Charge-Transfer Studies of the Availability of Aromatic Side Chains of Proteins in Guanidine Hydrochloride†

Carol R. Coan, Lois M. Hinman, and David A. Deranleau*

ABSTRACT: The ability of aromatic tryptophyl and tyrosyl side-chain donors to form charge-transfer (CT) complexes with the acceptor 1-methyl-3-carbamidopyridinium chloride has been used to investigate the degree of exposure of these aromatic residues in denatured proteins. The coplanar geometry of the CT complexes requires that virtually a full ring face of the donor be available for interaction with the acceptor, and the aromatic donor residues of lysozyme, trypsin, chymotrypsin, and the zymogens of the latter two enzymes do not appear to be wholly "exposed" in 6 *M* guanidine hydrochloride. Comparison of the CT properties of the proteins with the corresponding properties of model complexes suggests that the incomplete exposure is due at least in part to statistical fluctuations in the continuously mobile, randomly coiled polypeptide chain which result in residues being alternately fully exposed and partly covered. Reduction and alkylation of the disulfide cross-links increase the apparent availability of the aromatic residues, but the exposure is still less than that expected from a comparable mixture of tryptophan and tyrosine residues. Previous studies on the exposure of the aromatic residues of lysozyme and trypsin in aqueous salt solutions, when taken together with the present results, further suggest that there are two distinct kinds of surface environment possible on *native* proteins in solution. Some residues appear to be located in areas of the protein surface which are characterized by relatively fixed or stable local conformations, and have apparent CT association constants closely resembling those of comparable model complexes. Other residues may be located in a region where the protein conformation is flexible or continuously mobile, as evidenced by their smaller apparent association constants. It is probably significant that Trp-62 of lysozyme and Trp-215 of trypsin, both specificity site residues, appear to belong to the class of residues which can be considered as being in a flexible environment on the protein surface.

In the loose, more or less flexible model of a denatured protein—whose description varies from "partially unfolded" to "random coil"—it is conceivable that various folding centers remain more or less intact even in denaturing solvents. The particular state of the unfolded protein is central to the issue of how a protein acquires its unique three-dimensional configuration in the biosynthetic process, and considerable effort has been expended in search of ways and means to characterize both denatured proteins themselves and the process by which they can be refolded into their native state (cf. reviews by Tanford (1968, 1970) and by Wetlaufer and Ristow (1973)). The aromatic residues of a protein could presumably act as efficient initiators of the folding process, since they have relatively large permanent dipole moments (except for phenylalanine) for long-range mutual orientation interactions with other dipoles, and relatively large, rigid contact surfaces for stable short-range van der Waals (dispersion) interactions. Solvent perturbation studies indicate that relative to mixtures of model compounds, the aromatic residues are incompletely exposed in proteins which are considered to be completely denatured according to other criteria. The most obvious explanation of these results is that thermally induced statistical fluctuations in the completely unfolded polypeptide chain alternately cover and expose the residues and thus decrease their *apparent* availability. On the other hand, relatively unfolded regions of the protein could possibly exist even in denaturing solvents, and the solvent perturbation data could also be interpreted on the basis of nonuniform exposure of the aromatic residues. Since the molecular details of the interaction geometries necessary to produce a given effect on the spectroscopic properties of the common solvent–perturbant–chromophore complexes are virtually unknown, conclusions drawn from these studies alone must of necessity be partly subjective at the present time.

A similar method for detecting exposed tryptophan and tyrosine residues in proteins takes advantage of the fact that the very weak complexes between indole or phenol chromophores (donors) and the 1-alkyl-3-carbamidopyridinium ion (acceptor) exhibit a broad, reasonably intense charge-transfer (CT)[1] absorption well separated from the near-ultraviolet absorption bands of the chromophores themselves. The geometrical requirements for such π_D–π_A* complexes are relatively well understood: a highly overlapping face-to-face contact between the donor and acceptor rings appears to be a necessary condition for efficient light-mediated charge transfer (principle of optimum overlap and orientation; see Mulliken and Person, 1969), and thus there is an all or none relationship between complex formation and light absorption. Specific complexes between the indole moiety of tryptophan and the 1-alkyl-3-carbamidopyridinium ion have been studied crystallographically (Herriott et al., 1974; Ash et al., 1975), and insofar as the crystal structure can be taken as an indication of the probable solution conforma-

† From the Department of Biochemistry, University of Washington, Seattle, Washington 98195. *Received March 31, 1975.* We acknowledge the financial support of the National Science Foundation (Grant No. GB 18016).

[1] Abbreviations used are: CT, charge transfer; MeNCl, 1-methyl-3-carbamidopyridinium chloride (1-methylnicotinamide chloride); Gdn·HCl, guanidine hydrochloride; CH_2CO_2H, carboxymethyl; $(CH_2)_2NH_2$, aminoethyl; EDTA, ethylenediaminetetraacetic acid.

tion, the X-ray results give a picture of the CT interaction geometry which is in complete agreement with the optimum overlap and orientation hypothesis. Previous CT studies with proteins using 1-methyl-3-carbamidopyridinium chloride (MeNCl)[1] have mainly been concerned with the surface availability of tryptophan and tyrosine residues in native proteins (Deranleau et al., 1969, 1975; Bradshaw and Deranleau, 1970; Robbins and Holmes, 1972; Ikeda and Hamaguchi, 1973; Hinman et al., 1974).

