J. Biochem. **138**, 79–88 (2005)
DOI: 10.1093/jb/mvi101

Improved Estimation of the Secondary Structures 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, and [2]Hiroshima Synchrotron Radiation Center, Hiroshima University, Higashi-Hiroshima 739-8526, Japan

Received February 10, 2005; accepted April 30, 2005

The vacuum-ultraviolet circular dichroism (VUVCD) spectra of 16 globular proteins (insulin, lactate dehydrogenase, glucose isomerase, lipase, conalbumin, transferrin, catalase, subtilisin A, α-amylase, staphylococcal nuclease, papain, thioredoxin, carbonic anhydrase, elastase, avidin, and xylanase) were successfully measured in aqueous solutions at 25°C from 260 to 160 nm under a high vacuum using a synchrotron-radiation VUVCD spectrophotometer. These proteins exhibited characteristic CD spectra below 190 nm that were related to their different secondary structures, which could not be detected with a conventional CD spectrophotometer. The component spectra of α-helices, β-strands, turns, and unordered structures were obtained by deconvolution analysis of the VUVCD spectra of 31 reference proteins including the 15 proteins reported in our previous paper [Matsuo, K. *et al.* (2004) *J. Biochem.* 135, 405–411]. Prediction of the secondary-structure contents using the SELCON3 program was greatly improved, especially for α-helices, by extending the short-wavelength limit of CD spectra to 160 nm and by increasing the number of reference proteins. The numbers of α-helix and β-strand segments, which were calculated from the distorted α-helix and β-strand contents, were close to those obtained on X-ray crystallography. These results demonstrate the usefulness of synchrotron-radiation VUVCD spectroscopy for the secondary structure analysis of proteins.

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

Abbreviations: CD, circular dichroism; HSA, human serum albumin; LDH, lactate dehydrogenase; PPII, poly-L-proline type II; RNase A, ribonuclease A; SNase, staphylococcal nuclease; STI, soybean trypsin inhibitor; VUV, vacuum-ultraviolet; VUVCD, vacuum-ultraviolet circular dichroism.

Determining the structure of a protein is fundamental for understanding its biological function. The primary structure of a protein can now be easily determined from genomic data. The three-dimensional structure can be determined by X-ray crystallography and NMR spectroscopy, but the former requires crystalline proteins, many of which are difficult to obtain, and the latter is still limited to proteins of relatively low molecular weight. Although it exhibits no atomic level resolution, circular dichroism (CD) spectroscopy is a useful technique for filling the crucial gap between the primary structures and three-dimensional structures of proteins, since CD is sensitive to the backbone conformation or secondary structure and is applicable to non-crystalline proteins of any molecular weight (*1*, *2*). CD spectroscopy also has the advantage of allowing the conformational analysis of both native and non-native proteins. Hence this technique is widely used and becoming increasingly useful in structural biology through the increasing number of protein structures deposited in the Protein Data Bank (PDB), the development of programs to extract protein secondary structures from atomic coordinates (*3*, *4*), and advancements in the software used to analyze CD spectra (*5*, *6*).

The "pure component spectra" obtained from a CD database of reference proteins with known X-ray structures have been used to estimate the contents of secondary structures (*7*, *8*), and the numbers of α-helix and β-strand segments (*9*), and to assign tertiary-structure classes (*10*, *11*). Recently, Sreerama and Woody estimated the contents and segment numbers of secondary structures using three programs—CONTIN, SELCON3, and CDSSTR—with five sets of reference spectra (comprising 29, 37, 42, 43, and 48 proteins) extending down to 178 nm (*8*). They found that a larger reference data set and CD data at shorter wavelengths both improved the prediction of secondary structures. Toumadje *et al.* indicated that extending CD spectra to 168 nm, rather than stopping at 178 nm, could improve the prediction of secondary-structure contents (*7*). Thus the establishment of a definitive CD database in the vacuum-ultraviolet (VUV) region is very important for the CD spectroscopy of proteins.

A synchrotron is an excellent high-flux source of photons, yielding higher signal-to-noise ratios in the VUV region that cannot be attained with a conventional CD spectrophotometer. Therefore, since the 1980s, VUVCD spectrophotometers have been constructed at several synchrotron-radiation facilities to extend the short-wavelength limit of CD spectra (*12–18*). Sutherland *et al.* and France *et al.* measured the CD spectra of proteins down

*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		
insulin	Bovine pancreas	4INS	52.9	5.9	4.9	10.0	2.2
glucose isomerase	*Streptomyces rubiginosus*	1OAD	49.5	9.0	16.6	10.4	7.2
lactate dehydrogenase	Bovine heart	9LDT	45.5	17.0	13.6	14.9	7.2
lipase	*Pseudomonas cepacia*	3LIP	37.8	17.5	21.6	11.2	7.2
transferrin	Human serum	1LFG	34.3	17.9	26.8	14.1	5.5
conalbumin	Chicken egg white	1OVT	32.8	17.6	27.0	11.6	5.7
thioredoxin	*Escherichia coli*	2TRX	32.4	27.3	25.5	11.4	5.8
catalase	Bovine liver	7CAT	31.8	14.0	21.2	12.9[b]	7.2
α-amylase	*Bacillus subtilis*	1BAG	31.5	20.7	20.7	25.3	7.7
subtilisin A	*Bacillus* sp.	1SBC	29.5	17.9	23.4	8.6	5.0
papain	Papaya	9PAP	25.9	17.0	17.5	25.0[c]	6.2
nuclease	*Staphyllococcal*	1EY0	24.2	26.8	20.1	9.3	6.5
carbonic anhydrase	Bovine erythrocytes	1G6V	13.1	26.5	27.7	19.0	7.0
elastase	Porcine pancreas	3EST	10.8	34.2	20.8	22.0	5.0
xylanase	*Trichoderma* sp.	1ENX	5.3	62.1	18.9	26.9	7.2
avidin	Chicken egg white	1AVE	3.5	44.9	18.4	15.7	7.0

