformational differences were found between α- or γ-chymotrypsin structures.

g. Trypsin.

(1) Bovine Trypsin and Trypsinogen.

Figure II-60 presents the intensity of Trp B_b CD at 231 nm, the position of the long-wavelength extremum of the couplet, in trypsin crystals, and Figure II-61 in trypsinogen crystals. The trypsinogen structure differs from that of trypsin by only six N-terminal residues, which are removed upon activation of the enzyme. Both structures contain four Trp, Trp51, Trp141, Trp215 and Trp237, the conformations of which in trypsin and trypsinogen structures coincide with those in α-chymotrypsin (Fig.III-8 in Chapter III). Trypsin bound to trypsin inhibitor shows a slight increase in Trp CD. Trp CD in trypsins is about two-fold less intense than in chymotrypsins, probably due to the lack of Trp172, which makes a large contribution to chymotrypsin CD as will be

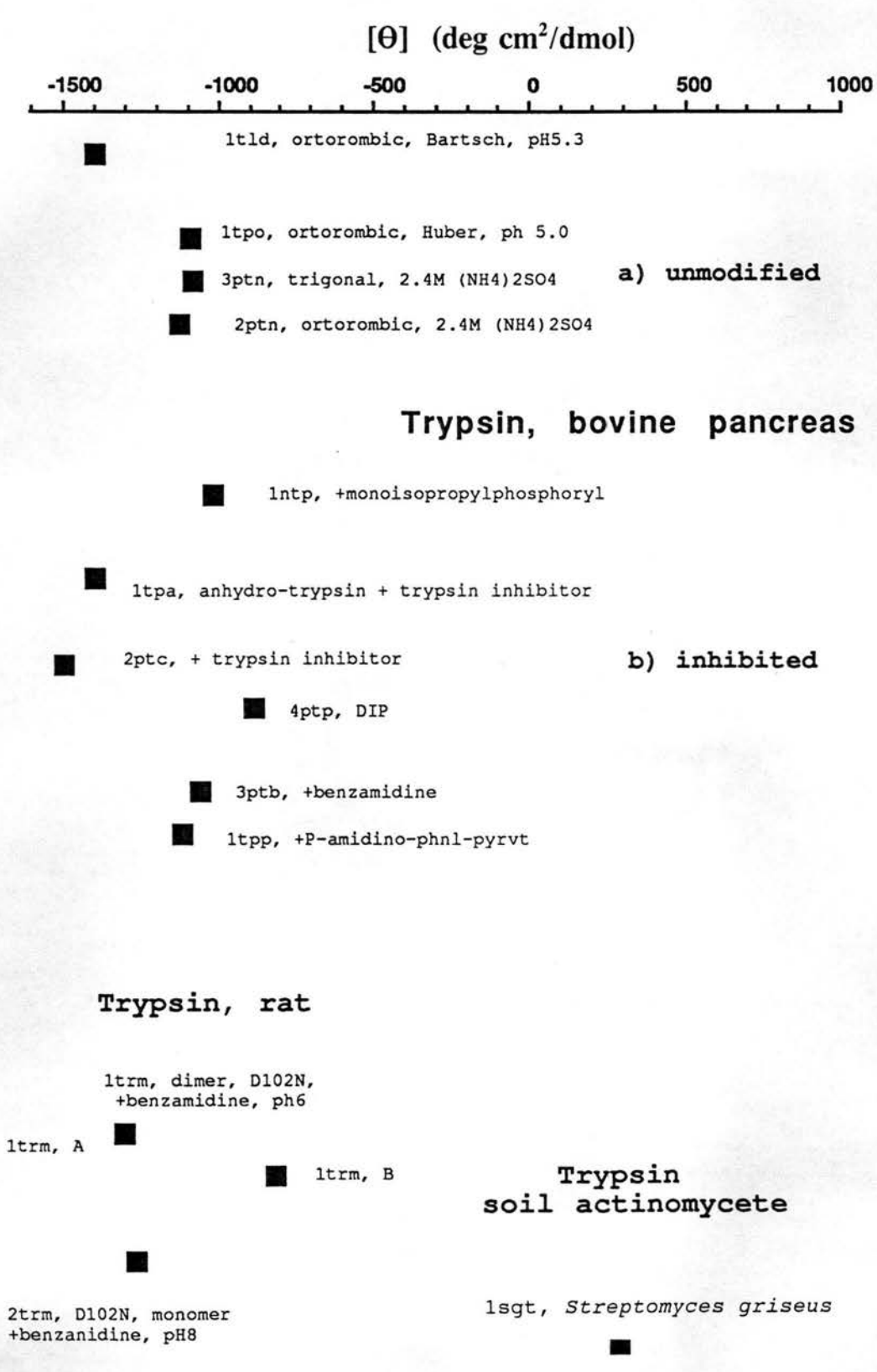


Figure II-60. B_b Trp CD of trypsins. Amplitude at long-wavelength band of the couplet.

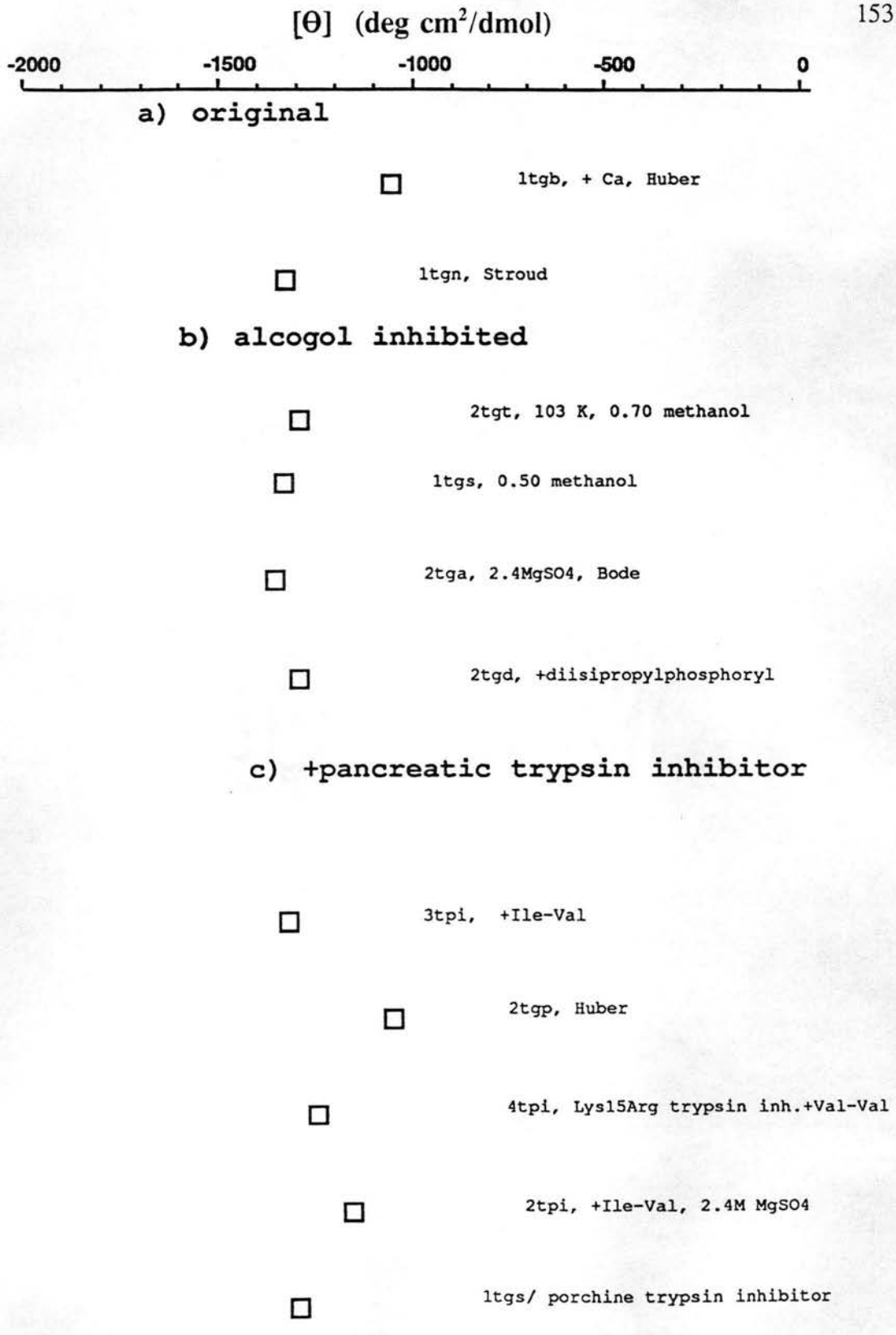


Figure II-61. B_b Trp CD of trypsinogens. Amplitude at long-wavelength band of the couplet.

discussed in Chapter III. Neither crystallization conditions nor substrate have any detectable effect on Trp CD of trypsinogen. Analysis of crystal structures shows almost no conformational differences among the Trp residues in the whole set of trypsinogen structures analyzed. Moreover the conformations of Trp51, Trp141, Trp215 and Trp237 in trypsin and trypsinogen structures coincide with those in chymotrypsin.

(2) Rat trypsin.

CD of rat trypsin differs slightly for the B molecule of the dimeric form and for the monomeric enzyme (Fig.II-60). Dimer crystals were grown at pH 6, whereas monomeric crystals were obtained at pH 8. Rat trypsin CD coincides with that of the bovine enzyme, which is not surprising since all Trp residues in these species are conserved and so are 74% of the remaining amino acids.

(3) Actinomycete trypsin.

Trp CD of trypsin from *Streptomyces griseus* is significantly different from that of the bovine enzyme (Fig.II-60) and so are the amino acid sequences of these structures. Both enzymes have four Trps, but only one of these occurs in the same position in both proteins.

Chapter III. Estimation of the Secondary Structure Content and Aromatic Contributions in Globular Proteins from CD.

1. Introduction

Interest in the circular dichroism of biopolymers has expanded enormously in recent years. In the area of proteins, the renewed interest stems from several sources. The protein folding problem has emerged as a central problem of biochemistry and biophysics, and CD is a highly sensitive method for following folding and unfolding transitions in proteins. Site-directed mutagenesis is a powerful technique for exploring the role of specific residues in protein structure and function. It is necessary to establish that the mutation has not significantly perturbed the protein conformation if function is to be elucidated, or to detect and characterize the conformational change if structural effects are being studied. New peptides with hormonal, neurotransmitter and other activities related to protein-protein and protein-DNA interactions are being isolated at an astounding rate. The conformational characterization of these peptides in solution is extremely convenient with CD, even if further thorough NMR analysis is possible.

2. Approaches for Secondary Structure Determination from CD.

a. Fixed Reference Spectra.

CD is an experimentally simple method for studying protein tertiary and secondary structure in solution. For instance, the interactions of peptides within different secondary structures, such as α-helices and β-sheets etc., give rise to characteristic CD bands in the far UV region (250 -175 nm). The α-helix exhibits two negative bands at

208 nm and 222 nm and a positive band at ~190 nm. The β-structure also exhibits a positive band ~ 190 and a negative band at ~ 217 nm. The original approaches to quantitative analysis of far-uv CD of proteins consisted in deconvolution of the protein spectrum into a linear combination of spectra for different secondary structural types (α-helix, β-structure and unordered) obtained as the CD of polypeptides under various solvent conditions (Greenfield & Fasman, 1968; Sarkar & Doty, 1967). Brahms & Brahms (1977; 1980) also attempted to account for the β-turn and even for aromatic side-chain contributions in the framework of linear deconvolution. Improvement was achieved when, instead of model systems, the reference secondary structure spectra were obtained from protein CD spectra (Greenfield & Fasman, 1969; Chen et al., 1974; Chang et al., 1978; Bolotina et al., 1980a,b,c). The CD spectra of a set of proteins of known structure are measured at regular intervals, usually in steps of 1 nm between 185 nm and 260 nm, so m points are used in the calculations. Then a system of M equations is solved, including M different proteins, for which CD spectra are known. In principle, any number of $r \leq M$ reference spectra, corresponding to CD of particular pure secondary structures, can be obtained from such system :

$$[\Theta]_{\lambda j} = f_{j\alpha}\,\alpha_\lambda + f_{j\beta}\,\beta_\lambda + \ldots + f_{jU}\,U_\lambda \qquad \text{(III-1)}$$

or, in generalized form :

$$[\Theta]_{\lambda j} = \sum_{l=1} f_{jl}\,k_{\lambda l} \qquad \text{(III-2)}$$

where $[\Theta]_{\lambda j}$ is the ellipticity per residue of protein j at wavelength λ; f_{jl} are the fractions of different secondary structures l ($l=1,\ldots,r$) in protein j; and $k_{\lambda l}$ (α_λ, β_λ, U_λ) are the reference spectra for particular secondary structures l at wavelength λ. The relative

contribution of each reference spectrum $k_{\lambda l}$ to the overall CD curve deconvoluted through Equation III-2 is proportional to the amount f_{jl} of that structural feature in the protein.

In a number of deconvolution approaches (Greenfield & Fasman, 1969; Chen et al., 1974; Chang et al., 1978; Bolotina et al., 1980a,b,c) the reference spectra, once obtained, were fixed. Thus the CD contribution from a residue in a definite secondary structural conformation to the overall CD spectrum was assumed to be the same for any protein. Consequently the conformation of the peptide groups in a definite secondary structure was not supposed to vary sufficiently from protein to protein as to affect the CD.

b. Unique Protein Contributions in Far-UV CD.

Reference spectra reflect the content of average regular structures and cannot thoroughly account for individual unique protein CD features. These contributions might arise due to irregularities in the secondary structures, such as length dependence; end distortions in α-helices; rare helix types such as 3_{10}-, π-, ω- (Chang et al., 1978; Manning et all., 1988; Manning & Woody, 1991); distortions in β-sheet structures (Manning & Woody, 1987; Manning et al., 1988); and the broad distribution of the β-turn conformations accounted for on average in the reference spectra (Woody, 1974; Perczel et al., 1991; Fasman & Potter, 1967; Stevens et al., 1968; Laiken et al., 1967). Also, the conformations considered might exhibit various CD spectra dependent on the environment and amino acid composition (Woody, 1992; Privalov et al., 1988). On the other hand, model compound studies have revealed that aromatic amino acids might exhibit high-intensity bands in the far-uv CD (Shiraki & Imahori, 1966; Quadrifoglio et

al., 1970; Peggion et al., 1968; Urry, 1968; Auer, 1973; Woody, 1978). Some proteins, such as snake venom neurotoxins (Menez et al., 1976; Yang et al., 1968; Cameron & Tu, 1977), gene 5 protein from filamentous bacteriophages (Day, 1973), acid DNAase (Timasheff & Bernardi, 1970), and avidin (Green & Melamed, 1966), exhibit a positive band around 220-230 nm, which cannot be attributed to any known secondary structure, but is consistent with Tyr contributions from the L_a band and Trp contributions from the B_b band identified in model compounds. In other cases, aromatic chromophores might make sizable contributions, either positive or negative, to the far-uv CD without exhibiting any obvious anomaly, such as concanavalin (Brahms & Brahms, 1980) and trypsin inhibitor (Kosen et al., 1983; Ikeda et al., 1968). In this case, the accuracy of secondary structure analysis can be seriously compromised.

c. Indirect Accounting for Unique Protein CD.

Important progress has been made recently in the empirical analysis of protein secondary structure by CD (Provencher & Glöckner, 1980; Johnson, 1988, 1990; Manavalan & Johnson, 1987; van Stockkum et al., 1990; Sreerama & Woody, 1993).

Provencher & Glöckner (1980), while still using a fixed basis set of proteins, weighted their individual contributions according to the fit to the studied spectrum. Several recently developed methods (van Stokkum et al., 1990; Sreerama & Woody, 1993) use the variable selection principle introduced by Manavalan & Johnson (1987). In this approach, various combinations of protein CD spectra are used to fit the spectrum of the unknown protein, rather than a single set. This has the advantage that proteins that have CD features in common with the unknown protein are given additional weight in

the analysis while those that have very different features are deleted from the basis set. Thus if one or more proteins in the total basis set have aromatic CD contributions similar to those in the spectrum of the unknown protein, the variable selection method will weight those proteins more heavily in the analysis. If none of the proteins in the initial basis set have such contributions, the variable selection method will not lead to an improved result. In such cases, if the aromatic feature is very prominent, the analysis can generally be recognized as flawed by the failure of the secondary structure fractions to be positive or to sum to one.

The self consistent approach (Sreerama & Woody, 1993) incorporates the studied spectrum into the basis set, while assuming some initial values for its structural content: either those of the protein from the basis set with the best fit to the unknown protein spectrum or by choice of the researcher. Then the content fractions are optimized through the local linearized version (van Stokkum et al., 1990) of the variable selection method (Manavalan & Johnson, 1987).

Current methods give good accuracy in estimating α-helix content for a wide range of proteins and, with somewhat lower accuracy, provide useful estimates of β-sheet, β-turn and unordered conformations (Table III-1). There remain some proteins, however, that have anomalous far-uv CD spectra and therefore do not give reliable analyses. Contributions from aromatic side chains are often the primary source of those anomalies. The introduction of ideas from artificial intelligence in CD analysis (pattern recognition and self training) has been marked by incorporation of the so-called "neural networks" (Pancoska et al., 1992; Bohm et al., 1992). The "neural network" might be

Table III-1. Accuracy of secondary structure estimations by CD and combined CD/IR approaches. Comparison of Pearson correlation coefficients.

Structure	PG	IR/CD	MJ	HJ	BL	LL	SC	CCA
α-helix	0.95	0.86 0.98 0.99	0.97	0.98	0.98	0.96	0.95	0.93
β-structure	0.72	0.90 0.39 0.91	0.73	0.88	0.96 0.80 0.52	0.85	0.84	0.71
β-turns	0.31	-0.21 0.34 0.72	0.37	0.56	0.33	0.80	0.77	0.73
unordered	0.41	0.38 0.34 0.80	0.77	0.76	0.72	0.79	0.72	0.48

PG - Provencher & Glöckner (1980);

IR/CD - infrared/CD, Sarver & Kruger (1991b); the first line corresponds to deconvolution on IR reference spectra only, the second line is the deconvolution on CD spectra only, and the third line corresponds to the deconvolution on combined IR and CD reference spectra;

MJ - Manavalan & Johnson (1988);

HJ - CD data to 168 nm included in Manavalan & Johnson (1988) method (Toumadje et al., 1992);

BL - Bolotina & Lugauskas (1985); in the column of the correlation coefficients for β-structure, the first line corresponds to the total β-structure (β-parallel and β-antiparallel structures); the second line corresponds to β-antiparallel structure and the third line to β-parallel structure.

LL - local linearized method (Van Stokkum et al., 1990) tested by N. Sreerama.

SC- self-consistent method, as initial guess the structure of the basis set protein is taken with smallest RMS deviation to the unknown (Sreerama & Woody, 1994).

CCA - convex constraint analysis method (Perzel et al., 1991).

considered as a generalized deconvolution approach so that all previously elaborated methods might be described as a restricted form. In the neural network method, the input (protein spectra) and the output (secondary structure content) are connected through a set of coefficients (functions) established for each connection between input and output elements. The type of functions, the number of intermediate layers and the number of elements (nodes) in the intermediate and input layers, can be varied. Once trained (optimized for the input-output connections), the network can be applied to proteins with unknown output. Here the idea of reference spectra is being greatly generalized. A linear network with no hidden layers and a fixed input set corresponds to deconvolution with fixed reference spectra. As in all approaches with variable weights of protein structures in the basis set, it is possible, to a certain extent, to account for unique protein CD features without explicitly deriving them or giving any explanation as to their particular origin.

d. Explicit Accounting for Unique Protein CD.

Few CD secondary structure analysis methods have attempted to explicitly account for the CD of aromatic side chains in the far uv. **Brahms & Brahms (1980)** measured the CD of model dipeptides of Trp, Phe and Tyr and combined these spectra with amino acid composition data to yield a simple correction to the observed spectrum. Although this led to a slightly better least-squares fit to the spectrum in some cases, as manifested by a smaller r.m.s. deviation from the experimental curve, the content of various secondary structural elements changed by only a few percent at most. It seems unlikely that a simple dipeptide can serve as an adequate model for all the aromatic chromophores

in a protein, especially in the linear way suggested by Brahms & Brahms, (1980), since much if not all aromatic CD might come from exciton type interactions between aromatic chromophores (Chapter II, Chapter III.4).

Perczel et al. (1992) applied their convex constraint analysis (CCA) to a set of proteins which included four proteins rich in β-sheet, in addition to the 18 proteins used in the initial work (Perczel et al., 1991). The CCA method does not require any input data from x-ray structures. Instead, the CD data are treated as vectors that are subjected to rotations that yield a set of pure component spectra, the number of which can be varied. When the β-protein-enriched data set was analyzed to obtain five components, four of the components were spectra resembling secondary structural elements such as α-helix, antiparallel β-sheet, etc. One component, however, showed only one significant feature, a positive band centered at 230 nm. Perczel et al. (1992) suggested that this band arises from aromatic and disulfide side chain groups and showed that its amplitude in the CD spectra of the basis set proteins showed a modest correlation ($r = 0.48$) with the fraction of aromatic residues + disulfide groups.

Bolotina & Lugauskas (1985) attempted to account for aromatic contributions in the far-uv CD with additional functions. The authors were able to obtain improved agreement between the CD analyses and x-ray diffraction for proteins outside of the basis set. The method of Bolotina and Lugauskas (1985) represents an improvement over the analysis omitting the aromatic contributions. The allowance for unique aromatic contributions for each protein represents a major advance which may provide useful information about these contributions, as well as improving the accuracy of secondary

structural estimates from CD. Since the Bolotina & Lugauskas (1985) method and modifications of it are extensively used in this thesis, a more thorough description will be given.

e. Method of Bolotina & Lugauskas (1985).

(1) Incorporation of Log-normal Functions in Deconvolution.

In the Bolotina & Lugauskas (1985) method, the observed CD spectrum of a protein is considered to be made up of a contribution from each secondary structural type, weighted by the fraction of that structural type in the protein, plus aromatic contributions. This method extends the nonlinear least squares approach to include in the system of Equations III-2 information on contributions unique to each protein :

$$[\Theta]_{\lambda j} = \Sigma_l f_{jl} k_{\lambda l} + \Sigma_i A_{ji} \phi (\nu^0_{ji}, \Delta\nu_{ji}, \rho_{ji}, \nu_\lambda), \qquad \text{(III-3)}$$

where $l=1,\ldots,r$; $i=1,\ldots,3$; $j = 1,\ldots,M$; $\lambda=1,\ldots,m$. The CD unique for the particular protein is described by a sum of three bands ($i=1,\ldots,3$), with wavelengths in the ranges 190 - 210 nm, 210 - 226 nm and 226 - 235 nm. Each of the bands is taken to have a shape function which is the log-normal distribution (Siano & Metzler, 1969) :

$$A_{jl} \phi (\nu^0_{ji}, \Delta\nu_{ji}, \rho_{ji}, \nu_\lambda) = A_{jl} \exp\left\{ \frac{-\ln 2}{\ln \rho_{ji}} \left(\ln\left[\frac{(\nu_\lambda - \nu^0_{ji})(\rho_{ji}^2 - 1)}{\Delta\nu_{ji} \quad \rho_{ji}} + 1 \right] \right)^2 \right\} \qquad \text{(III-4)}$$

when $\nu_\lambda > \nu^0_{ji} - \Delta\nu_{ji}\rho_{ji} / (\rho_{ji}^2 - 1)$,

and $A_{jl} \phi (\nu^0_{ji}, \Delta\nu_{ji}, \rho_{ji}, \nu_\lambda) = 0$, when $\nu_\lambda \leq \nu^0_{ji} - \Delta\nu_{ji}\rho_{ji} / (\rho_{ji}^2 - 1)$.

The ν_λ in Equation III-4 is the current value of the wave number, j is the number of the protein and i is the number of the log-normal band. Each band $A_{jl} \phi (\nu^0_{ji}, \Delta\nu_{ji}, \rho_{ji}, \nu_\lambda)$ requires four parameters to define its position, ν^0_{ji}; amplitude, A_{jl}; half bandwidth, $\Delta\nu_{ji}$

at e^{-1} of height; and asymmetry, ρ_{ji}. When the asymmetry parameter, ρ_{ji} , approaches one, the log-normal band reduces to a Gaussian function.

The part of the protein CD spectrum described by log-normal functions is considered to contain mostly aromatic contributions for four reasons: 1) the basis set was maintained constant and was composed of proteins with common secondary structural features; 2) the secondary structures were assigned from the x-ray data by a rigid method, i.e., they were narrowly defined; 3) the similarity of the reference spectra for α-helical, β-structural, unordered and, to a lesser extent for β-turn, conformations with those obtained on model systems; 4) the simulations performed for arbitrary combinations of CD spectra of Glu-Trp, Glu-Tyr and Lys-Phe dipeptides, taken from Brahms & Brahms (1980), supported this view. It was shown (Bolotina & Lugauskas, 1985) that all linear combinations of those model compound spectra could be fitted, within the error of experimental CD measurements, by a sum of three log-normal bands, the maxima of which lie in non-overlapping spectral regions : in the ranges 190 - 210 nm, 210 - 226 nm and 226 - 235 nm.

(2) Assignment of Secondary Structures. Rigid Method.

The assignment of the secondary structures from the crystallographic coordinates was performed by the rigid method (Bolotina et al., 1980a; Finkel'shtein et al., 1977) that considers mostly regular secondary structures and is close to the Kabsch & Sander (1983) method. In this method, residues are counted in the α-helical conformation if a 1→5 hydrogen bond is present, or if a 1→4 hydrogen bond is present on the ends of the helix, or it is surrounded by two 1→5 hydrogen-bonded residues inside the helix. Only

residues surrounded on both sides with peptide groups that participate in the hydrogen bonds of the β-sheet, are counted in β-structure. Parallel and antiparallel β-sheet conformations are distinguished. Turns are counted (Bolotina et al., 1980b) following the Chou and Fasman (1977) turn assignments. Turns of types I, II and III are added together, while their mirror images are subtracted, as the latter are assumed to have spectra of the same shape and the opposite sign. Only the second and third residues, of the four residues forming the turn, are counted. If turn residues may also be attributed to an α-helix or β-structure, preference is given to the latter. All the residues that do not fit in one of the categories discussed are considered to be in the unordered conformation. The assignment of secondary structure from x-ray coordinates by the rigid criteria has been performed for proteins included in the basis set and for those used for the testing of this CD spectra deconvolution method (Bolotina & Lugauskas, 1985). The rigid method was used for the assignment of secondary structure in the proteins discussed in Chapter IV, ribonuclease from *Bacillus intermedius* 7P (Grishina et al., 1989), porcine pepsin and pepsinogen (Makarov et al., 1991).

(3) Reference Spectra Corrected for Aromatic Contributions.

An iterative nonlinear squares procedure (Fraser & Suzuki, 1973) was used to obtain the peptide reference spectra for the structural types, assumed common to all proteins, and the twelve parameters describing the aromatic contribution, unique to each protein. Seven proteins were chosen for the basis set (M = 7): metmyoglobin (sperm whale); papain (Hg modified, papaya); lactate dehydrogenase (rabbit); lysozyme (hen

egg-white); ribonuclease A (bovine pancreas); subtilisin BPN' and thermolysin (*Bacillus thermoproteolyticus)*.

The non-linear system of equations (III-3) consists of m x M equations and contains [(m x r) + (4n) x M] unknowns. To obtain a solution the condition should be fulfilled :

$$m \times M \geq (m \times r) + (4n) \times M. \qquad \text{(III-5)}$$

Experimental CD spectra were obtained at 51 wavelengths from 200 nm to 250 nm at 1 nm interval (m=51). Three log-normal bands (n=3) and five secondary structural conformations (α-helix, antiparallel β-structure, parallel β-structure, β-turns and irregular) were considered (r=5).

Applying the non-linear least squares procedure to the system of Equations III-3, the minimum of the following function was sought :

$$S = \sum_{\lambda}\sum_{j}\left(\sum_{l} f_{jl}\, k_{\lambda l} + \sum_{i} A_{jl}\, \phi\,(\nu^{0}_{ji},\, \Delta\nu_{ji}, \rho_{ji}, \nu_{\lambda}) - [\Theta]_{\lambda j}\right)^{2} \qquad \text{(III-6)}$$

$$(\lambda = 1,\ldots,m;\ j = 1,\ldots,M;\ l = 1,\ldots,r\ ;\ i = 1,\ldots,n)$$

Since the parameters $k_{\lambda l}$ and A_{jl} enter into the functional linearly, their values were found at a fixed log-normal set (ν^{0}_{ji}, $\Delta\nu_{ji}$, ρ_{ji}, ν_{λ}) by solving a system of linear Equations (III-7) obtained by setting the derivatives $\partial S/\partial k_{\mu\chi} \equiv 0$ and $\partial S/\partial A_{tv} \equiv 0$:

$$\sum_{j}\sum_{l} f_{jl} f_{j\chi} k_{\mu l} + \sum_{j}\sum_{i} A_{jl}\, f_{j\chi}\, \phi\,(\nu^{0}_{ji},\, \Delta\nu_{ji}, \rho_{ji}, \nu_{\lambda}) = \sum_{j} [\Theta]_{\mu j} f_{j\mu} \qquad \text{(III-7)}$$

and

$$\sum_{\lambda}\sum_{l} f_{tl}\, \phi\,(\nu_{tv},\, \Delta\nu_{tv}, \rho_{tv}, \nu_{\lambda}) k_{\lambda l} + \sum_{\lambda}\sum_{i} \phi\,(\nu_{ti},\, \Delta\nu_{ti}, \rho_{ti}, \nu_{\lambda})\, \phi\,(\nu_{tv},\, \Delta\nu_{tv}, \rho_{tv}, \nu_{\lambda})\, A_{li} =$$

$$= \sum_{\lambda} \phi\,(\nu^{0}_{ti},\, \Delta\nu_{ti}, \rho_{ti}, \nu_{\lambda}) [\Theta]_{\lambda i}$$

where $\lambda,\ \mu = 1,\ldots,m;\ l,\ \chi = 1,\ldots,r\ ;\ j,\ t = 1,M;\ i,\ v = 1,n.$

After fixing the parameters $k_{\lambda l}$ and A_{jl}, the following system of non-linear Equations (III-8) was solved using the Gauss-Newton regularized iteration process, and the parameters for the log-normal bands were obtained :

$$[\Theta]_{\lambda j} - \Sigma_l f_{jl} k_{\lambda l}^{(l)} = \Sigma_i A_{jl}^{(l)} \phi (\nu^0_{ji}, \Delta\nu_{ji}, \rho_{ji}, \nu_\lambda) \quad \text{(III-8)}$$

After fixing the obtained parameters for log-normal bands, the system (Eq.III-7) is solved again to obtain a new approximation for the $k_{\lambda l}$ and A_{jl}. The process is repeated until the system of linear Equations (III-7) and nonlinear Equations (III-8) become self-consistent. Self-consistency is judged by the following criteria :

$$\nu^{0\,(k)}_{ji} - \nu^{0\,(k-1)}_{ji} \leq \epsilon_1 \; ; \quad \Delta\nu^{0\,(k)}_{ji} - \Delta\nu^{0\,(k-1)}_{ji} \leq \epsilon_2 \; ; \quad \rho^{0\,(k)}_{ji} - \rho^{0\,(k-1)}_{ji} \leq \epsilon_3 \quad \text{(III-9)}$$

where k is the number of the iteration and ϵ_1 ,ϵ_2 ,ϵ_3 are arbitrary small numbers set to bring the deviation from self-consistency of the systems (III-7) and (III-8) within the order of experimental error of CD measurements. It has to be pointed that due to uncertainty in the sign of the log-normal functions, several intermediate alternative solutions might be obtained through the procedure. Thus, only those reference spectra $k_{\lambda l}$ were considered, which were close to those predicted (Madison & Schellman, 1972) or experimentally obtained on model compounds (Greenfield & Fasman, 1969). The final reference spectra ($k_{\lambda l}$) obtained through those procedures will be referred to as CAC (corrected for aromatic contribution) or $k_{\lambda l}$ to distinguish them from those obtained from system III-2, which will be refered to as NCAC (not corrected for aromatic contribution) or $k'_{\lambda l}$.

(4) Manual for the Bolotina & Lugauskas (1985) Method.

The far-uv CD spectra of proteins of unknown conformation can be analyzed to obtain the fractions of secondary structural elements and the unique aromatic contribution. The method works in two steps. First, initial values are assigned for the structure content and parameters of log-normal bands. Second, all the variables are optimized to achieve the best fit of the calculated and experimental protein spectra.

The initial estimates are obtained through the following procedures:

1) The initial values (f'_{ij}) for structural content are calculated by deconvolution on the set of NCAC ($k'_{\lambda l}$) using Equation III-2.

2) A linear combination of products of f_{ij}' and the corresponding CAC is obtained, $\Sigma_l f_l' k_{\lambda l}$, and treated further as the first approximation of the peptide part of the protein spectrum.

3) The peptide part, $\Sigma_l f_l' k_{\lambda l}$, is subtracted from the protein spectrum, $[\Theta]_\lambda$. The result is treated as a first approximation for the non-peptide, i.e., aromatic part of the protein spectrum.

4) The aromatic part is deconvoluted into three log-normal curves, whose parameters are used as initial values in the final optimization procedure.

After the initial values for all parameters are obtained, the final optimization of all parameters together is carried out until the best fit of the calculated spectrum to the experimental is achieved. The restrictions used in this optimization for the parameters of the log-normal curves are the following, unless otherwise specified: $185 \leq \lambda^0_1 \leq 210$; $215 \leq \lambda^0_2 \leq 226$; $226 \leq \lambda^0_3 \leq 235$; $0.1 \leq \Delta\nu \leq 6.0$; $0.1 \leq \rho \leq 1.9$.

3. Extension of the Bolotina & Lugauskas (1985) method.

a. Combining CD and infrared data.

Analysis of infrared absorption spectra are an alternative to CD methods in estimation of protein secondary structure in solution (Dong et al., 1990, 1992; Kalnin et al., 1990; Sarver & Krueger, 1991a,b; Venyaminov et al., 1990). CD and infrared methods have specific advantages in analyzing for α-helical and β-structural conformations, respectively (Sarver & Krueger, 1991a,b).

I explored the possibilities of combining the far uv CD (200 - 250 nm) and infrared (1600-1700 cm^{-1}) data (Sarver & Krueger (1991a,b) in estimating the protein secondary structure content. The set of proteins was chosen to test the extended CD/IR reference spectra from the proteins for which IR spectra were reported (Krueger et all, 1991). The chosen test set includes: RNAse S, hemoglobin, elastase, parvalbumin, prealbumin, triose phosphate isomerase and α-chymotrypsin. Reference spectra for the infrared interval were calculated from a basis set which included: myoglobin, lysozyme, papain, trypsin inhibitor, cytochrome c, concanavalin and pepsinogen.

The results are summarized in Table III-2 for α-helical, β-sheet and β-turn conformations. The results on deconvolution on different sets of various spectra (Table III-2) were compared in terms of the Pearson correlation coefficient (r_p), which is a measure of coincidence of calculated (in our case obtained by deconvolution) and experimental (x-ray) results for a set of proteins. The Pearson correlation coefficient is calculated as:

$$r_p = \{\Sigma x_i y_i - (1/n)\Sigma x_i \Sigma y_i\}/\{[\Sigma x_i^2-(1/n)(\Sigma x_i)^2][\Sigma y_i^2-(1/n)((\Sigma y_i)^2]\}^{1/2}, \quad \text{(III- 10)}$$

Table III-2. Performance of combined CD/IR analysis. Comparison of Pearson correlaltion coefficients.

Conformation	NCAC	CAC	NCAC/IR	CAC/IR	CAC/IRs
α-helix	.99	.99	.98	.95	.97
β-structure	.75	.72	.99	.79	.84
β-turns	.48	.53	.25	.55	.38

NCAC - CD reference spectra not corrected for aromatic contributions.
CAC - CD reference spectra corrected for aromatic contributions.
IR- infra-red reference spectra.
IRs- infra-red reference spectra with account for side-chain contributions.

where x_i and y_i are the values of secondary structure contents obtained from CD spectra and x-ray crystallography, respectively; n is the number of proteins in the test set. A value of r_p equal to 1 corresponds to 100 % correlation, 0 to random agreement and -1 to 100 % anticorrelation.

Deconvolution has been performed on CD reference spectra not corrected for aromatic contributions (NCAC) and corrected for aromatic contributions (CAC), and for combined CD+IR extended reference spectra, where the CD part was corrected for aromatic contributions (CAC/IR) or was not corrected (NCAC/IR). Finally, side-chain contributions were considered in both CD and IR spectra (CAC/IRs). Trp and Tyr do not contribute to the Amide I region. The IR contributions of the side chains of Lys, Arg , Asn, Gln, Phe and His were taken into account. As the infrared spectra are actually absorbance spectra, the contributions of the individual side chains are likely to be additive. Side chain infrared spectra for basis and testing sets were taken from published data (Venyaminov et al, 1991).

As one can see no difference in correlation coefficient (r_p) has been obtained for the α-helical conformation. The correlation is close to unity in all cases. Extension of the reference set (NCAC/IR) did lead to a significant improvement in the correlation coefficient for β-structure, but very poor results were obtained for β-turns in that case (r_p= 0.25). Accounting for aromatic contributions in CD spectra (CAC and CAC/IR) gives the highest correlation for β-turns obtained in the method (r_p=0.53 and 0.55, respectively). The case in which we account for aromatic contributions in both CD and IR spectra gives probably the best solution, as the correlation for β-structure is

sufficiently high (r_p= 0.84), whereas correlation for the β-turns is not too low (r_p= 0.38).

Overall, I conclude that the approach of estimating secondary structure content by CD through deconvolution on fixed reference spectra accounting for side-chain contributions (Chapter III.2.e) has probably reached its limit of accuracy. Extending the reference spectra does not improve results significantly.

b. Standard Account of Aromatic Contribution.

Recently, we developed a new interpretation of the Bolotina & Lugauskas (1985) method (Morozov et al., 1994). In the Bolotina & Lugauskas method the secondary structure content and the unique, mostly aromatic, protein CD, referred to in the following as aromatic input (AI), are obtained for a particular protein by finding the minimum of the functional :

$$Q=\sum_{\lambda} [(\sum_{i} k_i f_{i\lambda} + \sum_{j} L_{j\lambda}) - \Theta_{\lambda}]^2. \qquad \text{(Eq.III-21)}$$

where index (λ) is the current wavelength (or wavenumber) from 200 nm to 250 nm, usually with a step of 1 nm; (i) refers to a secondary structure element (α-helix, antiparallel and parallel β-structure, β-turns and unordered conformation); j (j=1,..,3) is the index of the log-normal functions ($L_{j\lambda}$), the sum of which model the AI; $f_{i\lambda}$ is a reference spectrum of one of the secondary structure elements; k_i are the secondary structure contents; and Θ_{λ} is the experimental CD spectrum of the protein.

In order to minimize the functional (Eq.III-11), some initial parameters for all three $L_{j\lambda}$ are set as described in Chapter III.2.e. First, preliminary values for the

secondary structure content, k^0_i , are found using the standard least-square method and reference spectra $f_{i\lambda}$ (NCAC). Then a difference spectrum (D_λ) is calculated using k^0_i and $f_{i\lambda}$ (CAC):

$$D_\lambda = F_\lambda - \sum k^0_1 \ f_{i\lambda}. \qquad \text{(Eq.III-12)}$$

The D_λ is further deconvoluted into three log-normal curves, $L_{j\lambda}$. The amplitude, wavelength, bandwidth and asymmetry for each $L_{j\lambda}$ can be assigned automatically or manually after visual inspection of D_λ, and then optimized so that the sum of the three (or any number of) log-normal functions fit the first approximation of D_λ. Then the minimum of the functional Eq.III-11 is sought using the obtained values for the log-normal functions, and the k^0_i and thereby the final values for all parameters are obtained.

We (Morozov et al., 1994) searched for the possibilities of eliminating the step of searching for initial parameters for the log-normal functions by assigning some standard initial values that would converge to the solution within a reasonable error limit. In this way, the whole algorithm could be accelerated.

The difference spectrum D_λ was shown (Bolotina & Lugauskas, 1985) to be represented as the sum of three log-normal curves, their maxima being located in three nonoverlapping spectral regions: 185 - 210 nm, 215 - 226 nm, and 225 - 235 nm. Based on this and on the results of deconvolution of 16 protein spectra (Bolotina & Lugauskas, 1985), we formed a set of standard initial values for the parameters of $L_{j\lambda}$ (Table III-3) such that for any protein the RMS between the D_λ and its approximation with the sum of three log-normal functions does not exceed 15 %. The only parameter that is left for manual input after visual inspection of D_λ is the sign of the log-normal bands.

Table III-3. Initial values and their variation limits for the parameters of log-normal bands.

Number	ν (kK)	$\lvert A \rvert$	$\Delta\nu$ (kK)	ρ
1	43.0±1.5	0.25	2.2±1.0	0.9±0.4
1	45.0±1.5	0.65	3.0±1.5	1.1±0.5
1	49.5±1.5	1.00	2.5±1.5	0.9±0.5

ν (kK) is the location of the band; $|A|$ is the intensity of the band normalized to the $|D_\lambda|_{max}$, $\Delta\nu$ is the half bandwidth at e^{-1} of the amplitude of the maximum, and ρ is the asymetry of the band. See text for explanations.

The stability of the standard solution obtained for the initial values of $L_{j\lambda}$ has been evaluated both in calculation of (CAC) from the spectra of the basis proteins after subtraction of AI, and in determination of the secondary structure content (k_j) from protein CD spectra. Initial CAC were taken from Bolotina & Lugauskas (1985) and further approximations of CAC were obtained by fitting the reference spectra and the log-normal functions to the CD spectra of basis proteins with known secondary structure (lysozyme, hen egg-white; myoglobin, sperm whale; papain, papaya; ribonuclease A, bovine; subtilisin BPN′; thermolysin; lactate dehydrogenase).

Variation of the standard initial values within 5 % resulted in 2.6% variation of the α-helical spectrum; 19.3 % for the β-structure (parallel and antiparallel); 7.8 % of the spectrum of antiparallel β-structure; 7.0 % of the parallel β-structure; 9.4 % the β-turn spectrum; 12.2% of the spectrum for unordered conformation. The final spectra obtained for both parallel and antiparallel β-structure were of slightly larger amplitude when CAC, *ca.* -2.3 x 10^3 deg cm^2/dmol at 217 nm instead of *ca* -1.8 x 10^3 deg cm^2/dmol for the antiparallel β-structure, and *ca.* -3.6 x 10^3 deg cm^2/dmol at 217 nm instead of *ca* -2.4 x 10^3 deg cm^2/dmol for the parallel β-structure. A slight increase in the amplitude of the negative band of the β-turn spectrum at 204 nm has been observed from *ca.*-1.9 x 10^3 deg cm^2/dmol *to ca* -2.2 x 10^3 deg cm^2/dmol. No significant changes in the spectra of the α-helical or unordered conformation were obtained.

Determination of the secondary structure from CD has been analyzed for nine proteins excluded from the basis set (insulin, nuclease, trypsin inhibitor, carboxypeptidase, γ-crystallin, cytochrome C, concanavalin A, parvalbumin,

glyceraldehyde-3-P-dehydrogenase). Deconvolution of the protein spectra has been performed on four or five reference spectra by accounting for the parallel and antiparallel β-structure separately or together. The following values of the Pearson correlation coefficients have been obtained. Accounting for four structures: 0.87 for α-helices, 0.65 for β-structure, 0.32 for β-turns and 0.41 for unordered conformation. Accounting for five structures: 0.78 for α-helices, 0.90 for antiparallel β-structure, 0.93 for parallel β-structure, -0.14 for β-turns and for 0.31 unordered conformation. The correlation coefficients obtained in the original Bolotina & Lugauskas approach were: 0.98 for α-helices, 0.96 for the total β-structure, 0.80 for antiparallel β-structure, 0.52 for parallel β-structure, 0.33 for β-turns and 0.72 for unordered conformation. As can be seen, the correlation coefficients obtained with standard values are lower than if the original algorithm is applied, other than for antiparallel and parallel β-structure. This result indicates that the new reference spectra for the antiparallel and parallel β-structure are more adequate.

Overall, the standard initial values obtained for the parameters of the log-normal functions may be used in the Bolotina & Lugauskas (1985) method with the expectation of reasonable accuracy for the α-helical and β-structure conformations, and a simplified calculation procedure.

4. Theory and Experiment in Description of Aromatic CD in Globular Proteins.

In this section I compare the results of calculations of exciton Trp CD for some globular proteins with experimental CD that can be explained by aromatic contributions.

a. Dihydrofolate Reductase.

Dihydrofolate reductase (DHFR) is an enzyme important in the metabolism of the folates in the cell. DHFR is involved in cell proliferation through the synthesis of thymine nucleotides, and hence of DNA. Some recent studies in cancer research have been concentrated on folate analogs such as methotrexate, which inhibit DHFR and are currently used in the treatment of some forms of acute leukemia (Matthews & van Holde, 1990).

Recently the *E.coli* DHFR gene was cloned, and the wild-type protein and some mutants at the active site residues were expressed (C. R. Matthews, Pennsylvania State University). We obtained CD spectra in the range 185-350 nm of the wild-type and Trp74→Leu mutant from C. R. Mathews (Pennsylvania University). The spectra in the far uv are shown in Figure III-1. As can be seen, the intersection of the spectra of the wild-type and mutant implies a couplet-type difference spectrum (Fig. III-1). The near uv spectra of the wild-type and mutant differ significantly (Fig.III-2). Kuwajima et al. (1991) proposed that the exciton coupling interaction between Trp-74 and Trp47 might account for the differences in the far-uv spectra.

I calculated the difference spectrum (Fig. III-1), which in fact has a classic couplet shape in the interval 210-240 nm centered around 225 nm, which can be attributed to the B_b transition in indole (Chapters I.2.c). Several bands can be seen below 210 nm, which might be attributed to the B_a , V and VI transitions in indole (Fig. III-1; Chapters I.2.c). The near uv spectra (Fig.III-2) are complicated. The difference spectrum between wild type and W74L DHFR shows a broad negative band with a minimum

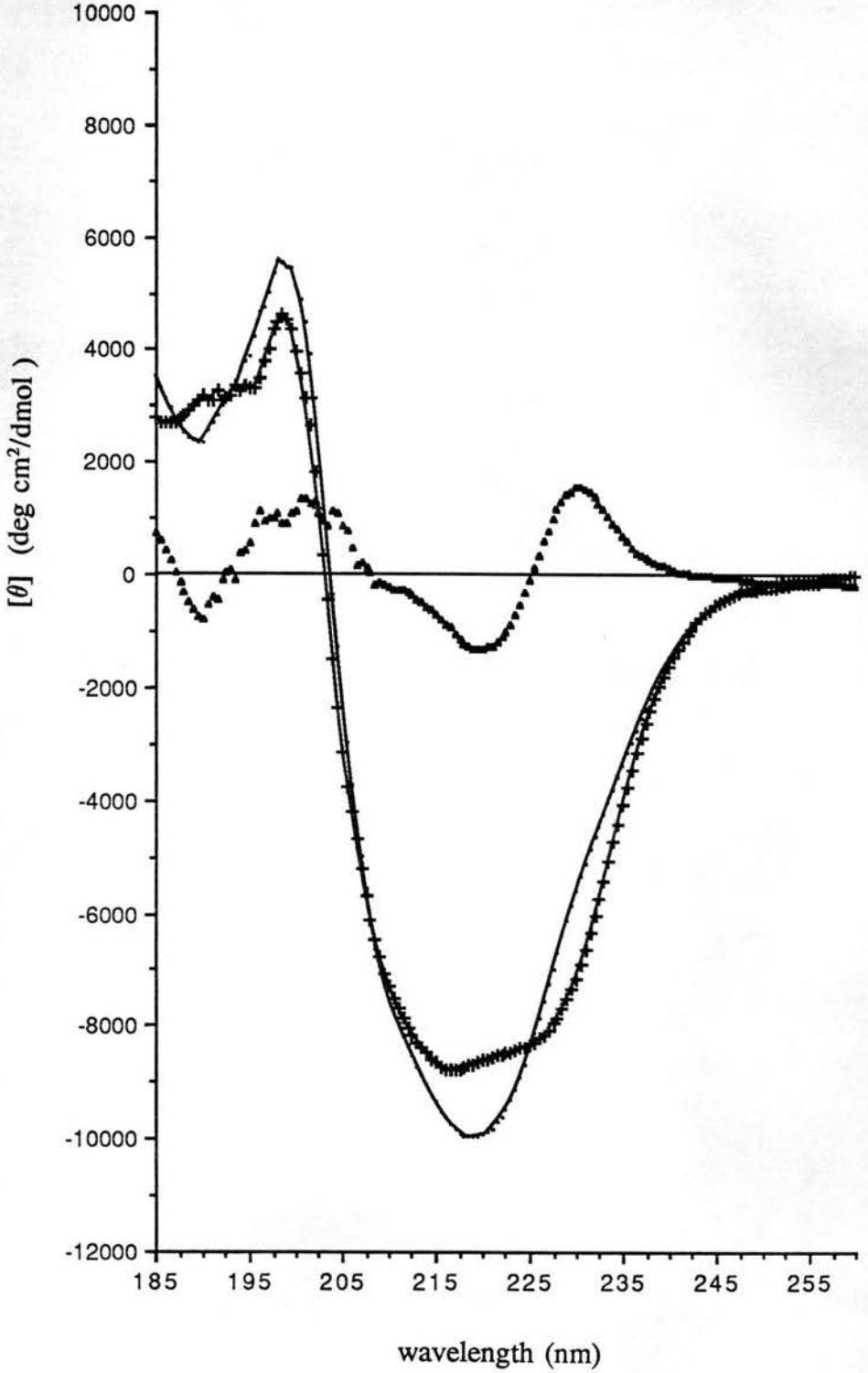


Figure III-1. Experimental far-uv CD of the wild-type (-) and W74L mutant (+) DHFR. Difference spectrum between W74L mutant and wild-type CD (▲).