In the present paper we examine the exposure of the aromatic side chains of tryptophan and tyrosine in several denatured proteins, and the effects of reduction and alkylation of the disulfide bridges of these proteins. The CT complexes have somewhat different properties in 6 *M* Gdn·HCl[1] than in the aqueous salt solutions studied previously (Hinman et al., 1974), and therefore appropriate model complexes were reexamined in 6 *M* Gdn·HCl to provide a basis for comparison with and interpretation of the protein data in the present paper.

Experimental Section

Materials. *N*-Acetyl-L-tryptophan and *N*-acetyl-L-tryptophanamide were purchased from Matheson Coleman and Bell and from Sigma Chemicals, respectively. L-Tryptophan, L-tyrosine, and *N*-acetyl-L-tyrosine were obtained from Calbiochem, and *N*-acetyl-L-tyrosinamide was a Fox Chemicals preparation. Chicken egg white lysozyme, bovine pancreatic ribonuclease A, α-chymotrypsin, chymotrypsinogen A, trypsin, and trypsinogen were all at least 2× crystallized salt-free preparations obtained from the Worthington Biochemical Corporation. 1-Methyl-3-carbamidopyridinium chloride (1-methylnicotinamide chloride, MeNCl) was synthesized by the method of Karrer et al. (1936). Extreme purity guanidine hydrochloride was purchased from Heico, Inc., and dithioerythritol and ethylenimine were obtained from the Pierce Chemical Co. All other chemicals were reagent grade preparations.

Reduced and Alkylated Proteins. The disulfide bridges of all proteins were reduced with dithioerythritol following the same general procedure. A 100-mg sample of the protein was dissolved in 5 ml of 6 *M* Gdn·HCl containing 0.13 *M* Tris-HCl and 0.1 mg/ml of EDTA at pH 7.6. The pH of the resulting solution was readjusted to 7.6 and after the addition of 68.3 mg (0.44 mmol) of dithioerythritol, the mixture was allowed to react with stirring for 2 hr under N_2.

S-Carboxymethylation of lysozyme with iodoacetic acid was carried out according to Hirs (1967). Iodoacetic acid (10 mg/ml) was added to the solution of dithioerythritol-reduced protein and the mixture was allowed to react for 20 min. A nitroprusside spot test was used to test the presence of unreacted SH groups, and upon completion of the reaction, a small amount of dithioerythritol was added to destroy excess iodoacetic acid.

S-Aminoethylation of the reduced disulfide bridges following the general procedure of Cole (1967) involved adding 0.2-ml portions of ethylenimine (4 mmol) at 10-min intervals to the reduced protein solutions. At least three such additions were made in each case, using the nitroprusside spot test to check for unreacted SH groups.

Both carboxymethylated (CH_2CO_2H) and aminoethylated [$(CH_2)_2NH_2$] proteins were dialyzed against 0.02 *M* acetic acid to remove excess reagents. CH_2CO_2H-lysozyme was collected as a precipitate, washed thoroughly, and redissolved in 6 *M* Gdn·HCl plus 0.05 *M* phosphate buffer at pH 7.0 for the CT titration studies. Solutions of $(CH_2)_2NH_2$ proteins (chymotrypsin, chymotrypsinogen, trypsin, trypsinogen) were first lyophilized and then redissolved in 6 *M* Gdn·HCl. The pH values of the solutions were adjusted to 7.0, and the solutions were then used directly in CT titration studies as discussed below.

Methods. A sequential titration procedure similar to that described previously for charge-transfer studies in aqueous solutions (Bradshaw and Deranleau, 1970) was used to investigate both amino acid and protein CT donors. Briefly, weighed portions of solid MeNCl were added sequentially to 2 ml of a 10^{-3}–10^{-4} *M* donor solution in 6 *M* Gdn·HCl at pH 7.0, or in 6 *M* Gdn·HCl plus 0.05 *M* phosphate buffer (pH 7.0). After each addition of MeNCl, the CT absorption spectrum of the mixture was recorded from 550 to 300 nm on a Cary 15 spectrophotometer. All measurements were taken at 25° in a thermostated cuvette holder. The data were generally analyzed at a single wavelength (350 nm), but other wavelengths were occasionally used as a check on internal consistency. The association constants k and extinction coefficients ϵ were estimated by fitting double intercept plots of $\bar{\epsilon}/[X]$ vs. $\bar{\epsilon}$, where $\bar{\epsilon}$ is the *statistical average* extinction coefficient of the complex (observed CT absorbance divided by the total model donor or protein concentration; Deranleau, 1975) and [X] is the concentration of uncomplexed acceptor.

When the ionic strength of the medium changes markedly as a result of increasing MeNCl concentration in a sequential titration, the association constant and extinction coefficient obtained from a *linearly fitted* double intercept plot can be in error. However, titrations carried out at a relatively high, constant ionic strength do not show anomalous effects (Hinman et al., 1974), nor could anomalous behavior be detected in the present experiments in 6 *M* Gdn·HCl even though the ionic strength varied somewhat during the titrations. The total ionic strength in 6 *M* Gdn·HCl appears to be high enough to effectively override the electrostatic charge effects previously noted for CT complexes at low ionic strengths (Hinman et al., 1974).