[a]From crystal data listed in PDB code. [b]At 278 nm. [c]At 276 nm.

to 175 nm under a nitrogen gas atmosphere using the VUVCD spectrophotometer at the Brookhaven National Laboratory (USA) to estimate secondary-structure contents with the VARSELEC program (*13*, *14*). Jones and Clark measured the CD spectra of 20 proteins down to 168 nm at the Daresbury Laboratory (UK) and estimated the contents of four secondary structures (α-helix, β-strand, turn, and unordered structure) of lipoproteins using the SELCON program (*17*). Wallace *et al.* used a VUVCD spectrophotometer to identify the conformational changes associated with mutations of a protein and to establish a database for β-type folding motifs (*18*).

We recently constructed a VUVCD spectrophotometer at the Hiroshima Synchrotron Radiation Center (HSRC, Japan), which can be used to measure CD spectra down to 140 nm in aqueous solutions by keeping all the optical devices under a high vacuum (*19–22*). We successfully measured the CD spectra of 15 globular proteins in the wavelength region of 260 to 160 nm using this spectrophotometer, and estimated the contents and segment numbers of the secondary structures (*23*). The results indicated that the estimation of secondary-structure contents can be improved by extending the CD spectra at least down to 165 nm and that the accumulation of VUVCD spectra should be useful for further characterizing the secondary structures of proteins. In the present study, we measured the VUVCD spectra of additional 16 globular proteins down to 160 nm and analyzed their secondary structures using the SELCON3 program with the VUVCD spectra of 31 reference proteins, including the 15 proteins previously reported (*23*). The improvements and limitations of secondary-structure estimation of proteins are discussed here on the basis of the performance indices between X-ray and VUVCD estimates of the contents and segment numbers of secondary structures.

MATERIALS AND METHODS

Materials—α-Amylase was purchased from Wako Pure Chemicals. Lactate dehydrogenase (LDH) and elastase were obtained from Serva Electrophoresis and Elastin Products, respectively. Glucose isomerase and xylanase were purchased from Hampton Research. Staphylococcal nuclease (SNase) was a generous gift from Dr. M. Kataoka of the Nara Institute of Science and Technology. Other proteins (insulin, lipase, conalbumin, transferrin, catalase, subtilisin A, papain, thioredoxin, carbonic anhydrase, and avidin) were purchased from Sigma. The origins of these proteins are listed in Table 1 along with their structural parameters. All proteins were used without further purification. The water-soluble proteins (conalbumin, transferrin, subtilisin A, SNase, and papain) were dialyzed against double-distilled water at 4°C. Insulin, α-amylase, and thioredoxin were dissolved in a HCl solution (pH 2.2), 50 mM Tris-HCl buffer containing 3.5 mM $CaCl_2$, and 10 mM glycine buffer, respectively. Glucose isomerase, LDH, lipase, catalase, carbonic anhydrase, avidin, and xylanase were dissolved in 10–40 mM potassium phosphate buffer, since these proteins are insoluble in pure water. The protein solutions were exhaustively dialyzed against the same buffer at 4°C. The dialyzed protein solutions were centrifuged at 14,000 rpm for 15 min to remove aggregates, and then adjusted to protein concentrations of 0.2–1.4% by dilution or concentration with a Centricut mini (V-10, Kurabo Industries) and a Centricon (YM-3, Millipore). Protein concentrations were determined by absorption measurement (V-560, Jasco) using the molar extinction coefficients in the literature, which are listed in Table 1.

VUVCD Measurements—The VUVCD spectra of proteins were measured in the wavelength region of 260 to 160 nm under a high vacuum (10^{-4} Pa) at 25°C, using the VUVCD spectrophotometer constructed at the HSRC (BL15). The optical devices of the spectrophotometer and the cell used were described previously (*19–22*). 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. 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 0.2–0.3%. To reduce the light absorption by water, no spacer was used for measurements below 175 nm with protein concentrations of 0.8–1.4%. The detailed conditions for measurements are given in a previous paper (*23*). The ellipticity was reproducible within an error of 5%, which was mainly attributable to noise and to inaccuracy in the light path length.

Secondary-Structure Analysis of Proteins—The secondary structures of 16 proteins in crystal form were assigned using the DSSP program (*3*) for the PDB codes listed in Table 1. The 3_{10}-helices were considered to be α-helices, the bends were treated as turns, and the single residues assigned as turns and bends were classified as unordered structures. Moreover, α-helices and β-strands were divided 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 (*9*). Thus the protein structures were classified 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 31 proteins including the 15 proteins reported previously (*23*) [myoglobin, hemoglobin, human serum albumin (HSA), cytochrome *c*, peroxidase, α-lactalbumin, lysozyme, ovalbumin, ribonuclease A (RNase A), β-lactoglobulin, pepsin, trypsinogen, α-chymotrypsinogen, soybean trypsin inhibitor (STI), and concanavalin A], using the SELCON3 program (*8*). This program was modified so as to extend it down to 160 nm. Its performance in the analysis of secondary structures was characterized by the root-mean-square deviations (δ) and the Pearson correlation coefficients (*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 following equations:

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

$$r = (\sum X_i Y_i - 1/N \sum X_i \sum Y_i) / (\sqrt{\sum X_i^2 - (\sum X_i)^2/N} \sqrt{\sum Y_i^2 - (\sum Y_i)^2/N})$$

where X_i and Y_i are the X-ray and CD estimates of a given type of secondary structure, i, 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 down to 160 nm of 16 proteins in aqueous solutions, which are divided into four panels for clarity. These spectra are superimposed on those down to 190 nm obtained with a conventional spectrophotometer, which indicates the good performance of the VUVCD spectrophotometer and the optical cell used. The spectral intensities did not decrease during the data acquisition (which lasted 3 h), indicating that the proteins were not damaged by synchrotron radiation (at 0.7 GeV).

