J. Biochem. **135**, 405–411 (2004)
DOI: 10.1093/jb/mvh048

Secondary-Structure Analysis of Proteins by Vacuum-Ultraviolet Circular Dichroism Spectroscopy

Koichi Matsuo[1], Ryuta Yonehara[1] and Kunihiko Gekko[1,2,*]

[1]Department of Mathematical and Life Sciences, Graduate School of Science, Hiroshima University, Higashi-Hiroshima 739-8526; and [2]Hiroshima Synchrotron Radiation Center, Hiroshima University, Higashi-Hiroshima 739-8526

Received November 19, 2003; accepted January 20, 2004

The vacuum ultraviolet circular dichroism (VUVCD) spectra of 15 globular proteins (myoglobin, hemoglobin, human serum albumin, cytochrome *c*, peroxidase, α-lactalbumin, lysozyme, ovalbumin, ribonuclease A, β-lactoglobulin, pepsin, trypsinogen, α-chymotrypsinogen, soybean trypsin inhibitor, and concanavalin A) were measured in aqueous solutions at 25°C in the wavelength region from 260 to 160 nm under a high vacuum, using a synchrotron-radiation VUVCD spectrophotometer. The VUVCD spectra below 190 nm revealed some characteristic bands corresponding to different secondary structures. The contents of α-helices, β-strands, turns, and unordered structures were estimated using the SELCON3 program with VUVCD spectra data on the 15 proteins. Prediction of the secondary-structure contents was greatly improved by extending the circular dichroism spectra to 165 nm. The numbers of α-helix and β-strand segments calculated from the distorted α-helix and β-strand contents did not differ greatly from those obtained from X-ray crystal structures. These results demonstrate that synchrotron-radiation VUVCD spectroscopy is a powerful tool for analyzing the secondary structures of proteins.

Key words: proteins, secondary-structure analysis, synchrotron-radiation, vacuum ultraviolet circular dichroism.

Abbreviations: CD, circular dichroism; HSA, human serum albumin; RNase A, ribonuclease A; STI, soybean trypsin inhibitor; VUVCD, vacuum ultraviolet circular dichroism.

Circular dichroism (CD) is very sensitive as to the backbone conformations of proteins, and hence various analytical methods have been developed for quantitative estimation of secondary structure contents using CD spectra (*1–5*). Early methods made use of the CD spectra of model polypeptides with specific secondary structures as "pure component spectra," the secondary-structure contents being estimated by least-squares fitting of the protein CD spectra. The model polypeptide spectra were replaced by the "pure component spectra" obtained from a set of proteins with known X-ray structures. Hennessey and Johnson estimated secondary-structure contents using reference CD spectra data on 15 proteins and a helical polypeptide in the wavelength region from 178 to 250 nm (*3*). They found that CD spectra down to 178 nm in comparison with those down to 190 nm only slightly affected the prediction of the α-helix content but significantly improved that of other structures. Toumadje *et al.* also suggested that extending CD spectra to 168 nm rather than stopping at 178 nm improved the prediction of the secondary-structure contents (*4*). Recently, Sreerama and Woody estimated secondary-structure contents using three programs—CONTIN, SELCON3, and CDSSTR—with five sets of reference spectra of 29, 37, 42, 43, and 48 proteins extending down to 178 nm (*5*). They found that a larger reference data set improves the prediction and recommended improvement of the CD data at shorter wavelengths. The accumulation of CD data on proteins in the vacuum ultraviolet region is required for significant improvements not only for predicting secondary-structure contents, but also for estimating the numbers of secondary-structure segments (*6*) and assignment of the tertiary-structure classes (*7*, *8*).

Since the 1980s, several groups have put considerable effort into the construction of vacuum ultraviolet circular dichroism (VUVCD) spectrophotometers using synchrotron radiation as an intense light source in order to extend the short-wavelength limit (*9–13*). However, these spectrophotometers only operate under a nitrogen gas atmosphere, and hence the short-wavelength limit is currently about 170 nm for aqueous solutions. We recently constructed a VUVCD spectrophotometer at the Hiroshima Synchrotron Radiation Center (HSRC) that is capable of measuring CD spectra down to 140 nm in aqueous solutions by keeping all the optical devices under a high vacuum (*14–17*). In this paper, we report the VUVCD spectra of 15 proteins measured down to 160 nm using this spectrophotometer and the results of secondary-structure analysis using the SELCON3 program. To our knowledge, this is the first systematic VUVCD analysis of protein secondary structures in aqueous solutions using a synchrotron-radiation spectrophotometer.