Results and Discussion

Model CT Complexes. The charge-transfer complexes between MeNCl and several derivatives of tryptophan and tyrosine were examined as possible models for the interaction between MeNCl and the aromatic side chains of a protein denatured in 6 *M* Gdn·HCl. Of the complexes studied, the uncharged *N*-acetyl amino acid amides are probably more representative of the side-chain environment in an unfolded polypeptide chain in 6 *M* Gdn·HCl than are the charged *N*-acetyl, amide, or zwitterionic models. The specific charge distribution characteristic of the native protein is presumably randomized in the denatured state, and the presence of a high concentration of added salt tends to minimize the effects of electrostatic interactions. The estimated oscillator strength of the CT absorption band of a particular amino acid, as well as the spectral distribution of the absorption, is largely independent of local charge effects. At the wavelength chosen for presentation of the titration data in this paper (350 nm), the differences in the extinction coefficients of the charged and uncharged model complexes do not exceed a few percent (10% maximum difference in the worst case; typically 2–4% deviation from the mean), and thus the choice of models is not crucial from the standpoint of the extinction coefficients to be used in estimating the extent of exposure of the aromatic residues in the pro-

Table I: Comparison of the Physical Properties of Amino Acid–MeNCl Complexes in 6 *M* Gdn·HCl and in Aqueous Salt at Constant Ionic Strength ([KCl] + [MeNCl] = 1 *M*).

Donor	Assoc. Constant k (cm^3/mmol)	Extinct. Coeff. ϵ (cm^2/mmol) at 350 nm	Spectral Band Area[a] $\int\epsilon(\nu)d\nu$	Estimated Oscillator Strength f
N-Acetyl-L-tryptophanamide:				
Aq salt[b]	4.01 ± 0.17	902 ± 43	8.31×10^6	0.038
6 *M* Gdn·HCl	1.67 ± 0.09	1070 ± 50	9.85×10^6	0.044
N-Acetyl-L-tyrosinamide:				
Aq salt[b]	0.95 ± 0.07	137 ± 10	1.70×10^6	0.010
6 *M* Gdn·HCl	0.26 ± 0.07	315 ± 70	3.31×10^6	0.015

[a] Integration limits: 20,000 cm^{-1} to 33,000 cm^{-1} (400 to 303 nm). [b] From Hinman et al. (1974); [MeNCl] + [KCl] = 1 *M*.

tein. On the other hand, the association constants of the model complexes depend on the net charge on the donor, increasing with increasing negative charge as might be expected (see also Hinman et al., 1974). While this effect may prove to be useful in differentiating between aromatic residues in different microenvironments on native proteins where the number of available residues is small (Ikeda and Hamaguchi, 1973), it is unlikely to be of much use in the study of denatured proteins where a large number of aromatic residues are potentially available for interaction with the acceptor.

The physical properties of the MeNCl complexes of the *N*-acetyl amino acid amides of tryptophan and tyrosine in aqueous salt and in 6 *M* Gdn·HCl are given in Table I. Compared to the results in aqueous salt, the association constants of both the tryptophan and tyrosine complexes are considerably smaller in 6 *M* Gdn·HCl. The difference is more than a factor of two for the tryptophan complex, and more than a factor of three for the tyrosine complex. The decreases correspond to changes in the apparent free energy of the interaction of from −0.8 to −0.3 kcal/mol (tryptophan complex) and from +0.3 to +0.8 kcal/mol (tyrosine complex) in passing from 1 *M* salt to 6 *M* Gdn·HCl, and are undoubtedly related to the properties of Gdn·HCl as a perturbant for disrupting hydrophobic contacts.

In contrast to the association constants, the extinction coefficients and spectral band areas (oscillator strengths) of the complexes are larger in 6 *M* Gdn·HCl than in aqueous salt solutions. The spectral distribution data are shown in Figure 1, and it can be seen that the increase in band area is relatively much larger for the tyrosine complex than for the tryptophan complex. In addition, the charge-transfer bands of both complexes appear to be red-shifted in Gdn·HCl. The shift is larger in the case of the tyrosine complex, as evidenced by the larger change in the position of the apparent maximum, and appears to be due to a shift in the position of the short-wavelength oscillator (Deranleau and Schwyzer, 1970).