As shown in Fig. 1a, four α-helix–rich proteins (insulin, glucose isomerase, LDH, and lipase)—with α-helix contents of 37.8–52.9%—exhibit three negative peaks at around 222, 208, and 170 nm, a positive peak at around 195 nm, and an expected positive peak below 160 nm, which are similar to those of α-proteins such as myoglobin (*23*). The negative peak at 222 nm is attributable to the n–π* transition of the peptide, and the negative and positive peaks at 208 and 195 nm, respectively, are attributable to the parallel and perpendicular excitons of the π–π* transition of the peptide (*5*, *24*). The CD peaks in the region of 175 to 150 nm probably arise from interpeptide charge transfer (*24*). Similar VUVCD spectra were observed for eight proteins having lower α-helix contents of 24.2–34.3% (transferrin, conalbumin, thioredoxin, catalase, α-amylase, subtilisin A, papain, and SNase), although the CD spectrum of thioredoxin was close to that of typical β-strand–rich proteins (Fig. 1b and c). Variations in the CD spectra between these α-helix–rich proteins would arise from not only differences in secondary-structure contents but also from ones in the arrangement of α-helices and β-strands in the tertiary structure (α/β and α+β classes).

Figure 1d shows the VUVCD spectra of two β-strand–rich proteins (carbonic anhydrase and elastase) and two β-proteins (xylanase and avidin). Elastase exhibits two negative peaks at around 200 and 170 nm, and a positive peak at around 185 nm, as found for α-chymotrypsinogen and trypsinogen, which have similar α-helix (10–14%) and β-strand (32–34%) contents (*23*). Xylanase containing 62.1% β-strand exhibits two negative peaks at around 220 and 175 nm, and two positive peaks at around 195 and 160 nm, which are similar to those for β-proteins such as β-lactoglobulin and concanavalin A (*23*). Interestingly, avidin (44.9% β-strands) exhibits a very different spectrum from those of xylanase and other β-strand–rich proteins: there are two positive peaks at around 230 and 195 nm, and two negative peaks at 210 and 175 nm. Manavalan and Johnson found that β-strand–rich proteins exhibit two types of CD spectra (*11*), reminiscent of either model β-sheet or poly-L-proline type II (PPII), which are respectively named β_I and β_{II} (*25*). Recently, Sreerama and Woody proposed that β-lactoglobulin and concanavalin A could be classified into the β_I type, and elastase, carbonic anhydrase, trypsinogen, and chymotrypsinogen into the β_{II} type (*26*). Xylanase and avidin could be classified into the β_I type, based on their characteristic CD spectra and X-ray structures. Thus, β-proteins and β-strand–rich proteins exhibit larger variations in VUVCD spectra compared with α-proteins and α-helix–rich proteins.

Component VUVCD Spectra of Secondary Structures—As shown in Fig. 1, the VUVCD spectra of proteins exhibit various characteristic peaks in the VUV region depending on the secondary structures present. Therefore, to evaluate the contribution of each secondary structure, the VUVCD spectra of 31 reference proteins were deconvoluted into the spectra of four components—α-helices, β-strands, turns, and unordered structures—using the SELCON3 program (*8*). The spectra for each of these components averaged across the 31 proteins are shown in Fig. 2. Such component spectra down to 160 nm were first successfully obtained in the present study. The overall

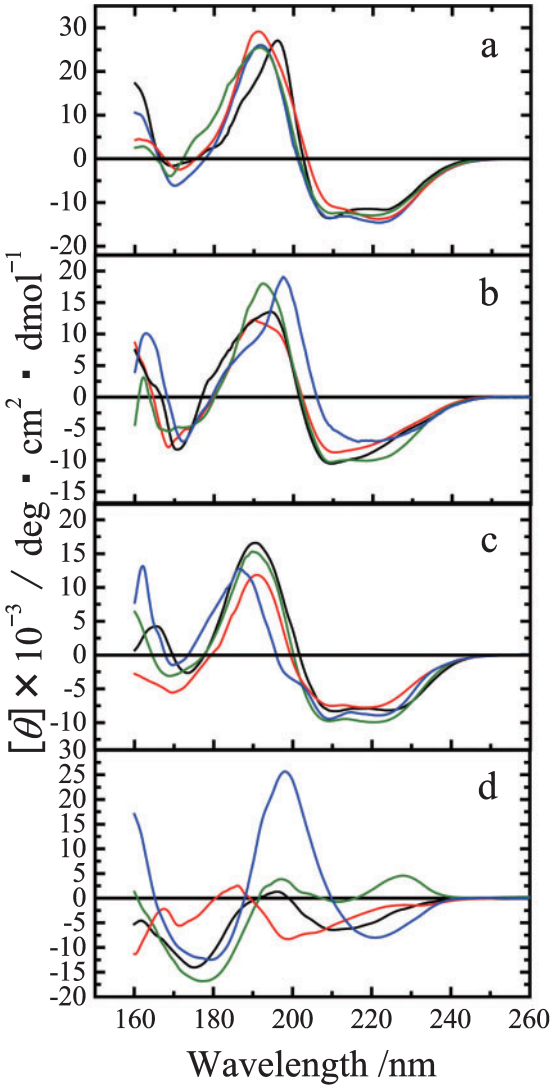

Fig. 1. **VUVCD spectra for 16 proteins in aqueous solutions at 25°C.** (a) Insulin (black), LDH (red), glucose isomerase (blue), and lipase (green); (b) transferrin (black), conalbumin (red), thioredoxin (blue), and catalase (green); (c) α-amylase (black), subtilisin A (red), papain (blue), and SNase (green); and (d) carbonic anhydrase (black), elastase (red), xylanase (blue), and avidin (green). A cell with a 50-μm path length was used for the measurements from 260 to 175 nm, and no spacer was used for those 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.