*To whom correspondence should be addressed. E-mail: gekko@sci.hiroshima-u.ac.jp

Table 1. **Origins and structural parameters of the proteins studied.**

Protein	Origin	PDB code	Secondary Structure (%)[a]			$OD_{280}^{1\%}$	pH
			α-Helix	β-Strand	Turn		
Myoglobin	Horse heart	1WLA	75.8	0.0	12.4	17.9	5.4
Hemoglobin	Bovine blood	1G08	75.0	0.0	14.0	106.9[b]	4.9
HSA	Human serum	1AO6	69.9	0.0	14.9	5.3	5.0
Cytochrome *c*	Horse heart	1HRC	41.0	0.0	21.9	18.4	7.1
Peroxidase	Horse radish	1ATJ	50.0	2.0	25.6	7.2	5.3
α-Lactalbumin	Bovine milk	1F6S	45.0	7.1	23.0	20.1[c]	7.3
Lysozyme	Hen egg white	1HEL	41.9	6.2	30.6	26.4	5.0
Ovalbumin	Hen egg white	1OVA	30.8	31.3	16.4	7.1	4.8
RNase A	Bovine pancreas	1FS3	21.0	33.0	21.8	7.1	6.1
β-Lactoglobulin	Bovine milk	1B8E	16.7	41.0	21.6	9.6	6.6[e]
Pepsin	Porcine stomach mucosa	4PEP	15.3	41.7	20.0	14.7	6.5[e]
Trypsinogen	Bovine pancreas	1TGN	10.1	32.3	25.3	13.9	6.6
α-Chymotrypsinogen	Bovine pancreas	2CGA	13.5	32.0	21.0	19.7[d]	5.6
STI	Soybean	1AVU	1.7	37.0	17.1	9.5	6.2[e]
Concanavalin A	Jack bean	2CTV	3.8	46.4	23.6	13.7	6.7[e]

[a]From crystal data listed in PDB code. [b]At 405 nm. [c]At 281.5 nm. [d]At 282 nm. [e]10 mM phosphate buffer.

MATERIALS AND METHODS

Materials—Peroxidase and ovalbumin were purchased from Tokyo Kasei Kogyo, and soybean trypsin inhibitor (STI) and lysozyme were obtained from Worthington and Seikagaku, respectively. The other proteins [myoglobin, hemoglobin, human serum albumin (HSA), cytochrome *c*, concanavalin A, β-lactoglobulin, pepsin, α-chymotrypsinogen, trypsinogen, α-lactalbumin, and ribonuclease A (RNase A)] were purchased from Sigma. The origins of these proteins are listed in Table 1 with their structural parameters. These proteins were used without further purification. The water-soluble proteins (myoglobin, hemoglobin, HSA, peroxidase, cytochrome *c*, α-chymotrypsinogen, trypsinogen, α-lactalbumin, lysozyme, RNase A, and ovalbumin) were dialyzed against double-distilled water at 4°C. Since β-lactoglobulin, pepsin, STI, and concanavalin A are insoluble in pure water, they were dissolved in 10 mM potassium phosphate buffer (pH 6.6) and then exhaustively dialyzed against the same buffer. The dialyzed protein solutions were centrifuged for 15 min at 14,000 rpm to remove the aggregates and then adjusted to protein concentrations of 0.2–1.0% by dilution or concentration with a Centricut mini (U-20, Kurabo Industries). Protein concentrations were determined by absorption measurement (V-560, Jasco) using the molar extinction coefficient values in the literature, which are listed in Table 1.

VUVCD Measurements—The VUVCD spectra of proteins were measured in the wavelength region from 260 to 160 nm under a high vacuum (10^{-4} Pa) at 25°C, using the VUVCD spectrophotometer constructed at the HSRC (BL15). The details of the optical devices comprising of the spectrophotometer are given in previous papers (*14–17*). The performance of the VUVCD spectrophotometer was confirmed by monitoring the CD spectrum of an aqueous solution of ammonium *d*-camphor-10-sulfonate, which exhibits positive and negative peaks at 291 and 192 nm in an intensity ratio of 1 to 2, respectively (*17*). All of the spectra were recorded with a 1.0-mm slit, a 16-s time constant, a 4-nm/min scan speed, and 4–16 accumulations. A commercial spectrophotometer (J-720W, Jasco) was also used for comparison of the CD spectra in the far-UV region from 260 to 200 nm.

The VUVCD measurements of each protein were performed using an assembled type optical cell with MgF_2 windows that can tolerate a high vacuum (10^{-4} Pa) and that is attached to the temperature-control unit (*17*). The path length of the cell was adjusted with a Teflon spacer to 50 μm for measurements from 260 to 175 nm with protein concentrations of 0.2–0.5%. To reduce the effect of light absorption by water, no spacer was used for measurements below 175 nm with protein concentrations of 0.5–1.0%. The spectra obtained without the spacer (1.7-μm path length) were calibrated by normalizing the ellipticity as to the spectra measured using a 50-μm spacer in the overlapping wavelength region from 260 to 175 nm. The ellipticity was reproducible within an error of ±5%, which was mainly attributable to signal noise and inaccuracy in the light path length.