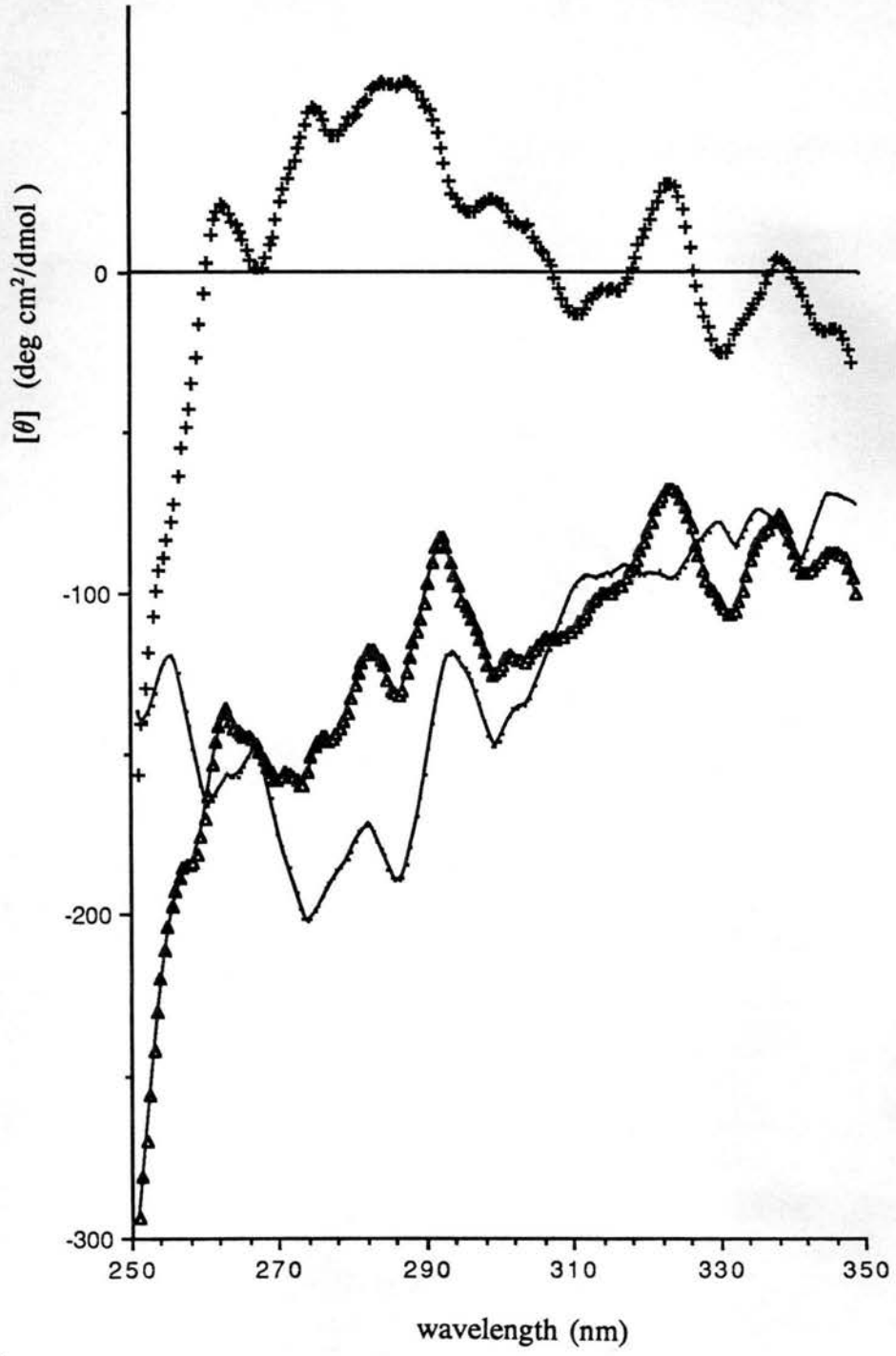


Figure III-2. Experimental near-uv CD of the wild-type (▵) and W74L mutant (+) DHFR. Difference spectrum between W74L mutant and wild-type CD (-).

around 280 nm. Both Trp and Tyr transitions could contribute to that difference spectrum. Sharp bands around 260 nm imply participation of Phe as well.

I estimated the secondary structure content of wild-type and mutant DHFR from CD spectra using two methods: the Bolotina & Lugauskas (1985) method (Chapter III.2.e) and the self-consistent approach (Sreerama and Woody, 1992). The self-consistent approach performs deconvolution of the experimental CD spectrum with a variable set of basis proteins (Manavalan & Johnson, 1987). The unique feature of the self-consistent approach is that it uses the secondary structure content of the basis protein with the most similar CD spectrum to that of the unknown protein as the initial guess in the deconvolution.

The secondary structure estimates for the wild-type DHFR performed through both deconvolution methods (Table III-4) are close to each other and agree with the results of the x-ray diffraction (Bolin et al., 1982) with respect to the α-helix and β-turn content. Both CD methods underestimated the β-sheet content by about 10% and overestimated the unordered conformation by a corresponding amount. There are no x-ray data for the W74L mutant of DHFR. Our estimates of the secondary structure content for the mutant protein showed a 7 - 8 % decrease in β-sheet (Table III-5). In fact Trp74 is submerged in the hydrophobic core in wild type DHFR (Bolin et al., 1982) and its replacement by Leu might account for the slight decrease in β-structure implied by the CD analysis. Parameters for the log-normal curves describing the unique protein CD, which is considered to arise mostly from aromatic side chains (Bolotina and Lugauskas, 1985), for wild-type (Table III-4) and W74L mutant proteins are shown (Table III-5).

Table III-3. Secondary structure content of wild type DHFR.

Method	Conformations (%)				Parameters for log-normal bands				
	α-helix	β-sheet	β-turns	unordered	N	λ_0 (nm)	$[\theta]_{max}$[4]	$\Delta\nu$ (Kk)	ρ
x-ray[1]	31	35	20	14					
CD, BL[2]	33	25	16	26	1	204.5	7.79	2.00	0.67
					2	225.2	-.50	2.00	0.60
					3	230.4	2.29	1.20	0.84
CD, SC[3]	31	24	22	24					

1. Bolin et al., 1982.
2. Bolotina and Lugauskas, 1985.
3. Self Consistent Method (Sreerama and Woody, 1993)
4. the units are (10^3 deg cm^2/dmol)

Table III-4. Secondary structure content of W74L mutant DHFR.

Method	Conformations (%)				Parameters for log-normal bands				
	α-helix	β-sheet	β-turns	unordered	N	λ_0 (nm)	$[\theta]_{max}$[5]	$\Delta\nu$ (Kk)	ρ
CD, BL	34	17	17	32	1	203.8	9.72	2.63	0.54
					2	225.6	-.86	2.42	0.15
					3	230.4	2.3	0.90	1.88
CD, SC	34	17	22	28					
CD, BL[4]	33	25	16	26	1	204.9	6.88	2.00	0.60
					2	226.2	-.66	0.40	1.03
					3	230.4	.57	0.80	1.03

1. Bolin et al., 1982.
2. Bolotina and Lugauskas, 1985.
3. Self Consistent Method (Sreerama and Woody, 1993)
4. Initial estimates for the mutant secondary structure taken as for the wild type DHFR.
5. the units are (10^3 deg cm^2/dmol)

Theoretical CD calculations for the Trp47 - Trp74 pair and for the system of all five DHFR Trp have been performed. The Trp47 - Trp74 interaction results in a positive couplet with an amplitude of about 4 x 10^3 deg cm^2 $dmol^{-1}$, with the negative extremum at 218 nm and the positive extremum at 230 nm. The position of the 225 nm couplet coincides with the difference spectrum (Fig.III-3). The CD difference spectrum between the wild-type and W74L is likely to be largely attributable to the interaction between Trp47 and Trp74. The calculated Trp spectrum of the W74L mutant has negligible CD (Fig.III-3). Trp 74 is located in close proximity to Trp47 (6.14 Å center-to-center), which explains the dominance of this couplet band in the Trp CD of DHFR. The theoretically predicted amplitude of the Trp band is about three times higher than that of the experimental difference spectrum between wild type and W74L mutant DHFR.

b. Concanavalin A.

The secondary structure of concanavalin A is dominated by β-structure, which constitutes about 50 % of the protein. Thus the CD of concanavalin is expected to have a minimum around 217-218 nm, corresponding to the minimum in β-structure CD. The observed minimum in the CD of concanavalin A is at 225 nm (Bolotina & Lugauskas, 1985; Herman et al., 1978). Various authors (Bolotina & Lugauskas, 1985; Herrmann et al., 1978) attempted to explain the anomaly in the concanavalin A CD spectrum by aromatic contributions and concluded that accounting for aromatic contribution significantly improves the estimation of the secondary structure of concanavalin A from CD (Bolotina & Lugauskas, 1985).

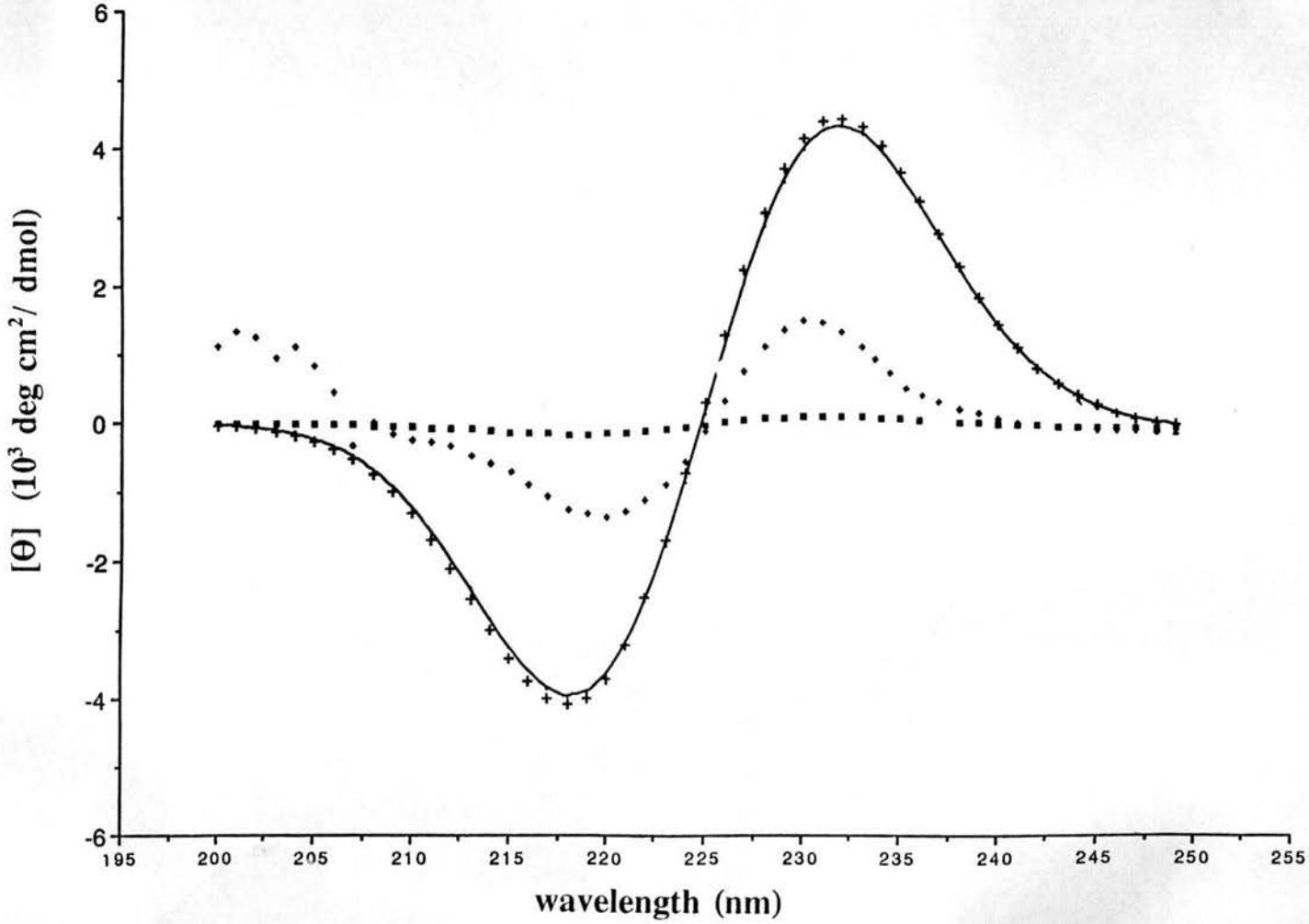


Figure III-3. Calculated B_b Trp CD of DHFR considering all five Trp (+), and excluding Trp74 (■). B_b Trp CD of the Trp47-Trp74 pair (-). Experimental difference spectrum of the wild type and W74L mutant DHFR (+).

I calculated the B_b Trp CD of concanavalin A (Figure III-4). The CD of the Trp88 - Trp182 pair in concanavalin gives one of the largest CD predicted during the scan of the CD of Trp pairs (Fig. II-2) in the Brookhaven Protein Databank (PDB). The Trp88 - Trp182 pair dominates in concanavalin Trp CD. The CD calculated considering all eight Trp of concanavalin almost coincides with that of this pair (Fig.III-4).

The positive Cotton effect of the calculated spectrum of concanavalin A correlates rather well with the sum of two Gaussian functions (Fig.III-4) assigned to aromatic CD by Herrmann et al. (1978) while performing curve fitting of the experimental CD in the range 200-250 nm to Gaussian components.

Bolotina & Lugauskas (1985) accounted for the shift of the concanavalin spectrum from 217-218 nm to 225 nm with a large log-normal band around 200 nm (Fig.III-5). The calculated concanavalin B_b Trp CD and CD of the Trp88-Trp182 pair show some correlation with the two minor log-normal bands (Fig.III-5). The B_a, V and VI indole transitions at energies higher than the B_b (5.51 eV) are less intense than the latter (Table I-2; Fig.II-19). Thus the band around 200 nm is unlikely to be due primarily to aromatic contributions, at least from Trp. The large positive CD in the interval 200-220 nm obtained by Bolotina & Lugauskas (1985) could be due to CD of distorted secondary structures in concanavalin (discussed in Chapter III.4), such as twisted β-sheets (Manning et al., 1988) or to the CD of β-turns (Woody, 1974; Brahms & Brahms, 1980).

c. Bovine α-chymotrypsin and Chymotrypsinogen A.

Chymotrypsin and chymotrypsinogen are attractive models for studying conformational and stereochemical changes involved in enzyme activation. The enzyme

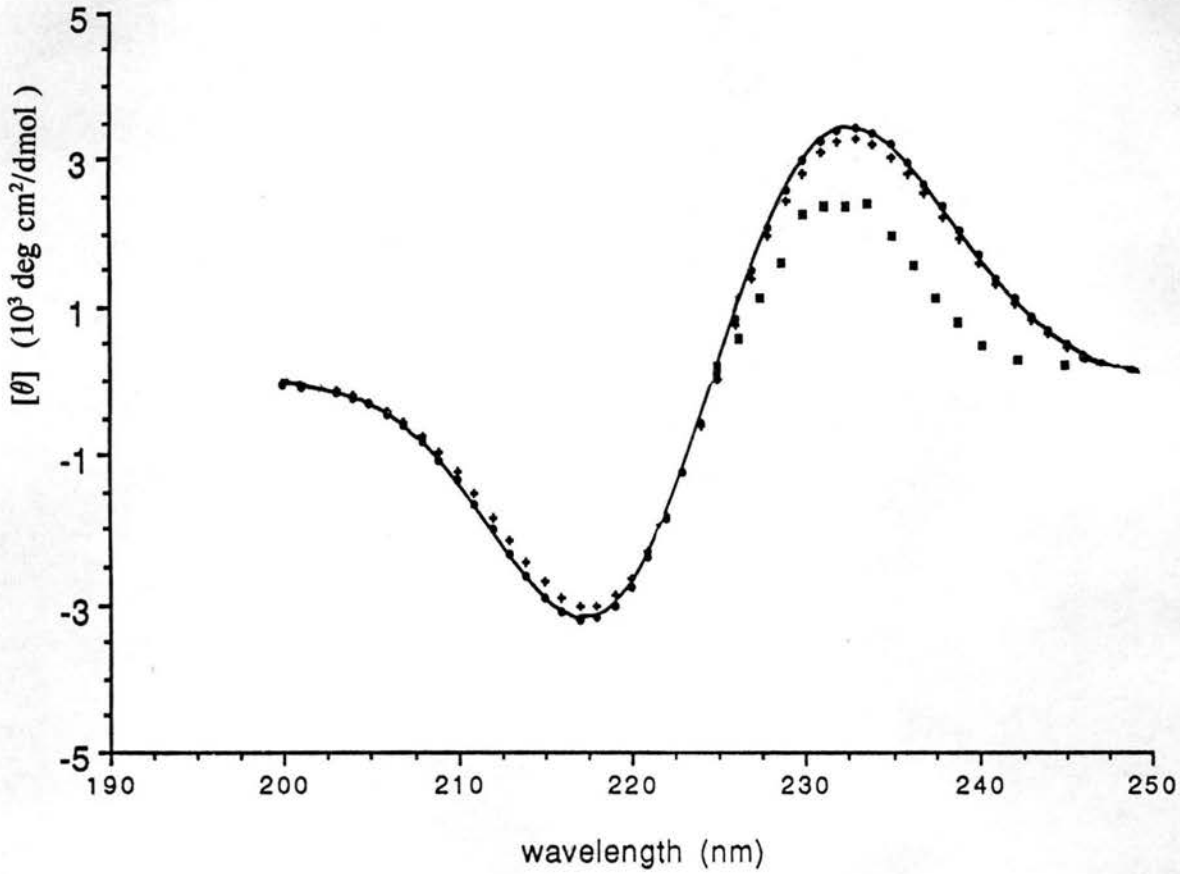


Figure III-4. Calculated B_b Trp CD of concanavalin accounting for all eight Trp (•). Calculated B_b Trp CD of the Trp88-Trp182 pair (+). Aromatic CD (■) obtained by curve-fitting of experimental concanavalin spectrum to a set of Gaussian components (Herrmann et al., 1978).

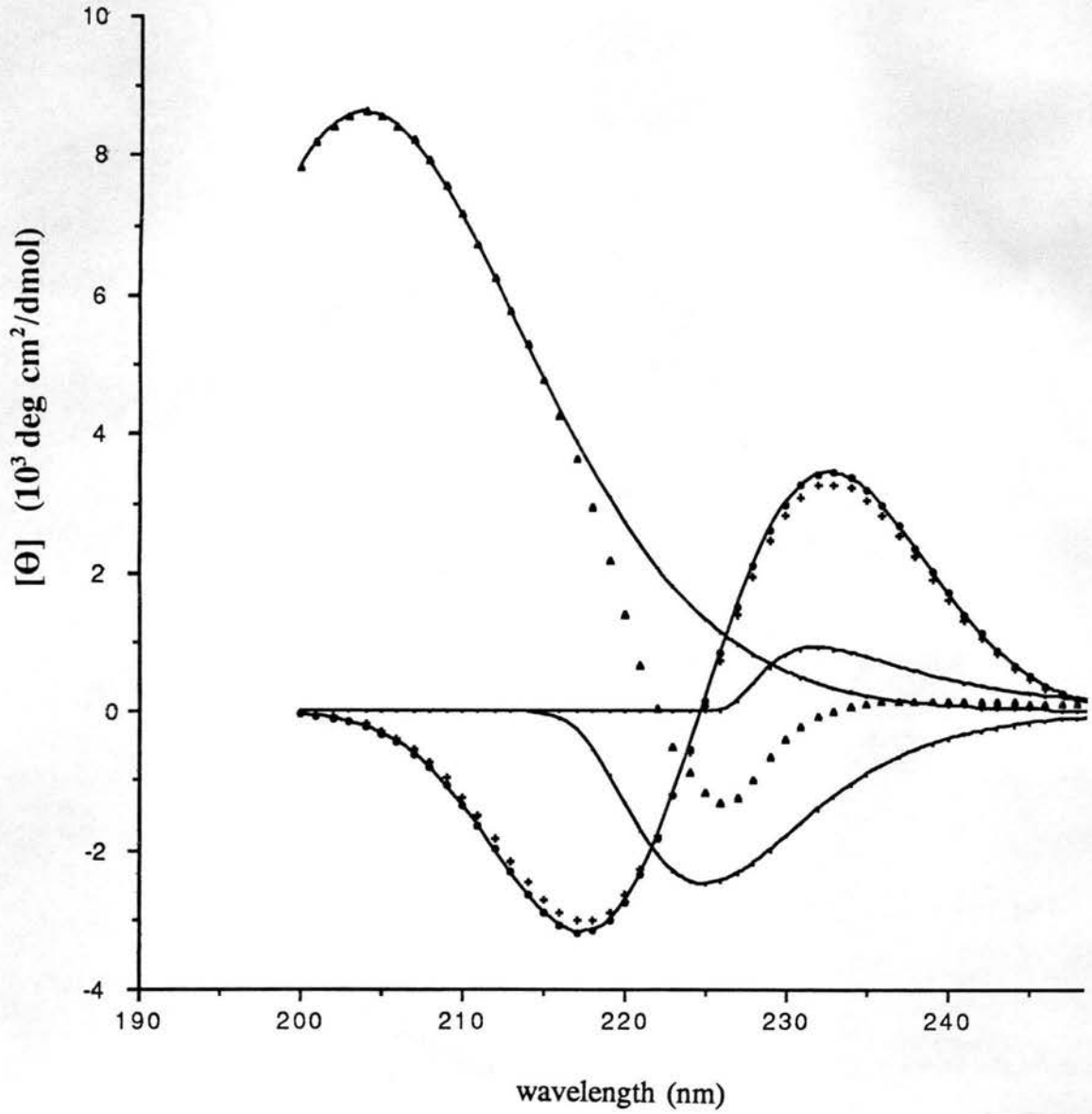


Figure III-5. Aromatic contribution in the CD spectrum of concanavalin A determined by the Bolotina & Lugauskas (1985) method as sum of three log-normal functions (▲). The three log-normal functions are shown in simple lines. Calculated B_b Trp CD of concanavalin accounting for all eight Trp (•); B_b Trp CD of the Trp88-Trp182 pair (+).

and zymogen in this case are almost identical in primary sequence. Formation of α-chymotrypsin from chymotrypsinogen proceeds upon excision of two peptides: Ser14 - Arg15 and Thr147 - Asn148 (Hess, 1971). The CD spectra of bovine α-chymotrypsin and chymotrypsinogen A show noticeable differences in the far uv (Fig.III-6; McConn et al., 1969; Cantor & Timasheff, 1982), while the near-uv CD spectra of the native enzyme and zymogen are very similar (Gertler et al., 1974). On the basis of ORD (optical rotatory dispersion, Chapter I.1.a) and CD studies it was originally concluded that some secondary and tertiary structure changes are involved in activation of chymotrypsin. It has been found that changes in CD upon formation of α-chymotrypsin from chymotrypsinogen A are identical to those in α-chymotrypsin upon change of pH from 7.0 to 10.5 and follow the ionization of a single group, the α-amino group of Ile16 (Oppenheimer et al., 1966; McConn et all., 1969). In α-chymotrypsin, the α-amino group of Ile16 forms an ion pair with Asp194 (Birktoft & Blow, 1972), which holds Ser195, important in substrate binding, in the appropriate position. In the absence of the Ile16 - Asp194 ion pair, as in chymotrypsinogen with the intact peptide chain at Ser14 - Arg15, or at higher pH when the Ile16 α-amino group is deprotonated, the Asp194 carboxylate protrudes into the solvent, disrupting the adjacent area of the substrate binding pocket (Sigler et al., 1968; Wang et al., 1985).

The most significant difference between the enzyme and zymogen CD spectra is the negative band around 230 nm in chymotrypsin, which is absent in chymotrypsinogen (Fig.III-6). Since the energy of this transition is close to the Cotton effect of the $n \rightarrow \pi^*$ transition in the α-helix at 222 nm (Sears & Beychok, 1973), it was suggested that

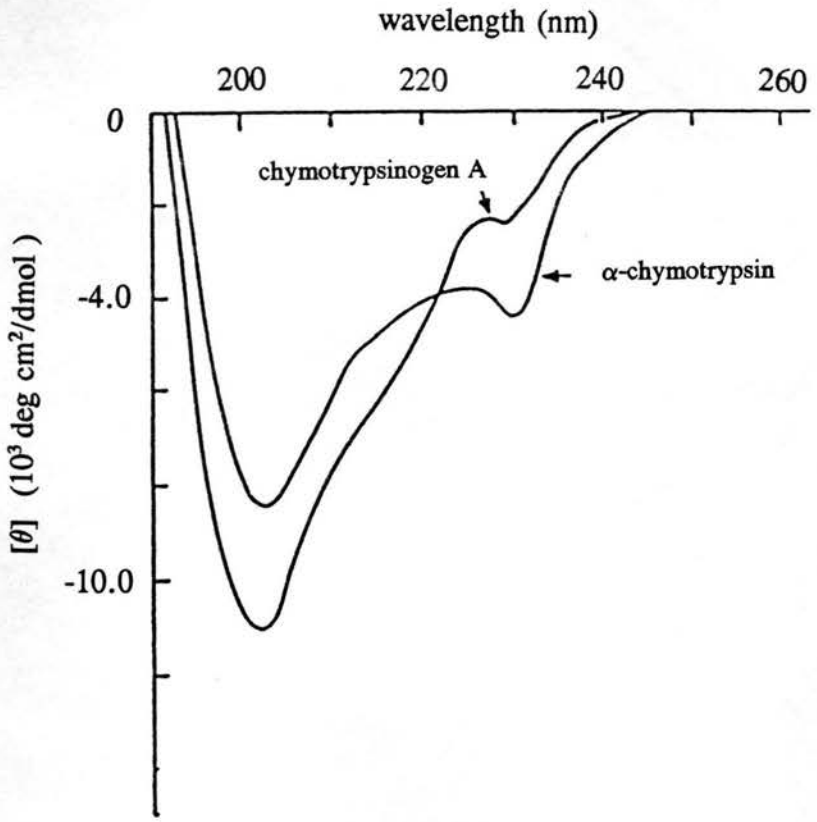


Figure III-6. Experimental CD spectra of bovine α-chymotrypsin and chymotrypsinogen A in the far uv. (Adapted from Cantor & Timasheff, 1982.)

chymotrypsinogen activation proceeds with an increase in α-helical content (Biltonen et al., 1965), which might include as many as 10-15 residues (Fasman et al., 1966). Chen et al. (1972) applied a deconvolution analysis to the CD and ORD spectra of chymotrypsin and chymotrypsinogen using a fixed set of reference spectra (Chapter III.2.a), and concluded that the α-helical content does not change upon activation, whereas the content of β-structure increases by about 15-20 %. No such changes were found upon examination of x-ray data. Only one region in the molecule has been found where extension of an α-helix by one turn (residues 169 - 173) might occur upon formation of α-chymotrypsin (Wright, 1973a,b). X-ray structures of α-chymotrypsin (Birktoft & Blow, 1972; Fujinaga et al., 1987) and chymotrypsinogen (Freer et al., 1970) revealed extensive similarity between the enzyme and zymogen. Significant differences were found only in the conformation of the segments that form the binding pocket (Wright, 1973a,b).

The similarity of the difference spectra between α-chymotrypsin and chymotrypsinogen (Fig.III-7) to the spectrum of N-acetyl-L-tryptophanamide (Shiraki, 1969) indicated the possibility that the band at 230 nm in chymotrypsin might be attributed at least in part to a Trp. Cantor & Timasheff (1982) suggested that Trp141 may be responsible since it hydrogen bonds with Asp194 in the zymogen and its conformation undergoes a substantial change when the Asp194 carboxylate moves about 12 Å towards N Ile16 upon activation (Birktoft & Blow, 1972).

I calculated the Trp CD that would result from exciton interactions between the B_b transitions of all eight Trp in chymotrypsinogen and chymotrypsin (Fig.III-7) and

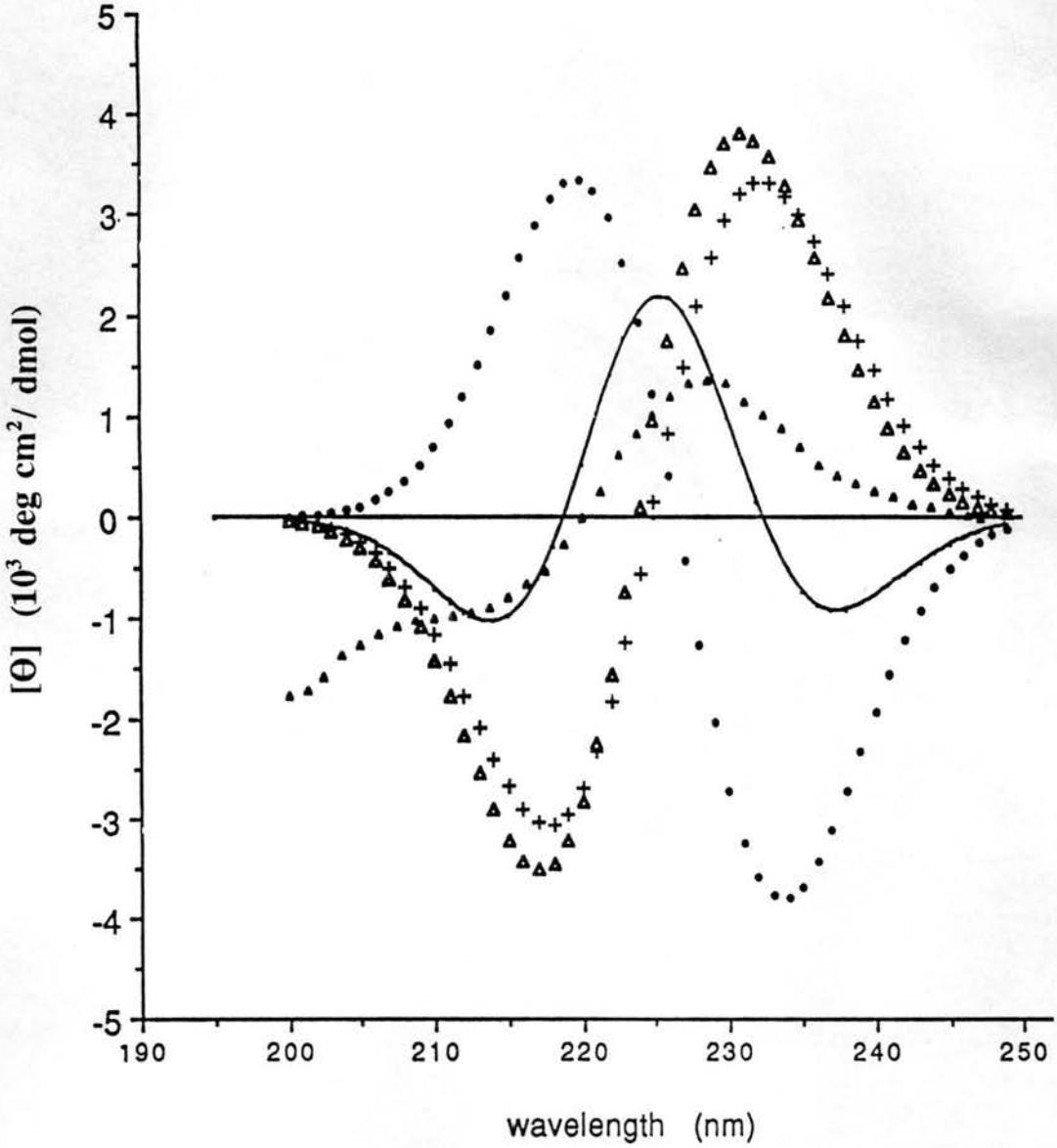


Figure III-7. Calculated B_b Trp CD of all eight Trp in monomer chymotrypsinogen A (-) and α-chymotrypsin (•). Difference spectrum between chymotrypsinogen A and α-chymotrypsin (▵). Calculated B_b Trp CD of the Trp172-Trp215 pair in monomer chymotrypsinogen A (+). Experimental CD difference spectrum between chymotrypsinogen A and α-chymotrypsin (▲).

between all the Trp pairs separately. Coordinates used in calculations for the Trp indole rings were taken from the Brookhaven Protein Data Bank (PDB) for bovine α-chymotrypsin (1cho) and chymotrypsinogen A (1chg). I found that the predicted Trp CD of chymotrypsin and its zymogen are markedly different (Fig.III-7). The chymotrypsin spectrum is a negative couplet of significant magnitude, *ca.* -3,500 deg cm^2/dmol at the long-wavelength extremum, whereas the chymotrypsinogen spectrum is a triplet with a positive central band at 225 nm *ca.* 2,000 deg cm^2/dmol, and negative bands of similar intensity, *ca.* -1,000 deg cm^2/dmol, on either side.

I obtained a good correlation between the experimental difference spectrum between chymotrypsin and chymotrypsinogen, and the difference between their predicted Trp B_b spectra (Fig.III-7), especially in the 210-235 nm region. The experimental difference CD below 210 nm probably contains peptide contributions, although it still matches the sign of the calculated difference spectrum.

I compared the predicted difference spectrum with CD of Trp pairs in chymotrypsinogen calculated separately, and found a remarkable coincidence with the pair Trp172 - Trp215 (Fig.III-7). This pair is the closest in chymotrypsinogen (4.9 Å) and provides the largest CD contribution. The CD of the Trp172 - Trp215 pair differs significantly from the overall Trp CD, unlike the case of the dominant Trp pairs in concanavalin and DHFR described above. In both chymotrypsin and chymotrypsinogen, other Trp-Trp exciton interactions contribute significantly to the spectrum, but those contributions do not change significantly upon activation. Thus, the difference between

the predicted CD of chymotrypsin and chymotrypsinogen is due to the change in CD of one Trp pair.

The amplitude of the predicted difference spectrum exceeds that of the experimental spectrum by a factor of two. The explanation for this discrepancy may be that the value of the B_b dipole transition moment we used (5.7 D; Table I-2) is too large, or that other groups, peptide or other aromatic side chains, might contribute to the difference spectrum. In particular, the interaction between the B_b transition in Trp, which we calculated at 5.51 eV or 225 nm, and the L_a transitions in Tyr at 230 nm might give significant CD contribution in the 210-240 nm interval. Considering all six indole transitions, as was shown in Chapter II, does not have any significant effect on the apparent intensity of the couplet in the B_b region.

The difference in the crossover between the predicted and experimental difference spectra (Fig.III-7) might be due to the somewhat arbitrary choice of 225 nm as the wavelength for the B_b band. If the Trp B_b bands Trp172 and Trp215 were centered at 220 nm, as a result of accounting for Trp exposure to the solvent (Chapter II.3.b.(1).), the experimental and predicted position of the bands would coincide (Fig.III-7).

I compared the conformations of Trp side chains in chymotrypsin (1cho) and chymotrypsinogen (1chg) (Fig.III-8). The eight Trp residues in monomer chymotrypsinogen (1chg, Freer et al., 1970) and chymotrypsin (1cho, Fujinaga et al., 1987) have been overlaid to give optimal superposition of all corresponding side-chain and backbone atoms. The RMS deviation between Trp residues in chymotrypsin and monomer chymotrypsinogen is 1.13 Å. The differences in calculated exciton Trp CD of

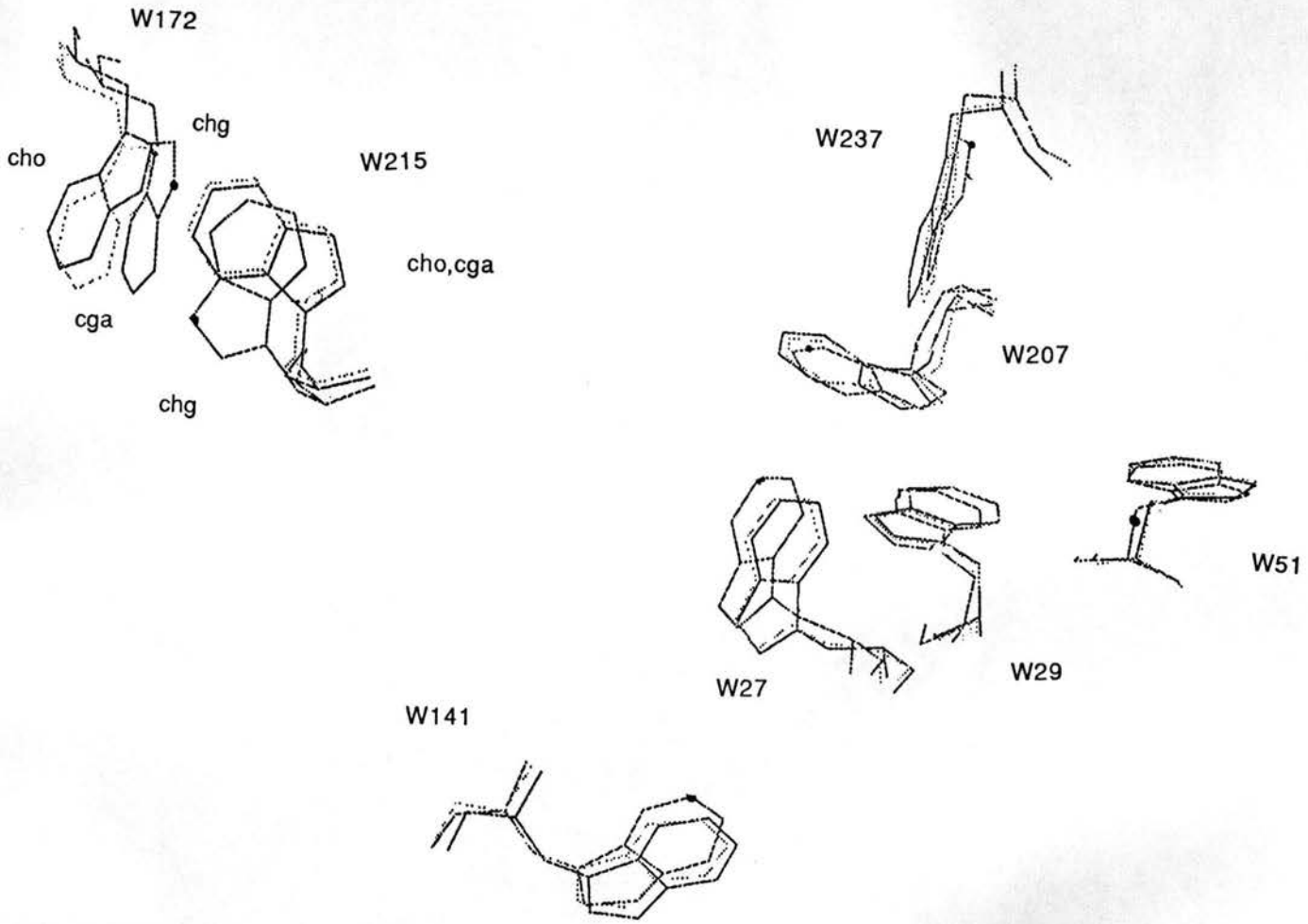


Figure III-8. Conformations of Trp residues in bovine α-chymotrypsin, 1cho (---); monomer chymotrypsinogen A, 1chg (-); molecule 1 of a dimer crystal of chymotrypsinogen, 2cga (...).

the monomeric chymotrypsinogen and chymotrypsin result from the differences in conformations of corresponding Trp residues. The most significant of these differences, both in the conformational sense and with respect to the effect on CD, are the tilt at Trp172 and the 90° twist of Trp215 (Fig.III-8), which corresponds to an approximately 180° change in the angle χ_2 (the torsional angle about the C_β-C_γ bond). The center-to-center distance between Trp172 and Trp215 also changes. It is 4.9 Å in chymotrypsinogen and 6.1 Å in chymotrypsin. The conformations of the other six Trps are very similar.

Trp215 in chymotrypsinogen belongs to the region containing residues from 214 to 220, which are involved in the formation of the substrate recognition site upon activation of the enzyme. Rotation of Trp215 in monomeric chymotrypsinogen (1chg), relative to chymotrypsin (1cho), is connected with conformational changes of the polypeptide backbone from Ser214 to Gly216, which bring the carbonyl oxygen of Ser214 into position to form the oxyanion hole (Steitz et al., 1969). The position of the carbonyl oxygen of Ser214 differs in chymotrypsinogen and α-chymotrypsin structures by 0.7 Å (Wright, 1973). Thus, the CD changes during activation of chymotrypsin correlate with conformational changes in the protein, mostly at Trp215, Trp172 and their surroundings.

d. Monomeric and Dimeric Chymotrypsinogen.

I also calculated the Trp CD for chymotrypsinogen on the basis of both monomeric (1chg) and dimeric crystal structure (2cga in PDB, Wang et al., 1985). Figure III-9 shows the Trp B_b CD of bovine chymotrypsinogen obtained for monomer

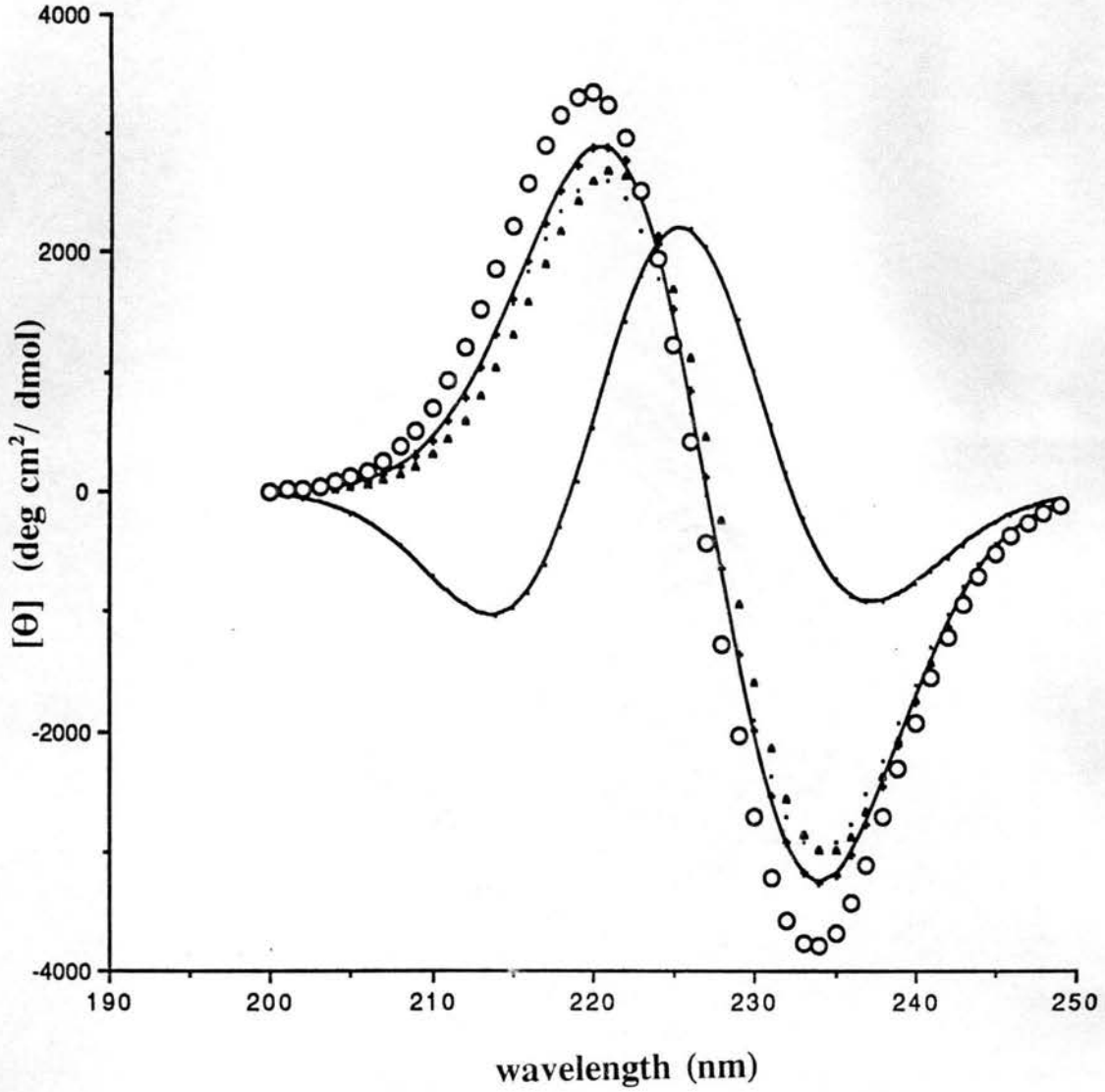


Figure III-9. Calculated B_b Trp CD of bovine α-chymotrypsin, 1cho (•); chymotrypsinogen A in monomer crystal form, 1chg (-); dimer chymotrypsinogen A, 2cga (+); molecule 1 in dimer chymotrypsinogen (▲); molecule 2 in dimer chymotrypsinogen (.).

(1chg, Freer et al., 1970) and dimer crystals (2cga, Wang et al., 1985). The predicted CD spectra for the nonequivalent chymotrypsinogen molecules in the dimer crystal are very similar to each other and to the predicted CD of chymotrypsin (Fig.III-9). The dimeric chymotrypsinogen structure was found to be more similar to that of chymotrypsin than to monomeric chymotrypsinogen (Wang et al., 1985). The difference in conformation between Trp side chains in the dimer crystal form of chymotrypsinogen (2cga) and α-chymotrypsin (1cho) is much less than the difference of either form compared with monomeric chymotrypsinogen (1chg) (Fig.III-8). The RMS deviation between the Trp side-chains is 0.32 Å between chymotrypsin and molecule 1 of chymotrypsinogen dimer (Fig.III-8), and 1.13 Å between chymotrypsin and monomeric zymogen. Thus, comparison of Trp atomic coordinates between chymotrypsin and the two chymotrypsinogen crystal forms (monomeric and dimeric) showed that the conformations of Trp residues in dimeric chymotrypsinogen are very close to those of the corresponding residues in chymotrypsin, including Trp172 and Trp215, for which the conformational differences between chymotrypsin and monomeric chymotrypsinogen and their role in CD were described above. The conformation of the whole segment 214-216 in dimeric chymotrypsinogen coincides with that in chymotrypsin (Wang et al., 1985). Monomeric chymotrypsinogen on the other hand shows significant deviations from the chymotrypsin conformation.

It is possible that the conformation of Trp215 might be affected by the dimerization process, as suggested for chymotrypsin on the basis of CD studies (Dorrington & Hofman, 1973). CD spectra of α- and δ- chymotrypsin under dimerization

conditions showed amplitude changes in the 287 nm band of α-chymotrypsin attributed to Trp, but no changes were observed in δ-chymotrypsin, which does not dimerize (Dorrington & Hofman, 1973). In order to explore the possibility of the effect of intramolecular contacts on the conformation of Trp172 and Trp215 in chymotrypsin and chymotrypsinogen crystals, I analyzed the available monomeric and dimeric crystals of both proteins. I found that the conformation of the pair Trp172 - Trp215 is almost identical in monomeric chymotrypsin inhibited with ovomucoid (1cho, Fujinaga et al., 1987), inhibited with eglin C (1acb, Frigerio et al., 1992), dimeric chymotrypsin (2cha, Birktoft & Blow, 1972; 4cha, Tsukada & Blow, 1983; 5cha, Blevins & Tulinsky, 1985; 6cha, Tulinsky & Blevins, 1987) and dimeric chymotrypsinogen (2cga, wang et al., 1985). Both ovomucoid and eglin C bind chymotrypsin at the active site and might form hydrophobic contacts with Trp172 and Trp215 (Fujinaga et al., 1987; Frigerio et al., 1992). Dimerization of chymotrypsin occurs in an area close to the active site, whereas the intermolecular contacts in dimeric chymotrypsinogen crystal occur on the opposite side of the molecule. Thus the conformation of the Trp172 - Trp215 pair in the dimeric chymotrypsin and chymotrypsinogen is defined by different local effects.

I should point out that chymotrypsinogen exhibits no tendency for dimer formation in solution at pH 4.0 (Schwert, 1949), in contrast to the clear evidence for dimerization of α-chymotrypsin under these conditions (Birktoft & Blow, 1972). We suggest that the monomeric chymotrypsinogen crystal conformation is in fact the one characteristic for the zymogen in solution, and that the three-dimensional structure of dimeric chymotrypsinogen reflects an intermediate form between chymotrypsinogen and

chymotrypsin. This proposal is based on: 1) preference for dimeric structure by chymotrypsin and not chymotrypsinogen; 2) availability of numerous intermediate and inactive, relatively stable enzyme forms between chymotrypsinogen and chymotrypsin; 3) significant differences between the experimental CD of chymotrypsinogen and chymotrypsin, which can not be explained from the crystal structure of dimeric chymotrypsinogen; 4) good correlation between the experimental difference spectrum, predicted Trp difference spectrum and CD of the Trp172-Trp215 pair. Based on our calculations, the experimental CD of dimeric chymotrypsinogen, if such can be obtained in solution, should be closer to that of chymotrypsin than to monomeric chymotrypsinogen. The crystallization conditions that enable one to obtain the dimeric chymotrypsinogen differ from those for monomer crystallization. In particular, the ethanol content in the mother liquor was increased from 10 % to 30 % and the pH was decreased from 6.3 to 4.0. The crystal form obtained at pH 4.0 might reflect an intermediate form between chymotrypsinogen and chymotrypsin, in which despite the fact that no cleavage of the polypeptide chain has yet occurred, some surface residues assume conformations close to those in the active enzyme.

e. Hen Egg-White Lysozyme.