features of these component spectra are similar to those previously estimated from model polypeptides (*27–29*) and reference proteins (*26*, *30*, *31*) in the far-UV region, but clearly distinct characteristics can be observed in the VUV region. It is therefore pertinent to briefly compare our component spectra with those in the literature, although detailed discussion must await the theoretical assignment of the VUVCD spectra.

The CD spectrum of α-helices exhibits a positive peak at around 192 nm, three negative peaks at around 222, 208, and 168 nm, and a shoulder at around 180 nm (Fig. 2). These peak wavelengths are very close to those found by Sreerama and Woody using two sets of reference pro-

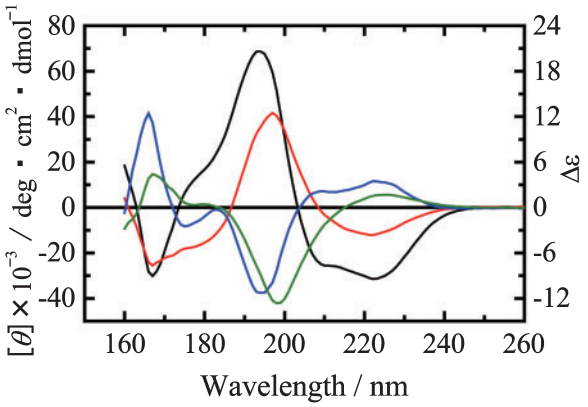

Fig. 2. **Component VUVCD spectra of four secondary structures deconvoluted from the 31 reference proteins.** α-Helix (black), β-strand (red), turn (blue), and unordered structure (green). RMSD, root-mean-square deviation (δ).

teins, comprising 37 and 16 (including a polypeptide) proteins, in the wavelength region of 260 to 178 nm (*26*, *30*). However, the peak intensities are not necessarily consistent among the three sets of reference proteins including ours, so the component spectrum of α-helix depends on the reference proteins used in the deconvolution. Our α-helix spectrum, as well as those obtained by Sreerama and Woody, is also similar to the spectra for α-helices of poly-L-lysine (*27*) and poly-L-glutamic acid (*28*) down to 190 nm, although the peak intensities are considerably smaller. In contrast, the intensity of the negative peak at around 170 nm is higher than that of poly-L-glutamic acid. These differing intensities in the α-helix spectra between proteins and polypeptides could be mainly due to differences in the chain lengths of α-helices. Our α-helix spectrum resembles the calculated and observed CD spectra, respectively, for α-helices consisting of 10 and 11 amino acid residues (*32*, *33*), which are close to the average chain length (10.4 residues) of α-helices in the 31 proteins used in this study.

The component spectrum of β-strands has two negative peaks at around 220 and 170 nm, and a positive peak at around 195 nm (Fig. 2). These peaks are typically observed in the VUVCD spectra of β-proteins such as xylanase. The overall spectrum of the β-strand is similar to those down to 178 nm obtained by Sreerama and Woody (*26*, *30*), except for a considerably higher intensity. Moreover, our β-strand spectrum is close to the β-sheet spectra of poly-(leucine-lysine) (*29*) and poly-L-lysine (*27*) down to 190 nm, although there is a little difference in the peak wavelength at around 220 nm.

The component spectrum of turns is characterized by three positive CD peaks at around 225, 208, and 165 nm, and two negative peaks at around 192 and 175 nm (Fig. 2). This spectrum is also very similar to that obtained by Sreerama and Woody (*30*) down to 178 nm, although the peak intensity at around 190 nm differs considerably. The component spectrum of unordered structures consists of two positive peaks at around 225 and 168 nm, and a negative peak at around 200 nm (Fig. 2), which are very similar to those of PPII (*27*) and of the PPII component deconvoluted from the CD spectra of 16 proteins down to 178 nm (*30*, *34*). This is probably because the PPII confor-

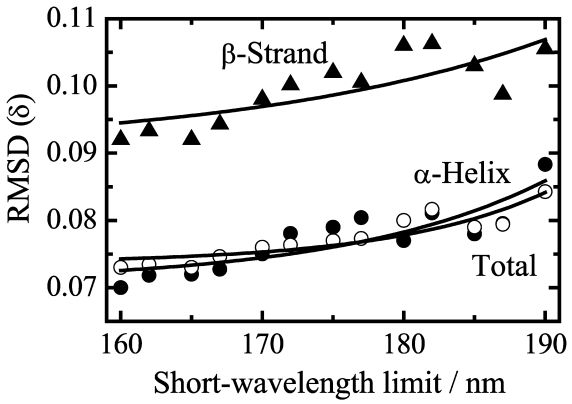

Fig. 3. **Plots of the δ values for α-helix (solid circles), β-strand (solid triangles), and total performance (open circles) against the short-wavelength limit.** Solid lines represent exponential fits to the data points.

mation is assigned as part of the unordered structures by the DSSP program used here. Thus the CD spectrum of unordered structure would mainly be due to the PPII conformation.