Secondary-Structure Analysis of Proteins—The secondary structures of 15 proteins in crystals were assigned using the DSSP program (*18*) for the PDB codes listed in Table 1. The 3_{10}-helix was grouped as a α-helix, the bends were treated as turns, and the single residues assigned as turns and bends were grouped as unordered. Moreover, α-helices and β-strands were split into regular (α_R and β_R) and distorted (α_D and β_D) classes, assuming that four residues per α-helix and two residues per β-strand were distorted (*6*). Thus the protein structure was grouped into six types: regular α-helix (α_R), distorted α-helix (α_D), regular β-strand (β_R), distorted β-strand (β_D), turn, and unordered structure.

The secondary-structure contents in solution were estimated from the VUVCD spectra of the 15 proteins, using the SELCON3 program developed by Sreerama and Woody (*2*). This program was modified to extend it down to 160 nm. Its performance in the analysis of secondary structures was characterized by the root-mean-square deviation (δ) and the Pearson correlation coefficient (*r*) between X-ray and CD estimates of the secondary-structure contents. For each of the secondary structures, the values of δ and *r* were calculated using the equations:

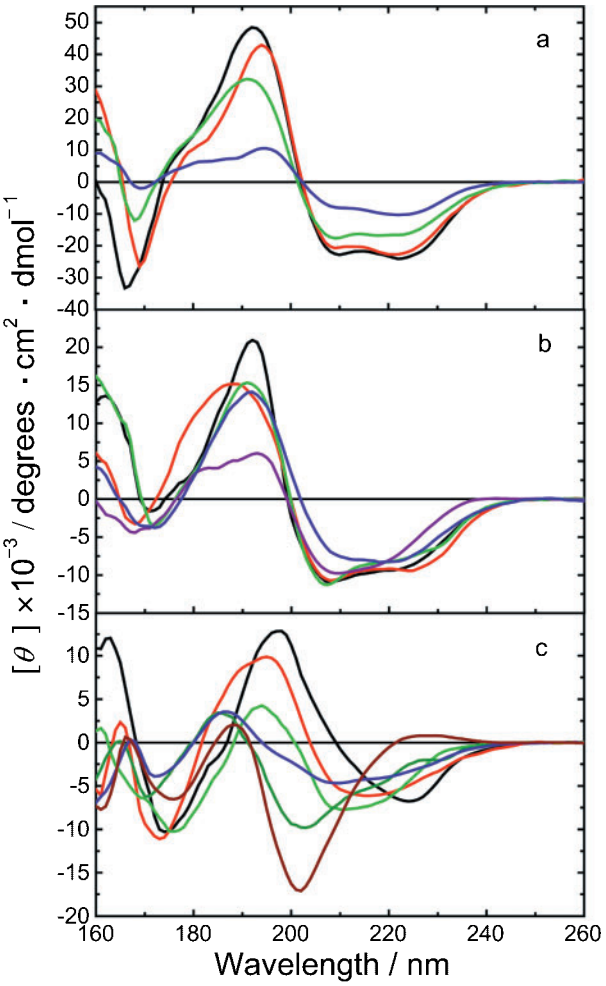

Fig. 1. **VUVCD spectra for 15 proteins in aqueous solutions at 25°C.** (a) Myoglobin (black), hemoglobin (red), HSA (green), and cytochrome *c* (blue); (b) peroxidase (black), α-lactalbumin (red), lysozyme (green), ovalbumin (blue), and RNase A (purple); (c) concanavalin A (black), β-lactoglobulin (red), pepsin (green), trypsinogen (blue), α-chymotrypsinogen (dark green), and STI (brown). A 50-μm path-length cell was used for the measurements from 260 to 175 nm, and no spacer was used below 175 nm. All spectra were recorded with a 1.0-mm slit, a 16-s time constant, a 4-nm/min scan speed, and 4–16 accumulations.

$$\delta = \sqrt{\sum (X_i - Y_i)^2 / N}$$

$$r = \{\sum X_i Y_i - (\sum X_i \sum Y_i)/N\} / \left(\sqrt{\sum X_i^2 - (\sum X_i)^2/N}\sqrt{\sum Y_i^2 - (\sum Y_i)^2/N}\right)$$

where X and Y are the X-ray and CD estimates of a given type of secondary structure in N reference samples, respectively. The overall performance of the analysis was determined by considering all the secondary-structure contents collectively.