Hen egg-white lysozyme (HEWL) is a relatively small (129 a.a.) protein, and contains six Trp residues (Trp28, Trp62, Trp63, Trp108, Trp111 and Trp123), three of which lie in the active site cleft and are involved in substrate binding (Goux & Hooker, 1980a). Most of the lysozyme Trp residues are in relatively close proximity (6 - 8 Å) to

at least one other Trp. Thus, multiple chromophore interactions are important in HEWL Trp CD (discussed in Chapter II.4).

The Trp CD of tetragonal hen egg-white lysozyme (2lyz, Diamond, 1974), calculated taking into account the exciton coupling and one-electron mixing for six indole transitions in Trp chromophores, was compared with the hen egg-white lysozyme (HEWL) experimental spectrum in the far uv (Yang et al., 1986). The L_b, L_a, B_b, B_a ,V and VI transitions were considered in the calculations performed by the origin-independent version (Goux & Hooker, 1980a) of the matrix method (Bayley et al., 1969), as described in Chapter I.2 and summarized in Chapter II.2. The predicted Trp CD of hen egg-white lysozyme (HEWL) in the far uv (Fig.III-9) is comparable in amplitude with its experimental CD spectrum (Yang et al., 1986) even though the experimental spectrum is dominated by the contributions of peptide groups. Trp CD constitutes as much as 30 % of the amplitude of the total experimental spectrum at 222 nm. Thus, aromatic contributions should be considered in the estimations of secondary structure content in solution in HEWL and other proteins with large predicted Trp CD discussed in Chapter II.4.

The predicted Trp CD satisfactorily matches the experimental CD of HEWL in the near uv (Fig.III-10). As one can see (Fig.III-10), the amplitudes of both the negative band in the L_a region (260-270 nm) and the positive band in the L_b region (270-280 nm) almost coincide in the predicted and experimental spectra. The position of the predicted and experimental bands coincide in the L_a region, but show a $\sim$ 15 nm shift at the maximum in the L_b region. This shift can be explained by the small value considered for

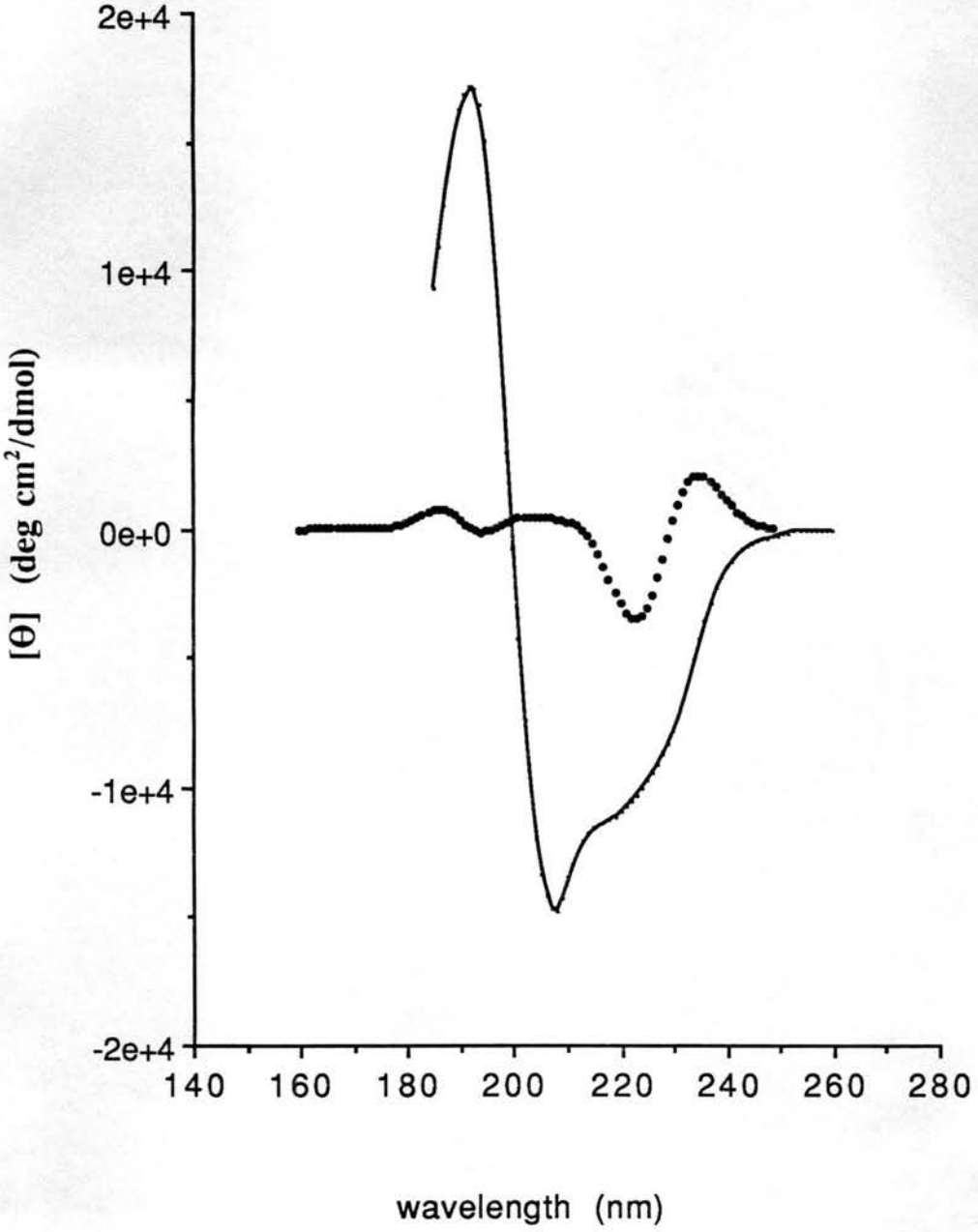


Figure III-9. Experimental CD of hen egg-white lysozyme (-) and calculated Trp CD (•) in the far-uv region.

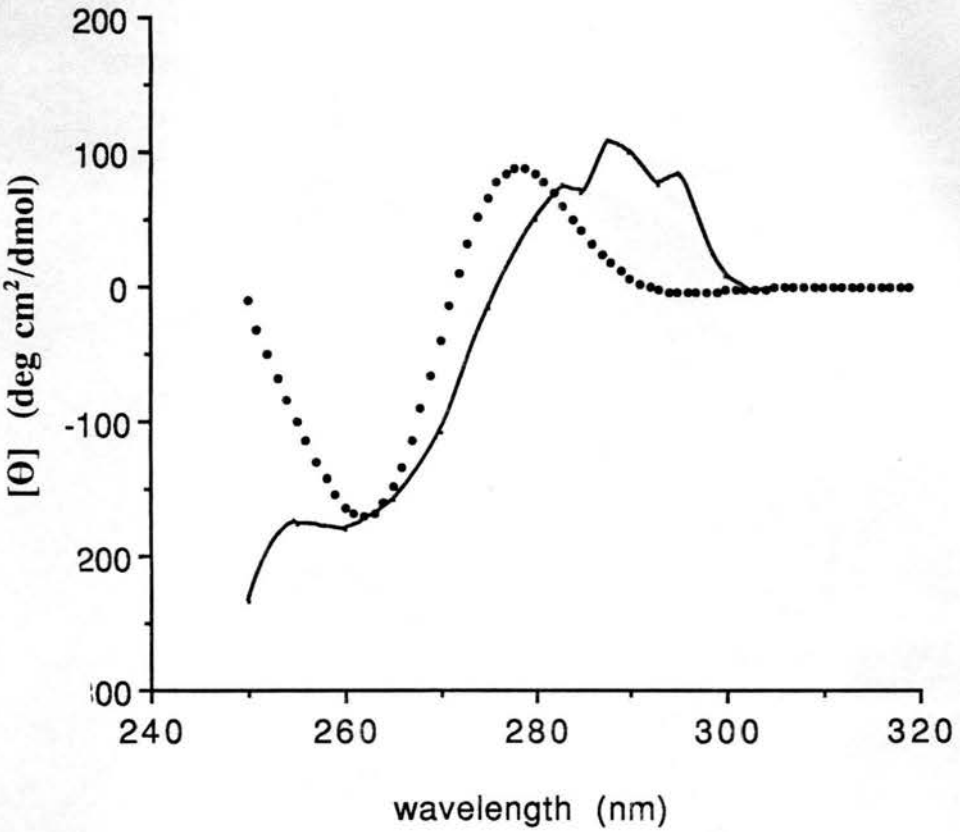


Figure III-10. Experimental CD of hen egg-white lysozyme (-) and calculated Trp CD (•) in the near-uv region.

the L_b dipole transition moment, 0.250 D. The band in the L_b region is dominated by the more intense L_a transition ($|\mu| = 0.774$ D).)

f. Human Lysozyme.

The structures of lysozyme and lactalbumin seem to be very similar in different vertebrate species (Cowburn et al., 1972), and the positions of Trp residues are largely conserved as well. Human lysozyme (Artymiuk and Blake, 1981; Acharya et al., 1989) has significant similarity to HEWL in backbone and side-chain conformations (Fig.III-11). Human lysozyme lacks Trp62 and Trp123. Trp62 (in hen numbering) is actually replaced with Tyr, the conformation of which is close to that of Trp62 in HEWL. In addition, Phe34 present in HEWL is replaced with Trp34 in human lysozyme (Fig.III-11). Based on the review of Trp side chains in lysozymes in different species (hen, human, turkey) and the related baboon lactalbumin, as well as the examples of other evolutionarily related proteins (Chapter II.4), it seems that positions of aromatic residues in globular structure are highly conserved. It can be seen (Fig.III-11) that positions of aromatic residues in a group of related proteins are likely to be conserved, but if mutated, the replacement is likely to be an aromatic residue as well. Mutations of aromatic to aromatic side chains apparently do not significantly distort the structure of the globule, like in case of hen and human lysozymes. On the other hand the intensities of electronic transitions in Tyr and, especially, Phe are one to two orders of magnitude lower than in Trp. Thus protein mutants can be produced with very similar overall structure and conformation of aromatic residues, while the intensity of aromatic CD contribution from a specific site of the globule could be intensified or made negligible

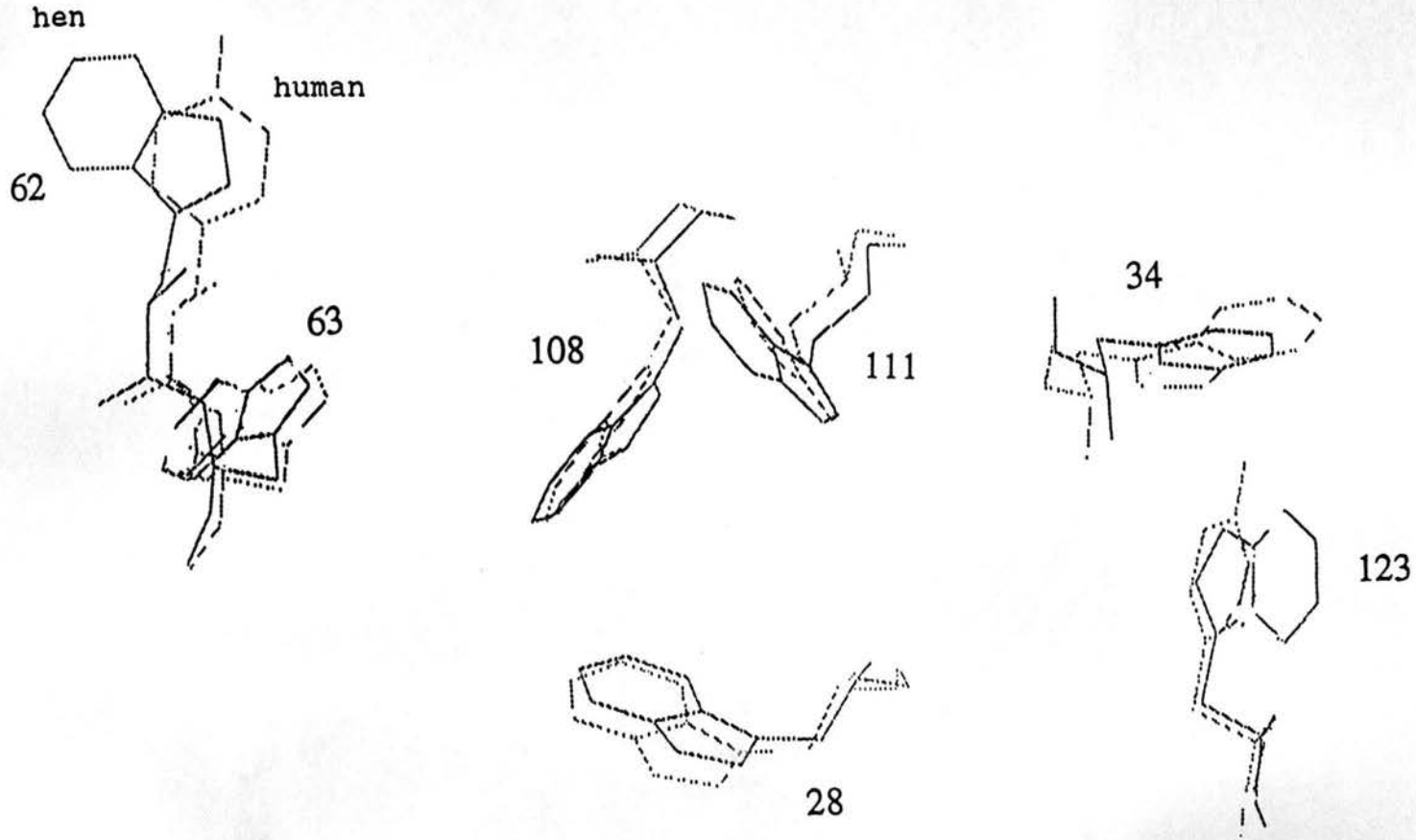


Figure II-11. Conformations of Trp residues in hen lysozyme (darker line) and human lysozyme (lighter line). Corresponding side chains at Trp positions in both proteins are shown. Numbering of Trp residues is shown for hen lysozyme.

by a proper arrangement of specific chromophores. A similar approach has already proved itself useful in the studies of ribonuclease from *Bacillus amiloliquefaciens* (Vuilleumier et al., 1993; Chapter IV.2.d.(5)) where each Trp and Tyr was mutated to Phe in order to reveal the specific CD contribution of individual side chains.

The B_b Trp CD of human lysozyme in the far uv differs significantly from that of HEWL (Fig.III-12). The Trp CD of human lysozyme is about three times smaller and is represented by a negative couplet, while HEWL has a triplet band with a large positive couplet on the longer wavelength side. The CD of human lysozyme is similar to that of HEWL in the 200-215 nm region, where free HEWL exhibits a characteristic positive band (Fig.III-12). The magnitude of the difference spectrum between the calculated B_b Trp CD of hen and human lysozyme in the 220 -230 nm region does correlate with the experimental difference spectrum (Fig.III-12), although the overall correlation throughout the far-uv spectral region is poor. The effect of other aromatic and/or peptide contributions has to be accounted for to explain the position and sign of the band at 230-233 nm in the experimental spectrum as well as the band around 210 nm.

The comparison of predicted Trp CD with the experimental near-uv CD spectra for hen and human lysozymes is given in Figure III-13. As can be seen the correlation between the calculated Trp CD and the experimental spectrum is considerably better for hen lysozyme than for human. Although the predicted and experimental band maxima for human lysozyme still match in amplitude almost as well as for hen lysozyme, the shapes of the two halves of the couplet are quite different in experiment *vs.* theory.

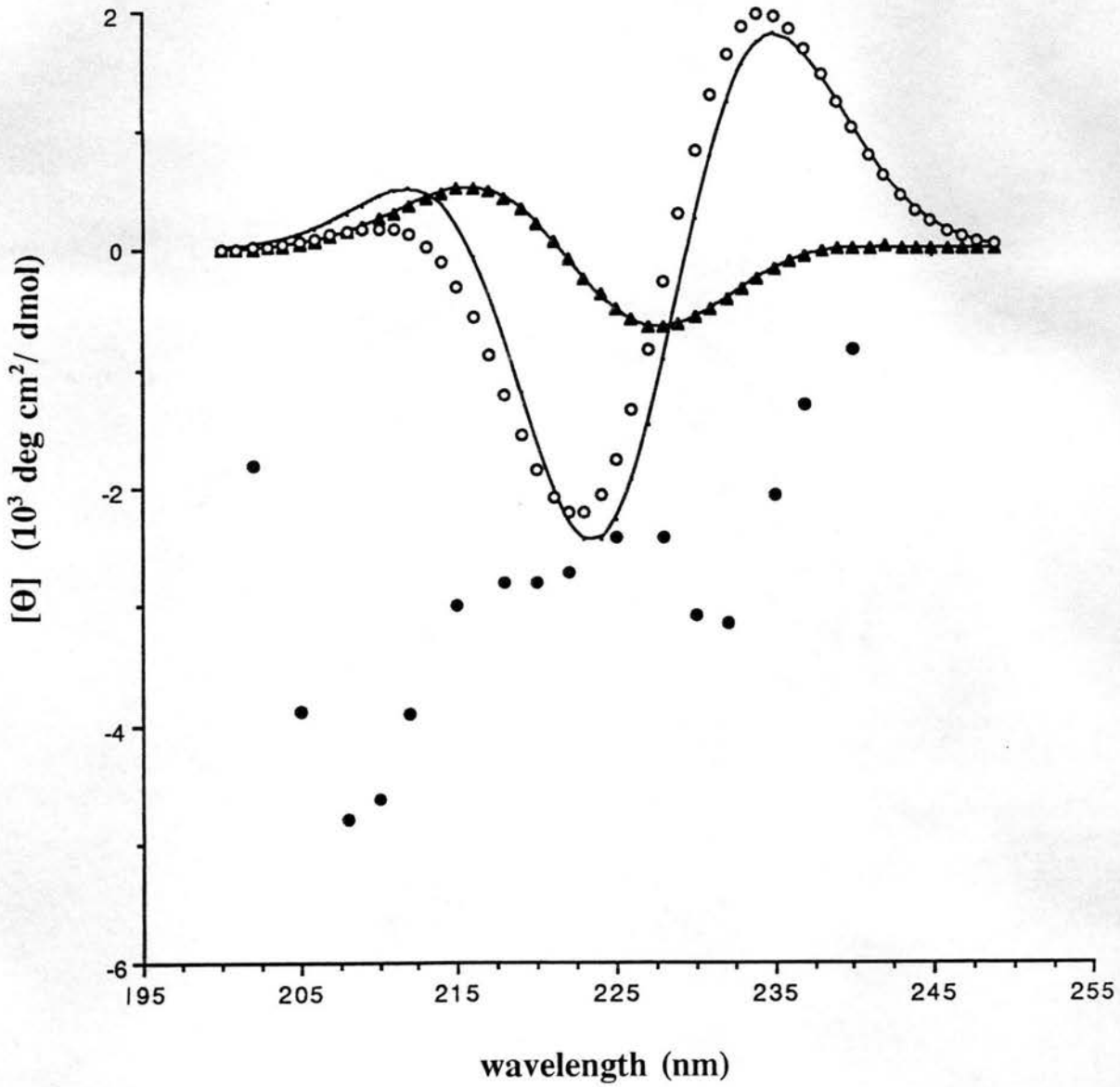


Figure III-12. Calculated B_b Trp CD of hen lysozyme (-) and human (▲) lysozyme. Difference between the predicted B_b spectra of hen and human lysozymes (o). Difference between experimental spectra of hen and human lysozyme (•).

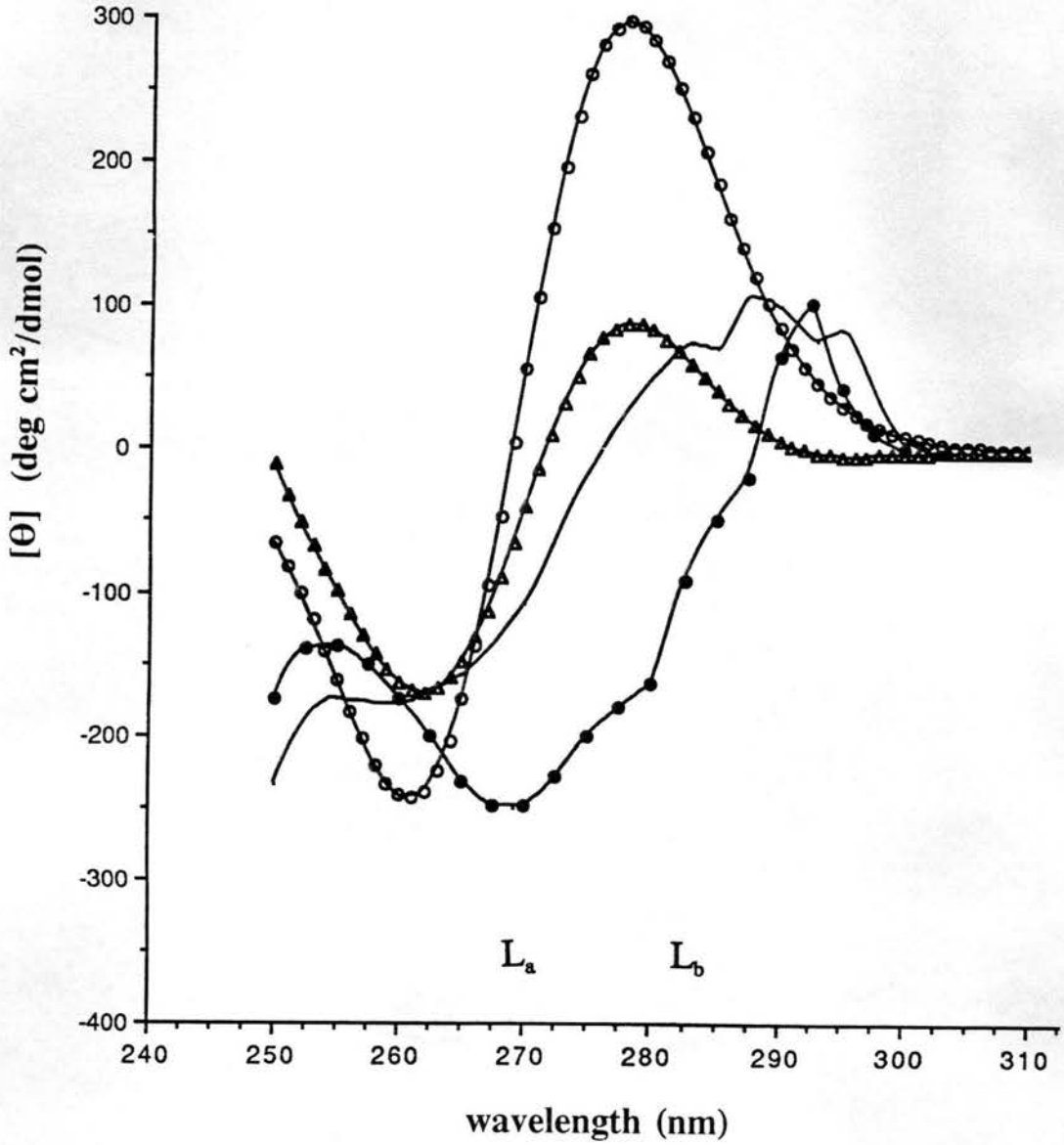


Figure III-13 Calculated Trp CD and experimental CD in the near uv for hen and human lysozymes. Calculated hen CD (▵); experimental hen CD (-); calculated human CD (o); experimental human CD (•). Wavelength positions for the L_a and L_b transitions assumed in calculations are shown.

g. Turkey Lysozyme.

The calculated Trp CD of turkey lysozyme has been discussed in detail in Chapter II.4. The Trp CD of hen and turkey lysozymes in the far uv, calculated including six transitions, is shown in Figure 14. The experimental CD spectra of hen and turkey lysozyme were obtained by Andreas Gruber, a visitor in this laboratory. Experimental far-uv CD spectra for both proteins are shown in Figure III-15. As can be seen, the large positive band around 230 nm in the difference spectrum between the predicted CD of hen and turkey lysozyme (Fig.III-14) coincides in sign with the difference between the experimental spectra in that region (Fig.III-15) and is of the same order of magnitude. However, the experimental curves do not exhibit the crossover expected from the theoretically predicted difference CD spectrum.

The calculated Trp CD spectra for both proteins in the near uv are shown in Figure III-16 and the experimental spectra in the near uv are given in Figure III-17. As can be seen (Fig.III-16), the predicted CD of hen and turkey differ significantly in the near uv, and the sign of their predicted difference spectrum (Fig.III-16) coincides with that of the experimental difference spectrum although calculations overestimate the amplitude (Fig.III-17).

i. Prognosis.

As was shown in Chapter III.4 our calculations of Trp CD spectra can often be useful in interpreting conformational changes in proteins such as those occurring upon activation of chymotrypsin, or CD changes upon Trp replacement such as in the case of DHFR, although in the general case it is hard to achieve good correlation between

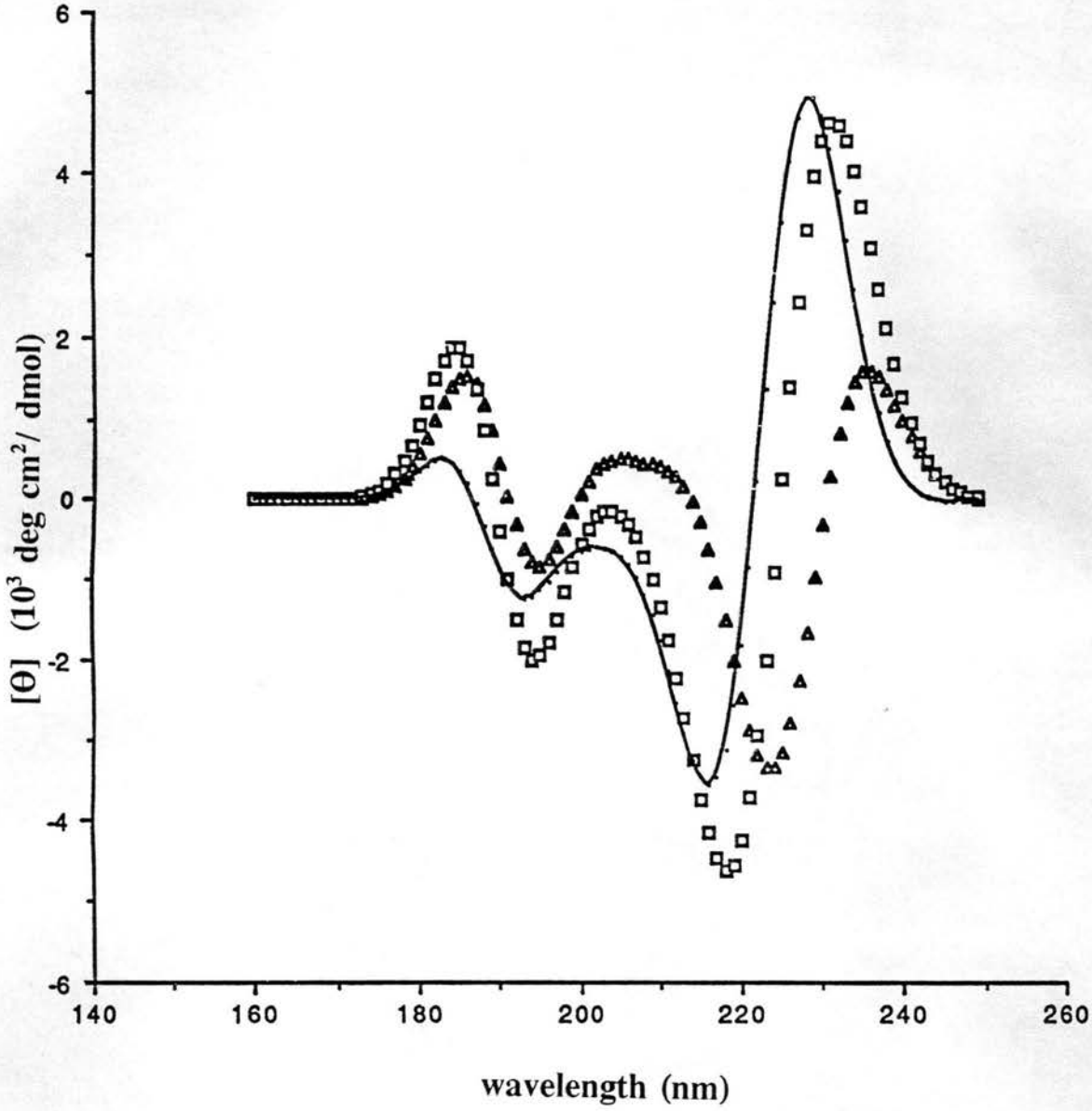


Figure III-14. Calculated Trp CD of hen lysozyme (△) and turkey lysozyme (□) in the far uv. Difference spectrum (-) between the predicted CD for hen and turkey.

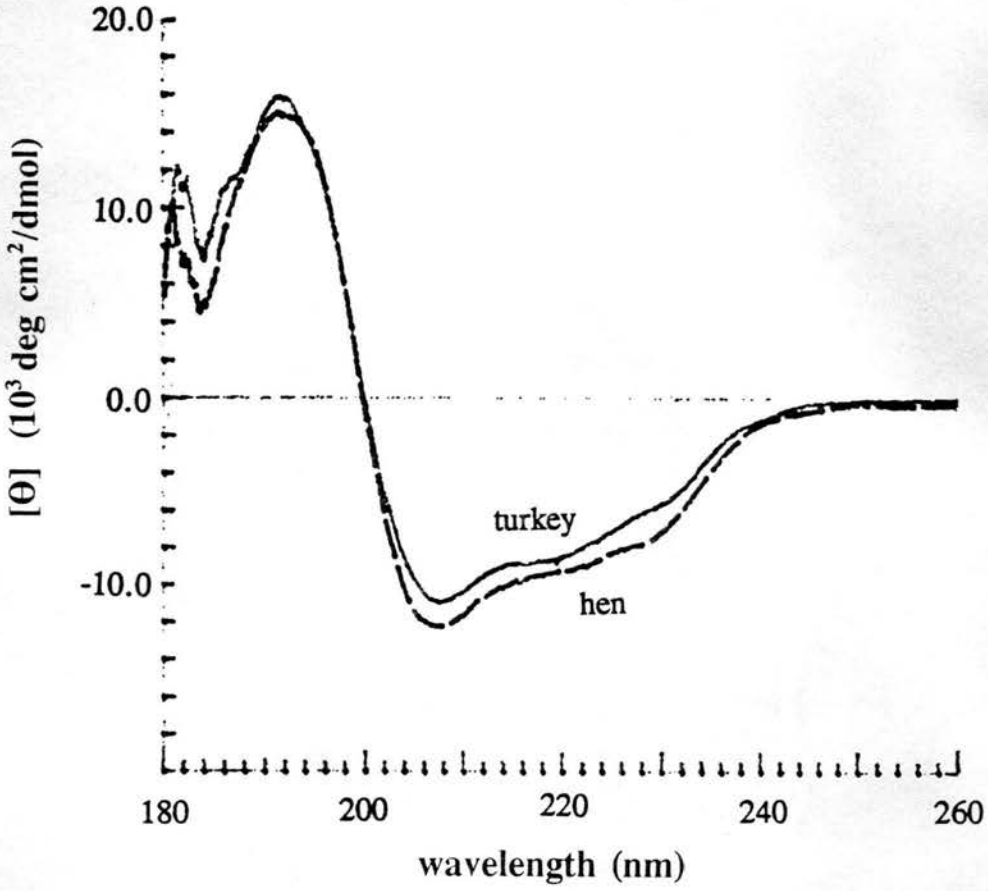


Figure III-15. Experimental CD of hen lysozyme and turkey lysozyme in the far uv. (Data provided by Andreas Gruber.)

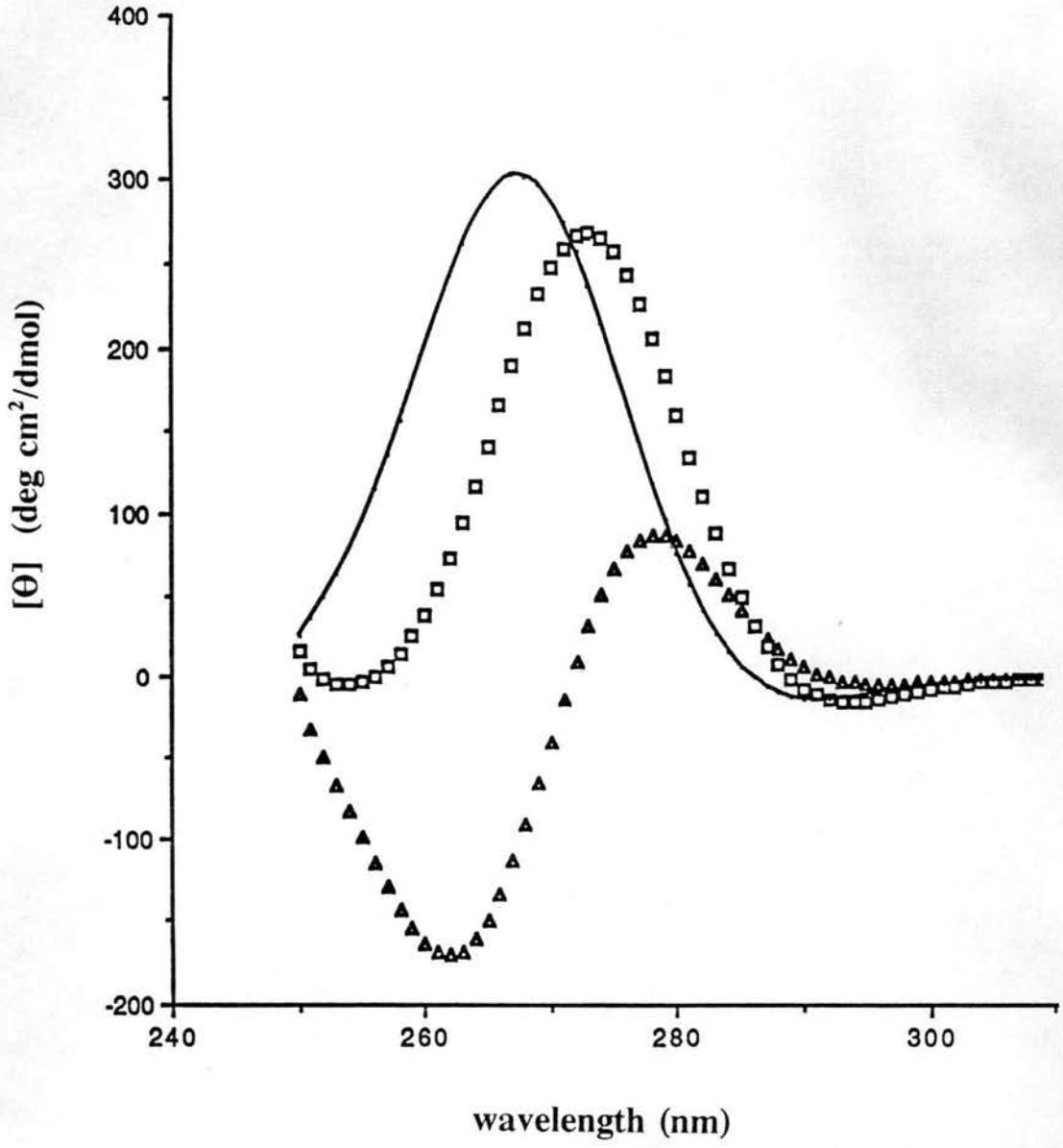


Figure III-16. Calculated Trp CD of hen lysozyme (△) and turkey lysozyme (□) in the near uv. Difference spectrum (-) between the predicted CD for hen and turkey.

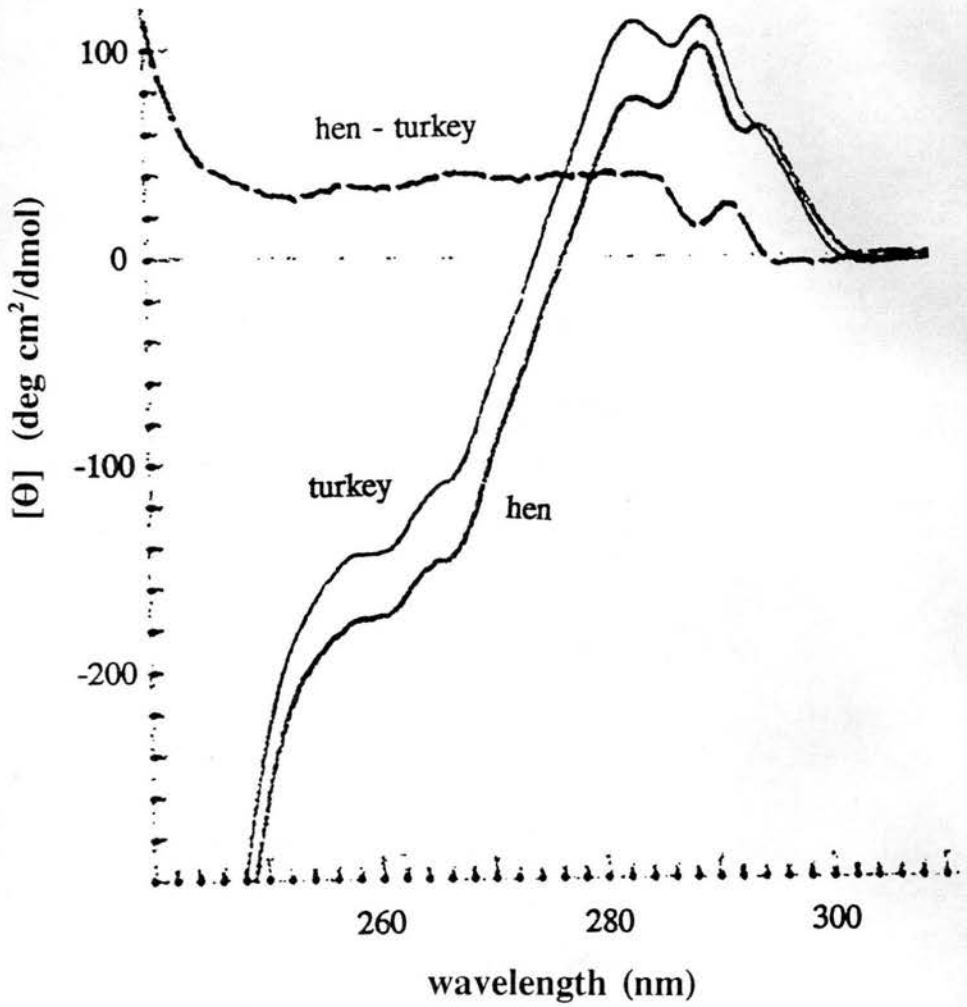


Figure III-17. Experimental CD of hen lysozyme and turkey lysozyme in the near uv. Difference spectrum between the experimental spectra. (Data provided by Andreas Gruber.)

predicted and experimental spectra accounting only for Trp chromophores. Here I will mention the main factors that may lead to improvement in calculations of aromatic protein CD. First is accounting for peptide and aromatic side-chain groups of Tyr and Phe since those might interact with Trp chromophores and give rise to additional aromatic CD. Second, accounting for vibronic transitions in indole, which might contribute to the fine structure observed in the near-uv spectra of proteins such as HEWL. Third, the dynamics of aromatic side chain and adjacent backbone might explain in several cases the difference between predicted CD, which is now calculated based on the x-ray structure while the experimental protein CD is measured in solution. The latter approach might be specifically useful in calculating CD of Trp-containing loops functionally important in DNA-binding proteins.

Chapter IV. Combining CD and Microcalorimetry in the Analysis of Protein Folding.

1. Introduction. Thermodynamics of Protein Folding.

a. Proteins as Thermodynamic Systems.

Globular proteins are known to adopt a unique three-dimensional structure throughout most of the molecule at the time they exhibit their biological function. Globular protein structure is stabilized by the balance of various non-covalent interactions: hydrogen bonds, van der Waals contacts, ion pairs and charge-dipole interactions within the α-helices. Thermodynamic analysis of globular proteins addresses such questions as protein stability under a particular range of conditions, interactions contributing to the stability of the globule, and the peculiar property of cooperativity of globular proteins, which will be discussed in detail below.

A protein molecule, composed of many thousands of atoms, represents a macroscopic system, and as such it should be specified thermodynamically, which means that the functional dependencies between its basic extensive thermodynamic parameters and their conjugate intensive variables should be defined. The thermodynamic description is the most general physical specification of any macroscopic system, without which one cannot judge its stability and behavior when the external conditions are varied. If one is interested in processes that are induced by temperature variation, it is necessary to know the dependence of the enthalpy on temperature, as this function includes all the thermodynamic information on the macroscopic states of the system that are realized in

the temperature range considered (Gill et al., 1985; Freire and Biltonen, 1978; Privalov and Pothekhin, 1986). If these states are known and specified, then, by studying their equilibrium within a given temperature range, one can deduce the enthalpy function of the system. However, if nothing is known about the macroscopic states of the macromolecules, such as in the case of a protein solution, then the enthalpy function cannot be found from the equilibrium studies and it can be determined only by direct calorimetric measurements. That is why scanning microcalorimetry, which permits direct measurements of the enthalpy dependence on temperature, has become such an important tool in thermodynamic studies of the temperature-induced processes in protein molecules.

b. Heat Capacity and Calorimetric Enthalpy.

The quantity that can be directly measured in scanning microcalorimetry is the heat capacity at constant pressure, C_p. The C_p is the energy required to raise the temperature of the sample (protein in solution in the calorimetric cell) by 1 degree.

$$C_p(\text{protein}) = C_p - C_p(\text{solvent}) \qquad \text{(IV-1)}$$

where C_p(solvent) is obtained when calorimetric measurements are performed with buffer only in the calorimetric cell. The C_p(protein) includes also the effect of the protein on the surrounding solvent. The C_p of a native protein is a linear function of temperature (Fig.IV-1). Figure IV-1 shows the results of calorimetric experiments on hen egg-white lysozyme at various pH (Privalov & Khechinashvili, 1974). The partial C_p shows a slight linear increase with temperature at low temperatures. Most of the energy during unfolding is absorbed in the narrow region that appears as a peak in the heat capacity curve. Below and above this denaturation peak, only slight changes in heat capacity are

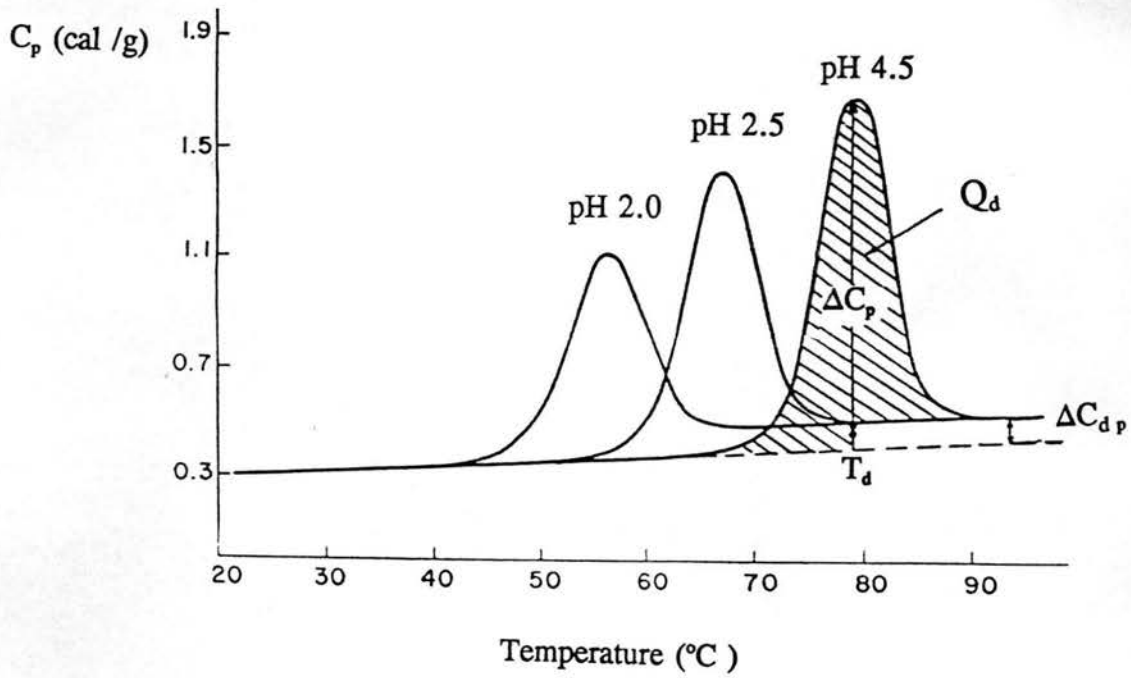


Figure IV-1. Temperature dependence of partial specific heat capacity of hen egg-white lysozyme in 0.04 M glycine buffer at pH 2.0, 2.5 and 4.5. The ΔC_P is measured from the dot at T_d between the heat capacities of the native and denatured protein to the maximum of the heat absorption. (Adapted from Privalov & Khechinashvili (1974).)

observed. When unfolding is complete, the C_p again shows only a small linear change with temperature, and is greater than the value for the native protein (Fig.IV-1). The partial heat capacity C_p is a measure of the temperature dependence of both the enthalpy and the entropy of a protein in solution. Studying the influence of temperature on the state of the protein requires the determination of the partial heat capacity of the protein in solution as a function of temperature. This became possible only after the development of a precise scanning microcalorimeter (Privalov, 1974; Privalov et al., 1975). The partial heat capacity of the protein can be determined experimentally from the difference ΔC_p^{app} between the recordings for the solvent and the protein solution at any temperature. Figure IV-1 is presented in units of specific heat capacity, which is heat capacity per unit of protein mass. Its value appears to be very similar for all compact globular proteins, 0.32 ± 0.02 cal /K g at 25°C.

The area under the transition peak gives the calorimetric enthalpy change upon unfolding, ΔH_{cal}. Heat capacity is the temperature derivative of the enthalpy, which is expressed in the Kirchhoff equation :

$$\partial \Delta H / \partial T_P = \Delta C_p \, . \qquad \text{(IV-2)}$$

All temperature-induced changes in the state of the protein molecule are reflected in the heat capacity curve. From Figure IV-1, one can see that as a first approximation the pre- and postdenaturational heat capacity changes can be described by linear functions of temperature. With this assumption, we can extrapolate the heat capacity of the native and denatured protein into the transition range. Then the following parameters can be obtained:

(a) The heat of the denaturation process Q_d corresponds to the peak area above the heat capacity functions extrapolated to the midpoint of the transition. The calorimetric enthalpy of the process is determined by integration of the heat capacity versus temperature through the area of the heat absorption peak. For the molar denaturation enthalpy we have :

$$\Delta H_d^{cal} = MQ_d \,, \qquad \text{(IV-3)}$$

where M is the relative molecular mass of the protein.

(b) The difference between the heat capacity of the native and denatured states can be calculated at the temperature of denaturation :

$$\Delta C_{d\,p}\,(T_d) = C_p^{\,D}(T_d) - C_p^{\,N}(T_d). \qquad \text{(IV-4)}$$

It is evident that since the heat capacity is a derivative of the enthalpy, the change in heat capacity must determine the temperature dependence of the enthalpy of denaturation. The temperature dependence of the enthalpy of denaturation can also be determined from the change in the enthalpy of denaturation at changing stabilities, for example at different pH (Fig. IV-1). In this case both the change in peak area δQ_d and temperature δT_d of denaturation are measured. By comparing $\delta Q_d/\delta T_d$ with $\Delta C_{d\,p}$ it is possible to decide whether the separation of the enthalpy of denaturation from the temperature dependencies of the heat capacities in the native and denatured states has been performed correctly. The values of $\delta Q_d/\delta T_d$ and $\Delta C_{d\,p}$ were shown to coincide for a large number of proteins (Privalov & Khechinashvili, 1974). Thus the linear interpolation of the heat capacities into the transition region satisfactorily reflects the properties of the system and the

extrapolated functions can in fact be considered as heat capacities of the pure native and pure denatured states.

c. Cooperativity. Effective Enthalpy.

The earliest observations of protein denaturation led to the belief that the protein globule exists in either the native or denatured state, and under certain conditions a rapid transition occurs from one state to the other. A system that can exist in two equilibrium states can be conveniently described by the equilibrium constant :

$$K^{eff} = (\theta_x - \theta_N) / (\theta_D + \theta_N) \qquad \text{(IV-5)}$$

where θ_N and θ_D are values of any observed indices characterizing the pure native and denatured states, respectively, and θ_x represents the value of this index under given conditions. It is tempting to treat the protein solution as a thermodynamic system, in which case the states of this system can be described with the following thermodynamic functions :

Gibbs energy	$-R\ T \ln K^{eff} = \Delta G^{eff}$	(IV-6)
enthalpy	$R\ T^2 (\partial \ln K^{eff} / \partial T)_P = \Delta H^{eff}$	(IV-7)
volume	$-R\ T (\partial \ln K^{eff} / \partial p)_T = \Delta V^{eff}$	(IV-8)
amount of bound ligand	$(\partial \ln K^{eff} / \partial \ln a_i)_{P,T} = \Delta \upsilon^{eff}.$	(IV-9)

Thermodynamic studies of protein denaturation were greatly accelerated when it was shown that even denaturation to a completely unfolded protein often is a reversible, thermodynamic process (Anfinsen, 1956); thus, equilibrium thermodynamics is applicable to the study of denaturation. It was also discovered that protein denaturation is accompanied by a rapid and extensive increase in enthalpy, which was considered an

indication that protein denaturation is a highly cooperative process involving the whole macromolecule (Anson, 1945).