The changes in the spectroscopic properties of the complexes in Gdn·HCl relative to aqueous salt solutions are not surprising in light of the well-known properties of Gdn·HCl as a hydrogen bond acceptor, and the fact that there is a considerable difference in the ease of hydrogen bond donation between the indole nitrogen or tryptophan and the phenolic oxygen of tyrosine. Similar red shifts in the normal ultraviolet (uv) absorption bands—due to specific hydrogen bonding between the heteroatoms of the indole and phenol chromophores and a hydrogen bond acceptor—have been observed for both tryptophan and tyrosine models by

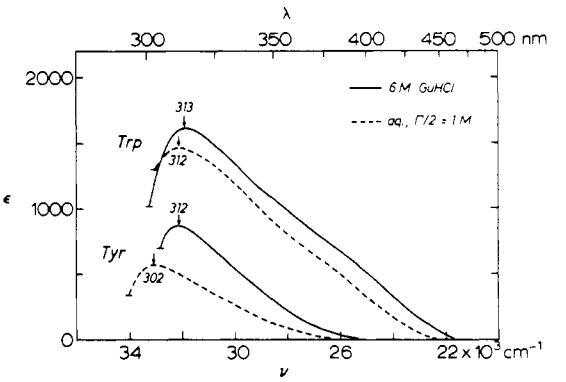

FIGURE 1: Charge-transfer absorption spectra of model amino acid–MeNCl complexes in 6 *M* Gdn·HCl (—) and in aqueous salt at constant ionic strength (- - -), [KCl] + [MeNCl] = 1 *M*): (upper curves) *N*-acetyl-L-tryptophanamide + MeNCl; (lower curves) *N*-acetyl-L-tyrosinamide + MeNCl. Peak positions are given in nanometers.

Strickland et al. (1972a,b). In the case of the tyrosine models, the H-bond red shift is accompanied by a significant increase in the oscillator strength (Strickland et al., 1972b), whereas only minor changes in the oscillator strengths of several substituted indoles could be observed in the presence of a hydrogen bond acceptor. On the basis of these observations, it appears that the present charge-transfer results can be reasonably well explained by differences in hydrogen bonding between the two solvents.

The observed changes in the spectroscopic properties of the model complexes between aqueous salt and Gdn·HCl solutions are also in keeping with expectations based on charge-transfer theory. The aromatic rings of tryptophan and tyrosine should exhibit larger ground and/or excited state dipole moments when their heteroatoms are H-bonded to the charged Gdn·H^+ molecule than when similarly bonded to water. This would tend to give rise to an increase in the first term of the theoretical expression for the transition moment of the complex (Mulliken and Person, 1969) and thus to a subsequent increase in oscillator strength, all other things being equal. Assuming that the relative orientation of the donor and the acceptor rings in the coplanar MeNCl–tryptophan complex in solution is similar to the relative orientation found in the crystal (Herriott et al., 1974; Ash et al., 1975), the dipole moments of the donor (μ_D) and the acceptor (μ_{A^+}) can be expected to lie in opposite directions. The so-called "no-bond" dipole moment of the complex (μ_0) is the vector sum of μ_D and μ_{A^+}, and an increase in μ_D due to hydrogen bonding of the type indole–NH $\cdots$ Gdn·H^+ will produce a corresponding increase in μ_0 (with little

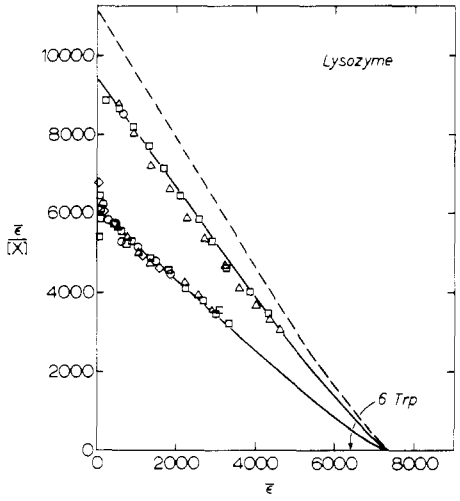

FIGURE 2: Double intercept plot for the lysozyme–MeNCl CT complex in 6 *M* Gdn·HCl at 350 nm: (lower curve) lysozyme (S–S bridges intact); (middle curve) CH_2CO_2H–lysozyme (S–S bridges broken); (upper curve) theory for a mixture of six Trp and three Tyr residues at the same concentrations as in the protein. Points are observed data, and the solid and broken lines are calculated from eq 3 using the parameters of Table II (see text). Different symbols are used to distinguish the experimental points arising from independent titrations of a given protein.

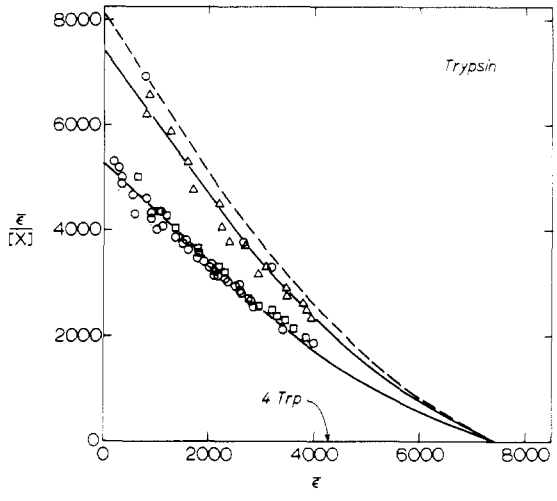

FIGURE 3: Double intercept plot for the trypsin–MeNCl complex in 6 *M* Gdn·HCl at 350 nm: (lower curve) trypsin, trypsinogen (S–S bridges intact); (middle curve) $(CH_2)_2NH_2$–trypsin (S–S bridges broken); (upper curve) theory for a mixture of four Trp and ten Tyr residues at the same concentrations as in the protein. Points and lines are as in Figure 2.