RESULTS AND DISCUSSION

VUVCD Spectra of Proteins—Figure 1 shows the VUVCD spectra of 15 proteins in aqueous solution. These spectra are superimposed on those down to 190 nm obtained using a commercial spectropolarimeter, indicating the good performance of the VUVCD spectrophotometer and the optical cell constructed. There was no decay of the spectra intensity during the accumulation (about 3 h), indicating that no protein was decomposed by synchrotron radiation.

Figure 1a shows the VUVCD spectra of four α-proteins (myoglobin, hemoglobin, HSA, and cytochrome *c*) that include only α-helices as the secondary structure. There are two negative peaks at around 222 and 208 nm, and a positive peak at around 190 nm, which are characteristic of α-helices. The negative peak at 222 nm is assigned to the n–π* transition of the carbonyl group of the peptide, and the negative and positive peaks at 208 and 190 nm, respectively, are assigned to the parallel and perpendicular excitons of the π–π* transition of the peptide (*1*, *19*). Evidently, the intensity of these spectra in the far-UV region from 260 to 190 nm increased with increasing α-helix content (Table 1): myoglobin (75.8%) > hemoglobin (75.0%) > HSA (69.9%) > cytochrome *c* (41.0%). Although the α-helix content differs only slightly between myoglobin and hemoglobin, there is a characteristic difference in the intensity and peak wavelength of the spectra below 200 nm. Extension of the VUVCD measurement to shorter wavelengths revealed a negative peak centered at around 170 nm and a shoulder at around 175 nm. A positive peak also appeared to be present below 160 nm. The CD peaks in the region from 150 to 175 nm probably arise from inter-peptide charge transfer (*20*).

Figure 1b shows the VUVCD spectra of three α-helix-rich proteins (peroxidase, α-lactalbumin, and lysozyme), and two proteins containing comparable amounts of α-helix and β-strand (RNase A and ovalbumin). The former three proteins exhibit similar spectra to α-proteins, although the molar ellipticity of the respective bands is smaller according to their α-helix contents (50%, 45%, and 41.9%, respectively). A negative peak centered at around 170 nm and a shoulder at around 175 nm were also observed, but they were small compared to those of α-proteins. The spectrum of ovalbumin retains the features of those of α-proteins, but RNase A, which contains less α-helix than ovalbumin (21% and 30.8%, respectively), exhibits a considerably different spectrum with a positive peak at around 180 nm. Thus the VUVCD spectra of proteins are very sensitive as to the α-helix content.

Figure 1c shows the VUVCD spectra of four β-strand-rich proteins (β-lactoglobulin, pepsin, α-chymotrypsinogen, and trypsinogen) and two β-proteins comprised almost entirely of β-strands (concanavalin A and STI). These proteins exhibit three negative peaks at around 200–230, 170–175, and 160 nm, and two positive peaks at around 185–200 and 165 nm, all of which are very sensitive as to the secondary-structure content. Pepsin and β-lactoglobulin exhibit considerably different molar ellipticity in the wavelength region from 190 to 210, and peak shifts at around 175 and 165 nm, although their α-helix and β-strand contents are similar (about 16% and 41%, respectively). Trypsinogen and α-chymotrypsinogen exhibit large variations in their spectra at around 200 nm although they have the same β-strand content (32%). Concanavalin A (comprising 46.4% β-strands) exhibits negative and positive peaks at around 220 and 195 nm,

Table 2. **Performance indices (δ and *r*) of four types of secondary structure determined from CD spectra in different wavelength regions.**

Wavelength	α-Helix		β-Strand		Turn		Unordered		Total	
	δ	*r*	δ	*r*	δ	*r*	δ	*r*	δ	*r*
260–185 nm	0.087	0.942	0.132	0.688	0.049	0.396	0.055	0.789	0.087	0.857
260–180 nm	0.074	0.958	0.109	0.795	0.047	0.365	0.068	0.667	0.078	0.887
260–175 nm	0.086	0.938	0.113	0.774	0.048	0.413	0.067	0.662	0.082	0.873
260–170 nm	0.087	0.937	0.108	0.799	0.048	0.428	0.068	0.648	0.081	0.878
260–165 nm	0.068	0.960	0.083	0.885	0.042	0.651	0.066	0.682	0.066	0.919
260–160 nm	0.082	0.943	0.123	0.727	0.055	0.601	0.083	0.607	0.089	0.851