During the scanning calorimetric recording it is possible to obtain the real (calorimetric) enthalpy of the process and the effective (van't Hoff) enthalpy simultaneously. The calorimetric enthalpy is determined as described in Equation 3. At the same time, the relative amount of heat absorbed at any given temperature,

$$q\,(T) = Q\,(T) / Q_d\,, \qquad \text{(IV-10)}$$

is a direct measure of progress of the temperature-induced reaction. Assuming that this reaction is a two-state transition, we will have for the equilibrium constant

$$K^{eff} = q / (1- q)\,, \qquad \text{(IV-11)}$$

which can be used to determine the effective enthalpy from Equation IV-7:

$$\Delta H^{eff} = R\,T^2\,(\partial \ln K^{eff} / \partial T) = [\,R\,T^2 / (q\,(1- q))]\,(\partial q / \partial T). \qquad \text{(IV-12)}$$

For the calorimetric curve,

$$\partial q / \partial T = [1/Q_d]\,\partial Q / \partial T = \Delta C_p / Q_d\,, \qquad \text{(IV-13)}$$

i.e., it is nothing other than the normalized intensity of heat absorption or excess heat capacity at a given temperature, divided by the total heat of denaturation. Thus, for the effective enthalpy at a given temperature we have :

$$\Delta_d H^{eff}\,(T) = [R\,T^2 / q\,(1- q)]\,[\,\Delta C_p\,(T) / Q_d\,]. \qquad \text{(IV-14)}$$

From the middle of the transition, where $q = 1/2$, and $T = T_d$,

$$\Delta_d H^{eff} = 4RT_d^2\;\Delta C_p\,(T_d) / Q_d. \qquad \text{(IV-15)}$$

One should also keep in mind that for an asymmetric peak, T_d is close to the temperature of maximum heat absorption, T_{max} , but is not generally equal to it (Privalov & Khechinashvili, 1974).

If the process considered is indeed of the two-state transition type the ratio $\Delta H_{cal}/\Delta H_{eff}$ should be close to unity. The average of this ratio has been obtained for several small compact globular proteins (Privalov & Khechinashvili, 1974; Privalov, 1979):

$$P \equiv \Delta H_{cal} / \Delta H_{eff} = 1.05 \pm 0.03 \qquad \text{(IV-16)}$$

As can be seen (Eq.VI-16), the observed deviation of P from unity is rather small and almost falls within the error limits. Thus, at equilibrium, we can consider denaturation as involving only two macroscopic states, the native and the denatured, i.e., as an all-or-none process. The possibility that some intermediates are present at concentrations of less than a few percent cannot be excluded, but in the first approximation, and within the error of the method, these can be neglected.

Calorimetric studies of small monomeric globular proteins and of individual domains of large proteins have also shown that thermal denaturation results in an essential increase in heat capacity by a value that does not depend significantly either on temperature or on environmental conditions (Fig.VI-1) and is specific for the given protein (Privalov, 1979; Privalov, 1982). It has been suggested that small proteins and single domains of large proteins can be considered as cooperative systems with only two stable macroscopic states, the native (N) and the denatured (D). The differences between

the thermodynamic parameters of these states can be described as temperature-dependent functions.

d. Thermodynamics of the Cooperative Domain.

One can describe this macroscopic system by two surfaces in the phase space corresponding to the extensive thermodynamic functions of these states. The transition between these states is determined by the differences of :

enthalpy, $\Delta^{D}_{N} H\ (T, pH, a_i) = H^{D}\ (T, pH, a_i) - H^{N}\ (T, pH, a_i)$, (IV-17)

entropy, $\Delta^{D}_{N} S\ (T, pH, a_i) = S^{D}\ (T, pH, a_i) - S^{N}\ (T, pH, a_i)$, (IV-18)

and Gibbs energy, $\Delta^{D}_{N} G\ (T, pH, a_i) = G^{D}(T, pH, a_i) - G^{N}\ (T, pH, a_i) =$

$= \Delta^{D}_{N} H\ (T, pH, a_i) - T\ \Delta^{D}_{N} S\ (T, pH, a_i)$. (IV-19)

The midpoint of the transition is determined by the condition

$$\Delta^{D}_{N} G\ (T, pH, a_i) = \Delta^{D}_{N} H\ (T, pH, a_i) - T\ \Delta^{D}_{N} S\ (T, pH, a_i) = 0. \quad \text{(IV-20)}$$

If the temperature is used as a variable, then the enthalpy of transition between the native and denatured states can be determined experimentally from the area of the heat absorption peak (Fig.VI-1). As follows from Equation IV-20 the entropy of transition is

$$\Delta^{D}_{N} S\ (T_G) = [\Delta^{D}_{N} H\ (T_G)]/T_G \quad \text{(IV-21)}$$

where T_G is the temperature at which the Gibbs energy change of the transition vanishes and the concentrations of the native and denatured proteins are equal. One of the most specific features of protein denaturation is a significant heat capacity increment that always accompanies this process (Fig.IV-1) so that :

$$\Delta C_p = (\partial \Delta H / \partial T)_P \quad \text{(IV-22)}$$

and

$$\Delta C_p / T = (\partial \Delta S / \partial T)_P. \quad \text{(IV-23)}$$

Then $$\Delta^{D}_{N}H\ (T) = \Delta^{D}_{N}H\ (T_G) + \int_{T_o}^{T} \Delta^{D}_{N}C_p\, dT \qquad \text{(IV-24)}$$

and $$\Delta^{D}_{N}S\ (T) = \Delta^{D}_{N}H\ (T_G) + \int_{T_o}^{T} [(\Delta^{D}_{N}C_p)\ /T]\ dT \qquad \text{(IV-25)}$$

The dependence of the denaturation enthalpy on the temperature can be observed directly by studying, at different temperatures, protein denaturation caused by varying the pH or denaturant concentration. It appears that the denatured states of a protein obtained by heating, pH or denaturants are similar from the thermodynamic point of view. In all cases the heat capacity increments of denaturation are indistinguishable (Privalov, 1982).

For all the proteins studied, the value of the $\Delta^{D}_{N}G$ function determining the stability of the native states does not exceed 12 kcal/mol (Pfeil, 1981; Privalov, 1979). Since the cooperative domain usually includes about 100 amino acid residues, it appears that the contribution of each of the residues in stabilization of the native structure does not exceed 0.12 kcal per mole of residue. This value is five time less than that of the energy of thermal motion at room temperature. It follows that a protein has an ordered native structure only because the protein is a cooperative system, whose components can only change their state cooperatively. In other words, the stability of such a system is determined by an integral contribution of all the components of the system. The stability of a cooperative domain exceeds by almost 20 times the energy of thermal motion. Thus it is sufficient to ensure the existence of the ordered structure. The secret of the stability of the native structure of protein is not in the magnitude of intramolecular interactions, which are always too weak to withstand individually the dissipative action of thermal

motion, but in the effective cooperation of these interactions. The requirement for stability seems to determine the lower limit of the size of the cooperative unit. Privalov (1982) suggested that any cooperative domain must include at least 50 amino acid residues to be stable enough at physiological temperature. The upper size limit of the cooperative unit is probably determined by difficulties in the formation of a completely integrated cooperative aperiodic structure, which increase rapidly with an increase in the number of amino acid residues. Therefore, the cooperative domain usually does not include more than 200 amino acid residues.

e. Universality of Specific Heat Capacity.

It appeared in earlier experiments that the heat capacity increment of denaturation does not depend on temperature (Fig.IV-2), and this was supported by the observed linear dependence of the denaturation enthalpy on temperature. However recent studies carried out over a broad temperature range (Privalov & Gill, 1988; Privalov et al., 1988) showed that the heat capacity of the native protein and that of the denatured protein do not change in parallel as the temperature increases. While the heat capacity of the native state is likely to be a linear function of temperature (in any case in the range from zero to 80°C, in which the native state can be practically studied), the heat capacity of the denatured state is a nonlinear function, which asymptotically approaches some constant level at high temperature. Linear extrapolation of the heat capacity of the native state above 80°C indicates that $\Delta^{D}_{N} C_p$ is likely to decrease to zero at about 140°C. The main consequence of the decrease in the heat capacity increment of denaturation with increasing temperature is that the enthalpy and entropy of protein denaturation are

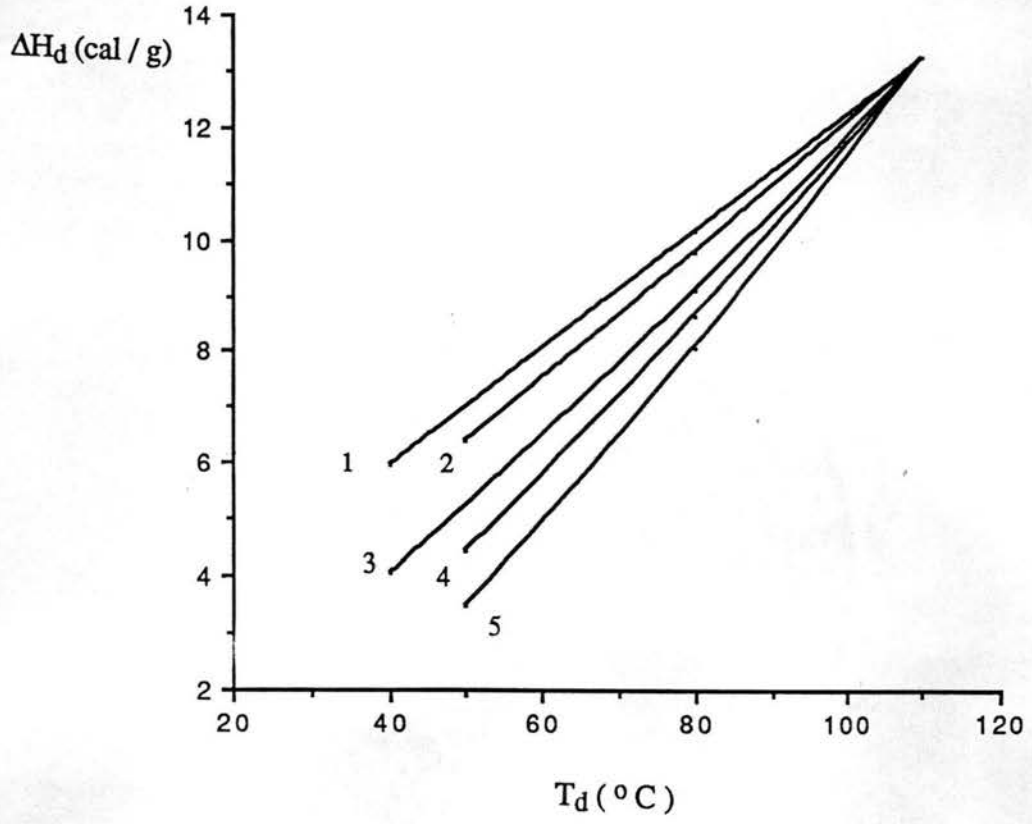


Figure IV-2. Dependence of the specific heat of denaturation on the transition temperature for : 1) ribonuclease A; 2) hen lysozyme; 3) α-chymotrypsin; 4) cytochrome c; 5) metmyoglobin.

increasing functions of temperature that asymptotically approach definite levels at about 140°C. These levels are likely to be universal for the specific values of the enthalpy and entropy of all compact globular proteins (Privalov, 1979; Privalov & Gill, 1988). If one neglected the temperature dependence of $\Delta^{D}_{N}C_p$, then the specific enthalpy and entropy of all globular proteins would come to the same values, but at about 110°C. However, in this case one could hardly suggest any physical meaning for these universal values, as the enthalpy and entropy functions would increase continuously above 110°C. Since in reality an indefinite increase of these functions is improbable, the assumption that $\Delta^{D}_{N}C_p$ decreases with temperature increase is more justified than the assumption that it is independent of temperature.

f. Temperature Dependencies of Thermodynamic Functions.

The stabilities of various small single-domain proteins can be compared by normalizing their thermodynamic quantities to correct for their different sizes. When this is done the entropies and enthalpies of protein unfolding converge at higher temperatures. Assuming the ΔC_p independent of temperature, the stabilities of the various small proteins studied thus far can be described at most temperatures T by a simple equation in which the only variable is the ΔC_p of unfolding:

$$\Delta G_{unfold} = \Delta H^* - T\Delta S^* + \Delta C_p [(T - T^*) - T\ln (T/T^*)], \qquad \text{(IV-26)}$$

where $\Delta H^* = 1.54$ kcal/(mol residue) and $\Delta S^* = 4.35$ cal/(mol residue), which are the common values at the convergence temperature T*. The value of T* is taken to be 112°C with the approximation that the ΔC_p is constant. The term by which ΔC_p is multiplied is proportional to the nonpolar surface exposed to water by unfolding. Equation IV-26

reflects the free energy of the hydration of the nonpolar surfaces upon exposure to solvent as a function of temperature. The temperature component of Equation IV-26 is negative for all temperatures less than T* which correlates with the hypothesis that water ordering opposes nonpolar solvation at low temperatures, and that burial of nonpolar residues in a hydrophobic core is a large driving force for protein folding (Spolar et al., 1989; Dill, 1990; Doig & Williams, 1991).

Another peculiarity that follows from Equations VI-24 and VI-25 is that both the $\Delta^D_N H$ and $\Delta^D_N S$ functions decrease with decreasing temperature and at some temperature become zero and then change their sign. The temperatures at which the entropy and enthalpy vanish, T_S and T_H , respectively, are constants independent of temperature. In this case:

$$\Delta^D_N H\ (T) = \Delta^D_N H\ (T_G) + (T - T_G)\ \Delta^D_N C_p \qquad \text{(IV-27)}$$

and

$$\Delta^D_N S\ (T) = [\Delta^D_N H\ (T_G)]/\ T_G + \Delta^D_N C_p \ln T/T_G \qquad \text{(IV-28)}$$

Thus for $\Delta^D_N G\ (T)$,

$$\Delta^D_N G\ (T) = \Delta^D_N H\ (T) - T\Delta^D_N S\ (T) = \qquad \text{(IV-29)}$$

$$= [(T_G - T) / T_G]\Delta^D_N H\ (T_G) + \int_{TG}^{T} \Delta^D_N C_p dT - T \int_{TG}^{T} \Delta^D_N C_p d\ (\ln T)$$

$$\simeq [(T_G - T) / T_G]\Delta^D_N H\ (T_G) + (T - T_G)\ \Delta^D_N C_p + T\ \Delta^D_N C_p \ln\ (T_G / T). \qquad \text{(IV-30)}$$

Then

$$T_H \simeq T_G - [\Delta^D_N H\ (T_G)]\ /\Delta^D_N C_p\ , \qquad \text{(IV-31)}$$

$$T_S \simeq T_G / \{[\Delta^D_N H\ (T_G)]\ /(\Delta^D_N C_p T_G) + 1\} = T_G^2 / (2T_G - T_H)\ , \qquad \text{(IV-32)}$$

and

$$T_S - T_H = \{[\Delta^D_N H\ (T_G)]\ /(\Delta^D_N C_p T_G)\}^2\ T_G \qquad \text{(IV-33)}$$

which is always positive. The temperature shift of the enthalpy and entropy functions is very important for stabilization of the native protein structure.

The stability of a protein is usually expressed in the Gibbs free energy values, since $\Delta^D_N G$ is the work required for disruption of the native protein structure. The extremum of $\Delta^D_N G$ (Eq.IV-29) is not very sensitive to the $\Delta^D_N C_p$ dependence on temperature. The maximum $\Delta^D_N G$ is reached when

$$\partial \Delta^D_N G \ / \ \partial T = - \Delta^D_N S = 0 \ , \qquad \text{(IV-35)}$$

i.e. at temperature T_S. Thus, the native state of protein is most stable at the temperature where the entropy difference of the native and denatured states is zero, and it is stabilized only by the enthalpy difference of these states (Becktel & Schellman, 1987; Privalov, 1979). At temperatures above and below T_S, $\Delta^D_N G$ should decrease and, correspondingly, the stability of the native state should decrease. One can expect, therefore, that zero stability of the native state is reached at two different temperatures: a high-temperature T_G , at which heat denaturation proceeds with heat absorption and thus with an increase of the enthalpy and entropy, and a low-temperature T_G at which cold denaturation proceeds with a release of heat, i.e. with enthalpy and entropy decreases, since both functions change their sign below T_H and T_S , respectively.

g. Denaturation as a Phase Transition

Since the native and denatured states of proteins differ not only in heat capacity, but also in enthalpy and entropy, the transition between the two states must proceed with a discontinuity of these functions, which are the first derivatives of the thermodynamic potential. According to the thermodynamic definition of changes in macroscopic systems,

denaturation of a single cooperative domain is a first-order phase transition. The native and denatured states of a single-domain protein can be regarded as two different phases of the macroscopic system (Privalov et al., 1986). The two phases differ qualitatively under all conditions, even when they are not distinguished by enthalpy or entropy which occurs at temperatures T_H and T_S , correspondingly. One phase cannot be gradually transformed into the other, but only be transformed abruptly and entirely. This shows that these two phases differ in symmetry (Landau & Lifshitz, 1958).

Thermodynamically, a protein molecule represents a unique macroscopic system with an aperiodic order. This system is unusually discrete in terms of its thermodynamic properties; it is subdivided into cooperative subsystems, domains each with only two macroscopic states - the ordered (native) and disordered (denatured). Unravelling the mechanism of the cooperation of all interactions within the domain is one of the most important problems, as this mechanism represents a physical basis for the principles of protein architecture, evolution, and functioning. The cooperative domain has a marginal stability provided by a delicate balance of various intramolecular forces. The integral enthalpy and entropy of stabilization of an ordered state of the domain largely depends on the temperature; thus the disruption of the structure occurs with a significant increase in the heat capacity. This heat capacity increment is primarily caused by the exposure of nonpolar groups of the domain to water.

h. Multidomain Proteins.

The previously discussed equality of the real and effective enthalpies obtained for small globular proteins is not universal. For instance, it was shown that the ratio

$\Delta H_{cal}/\Delta H_{eff}$ for papain, which is also a typical average-size globular protein (MW 23,000), is 1.80 $\pm$ 0.01 (Tiktopulo and Privalov, 1978). For the Bence-Jones protein (MW 48,000) it is 1.90 $\pm$ 0.01 (Zavyalov et al., 1977). Thus denaturation of these proteins could not be described as a two-state transition. The most probable explanation is that both proteins consist of two independent cooperative regions, a suggestion that is supported by their known three-dimensional structure. Indeed, papain has a very deep cleft that divides the molecule into two nearly equal domains (Drenth et al., 1971). The Bence-Jones protein consists of four domains (Edelman, 1970), two identical immunoglobulin light chains, each containing a variable and a constant part, linked by disulfide bonds (Edelman, 1970). Apparently, pairs of these domains are combined in cooperative regions.

The knowledge of the number of domains constituting the protein is in itself insufficient to predict the number of the cooperative regions in the globule. Parvalbumin, which has two calcium-binding domains, behaves as a single cooperative unit in the absence of calcium (Filimonov et al., 1978). The opposite phenomena has been demonstrated for the pancreatic trypsin inhibitor. The calorimetric enthalpy of pancreatic trypsin inhibitor, calculated per 6500-dalton monomer unit, is half as large as the effective enthalpy. This means that the cooperative unit of this protein is a dimer (Privalov, 1979).

It appears that in general all proteins represent macroscopic systems that are subdivided into more or less independent cooperative blocks, i.e. domains (Janin & Wodak, 1983; Privalov, 1982; Wetlaufer, 1973). Small proteins may consist of one

domain, while large proteins may comprise a number of distinct domains; for example, according to calorimetric studies the fibrinogen molecule may contain twelve domains (Privalov & Medved'ev, 1982). The cooperative domains in large proteins seem to behave in exactly the same way as small globular proteins. Their native structure is stable up to a certain critical temperature, which depends on the environmental conditions (pH, ionic strength, denaturants) and is then disrupted with intense heat absorption. The discreetness is apparently an intrinsic principle of protein structural organization. Its universality implies that it might have deep physical and biological grounds. On the other hand, it is just this unique thermodynamic property of the protein molecule that permits quantitative definition and thermodynamic description of protein structural stability (Privalov, 1982).

The family of calcium-binding proteins (parvalbumin, troponin C, calmodulin, myosin light chain) contain two or more highly homologous H domains. It has been suggested that these domains arose from gene duplication (Privalov, 1982) of a predecessor globular polypeptide, and thus are probably self-sufficient to fold into energetically independent globular structures. Parvalbumin has two H regions, each of which is folded into two α-helices, with the loop between those helices coordinated by Ca^{2+} ion (Moews & Kretsinger, 1975), representing, therefore, a calcium-binding domain. Thermal denaturation of parvalbumin was found to proceed as a two-state transition (Filimonov et al., 1978). Both domains seem to be consolidated into a single cooperative block, i.e., in the course of evolution they evolved in a way to perfectly fit each other.

Ovomucoid, a protease inhibitor, has three homologous regions (Kato et al., 1976). The calorimetric studies suggest that all three of these homologous regions unfold independently and react differently on the change of pH (Privalov, 1982).

Plasminogen contains five homologous regions bordered by a nonhomologous 80 a.a region at the N-terminus that is cut off during activation, and a 230 a.a. C-terminal that is S-S bonded to the preceding H region in the zymogen, but not in the active form (Collen, 1980). Deconvolution of the complicated melting profile of plasminogen resulted in seven transitions (Novohatni et al., 1982). Calorimetric studies of fragments confirmed that five of these transitions correspond to melting of individual homologous repeats and the other two to the melting of the C-terminal part. It was observed that the stability of one of the C-terminal cooperative regions was decreased upon conversion to plasmin when the S-S bond, connecting it to the H region, is cleaved.

Thermodynamic analysis of the termal denaturation of immunoglobulins revealed three transitions in which F_{ab} domains melted and four for F_c regions. Comparison of the F_{ab} unfolding with the melting profile of the Bence-Jones protein (see above, Zav'alov et al., 1977) indicated that two of the transitions correspond to independent unfolding of light and heavy chains. As for the third one, which occurs at lower temperature, it might correspond to disruption of the contact between the heavy and light chain domains or to a subdomain structure of one of the parts (Privalov, 1982). The latter suggestion was based in part on the knowledge that the variable chain is coded by two different genes (Gough, 1981), whose products might be able to form cooperative structures independently. The F_c fragment consists of two identical S-S linked chains, yet displays

four transitions in its melting profile. It was found that the dimer of two identical C_H^3 domains still exhibits a complex profile (Tischenko et al., 1982), in which one of the transitions is concentration dependent, i.e., represents a bimolecular reaction of separation of the two fragments. It was concluded that each of the C_H^3 domains consists of two nonidentical subunits that are combined in the dimer in a nonsymmetrical manner. This seems to be the case for the C_H^2 domain as well. At the same time, the components of the C_H^2 and C_H^3 domains could not be defined in the crystallographic model (Deisenfhofer, 1981). The subunits seem to fit each other perfectly, while the memory of their evolutionarily earlier independent existence is conserved in their separate cooperative folding.

i. Structural and Energetic Domains in Globular Proteins.

According to x-ray diffraction, most of the structure in globular proteins is concentrated in regions of higher density that generally constitute hydrophobic cores. The core and structurally connected adjacent protein regions are defined as crystallographic structural domains. Small globular proteins are usually composed of a single domain and undergo cooperative denaturation. Proteins that contain several structural domains usually exhibit complex denaturation profiles, corresponding to more or less independent unfolding of the domains and changes in interactions (separation) between those domains. Domain structure is generally assumed to be constant for the particular protein and independent of the medium.

Recent studies on thermal denaturation of binase, pepsin and pepsinogen indicated that the number of cooperatively melting regions within the protein might vary,

depending on the pH (Protasevich et al., 1985; Protasevich 1985; Protasevich et al., 1987; Grishina et al., 1989; Makarov et al., 1991). It was suggested (Protasevich et al., 1987; Grishina et al., 1989; Makarov et al., 1991) that the number of cooperative regions in globular proteins might be pH-dependent as a consequence of the distribution of charge on the surface of the protein. Electrostatic charge-charge interactions are the longest range forces in the protein macromolecule (Quiocho et al., 1987; Sternberg et al., 1987) and the most strongly dependent on the electrostatic properties of the medium. It was proposed to call the cooperative portions in globular proteins energetic domains (Protasevich et al., 1985), to distinguish them from the structural domains defined from x-ray diffraction. Energetic domains were defined as regions of globular structure that react to a change in electrostatic conditions of the environment by separation or unification with other energetic domains, thus changing their thermostability (Protasevich et al., 1987; Grishina et al., 1989). Studies by other groups (Brandts et al., 1989; Battistel et al., 1991; Griko et al., 1989; 1992) confirmed that number and size of cooperative regions in proteins might vary, depending on the pH of the solvent.

There are a number of proteins whose crystal structures show two lobes partially separated by a cleft. Some of these (RNase A, lysozyme, chymotrypsinogen, parvalbumin) have been reported to always exhibit a single cooperative transition. Some (papain, phage λ repressor) always show two uncoupled transitions, while still others (hexokinase, PGK) show a single coupled transition under some conditions and two transitions under other conditions. The possibility of uncoupling of strongly interacting domains has been tested on RNase A (Brandts et al., 1989). It has been found by DSC

(differential scanning calorimetry) that the RNase A transition can in fact be split into two by either raising the amount of methanol in the solvent at constant apparent pH or by lowering the apparent pH for RNase solutions containing 50 - 70 percent methanol. In all cases, the transitions were reversible. It is known that RNase retains activity at low temperature in solvents containing up to 50 percent methanol (Fink & Grey, 1978) and appears to have a structure similar to that in pure water (Biringer & Fink, 1982). Thus the two transitions in RNase A in methanol-rich solvent may correspond to separate unfolding of two native-like domains.

It was found that covalent attachment of RNase A to silica beads causes significant changes in the thermal stability and thermodynamics of unfolding of the immobilized enzyme (Battistel et al., 1991). The deconvolution of the thermograms reveals that the thermal transition represents a sum of two overlapping processes. Immobilization leads to a general stabilizing effect. At low pH the stabilization effect becomes larger. At pH 4.0 the T_m of immobilized enzyme is six degrees higher than for the enzyme in solution. The pH has less effect on the ΔH and T_m of the immobilized enzyme. Battistel et al. (1991) suggested that, to a first approximation, this effect might be due to a decrease in the number of charged amino groups, since one or two of these (Koch-Schmidt & Mosbach, 1977) are masked in the linkage between the protein and the support. Also, after immobilization, the presence of one or more covalent bonds with the support may change the relative intrinsic stability by stabilizing one domain over the other, which leads to the splitting of the unfolding transition. On the other hand, interdomain interactions might be weakened. The N-terminal amino group is expected

to be the most reactive, although this has not been shown directly. If we assume that the protein is bound to the silica through the N-terminal amino group, then the N-terminal domain should be stabilized. This is in agreement with NMR studies (Westmoreland & Matthews, 1973) that showed that the thermally labile part of the protein is the C-terminal region. Evidence of coupling between the domains has been observed in experiments with ligand binding to immobilized RNase A (Battistel et al., 1991), which stabilizes both of the domains. It has been suggested that strong coupling of RNase A domains in the native enzyme makes their detection difficult in solution.

It has been shown (Griko et al., 1989) that the denaturation of phosphoglycerate kinase (PGK) can be observed not only when the solution is heated above 30°C, but also when it is cooled below this temperature. This disruption of the native PGK structure upon cooling and the subsequent formation of this structure upon heating both proceed in two distinct stages that correspond to the independent disruption or reformation of each of its domains. In contrast, the heat denaturation of PGK proceeds in one stage, showing that the two domains of the molecule are associated into a single complex which figures in the denaturation complex as a cooperative unit. Thus, at elevated temperatures there is a positive interaction between the domains, which disappears at lower temperatures. This might be due to hydrophobic interactions, which are known to be temperature dependent. They decrease in strength as the temperature decreases and may even change sign at sufficiently low temperatures (Privalov & Gill, 1988; 1989). The temperature decrease, though, might lead to a decrease in inter- and intradomain interactions, which results in an increase of the independence of the domains and a

decrease in their stability. The observation of the structural changes in PGK during heat and cold denaturation revealed a cooperative transition in the molecule structure on heat denaturation as observed by CD and fluorescence (Griko et al., 1989). On the other hand, PGK shows more complex behavior during cold denaturation. CD changes and fluorescence show a transition almost as cooperative as for the heat denaturation and in the temperature range corresponding to the first peak (lower temperatures) in the cold denaturation profile. As the only two Trp of PGK are located in its C-terminal domain (Watson et al., 1982), it has been concluded that the first peak of the cold denaturation profile corresponds to the C-terminal domain and the second to the N-terminal, which denatures independently upon cooling.

The Cro repressor of phage λ is the smallest member of the family of DNA-binding proteins with a characteristic helix-turn-helix motif that interacts with DNA (Brenman & Matthews, 1989; Pabo & Sauer, 1984). Cro repressor is active only in the dimeric form (Pakula & Sauer, 1989). It has been shown that the thermal denaturation of the dimer of Cro repressor can be described as a two-state transition (Gitelson et al., 1991). This indicates that the interaction between the two subunits are so strong that the dimer forms a single cooperative block. Thermodynamic properties of a mutant λ Cro repressor with Cys replacing Val55 were studied calorimetrically at various pHs (Griko et al., 1992). It was found that formation of the disulfide bond between neighboring Cys55 in the mutant dimeric molecule leads to stabilization of the structure formed by the two C-terminal parts. This structure behaves as a single cooperative domain during termal denaturation. Under all conditions, the V55C mutant denaturates in two stages,

shifting to lower temperatures with decreasing pH. These temperature shifts were found to be distinctly different for the two peaks, which means that the portion of the dimer corresponding to one of the peaks is stabilized more strongly by an increase of pH than the rest of the molecule, corresponding to the other peak.

At pH 3.0, both transitions occur at the same temperature. The enthalpy of the more pH-sensitive transition is significantly smaller than that of the less sensitive one. The total denaturation enthalpy of the mutant is close to that of the wild type and, notably, the denaturation of the mutant Cro dimer proceeds as a two-state transition.

The N-terminal part of the λ Cro repressor does not form any stable structure when isolated (Griko et al., 1992). It was found that stability of the C55V dimer depends on the presence of the C-terminal domain. The C-terminal domain influences the stability of the N-terminal even when its own enthalpy of denaturation is equal to zero, which indicates the entropic character of the stabilization. The removal of the unstructured N-terminal part of the polypeptide chain stabilizes the remaining structure. Thus there is a negative cooperativity between the folded central part of the dimeric molecule and the unfolded N-terminal parts of the polypeptide chains. In an S-S cross-linked dimer, the N-termini of the monomers interact to form a single cooperative block that melts in an all-or-none way 9°C higher than the un-crosslinked dimer.

Based on the crystal structure (Anderson et al., 1981), it is not possible to explain the formation of a single cooperative domain from the two N-termini by assuming any stabilizing interaction between them. The C-termini form the joint β-structural central domain, relative to which the N-termini are located on opposite sides, far apart from

each other. Griko et al. (1992) studied the possibility of tetramer formation in solution by equilibrium centrifugation under native conditions and upon adding 2M urea, and by gel filtration at increasing temperatures from 25°C to 50°C. No indications were found of the presence of tetramer under native conditions nor at increased temperatures. The results obtained raise questions about the relevance of the Cro repressor crystal structure to the structure in solution. They also indicate that domains in DNA-binding proteins responsible for separate functions like DNA-binding and transcriptional activation/repression, and which can be separated and interchanged between different proteins (Brent & Ptashne, 1984; Fields and Song, 1989), are most probably energetically independent regions of the globule.

From the examples of multidomain proteins reviewed in sections (IV.1.h) and (VI.1.i), one can see that in some cases the domains unfold independently, while in others, stabilizing or destabilizing interaction between them is apparent. Brandts et al. (1989) attempted to provide a quantitative description of the interaction between cooperative regions in multidomain proteins. The interaction between domains was quantified by considering an interface free energy, ΔG_{AB}. It is assumed that ΔG_{AB} becomes zero whenever either of the interacting domains unfolds. In this case, the interaction term stabilizes only the domain with lower T_m.

This model has been applied to estimate the free energy of interaction between the regulatory and catalytic subunits of native aspartate transcarbamoylase, and the value of ΔG_{AB} was found to be -25 Kcal/mol. On the other hand no interaction ($\Delta G_{AB}=0$) has

been detected between the transmembrane and cytoplasmic domains of band three of the human erythrocyte membrane.

In the study of dimeric ligand-binding proteins that consist of a ligand-binding and a regulatory domain, it has been shown (Brandts et al., 1989) that whenever the ΔG_{AB} term is ligand-dependent, ligand binding is affected also by the other, regulatory domain. It was found that binding of MgATP to the carboxy-terminal domain of phosphoglycerate kinase is stabilized by about 20 percent in free-energy by the presence of the regulatory domain, which increases the binding constant 25-fold. Similar results were obtained upon binding of MgADP to creatine kinase. However, the contribution from the regulatory domain of creatine kinase was found to contribute only about 5 percent of the free energy of binding, and only a 2-fold increase in binding constant has been detected.

Application of this model requires obtaining calorimetric data under conditions that destabilize the interacting domains to a different extent. The most convenient quantitative definition for ΔG_{AB} :

$$\Delta G_{AB} = -\Delta H_B (1 - T_B / T'_B) \qquad \text{(IV-35)}$$

requires the knowledge of the temperature of the thermal transition of one of the domains both under conditions when this domain is more (T_B) and less (T'_B) stable than the other. Such conditions might be achieved by denaturing agents, hydrophobic compounds, pH or mutations destabilizing the interface between the domains (Brandts et al., 1989).

In the present Chapter, I will describe some applications of structural (CD) (Bolotina et al., 1985; Morozov et al., 1994) and thermodynamic (microcalorimetry) approaches to the study of variation in number, size and stability of cooperative regions

(energetic domains) in two small one-domain proteins, bacterial ribonucleases from *Bacillus intermedius 7P* (Grishina et al., 1989; Grishina et al., 1993; Makarov et al., 1994) and *Bacillus amyloliquefaciens*, and two multidomain proteins, the gastric protease, porcine pepsin (Makarov et al., 1991; Esipova et al., 1993; Makarov et al., 1994), and its zymogen (Makarov A.A, Grishina, I.B., Protasevich, I.I., Esipova,N.G., in preparation).

2. Localization of Energetic Domains in *Bacillus Intermedius* 7P Ribonuclease.

a. Introduction.

Scanning microcalorimetry studies of *Bacillus intermedius* 7P ribonuclease (binase) have shown (Protasevich et al., 1987) that in the pH interval 4.0-7.0 the globule undergoes cooperative thermal denaturation (Fig.IV-3). Binase enzymatic activity is stable in the pH interval 4.0-7.0 and decreases abruptly outside this interval (Karpeiskii et al., 1981). The pI of binase is 8.5 (Golubenko et al., 1979). At pHs below 4.0, the heat absorption peak is no longer symmetric. Two heat absorption maxima could be readily distinguished in the binase denaturation profile at pH 2.4 (Fig.IV-3). The ratio of calorimetric to van't Hoff enthalpy at this pH approaches two (Privalov, 1982; Protasevich et al., 1987). Therefore it has been concluded that at acidic pH values the cooperative binase globule separates into two energetic domains (Protasevich et al., 1987). It has been determined that the far-uv CD spectrum of binase, which reflects its secondary structure content, remains unchanged throughout the pH interval 2.8-11.5 (Bolotina et al., 1979). It has been suggested that changes in binase tertiary structure might be responsible for separation of the cooperative globule into two energetic domains

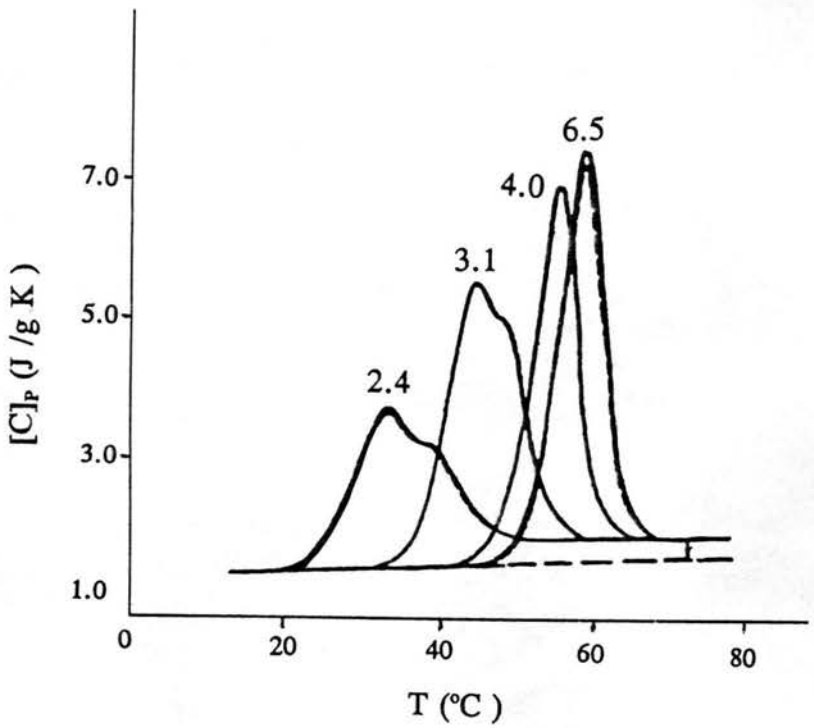


Figure IV-3. Temperature dependence of the specific heat capacity of *Bacillus intermedius* 7P ribonuclease in 10 mM glycine buffer at different pH shown in the figure. Results of the first heating are shown in solid line. Results of the second heating of the sample are shown in broken line. Broken line also shows extrapolation of the heat capacity of native binase. (Adapted from Protasevich et al. (1985).)

(Protasevich et al., 1987). By combining CD and analysis of the x-ray coordinates of binase I was able to identify the regions of the globule that form energetic domains at acidic pH and the interactions that are likely to regulate the discrete change in cooperativity of binase, and characterize the globular structure following those thermodynamic changes (Grishina et al., 1989; Grishina et al., 1993; Makarov et al., 1994).

b. Materials and Methods.

(1) Materials. Bacterial ribonuclease from *Bacillus intermedius* 7P (EC 3.14.23, binase) was obtained 80-90 percent pure from Medical Preparations Co. (Institute of Organic Synthesis, Latvian Academy of Sciences, Riga, Latvian Republic). The relative molecular mass of the protein molecule is 12.3 kDa. Binase was additionally purified by column chromatography and desalted as previously described (Golubenko et. al., 1979) by Research Associate Nina Chepurnova (Engel'hardt Institute of Molecular Biology, Russian Academy of Sciences, subsequently referred to as IMB). The lyophilized sample was stored at -20°C, and defrosted and stored at -6°C for current use over a few days. Binase samples for CD experiments were prepared in double-distilled deionized water.

(2) Protein concentration was measured by absorption at 280 nm on the spectrophotometer Specord UV-VIS Zeiss (GDR). The extinction coefficient ($E_{1cm}^{1\%}$) for the protein solution at neutral pH was reported to equal 18 (Golubenko et. al., 1979).

(3) The pH of aqueous binase solutions was adjusted by addition of 0.1 M HCl and measured with a Radiometer pH-meter (Denmark).

(4) CD spectra in far-uv region (200-250 nm) were recorded with a spectropolarimeter DKhR-02 (Russia) in a 1-mm quartz cell. Protein concentrations were kept at 0.5-0.6 mg/ml. The polarimeter was calibrated with (+)-10-camphorsulfonic acid. The instrument sensitivity was set at 2.225×10^{-5} $\Delta\epsilon$/mm. Measurements were performed with a time constant of 30 sec and a scanning rate 1 nm/min. The data obtained were presented in the form of molar ellipticity per amino acid residue of the protein. The average relative molecular mass of a residue in binase was calculated to be 112.8 Da.

(5) CD spectra in the near-uv region were recorded with a spectropolarimeter Jobin-Yvon Mark III (France) at protein concentrations of 0.4-0.5 mg/ml in a 0.5-cm quartz cell. The polarimeter was calibrated with (+)-10-camphorsulfonic acid. The sensitivity of the instrument was 2×10^{-6} $\Delta\epsilon$/mm. The time constant was adjusted to 30 sec and the scanning rate was 1 nm/min.

(6) Temperature inside the cell was adjusted with thermostated water flow. With the DKhR-02, the temperature in the cell was determined by a thermometer in contact with a metal coat around the cell. In the Mark III, the temperature in the cell was monitored by a thermocouple in contact with the metal coat. In all measurements, sufficient time was allowed to adjust and equilibrate the temperature of the protein sample in the cell (10 min).

(7) Estimation of the secondary structure content in experimental protein CD spectra and accounting for aromatic CD has been performed according to Bolotina & Lugauskas (1985), using the software developed by these authors. Calculations of initial values for

secondary structure content and parameters for log-normal curves were performed on a Hewlet-Packard 9830A (USA). Final fitting of all parameters was performed on a BESM-6 (Russia). Deconvolution of the CD spectra was also performed as described by by Provencher & Glöckner method (1981) on an IBM AT personal computer.

(8) Crystal structure of the binase molecule and molecular graphic studies. The spatial structure of the binase molecule was determined at 2.0 Å by A. Pavlovsky (Pavlovsky et al., 1989; Grishina et al., 1990). The crystals are rhombic and contain two molecules per asymmetric unit. The symmetry group of the crystal is $P2_12_12_1$. The unit cell parameters of the crystal are: a=111.4 Å, b=69.6 Å, c=33.5 Å. 15,000 independent reflections were analyzed with the Hendrickson-Konnert program (Konnert & Hendrickson, 1980) on a Nord-560 computer (Norway). In the latter stages of refinement, the coordinates and individual temperature factors for all non-hydrogen protein atoms (except Ala1) and 223 water molecules were accurately determined. The final R-factor was 0.15, and the average mean square deviation of the bond lengths of the model from the ideal values was 0.019 Å.

The atomic coordinates of binase (Pavlovsky et al., 1989) were refined by D. Vassylyev using the molecular graphic package FRODO (Jones, 1978; Pflugrath, 1984) on a PS390 Evans & Sutherland graphic station (in the laboratory of R. Huber, Max Planck Institute of Biochemistry, München, Germany) and using the modified package Tek_FRODO (Jones, 1978; Pflugrath, 1984; Vassylyev & Adzubey, 1992) on a Tektronix (USA) graphic station. Optimization of coordinates, following the manual adjustment, was performed with the program TNT (Tronrud & Ten Eyck, 1990).

Calculation of the hydrogen-bond angles defined by the atoms donor-proton-acceptor, was performed by A. Pavlovsky on the computer Nord-560 (Norway) using his program, which calculates the position of the hydrogen-bonded proton from geometrical considerations.

I performed all the visual analysis on the binase 3D structure using the graphic package VisiCoor (Kuznetsov & Lim, 1992), which allowed high quality interactive manipulations with the protein molecule. This program allows the use of any of the popular space filling elements and allows the user to operate separately with defined fragments of the model. The VisiCoor program is built on VAX/VMS media and was executed on the graphics terminal Tektronix (USA).

The set of atomic coordinates was completed by adding protons in the appropriate positions with a program for determination of proton coordinates (by Dmitriy Kuznetsov, IMB), which discriminates between six possible hydrogenation types for a specific protein atom: one, two or three protons on an sp^3 atom, one or two protons on an sp^2 atom, and a single proton on the oxygen of the hydroxyl (-OH) group. The algorithm according to which the proton coordinates were calculated assumes an ideal geometry for the positions of the proton atoms. Among the possible isomers of the methyl groups, the staggered conformation was assumed. The protons of the hydroxyl groups were placed in the plane of the supporting atoms (for example, in the plane defined by C_α-C_β-O_γ in Ser). The resulting proton coordinates were added to the file of protein coordinates in Brookhaven Data Bank format. This program was also realized on VAX/VMS media.

c. Results and Discussion.

(1) Analysis of the Near-UV CD of Binase.

My first objective was to determine *what kind of changes in tertiary structure of binase might be responsible for the separation of energetic domains.* This goal was approached by obtaining the near-uv CD spectra of binase in H_2O throughout the pH region in which the separation of the energetic domains takes place. The near-uv CD spectrum of native binase is shown in Figure IV-4. The pH dependence of the CD at 278 nm and 295 nm is shown in Figure IV-5 and the pH dependence of the positions of the two maxima is shown in Figure IV-6. As can be seen from Figures IV-5 and Figure IV-6, changes in pH do not cause any shifts in position or changes in CD amplitude in the pH range 2.2 - 4.8. I concluded that acidic pH *per se* does not cause any changes in the local surrounding of aromatic chromophores that contribute to the near-uv CD. I assumed here, as is widely accepted (Sears & Beychok, 1973), that only the local surroundings of aromatic chromophores contribute to their electrostatic and hydrophobic environment, and to the formation of hydrogen bonds with aromatic side chains. Thus, according to my results, only those changes in the binase tertiary structure occur that leave the areas around the aromatic chromophores intact.

I then attempted to identify the chromophores that are the least significant in terms of CD contribution. The binase spectrum (Fig.IV-4) displays two clearly resolved bands at 295 nm and at 278 nm with shoulders at 268 nm and 286 nm. Model compounds of Trp exhibit CD at 285 nm and 292 nm in aqueous solutions, and those of Tyr a maximum around 278 nm with a shoulder ~286 nm (Sears & Beychok, 1973). I attributed

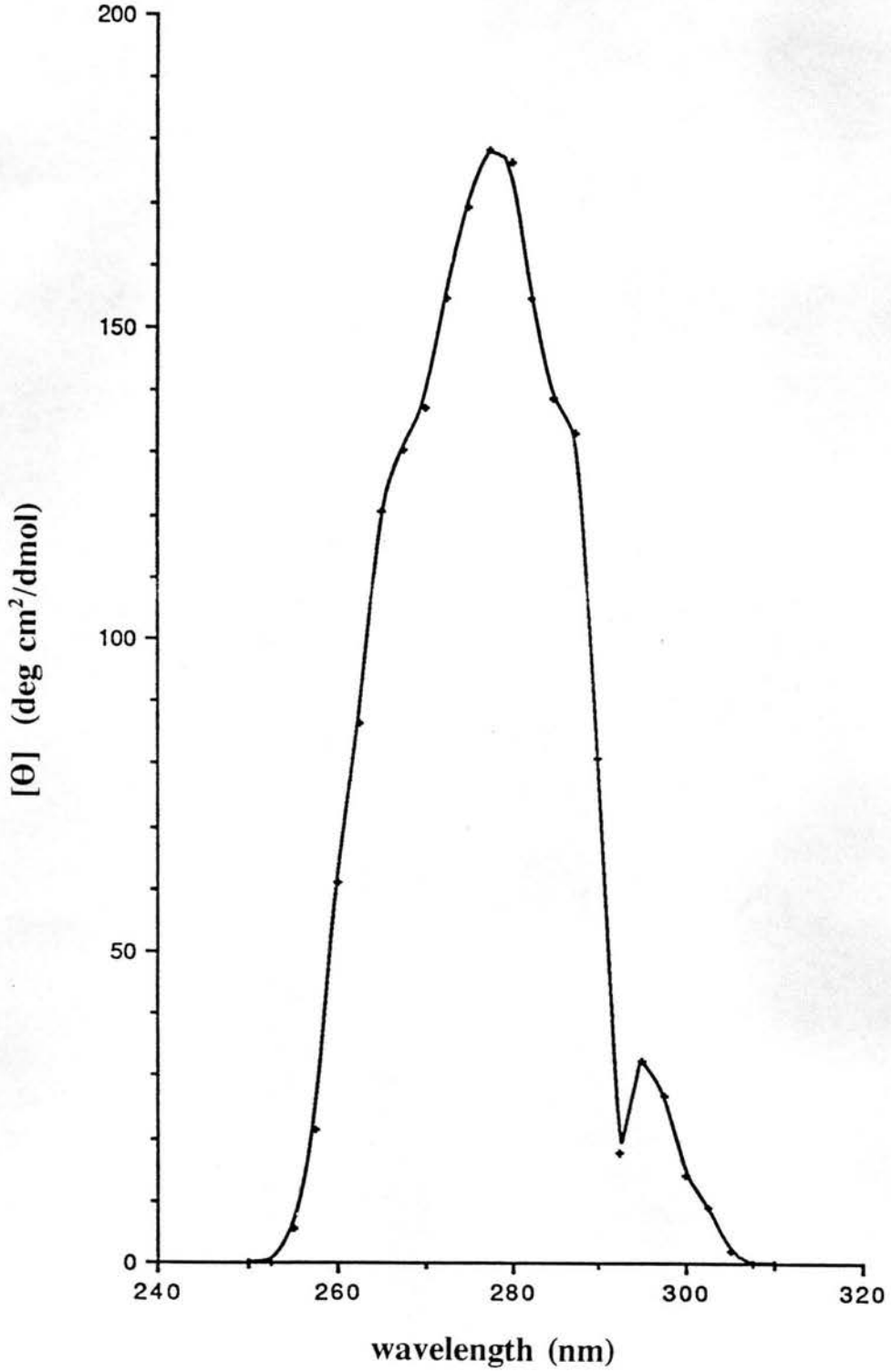


Figure IV-4. Binase experimental near-uv CD.