It is evident that the VUVCD spectra of the four components exhibited new peaks (below 185 nm) that are difficult to detect in solution on conventional CD spectroscopy. The CD spectrum of α-helices is positive at around 180 nm, whereas it is negative for β-strands and turns. The CD spectra of α-helices and β-strands exhibit a negative peak with a similar intensity at 170 nm, and become positive below 160 nm. In contrast, turns and unordered structures exhibit a positive peak near 170 nm, and are negative below 160 nm. Thus the CD spectra of proteins below 185 nm are greatly affected by not only α-helices and β-strands but also by turns and unordered structures as well as those in the far-UV region. These component spectra should form the basis for future theoretical and *ab initio* assignments of VUVCD spectra.

Estimation of Secondary-Structure Contents—The secondary-structure analysis of proteins was performed using the CD spectra of model polypeptides with specific secondary structures as pure component spectra down to 165 nm, the secondary-structure content being estimated by least-squares fitting of the protein CD spectra (*35*). The model peptide spectra were replaced by the pure component spectra down to 168 nm obtained from a set of proteins with known X-ray structures, a helical polypeptide being used in the secondary structure analysis (*7*). These results indicate the improvement of secondary-structure analysis by extending the lower wavelength limit of CD measurements, but they cannot be directly compared with ours since the pure component spectra obtained from polypeptides are clearly different from those for proteins in the VUV region. As indicated above, our VUVCD spectra for proteins down to 160 nm provide new and detailed information on secondary structures that cannot be obtained from far-UV CD spectra down to only 190 nm. It is therefore of interest to determine how secondary-structure analysis can be improved by extending the CD spectra to shorter wavelengths and by increasing the number of reference spectra. 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 (*8*) with the VUVCD spectra of 31 reference proteins including the 15 proteins examined in our previous study (*23*) in the wavelength regions of 260 nm to various short-wavelength limits. The results of the calculations are listed in Table 2, and the δ values for α-helix, β-strand and overall performance are plotted against the short-wavelength limit in Fig. 3. For all the short-wavelength limits, the accuracy of prediction was better for α-helices than β-strands. Evidently, δ decreases and *r* increases for α-helices and β-strands as the short-wavelength limit decreases, while these parameters are not significantly dependent on the wavelength limit for turns and unordered structures, resulting in improved overall estimation of the secondary-structure contents. The overall δ value (0.073) at 160 nm is considerably smaller than that (0.089) obtained from 15 reference proteins (*23*). This implies that prediction of the secondary-structure content can be improved by increasing the number of reference proteins, which is consistent with the results of Sreerama and Woody (*8*).

The validity of the VUVCD spectra obtained on the SELCON3 analysis was further examined with six types of secondary structure, which were obtained by splitting α-helices and β-strands into regular (α_R and β_R) and distorted (α_D and β_D) classes, 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. The estimation of α_R, β_R, and β_D structures is much improved, but not that of α_D structures, turns, and unordered structures. As shown by the overall performance indices in the last column of the table, the accuracy of secondary-structure estimation improves as the short-wavelength limit decreases: the best overall performance (δ = 0.054) was obtained by extending the short-wavelength limit to 160 nm. The large δ values for β_R structures and turns compared with

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	0.078	0.922	0.103	0.762	0.057	0.296	0.069	0.534	0.079	0.860
260–180 nm	0.077	0.924	0.106	0.744	0.058	0.245	0.070	0.520	0.080	0.832
260–175 nm	0.079	0.922	0.102	0.766	0.055	0.165	0.063	0.636	0.077	0.828
260–170 nm	0.075	0.927	0.098	0.789	0.058	0.168	0.067	0.578	0.076	0.860
260–165 nm	0.072	0.934	0.092	0.818	0.058	0.240	0.068	0.552	0.073	0.877
260–160 nm	0.070	0.938	0.092	0.818	0.056	0.233	0.069	0.540	0.073	0.865