respectively, that are typical of β-strands (*21*), and also negative and positive peaks at around 175 and 165 nm, respectively. The negative band at 220 nm is assigned to the n–π* transition of the peptide carbonyl group, as observed for α-helices, and the positive and negative bands at 195 and 175 nm, respectively, should arise from the exciton splitting of the lowest peptide π–π* transition (*19*). Interestingly, STI (comprising 37% β-strands) exhibits three small positive peaks at around 220, 190, and 175 nm, and three negative peaks at 200, 175, and 160 nm, which is very different from the spectra of concanavalin A and other β-strand-rich proteins. The spectrum of STI closely resembles that of poly-L-proline type-II (PP-II), which is a left-handed three-fold helix (*22*). It is known that short stretches of the PP-II helical conformation exist in unordered polypeptides such as poly-L-lysine and poly-L-glutamic acid at pH 7 (*23*, *24*). These results suggest that the spectrum of STI is largely attributable to the PP-II conformation in unordered structures. Recently, Sreerama and Woody proposed that β-strand-rich proteins exhibit two CD spectra reminiscent of either model β-sheets [β(I)] or PP-II [β(II)], and that STI and concanavalin A can be classified into the β(II) and β(I) types, respectively (*25*). Thus, β-proteins and β-strand-rich proteins exhibit variations in VUVCD spectra that are larger than those exhibited by α-proteins and α-helix-rich proteins.

Secondary-Structure Analysis with VUVCD Spectra—As shown above, examining the VUVCD spectra of proteins down to 160 nm can provide new and detailed information on secondary structures that cannot be obtained from far-UV CD spectra down to 190 nm. It is of interest to determine how secondary structure analysis can be improved by extending the VUVCD spectra to shorter wavelengths. To examine this, the root-mean-square deviation (δ) and Pearson correlation coefficient (*r*) between X-ray and CD estimates of α-helices, β-strands, turns, and unordered structures were calculated using the SELCON3 program (*2*, *5*) with VUVCD spectra of 15 proteins in the wavelength region from 260 nm to various short wavelength limits. Table 2 lists the results of this calculation. As expected (*26*), δ decreases and *r* increases for each secondary structure as the lower wavelength limit decreases, indicating improved estimation of the secondary-structure contents. Toumadje *et al.* (*4*) reported that the extension of CD measurement of 15 proteins and an α-helical polypeptide from 178 to 168 nm improved the prediction of α-helices and turns, while the prediction is conversely worse for β-strands and unordered structure. However, extension from 170 to 165 nm greatly improved the prediction for all of α-helices, β-strands, turns and unordered. This is confirmed by the overall performance indices listed in the last column of Table 2, which were obtained by considering the contributions of the four structural contents collectively. In fact, the prediction for most proteins is improved by extension to 165 nm although that for only half of the proteins is improved by extension to 168 nm (*4*). These results indicate that characteristic CD peaks in the VUV region between 170 and 165 nm significantly contribute to prediction of each secondary structure content, though the δ and *r* values indicate that further extension to 160 nm does not necessarily lead to a better performance.

The validity of VUVCD spectra based on SELCON3 analysis was further examined with six types of secondary structures, which were obtained by splitting α-helices and β-strands into regular (α_R and β_R) and distorted classes (α_D and β_D), in addition to turns and unordered structures. Table 3 lists the δ and *r* values for each structural component, and the overall performance in various wavelength regions. As shown by the overall performance indices in the last column of the table, the accuracy of secondary-structure prediction also tends to improve when the lower wavelength limit is decreased: the prediction gradually improved for α_R and turns, and greatly improved for β_R and β_D, but such improvement was not

Table 3. **Performance indices (δ and *r*) of six types of secondary structure determined from CD spectra in different wavelength regions.**[a]

Wavelength	α_R		α_D		β_R		β_D		Turn		Unordered		Total	
	δ	*r*	δ	*r*	δ	*r*	δ	*r*	δ	*r*	δ	*r*	δ	*r*
260–185 nm	0.069	0.930	0.055	0.708	0.085	0.733	0.045	0.696	0.047	0.442	0.053	0.803	0.061	0.788
260–180 nm	0.056	0.953	0.058	0.658	0.076	0.792	0.037	0.799	0.050	0.323	0.068	0.677	0.059	0.816
260–175 nm	0.055	0.955	0.067	0.520	0.076	0.785	0.035	0.818	0.047	0.412	0.066	0.679	0.059	0.825
260–170 nm	0.059	0.948	0.060	0.645	0.070	0.829	0.034	0.836	0.046	0.468	0.066	0.674	0.057	0.825
260–165 nm	0.053	0.960	0.059	0.674	0.061	0.868	0.027	0.896	0.041	0.650	0.064	0.698	0.052	0.849
260–160 nm	0.062	0.944	0.076	0.536	0.088	0.692	0.034	0.836	0.052	0.602	0.078	0.628	0.067	0.754

[a]α_R, regular α-helix; α_D, distorted α-helix; β_R, regular β-strand; β_D, distorted β-strand.