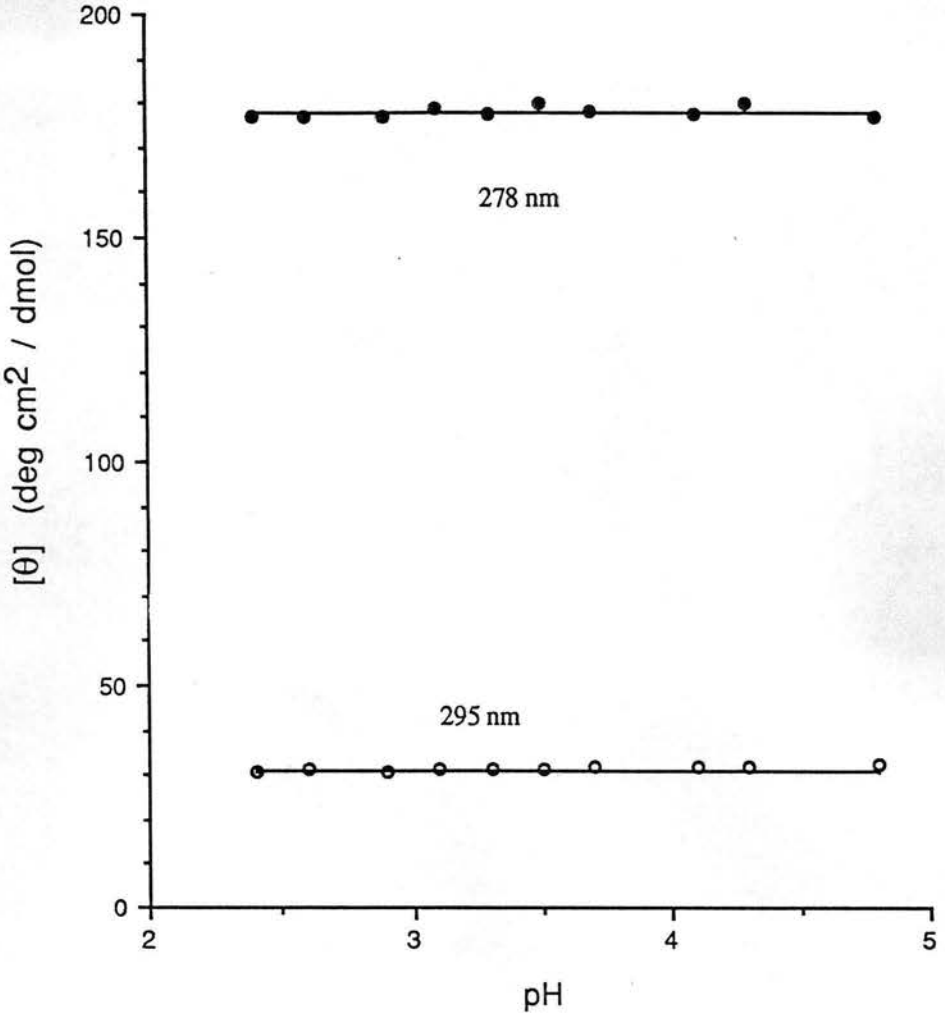


Figure IV-5. Binase CD at pH from 2.4 to 4.8 at 295 nm (o) and 278 nm (•).

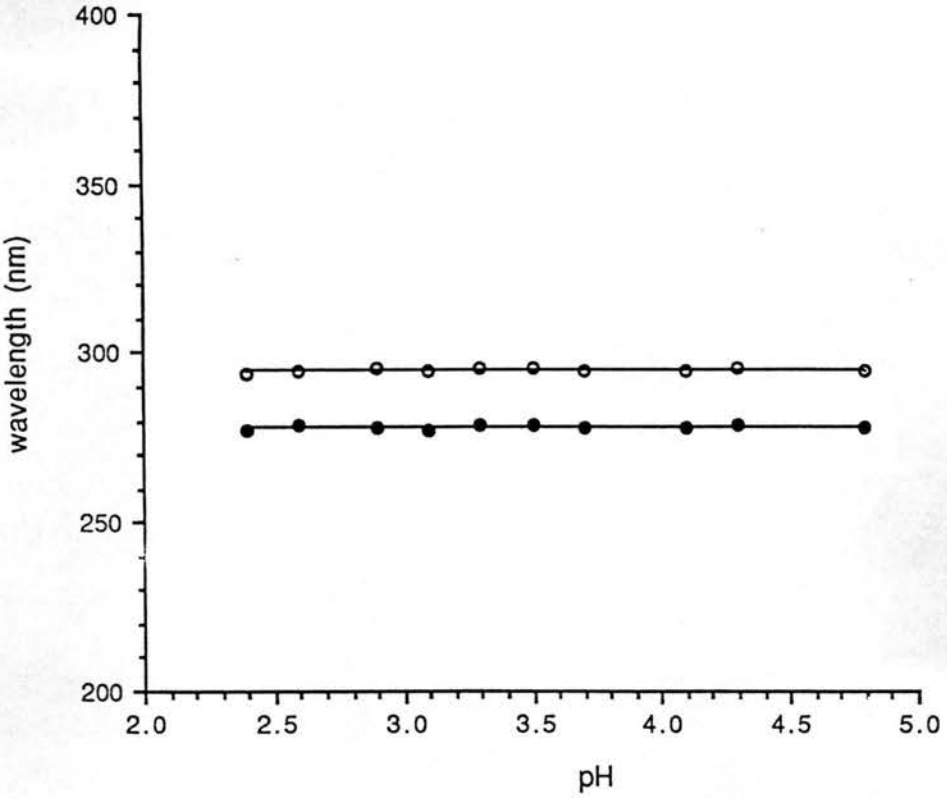


Figure IV-6. Positions of the 278 nm (•) and 295 nm (o) bands in binase CD spectrum at pH from 2.4 to 4.8.

the band at 295 nm to the ^1L_b transition of Trp. The large band at 278 nm was assigned to the L_b transition of Tyr, although Trp might also contribute in this interval. The shoulder around 286 nm could arise from Trp or Tyr contributions, or both. The red shift in CD of the Trp transition in binase spectrum (at 295 nm) relative to the corresponding transition in model compounds in aqueous solution (at 292 nm) might be explained by an apolar environment of the Trp contributing to the CD.

As can be seen from Figure IV-7, Trp93 is exposed to the solvent. Chemical modification of Trp93 does not effect the enzyme activity (Golubenko et al., 1981). I concluded conformational changes at Trp93 are not likely to effect the structure of the globule, and no changes in its CD are likely to be detected due to its flexible conformation.

Trp70 is buried in the hydrophobic β-structure core of the globule (Fig. VI-7) and remains partially buried even in the presence of strong denaturing reagents (Khandanyan & Dudkin, 1979). There are two tyrosines (Tyr89 and Tyr96) located in the β-structure layer, and two other (Tyr77 and Tyr102) in the β-turn regions adjacent to this layer (Table IV-1, Fig.IV-7). The local environment of these residues (Trp70 and Tyr77, Tyr89, Tyr96, Tyr102) is determined by the integrity of the β-structure in this region. In principle, these aromatic chromophores should exhibit significant CD and most of the binase aromatic CD may arise from their contributions.

Tyr23 and Trp34 are located in the N-terminal part of the binase globule, 13 Å apart from each other. Exciton coupling between them might contribute to the CD. The N-terminal part of the globule is mostly α-helical, but includes a short parallel β-

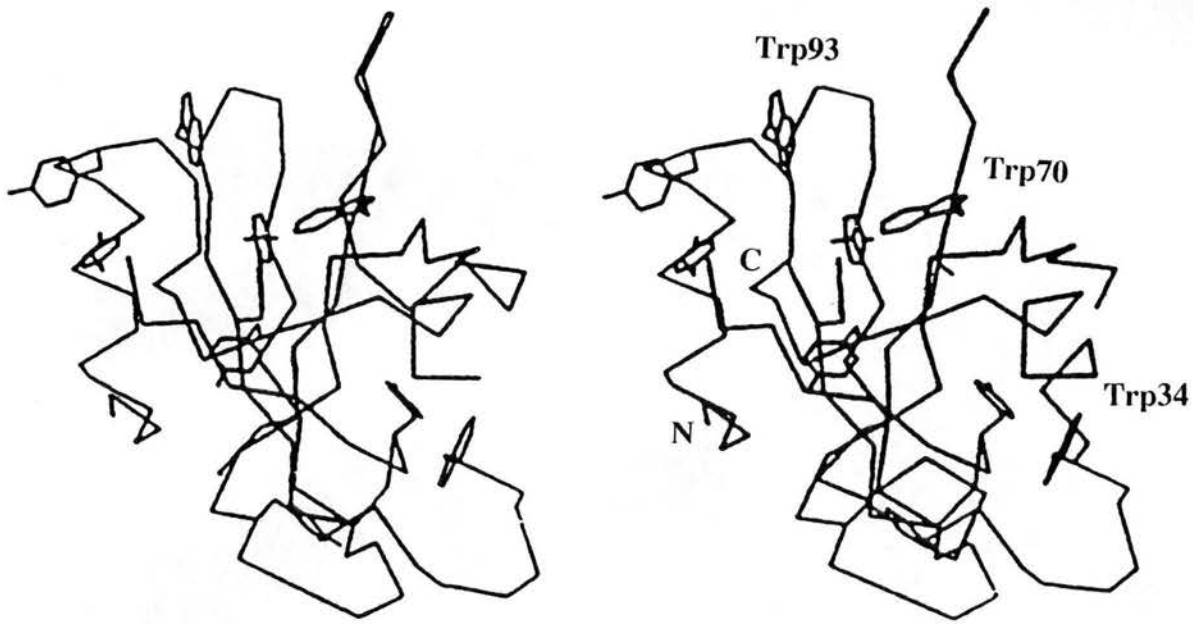


Figure IV-7. Stereopair of the polypeptide chain of binase molecule with aromatic side chains.

Table IV-1. Location of aromatic side chains and ion pairs in the secondary structure of binase.

Residues	Secondary structure	Aromatic side chains	Ion pairs
1-5	unordered		
6-16	α-helix	Phe6, Tyr12, Tyr16	Asp7, Asp11
17-22	undordered		
23-24	parallel β-structure	Tyr23	
26-32	α-helix		Lys26
33-35	unordered	Trp34	
36-39	3_{10}-helix		
40	unordered		
41-44	α-helix		
45-48	unordered		
49-50	parallel β-structure		
51,52,54	antiparallel β-structure		
53,55-69	unordered	Phe55	Arg68 Asp53
70-74	antiparallel β-structure	Trp70	Asp74 Glu72
75-85	unordered	Tyr77, Phe81	Arg82
86-90	antiparallel β-structure	Tyr89	Arg86
91-94	unordered	Trp93	Asp92
95-98	antiparallel β-structure	Tyr96	
99-105,108	unordered	Tyr102, Phe105	
106,107,109	antiparallel β-structure		Arg109

structure (residues 23-24, hydrogen bonded to 49-50), adjacent to the antiparallel β-layer in the C-terminal region (Table IV-1, Fig.IV-7). Trp34 is located in a relatively poorly structured loop (residues 34-47) which contains two short regions of 3_{10}-helix. Trp34 is adjacent to the second α-helix (25-33) and the conformation of the aromatic ring might be stabilized by the local rigidity of the backbone as well as by the interaction with the helix dipole (Šali et al, 1988). It seems likely that conformational changes in the region 20-40 could have been reflected in CD, provided the contribution of Trp34 is sufficiently large. Also, local conformational changes around Trp34 do not have to involve structured regions and no detectable changes in secondary structure might occur, unlike the situation in the β-structure region.

I calculated Trp CD in binase considering all three Trp (Trp34, Trp70 and Trp93), and coupling in each pair separately (Fig.IV-8). It seems that most of the Trp CD in binase results from coupling of the Trp70 and Trp93 in the β-structure region. Coupling of Trp34 with the other Trps is much weaker. Most likely Trp34 CD should arise from an exciton interaction with Tyr23 located in the parallel β-structure at 13.5 Å from Trp34, in a conformation close to P45 (Chapter II) for the coupling between the B_b transition of Trp34 and the L_a transition in Tyr23.

In summary, the analysis of locations of Trp and Tyr in the structure of binase leads to the conclusion that changes in the tertiary structure in the β-structure region are more likely to be detectable by aromatic CD because they could disrupt the coupling between Trp70 and Trp93, and between the Trp and Tyr in this region. Since no changes in CD were detected in the pH range 4.8 to 2.4, we can conclude that if any changes in

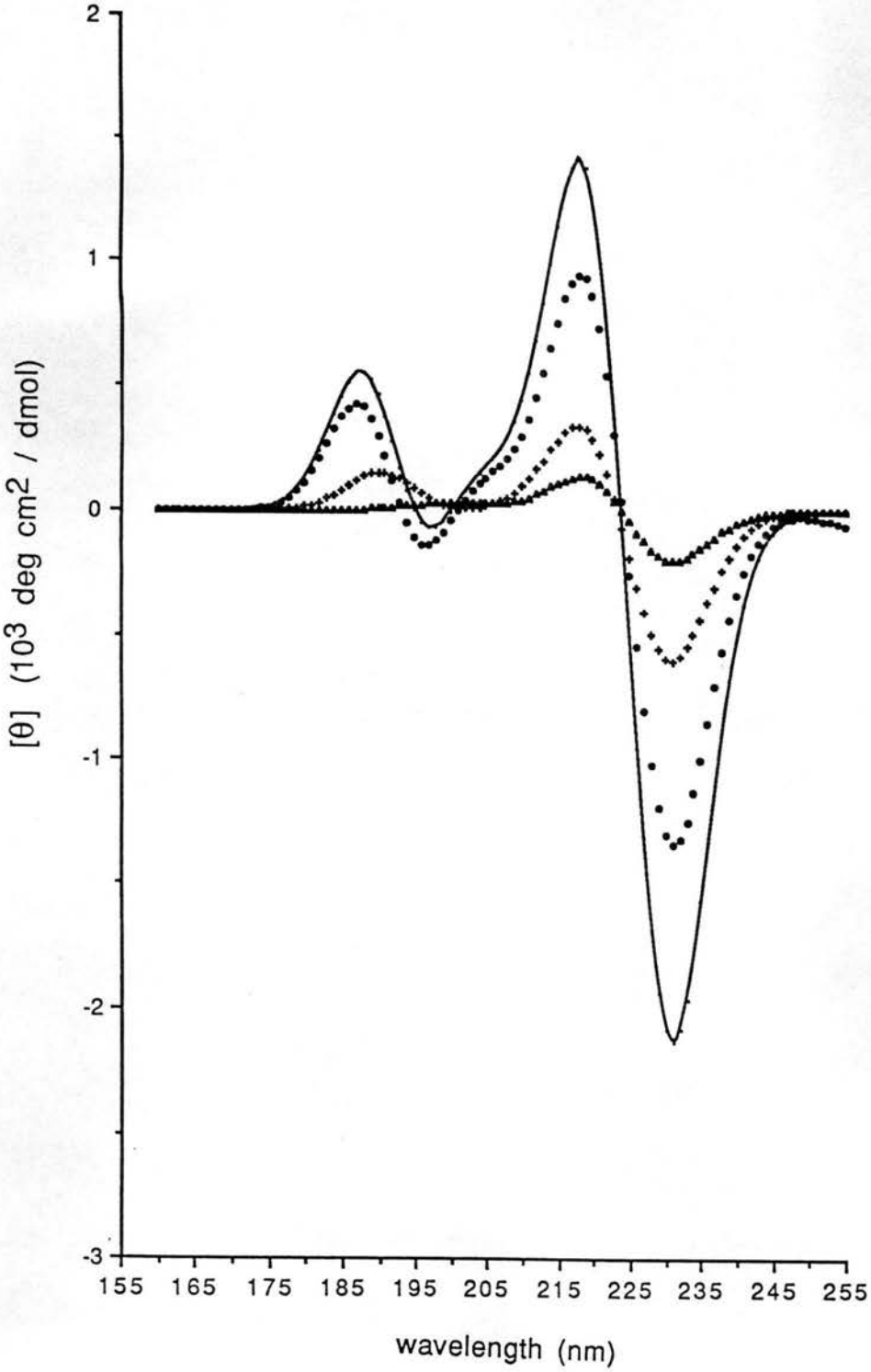


Figure IV-8. Calculated binase Trp CD in the far uv considering all indole transitions (-). Calculated CD of Trp34-Trp70 pair (+), Trp34-Trp93 pair (▲), and Trp70-Trp93 pair (•).

tertiary structure do take place, they are likely to occur in the loop (34-47) between the α-helical and β-structural regions of binase.

(2) Analysis of Ionic Interactions in Binase.

I analyzed ionic interactions in binase in order to determine which ones could be critical in holding together large portions of the globule, and what portions of the globule (energetic domains) those might be. Binase x-ray atomic coordinates were determined and refined earlier to 2 Å resolution (Pavlovsky, 1989) without application of computer graphics. Our goal at this point was to incorporate in the binase structure those residues that were not fitted satisfactorily by previous automatic optimization (Pavlovsky, 1989). It was also necessary to make a few corrections in amino acid sequence of the model structure, which were made after sequencing of the binase gene (Nukrianova et al., 1989). The following positions in the sequence were corrected, relative to the previously published sequence (Golubenko et al., 1979): N21E, E22N, E40N, G66S, S67G. These corrections were confirmed by S. Shliapnikov (Institute of Molecular Biology, Russian Academy of Sciences, Moscow) by sequencing of portions of the protein. Fitting of 30 amino acid residues in the binase electron density has been improved in the process of refinement. Automatic refinement with the TNT program (Tronrud & Ten Eyck, 1990), performed after the manual fitting, did not give any significant improvement in the R-factor, although the changes made in the model helped in my analysis of the hydrogen bonds and ion pairs, and the distribution of the charged residues in the protein.

Analysis of the binase crystal structure was performed with the following objectives: to determine whether, according to the electron density distribution in the

protein, separate structural domains may be distinguished, the thermal denaturation of which would correspond to the two peaks in the denaturation profile (Protasevich, 1985); to determine the regions of secondary structure; to characterize the ionic interactions; and to describe the positions of the aromatic side chains in the secondary and tertiary structure of the molecule, and relative to the ion pairs and potentially charged side chains.

The analysis of binase x-ray data by Pavlovsky et al. (1989) showed that it was not possible to identify sites with sufficiently low electron density that the structural domains could be revealed. Thus, the binase molecule consists of a single structural domain (Chapter IV.1.i.).

The regions of secondary structure have been determined by "rigid" criteria (Bolotina et al., 1980a; Finkel'shtein et al., 1977), together with A. Finkelshtein and A. Pavlovsky. It can be seen (Fig.IV-7, Table IV-1) that binase is a typical $\alpha+\beta$ protein (Levitt & Chotia, 1976); the α-helical and β-structural regions are distinct and are separated along the amino acid sequence.

Since the number of cooperatively melting regions in the binase molecule depends on pH, we assumed that the properties of energetic domains depend on the electrostatic interactions in the globule and changes in the number of ion pairs may be a major factor in the separation of the energetic domains. I should point out that neutralization of one of the charged residues in the ion pair actually creates an uncompensated charge, located on the other partner.

Barlow & Thornton (1983) found two maxima in the statistical distribution of distances between charged side-chains of Arg, Lys, His, Asp, Glu, or the N- and C-terminus in globular proteins, which were at 4 Å and at 6 Å, and proposed to define ion contacts as those which lie in that range. This definition does not give any consideration to the possible electrostatic state of the group due to formation of hydrogen bonds in the protein globule. The strength of salt bridges is determined by a number of factors: the geometry and distances of the interaction; the degree of exposure to the solvent; the effect of neighboring charged residues (Horovitz et al., 1990).

I identified the salt bridges in binase (Table IV-2) and analyzed, whenever possible, their local environment. Simultaneously, the ionic interactions and hydrogen bond formation which involve those charged residues were analyzed. The thorough study has been performed for molecule 1 (Table IV-2), but I also compared the two binase molecules in the asymmetric unit (Table IV-3).

Most of the contacts formed by charged side chains (i.e. potentially charged side chains of Asp, Glu, Arg, Lys, His, C-terminus and N-terminus) in the binase structure refined with FRODO coincide with the previous structure within a 0.05 Å margin, although I detected differences in some contact distances of up to 2 Å. In those cases, ionic or hydrogen bonds determined in the previous refinement would correspond to contacts of a different type after FRODO refinement (for instance ionic instead of hydrogen-bonded), or not be considered as physical contacts at all. I performed all the visual analysis on the binase 3D structure using the graphic package VisiCoor (Kuznetsov

Table IV-2. Interactions of the charged residues in binase. (molecule 1).

CLUSTER	Acceptor	Donor	Angle, degrees**	Distance, Å	Contact type
	OD2 Asp7 (13)	NH1 Arg109 (18)	179	3.4	ionic
	OD1 Asp11 (19)	NH1 Arg109 (18)	173	2.5	H-bond
I	OD2 Asp11 (9)	NH1 Arg15 (27)		6.1	ionic
	O Ile108 (15)	NH2 Arg109 (13)	155	3.0	H-bond
II	OD2 Asp21 (34)	NH1 Arg18 (44)		3.2	ionic
	O Asp21 (19)	NZ Lys48 (30)	153	3.2	H-bond
	OD1Asp53 (15)	NH1Arg71 (22)		6.5	ionic
	OD2 Asp53 (12)	N Lys26 (9)	157	2.7	H-bond
	OD2 Asp53 (12)	NZ Lys26 (27)		5.3	ionic
	OD1 Asp53 (15)	NZ Lys26 (27)		5.3	ionic
	OE1 Glu72 (45)	NZ Lys26 (27)		4.5	ionic
	OE2 Glu72 (40)	NZ Lys26 (27)		4.5	ionic
III	OE2 Glu72 (40)	OH Tyr102 (33)	119	3.0	H-bond
	OD1 Asp74 (9)	NH2 Arg82 (15)	168	2.7	H-bond
	OD2 Asp74 (14)	NH1 Arg82 (13)		5.0	ionic
	OD2 Asp74 (14)	NH1 Arg86 (7)		4.3	ionic
	OD2 Asp74 (14)	NH2 Arg86 (8)		4.6	H-bond
	OD2 Asp85 (21)	ND2 Asn76 (14)	154	2.6	H-bond

Table IV-2 (continued). Interactions of the charged residues in binase (molecule 1).

CLUSTER	Acceptor	Donor	Angle, degrees	Distance, Å	Contact type
IV	OE1 Glu59 (77)	NH1 Arg58 (45)		5.4	ionic
	OE1 Glu59 (77)	NH1 Arg61 (52)		3.9	ionic
V	OD1 Asp92 (34)	NH1 Arg68 (25)	179	2.3	H-bond
	OD2 Asp92 (12)	OG Ser90 (17)	177	2.5	H-bond
	OD2 Asp92 (12)	N Asp92 (25)	91	2.5	H-bond

* Atomic B-factors are shown in parenthesis

** Angle is calculated for donor-proton-acceptor of a hydrogen bond at distances $\leq$ 3.5 Å.

Table IV-3. Interactions of the charged residues in binase. Comparison between molecules 1 and 2.

	Molecule 1				Molecule 2			
CLUSTER	Acceptor	Donor	Distance, Å	Contact	Acceptor	Donor	Distance, Å	Contact
I	OD2 Asp7	NH1 Arg109	3.4	ionic	OD2 Asp7	NH1 Arg109	2.4	H-bond
	OD1 Asp11	NH1 Arg109	2.5	H-bond	OD1 Asp11	NH1 Arg109	3.6	ionic
	OD2 Asp11	NH1 Arg15	6.1	ionic	OD2 Asp11	OG1 Thr15	6.1	ionic
	O Ile108	NH2 Arg109	3.0	H-bond	O Ile108	NH2 Arg109	2.7	H-bond
II	OD2 Asp21	NH1 Arg18	3.2	ionic	OD1 Asp21	NH1Arg18	2.8	ionic
	O Asp21	NZ Lys48	3.2	H-bond	O Asn21	NZ Lys48	3.4	weak H-bond
III	OD1Asp53	NH1Arg71	6.5	ionic	OD1 Asp53	NH1 Arg71	5.3	ionic
	OD2 Asp53	N Lys26	2.7	H-bond	OD2 Asp53	N Lys26	2.9	H-bond
	OD2 Asp53	NZ Lys26	5.3	ionic	OD2 Asp53	NZ Lys26	4.9	ionic
	OE1 Glu72	NZ Lys26	4.5	ionic	OE1 Glu72	NZ Lys26	3.7	ionic
	OE2 Glu72	OH Tyr102	3.0	H-bond	OE2 Glu72	OH Tyr102	3.6	H-bond
	OD1 Asp74	NH2 Arg82	2.7	H-bond	OD1 Asp74	NH2 Arg82	2.7	H-bond
	OD2 Asp74	NH1 Arg82	5.0	ionic	OD2 Asp74	NH1 Arg82	4.8	ionic
	OD2 Asp85	ND2 Asn76	2.6	H-bond	OD2 Asp85	ND2 Asn76	2.9	H-bond
	OD2 Asp74	NH1 Arg86	4.3	ionic	OD2 Asp74	NH1 Arg86	3.6	ionic
	OD2 Asp74	NH2 Arg86	4.6	ionic	OD2 Asp74	NH2 Arg85	4.0	ionic
*	NZ Lys26	NH2 Arg82	4.9	ionic	NZ Lys26	NH2 Arg82	5.1	ionic

* Like charges.

Table IV-3. (continued).

CLUSTER	Molecule 1 Acceptor	Donor	Distance, Å	Contact	Molecule 2 Acceptor	Donor	Distance, Å	Contact
IV	OE1 Glu59	NH1 Arg58	5.4	ionic	OE1 Glu59	NH1 Arg59	5.7	ionic
	OE1 Glu59	NH1 Arg61	3.9	ionic	OE1 Glu59	NH1 Arg62	5.6	ionic
	OE2 Glu59	N Glu59	2.7	H-bond	OE2 Glu59	N Glu59	2.8	H-bond
V	OD1 Asp92	NH1 Arg68	2.3	ionic**	OD1 Asp92	NH1 Arg68	2.5	ionic
	OD2 Asp92	OG Ser90	2.5	H-bond	OD2 Asp92	OG Ser90	2.7	H-bond
	OD2 Asp92	NH2 Arg68	4.9	weak ionic	OD2 Asp92	NH2 Arg68	3 .5	ionic
	OD2 Asp92	N Asp92	2.5	H-bond	OD2 Asp92	N Asp92	2.7	H-bond

**see text.

& Lim, 1992). I completed the set of crystallographic atomic coordinates by adding protons in the appropriate positions (see Methods,Chapter IV.2.b).

The carboxyl groups may be a donor of a proton in the contacts with O, or an acceptor in contacts with NH_2 and NH_3^+. While characterizing contacts formed by a carboxyl group, I considered several situations, corresponding to different electrostatic conditions. If the side-chain carboxyl group is ionized, it might participate in the following type of contacts: 1) one or several ionic contacts through the charge delocalized between the carboxyl oxygens, while no hydrogen bonds with uncharged atoms are formed; 2) one of the oxygens participates as acceptor in a hydrogen bond with an uncharged atom, which induces a shift of negative charge towards the other oxygen, which may form an ion pair with a positively charged atom ; 3) both oxygens form hydrogen bonds with uncharged atoms, thus shielding the negative charge of the group and weakening possible ionic interactions. The case in which one of the oxygens participates in a hydrogen bond as a proton donor is essentially indistinguishable from (1), since it is hard to discriminate between a polarized hydrogen bond and an ion pair, and because the negative charge is likely to be delocalized between the two oxygens, according to the distribution of electronic density of the 1.5 bond.

The screening of the binase structure for possible ionic contacts was performed in two stages. First a spherical volume of 3.5 Å radius around every side-chain and peptide C-terminal carboxyl oxygen was searched for possible hydrogen bonds formed by those oxygens. Then those Asp and Glu side-chain carboxyl groups were tested for whether they are within 6.5 Å of a nitrogen atom of Arg or Lys. Contacts with possible

partners were verified for all non-hydrogen atoms of the selected carboxyl groups, guanidinium groups and NZ of Lyz. Lengths and angles for all possible hydrogen bonds formed by selected atoms were calculated. Asp and Glu carboxyl groups were considered to participate in ionic contacts if they fall into cases (1) or (2) described above, and if at least one of the atoms in the carboxyl group was within 6.0 Å of one of the atoms of the Arg guanidinium group or the NZ of Lys. The B-factors of the atoms participating in ion pairs were also taken into consideration, since the groups participating in ionic contacts (or hydrogen bonds) are usually more rigid than the residues of the same type that do not form those contacts (Barlow & Thornton, 1983).

I found that most of the charged side chains are concentrated near other charged side chains with the same or opposite charge, participating in multiple ionic interactions, forming what will be subsequently referred to as clusters. Clusters of charged residues in the binase structure are presented in Table IV-2, in order of the Asp and Glu in the binase sequence.

Cluster I includes Asp7, Asp11, Arg15, Lys97 and Arg109. Most interactions in this cluster would be favorable for holding together the α-helical N-terminal region (amino acids 1-40) and the C-terminal β-structure part of binase (amino acids 40-109). NH2 of Arg109 forms a hydrogen bond with the main-chain carbonyl of O Ile108, while NH1 Arg109 may form ionic contacts with OD2 Asp7 and OD1 Asp11 (Table IV-2, Fig.IV-9). The long ionic contact (6.1 Å) between OD2 Asp11 and NH1 Arg15 may play a weakly stabilizing role for the N-terminal α-helix.

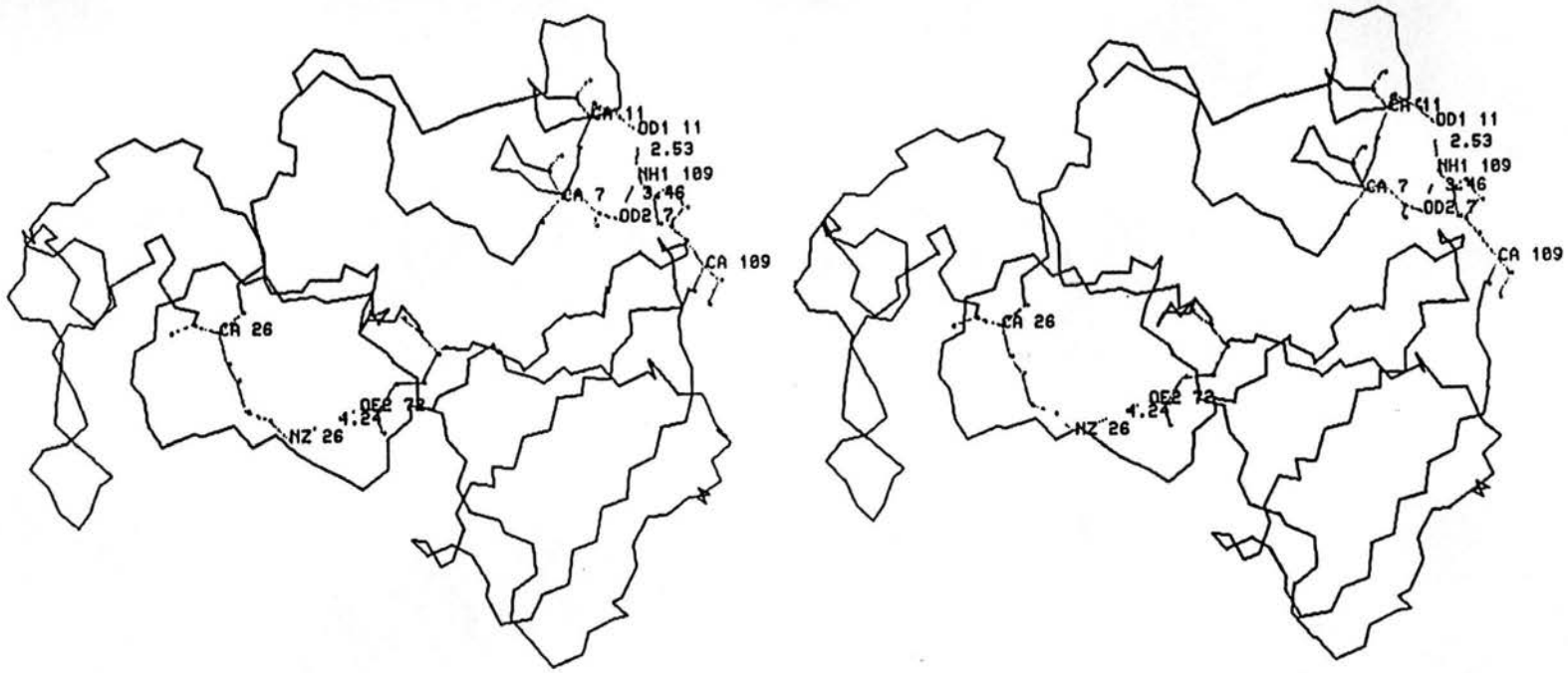


Figure IV-9. Stereopair of the polypeptide chain of binase molecule. Ionic triads Asp7, Asp11 - Arg109 and Lys26 - Asp53, Glu72 between N-terminal α-helical and C-terminal β-structural regions are shown.

The charged residues forming Cluster II - Arg18, Asp21 and Lys48 - are also located between the α-helical and β-structural parts of binase. Lys48 and Asp21 do not form an ion pair, since the distance from NZ Lys48 to the carboxyl group of Asp21 is 7.5 Å. However, NZ Lys48 forms a hydrogen bond with the O of the backbone carbonyl Asp21. OD2 Asp21 may form an ionic contact on the surface of the globule with Arg18. Thus we have found an ionic cluster that might play a critical role in the region connecting the α-helical and β-structural parts of binase, and we have also found at least two ionic contacts on the surface of the N-terminal α-helices that might stabilize them.

Cluster III incorporates side chains that are important for the enzyme activity (Pavlovsky & Sanishvili, 1988; Sanishvili, 1989): Lys26, Glu72, Arg82, Arg86 and His101. Other charged residues in this cluster are Asp53, Arg71, Asp74 and Asp85. Cluster III includes all the charged residues from the active site of binase and is the largest cluster of charged residues in the protein (Fig.IV-9, Table IV-2). Is this a coincidence? The core of the cluster is composed of a short triple ionic contact Asp74 - Arg82, Arg86, and two ion pairs, in each of which one component belongs to the N-terminal and the other to the C-terminal part of the globule: Lys26 - Glu72 and Lys26 - Asp53 (Fig.IV-9). One of the bonds formed by the carboxyl group of Asp74, between OD1 Asp74 and NH2 Arg82, is rather short and its geometry corresponds to that of a hydrogen bond, although since we assume that this contact is formed by charged atoms we might consider it an ion pair (Table IV-2). The N of Lys26 is a proton donor in the hydrogen bond formed with OD2 Asp53. The negative charge of the carboxyl group of Asp53 might still be considered as localized on the other oxygen (OD1) or between

oxygens. This cases satisfies condition (2). The carboxyl group might still form an ionic contact with NZ of Lyz26.

Protonation of the carboxyl groups of Asp74 and Glu72 would lead to the disruption of the ionic contacts Asp72 - Arg82, Arg86 and Lys26 - Asp53, Glu72. This would leave the charges on Lys26, Arg82 and Arg86 uncompensated. The charges on Lys26 and Arg82 are relatively close (at 4.9 Å) in the crystal structure of native binase (Table IV-2). Repulsion between those charges might contribute to the separation of the N-terminal (which includes Lys26) and C-terminal (which includes Arg82 and Arg86). Protonation of Asp74 would also lead to strong repulsion between charges on Arg82, Arg86 and His101 located in close proximity to each other in the active site (Table IV-4).

Cluster IV consists of Arg58, Glu59 and Arg61 located on the surface of the binase globule. These charged residues might be involved in binding the substrate (Sanishvili, 1989). The negative charge on OE1 Glu59 could electrostatically interact with Arg58 and Arg61. OE2 Glu59 is almost at hydrogen bonding distance to the backbone N of Glu59 (3.5 Å), although the angle between donor-proton-acceptor (123°) would be unfavorable to form a hydrogen bond.

The binase crystals contain two molecules per asymmetric unit and most of our conclusions are drawn for molecule 1 (Pavlovsky & Sanishvili, 1988). I compared all the contacts of interest between the two molecules (Table IV-3, B-factors not shown). Sometimes local differences between the two molecules of binase were found. Especially in the case of the exposed side chain of Glu59, the B-factors for OE1 and OE2 differ significantly between the two molecules. They equal 77 and 72, respectively, for the first

Table IV-4. Pairs of close positively charged residues in binase (molecule 1).

Interacting Residues		Distance (Å)
NZ Lyz26 (27)*	NH2 Arg82 (15)	4.9
ND1 His101 (50)	NH2 Arg86 (8)	3.4
NE2 His101 (59)	NH2 Arg82 (15)	4.2
NH2 Arg82 (15)	NH2 Arg86 (8)	3.4

* Atomic values for the B-factors are given in parenthesis.

molecule, which means those positions are very flexible in the crystal. B-factors above 40 generally are considered to denote flexibility of the corresponding atom. The same atoms in the second molecule have B-factors of 19 and 12, which indicates rigidity of the region and, most probably, participation in hydrogen bonds (generally characterized by B-factors less than 25). Thus the side chain is rigidly hydrogen bonded in one molecule and free in the other. Such local differences between two molecules in the same crystal indicates possible variety in local stable conformations in the native structure, which could correspond to different local energy minima. The two binase molecules also have some differences in intermolecular contacts, which might play some role in stabilizing specific conformations of the surface residues. For instance OE1 Glu59 in molecule 2 (B=19) is hydrogen bonded to N Phe81 of the first molecule. No such contact stabilizing the local conformation exists for molecule 1.

The RMS (root mean square) deviation for corresponding N, C_α and C atoms, calculated after superposition of the two binase molecules, was 0.27 Å before refinement with FRODO and 0.31 Å for the refined crystal. The differences between the two molecules actually became more pronounced after refinement (Frank Eisenmenger, personal communication; Grishina et al., 1993). This also favors my conclusion on the variety of local conformations in the native state.

Cluster V is represented by an isolated ion pair in binase. The close ion pair Arg68 - Asp92 is partially exposed on the surface of the globule and connects two β-strands. OD2 Asp92 might be involved in two hydrogen bonds: with OG of Ser90 and N Asp92, and probably plays an acceptor role in at least one of the contacts. Thus the

ionic contact with Arg68 is more likely to be formed through OD1 Asp92, even though the distance between OD2 of Asp92 and NH2 of Arg68 (4.9 Å) satisfies the criteria for a weak ionic contact. The unusually short distance of the ionic contact between OD2 of Asp92 and NH1 of Arg68 (2.3 Å) is a source of concern, although it is most probably an artifact of the refinement. The fact that OD1 Asp92 was not found to form any hydrogen bonds with uncharged atoms confirms the classification for the state of the carboxyl groups proposed above.

In summary, I found five regions in the structure of binase with concentrations of charge-charge interactions referred to as clusters. In several cases these are isolated ionic contacts that might stabilize structured regions. Ionic clusters obviously play an active role in substrate binding and enzymatic activity of the protein, and probably in regulation of both functions in response to the pH of the medium. Two triple ionic contacts that belong to separate clusters may regulate interactions between the N-terminal α-helical and C-terminal β-structural lobes of binase, which could represent the energetic domains of the protein (Protasevich et al., 1987; Grishina et al., 1989, 1993). The summary of ionic contacts between the α-helical and β-structural parts of binase is given in Figure IV-10. Acidification of binase solution below pH 4.0, leads to protonation of Asp (pK 3.9) and Glu (pK 4.2). The disruption of ionic interactions in the hinge region between energetic domains and repulsion between uncompensated positive charges on Arg and Lys might explain the independent thermal denaturation of those parts of the binase globule at acidic pH. Repulsion between positive charges on Arg 82,

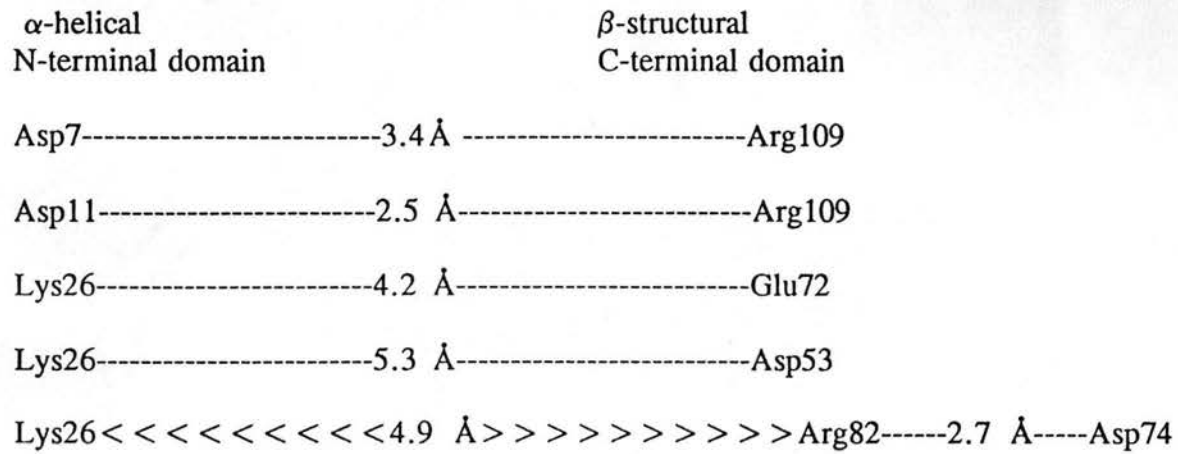


Figure IV-10. Most important ionic interactions between the N-terminal α-helical and C-terminal β-structural parts of binase, constituting energetic domains. Attraction (--). Repulsion (< >).

Arg86 and His 101 might contribute to the relative decrease in thermostability of the β-structural domain compared to the α-helical domain (Grishina, 1989; 1993).

(3) Estimation of Possible Secondary Structure Distribution in Energetic Domains.

I found several electrostatic contacts in the structure of binase. Weakening or disruption of some or all of those ionic pairs at acidic pH might explain why the binase globule, which denatures cooperatively at neutral pH, separates into two independent cooperative systems below pH 4.0 (Fig.IV-3). My goal at this point was to determine what ionic pairs might be crucial for separation of the energetic domains in binase, and what regions of the molecule correspond to the energetic domains.

The mass ratio of the binase domains that melt independently at pH 2.4 can be obtained using the rule (Privalov & Khechinashvili, 1974) that states that for all cooperative protein globules ΔH in cal/g (enthalpy per unit mass) equals an empirical constant at 110°C (Chapter IV.1.e). Indeed, since parts of the globule melt cooperatively at acidic pH, as does the whole globule at neutral pH, the following relation at 110°C would apply:

$$\Delta Q / x = \Delta Q_1 / x_1 = \Delta Q_2 / x_2 = \text{Constant}, \qquad \text{(IV-36)}$$

where ΔQ, ΔQ_1 ,ΔQ_2 is the heat of denaturation of the whole molecule and its parts, respectively at 110°C; x, x_1, x_2 are the masses of the whole protein and its parts, in grams.

$$x = Mn, \quad x_1 = M_1 n_1, \quad x = M_2 n_2 , \qquad \text{(IV-37)}$$

where M, M_1, M_2 are the relative molecular masses of the whole molecule and its parts, and n, n_1, n_2 are numbers of protein molecules or their parts in the sample. We assume that each protein globule separates at acidic pH into two cooperative domains. Then

$$n = n_1 = n_2. \qquad \text{(IV-38)}$$

Consequently, at 110°C

$$\Delta Q_1 / \Delta Q_2 = M_1 / M_2 = \text{constant}. \qquad \text{(IV-39)}$$

ΔQ_1 and ΔQ_2 could be obtained from extrapolation of $Q_i(T)$ to 110°C, and the ratio of relative molecular masses for the cooperatively melting regions of binase could be obtained from Equation IV-39. This procedure was performed for the two energetic domains in binase (Protasevich et al., 1987), and the value of the constant in Equation IV-39 was obtained as 7.3/5.0. Thus, the portion of the globule that denaturates at the lower temperature, giving rise to the first peak of heat capacity, is 1.46 times larger by weight than the portion that denatures at higher temperatures.

I calculated the mass of binase fragments that satisfy this ratio and account for the complete polypeptide chain, on the basis of the known molecular weights of the amino acids. There might be two possibilities, as either the N- or C-terminal portions of binase could correspond to the larger energetic domain. One case would correspond to separation of the 1-69 region from the 70-109, and the other to separation of the 1-46 region from 47-109. Assuming a 10 percent error in estimation of the domain mass ratio from calorimetric measurements (Privalov, 1974; Protasevich, 1985b), the border of the N-terminal energetic domains could extend in the first case from residues 65 to 72, and in the second case from residues 40 to 51.

If the separation into energetic domains generates a 1-69 N-terminal domain, then disruption of most of the ionic interactions between N- and C-terminal parts of the molecule might be taking place (Table IV-2). In this case, the ion pair Arg68 - Asp92

and triad Asp74-Atg82, Arg86 are the closest ionic interactions near the interface, thus their disruption would be most destabilizing. In this case the energetic domain of the larger mass would be largely α-helical.

In the other case, if the N-terminal energetic domain is created in the region 1 to 40-51, then the disruption of the close Asp7, Asp11 - Arg109 ionic contact would be the most critical, which might be accompanied by weakening or disruption of the looser contact Lys26 - Asp53, Glu72, enhanced by the repulsion between Lys26 and Arg82 (Table IV-4). In this case the α-helical part of binase will be localized within the smaller energetic domain, the larger domain being β-structural. The ionic interactions in the β-structural domain (Arg68 - Asp92 and Asp74 - Arg82, Arg86) might stay intact in this case. We do not have enougth data about the conformation of the active site of binase at acidic pH. According to our results, it might be preserved in either model of the domain separation, although the wild-type binase is deactivated below pH 4.0 (Karpeiskii et al., 1981), probably due to protonation of the imidazole of His101.

If we assume that any of the Asp and Glu residues involved in ionic interactions are protonated at pH 2.4 at 15°C, then we would need more structural information on binase behavior during the thermal denaturation in order to discriminate between the ion pairs based on their role in separation of the energetic domains. If ionic contacts within the β-structure are disrupted at pH 2.4 (Asp74 - Arg82, Arg86), we might expect at least partial acidic denaturation of the β-structural domain. This does not occur, so we have to assume that, at least until thermal denaturation takes place, Asp74 is not protonated.

(4) Estimation of Secondary Structure Content in Native Binase.

The far-uv CD spectrum of binase at pH 4.8 at 15°C is shown in Fig.IV-11a and the spectrum at pH 2.4 and 15°C is shown in Fig.IV-11b. The calculated (Bolotina & Lugauskas, 1985; Chapter III.2.e) "aromatic" parts of each spectrum are also presented. The secondary structure content of binase estimated by this method is shown in Table IV-5. It is clear from Table IV-5 and from direct comparison of the spectra that no changes occur in the secondary structure of binase from pH 4.8 to pH 2.4 at 15°C. The assignment of secondary structure from x-ray coordinates has been performed using the rigid criteria (Chapter III.2.e.(2)) in collaboration with A. Finkel'shtein (Insitute of Protein, Puschino, Moscow region). My estimates of the binase secondary structure content are very close to those determined by x-ray, $f_\alpha = 0.20$ and $f_\beta = 0.22$

I analyzed the accuracy of the CD method. The correlation coefficients between CD estimates and x-ray analyses for nine proteins excluded from the basis set were 0.977 for α-helical and 0.964 for β-structural regions, and the RMS found were 0.04 for both conformations. Thus, my estimates of the binase secondary structure content from CD coincide with x-ray within the error of the method. My estimates are considerably better than those obtained through deconvolution on the same fixed reference spectra, but without accounting for aromatic contributions (NCAC, Chapter III.2.e.(4)). Bolotina et al. (1979) obtained only 0.06 fraction of α-helix.