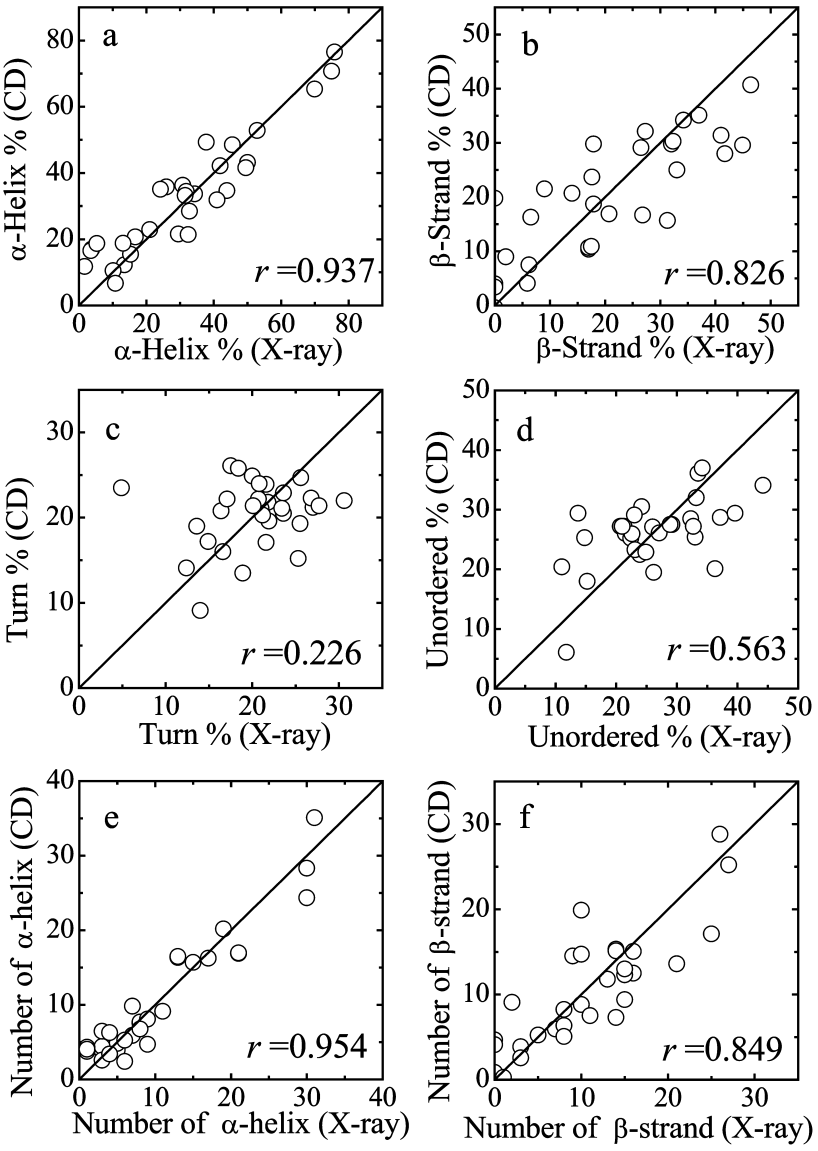

Fig. 4. **Plots of the contents and segment numbers of secondary structures predicted from the 31 VUVCD spectra down to 160 nm against the X-ray estimates.** (a) α-helix content, (b) β-strand content, (c) turn content, (d) unordered-structure content, (e) number of α-helix segments, and (f) number of β-strand segments.

α-helix components (α_R and α_D) may be partly due to the existence of two types of conformers in β-strands (parallel and antiparallel forms) and turns (types I and II), which originate from large geometric variations in their dihedral angles (ϕ and φ) (*36*, *37*).

Comparison of X-Ray and VUVCD Estimations of Secondary-Structure Contents—Table 4 lists the secondary-structure contents of the 31 proteins that were determined by SELCON3 analysis with the VUVCD spectra down to 160 nm, and the X-ray crystal structures assigned with the DSSP program. As listed in the last column of the table, the overall root-mean-square deviations (δ) between the X-ray and VUVCD estimates are less than 0.04 for 11 proteins, indicating good VUVCD estimation of the secondary-structure contents. In contrast, the prediction is inadequate for insulin and xylanase, whose δ values are larger than 0.10. The poor prediction for insulin (δ = 0.103), an α-helix–rich protein,

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	0.053	0.934	0.044	0.748	0.078	0.735	0.035	0.698	0.057	0.301	0.070	0.514	0.058	0.842
260–180 nm	0.048	0.945	0.041	0.775	0.078	0.736	0.033	0.731	0.058	0.259	0.071	0.487	0.057	0.847
260–175 nm	0.048	0.946	0.046	0.712	0.078	0.732	0.032	0.746	0.056	0.179	0.064	0.624	0.056	0.855
260–170 nm	0.049	0.944	0.040	0.072	0.072	0.775	0.031	0.775	0.058	0.189	0.067	0.576	0.055	0.861
260–165 nm	0.048	0.946	0.042	0.776	0.072	0.797	0.031	0.783	0.057	0.259	0.067	0.558	0.055	0.862
260–160 nm	0.046	0.953	0.043	0.749	0.070	0.804	0.030	0.795	0.057	0.226	0.068	0.563	0.054	0.864

[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 31 reference proteins determined by X-ray analysis and CD spectra from 260 to 160 nm.**[a]