Table 4. **Secondary-structure contents of 15 reference proteins determined by X-ray analysis, and CD spectra in the wavelength region from 260 to 165 nm.**

Protein		α_R	α_D	β_R	β_D	Turn	Unordered	Total	δ
Myoglobin	X-ray	54.9	20.9	0.0	0.0	12.4	11.8	100	0.035
	CD	60.2	20.5	1.5	–0.7	6.5	8.8	96.8	
Hemoglobin	X-ray	54.0	21.0	0.0	0.0	14.0	11.0	100	0.021
	CD	50.5	20.6	–1.4	–0.8	14.3	14.7	97.9	
HSA	X-ray	49.1	20.8	0.0	0.0	14.9	15.2	100	0.036
	CD	42.3	25.3	1.4	0.6	17.4	14.4	101	
Cytochrome *c*	X-ray	21.9	19.1	0.0	0.0	21.9	37.2	100	0.092
	CD	14.7	18.7	12.0	6.2	26.4	21.4	99.4	
Peroxidase	X-ray	29.2	20.8	0.7	1.3	25.6	22.4	100	0.030
	CD	23.4	20.5	3.4	4.8	25.8	22.1	100	
α-Lactalbumin	X-ray	19.5	24.4	1.6	4.9	23.6	26.0	100	0.069
	CD	26.8	16.4	5.4	3.0	21.1	27.4	100	
Lysozyme	X-ray	20.2	21.7	1.5	4.7	30.6	21.3	100	0.042
	CD	24.7	17.5	6.8	4.1	25.0	24.1	102	
Ovalbumin	X-ray	17.9	12.9	23.0	8.3	16.4	21.5	100	0.056
	CD	15.1	18.1	12.7	6.9	22.9	23.2	98.9	
RNase A	X-ray	11.3	9.7	21.7	11.3	21.8	24.2	100	0.044
	CD	12.2	14.7	14.1	8.7	20.6	28.9	99.2	
β-Lactoglobulin	X-ray	5.6	11.1	28.7	12.3	21.6	20.7	100	0.040
	CD	4.3	5.1	35.7	15.6	20.9	20.4	102	
Pepsin	X-ray	3.0	12.3	26.4	15.3	20.0	23.0	100	0.056
	CD	1.6	1.8	31.6	16.5	20.2	29.9	102	
Trypsinogen	X-ray	5.3	4.8	20.9	11.4	25.3	32.3	100	0.033
	CD	1.9	10.1	21.9	11.7	20.6	30.6	96.8	
α-Chymotrypsinogen	X-ray	5.1	8.4	20.0	12.0	21.0	33.5	100	0.015
	CD	3.5	6.9	19.0	13.7	20.0	35.5	98.6	
STI	X-ray	0.0	1.7	19.3	17.7	17.1	44.2	100	0.080
	CD	1.8	8.3	19.7	15.2	25.9	28.4	99.3	
Concanavalin A	X-ray	0.0	3.8	32.9	13.5	23.6	26.2	100	0.080
	CD	8.6	11.0	22.1	8.6	21.4	24.0	95.9	

observed for α_D and unordered structures. The best performance was obtained by extending the lower wavelength limit down to 165 nm, but extension to 160 nm reduced the prediction accuracy, as found above for the four types of secondary structure (Table 2). Such decreased performance at 160 nm is mainly attributable to the low signal-to-noise ratio of VUVCD spectra due to strong absorption of water. This problem could be improved by extending the CD measurement to shorter wavelengths below 160 nm (*e.g.*, 140 nm) and increasing the accumulation time for the VUVCD measurements, because our VUVCD spectrophotometer establishes accurate control of a photoelastic modulator and a lock-in amplifier with high stability under a high vacuum with an optical servo-control system in a double-beam configuration.

Comparison Between X-Ray and VUVCD Estimations of Secondary-Structure Contents—Table 4 shows the secondary-structure contents of the 15 proteins that were determined by SELCON3 analysis with the VUVCD spectra down to 165 nm and by assignment of the X-ray crystal structures using the DSSP program. In this table, the small negative values of the β_R and β_D contents for myoglobin and hemoglobin can be regarded as zero. The overall root-mean-square deviations (δ) between the X-ray and VUVCD estimates are listed in the last column of the table. Myoglobin, hemoglobin, HSA, peroxidase, α-chymotrypsinogen, and trypsinogen show small δ values, indicating good performance as to VUVCD estimation for these proteins. In contrast, the prediction is not good for cytochrome *c*, concanavalin A, STI, and α-lactalbumin. These results suggest that VUVCD analysis is more suitable for α-proteins than for β-proteins, although the δ values of β-proteins also decrease as the lower wavelength limit decreases. The exceptionally large δ value for cytochrome *c*, which is classified as a α-protein, may be caused by a covalently bonded heme group. The large δ values for β-proteins are evidently caused by the large variations in the VUVCD spectra evident in Fig. 1c. As suggested by Sreerama and Woody (*5*), the prediction would be improved by using a larger reference set of CD spectra and considering the contribution of the PP-II conformation. Improving the VUVCD estimation of the secondary-structure contents requires the acquisition of further VUVCD data on proteins, especially β-strand-rich proteins, as well as further extension to shorter wavelengths.