change in direction, since the Gdn·H^+ charge will be approximately in line with μ_D). Naively, if there are no other effects (changes in the dipole moment of the dative state μ_1 or in the orbital overlap S_{01}, etc.) this will result in an increase in the first term of the approximate expression for the transition moment (eq 1) because μ_0 and μ_1 are oppositely directed. The increase in oscillator strength f then follows from eq 2, where ν is the absorption frequency in wave numbers (cm^{-1}). A similar analysis holds for MeNCl–tyrosine complexes.

$$\mu_{VN} \simeq a^*b(\mu_1 - \mu_0) + \tfrac{1}{2}aa^*S_{01}\mu_1 \quad (1)$$

$$f = (8\pi^2mc/3he^2)\nu_{VN}|\mu_{VN}|^2 \simeq 4.317 \times 10^{-9} \int \epsilon(\nu)d\nu \quad (2)$$

In addition to the above considerations, the excited state energy of the donor H-bonded to Gdn·HCl should be lowered more (relative to water) than the ground-state energy is lowered—and more so for tyrosine than for tryptophan complexes—because of the increased ionic character of the donor in the dative state (the π_D–π_A^* transition corresponds to the electron-transfer reaction D·$A^+ \rightarrow D^+$–A). Under these conditions the transition energies will be smaller, leading to the observed red shifts in the spectra. In summary, the present experimental results with model tryptophan and tyrosine complexes are in good agreement with prior studies on the effects of hydrogen bonding on the spectra of isolated chromophores, and with expectations based on the application of charge-transfer theory to these complexes.

Protein CT Complexes, Preliminary Considerations. The characteristic extinction coefficients and association constants for model CT complexes can generally be obtained unambiguously from linearly fitted double intercept plots of $\bar{\epsilon}/[X]$ vs. $\bar{\epsilon}$, where $\bar{\epsilon}$ is the average extinction coefficient of the mixture (identical with the absorbance divided by the total protein concentration; Deranleau, 1975) and [X] is the concentration of unbound acceptor. With proteins, however, both tryptophan and tyrosine complexes may be observed simultaneously, and the average extinction coefficient is that due to a mixture of the different contributing chromophores. We suppose here that the aromatic residues of a denatured, randomly coiled protein can be considered as being made up of two sets, each set having identical intrinsic association constants k and extinction coefficients ϵ, and consisting of n_W tryptophyl complexes and n_Y tyrosyl complexes, respectively (phenylalanyl and histidyl complexes make negligible CT contributions by comparison; Hinman et al., 1974). The average extinction coefficient of the mixture is then given by (Deranleau, 1975; Hinman et al., 1974):

$$\bar{\epsilon} = \frac{n_W\epsilon_W k_W[X]}{1 + k_W[X]} + \frac{n_Y\epsilon_Y k_Y[X]}{1 + k_Y[X]} \quad (3)$$

Obviously the linear fitting procedures which are relevant when 1:1 model complexes of a single type of residue are being considered are inadequate for mixtures except under special circumstances. Certain assumptions are necessary in order to fit the curved double intercept plots, and we use the following.

(1) The number of exposed residues in each case is an integral number, and equal to the total number of tryptophan or tyrosine residues in the protein. Both chemical modification and solvent perturbation studies, as well as other physical studies, indicate that for the proteins considered here, the number of tryptophan residues which are at least to some degree solvent available under denaturing conditions corresponds to the total number in the protein (Shimaki et al., 1971; Villaneuva and Herskovits, 1971; Lehrer, 1971; Bello, 1970; see also the review articles by Tanford, 1968 and 1970, and by Timasheff and Gorbunoff, 1967).

(2) The extinction coefficients of the protein complexes are the same as those of the model complexes. This is an oversimplification of the actual molecular situation, but except for changes induced by high concentrations of strong H-bond acceptors like Gdn·HCl, spectral perturbations arising from differences in environment of the CT complexes are relatively small (Deranleau and Schwyzer, 1970).

(3) The association constants are considered to be vari-

Table II: Parameters Used in Curve-Fitting CT Titration Data in 6 *M* Gdn·HCl.[a]

Protein	No. of Exposed Residues W(Trp)	Y(Tyr)	App. Assoc. Constant (cm^3/mmol) k_W	k_Y
Lysozyme	6	3	0.90	0.14
CH_2CO_2H-Lysozyme	6	3	1.44	0.22
Trypsin or trypsinogen	4	10	1.1	0.17
$(CH_2)_2NH_2$-Trypsin	4	10	1.54	0.24
Chymotrypsin or chymotrypsinogen	8	4	1.0	0.15
$(CH_2)_2NH_2$-Chymotrypsin	8	4	1.32	0.20
Approximate Ratio of Protein-to-Model Association Constants (Averages)				
(a) Unmodified proteins			0.59	0.58
(b) Reduced and alkylated proteins			0.82	0.85

[a] Extinction coefficients at 350 nm are given in Table I.

able parameters. Thus, in curve fitting the data on denatured protein CT complexes according to eq 3, only k_W and k_Y are varied. The association constants obtained in this manner are *apparent* constants, reflecting both the actual site-specific association constants of the residues and their statistical availability in the randomly coiled, continuously mobile polypeptide chain.