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	0.030
	CD	56.9	19.6	3.7	0.3	14.1	6.1	100.7	
hemoglobin	X-ray	54.0	21.0	0.0	0.0	14.0	11.0	100	0.049
	CD	49.2	21.5	2.2	1.2	9.1	20.4	103.6	
HSA	X-ray	49.1	20.8	0.0	0.0	14.9	15.2	100	0.039
	CD	41.3	24.0	–2.0	1.6	17.2	18.0	100.1	
cytochrome *c*	X-ray	21.9	19.1	0.0	0.0	21.9	37.2	100	0.073
	CD	16.7	15.2	11.9	7.9	19.6	28.7	100.0	
peroxidase	X-ray	29.2	20.8	0.7	1.3	25.6	22.4	100	0.038
	CD	22.1	21.1	3.1	5.9	24.7	25.2	102.1	
α-lactalbumin	X-ray	19.5	24.4	1.6	4.9	23.6	26.0	100	0.053
	CD	19.4	15.3	10.0	6.3	20.5	27.1	98.6	
lysozyme	X-ray	20.2	21.7	1.5	4.7	30.6	21.3	100	0.048
	CD	23.9	18.3	3.5	4.0	22.0	27.2	98.9	
ovalbumin	X-ray	17.9	12.9	23.0	8.3	16.4	21.5	100.0	0.065
	CD	19.3	17.0	9.2	6.5	20.8	26.0	98.8	
RNase A	X-ray	11.3	9.7	21.7	11.3	21.8	24.2	100.0	0.039
	CD	10.5	12.4	15.4	9.6	21.8	30.5	100.2	
β-lactoglobulin	X-ray	5.6	11.1	28.7	12.3	21.6	20.7	100.0	0.046
	CD	8.8	11.9	20.5	10.9	23.9	27.2	103.2	
pepsin	X-ray	3.0	12.3	26.4	15.3	20.0	23.0	100.0	0.053
	CD	4.3	11.2	17.5	10.5	24.9	29.1	97.5	
trypsinogen	X-ray	5.3	4.8	20.9	11.4	25.3	32.3	100.0	0.047
	CD	2.8	7.7	20.0	10.3	15.2	28.5	84.5	
α-chymotrypsinogen	X-ray	5.1	8.4	20.0	12.0	21.0	33.5	100.0	0.016
	CD	3.7	8.6	17.3	12.5	21.1	36.1	99.3	
STI	X-ray	0.0	1.7	19.3	17.7	17.1	44.2	100.0	0.059
	CD	3.5	8.3	21.5	13.6	22.2	34.1	103.2	
concanavalin A	X-ray	0.0	3.8	32.9	13.5	23.6	26.2	100.0	0.051
	CD	6.1	10.9	28.0	12.7	22.9	19.5	100.1	
insulin	X-ray	29.4	23.5	2.0	3.9	4.9	36.3	100.0	0.103
	CD	32.6	20.2	3.0	1.1	23.5	20.1	100.5	
glucose isomerase	X-ray	27.8	21.7	3.9	5.1	16.6	24.9	100.0	0.049
	CD	24.1	17.5	13.9	7.6	16.0	22.9	102.0	
LDH	X-ray	27.4	18.1	9.0	8.0	13.6	23.9	100.0	0.031
	CD	29.6	18.9	6.3	4.4	19.0	22.5	100.7	
lipase	X-ray	21.6	16.2	10.6	6.9	21.6	23.1	100.0	0.044
	CD	28.7	20.6	6.2	4.7	17.1	23.3	100.6	
transferrin	X-ray	16.9	17.4	10.1	7.8	26.8	21.0	100.0	0.032
	CD	17.3	16.4	11.4	7.3	22.3	27.2	101.9	
conalbumin	X-ray	15.3	17.5	10.0	7.6	27.0	22.6	100.0	0.038
	CD	14.3	14.2	15.3	8.4	21.2	26.0	99.4	
thioredoxin	X-ray	17.6	14.8	18.0	9.3	25.5	14.8	100.0	0.065
	CD	8.8	12.6	22.4	9.7	19.3	25.3	98.1	
catalase	X-ray	16.8	15.0	10.1	3.9	21.2	33.0	100.0	0.038
	CD	18.4	16.0	12.8	7.9	20.3	25.4	100.8	
α-amylase	X-ray	11.8	19.7	10.8	9.9	20.7	27.1	100.0	0.032
	CD	17.3	15.9	10.5	6.4	22.2	26.1	98.4	
subtilisin A	X-ray	16.4	13.1	11.3	6.6	23.4	29.2	100.0	0.047
	CD	9.8	11.8	19.2	10.6	21.1	27.5	100.0	
papain	X-ray	12.7	13.2	9.4	7.6	17.5	39.6	100.0	0.065
	CD	17.3	18.5	4.4	6.0	26.1	29.4	101.7	
SNase	X-ray	13.4	10.8	16.1	10.7	20.1	28.9	100.0	0.044
	CD	18.3	16.8	9.9	6.8	21.4	27.5	100.7	
carbonic anhydrase	X-ray	0.8	12.3	15.0	11.5	27.7	32.7	100.0	0.050
	CD	8.4	10.4	19.1	10.0	21.4	27.2	96.5	
elastase	X-ray	0.8	10.0	22.5	11.7	20.8	34.2	100.0	0.031
	CD	2.7	4.0	21.6	12.6	24.0	37.0	101.9	
xylanase	X-ray	3.2	2.1	46.3	15.8	18.9	13.7	100.0	0.109
	CD	10.2	8.5	28.5	9.9	13.5	29.4	100.0	
avidin	X-ray	0.0	3.5	32.0	12.9	18.4	33.2	100.0	0.082
	CD	3.1	13.5	16.7	12.9	25.8	32.0	104.0	

[a]The results for the top 15 proteins in the table were obtained using the VUVCD spectra reported previously (*23*).

Table 5. **Numbers of α-helix and β-strand segments determined by X-ray analysis and CD spectra from 260 to 160 nm.**

Protein	α-helix		β-strand	
	X-ray	CD	X-ray	CD
myoglobin	8	8.0	0	0.0
hemoglobin	8	8.0	0	1.0
HSA	31	35.0	0	4.0
cytochrome *c*	5	4.0	0	4.0
peroxidase	17	16.0	2	9.0
α-lactalbumin	9	5.0	3	4.0
lysozyme	7	6.0	3	3.0
ovalbumin	13	16.0	16	13.0
RNase A	3	4.0	7	6.0
β-lactoglobulin	5	5.0	10	9.0
pepsin	11	9.0	25	17.0
trypsinogen	3	4.0	13	12.0
α-chymotrypsinogen	6	5.0	14	15.0
STI	1	4.0	15	12.0
concanavalin A	3	6.0	16	15.0
insulin	3	3.0	1	0.0
glucose isomerase	21	17.0	10	15.0
LDH	15	16.0	14	7.0
lipase	13	16.0	11	8.0
transferrin	30	28.0	27	25.0
conalbumin	30	24.0	26	29.0
thioredoxin	4	3.0	5	5.0
catalase	19	20.0	10	20.0
α-amylase	21	17.0	21	14.0
subtilisinA	9	8.0	9	14.0
papain	7	10.0	8	6.0
SNase	4	6.0	8	5.0
carbonic anhydrase	8	7.0	15	13.0
elastase	6	2.0	14	15.0
xylanase	1	4.0	15	9.0
avidin	1	4.0	8	8.0

may be attributable to the large deviation in the contents of turns and unordered structures between the X-ray and CD estimates. The large δ value for xylanase (0.109) may be due to its exceptionally high β-strand content. The rather large δ value for avidin (0.082), whose VUVCD spectrum differs significantly from those of others, is attributable to poor estimation of its β_R and α_D contents.