Estimation of the Numbers of α-Helix and β-Strand Segments—It is of interest whether or not the numbers of α-helix and β-strand segments can be determined from VUVCD spectra, since they are difficult to estimate from far-UV CD spectra down to 190 nm. The numbers of α-helix and β-strand segments have been estimated using two approaches. Pancoska *et al.* used a matrix descriptor

Table 5. **Numbers of α-helix and β-strand segments determined by X-ray analysis, and CD spectra in the wavelength region from 260 to 165 nm.**

Protein	α-Helix		β-Strand	
	X-ray	CD	X-ray	CD
Myoglobin	8	8.0	0	0.0
Hemoglobin	8	7.5	0	0.0
HSA	31	37.0	0	2.0
Cytochrome *c*	5	5.0	0	3.0
Peroxidase	17	16.0	2	7.0
α-Lactalbumin	9	5.0	3	2.0
Lysozyme	7	6.0	3	3.0
Ovalbumin	13	17.5	16	13.0
RNase A	3	4.5	7	5.5
β-Lactoglobulin	5	2.0	10	12.5
Pepsin	11	1.5	25	26.0
Trypsinogen	3	5.5	13	13.0
α-Chymotrypsinogen	6	4.5	14	17.0
STI	1	3.5	15	14.0
Concanavalin A	3	6.5	16	10.5

of secondary-structure segments for neutral-network-based analysis of protein CD spectra (*27*). Sreerama *et al.* estimated the numbers of α-helix and β-strand segments from the distorted residues in α-helices and β-strands, assuming the average number of distorted residues to be four per α-helix and two per β-strand (*6*). Since the results of these two analyses are comparable, in the present study the numbers of α-helix and β-strand segments for the 15 proteins were estimated by the method of Sreerama *et al*. using the α_D and β_D fractions obtained from the VUVCD spectra in the wavelength region from 260 to 165 nm (Table 3). The results of the calculation are listed in Table 5 in comparison with those obtained for the X-ray structure. The numbers of the two segments estimated by these methods are comparable, although large deviation was observed for α-helices in pepsin and for β-strands in peroxidase. The root-mean-square differences between the VUVCD and X-ray estimations were calculated to be 3.6 and 2.5 for α-helices and β-strands, respectively. These values are not smaller than those (3.2 and 2.5) obtained from CD spectra down to 178 nm by Sreerama *et al*. (*6*). However, this does not necessarily mean that our VUVCD data do not improve the estimation of the numbers of secondary-structure segments, because Sreerama *et al*. used twice as many reference proteins as we did. It is likely that the prediction of the numbers of the segments would be greatly improved by increasing the number of reference proteins, and by using more accurate numbers of distorted residues per α-helix and β-strand.

CONCLUDING REMARKS

The present paper describes the first successful measurement of the VUVCD spectra of 15 proteins down to 160 nm in aqueous solutions using a synchrotron-radiation spectrophotometer. These VUVCD spectra improved the estimation of the secondary-structure contents using the SELCON3 program, and allowed us to determine the numbers of α-helix and β-strand segments. Further extension to shorter wavelengths and accumulation of VUVCD data in progress using synchrotron radiation should lead to new methods of protein structure analysis that are superior to the use of existing CD spectrophotometers.

This work was financially supported by the HSRC, the Japan Atomic Energy Institute, and a Grant-in-Aid for Scientific Research from the Ministry of Education, Science, Sports, and Culture of Japan (No. 12559007).