I compared my estimates of binase secondary structure obtained with the Bolotina & Lugauskas (1985) method, which is normally performed on spectra in the 200 - 250 nm wavelength range, with the PG method (Provencher & Glöckner; 1981) in the 195 -

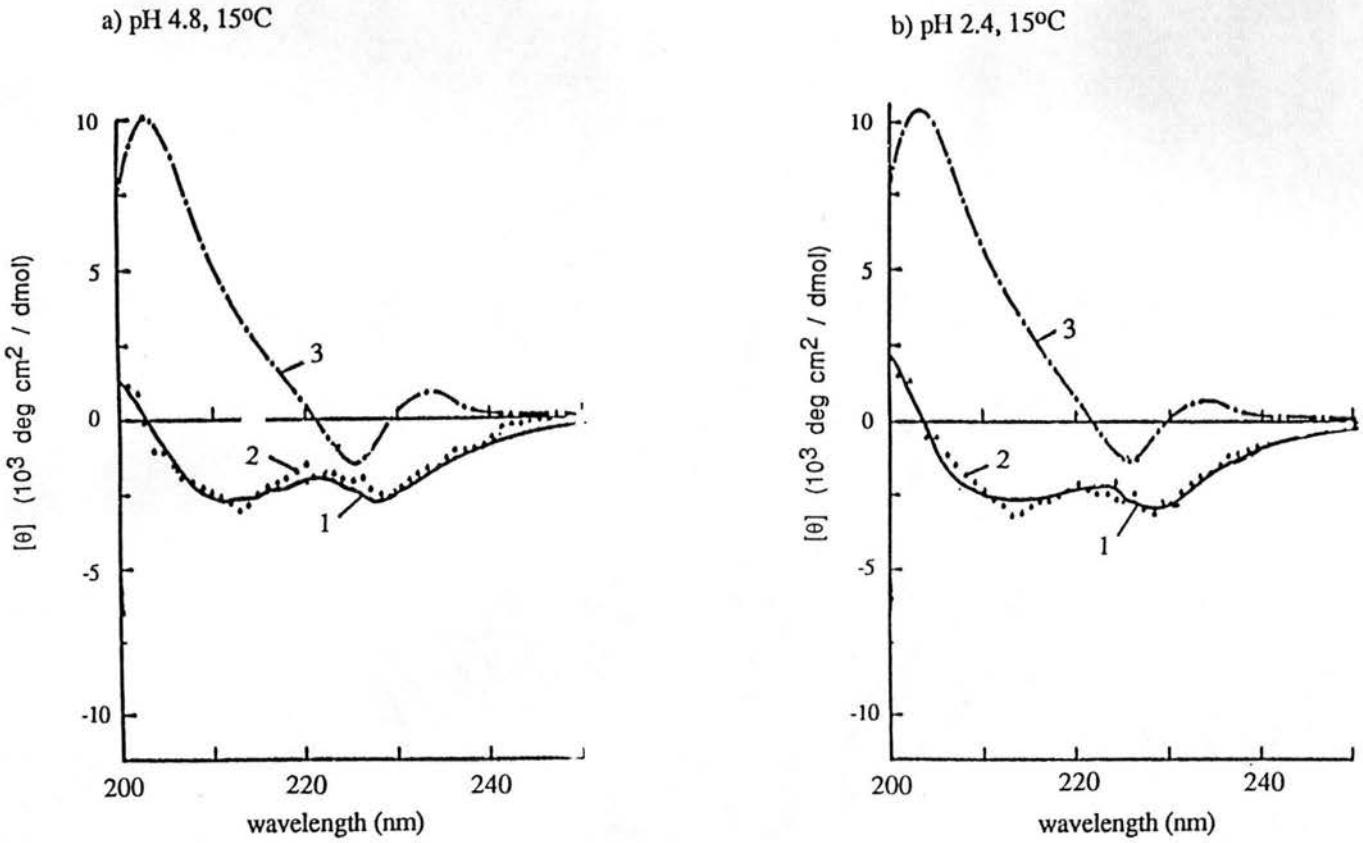


Figure IV-11. Far-uv CD of binase at 15°C at pH 4.8 (a) and pH 2.4 (b). Experimental spectrum (1), calculated spectrum (2), calculated "aromatic part" of the spectrum (3). (Adapted from Grishina et al. (1989).)

Table IV-5. Secondary structure content of binase determined from CD spectra. Values are given in molar fractions. The error for the values in the table is $\pm$ 0.02.

Conditions	f_{α}	f_{β}	$f_{\beta\text{-turns}}$	f_{u}
pH 4.8, 15°C	0.18	0.18	0.25	0.39
pH 2.4, 15°C	0.18	0.18	0.24	0.40
pH 2.4, 30.5°C	0.17	0.09	0.20	0.54

240 nm range, which gave f_α = 0.00, f_β = 0.60, f_t= 0.31 and f_u = 0.09. The discrepancies between the PG estimates and x-ray could be explained by the significant deviation of the binase spectrum from those in the PG basis set with similar secondary structure content, and by the inability of the PG method to give good results for proteins with significantly different CD from those in the selected basis set.

The binase far-uv CD is truly quite unique for a protein with quite ordinary secondary structure content. The negative peak at 229 nm would be especially hard to explain based solely on the secondary structure CD. Positive bands around 230 nm observed in several snake toxins (Menez et al., 1976; Yang et al., 1968; Cameron & Tu, 1977), gene 5 protein from filamentous bacteriophages (Day, 1973), acid DNAase (Timasheff & Bernardi, 1970) and avidin (Green & Melamed, 1966), have been attributed to Trp and Tyr, since electrically allowed transitions are observed in both of those chromophores near 225 nm, the B_b transition in Trp and L_a transition in Tyr (Platt, 1949). Binase contains 3 Trp, 7 Tyr and 4 Phe. The secondary structure content of binase is similar to that of pancreatic ribonuclease A, although the CD of ribonuclease A is almost four times larger than that of binase in the 210 -220 nm region (Fig.IV-12) Ribonuclease A contains no Trp. The estimated "aromatic" CD for ribonuclease A is smaller than that of binase (Fig.IV-12). One can see that accounting for aromatic contributions is especially important for the analysis of binase CD. Experimental and "aromatic" difference spectra between binase and pancreatic ribonuclease A are shown in Figure IV-13 and compared with calculated Trp CD in binase. One can see a good correlation between the couplet feature in the "aromatic" difference spectrum around 215

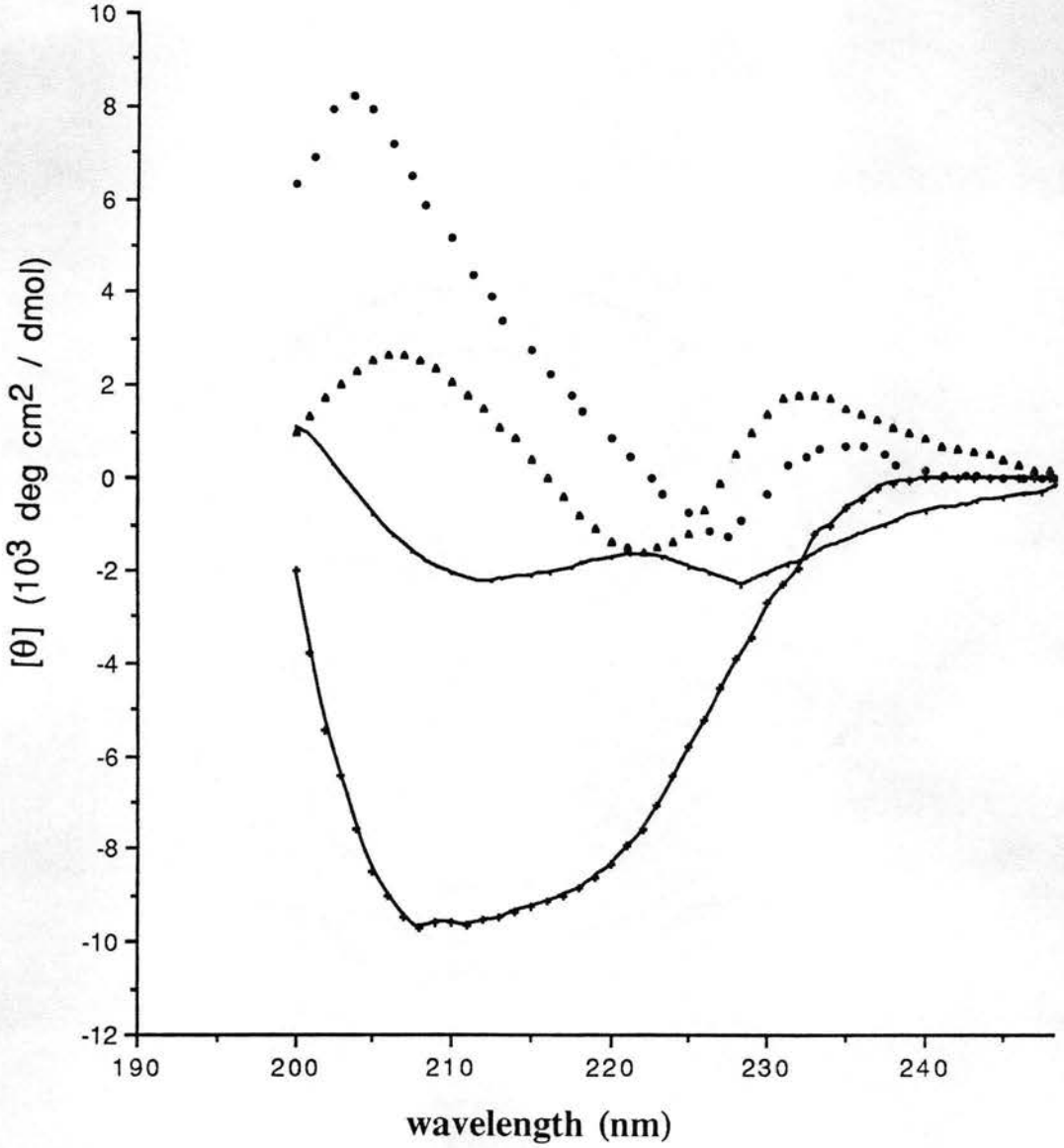


Figure IV-12. Experimental CD spectra of binase (_) and pancreatic ribonuclease A (+). Aromatic contribution obtained by Bolotina & Lugauskas (1985) method for binase (•) and pancreatic ribonuclease A (▲)

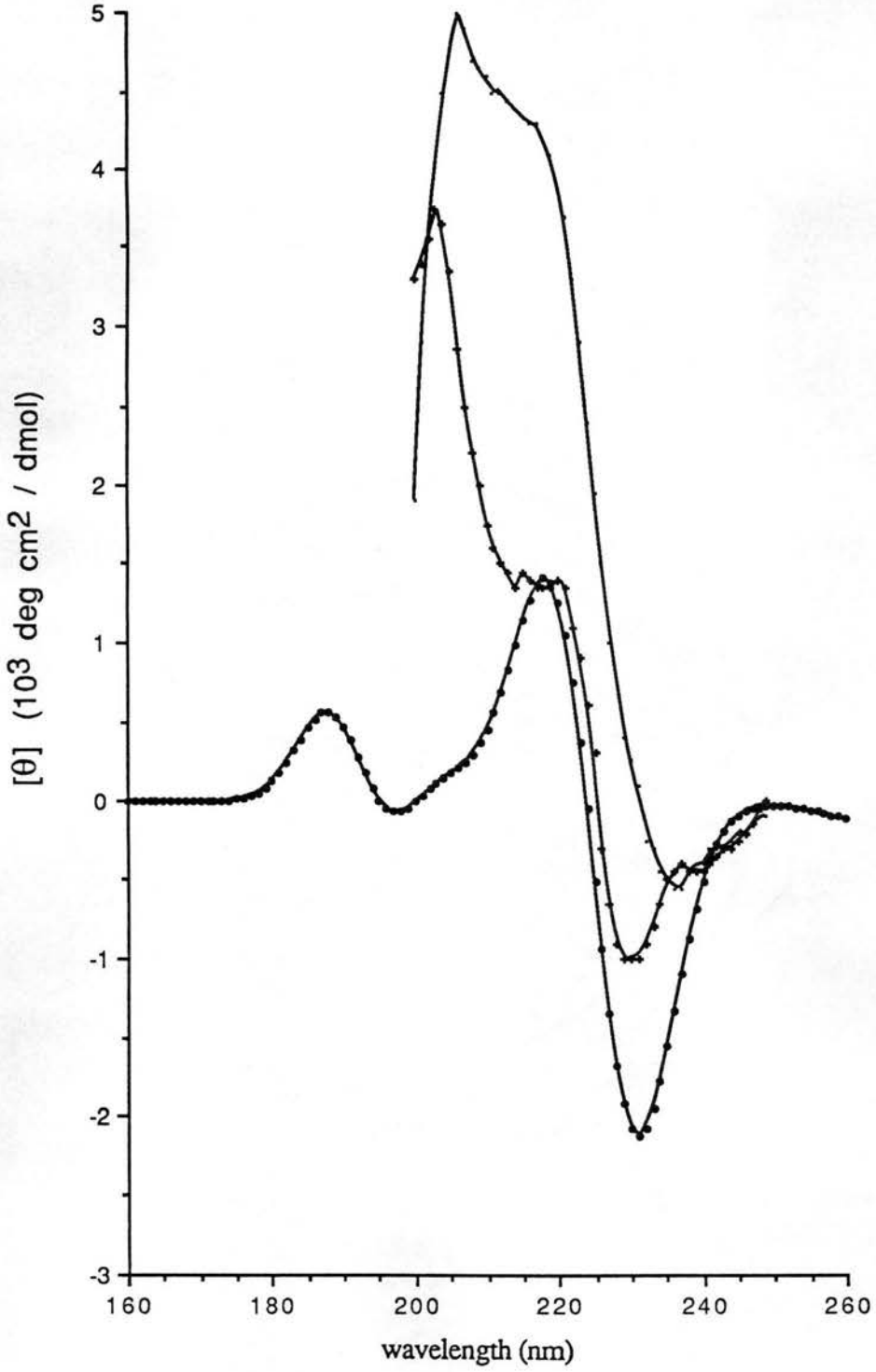


Figure IV-13. Experimental difference CD spectrum between binase and pancreatic ribonuclease A (_). Difference in aromatic contribution obtained by Bolotina & Lugauskas (1985) method (+) between binase and pancreatic ribonuclease A. Binase calculated Trp CD (•).

nm - 240 nm. Accounting for coupling between Tyr and Trp is likely to lead to better agreement of calculated aromatic spectrum and the one deduced from the experimental CD around 230 nm. The band around 200 nm in the experimentally deduced "aromatic spectra" (Bolotina & Lugauskas, 1985) could in fact arise from peptide contributions, such as twisting of β-sheet or particular β-turns, which are not accounted for in the reference spectra.

(5) Dynamics of the Secondary Structure Changes in Binase During Semi-Independent Denaturation of Energetic Domains.

To analyze the assumptions on the secondary structure content of the energetic domains in binase, I performed CD measurements of binase in the far uv and estimated the secondary structure content with the Bolotina & Lugauskas (1985) method, accounting for aromatic contributions in the far uv. Estimates for binase secondary structure in solution were compared for native protein at pH 4.8, 15°C; structurally native-like protein at pH 2.4, 15°C; and partially denaturated protein at pH 2.4, 30.5°C (Fig.IV-14, Table IV-5). As discussed above, binase denatures at pH 2.4 as two cooperative domains (Fig.IV-3). The temperatures of denaturation for those cooperative regions are 31.5°C and 39°C (Protasevich et al., 1987), so at 30.5°C almost half of the first cooperative domain would be denaturated and the second domain should be almost intact. We assume here that the energetic domains become independent at acidic pH before denaturation occurs. Thus, denaturation of the first domain should not lead to additional destabilization of the other one. All CD experiments were performed for deionized aqueous solutions of binase in order to minimize possible effects of ionic

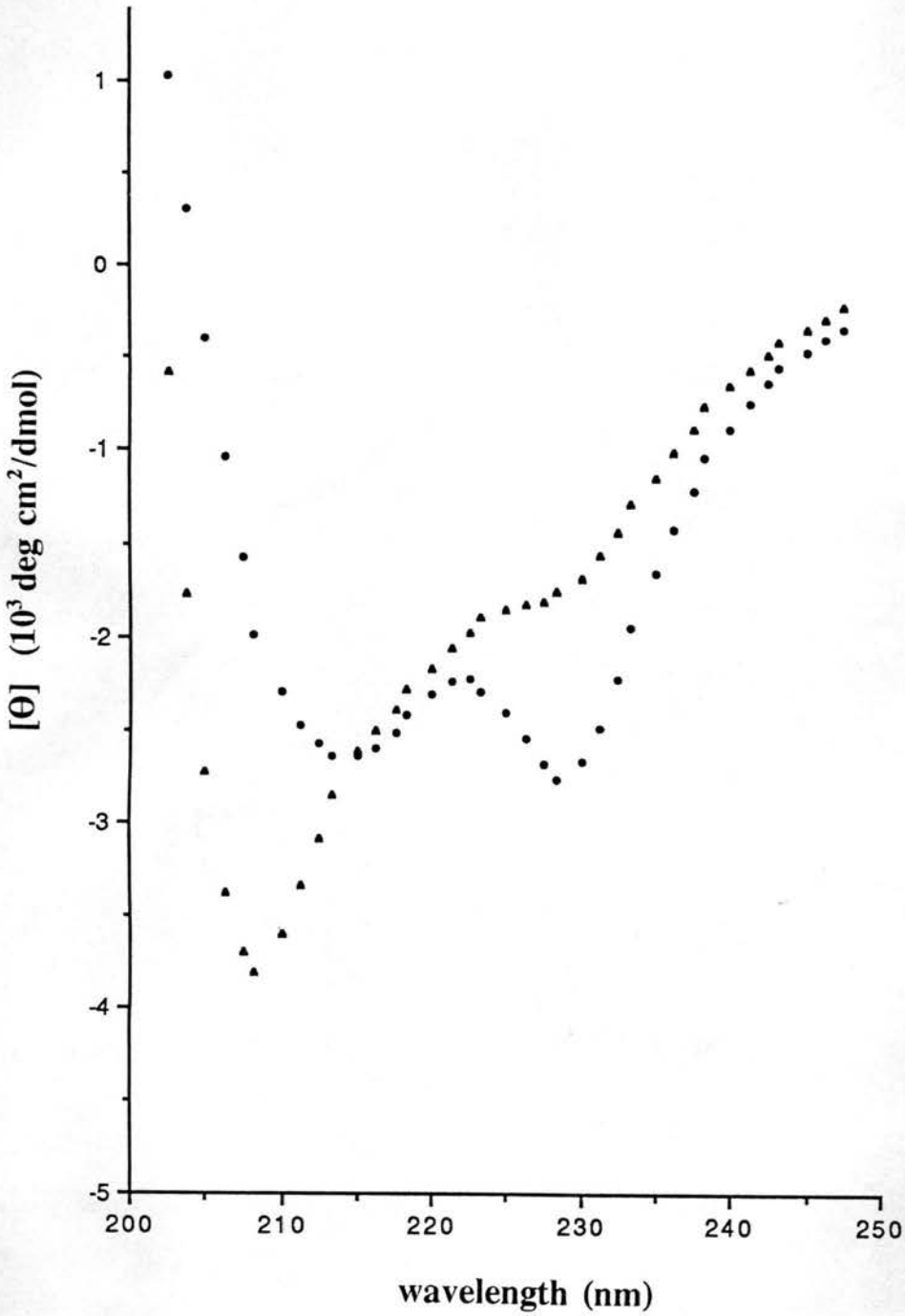


Figure IV-14. Experimental far uv CD of binase at pH 2.4 and 15°C (•); at pH 2.4 and 30.5°C (▲).

strength on ionic interactions. The CD spectra of binase at pH 2.4 at 15^{0}C, which is well below the denaturation temperature at this pH (Protasevich et al., 1987), and at 30.5^{0}C (Fig.IV-14) are shown in Figure IV-14. Secondary structure estimates for the native-like and partially denatured binase are presented in Table IV-5. It is clear that at 30.5^{0}C half of the β-structure present within the native globule has been destroyed, whereas the α-helical content remains unchanged. The estimated change in β-structural content (0.09) is more than two times the error of the method (RMS = 0.04). It should be noted though, that this CD experiment was performed with the same preparation of the protein (except for adjustment of the pH) and only temperature was varied. What we are interested in here is not the absolute error of the method, but the statistical error under the compared conditions. This error did not exceed 0.02 mole fractions for α-helical and β-structural conformations.

The results obtained support the second proposed model, which is that the larger and less stable energetic domain contains the β-structural C-terminal region of binase (*ca.* 47-109), and the separation of energetic domains occurs between the N-terminal α-helical (*ca.* 1-46) and C-terminal β-structural (*ca.* 47-109) parts of binase. Disruption of the ionic contacts Asp7, Asp11 - Arg109 is probably the crucial point in separating the energetic domains. Disruption of the weaker Lys26 - Asp 53, Glu72 contact on the other side of the globule might also contribute in regulating interactions between the energetic domains, as well as the repulsion between Lys26 and Arg82 (Fig.IV-10). The lower thermal stability of the β-structural domain might be also attributed to electrostatic interactions. When Asp74 is protonated in the course of thermal denaturation, repulsion

between uncompensated charges on Arg82, Arg86 and His101 should disrupt the rest of the structure. Protonation of Asp74 may stabilize the denatured state, which could be the thermodynamic explanation for the lower thermal stability of the C-terminal part of binase.

(6) Conclusion.

Combining CD, x-ray analysis and calorimetric data, we were able to identify the parts of a small protein, ribonuclease from *Bacillus intermedius* 7P (binase), which separate into energetic domains at acidic pH, while combined in a single cooperative system in the pH interval where binase is active. The energetic domains in binase consist of the N-terminal α-helical region (*ca.* 1-39) and the C-terminal region containing antiparallel β-sheet (*ca.* 40-109).

Concentrations of acidic and basic side chains occur throughout the protein, mainly on the surface, in the area between α-helical and β-structural energetic domains, and within the β-structural domain close to or in the active site. Most of the acidic and basic side chains are concentrated in clusters of three or more. The latter feature is so pronounced that it should be explored as a general characteristic of the arrangement of charged residues in the tertiary structure of globular proteins.

Ionic interactions were identified that might contribute to the regulation of cooperativity of binase. These are two ionic triads, Asp7, Asp11 - Arg109 and Lys26 - Asp 53, Glu72. Neutralization of ionic contacts is probably enhanced by the consequent repulsion between uncompensated positive charges, mainly Lys26 and Arg82. Both events combined lead to the separation of energetic domains.

It was found that the β-structural energetic domain is of lower thermal stability at acidic pH than the α-helical, whereas at neutral pH the globule is cooperative (Fig. IV-3), thus denaturation of both domains occurs at the same T_d. The lower stability might also be caused by electrostatic interactions within the domain. Protonation of Asp74 at acidic pH in the course of thermal denaturation would enhance the repulsion between the uncompensated charges on Arg82, Arg86 and His101, and thus contribute to the disruption of the β-sheet.

It appears that in a protein molecule consisting of a single structural domain, specific sites may contain less extensive van der Waals contacts and hydrogen bonds. Over a certain pH range, cooperativity at these sites is compensated by favorable interactions between charged residues. Neutralization of acidic side chains upon decrease in pH would lead to the disruption of ionic contacts due to protonation of Glu or Asp and to increased repulsion between like charges on Lys or Arg on the surface or partially buried in the globule in that region. Weakening or disruption of ionic interactions in those sites, where they play a critical role in the formation of the tertiary structure, may lead to the decrease in cooperativity at that site, followed by the formation of separate cooperative domains in the adjacent parts of the globule.

Separation of energetic domains, observed microcalorimetrically in the pH range from 3.1 to 2.4 (Fig.IV-3), is not reflected in any detectable changes in secondary or tertiary structure at temperatures below the thermal denaturation interval. Apparently, disruption of ionic contacts and increased repulsion between uncompensated charges at

acidic pH is not sufficient to overcome the energy of hydrogen bonds and van der Waals contacts in the particular region and thereby induce structural changes.

d. Parallels between Binase and Barnase.

(1) Introduction. Structural and Functional Similarities.

Barnase, a ribonuclease from *Bacillus amyloliquefaciens*, has a sequence very similar to that of binase (Hill et al., 1983). The homology between the two reaches 85 percent if the frame shift due to insertion of Q2 in barnase is taken into account. I will use the numbering for binase in the following.

Most of the mutations between the two sequences are conservative (Fig.IV-15). One of the mutations from barnase to binase that could be structurally significant is the mutation from histidine to lysine, $H17_{ba} \rightarrow K17_{bi}$, which would stabilize the charge-dipole interaction between the charge on residue 17 and the α-helix 6-16 dipole in a wider range of pH in binase, compared to barnase. The H17 pK_a was reported to increase from 6.5 in unfolded barnase to 7.1 in the native protein, which has been shown by H17 titration and H17K mutation to be due to the interaction with the helix dipole (Šali et al.,1988).

Mutations that might change the charge-charge interactions from barnase to binase are T15R, N21D, D22N, E28Q and K65A. Mutations contributing mainly to the change in hydrophobic interactions are T5I, Q14I, I53V, G64S, T78V, S84A, I87L, L88V, Q103A. Despite the similarities in the sequence (Fig.IV-15), the surface properties of binase and barnase are probably somewhat different. During purification, it was found that barnase could not be eluted from the ion-exchange cellulose column with aqueous solution of pH above 4.3, unless a buffer is used (N.K.Chepurnova, Institute of

						5					10
Bi	A	-	V	I	N	I	F	D	G	V	A
Ba	A	Q	V	I	N	T	F	D	G	V	A
						15					20
Bi		D	Y	L	**I**	**R**	Y	**K**	**R**	L	P
Ba		D	Y	L	**Q**	**T**	Y	**H**	**K**	L	P
						25					30
Bi		**D**	**N**	Y	I	T	K	S	**Q**	A	**S**
Ba		**N**	**D**	Y	I	T	K	S	**E**	A	**Q**
						35					40
Bi		A	L	G	W	V	A	S	K	G	N
Ba		A	L	G	W	V	A	S	K	G	N
						45					50
Bi		L	A	**E**	V	A	P	G	K	S	I
Ba		L	A	**D**	V	A	P	G	K	S	I
						55					60
Bi		**G**	**G**	**D**	**V**	F	S	N	R	E	G
Ba		**G**	**G**	**D**	**I**	F	S	N	R	E	G
						65					70
Bi		R	L	P	**S**	**A**	S	G	R	T	W
Ba		K	L	P	**G**	**K**	S	G	R	T	W
						75					80
Bi		R	E	A	D	I	N	Y	**V**	S	G
Ba		R	E	A	D	I	N	Y	**T**	S	G
						85					90
Bi		F	R	N	**A**	D	R	**L**	**V**	Y	S
Ba		F	R	N	**S**	D	R	**I**	**L**	Y	S
						95					100
Bi		S	D	W	L	I	Y	K	T	T	D
Ba		S	D	W	L	I	Y	K	T	T	D
						105					
Bi		H	Y	**A**	T	F	T	**R**	I	R	
Ba		H	Y	**Q**	T	F	T	**K**	I	R	

Figure IV-15. Alignment of amino acid sequences of binase (Bi) and barnase (Ba). Differences are shown in bold. Numbering for barnase is kept as for binase. Insertion of Q2 in barnase is shown.

Molecular Biology, Moscow, personal communication). Binase, on the other hand, can be easily eluted with water at neutral pH.

The active site residues Glu72, Arg86 and His101 are conserved (Hartley, 1989) and so are Lys26 and Arg82, which participate in substrate binding (Sanishvili, 1989; Pavlovsky et al., 1989). Nevertheless the two ribonucleases have slightly different substrate specificity (Hartley, 1989): barnase is nonspecific and binase prefers G > A > T ~ U ~ I (Karpeisky et al., 1981; Afanasenko et al., 1979).

The crystal structures of binase (Pavlovsky et al., 1989) and barnase (Maugen et al., 1982) obtained from x-ray diffraction are very similar. The least-squares fit for the C_α carbons from residues 3 to 110 in barnase and from 2 to 109 in binase gives an average deviation of 0.513 Å (Hartley, 1989). The most structurally homologous region is the C-terminal β-sheet.

Barnase has been intensively studied by Alan Fersht's group in Cambridge (UK), which has applied site-directed mutagenesis to study general questions of globular stability and folding, some of which will be discussed in the following section. I will analyze the applicability of the conclusions obtained on binase (Chapter IV.2.c) to barnase and *vice versa*.

(2) Ionic Interactions.

Comparing the ionic interactions in binase and barnase (Table IV-6), I considered two sets of barnase coordinates, obtained by Chris Hill from a dimeric crystal and by Yves Maugen from a hexameric crystal. Both sets, refined to 2.0 Å, were kindly provided by Robert Hartley upon agreement with the authors. The measurements based

Table IV-6. Interactions of the charged residues in binase and barnase.

	Binase				Barnase[1]			
CLUSTER	Acceptor	Donor	Distance, Å	Contact	Acceptor	Donor	Distance, Å	Contact
I	OD2 Asp7	NH1 Arg109	3.4	ionic	OD2 Asp7	NH1 Arg109	2.7	H-bond
	OD1 Asp11	NH1 Arg109	2.5	H-bond	OD1 Asp11	NH1 Arg109	3.3	ionic
	OD2 Asp11	NH1 Arg15	6.1	ionic	OD2 Asp11	OG1 Thr15	6.9	none
	O Ile108	NH2 Arg109	3.0	H-bond	O Ile108	NH2 Arg109	2.5	H-bond
II	OD2 Asp21	NH1 Arg18	3.2	ionic	OD1 Asn21	NH1Arg18[2]	8.2	none
	O Asp21	NZ Lys48	3.2	H-bond	O Asn21	NZ Lys48	4.4	none
III	OD1Asp53	NH1Arg71	6.5	ionic	OD1 Asp53	NH1 Arg71	6.3	ionic
	OD2 Asp53	N Lys26	2.7	H-bond	OD2 Asp53	N Lys26	2.7	H-bond
	OD2 Asp53	NZ Lys26	5.3	ionic	OD2 Asp53	NZ Lys26	4.8	ionic
	OE1 Glu72	NZ Lys26	4.5	ionic	OE1 Glu72	NZ Lys26	4.9	ionic
	OE2 Glu72	OH Tyr102	3.0	H-bond	OE2 Glu72	OH Tyr102	2.5	H-bond
	OD1 Asp74	NH2 Arg82	2.7	H-bond	OD1 Asp74	NH2 Arg82	2.8	H-bond
	OD2 Asp74	NH1 Arg82	5.0	ionic	OD2 Asp74	NH1 Arg82	5.0	ionic
	OD2 Asp85	ND2 Asn76	2.6	H-bond	OD2 Asp85	ND2 Asn76	2.9	H-bond
	OD2 Asp74	NH1 Arg86	4.3	ionic	OD2 Asp74	NH1 Arg86	4.0	ionic
	OD2 Asp74	NH2 Arg86	4.6	ionic	OD2 Asp74	NH2 Arg86	4.5	ionic
[3]	NZ Lys26	NH2 Arg82	4.9	ionic	NZ Lys26	NH2 Arg82	6.0	ionic

Table IV-6 (continued).

	Binase				Barnase			
CLUSTER	Acceptor	Donor	Distance, Å	Contact	Acceptor	Donor	Distance, Å	Contact
	OE1 Glu59	NH1 Arg58	5.4	ionic	OE1 Glu59	NH1 Arg58	9.4	none
IV	OE1 Glu59	NH1 Arg61	3.9	ionic	OE1 Glu59	NH1 Arg61	3.6	ionic
	OE2 Glu59	N Asp59	2.7	H-bond	OE2 Glu59	N Glu59	4.6	none
	OD1 Asp92	NH1 Arg68	2.3	ionic	OD1 Asp92	NH1 Arg68	3.3	ionic
V	OD2 Asp92	OG Ser90	2.5	H-bond	OD2 Asp92	OG Ser90	4.7	none
	OD2 Asp92	NH2 Arg68	4.9	weak ionic	OD2 Asp92	NH2 Arg68	5.1	weak ionic
	OD2 Asp92	N Asp92	2.5	H-bond	OD2 Asp92	N Asp92	4.3	none

1 Numbering for barnase residues is consistent with that for binase. See Figure IV-15.
2 Like charges.
3 Atomic coordinate were not available in the Chris Hill set. Included later from 1rnb PDB file (Baudet & Janin, 1991).

on atomic coordinates from Chris Hill were included in Table IV-6, unless unavailable, which was the case for NH1 of Arg18. The distances from the guanidinum group of Arg18 were first considered from the Yves Maugen coordinates and replaced later by the data of Baudet & Janin (1991) submitted to the Brookhaven Protein Data Bank (1rnb). The main ionic contacts are conserved in binase and barnase, as well as their environment, although some differences were found (Table IV-6).

The two ionic triads, Asp7, Asp11 - Arg109 and Lys26 - Asp53,Glu72, which might play an important role in regulation of binase cooperativity, are present in barnase as well. The ionic contact Asp7, Asp11 - Arg109 was reported to participate in the tertiary structure folding in barnase by facilitating packing of the N-terminal α-helix on the β-structural core (Šali et al., 1988).

Horovitz et al. (1990) studied the role of cooperativity of the ionic triad Asp7, Asp11- Arg109 in stabilization of the barnase globule. All three residues have been mutated to alanine to give single, double, and triple mutants. It has been found that in the intact triad, the apparent contribution to the stabilization energy of the protein of the salt bridge between Asp11 and Arg109 is -1.25 kcal/mol, whereas that of the salt bridge Asp7 - Arg109 is -0.98 kcal/mol. The two salt bridges are energetically coupled; the energy of each is reduced by 0.77 kcal/mol when the other is absent. It was found that both the double mutant D11A, R109A and the triple mutant are a little more stable than the wild type (1-5 percent increase in ΔG). This correlates with my observation on binase that its structure at low temperatures at neutral pH (when ion contacts are present) and

at pH 2.4 (when the carboxyl side chains of Asp and Glu are likely to be protonated) is unchanged.

The above results obtained for barnase indicate that multiple salt bridging is energetically more favorable than the sum of single bridges. This idea is completely supported by the tendency of charged residues to group in clusters as found on binase (as in previous section) and barnase (Table IV-6).

It has been shown the salt triad Asp7, Asp11 - Arg109 in barnase still retains some cooperativity in the transition states for unfolding and folding intermediates (Horovitz & Fersht, 1992). The excess energy due to cooperativity (over the sum of pairwise interactions) for this triad was found to be 0.77 kcal mol^{-1} in the native protein and 0.60 kcal mol^{-1} in both intermediates. These results do not contradict my conclusions on the role of the triad in binase since they were performed under neutral conditions, at pH 6.3, at which both binase (Fig.IV-3) and barnase (Hartley, 1989; Ira Protasevich, personal communication) are assumed to have intact salt bridges and denature cooperatively. In fact, retention of cooperativity in folding intermediates supports the idea that the ionic triad contributes to the cooperativity of the globule at neutral pH.

There are a number of examples of proteins, in which the transition from one allosteric state to the other is accompanied by breaking of salt bridges, with the missing contacts being replaced by water. The T (tense) state of haemoglobin, for example, is constrained in part by salt bridges that link the C-terminal Arg141 to both Asp126 and Lys127, and the His146 to Asp94 and Lys40 (Perutz, 1989). These salt bridges are disrupted following oxygen binding, thus inducing the conversion from the T state to the

R (relaxed) state. Another example is phosphofructokinase, in which in the R state the dimers are linked by salt bridges from the side chains of Arg162 and Arg243 to the phosphate group of fructose-6-phosphate (Schirmer & Evans, 1990). It was suggested (Horovitz et al., 1990) that the cooperativity of the surface salt bridges in barnase could be part of some mechanism of allosteric regulation as in haemoglobin or phosphofructokinase. This suggestion correlates with my study of the role of ionic contacts in binase, although regulation of the globular cooperativity itself rather than allosteric conformational changes has been observed.

(3) Stability of N-terminal and C-terminal Domains.

Fragments of barnase comprising amino acids 1 to 36 (Sancho et al., 1992) and 37-110 (Sancho & Fersht, 1992) were obtained, which correspond to the N- and C-terminal energetic domains in our definition. Neither of the parts retained catalatic activity. Fluorescence, circular dichroism and ^{1}H NMR data indicate that conformations of both fragments are close to that of the random coil (Sancho & Fersht, 1992). The judgment on the secondary structure content from CD in this work was qualitative, and no deconvolution of the spectra has been performed. It has been shown (Sancho et al., 1992) by two-dimensional ^{1}H NMR that a fraction of the molecular population of the 1-36 fragment does have helical structure, spanning amino acid residues 8 to 18, and this occurs in the absence of tertiary interactions. Also in studies of the intermediate and transition state in barnase folding by protein engineering and NMR spectroscopy (Matouchek et al., 1989; 1990; Bycroft et al., 1990a; Horovitz & Fersht, 1992; Sancho et al., 1991) it has been suggested that the α-helix 6-16 could be partly formed early in

the folding process. This finding correlates with the result obtained under acidic conditions on binase, when cooperative unfolding of the α-helical domain occurs at a higher T_d than the β-structural. Under these conditions the α-helical domain remains folded through the part of temperature interval where the β-structural domain denatures (Fig.IV-3; Grishina, unpublished). Thus the α-helical domain in binase may exist without the stabilizing effect from the β-structural region.

Sancho et al. (1992) concluded that separate formation of α-helical and β-sheet secondary structure in barnase (thus in binase) might precede the tertiary docking of those domains. It also was found that the two fragments of barnase associate, forming a tight complex that displays spectroscopic properties close to those of uncleaved wild-type barnase, and regains catalytic activity with a K_m similar to that of the native enzyme but with a relative k_{cat} of only about 25-30 percent.

(4) Analysis of Barnase Structure in Solution from Far-UV CD.

The experimental far-uv CD spectra of barnase under native conditions at pH 4.8 and 20°C, is shown in Figure IV-16. I should point out that the shapes of binase and barnase far-uv CD spectra are not usual for proteins of the $\alpha+\beta$ class due to their very low intensities and a minimum around 228-231 nm. Despite the general similarity of the far-uv spectra of binase and barnase (Fig.IV-14; Fig.IV-16), it is worth pointing out some differences. The long-wavelength minimum in barnase is at 231 nm, whereas in binase it is at 228 nm. The minimum at 231 nm was associated with the presence of Trp93 (Vuilleumier et al., 1993). The two minima in binase are similar in intensity

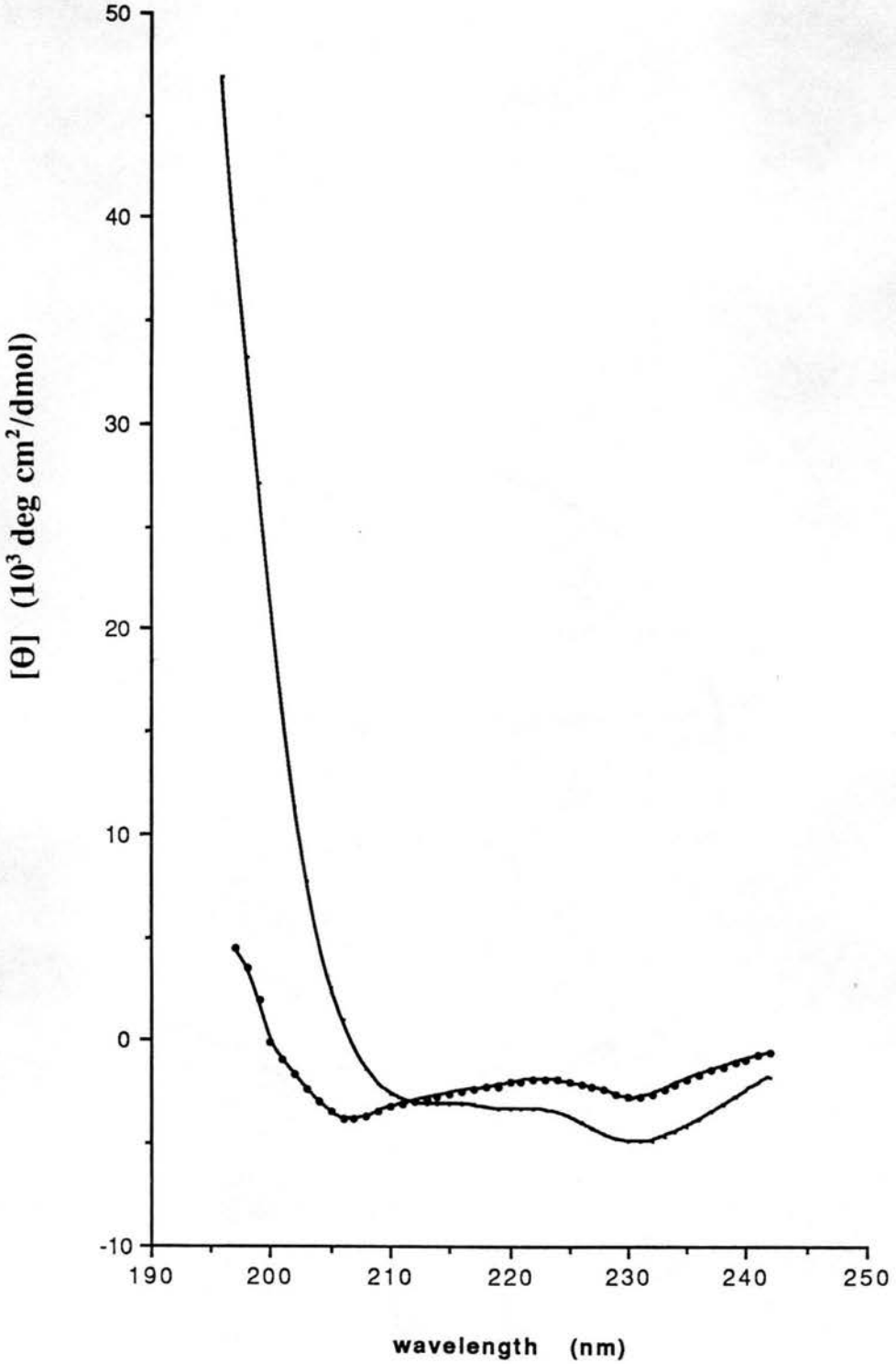


Figure IV-16. Far-uv CD of barnase at native conditions, at pH 4.8 and 20°C (◡), and at the midpoint of thermal transition at acidic conditions, at pH 2.0 and 27°C (•).

(Fig.IV-14), whereas in barnase the short-wavelength minimum is less intense and almost flat (Fig.IV-16) which correlates with observations of Vuilleumier et al. (1993).

The experimental far-uv CD spectra of barnase at the midpoint of thermal denaturation at acidic conditions, at pH 2.0 and 27°C, is shown in Figure IV-16. Barnase is more stable under acidic conditions than binase. The T_d of barnase denaturation at pH 2.0 (27°C) is close to that of binase at pH 2.4, where T_d is 31.5°C (Fig.IV-3) for the first peak of denaturation. The relation between experimental CD under native and partially denatured conditions are similar for barnase (Fig.IV-16) and binase (Fig.IV-14). In the course of denaturation, the long-wavelength minimum is decreased and the short-wavelength band is intensified. Estimation of the changes in secondary structure in barnase during denaturation by the Bolotina & Lugauskas (1985) method gave essentially the same results as for binase (Table IV-5). At the midpoint of the denaturation process at pH 2.0 and 27°C, half of the β-structure had melted (8 percent left from 16 percent estimated from the native spectrum) and all the α-helical content was preserved (16 percent as estimated at native conditions). The secondary structural content of native barnase estimated from CD is a little lower than the assignment from the x-ray data by the rigid method (Chapter III.2.e.(2)) which gave 20 percent in the α-helical conformation and 22 percent in β-structure, as for binase.

Overall, despite extensive sequence and structural similarities, binase and barnase differ both in thermal stability and in intensity of aromatic CD. The dynamics of secondary structure folding seem to be similar: the β-structure domain is the first to undergo denaturation, while the α-helical part is still folded.

(5) Spectroscopic Properties of Aromatic Residues in Barnase.

I obtained the near and far-uv CD spectra for barnase, the shape of which appeared very similar to those of binase. The band at 278 nm in the barnase spectrum has 35 % higher intensity than in binase, and is close to the spectrum reported by Vuilleumier et al. (1993), the difference constituting about 6 % of the amplitude.

All three Trp are conserved in binase and barnase and, judging from the similarity of x-ray structure, so are most features in their local environment. In both proteins Trp34 lies near the C-terminal end of the second α-helix and relatively far (22 - 25 Å) from the other two Trp. Trp70 is in a hydrophobic region at the beginning of the second strand of the β-sheet. Trp93 is at the beginning of the fourth strand of the β-sheet and only 10 Å from Trp70. The fractions of solvent-exposed surface of the side chains of the three Trp are 0.075 for Trp34, 0.029 for Trp70 and 0.219 for Trp93 as calculated for barnase (Loewenthal et al., 1991). Thus both Trp34 and Trp70 are buried residues and Trp93 is a more exposed residue. The same conclusion is true for the native structure of binase at neutral pH, but the area around Trp34 was reported more susceptible to perturbation by pH or denaturants than the β-sheet area around Trp70 (Khandanyan & Dudkin, 1979).

Fluorescence studies of barnase (Loewenthal et al., 1991) led to several conclusions on the individual contributions of the three Trp residues. The spectrum is dominated by the contribution of Trp34. Trp 93 is heavily quenched by the neighboring imidazole group of His17 as determined by site-directed mutagenesis and pH dependence studies (Loewenthal et al., 1991), and also from a multifrequency phase fluorometric study, in which fluorescence lifetimes were determined (Willaert et al., 1992). Analysis

of the x-ray coordinates detected numerous close contacts between Trp93 and His17. Loewenthal et al. (1991) even suggested use of the quenching effect as a marker in protein folding. Residue 17 in binase is lysine, so different fluorescent properties could be expected, which has been confirmed (N. Kuznetsova, Institute of Molecular Biology, Moscow, personal communication). However, it was found by site-directed mutagenesis (Vuilleumier et al., 1993) that this close charge-aromatic interaction between Trp93 and His17 contributes little to the CD spectrum of barnase.

Trp70 and Trp93 were found to behave as an energy-transfer couple (Willaert et al., 1992) with energy transfer in both directions. At pH 8.9, 71 % efficiency was found for transfer from Trp70 to Trp93 and 36 % for the reverse transfer. At pH 5.8 the transfer efficiency was 86 % for forward transfer and only 4 % for reverse transfer.

Vuilleumier et al. (1993) studied a series of barnase mutants in which every single aromatic residue has been mutated, and concluded that the major features in the near-uv spectra of barnase arise from the three Trp residues Trp34, Trp70 and Trp93. Tyrosine contributions were found to be less prominent, with Tyr77 and Tyr96 contributing the most to the CD spectrum, which coincides with my conclusions on binase.

In the far uv, the difference CD spectra between the mutant and wild-type proteins lead to the conclusion that Trp94 contribute *ca.* -180,000 deg cm^2/dmol to the 231 nm CD band (Vuilleumier et al., 1993). This value is close to my estimates for the Trp CD in binase (Fig.IV-8). In fact, I calculated that the total Trp contribution to the far-uv CD in binase at 231 nm is -232,000 deg cm^2/dmol. A CD of -66,000 deg cm^2/dmol was predicted for the W34W70 pair. The contribution of the W34W93 pair

was calculated as -22,000 deg cm^2/dmol and -146,000 deg cm^2/dmol was obtained for the W70W93 pair. Thus elimination of Trp70 would lead to a change in CD at 231 nm of *ca.* 210,000 deg cm^2 /dmol which deviates by only 30,000 deg cm^2/dmol from the experimentally estimated contribution of Trp70 in barnase.

3. Application of CD and Microcalorimetry to the Study of Multidomain Structure in Porcine Pepsin and Pepsinogen.

a. Domain Structure of Pepsin under Various Conditions.

(1) Introduction.

Pepsin is a proteolytic enzyme that belongs to the class of aspartic proteinases. Pepsin functions under very acidic conditions, at pH ~ 2 (Sodek & Hofmann, 1970). Pepsin is characterized by an unusually high concentration of acidic side chains: 30 Asp, 13 Glu, and a phosphoryl group covalently attached to Ser68 (Sielecki et al., 1990). The number of basic side chains is only five. Electrophoretic studies have shown that pepsin migrates as a homogeneous protein with a net negative charge in the pH range 1.01 to 4.57 (Tiselius et al., 1938). Thus, the isoelectric point of pepsin seems to be lower than one.

Pepsin is a multidomain protein comprised of two domains, as was found by x-ray diffraction on crystals grown at pH 2.0 (Andreeva et al., 1984; Sielecki et al., 1990). On the other hand, calorimetric experiments performed by Privalov et al. (1981) at neutral pH have indicated four regions in pepsin that denature more or less independently. We (Makarov et al., 1991; Esipova et al., 1993; Makarov et al., 1994) suspected that the pepsin domain structure might depend on pH or other solvent or crystal conditions and

performed a thermodynamic and structural analysis of pepsin under different solvent conditions.

It has been suggested earlier (Chapter IV.2; Protasevich et al., 1985; Grishina et al., 1989) that interaction between cooperative (or energetic) domains might be stabilized by ion pairs. This would explain the dependence of the domain structure upon the pH of the solvent. I analyzed the near and far-uv CD spectra of pepsin by the Bolotina & Lugauskas (1985) deconvolution approach to follow the structural changes in the pH range from 2.0 to 6.5. Simultaneously, the cooperativity of pepsin was studied under the same conditions in close collaboration with the calorimetrist Irina Protasevich (Makarov et al., 1991).

(2) Materials and Methods.

Porcine pepsin (EC 3.4.23.1) was purchased from Sigma (Lot 64F-8080, U.S.A.) and was purified by K. Frank (Makarov et al., 1991). Protein was purified by gel filtration through a Toyopearl HW-55 column (2 x 50 cm) equilibrated with 20 mM sodium phosphate buffer (pH 6.5). Pepsin (100 mg) was dissolved in 20 mM sodium-phosphate buffer (pH 6.5, 1 ml), and the same buffer was used in elution. The rate of elution was 30 ml/h, and fractions were collected at a 6 min interval. Figure IV-17 shows the profile of pepsin gel filtration from the Toyopearl HW-55 column: the main peak, and peaks at higher and lower relative molecular masses. Fractions 12-21 of the main peak were combined to yield 16 ml of the pepsin solution with a concentration of approximately 4 mg/ml.