Individual Protein CT Complexes. The results of CT titrations on several proteins are shown in Figures 2 through 4 in the form of double intercept plots. The slopes of these plots are concentration dependent because of the presence of two different sets of nonidentical sites, and are therefore not directly proportional to the association constants of the individual sites as in the case of 1:1 model complexes. Qualitatively, however, it is clear that in each of the proteins studied—lysozyme, trypsin, chymotrypsin, and the zymogens of the latter two—the apparent binding strength is smaller than would be expected for a mixture of tryptophan and tyrosine model complexes in the same proportion. Reduction and alkylation of the disulfide bridges of the proteins diminish the discrepancy but the curves are still below those expected on theoretical grounds.

An analysis of the data in terms of eq 3 and the assumptions discussed above shows that in the unmodified proteins, the apparent association constants for the two classes of sites are considerably smaller than those of the model complexes. The association constants for the curve-fitted data of Figures 2 through 4 are given in Table II, and the entire charge-transfer spectra of the complexes calculated on the basis of those constants are compared to the observed CT spectra in Figure 5. The agreement between the calculated and observed spectra implies that none of the parameters chosen in the fitting of eq 3 to the single-wavelength data of Figures 2–4 can be seriously in error.

Referring back to Table II and comparing the results on modified and unmodified proteins, it can be seen that reduction and alkylation of the disulfide bridges bring the apparent association constants more nearly in line with those of the model complexes. The effect of the disulfide bridges in restricting the random movements of the polypeptide chains in 6 *M* Gdn·HCl seems to be qualitatively similar in each of the proteins studied. The ratios of the apparent association constants of the denatured proteins to the association constants of the model compounds suggest that the ring

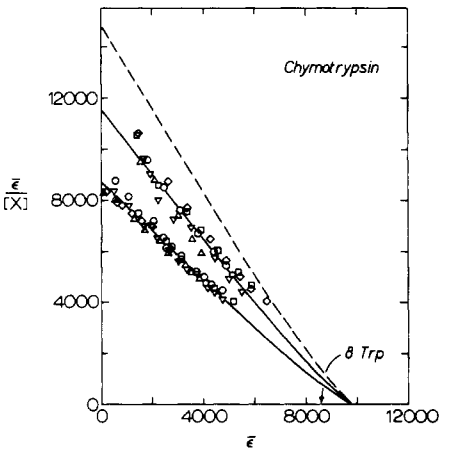

FIGURE 4: Double intercept plot for the chymotrypsin–MeNCl CT complex in 6 *M* Gdn·HCl at 350 nm: (lower curve) chymotrypsin, chymotrypsinogen (S–S bridges intact); (middle curve) chymotrypsin (S–S bridges broken); (upper curve) theory for a mixture of eight Trp and four Tyr residues at the same concentrations as in the protein. Points and lines are as in Figure 2.

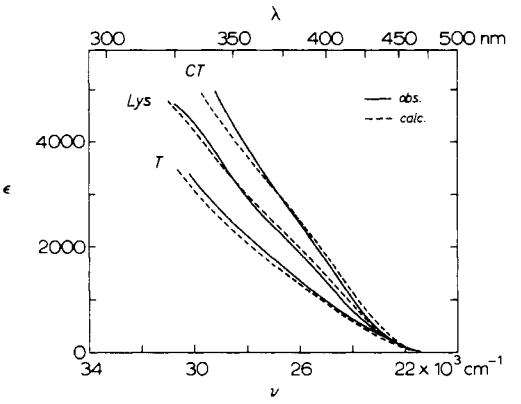

FIGURE 5: Observed and calculated CT spectra of protein–MeNCl complexes in 6 *M* Gdn·HCl: (lower curves) reduced and aminoethylated trypsin; (middle curves) reduced and carboxymethylated lysozyme; (upper curves) reduced and aminoethylated chymotrypsin. In each case the solid line is the observed spectrum, and the broken line is the spectrum calculated from eq 3 with the association constants of Table II and a list of model compound extinction coefficients as a function of wavelength. The ordinate is the *average* extinction of an arbitrary mixture, and since the data for the three protein systems were obtained at different MeNCl:protein ratios, the curves for a given protein are not comparable with those for the others in terms of total extinction (binding sites not fully saturated).

faces of the aromatic donor residues are completely uncovered about 60% of the time when the disulfide bridges are intact, and about 85% of the time when the disulfide bridges are broken.

There is also a correlation between the apparent degree of exposure of tryptophan residues and their average positions in the cross-linked polypeptide chain. The relative exposure of tryptophan side chains in proteins with intact vs. broken disulfide bonds is shown in Table III as the ratio of the apparent association constants of the MeNCl–protein complexes listed in Table II (tyrosine residues were not considered in this context because of the very much smaller sensitivity of their association constants to possible variations in structure; see Hinman et al., 1974). As can be seen from Table III, the relative exposure of tryptophan increases in the order lysozyme < trypsin < chymotrypsin, and roughly parallels the increase in the average number of

residues between tryptophan and the nearest disulfide bridge. This is the expected result, considering that the closer a residue is to a cross-link, the more restricted will be its movement and subsequent apparent exposure.