REFERENCES

1. Woody, R.W. (1995) Circular dichroism. *Methods Enzymol.* **246**, 34–71
2. Sreerama, N. and Woody, R.W. (1993) A self-consistent method for the analysis of protein secondary structure from circular dichroism. *Anal. Biochem.* **209**, 32–44
3. Hennessey, J.P. Jr. and Johnson, W.C., Jr. (1981) Information content in the circular dichroism of proteins. *Biochemistry* **20**, 1085–1094
4. Toumadje, A., Alcorn, S.W., and Johnson, W.C., Jr. (1992) Extending CD spectra of proteins to 168 nm improves the analysis for secondary structures. *Anal. Biochem.* **200**, 321–331
5. Sreerama, N. and Woody, R.W. (2000) Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal. Biochem.* **287**, 252–260
6. Sreerama, N., Venyaminov, S.Y., and Woody, R.W. (1999) Estimation of number of α-helical and β-strand segments in proteins using circular dichroism spectroscopy. *Protein Sci.* **8**, 370–380
7. Venyaminov, S.Y. and Vassilenko, K.S. (1994) Determination of protein tertiary structure class from circular dichroism spectra. *Anal. Biochem.* **222**, 176–184
8. Manavalan, P. and Johnson, W.C., Jr. (1983) Sensitivity of circular dichroism to protein tertiary structure class. *Nature* **305**, 831–832
9. Snyder, P.A. and Rowe, E.M. (1980) The first use of synchrotron radiation for vacuum ultraviolet circular dichroism measurements. *Nucl. Instrum. Methods* **172**, 345–349
10. Sutherland, J.C., Desmond, E.J., and Takacs, P.Z. (1980) Versatile spectrometer for experiments using synchrotron radiation at wavelength greater than 100 nm. *Nucl. Instrum. Methods* **172**, 195–199
11. Sutherland, J.C., Emrick, A., France, L.L., Monteleone, D.C., and Trunk, J. (1992) Circular dichroism user facility at the National Synchrotron Light Source: estimation of protein secondary structure. *Biotechniques* **13**, 588–590
12. Wallace, B.A. (2000) Conformational changes by synchrotron radiation circular dichroism spectroscopy. *Nat. Struct. Biol.* **7**, 708–709
13. Wallace, B.A. and Janes, R.W. (2001) Synchrotron radiation circular dichroism spectroscopy of proteins: secondary structure, fold recognition and structural geomics. *Curr. Opin. Chem. Biol.* **5**, 567–571
14. Ojima, N., Sakai, K., Fukazawa, T., and Gekko, K. (2000) Vacuum-ultraviolet circular dichroism spectrophotometer using synchrotron radiation: optical system and off-line performance. *Chem. Lett.* 832–833
15. Ojima, N., Sakai, K., Matsuo, K., Matsui, T., Fukazawa, T., Namatame, H., Taniguchi, M., and Gekko, K. (2001) Vacuum-ultraviolet circular dichroism spectrophotometer using synchrotron radiation: optical system and on-line performance. *Chem. Lett.* 522–523
16. Matsuo, K., Matsushima, Y., Fukuyama, T., Senba, S., and Gekko, K. (2002) Vacuum-ultraviolet circular dichroism of amino acids as revealed by synchrotron radiation spectrophotometer. *Chem. Lett.* 826–827
17. Matsuo, K., Sakai, K., Matsushima, Y., Fukuyama, T., and Gekko, K. (2003) Optical cell with a temperature-control unit

for a vacuum-ultraviolet circular dichroism spectrophotometer. *Anal. Sci.* **19**, 129–132
18. Kabsch, W. and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen bonded and geometric feature. *Biopolymers* **22**, 2577–2637
19. Koslowski, A., Sreerama, N., and Woody, R.W. (2000) Circular dichroism of peptides and proteins in *Circular dichroism: Principles and Applications* (Berova, N., Nakanishi, K., and Woody, R.W. , eds.) pp. 55–95, Willy-VCH Press, New York
20. Woody, R.W. and Koslowski, A. (2002) Recent development in the electronic spectroscopy of amides and α-helical polypeptides. *Biophys. Chem.* 101–102, 535–551
21. Brahms, S. and Brahms, J. (1980) Determination of protein secondary structure in solution by vacuum ultraviolet circular dichroism. *J. Mol. Biol.* **138**, 149–178
22. Young, M.A. and Pysh, E.S. (1975) Vacuum ultraviolet circular dichroism of poly(L-proline) I and II. *J. Amer. Chem. Soc.* **97**, 5100–5103
23. Tiffany, M.L. and Krimm, S. (1968) Circular dichroism of poly-L-proline in an unordered conformation. *Biopolymers* **6**, 1767–1770
24. Woody, R.W. (1992) Circular dichroism and conformation of unordered peptides. *Adv. Biophys. Chem.* **2**, 37–79
25. Sreerama, N. and Woody, R.W. (2003) Structural composition of βI- and βII-proteins. *Protein Sci.* **12**, 384–388
26. Venyaminov, S.Y., Baikalov, I.A., Wu, C.-S.C., and Yang, J.T. (1991) Some problems of CD analyses of protein conformation. *Anal. Biochem.* **198**, 250–255
27. Pancoska, P., Janota, V., and Keiderling, T.A. (1999) Novel matrix descriptor for secondary structure segments in proteins: demonstration of predictability from circular dichroism spectra. *Anal. Biochem.* **267**, 72–83