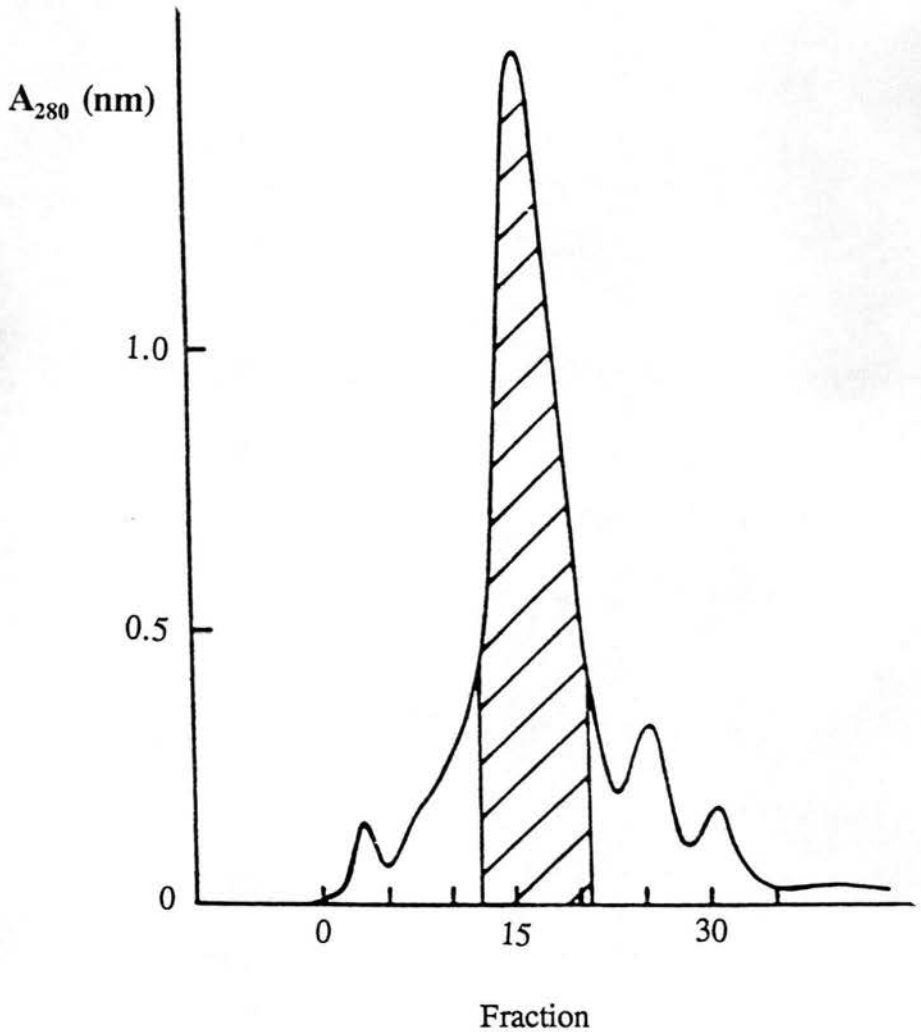


Figure IV-17. Gel filtration profile of porcine pepsin (Sigma) passed through a Toyopearl HW-55 column. The zone of the fractions 12 - 21 taken for the study is shaded. (Adapted from Makarov et al. (1991).)

Pepsin eluted from the column was diluted to obtain a 5 mM solution in sodium phosphate buffer at pH 6.5 and used immediately. The remaining protein solution was dialyzed against 500 ml of 20 % EtOH (v/v) in 5 mM sodium phosphate (pH 6.5) at 4^0C for 24 h. Pepsin solutions at other pHs were prepared by taking a 2-ml aliquot from the above solution at pH 6.5 and dialyzing it against 500 ml of 20 % EtOH in 5 mM sodium phosphate with an appropriate pH at 4° C for 24 h. The pH was controlled with a Radiometer pH-meter (Denmark).

Grade I-S porcine pepsinogen, chromatographically free of pepsin activity, was purchased from Sigma (Lot 24F-8090, U.S.A.) as a lyophilized powder. Protein solutions were prepared by dissolving the lyophilized powder in 5 mM sodium phosphate buffer (or 20 % (v/v) ethanol in buffer) and, prior to calorimetric experiments, they were dialyzed against solvent at 4°C for 24 h. The concentration of pepsinogen was in the range of 0.3-0.4 mg/ml and was determined spectrophotometrically at 278 nm using the molar extinction coefficient of 51,000 $M^{-1}cm^{-1}$ (Fruton, 1971).

Calorimetric measurements were performed with a differential adiabatic scanning microcalorimeter DASM-lM at a heating rate of 1 K/min and pepsin concentration of 1.1-1.6 mg/ml. The thermodynamic parameters of heat denaturation: denaturation temperature, T_d , calorimetric denaturation enthalpy, ΔH_{cal} , and effective or van't Hoff denaturation enthalpy, ΔH_{eff} , were calculated using experimental calorimetric curves as described (Chapter IV.1; Privalov & Khechinashvili, 1974). ΔH_{cal} was determined by measuring the area under the heat absorption peak and using a calibration coefficient. The value of the effectively enthalpy was calculated from the equation $\Delta H_{eff}=$

$4RT^2_d\delta\Delta C_p^{max}/Q_d$, where ΔC_p^{max} is the maximal change of protein heat capacity at T_d , and Q_d is the area under the heat absorption peak in energy units. The ratio $\Delta H_{cal}/\Delta H_{eff}$ was used as a criterion for the extent of heat denaturation cooperativity (Privalov, 1982). The reversibility of heat denaturation was defined as a percentage ratio between the denaturation heats of a protein in the course of its first heating and in its reheating after cooling.

CD spectra at 200 to 250 nm were recorded with a DHR-02 dichrograph (USSR) at a pepsin concentration of 0.3 mg/ml in a 1mm pathlength quartz cell at an instrument sensitivity of $2.5x10^{-5}$ A/mm. CD spectra at 240 to 310 nm were monitored with a Mark III dichrograph (Jobin-Yvon, France) at pepsin concentrations of 0.65 to 0.75 mg/ml in a 0.5-cm quartz cell at an instrument sensitivity of $2x10^{-6}$ A/mm. The results were expressed in units of molar ellipticity based upon the mean amino acid residue, the molecular mass of which has been calculated to be 106 Da for the pepsin molecule. The secondary structure of pepsin was calculated from CD spectra with a Hewlett-Packard 9830A computer (USA) and a BESM-6 computer (USSR) as described (Chapter III.2.e).

(3) Thermal Denaturation of Pepsin at Various pH.

Pepsin represents a special challenge in a thermodynamic study since its denaturation is only partially reversible, as can be seen from the recordings of the first and second heatings in Figure IV-18. However, pepsin denaturation still satisfies the conditions under which equilibrium thermodynamics can be applied, namely a linear dependence of the C_P of the denatured protein on temperature, and invariance of the ΔH_{cal} on the concentration of the sample and the heating rate (Edge et al., 1985; Manly et al.,

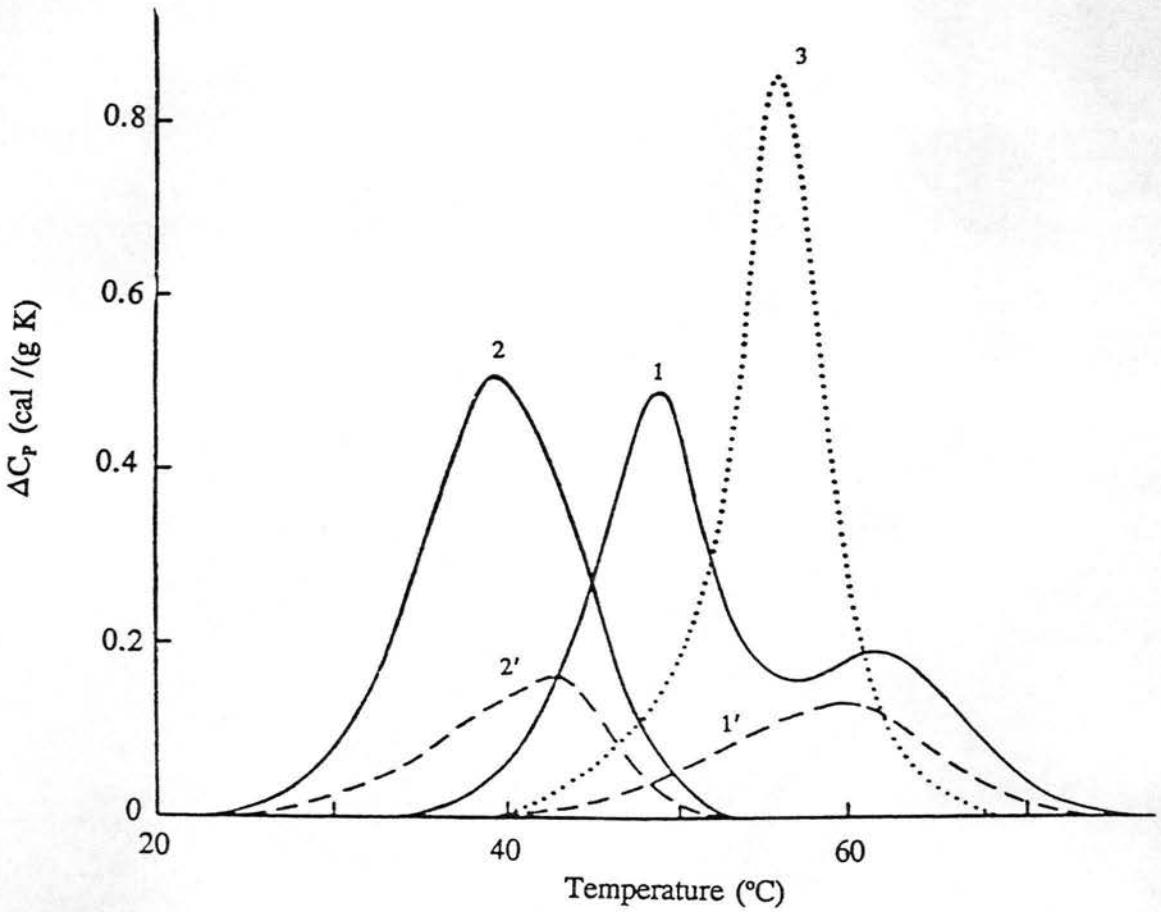


Figure IV-18. Pepsin in 5mM sodium phosphate. The excess heat capacity dependence on temperature at pH 6.5 during first (1) and second (1′) heating; at pH 6.5 in 20 % EtOH during first (2) and second (2′) heating; at pH 2.0 in 20 % EtOH (3). (Adapted from Makarov et al. (1991).)

1985; Hu & Sturtevant, 1987; Sturtevant, 1987; Edge et al., 1985). If equilibrium thermodynamics is to be applied, the protein should not undergo an irreversible aggregation during or after the thermal denaturation. Then, the irreversibility of its denaturation can be considered as a kinetic rather than a thermodynamic property of the system. The aggregation of the denatured form of the protein can be seen on the calorimetric recording as a sharp drop of the heat capacity to negative values due to extensive heat release. On the other hand, if the system is in thermodynamic equilibrium, then results should not depend on concentration or the rate of the process. Ideally, the calorimetric parameters obtained at a given rate of heating should be the same as for an infinitesimally slow, i.e., equilibrium, process. Concentrations of pepsin in solution were varied in the range of 1.1 to 3.9 mg/ml. Measurements were also carried out at the heating rate of 0.25 K/min and at $\sim$ 1 K/h with an absolute adiabatic calorimeter (Makarov et al, 1976; Esipova et al., 1978). In all the cases the value of ΔH_{cal} was the same and the C_P (T) showed the usual slight linear increase with the temperature. Thus the pepsin thermal denaturation profile during the first heating can be treated as a thermodynamic process.

We have studied pepsin thermal denaturation and CD spectra within a wide pH range, including the pH at which it functions, pH 2.0, and the pH at which it assumes a structurally stable inactive form, pH 6.5. Ethanol is a pepsin inhibitor. It prevents autolysis (Northrop, 1946) and therefore allows one to work in the pH region at which pepsin functions. Experiments were performed in a 20 % aqueous ethanol solution. Since the dielectric properties of ethanol differ from water solution, we also intended to clarify

the effect of ethanol as a solvent on the thermodynamic properties of pepsin and, specifically, on the behaviour of the energetic domains.

Figure IV-18 presents the excess heat capacity of pepsin as a function of temperature at pH 6.5 with and without ethanol. The reversibility of denaturation was about 30 % at pH 6.5 (Fig. IV-18, curves 1', 2'), and the reversibility of the high temperature peak at pH 6.5 (Fig. IV-18, curve 1') was 75 %. The thermodynamic parameters for the two peaks of heat absorption at pH 6.5 (Table IV-7) are in good agreement with those reported previously (Privalov et al., 1981). The first peak results from the denaturation of the N-terminal domain in pepsin and the second peak from that of the C-terminal domain (Privalov et al., 1981).

When 20 % EtOH was added to the protein solution, the denaturation curve shifted towards lower temperatures (Fig. IV-18, curve 2) and the peaks came closer to each other. However, we did not observe any noticeable changes in ΔH_{cal}. The ratio of the calorimetric enthalpy to the effective enthalpy is a criteron which allows one to estimate how much the average size of a cooperative units differs from the size of the globule (Privalov, 1982). We obtained the ratio R equal to 3.7 at pH 6.5, which indicates four cooperative units in 20 % EtOH, just as in aqueous solution at that pH (Privalov, 1982). A similar effect of EtOH was observed on ribonuclease A, where addition of 50 % methanol decreased T_d but did not change ΔH_{cal} or the value of the ratio $\Delta H_{cal}/\Delta H_{eff}$ (Brands et al., 1989). We obtained the difference between the heat capacities of native and denatured states of pepsin ($\delta\Delta C_p$) at pH 6.5 in aqeous solution as 0.12 ± 0.01 cal/(K g) which is close to the value 0.13 ± 0.01 cal/(K g) reported by Privalov et al., 1981.

When EtOH was added at pH 6.5, we observed the value of $\Delta\delta C_p$ drop to 0.08 $\pm$0.02 cal/(K g), which we attribute to changes in the environment of hydrophobic groups upon denaturation.

The x-ray diffraction data on pepsin have been obtained from crystals grown at pH 2 in 20 % EtOH (Andreeva et al., 1984; Sielecki et al., 1990). As mentioned above, this pH corresponds to the region of protein function. EtOH exhibits an inhibitory effect. Its addition is necessary to prevent the autolytic activity of pepsin. We performed calorimetric recording of the thermal denaturation of pepsin at pH 2.0 in 20 % EtOH and obtained the $\delta\Delta C_p$ of 0.08 cal/(Kg). The same $\delta\Delta C_p$ was obtained in EtOH at pH 6.5. No changes in $\Delta\delta C_p$ between pH 6.5 and pH 2.0 were found in aqueous solution (Privalov et al., 1981) as well. Therefore, changes in the heat capacity of pepsin upon denaturation do not depend on the pH in solutions with or without EtOH. Comparison of pepsin thermal denaturation at pH 6.5 and 2.0 reveals an increase in the temperature of denaturation and the effective enthalpy. The ΔH_{cal} hardly changes (Fig. IV-18, Fig.IV-19, Table IV-7).

The ΔH_{cal} does depend on T_d in the pH region from 5 to 2. Only one peak of heat absorption has been observed in this region. The ratio R changes from 3 to 2, indicating less than four energetic domains. Fluorescence polarization studies showed that pepsin is in the monomeric form by structural criteria within the pH range 2.0 to 6.5 (Glotov et al., 1976). However, several energetic domains are detected throughout that pH interval: two at pH 2.0, four at 6.5 and from three to two in the pH interval 5.0 - 2.0.

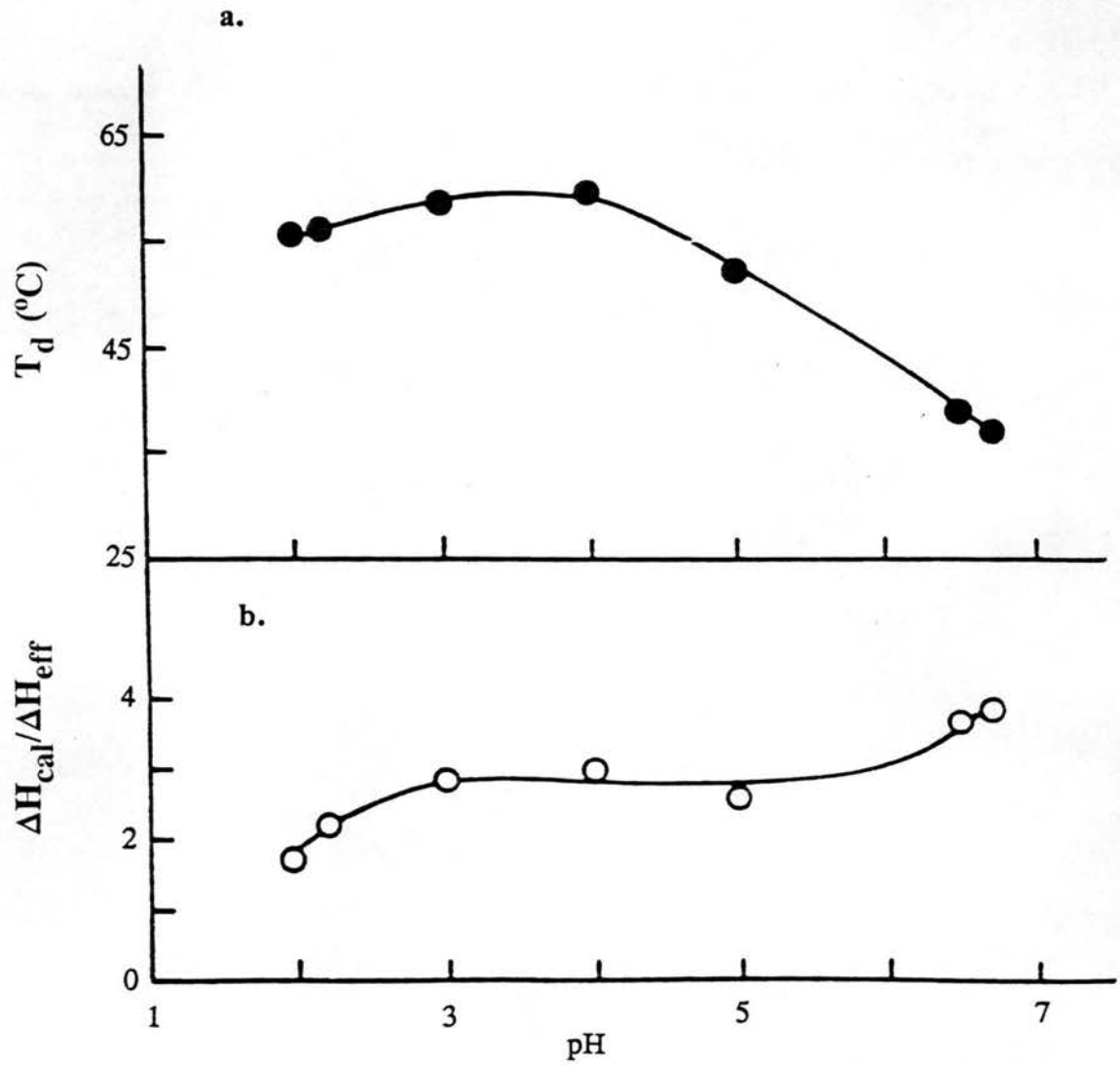


Figure IV-19. The pH dependence of the denaturation temperature of pepsin (a) and the ratio $\Delta H_{cal}/\Delta H_{eff}$ (b). (Adapted from Makarov et al. (1991).)

Table IV-7. Parameters of thermal denaturation of pepsin at different pH.

Conditions	ΔH_{cal} (kcal/mol)*	ΔH_{eff} (kcal/mol)	R	T_d (°C)
H_2O **				
pH 6.5	$\Delta H_{cal}^1 = 135$	$\Delta H_{eff}^1 = 74$	$R^1 = 1.8$	$T_d^1 = 49.0$
	$\Delta H_{cal}^2 = 85$	$\Delta H_{eff}^2 = 50$	$R^2 = 1.7$	$T_d^2 = 62.0$
+20 % ethanol				
pH 6.5	227	62	3.7	38.8
pH 5.0	215	87	2.5	52.0
pH 4.0	233	77	3.0	59.5
pH 3.0	238	82	2.9	58.5
pH 2.2	212	95	2.2	56.0
pH 2.0	214	128	1.7	56.0

* Experimental error does not exceed ± 10 kcal/mol for ΔH_{cal} ; ± 5 kcal/mol for ΔH_{eff}; $\pm$ 5°C for T_d ; ± 0.2 for $R = \Delta H_{cal} / \Delta H_{eff}$.

** Parameters for the two peaks are given.

We found that the T_d is pH-dependent, both with and without EtOH. Pepsin is much more stable at acidic pH than in the neutral region (Fig.IV-18, Fig.IV-19).

Pepsin denaturation is only partially reversible at pH 6.5 (Fig.IV-18) and irreversible in the pH interval 5.0 - 2.0. The cause of irreversibility is unclear. We suggest that it might be associated with kinetic factors such as protonation of some of the carboxylates buried in the globule, which will be discussed later. It is interesting that the autocatalytic activation of the pepsin precursor, pepsinogen, occurs around pH 3.0 - 4.0 (Herriott, 1939), that is in the same pH interval where unfolding of pepsin is irreversible. Both phenomena imply a certain ambiguity, i.e. certain transitional characteristics, in structural and stereochemical properties of pepsin/pepsinogen globule at those pHs, and might be interconnected. Some studies on pepsinogen structure and cooperativity are reported in Chapter IV.2.b.

The increase of ΔH_{eff} as pH decreases indicates that the average size of cooperative regions increases. The ΔH_{cal} hardly changes with the pH decrease, indicating that: 1) the globule stays in its monomeric state, and 2) no significant changes have occurred in the number of hydrogen bonds or hydrophobic contacts. The ratio R is almost twice as low at acidic pH values than in the neutral pH region, and approaches 2 (Fig.IV-19, Table IV-7). Deconvolution results at pH 2 (20 % ethanol) are consistent with two overlapping two-state transitions.

(4) Pepsin CD Spectra at Various pH.

I have analyzed the near- and far-uv CD spectra of pepsin under conditions where different domain structure has been established: at pH 2.0, 3.5, 4.5, 5.5 and 6.5, in

solutions with and without EtOH. Figure IV-20 shows CD spectra for pepsin solutions in 20 % ethanol in the far UV region at different pH. One can see that these spectra are very close to one another within the pH range of 6.5 to 2.1. Ethanol has no effect on the CD spectra at any pH studied (Fig.IV-20).

Pepsin contains 5 Trp, 16 Tyr and 13 Phe distributed throughout the 327-amino acid globule. The near-uv CD spectra of pepsin in 20 % EtOH does not depend on pH (Fig. IV-21) indicating that pH does not affect the environment of the aromatic side chains. A similar conclusion has been reached in fluorescence experiments (Kozlov et al., 1979) for pepsin in aqueous solotion. The near-uv CD changes slightly on going from the aqueous solution to 20 % EtOH (Fig.IV-21). This may be explained by the effect of the polarity of the solvent on the CD of aromatic chromophores or on their local conformation.

The secondary structure of pepsin in solution has not been determined previously. I used the deconvolution approach of Bolotina & Lugauskas (1985), which allows one to account separately for the changes in aromatic CD and in the content of regular secondary structures. The secondary structure of pepsin was found virtually unchanged in the pH region studied. The following fractional content of secondary structures was obtained: 0.16 ± 0.03 in α-helices, 0.34 ± 0.02 in β-sheets, and 0.18 ± 0.02 in β-bends.

The assignment of the secondary structure from x-ray has been performed in collaboration with A. Fedorov, who was involved in the refinement of the pepsin crystal structure (Sielecki et al., 1990). By applying the criteria of the rigid method (Chapter III.2.e.(2); Bolotina et al., 1980) we determined the fractional content of amino acid

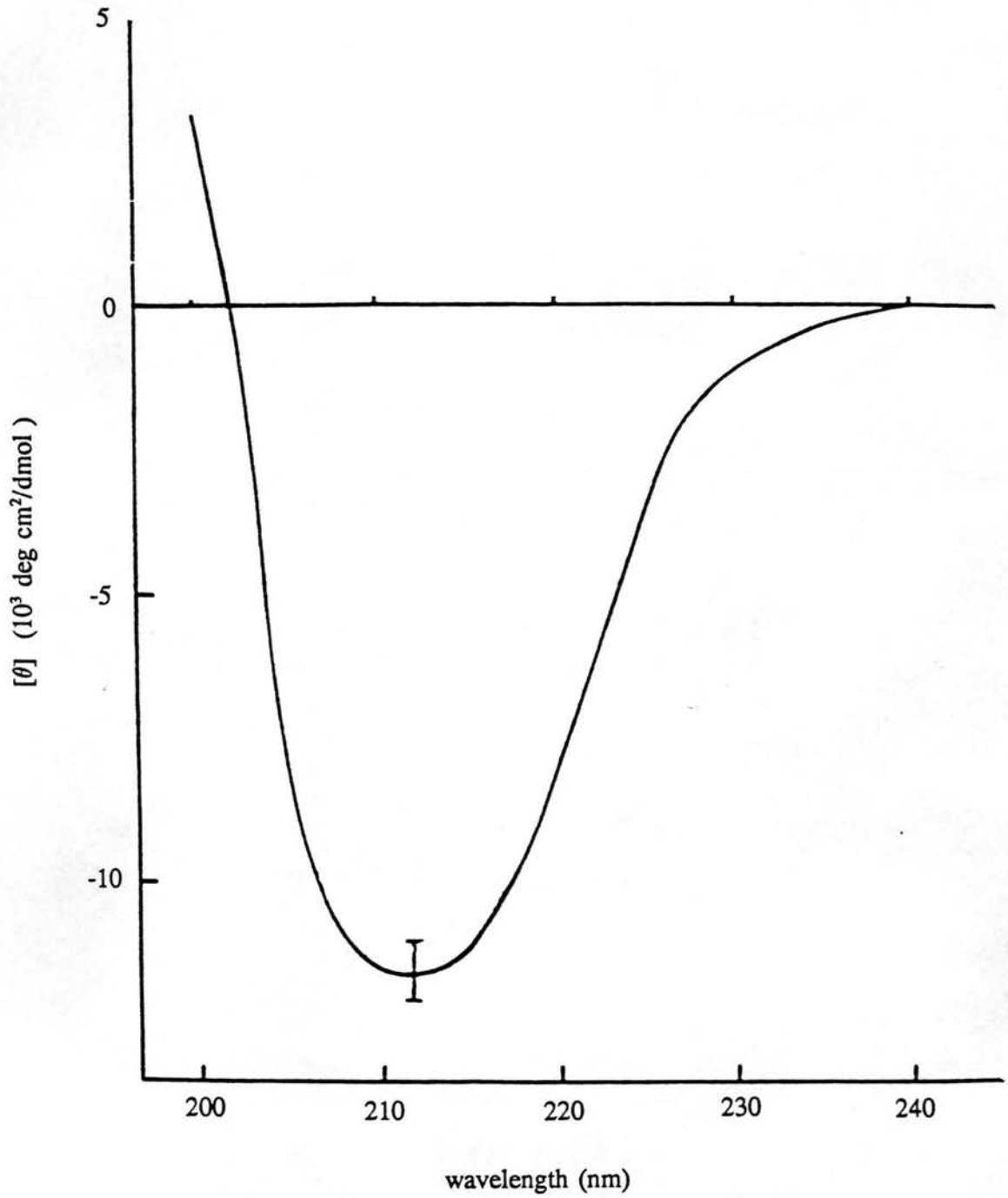


Figure IV-20. Far-uv CD of pepsin in 20 % EtOH at pH 2.1 - 6.5. (Adapted from Makarov et al. (1991).)

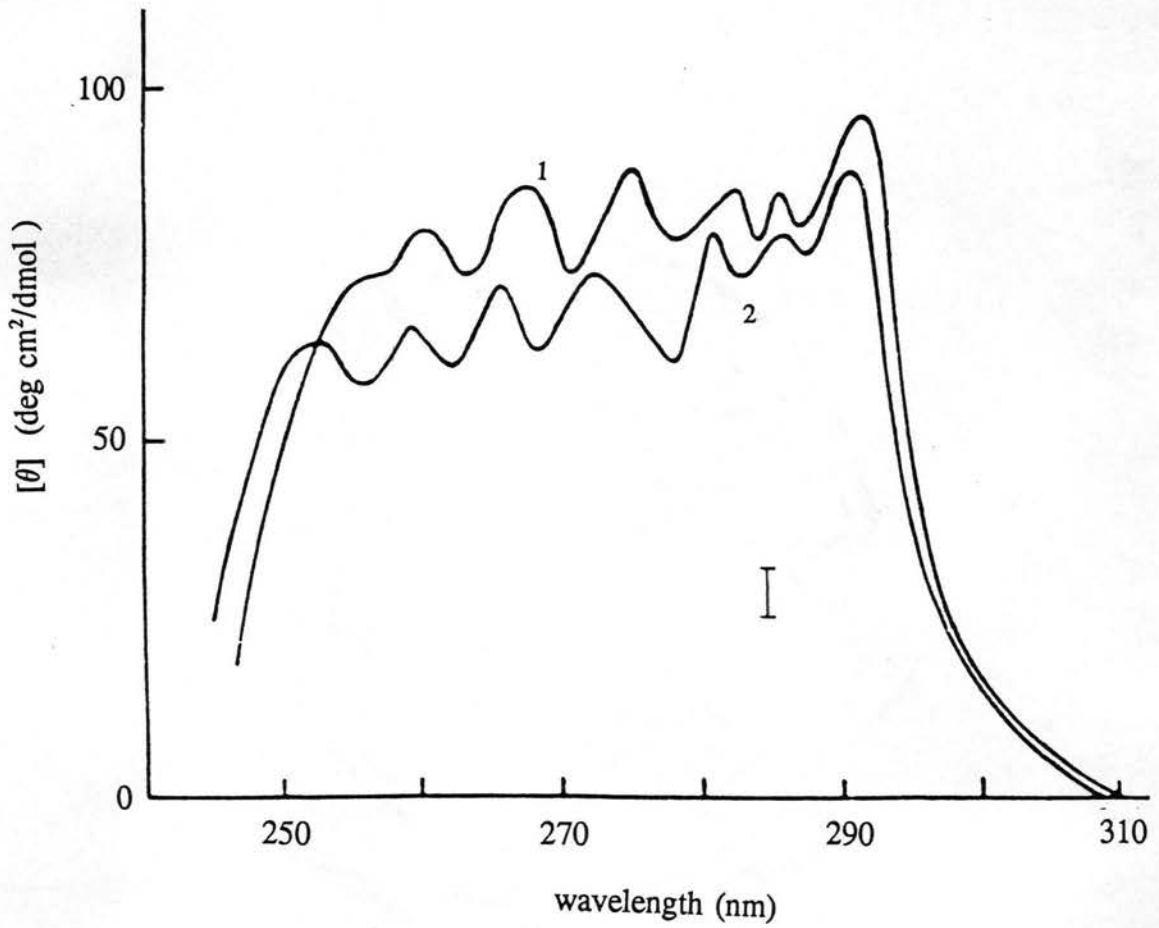


Figure IV-21. Near-uv CD of pepsin in 5mM sodium phosphate at pH 6.5 (1) and in 5mM sodium phosphate and 20 % EtOH at pH 2.0 - 6.5 (2). (Adapted from Makarov et al. (1991).)

residues as 0.12 in α-helices, 0.36 in β-sheets, and 0.14 in β-bends. One can see that I obtained a good correlation between CD and x-ray estimates.

(5) Conclusion.

Therefore, we found that: 1) changes in the number of cooperative domains in pepsin globule do not involve any modifications either in its secondary structure or in the local surroundings of aromatic amino acid residues; 2) cooperative properties of the pepsin globule are pH-dependent. Changes in the net charge of the molecule most likely arise from protonation/deprotonation of Asp and Glu side-chain carboxyl groups, since there are only five positively charged side chains and 44 possible negatively charged positions: 30 Asp, 13 Glu, and a phosphoryl group covalently attached to Ser68 (Sielecki et al., 1990). As the net negative charge decreases (below pH 4), the repulsion between like-charged surface regions of the globule should weaken, which would explain the reduction of the number of energetic domains from four to two.

Our experiments have shown that the thermal stability of pepsin decreases (at a given pH) when its inhibitor ethanol is added, while the number of cooperative regions does not change. The number of cooperative regions in a pepsin molecule changes gradually with its net charge. Just as another pepsin inhibitor, pepstatin (Privalov, 1982), ethanol influences the transition temperature and not its enthalpy. Therefore, in thermodynamic terms, the effect of protein inhibition in case of pepsin is exerted on the entropy of a transition, but not on its enthalpy.

Ethanol might favor charge repulsion due to its low dielectric constant, which would shift pepsin denaturation to lower temperatures. When the pH is raised from 4 to

the alkaline region, the temperature of denaturation abruptly drops (Fig.IV-19b), i.e. becomes strongly pH-dependent. The charge of pepsin changes most rapidly around pH 4.0 due to deprotonation of numerous Asp and Glu. This confirms that the pH dependence of the thermal transitions in pepsin is associated with the change in the net charge of the molecule.

The number of cooperative regions in pepsin changes monotonically within the entire range of the pH interval studied (2.0 - 6.5) and approaches unity around the isoelectric point at pH 1 (Fig.IV-19b). In other words, pepsin may act as an integral cooperative system at a very low pH. The number of energetic domains in pepsin at its optimal pH coincides with the number of structural domains.

The specific effect of ethanol as an inhibitor would imply its specific binding to the protein molecule. However no bound ethanol was identified in the pepsin crystal structure (Sielecki et al., 1990). It has been suggested that the inhibiting action of ethanol may be associated with disruption of ordered solvent in the pepsin active site. We found that ΔH_{cal} of pepsin in water-ethanol solution is the same as in an aqueous solution at the same pH. Hence, any contacts of native pepsin molecules with ethanol molecules either remain unchanged upon denaturation or are compensated.

We have demonstrated for the two proteins with different functions and domain structure, pepsin (Makarov et al., 1991; Esipova et al., 1993) and binase (Protasevich et al., 1987; Grishina et al., 1989; Grishina et al., 1993; Makarov et al., 1994) that in both cases the protein globule contains fewer energetic domains under the conditions optimal for the activity. The number of energetic domains decreases for different reasons

in binase and pepsin: binase becomes an integral cooperative system when ion pairs are switched on (Grishina et al., 1993), whereas the size of the energetic domains in pepsin increases because the electrostatic repulsion between regions in the molecule is weakened (Makarov et al., 1991).

Since various types of non-covalent interactions stabilizing the native protein globule have characteristic ranges, the energetics of a protein globule could be described as a hierarchic function of those interactions. Following this hypothesis, the weakening of a certain group of interactions will lead to the separation of the globule into several cooperative regions. Thus the dimensions of energetic domains within a globule as well as their geometry could be, as we observed, a function of the environmental conditions.

b. Application to the Study of Thermal Denaturation in Pepsinogen.

(1) Introduction.

Pepsinogen is an inactive precursor of the proteolytic enzyme pepsin. Their amino acid sequences differ only by the 44-residue prosegment at the N-terminus of pepsinogen. This prosegment is cleaved during activation of pepsinogen. Activation proceeds as a second order reaction at pH below 5.0 and is autocatalytic below pH 4.0 (Herriott, 1939). The prosegment plays a dual role in the structure of pepsinogen. It partially neutralizes the negative charge of the pepsin moiety by adding eleven basic side chains. It also forms numerous hydrogen bonds and ion pairs with the residues of the active site, thereby inhibiting the enzyme activity.

The addition of ethanol, pepstatin or the intact prosegment all inhibit the active site of pepsin, although the thermodynamic effects of ethanol, pepstatin and prosegment

differ. Two cooperative transitions are detected in pepsinogen at pH 6.5 in water (Mateo & Privalov, 1981), while four transitions are observed in pepsin under those conditions (Privalov et al., 1981). The pepsin inhibitor, pepstatin, changes the number of cooperative regions in the N-terminal domain of pepsin as the transition temperature increases (Privalov et al., 1982). We have shown (Chapter IV.2.a) that ethanol decreases the thermal stability of pepsin, while the number of cooperative regions do not change. I combined scanning microcalorimetry and CD to study the specific effects of the prosegment and ethanol on the stability of the pepsinogen globule.

(2) Materials and Methods.

Grade I-S porcine pepsinogen, chromatographically free of pepsin activity, was purchased from Sigma (Lot 24F-8090, USA.) as a lyophilized powder. Protein solutions were prepared by dissolving the lyophilized powder in 5 mM sodium phosphate buffer (or 20 % (v/v) ethanol in buffer). Prior to calorimetric experiments, the solutions were dialyzed against solvent at 4°C for 24 h. The concentrations of pepsinogen solutions were in the range of 0.3-0.4 mg/ml and were determined spectrophotometrically at 278 nm using the molar extinction coefficient of 51,700 $M^{-1}cm^{-1}$ (Privalov, 1981). Calorimetric and CD measurements were performed as for pepsin (Chapter IV.3.a.(2)) in collaboration with I. Protasevich (Makarov et al., 1991).

(3) Structure of Pepsinogen under Various Solvent Conditions.

Calorimetric measurements of pepsinogen were performed at pH 6.5 in water solution (5 mM sodium phosphate) and ethanol (Fig.IV-22, Table IV-8). EtOH causes T_d to decrease compared with measurements without EtOH (Fig.IV-22). The reversibility

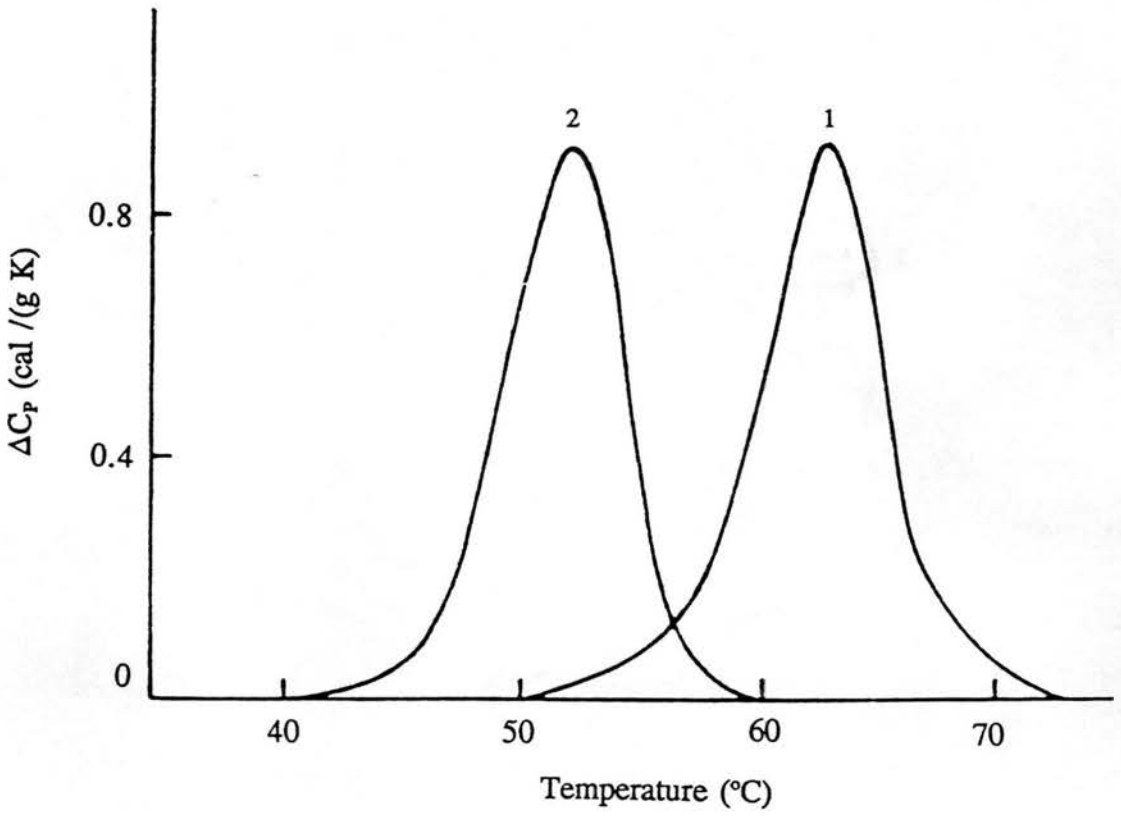


Figure IV-22. Pepsinogen excess heat capacity dependence on temperature at pH 6.5 : in 5mM sodium phosphate (1), and same with 20 % EtOH (2). (Adapted from Makarov et al. (1991).)

Table IV-8. Parameters of thermal denaturation of pepsinogen.

Conditions	ΔH_{cal} (kcal/mol)	ΔH_{eff} (kcal/mol)	R	T_d (°C)
H_2O				
pH 6.5	248	131	1.9	62.8
pH 8.0	169	113	1.5	51.1
+20 % ethanol				
pH 6.5	248	118	2.1	52.1
pH 8.0	169	125	1.4	35.3

Experimental error does not exceed $\pm$10 kcal/mol for ΔH_{cal}; $\pm$5 kcal/mol for ΔH_{eff} ; $\pm$5°C for T_d ; $\pm$0.2 for $R = \Delta H_{cal} / \Delta H_{eff}$.

of denaturation was no less than 82 % and did not change upon ethanol addition. The thermodynamic parameters of pepsinogen thermal denaturation in 20 % EtOH coincided with those reported for the same pH in aqueous solutions (Mateo & Privalov, 1981). The extrapolation of the plot ΔH_{cal} vs. T_d for a pepsinogen solution in ethanol to 110°C yields 12.9 cal/g, which coincides with the value found in the absence of ethanol (Mateo & Privalov, 1981). We found that the action of ethanol on pepsinogen is similar to its effect on pepsin: ethanol makes T_d decrease, but has no effect on ΔH_{cal} and R (Table IV-8). The effect of ethanol on T_d increases with pH. The difference in denaturation temperatures is 10.7°C at pH 6.5 and 15.8°C at pH 8.0. The isoelectric point of pepsinogen was found to be pH 3.8 (Herriott, 1939), so on moving from pH 6.5 to pH 8.0 we move further from the isoelectric point. Thus the effect of ethanol increases with the net charge on the surface.

The pepsinogen far-uv CD does not seem to depend on pH and/or addition of 20% ethanol in the pH interval 6.1- 7.9 and at low temperatures (below the thermal transition). All spectra both in water and in 20 % ethanol lie in a narrow interval within a $\pm$ 5 % relative deviation of the amplitude at the maximum. The average of the secondary structure estimations by the Bolotina & Lugauskas (1985) method from all the six spectra at three pHs (6.1, 6.5 and 7.9) and in two kind of solvents (5mM sodium phosphate, and same with 20 % ethanol) is presented in Table IV-9 with corresponding mean deviations. The assignment of secondary structure from the x-ray coordinates (Sielecki et al., 1991) was performed using the rigid criteria (Chapter III.2.e.(2); Bolotina et al., 1980c) in collaboration with A. Fedorov (Institute of Molecular Biology,

Table IV-9. Estimation of the secondary structure content in pepsinogen.

Method	f_{α}	f_{β}	$f_{\beta\text{-turns}}$	$f_{\text{unord.}}$
x-ray[1]	0.159	0.323	0.157	0.361
CD[2]	0.10	0.65	0.05	0.20
CD[3]	0.13	0.24	0.18	0.45
CD[4]	0.23±0.03	0.29±0.03	0.21±0.02	0.27±0.03

1 - crystallographic data reported by Sielecky et al. (1991) analyzed by the rigid method (see text); 2 - analysis of CD data by the method of Yang et al. (1986); 3 - analysis of CD data by method of convex constraint analysis (Perczel et al., 1991); 4 - analysis of CD data by method of Bolotina & Lugauskas (1985).

Moscow) (Table IV-9). As can be seen the CD deconvolution method that I used (Chapter III.2.e; Bolotina & Lugauskas, 1985) gives estimates closer to x-ray than the method of Yang et al. (1986) does. My estimates are comparable in accuracy to those obtained by the "natural deconvolution" method (Perzel et al., 1990). The α-helical content of native pepsinogen as determined from CD is slightly higher than that estimated from the crystal structure. Some difference in estimations of the secondary structure content by various authors could be attributed to the differences in the samples of pepsinogen, or due to the purification as accidental activation of pepsinogen would lead to autolysis. The discrepancies in the estimates between the CD methods (Table IV-9) might be also due to calibration or bias in the deconvolution processes.

In addition to the 5 Trp, 16 Tyr and 13 Phe in the pepsin part of the molecule, pepsinogen contains one Tyr and two Phe side chains in the prosegment. It is worth mentioning that the authors of the convex constraint analysis method considered it important to account for aromatic contributions in the spectrum of pepsinogen (Perzel et al., 1990), although in contrast to our method, they made an attempt to extract some kind of universal reference spectrum to account for aromatic contributions. In the Bolotina & Lugauskas (1985) approach that I used, the aromatic contribution to the CD of each protein is considered to be unique to that protein and is described as such, although with standard functions (log-normal). Such an individual approach should better reflect the fact that aromatic CD depends on the local environment and relative arrangement of the aromatic side-chains in the globule, and both of these factors are specific for each protein.

I should point out that the pepsinogen CD spectrum in the near uv (not shown) differs considerably from the CD spectrum of pepsin (Fig.IV-21), whereas their spectra in the far uv are similar (Fig.IV-20; Fig.IV-23). This indicates that while the secondary structure does not change considerably upon activation of the pepsinogen, which correlates with the x-ray data (Sielecki 1990; 1991), the factors defining the CD of the aromatic side-chains undergo significant changes, which results in changes in the amplitude ratios and signs of CD bands in the near uv. The main features observed on comparing the near-uv CD spectra of pepsin and pepsinogen are conservation of the sign and only a slight amplitude change for the band at 292 nm (Fig.IV-21), which can be assigned to the L_b transition in Trp, and a drastic change of the large negative band around 270-290 nm in the pepsinogen spectrum, which upon formation of pepsin decreases two-fold in intensity, changes sign and acquires fine structure (Fig.IV-21).

(4) Structure of Partially Denatured Pepsinogen.

The CD spectra of pepsinogen below the thermal transition at pH 6.5 in water, and above the thermal transition (Fig.IV-22) at 80°C in water and at 65°C in ethanol, are shown in Figure IV-23. Secondary structure of pepsinogen estimated under conditions below and above the thermal transition are shown in Table IV-10. The secondary structure content of pepsinogen estimated from the CD spectrum at pH 6.5 and room temperature (24°C), i.e. under the same conditions as used for the growth of the crystals, is in very good agreement with the x-ray results (Sielecki et al., 1991). I also obtained the pepsinogen CD spectrum roughly in the middle of the thermal transition in water solution (Figure IV-22), at pH 6.5 and 65°C, which was similar to the spectra above the

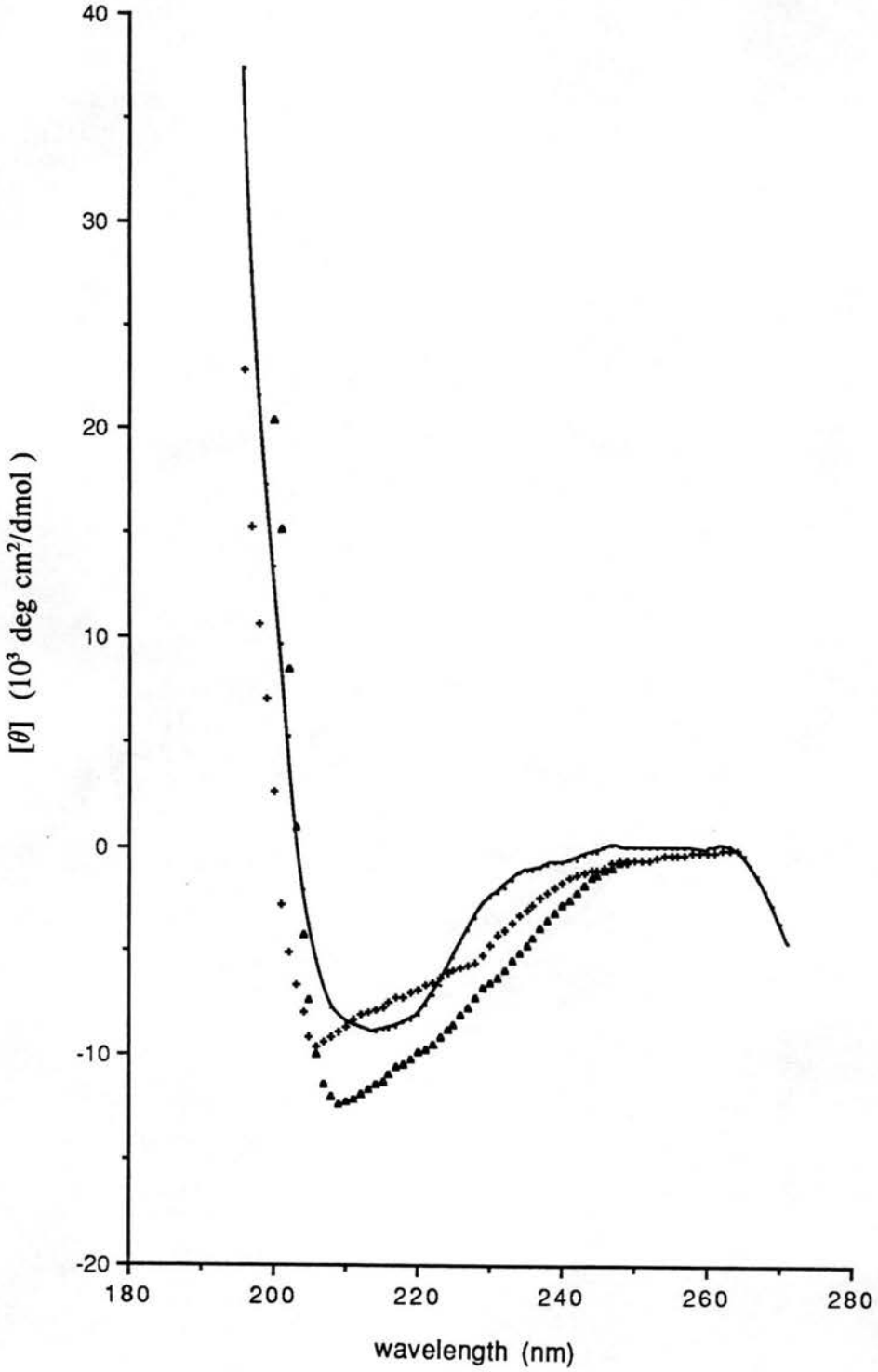


Figure IV-23. Far-uv CD of native porcine pepsinigen, in H_2O at 24°C (◡) , and temperature denatured porcine pepsinigen, in H_2O at 80°C (▲) and in 20 % EtOH at 65°C (+).