The present results are in full agreement with the results of solvent perturbation studies, and provide additional molecular detail because of the all or none nature of the CT interactions. Solvent perturbation of indole or phenol chromophores can arise either through specific changes in hydrogen bonding at the ring heteroatoms (Strickland et al., 1972a,b) or by mutual polarization interactions between the perturbant and the aromatic rings. While hydrogen bonding is an all or none phenomenon like charge transfer (either there is, or is not, a hydrogen bond or a complex with specific geometry present prior to excitation), mutual polarization interactions depend on the relative orientation and distance between the chromophore and the perturbant at the instant of excitation by light. The latter effect is thus averaged over all of the interaction geometries possible under a given degree of exposure for a particular chromophore, as well as being averaged over the different degrees of exposure possible in a collection of chromophores. The average "exposure" of a residue obtained from a particular type of measurement is thus related to both the time-averaged (statistical) availability of the residue during the course of the measurement, and to the molecule-averaged interaction geometries giving rise to specific changes in the spectroscopic or other properties of the system.

Native vs. Denatured Proteins. In the limit when a protein becomes refolded into its native conformation only those aromatic residues which remain as surface residues are potentially available for interaction with a CT probe. Evidently a certain degree of flexibility or mobility appears to be associated with some surface residues even in native proteins. The apparent association constant for Trp-62 in native lysozyme appears to be from 5 to 15% below that expected on the basis of model compounds (Hinman et al., 1974), and the apparent association constant for Trp-215 in native trypsin is between 40 and 80%—depending on pH—below the value for model complexes (Deranleau et al., 1975). By contrast, another titratable tryptophan residue in native trypsin (and in PTI-trypsin) has a pH-independent apparent association constant closely resembling the association constants of the model complexes. Taken together with the present results, these findings suggest that there are both continuously mobile and relatively fixed conformations associated with different areas of the *native* protein surface. *It is intriguing that in both lysozyme and trypsin, the tryptophan residue which appears to be in a more or less flexible environment is one which is associated with the specificity site of the enzyme.* These results tend to support the concept of intrinsic flexibility in the active-site regions of enzymes in solution previously inferred on the basis of X-ray crystal studies (Vandlen and Tulinsky, 1973; Tulinsky et al., 1973; see also Krieger et al., 1974).

Possible Folding Centers in Proteins. The present work does not answer the question as to whether the aromatic residues of a protein act as possible folding centers in proteins. Compared to a theoretical mixture of aromatic amino acids in the same proportion, a significant reduction in the average exposure of the aromatic residues is observed in 6 *M* Gdn·HCl, but it is not possible to make an unequivocal statement concerning the relative availability of the aromatic residues vs. other types of residues. However, two of the proteins considered here (lysozyme and chymotrypsin) show viscosity increases amounting to about 40% when the disulfide bonds of the denatured protein are broken (Aune, 1968; Tanford et al., 1967), whereas the CT data indicate that the apparent availability of tryptophan increases by a larger amount—about 60% for lysozyme and about 75% for chymotrypsin—under similar conditions (Tables II and III). Insofar as the comparison can be trusted, these results may be an indication that the aromatic residues in denatured proteins with intact disulfide bonds are *relatively* less exposed, on the average, than the other residues of these proteins. While this could conceivably be interpreted as implicating the aromatic residues in more structured, less random regions of the denatured protein (potential folding centers), the comparison between intrinsic viscosity and CT data may be wholly invalid and other experiments concerning the availability of nonaromatic residues would have to be devised before such a tenuous argument could be supported. In defense of the idea itself, Tulinsky et al. (1973) have suggested that the aggregation of aromatic residues within the molecular structure of *native* chymotrypsin may play an important role in imparting stability to the molecule as a whole. If this is indeed the case, it is not unreasonable to suppose that the aromatic residues may also play a key role in the folding process.

Table III: Apparent Relative Exposure of Tryptophan Residues in Proteins with Intact vs. Broken Disulfide Bridges.

Protein	Ratio of k_W, Intact/Broken S–S Bridges[a]	Av No. Residues between Trp and Nearest S–S Bridge	Av Distance between Trp and Nearest S–S Bridge (Å)[b]
Lysozyme	0.62	3.3	11
Trypsin	0.71	5.5	16
Chymotrypsin	0.76	9.0	24

[a] k_W from Table II. [b] Unperturbed root-mean-square end-to-end distance $\langle r^2\rangle_0^{1/2}$ for $\langle r^2\rangle_0/n_{p\to\infty}l_p^2 = 9$ (random coil); Brant and Flory (1974); Miller and Goebel (1968).