Table IV-10. Secondary structure content of native and thermally denatured pepsinogen in 5mM sodium phosphate at pH 6.5 estimated by the method of Bolotina & Lugauskas (1985).

Conditions	f_{α}	f_{β}	$f_{\beta\text{-turns}}$	$f_{\text{unord.}}$
24°C	0.19	0.32	0.18	0.31
80°C	0.24	0.20	0.25	0.31
65°C, 20 % EtOH	0.28	0.06	0.23	0.43

thermal transition. As can be seen from Figure IV-23, the CD spectra of denatured pepsinogen in water and in 20 % ethanol, even though they differ significantly from the CD of the native protein, do not resemble the usual spectra of denatured proteins or polypeptides in unordered conformation (Woody,1992). The near-uv spectra of pepsinogen obtained under thermally denaturing conditions showed about a 90 % decrease in intensity, rather typical for denatured proteins. We might assume that the content of the unordered conformation in pepsinogen does not increase significantly when the cooperative thermal transition is completed, although some changes do occur both in tertiary and secondary structure. The properties of thermally denatured pepsinogen seem similar to the "molten globule". In fact, estimates of the secondary structure content by the Bolotina & Lugauskas (1985) method have shown no significant increase in the content of unordered conformations in pepsinogen above the thermal transition in water, and only a 12 % increase in unordered conformation was detected in 20 % ethanol (Table IV-10). It seems that the thermal transition increases the amount of β-turns and residues in α-helical conformations in pepsinogen. We may suspect that the accuracy of Bolotina & Lugauskas (1985) method, as with any similar approach, is decreased when applied to non-native protein structures. Therefore, we suggest caution with respect to interpreting quantitative data obtained for non-native conditions. The apparent increase in α-helicity with temperature might be associated with partial loss of integrity in the β-sheet, which might occur at the N-terminus, where the prosegment is joined to the pepsin part of the globule. The only decrease in the content of secondary structure in pepsinogen upon the thermal denaturation is detected in the β-structure regions. As can

be seen from Table IV-10, the fraction of residues in β-structural conformations decreases by 12 % during the thermal denaturation in water and by 26 % in 20 % ethanol.

As mentioned above, pepsinogen exhibits a large negative band in the near-uv CD spectrum around 280 nm, which could be attributed to the L_b transition in Tyr. In fact all Trp residues are located in the pepsin part of pepsinogen, the conformation of which is very close to that in pepsin, as judged from the overlap of 284 topologically identical C_α atoms (r.m.s. deviation 0.90, Sielecki et al., 1990) and the secondary structure assignments (Chapter IV.2.a.3; Table IV-9). Our calculations of Trp CD (Fig.IV-24a,b) confirm that the Trp contribution in pepsin and pepsinogen remain almost unchanged. The negative couplet in the B_b region (Fig.IV-24a) might contribute to the slight shift in the position of the minimum in the far-uv spectrum, which occurs at 212 nm in pepsin and at 214 in pepsinogen, but the predicted difference in intensity of Trp spectra ($\sim$ 250 deg cm^2/dmol) is too small to account for the effect. It is tempting, based on the results obtained and on the similarity of the pepsinogen near-uv spectrum to that of AcTyr in water (Sears & Beychok, 1973), to ascribe the differences in the near-uv spectra to the Tyr37p of the prosegment. On the other hand Tyr37p is hydrogen-bonded to Asp215 in the active site, so it is not clear how flexible its side-chain is and what mechanism would determine its optical activity in solution. In cases like this, calculation of the aromatic CD with both Tyr and Trp contributions will be useful.

Therefore, we found (Makarov et al., 1991) that the effect of ethanol on thermal stability is similar in pepsin and pepsinogen. The stability of both proteins, at any pH

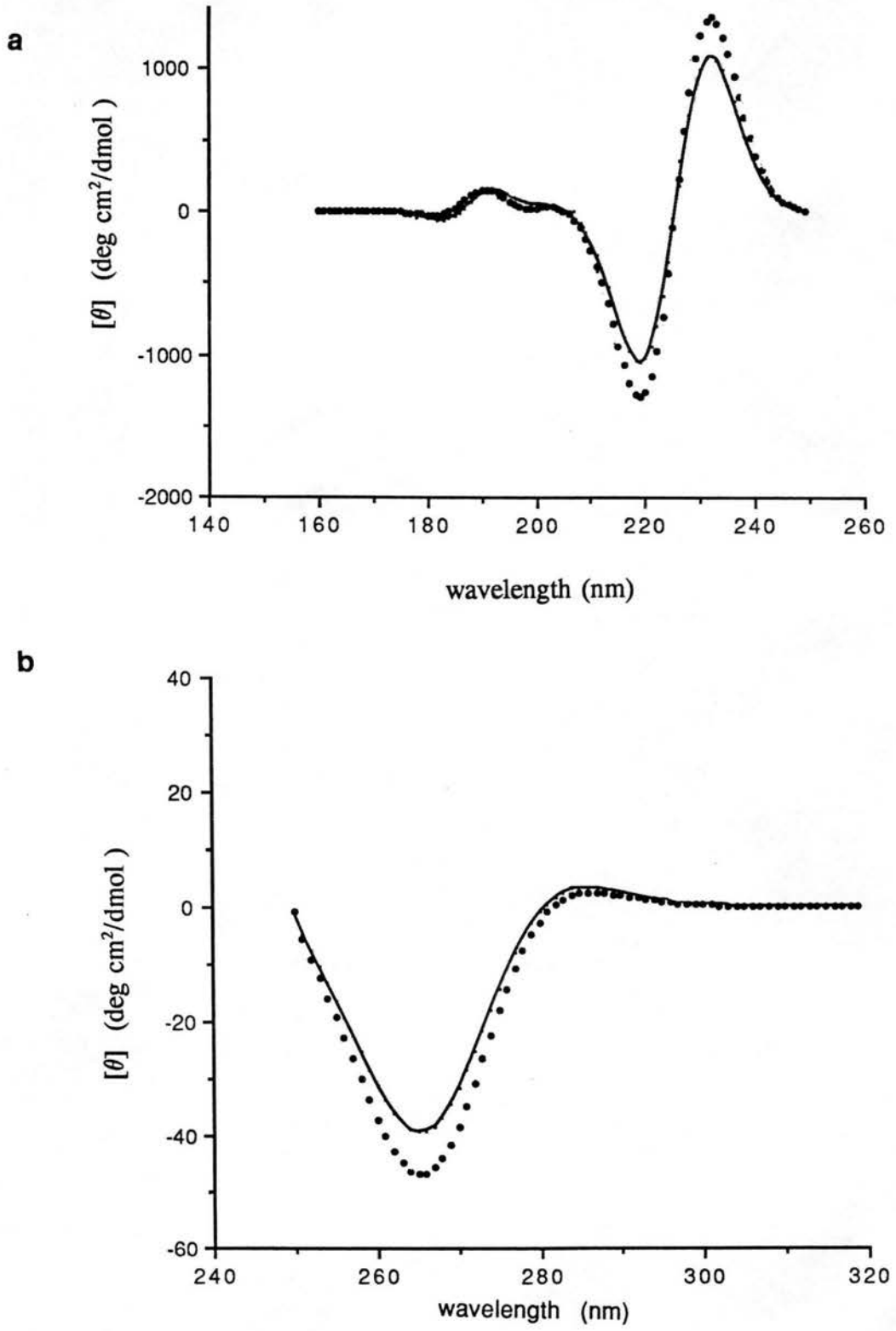


Figure IV-24. Calculated Trp CD of pepsinogen, 2pgs (_), and pepsin, 4pep (•) in the far uv (a) and in the near uv region (b).

considered, decreased when 20 % EtOH was added, while the number of cooperative regions does not change. This positively shows that the effect of ethanol on protein stability is separate from its inhibiting function on the pepsin active site. The effect of ethanol on the stability of pepsinogen increases with pH in the interval 6.5 - 8.0, i.e. on moving further from the isoelectric point. Thus ethanol accentuates the effect of the charge on the globular surface on its stability. Neither the effect of pH nor ethanol on the thermodynamic properties of pepsinogen is reflected in its secondary structure at temperatures below the thermal transition.

The transition from pepsinogen to pepsin causes a drastic change in the aromatic CD spectra in the near uv, whereas the secondary structure content remains almost unchanged. CD changes in the near uv might be due to the Tyr37p of the prosegment, which forms hydrogen bonds with Asp215 in the active site.

Pepsinogen was found to retain most of the secondary structure at temperatures above the thermal (cooperative) transition in aqueous solution or in 20 % EtOH (Makarov, A.A, Grishina I.B., Protasevich, I.I., Esipova, N.G., in preparation). Aromatic spectra in the near uv under these conditions are typical for denatured proteins or moieties with poorly defined tertiary structure. Thus, pepsinogen at high temperatures seems to assume the "molten globule" state .

Chapter V. Synopsis.

The exciton couplet approach was applied to estimate the circular dichroism (CD) of Trp side-chains in proteins. Calculations were performed by the origin-independent version of the matrix method, either for the indole B_b transition only or for the six lowest energy indole transitions.

A survey of a large number of proteins from the Protein Data Bank reveals a number of cases where readily detectable exciton couplets are predicted to result from the exciton coupling of Trp B_b bands. The predicted CD spectra are generally couplets, often dominated by the contributions of the closest pair, but sometimes exhibit three distinct maxima. It was found that Trp pairs with distances $\leq$ 13 Å, in favorable conformation, may contribute 15-20 % of the amplitude of the protein CD. The analyzed proteins were arranged according to functional classes and related structures.

Our calculations show that aromatic CD, Trp CD in particular, might contribute significantly to the far-uv CD of globular proteins, thus aromatic contributions should be considered in estimations of the secondary structure content from the far-uv CD.

Exciton Trp CD depends on the distance and relative orientation of Trp pairs, and reflects the spatial arrangement of Trp residues in the protein. Thus, conformational changes of Trp residues, and adjacent regions, might be deduced from Trp CD with further development of this approach.

The distance dependence of exciton splitting, rotational and couplet strengths of Trp pairs in globular proteins show general agreement with theoretical predictions. No indication of any preferential conformation of Trp pairs in proteins were found.

The dependence of the CD of a Trp pair upon its distance and geometry has been analyzed and tabulated. The effects of varying exposure of Trp chromophores and nearby charges on Trp CD have been examined.

Some of our calculations of the near-uv CD have been successful. The calculated near-uv CD for hen lysozyme was found to match the experimental amplitude at the L_a and L_b bands and to coincide with experimental position at the L_a band. It was predicted that mixing with far-uv transitions are as important in determining the CD intensity of the near-uv transitions as the coupling among near-uv transition.

It was found that in several cases, changes in protein Trp CD can be attributed to a specific Trp pair and explained as a definite change in its conformation. The differences in predicted Trp CD of hen lysozyme, turkey lysozyme and hen lysozyme bound to monoclonal antibody were attributed to conformational changes at a single residue, Trp62, which were discussed in relation to mutations and conformational changes in the adjacent area of the protein structure. Applications of the exciton couplet approach are discussed for various crystal forms of hen lysozyme, turkey and human lysozyme. Trp62 in hen lysozyme was found to be sensitive to the perturbations of the protein surface due to binding of substrate, antibodies and intermolecular contacts in the crystal. Conformational changes of Trp62 are predicted to have a strong effect on the overall Trp CD of lysozyme.

The predicted Trp CD of α-chymotrypsin and chymotrypsinogen A explains the differences in experimental CD of those proteins as attributable to the changes of relative geometry of the Trp172-Trp215 pair triggered by the movement of the 214-220 portion of the chain that leads to formation of the substrate recognition site of the enzyme.

It was shown that dimerization of hen lysozyme and bovine chymotrypsinogen A affects the conformation of the surface Trp residue which leads to significant changes in CD spectra.

Predicted Trp CD was compared with experimental results for various lysozymes, α-chymotrypsin and chymotrypsinogen A, concanavalin, dihydrofolate reductase and ribonuclease from *Bacillus intermedius* 7P (binase). We found that the exciton couplet approach might be useful in relating Trp CD and changes in protein structure upon local mutations and conformational changes involved in enzyme activation.

Predicted Trp CD was compared with unique CD contributions, obtained through deconvolution of the CD spectra on a set of secondary structure reference spectra while accounting for unique CD contributions with three log-normal functions, for various proteins: concanavalin, dihydrofolate reductase and binase. Possibilities of extension of the Bolotina & Lugauskas (1985) method by combining CD and infrared data, and by using standard values for aromatic contributions were discussed.

It was found that aromatic residues in globular structure are highly conserved, and if mutated, the replacement is more likely to be an aromatic residue as well. Mutations of aromatic to aromatic side chains may not distort the local structure. Thus, protein mutants can be produced with very similar overall structure and conformation of aromatic

residues, while the intensity of aromatic CD contribution from a specific site of the globule could be intensified or made negligible by a proper arrangement of specific chromophores.

Cooperative regions (energetic domains) were studied in binase and porcine pepsin by combining a CD analysis of the structural changes in the proteins during thermal denaturation and under various solvent conditions with thermodynamic properties observed by scanning microcalorimetry. Estimates of secondary structure were obtained from CD spectra, taking side-chain CD into account. It was found that neither of the proteins show any changes in secondary structure or local environment of aromatic amino acids upon separation of the energetic domains.

The structural regions in binase corresponding to energetic domains were identified through the analysis of the secondary structure content of the partial denatured globule. Preservation of part of the secondary structure at temperatures above the thermal transition was found for pepsinogen.

It was shown that binase is converted from a single cooperative system into two separate energetic domains when ion pairs are disrupted, whereas the size of cooperative units in pepsin decreases as the electrostatic repulsion between regions in the molecule increases.

Overall, theoretical, experimental and deconvolution analysis of the aromatic, particularly Trp, CD spectra were applied to study local conformational changes in globular proteins and problems of protein cooperativity in combination with calorimetric studies and analysis of the x-ray diffraction data.

REFERENCES

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

Acharya, K.R., Stuart, D.I. , Walker, N.P.C., Lewis, M. & Phillips, D.C. (1989) *J. Mol. Biol.* **208**, 99-127.

Aphanasenko, G.A., Dudkin, S.M., Kaminir, L.B., Leshinskaya, I.B. & Severin, E.S. (1979) *FEBS Lett.* **97**, 77-80.

Anderson, W.F., Ohlendorf, D.H., Takedo, Y. & Matthews, B.W. (1981) *Nature* **290,** 744 - 749.

Anfinsen, C. B. (1956) *J. Biol. Chem.* **221**, 405 - 412.

Anson, M. L. (1945) *Adv. Prot. Chem.* **2,** 361-386.

Artimiuk, P.J. & Blake, C.C.F. (1981) *J. Mol. Biol.* **152**, 737 - 762.

Artimiuk, P.J., Blake, C.C.F., Rice, D.W. & Wilson, K.S. (1982) *Acta Cryst.* **B38**, 778-783.

Andreeva, N.S. Zdanov, A.S., Gustchina, A.E., Fedorov, A.A. (1984) *J. Biol. Chem.* **259**, 11353-11365.

Antonov, V.K. (1977) in *Acid Proteases, Structure, Function and Biology* (Tang, J., ed.) 179-198, Plenum Press, New York.

Auer, H.E. (1973) *J. Am. Chem. Soc.* **95**, 3003-3011.

Banaszak, L.S., Andrews, P.A., Burgner, J.W., Eylar, E.H. & Gurd, F.R.N. (1963) *J. Biol. Chem.* **238**, 3307-3314.

Barlow, D. J. & Thornton, J., M. (1983) *J. Mol. Biol.* **168**, 867-885.

Battistel, E., Bianchi, D. & Rialdi, G. (1991) *Pure & Appl. Chem.* **63**, 1483-1490.

Bayley, P.M. (1973) *Progr. Biophys. Mol. Biol.* **2**, 1 - 76.

Bayley, P.M., Nielsen, E.B. & Schellman, J.A (1969) *J. Phys. Chem.* **73**, 228-243.

Baudet, S. & Janin, J. (1991) *J. Mol. Biol.* **219**, 123-132.

Becktel, W.J. & Schellman , J.A. (1987) *Biopolymers* **26**, 1859 - 1877.

Beddell, C.R., Blake, C.C.F. & Oatley, S.J. (1975) *J. Mol. Biol.* **97**, 643-654.

Biltonen, R., Lumry, R, Madison, V. & Parker, H. (1965) *Proc. Natl. Acad. Sci.* USA **54**, 1018-1024.

Birktoft J.J. & Blow D.M. (1972) *J. Mol. Biol.* **122**, 535- 542.

Blake, C.C.F., Mair, G.A., North, A.C.T., Phillips, D.C. & Sarma, V.R. (1967) *Proc. R. Soc.* London Ser.B, **167**, 365-377.

Blevins, R.A. & Tulinsky, A. (1985) *J. Biol. Chem.* **260**, 4264-4275.

Böhm, G., Muhr, R. & Jaenicke, R. (1992) *Prot. Eng.* **5**, 191-195.

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

Bolotina, I.A. & Lugauskas, V.Yu (1985) *Mol.Biol.* (USSR, engl. transl) **19**, 1154-1166.

Bolotina, I.A., Dudkin, S., Lugauskas, V.Yu., Khandanyan, A. G., Leshinskaya, I.B. (1979) *Bioorg. Chimia* (USSR) **5**, 203-209.

Bolotina, I.A., Chekhov, V.O., Lugauskas, V.Yu. & Ptitsyn, O.B. (1980a) *Mol.Biol.* (USSR) **15**, 167-175.

Bolotina, I.A., Chekhov, V.O., Lugauskas, V.Yu., Finkel'shtein, A.V. & Ptitsyn, O.B. (1980b) *Mol. Biol.* (USSR) **14**, 701-709.

Bolotina, I.A., Chekhov, V.O., Lugauskas, V.Yu. & Ptitsyn, O.B. (1980c) *Mol.Biol.* (USSR) **14**, 891-901.

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

Brahms, S., Brahms, J., Spach, G. & Brack, A. (1977) *Proc. Natl. Acad. Sci.* USA **14**, 3208-3212.

Brandts, J. F., Hu, C. Q., Lin, L.N. & Mas, M.T. (1989) *Biochemistry* **28**, 8588-8596.

Brenman, R.G. & Matthews, B.W. (1989) *J. Biol. Chem.* **264**, 1903 - 1906.

Brown, A., Kemp, C.M. & Mason, S.F. (1971) *J. Chem. Soc.* (**A**) 751 - 755.

Burley, S.K. & Petsko, G.A. (1988) *Advan. Protein Chem.* **39**, 125-192.

Cameron, D.L. & Tu, A.T. (1977) *Biochemistry* **16**, 2546-2553.

Cantor, C.R. & Schimmel, P.R. (1980) *Biophysical Chemistry*, Freeman & Co., New York, part **II**, 409-433.

Cantor, C.R. & Timasheff, S.N. (1982) *Proteins* **5**, 202.

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

Chen, A.K. & Woody, R.W. (1971) *J. Am. Chem. Soc.* **93**, 29-37.

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

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

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

Cooper, T.M. & Woody, R.W. (1990) *Biopolymers* **30**, 657-676.

Cosani, A., Peggion, E., Verdini, A.S., Terbojevich, M. (1968) *Biopolymers* **6**, 963 - 971.

Cowburn, D.A., Brew, K. & Gratzer, W.B. (1972) *Biochemistry* **11**, 1228-1234.

Creighton, T.E. (1985) *J. Phys. Chem.* **89**, 2452 - 2459.

Davydov, A.S. (1971) *Theory of Molecular Excitons*, Plenum Press, New York - London.

Dayhoff, M.O. (1972) *Atlas of protein sequence and structure*, NBRF, Georgetown University Medical Center, Washington, D.C., vol.**5**, p. D365.

Day, L.A. (1973) *Biochemistry* **12**, 5329 -5339.

Del Bene, J. & Jaffe, H.H. (1968) *J. Chem. Phys.* **48**, 1807-1813.

Diamond, R (1974) *J. Mol. Biol.* **82**, 371-391.

Dill, K.A. (1990) *Science* **250**, 297.

Doig, A.J. & Williams, D.H. (1991)*J. Mol. Biol.* **217**, 389-398.

Dong, A., Huang, P. & Caughey, W.S. (1990) *Biochemistry* **29**, 3303-3308.

Dong, A., Caughey, B., Caughey, W.S., Bhat, K.S. & One, J.E. (1992) *Biochemistry* **31**, 9364-9370.

Dorrington, K.J. & Hoffman, T. (1973) *Can. J. Biochem.* **51**, 1059 - 1065.

Drenth, J., Jansonius, J.N., Koek, R. & Wolthers, B.G. (1971) In *The Enzymes* (P.D.

Boyer, ed.), 3rd ed., vol.**3**, pp. 485 - 499. Academic Press, New York.
Edelman, G.M. (1970) *Biochemistry* **9**, 3197 - 3204.
Edge, V., Allewell, N.M. & Sturtevant, J.M. (1985) *Biochemistry* **24**, 5899-5906.
Edge, V., Allewell, N.M. & Sturtevant, J.M. (1988) *Biochemistry* **27**, 8081-8087.
Ellis, R.L., Kuehnlenz, G. & Jaffe, H.H. (1972) *Theor. Chim. Acta* **26**, 131-140.
Esipova, N.G., Makarov, A.A., Mageladze, G.N., Madzhagaladze, G.V, & Volkenshtein, M.V. (1978) *Mol. Biol.* (USSR) **12**, 1152-1162.
Esipova, N.G., Protasevich, I.I., Platonov, A.L., Makarov, A.A., Pavlovsky, A.G., Chekhov, O.V. (1985) *Proceedings of II International Conference on Biothermodynamics*, Laggner, P. ed., Austria,15.
Esipova, N, Grishina, I, Protasevich, I, Frank, E., Lobachov, V. & Makarov, A. (1993) In "*Stability and stabilization of enzymes*", W.J.J. van den Tweel, A.Harder, Eds, pp.309-314, Elsevier Publishers B.V.
Fasman, G.D. & Potter, J. (1967) *Biochem. Biophys. Res. Commun.* **27**, 209.
Fasman, G.D., Foster, R.J. & Beychok, S. (1966) *J. Mol. Biol.* **19**, 240-253.
Filimonov, V.V., Pfeil, W., Tsalkova, T.N. & Privalov, P.L. (1978) *Biophys. Chem.* **8**, 117 - 122.
Finkel'shtein, A.V., Ptitsyn, O.B. & Kozitsyn S.A. (1977) *Biopolymers* **16**, 497-524.
Fraser, R.D.B. & Suzuki, E. (1973) in *Physical Principles and Techniques of Protein Chemistry* (Leach, S.J., ed.) Academic Press, New York, Part **C**,301-355.
Freire, E. & Biltonen, R.L. (1978) *Bio*polymers **17**, 463-479.
Freer, S.T., Kraut, J., Robertus, J.D., Wright, H.T. & Xuong, N.H. (1970) *Biochemistry* **9**, 1997-2009.
Fresnel, A. (1825) *Ann. Chim. Phys.* **28**, 147.
Frigerio, F., Coda, L., Pugliese, L., Lionetti, C., Menegatti, E., Amiani, G. & Schnebli, H.P., Aschenzi, P. & Bolognese, M. (1992) *J. Mol. Biol.* **225**, 107-123.
Fruton, J.S. (1971) in *The Enzymes* (Boyer, P.D., ed.) **3**, pp. 127-128, Academic Press, New York.
Fujinaga, M., Sielecki, A.R., Read, Ardelt, W., Laskowski, M.Jr. & James, M.N.G. (1987) *J. Mol. Biol.* **195**, 397- 418.
Gertler, A., Walsh, K.A. & Neurath, H. (1974) *Biochemistry* **13**, 1302 - 1310.
Gill, S.J., Rickey, B., Bishop, G. & Wyman, J. (1985) *Biophys. Chem.* **21**, 1 - 14.
Gitelson, G.I., Griko, Yu.V., Kurochkin, A.V., Rogov, V.V., Kutyshenko, V.P., Kirpichnikov, M.P. & Privalov, P.L. (1991) *FEBS Letters* **289**, 201 - 204.
Griko, Yu. V., Rogov, V.V. & Privalov, P.L. (1992) *Biochemistry* **31**, 12701 - 12705.
Golubenko, I.A., Balaban, N.P., Leshinskaya, T.I., Volkova, G.I., Kleiner, Chepurnova, N., K., Afanasenko, G., A. & Dudkin, S., M. (1979) *Biokhimia* (USSR) **44**, 640-647.
Golubenko, I., A., Leshinskaya, T., I. & Dudkin, S., M. (1979) *Bioorg. Khimia* (USSR) **7**, 1201-1208.
Goodman, M. & Toniolo C. (1968) *Biopolymers* **6**, 1673-1689.
Gorbunoff, M.J. (1971) *Biochemistry* **10**, 250-257.
Goux, W.J. & Hooker, T.M.,Jr. (1975) *J. Am. Chem. Soc.* **97**, 1605-1606.

Goux, W.J. & Hooker, T.M.,Jr. (1980a) *J. Am. Chem. Soc.***102**, 7080-7088.
Goux, W.J. & Hooker, T.M.,Jr. (1980b) *Biopolymers* **19**, 2191-2208.
Goux, W.J, Cooke, D.B., Rodriguez, R.E. & Hooker, T.M.,Jr. (1974) *Biopolymers* **13**, 2315-2329.
Glotov, B.O., Kozlov, L.V. & Zavada, L.L. (1976) *Mol. Biol.* (USSR) **10**, 530-537.
Grishina, I.B., Bolotina, I.A., Esipova, N.G., Pavlovsky, A.G. & Makarov, A.A. (1989) *Mol. Biol.* (USSR, engl. transl.) **23**, 1157-1174.
Grishina, I.B., Makarov, A.A., Kuznetsov, D.V., Vasiliev, D.G., Pavlovskii, A.G., Aizenmenger, F. & Esipova, N.G. (1993) *Biofizika* (Russia) **38**, 22-30.
Green, N.M. & Melamed, M.D. (1966) *Biochem. J.* **100**, 614-621.
Greenfield, N. & Fasman, G.D. (1969) *Biochemistry* **8**, 4108-4116.
Gustshina, A.E., Andreeva, N.S. & Antonov, V.K. (1983) *Bioorg. Khimia* (USSR) **9**, 1620-1624.
Harada, N. & Nakanishi, K. (1972) *Acc. Chem. Res.* **5**, 257-263.
Harada, N. & Nakanishi, K. (1983) *Circular Dichroism Spectroscopy. Exciton Coupling in Organic Stereochemistry.* University Science Books .
Hartley, R.W. (1989) *Trends Biochem. Sci.* **14**, 450-456.
Hennessey, J.P. Jr & Johnson W.C., Jr (1981) *Biochemistry* **20**, 1085-1094.
Hennessey, J.P. Jr & Johnson W.C., Jr (1982) *Anal. Biochem.* **125**, 177-188.
Hermann, M.S., Richardson, C.E., Setzler, L.M., Behnke, D.W. & Thomson, R.E. (1978) *Biopolymers* **17**, 2107-2120.
Herriott, R.M. (1939) *J. Gen. Physiol.* **22**, 65 - 78.
Hess, G.P. (1971) In *The Enzymes* 3rd Ed. Vol. 3. (P.D. Boyer, ed.) Acad. Press, New York, p.171.
Hill, C., Dodson, G., Heinemann, U., Saenger, W., Mitsui, Y., Nakamura, K., Borisov, S., Tishenko, G., Polyakov, K. & Pavlovsky, S. (1983) *Trends Biochem. Sci.* **8**, 364 - 369.
Hogle, J., Rao, S.T., Mallikarjunan, M., Beddel, C., McMullan, R.K. & Sundaralingam, M. (1981) *Appl. Cryst.* **B37**, 591-597.
Hooker, T.M., Jr. & Schellman, J.A. (1970) *Biopolymers* **9**, 1319-1348.
Holzwarth, G. & Doty, P. (1965) *J. Am. Chem. Soc.* **87**, 218.
Horovitz, A. & Fersht, A. (1992) *J. Mol. Biol.* **224**, 733-740.
Horovitz, A., Serano, L. & Avron,B (1990) *J. Mol. Biol.* **216**, 1031-1044.
Hu, C.Q. & Sturtevant, J.M. (1987) *Biochemistry* **26**, 178-182.
Hunter, C.A., Singh, J. & Thornton, J.M. (1991) *J. Mol. Biol.* **218**, 837-846.
Ikeda, K., Hamaguchi, K., Yamamoto, M. & Ikenaka, T. (1968) *J. Biochem.* **63**, 521-531.
Ikegami, A. (1977) *Biophys. Chem.* **6**, 117-130.
Jackson, W.M. & Brandts, J.F. (1970) *Biochemistry* **9**, 2294-2301.
Janin, J. & Wodak, S.J. (1983) *Prog. Biophys. Mol. Biol.* **42**, 21-78.
James, M.N.G. & Sielecki, A.R. (1986) *Nature* **319**, 33-38.
Jaques, P., Faure, J., Chavlet, T.E. & Jaffe, H.H. (1981) *J. Phys. Chem.* **85**, 473- 479.
Johnson W.C. Jr (1988) *Ann. Rev. Biophys. Biophys. Chem.* **17**, 145-166.

Johnson, W.C., Jr. & Tinoco, I., Jr. (1969) *Biopolymers* **8**, 715-732.
Jones, T.A. (1978) *J. Appl. Cryst.* **11**, 268.
Kabsch W. & Sander C. (1983) *Biopolymers* **22**, 2511-2637.
Kalnin, N.N., Baikalov, I.A. & Venyaminov, S.Yu. (1990) *Biopolymers* **30**, 1271-1280
Karpeiskii, M., Ya, Khandanyan, A., Zh., Chepurnova, N., K., Platonov, A.,L., Yakovlev, G., I. (1981) *Bioorg. Khimia* (USSR) **7**, 1669-1679.
Khandanyan, A., Zh. & Dudkin, S., M. (1979) *Bioorg. Khimia* (USSR) **5**, 1700-1709.
Konnert, I., N. & Hendrickson, W., A. (1980) *Acta Cryst.* **A36**, 344-349.
Kosen, P.A., Creigton, T.E. & Blot, E.R. (1983) *Biochemistry* **22**, 2433-2440.
Kozlov, L.V., Meshcheryakova, E.A., Zavada, L.I., Efremov, E.S., Rashkolevetskij, L.G. (1979) *Biokh*imia (USSR) **44**, 338-349.
Kraut J. (1971) In *The Enzymes* 3rd Ed. Vol. 3. *Edited by* P.D. Boyer. Acad. Press, New York, pp.165 - 183.
Kundrot, C.E. & Richards, F.M. (1987) *J.Mol. Biol.* **193**, 157-170.
Kundrot, C.E. & Richards, F.M. (1988) *J.Mol. Biol.* **200**, 401-410.
Kurachi, K., Sieker, L.C. & Jensen, L.H. (1976) *J. Mol. Biol.* **101**, 11-24.
Kuwajima, K., Garvey, E.P., Finn, E.B., Mattews, C.R. & Sugai, S. (1991), *Biochemistry* **30**, 7693-7703.
Kuznetsov, D.A. & Lim, H.A. (1992) *J. Mol. Graphics* **10**, 99.
Laiken, S., Printz, M. & Craig, L.C. (1967) *J. Biol. Chem.* **244**, 209.
Landau, L.D. & Lifshits, E.M. (1958) *Statistical Physics.* Reading, Mass: Addison-Wesley.
Levitt, M. & Chothia, C. (1976) *Nature* **261**, 552 - 558.
Levitt, M. & Greer, J. (1977) *J. Mol. Biol.* **114**, 181-239.
Loewenthal, R., Sancho, J. & Fersht, A.R (1991) *Biochemistry* **30**, 6775-6779.
Madison, V., Schellman, J. (1972) *Biopolymers* **11**, 1041-1076.
Makarov, A.A., Esipova, N.G., Monaselidze, D.R., Mgeladze, G.N. & Madzhagaladze, G.V. (1976) *Biochim. Biophys. Acta* **434**, 286-289.
Makarov, A.A., Protasevich, I.I., Frank, E.G., Grishina, I.B., Bolotina, I.A. & Esipova, N.G. (1991) *Biochim. Biophys.* Acta **1078**, 283-288.
Makarov, A.A., Protasevich, I.I., Frank, E.G., Grishina, I.B., Bolotina, I.A. & Esipova, N.G. (1991) *Biofizika* (USSR) **36**, 39-45.
Makarov, A.A., Protasevich, I.I., Grishina, I.B., Frank, E.G. & Esipova, N.G. (1994) In *Modern Enzymology: Problems and Trends*, B.I.Kurganov,V.I.Tishkov, Eds, NovaNova Science Publishers, Inc.. (in press).
Manavalan, P. & Johnson, W.C.,Jr. (1988) *Anal. Biochem.* **167**, 76-85.
Manly, S.P., Matthews, K.S. & Sturtevant, J.M. (1985) *Biochemistry* **24**, 3842-3846.
Manning, M.C. & Woody, R.W. (1987) *Biopolymers* **26**, 1731-1752.
Manning, M.C. & Woody, R.W. (1989) *Biochemistry* **28**, 8609 -8613.
Manning, M. C. & Woody, R. W. (1991) *Biopolymers* **31**, 569 -586.
Manning, M.C., Illangasekare, M. & Woody, R.W. (1988) *Biophys. Chem.* **31**, 87-88.
Mateo, P.L. & Privalov, P.L. (1981) *FEBS Lett.* **123**, 189-192.
Matthews, C.K. & van Holde, K.E. (1990) *Biochemistry*, The Benjamin/Cummings Publishing Company, Inc, Redwood City, CA.

Matsumura, M., Becktel, W.J. & Matthews, B.W (1988) *Nature* **334**, 406-410.
McConn, J., Fasman, G.D. & Hess, G.P. (1969) *J. Mol. Biol.* **39**, 569-586.
Menez, A., Bouet, F., Tamiya, N. & Fromageot, P. (1976) *Biochim. Biophys. Acta* **453**, 121-132.
Merck Index. An Encyclopedia of Chemicals and Drugs. (1968) (Paul G. Stecher, ed.) Merck &Co., Inc., Rahway, N.J., USA
Moffitt, W. (1956) *J. Chem. Phys.* **25**, 467- 478.
Moffitt, W. & Moscovitz, A. (1959) *J. Chem. Phys.* **30**, 648-660.
Moffitt, W.,Fitts, D.D. & Kirkwood, J.G. (1957) *Proc. Natl. Acad. Sci.* **43**, 723
Morozov, Yu.V., Bazhulina, N.P., Chekhov, V.O., Savitsky, A.V., Grishina, I.B. & Makarov, A.A.(1994) In Modern Enzymology: Problems and Trends, (B.I.Kurganov, V.I.Tishkov, eds.), NovaNova Science Publishers, Inc.. (in press).
Moult, J., Yonath, W., Traub, A., Smilansky, A., Podjarny, A., Rabinovich, D. & Saya, A. (1976) *J. Mol. Biol.* **100**, 179-195.
Murrell, J. N. & Harget, A.J. (1972) *Semi-Empirical Molecular Orbital Theory of Molecules*, Wiley, London.
Nakatani, H., Kitagishi, K. & Hiromi, K. (1980) *J. Biochem.* **87**, 563-571.
Nielsen, E.B., & Schellman, J.A. (1967) *J. Chem. Phys.* **23**, 1833-1840.
Nicholson, H., Becktel, W.J. & Matthews, B.W (1988) *Nature* **336**, 651-655.
Nicholson, H., Soderling, E., Trounrud, D.E. & Matthews, B.W (1989) *J. Mol. Biol.* **210**, 181-193.
Nishimoto, K. & Forster, L.S. (1965) *Theor. Chim. Acta* (Berlin) **3**, 407-417.
Nishimoto, K. & Forster, L.S. (1966) *Theor. Chim. Acta* (Berlin) **4**, 155 -165.
Northrop, J.H. (1946) *J. Gen. Physiol.* **30**, 177-185.
Nukrianova, K.M., Shulga, A.A. & Zakhar'ev, *Dokl. Akad. Nauk* (USSR) **309**, 1476.
Oppenheimer, H.L., Labouesse, B. & Hess, G.P. (1966) *J. Biol. Chem.* **241**, 2720.
Pabo, C.O. & Sauer, R.T (1984) *Ann. Rev. Biochem.* **53**, 293 - 321.
Pabo, C.O., Sauer, R.T., Sturtevant, J.M. & Ptashne, M. (1979) *Proc. Natl. Acad. Sci.* USA **76**, 1608 - 1612.
Padlan, A., Silverton, E., Sheriff, S., Cohen G., Smith-Gil, S. & Davies, D. (1989) *Proc. Natl. Acad. Sci.* USA **86**, 5938-5942.
Pakula, A.A & Sauer, R.T. (1989) *Proteins* **5**, 202 - 210.
Pancoska P., Blazek, M. & Keiderling T.A. (1992) *Biochemistry* **31**, 10250-10257.
Paskowsky, D.J., Stevens, E.S., Bonora, G.M. & Toniolo, C. (1978) *Biochim. Biophys. Acta* **535**, 188-192.
Pavlovsky, A.G. & Sanishvili, R.G (1988) *Dokl. Akad. Nauk* (USSR) **30**, 1254.
Pavlovsky, A.G., Sanishvili, R.G. & Borisova, S.N. (1989) *Krystallografia* (USSR) **34**, 137.
Peggion, E., Cosani, A., Verdini, A.S., Del Pra, A. & Mammi, M. (1968) *Biopolymers* **6**, 1477-1486.
Perczel, A., Hollosi, M., Tusnady, G. & Fasman, G.D. (1991) *Prot. Eng.* **4**, 669-679.
Perczel, A., Park, K. & Fasman G.D. (1992) *Proteins* **13**, 57-69.
Pfeil, W. (1981) *Mol. Cell. Biochem.* **40**, 3 - 28.

Pfeil, W. & Privalov, P.L. (1976) *Biophys. Chem.* **4**, 23 - 32.
Pflugrath, J.W., Saper, M.A. & Quiocho, F.A. (1984) in *Methods and applications in crystallographic computing.* (S. Hall and T. Ashiaka, eds.) Clarendon Press, Oxford, 407.
Phillips, D.C. (1966) *Sci. Am.* **215**, 78-90.
Phillips, L.A. & Levy, D.H. (1986a) *J. Chem. Phys.* **85**, 1327 - 1332.
Philllips, L.A. & Levy, D.H. (1986b) *J. Chem. Phys.* **90**, 4921-4923.
Platt, J.R. (1949) *J. Chem. Phys.* **17**, 484-495.
Privalov, P.L. (1974) *FEBS Letters* **40S**, S140 - S153.
Privalov, P.L. (1979) *Adv. Prot. Chem.* **33**, 167 - 241.
Privalov, P.L. (1982) *Adv. Prot. Chem.* **35**, 1-104.
Privalov, P.L. (1989) *Ann. Rev. Biophys. Biophys. Chem.* **18**, 47 - 69.
Privalov, P.L. & Gill, S.J. (1988) *Adv. Prot. Chem.* **39**, 193.
Privalov, P.L. & Khechinashvili, N.N. (1974) *J.Mol.Biol.* **86**, 665-684.
Privalov, P.L. & Medved'ev, L.V. (1982) *J. Mol. Biol.* **159**, 665 - 683.
Privalov, P.L. & Potekhin, S.A. (1986) *Methods Enzymol.* **131**, 4 -51.
Privalov, P.L., Plotnikov, V.V. & Filimonov, V.V. (1975) *J. Chem. Thermodyn.* **7**, 41-47.
Privalov, P.L., Mateo, P.L. & Khechinashvili, N.N. (1981) *J. Mol. Biol.* **152**, 445-464.
Privalov, P.L., Griko, Yu.V., Venyaminov, S.Yu. & Kutyshenko, V.P. (1986) *J. Mol. Biol.* **190**, 487 - 498.
Privalov, P.L., Tiktopulo, E.I., Venyaminov, S.Yu., Griko, Yu.V., Makhatadze, G.I. & Khechinashvili, N.N. (1988) *J.Mol.Biol.* **205**, 737-750.
Protasevich, I.I., Makarov, A.A., Platonov, A.L. & Esipova, N.G. (1985) *Biofizika* (USSR) **30**, 494.
Protasevich, I.I. (1985) Ph. D. Thesis, Dep. of Biology, Moscow State University, Moscow, Russia.
Protasevich, I.I., Platonov, A.L., Pavlovsky, A.G. & Esipova, N.G. (1987) *J. Biomol. Struct. Dyn.* **4**, 885-893.
Provencher, S.W. & Glöckner, J. (1981) *Biochemistry* **20**, 1085-1094.
Quadrifoglio, F., Ius, A. & Crescenzi, V. (1970) *Makromol. Chem.* **167**, 297-303.
Quiocho, F.A., Sack, J.S. & Vyas, N.K. (1987) *Nature* **329**, 561-564.
Rao, S.T., Hogle, J., & Sundaralingam, M. (1983) *Acta Cryst.* **C39**, 237-240.
Šali, D., Bycroft, M. & Fersht, A. (1988) *Nature* **335**, 740-743.
Sancho, J. & Fersht, A.R. (1992) *J. Mol. Biol.***224**, 741-747.
Sancho, J. , Niera, J.L. & Fersht, A.R. (1992) *J. Mol. Biol.***224**, 749-758.
Sanishvili, R.G (1989) Ph.D Thesis, Institute of Crystallography, Russian Academy of Sciences.
Sarver, R.W.Jr. & Krueger, W.C. (1991a) *Anal. Biochem.* **194**, 89-100.
Sarver, R.W.Jr. & Krueger, W.C (1991b) *Anal. Biochem.* **199**, 61-67.
Sears, D.W & Beychok, S. (1973) in *Physical Principles and Techniques of Protein Chemistry* (Leach, S.J., ed.) Academic Press, New York, Part C, 445-593.
Schellman, J.A. (1968) *Accts. Chem. Res.* **1**, 144 - 151.

Schellman, J.A. & Oriel, P. (1962) *J. Chem. Phys.* **37**, 2114-2124.
Schleker, W. & Fleischhauer, J. (1987) Z. *Naturforsch.* 42a, 361-366.
Schwert, G. W. (1949) *J. Biol. Chem.* **179**, 655 - 664.
Sielecki, A.R., Fedorov, A.A., Boodhoo, A., Andreeva, N.S., James, M.N.G. (1990) *J. Mol. Biol.* **214**, 143-170.
Sielecki, A.R., Fujinaga, M. Read, R.J. & James, M.N.G. (1991) *J. Mol. Biol.* **219**, 671 - 692.
Sigler, P.B., Blow, D.M., Matthews, B.W. & Henderson, R. (1968) *J. Mol. Biol.* **35**, 143-164.
Simpson, R.T., Riordan, J.F. & Vallee, B.L (1963) *Biochemistry* **2**, 616-622.
Sheriff, S., Silverton, E., Padlan, A., Smith-Gil, S., Finzel, B., Davies, D. R. (1987) *Proc. Natl. Acad. Sci.* USA **84**, 8075-8079.
Shiraki, M (1969) *Sci. Pap. Coll.Gen. Educ., Univ. Tokyo* **19**,151-173.
Shiraki, M. & Imahori, K. (1966) *Sci. Pap. Coll. Gen. Educ., Univ. Tokyo* **16**, 215-229.
Shoemaker, K.R., Kim, P.S., York, E.J., Stewart, J.M. & Baldwin, R.L. (1987) *Nature* **326**, 563 - 567.
Siano, D.B. & Metzler, D.E. (1969) *J. Chem. Phys.* **51**, 1856-1861.
Silverton, E.W., Padlan, E.A., Davies, D.R., Smith-Gill, S. & Potter, M. (1984) *J. Mol. Biol.* **180**, 761-765.
Sodek, J. & Hofmann, T. (1970) *Can. J. Biochem.* **48**, 1014 - 1017.
Spolar, R.S.,Ha, J-H. & Record, M.T. (1989) *Proc. Natl. Acad. Sci.* **86**, 8382-8385.
Sreerama, N. & Woody, R.W. (1993) *Anal. Biochem.* **209**, 32-44.
Sreerama, N., Manning M.C. & Woody, R.W. (1991) in *Proc. IV Int.Conf. CD*, Bochum, Germany., pp. 186-201.
Stevens, L., Townend, R., Timasheff, S.M., Fasman, G.D. & Potter, J. (1968) *Biochemistry* **1**, 3717-3720.
Steitz, T.A., Henderson, R. & Blow, D.M. (1969) *J. Mol. Biol.* **46**, 337-348.
Sternberg, M.J.E., Hayes, F.R.F., Russel, A.J., Thomas, P.G., Fersht, A.R. (1987) *Nature* **330**, 86-88.
Strassburger, W., Glatter U., Wollmer A., Fleischhauer, J., Mercola, D.A., Blundell T.L., Glover, I., Pills, J.E., Tickle, I.J. & Wood, S.P. (1982) *FEBS Letters* **139**, 295-299.
Strickland, E.H. & Mercola, D. (1976) *Biochemistry* **15**, 3875-3884.
Sturtevant, J.M. (1987) *Ann. Rev. Phys. Chem.* **38**, 463-488.
Tang, J. (1965) *J. Biol. Chem.* **240**, 3810-3815.
Tiktopulo, E.I. & Privalov, P.L. (1978) *FEBS Letters* **91**, 57 - 58.
Timasheff, S.N. & Bernardi, G. (1970) *Arch. Biochem. Biopys.* **141**, 53-58.
Tinoco, I., Jr. (1962) *Adv. Chem. Phys.* **4**, 113-161.
Tinoco, I., Jr. and Bush, C.A. (1964) *Biopolym. Symp.* **1**, 235-250.
Tinoco, I., Jr. , Woody, R.W. & Bradley, D.F. (1963) *J. Chem. Phys.* **38**, 1317.
Tiselius, A., Henschen, G.E. & Svenson, H. (1938) *Biochem. J.* **32**, 1814-1818.
Toumadje, A., Alcorn, S.W. & Johnson, W.C., Jr (1992) *Anal. Biochem.* **200**, 321-331.
Tronrud, D., Ten Eyck, L. (1989) *Ten Eyck-Tronrud refinement program. Description*

manual. Inst. of Mol. Biol., University of Oregon.
Tsong, T.Y., Hearn, R.P., Wrathall, D.P. & Sturtevant, J.M. (1970) *Biochemistry* **9**, 2666 - 2677.
Tsukada, H. & Blow, D.M. (1983) *J. Mol. Biol.* **184**, 703-711.
Tulinsky, A. & Blevins, R.A. (1987) *J. Biol. Chem.* **262**, 7737-7743.
Urry, D.W. (1968) *Ann. Rev. Phys. Chem.* **19**, 447-530.
van Stokkum, I.H.M., Spoelder, H.J.W., Bloemendal, M., van Grondelle, R. & Groen F.C.A. (1990) *Anal. Biochem.* **191**, 110-118.
Vassylyev, D.G. & Adzhubei, A.A. (1992) *J. Molec. Graphics* **10**, 278.
Venyaminov, S.Yu. & Kalnin, N.N. (1990) *Biopolymers* **30**, 1259-1271.
Vuilleumier, S., Sancho, J., Loewenthal, R. & Fersht, A.R (1993) *Biochemistry* **32**, 10303-10313.
Wang, D., Bode, W. & Huber, R. (1985) *J. Mol. Biol.* **185**, 595-624.
Weather, L.H., Gray, T.M., Gruetter, M.G., Anderson, D.E., Wozniak, J.A., Dahlquist, F.W. & Matthews, B.W. (1989) *Biochemistry* **28**, 3793-3797.
Wetlaufer, D.B. (1973) *Proc. Natl. Acad. Sci.* USA **70**, 697 - 701.
Willaert, K., Loewenthal, R., Sancho, J., Froyen, M., Fersht, A.R & Engelborghs, Y. (1992) *Biochemistry* **31**, 711-716.
Woody, R.W. (1973) *Tetrahedron* **29**, 1273-1283.
Woody, R.W. (1974) *Pept. Polypept. Proteins* (Blout, E.R., Bovey, F.A., Goodman, M. & Lotan, N., Eds.) Wiley-Interscience, New York, 338-349.
Woody, R.W. (1978) *Biopolymers* **17**, 1451-1461.
Woody, R.W. (1985) *The Peptides*, vol. **7**, (Hruby, V.J., ed.), Academic Press, Orlando, FL, pp. 51-104.
Woody, R.W. (1987) *Proc. Fed. Eur. Chem. Soc. 2nd Intl. Conf. Circular Dichroism*, Budapest, Hungary, 38-51.
Woody, R.W. (1992) *Adv. Biophys. Chem.* **2**, 31-79.
Woody, R.W. (1994) *Eur. Biophys. J.* (in press).
Wright, H.T. (1973a) *J. Mol. Biol.* **79**, 1-12.
Wright, H.T. (1973b) *J. Mol. Biol.* **79**, 13-23.
Yamamoto, Y. & Tanaka, J. (1972) *Bull. Chem. Soc. Jpn.* **45**, 1362 -1366.
Yang, C.C., Chang, C.C., Hayashi, K., Suzuki, T., Ikeda, K. & Hamaguchi, K. (1968) *Biochim. Biophys. Acta* **168**, 373-376.
Yang, J.T., Wu, C.-S.C. & Martinez, H.M. (1986) *Methods in Enzymology* **130**, 209-269.
Zav'yalov, V.P., Troitsky, G.V., Khechinashvili, N.N. & Privalov, P.L. (1977) *Biochim. Biophys. Acta* **492**, 102 -111.