Part of the discrepancy between the apparent availability of Trp and Tyr in model compounds and in fully denatured proteins with broken disulfide bridges may be due to nearest-neighbor aggregation in the primary structure. Although there is no specific evidence to support this conjecture, it is suggestive that lysozyme has groupings of aromatic residues such as Tyr-20,23, Trp-62,63, and Trp-108,111, and chymotrypsin has Trp-27,29, Tyr-171, and Trp-172. Trypsin is less obvious, with Tyr-228 and Trp-234,238, but it also shows the closest correspondence with model compounds (Tables I and II). It is not unlikely that if such nearest-neighbor interactions exist at all, they would persist in the denatured protein when the disulfide bridges were reformed, and could serve as specific initiators of the overall folding process in accordance with the premise that the information for folding is contained in the primary structure. On this basis it may not be surprising to find clustering of aromatic residues in the tertiary structure of such diverse proteins as rubredoxin, lysozyme, trypsin and chymotrypsin, carbonic anhydrase, and perhaps many others. With respect to trypsin and chymotrypsin which have generally similar tertiary structures, it may be significant that about half of the aromatic residues are either the same or are replaced by another aromatic residue, that another 30% or so are replacements involving a nonpolar or neutral residue,

and that the pattern of aromatic clustering in the tertiary structure is similar in the two proteins.

The present findings confirm, *for specific residues,* the current view that fully denatured proteins with no cross-links can be closely described by the true random coil model. For denatured proteins with intact disulfide bridges, the tryptophan residue exposure appears to conform—at least qualitatively—to a cross-linked random coil model. However, it is not known whether these residues are more exposed or less exposed than would be expected on the basis of specific statistical models which take into account the size of the protein, the number of cross-links, and the lengths of the polypeptide loops.

Acknowledgment

We gratefully acknowledge the suggestions made by one of the referees.

References

Ash, R. P., Herriott, J. R., and Deranleau, D. A. (1975), manuscript in preparation.

Aune, K. C. (1968), Ph.D. Thesis, Duke University, Durham, N.C.

Bello, J. (1970), *Biochemistry 9,* 3562–3568.

Bradshaw, R. A., and Deranleau, D. A. (1970), *Biochemistry 9,* 3310–3315.

Brant, D. A., and Flory, P. J. (1965), *J. Am. Chem. Soc. 87,* 2791–2800.

Cilento, G., and Tedeschi, P. (1961), *J. Biol. Chem. 236,* 907–910.

Cole, R. D. (1967), *Methods Enzymol. 11,* 315–317.

Deranleau, D. A. (1975), *J. Am. Chem. Soc. 97,* 1218–1224.

Deranleau, D. A., Bradshaw, R. A., and Schwyzer, R. (1969), *Proc. Natl. Acad. Sci. U.S.A. 63,* 885–889.

Deranleau, D. A., Hinman, L. M., and Coan, C. R. (1975), *J. Mol. Biol. 94,* 567–582.

Deranleau, D. A., and Schwyzer, R. (1970), *Biochemistry 9,* 126–134.

Herriott, J. R., Camerman, A., and Deranleau, D. A. (1974), *J. Am. Chem. Soc. 96,* 1585–1589.

Hinman, L. M., Coan, C. R., and Deranleau, D. A. (1974), *J. Am. Chem. Soc. 96,* 7067–7073.

Hirs, C. H. W. (1967), *Methods Enzymol. 11,* 199–203.

Ikeda, K., and Hamaguchi, K. (1973), *J. Biochem. (Tokyo) 74,* 221–230.

Karrer, P., Schwarzenbach, G., Benz, F., and Solmssen, U. (1936), *Helv. Chim. Acta 19,* 811.

Krieger, M., Kay, L. M., and Stroud, R. (1974), *J. Mol. Biol. 83,* 209–230.

Lehrer, S. S. (1971), *Biochemistry 10,* 3254–3263.

Miller, W. G., and Goebel, C. V. (1968), *Biochemistry 7,* 3925–3935.

Mulliken, R. S., and Person, W. B. (1969), Molecular Complexes, New York, N.Y., Wiley-Interscience.

Robbins, F. M., and Holmes, L. G. (1972), *J. Biol. Chem. 247,* 3062–3065.

Shimaki, N., Ikeda, K., and Hamaguchi, K. (1971), *J. Biochem. (Tokyo) 70,* 497–508.

Strickland, E. H., and Billups, C. (1973), *Biopolymers 12,* 1989–1995.

Strickland, E. H., Billups, C., and Kay, E. (1972a), *Biochemistry 11,* 3657–3662.

Strickland, E. H., Wilchek, M., Horowitz, J., and Billups, C. (1972b), *J. Biol. Chem. 247,* 572–580.

Tanford, C. R. (1968), *Adv. Protein Chem. 23,* 121–282.

Tanford, C. R. (1970), *Adv. Protein Chem. 24,* 1–95.

Tanford, C. R., Kawahara, K., and Lapanje, S. (1967), *J. Am. Chem. Soc. 89,* 729–736.

Timasheff, S. N., and Gorbunoff, M. J. (1967), *Annu. Rev. Biochem. 36,* 13–54.

Tulinsky, A., Vandlen, R. L., Morimoto, C. N., Mani, N. V., and Wright, L. H. (1973), *Biochemistry 12,* 4185–4192.

Vandlen, R. L., and Tulinsky, A. (1973), *Biochemistry 12,* 4193–4200.

Villaneuva, G. B., and Herskovits, T. T. (1971), *Biochemistry 10,* 3358–3365.

Wetlaufer, D. B., and Ristow, S. (1973), *Annu. Rev. Biochem. 42,* 135–158.