To clarify the overall performance of the secondary-structure analysis, the contents of α-helices ($\alpha_R + \alpha_D$), β-strands ($\beta_R + \beta_D$), turns, and unordered structures predicted from VUVCD spectra of the 31 proteins are plotted against those obtained from X-ray structures in Fig. 4. Figure 4a shows that there is good agreement between the estimates obtained with the two methods for the α-helix content ($r = 0.937$), as expected from the small δ values (Tables 2 and 3). The performance is also good for β-strands ($r = 0.826$) (Fig. 4b). In contrast, only a low correlation is observed for turns (probably because the content clusters within a small range at around 20%), while slightly better performance is observed for the unordered structures. Thus, the performance of secondary-structure prediction decreases in the following order: α-helices > β-strands > unordered structures > turns.

Estimation of the Numbers of α-Helix and β-Strand Segments—The numbers of α-helix and β-strand segments have previously been estimated by two methods. Pancoska *et al.* used a matrix descriptor of secondary-structure segments for the neural-network-based analysis of protein CD spectra (*38*). Sreerama *et al.* estimated the numbers of α-helix and β-strand segments from the distorted residues in α-helices and β-strands, assuming that on average there were four and two distorted residues per α-helix and β-strand, respectively (*9*). The results of these two analyses are comparable, and hence we estimated the numbers of α-helix and β-strand segments by the method of Sreerama *et al*. using the α_D and β_D fractions obtained from the VUVCD spectra of 31 proteins in the wavelength region of 260 to 160 nm (Table 3). The results of the calculation are listed in Table 5, and compared with those determined from the X-ray structures. The numbers of α-helix and β-strand segments are plotted against those determined from X-ray structures in Figure 4e and f, respectively. There is clearly agreement for α-helix segments ($r = 0.945$), while slightly larger deviation is observed for β-strand segments. The root-mean-square differences between the VUVCD and X-ray estimates of the segment numbers were calculated to be 2.6 and 4.0 for α-helices and β-strands, respectively; the corresponding values were calculated to be 3.6 and 2.5 for α-helices and β-strands, respectively, for the VUVCD spectra of 15 proteins to 165 nm (*23*). Similar values (3.2 and 2.5) were obtained from the spectra of 29 proteins down to 178 nm by Sreerama *et al*. (*9*). Thus, the prediction can be improved for α-helices but it is rather worse for β-strands on extending the short-wavelength limit and increasing the numbers of reference proteins, despite the improvement in the prediction for β_D (Table 3). However, this does not necessarily mean that our VUVCD data are not useful for predicting the number of β-strands because these root-mean-square differences depend on not only the number but also the type of reference proteins used, as discussed below.

Further Improvements and Limitations of Secondary-Structure Analysis—As indicated above, VUVCD spectra down to 160 nm provide a useful database for the secondary-structure analysis of proteins with the SELCON 3 program. The results of analysis are satisfactory for α-helices but not necessarily adequate for other structures, making it necessary to discuss the possibility of further improvements and limitations of the secondary-structure analysis with VUVCD spectra. As indicated in Tables 2 and 3, and Fig. 3, the δ value decreases when the short-wavelength limit of VUVCD spectra is lower. However, the plots in Fig. 3 appear to be exponential, and the overall δ values for the 31 proteins examined would be saturated at approximately 0.07 and 0.05 for four and six types of secondary structures, respectively. Therefore, it is unlikely that significant further improvements in the prediction would result from further extension of the short-wavelength limit below 160 nm (where the signal-to-noise ratios become lower), even though our VUVCD spectrophotometer works down to 140 nm.

The number and type of reference proteins are also important for improving the prediction of secondary structures. The δ values obtained down to 160 nm for a set of 15 reference proteins reported previously (*23*) are listed in Table 6 for comparison with those for the 31 proteins in this study. The δ values for all the component

Table 6. **Performance indices (δ and *r*) of six types of secondary structure calculated using different numbers of reference CD spectra of proteins from 260 to 160 nm.**

Reference proteins	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*
15[a]	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
	(0.041)[e]		(0.041)[e]		(0.066)[e]		(0.031)[e]		(0.045)[e]		(0.061)[e]		(0.049)[e]	
31[b]	0.046	0.953	0.043	0.749	0.070	0.804	0.030	0.795	0.057	0.226	0.068	0.563	0.054	0.864
28[c]	0.042	0.957	0.039	0.734	0.069	0.794	0.030	0.760	0.041	0.490	0.054	0.692	0.048	0.913
18[d]	0.053	0.916	0.037	0.282	0.054	0.465	0.035	0.356	0.061	0.396	0.076	0.453	0.054	0.875
13[d]	0.033	0.835	0.050	0.904	0.064	0.695	0.022	0.594	0.039	0.200	0.046	0.792	0.044	0.902

[a]Taken from a previous paper (*23*). [b]The present study. [c]Three proteins (insulin, STI, and avidin) were eliminated from the 31 proteins. [d]Eighteen proteins containing more α-helices than β-strands (α-group), and the remaining 13 proteins (β-group) (see text and Table 1). [e]Performance indices for the secondary structures of the 15 proteins predicted using the 31-protein data set.

structures except turns are smaller in the 31-protein data set than in the 15-protein data set, resulting in the large decrease in the overall δ values. As indicated by the parenthesized values in the table, the secondary structures of each protein in the 15-protein data set are better predicted when the 31-protein data set is used. These results demonstrate that increasing the number of reference VUVCD spectra improves the prediction of secondary structures of proteins.

As is evident in Fig. 1, some proteins exhibit significantly different VUVCD spectra. For example, avidin, whose β-strands are all in an antiparallel form, exhibits a positive peak at around 230 nm. As shown in a previous study (*23*), the spectrum of STI resembles that of PPII, which has a left-handed three-fold helix (*39*). If such special cases are excluded from the 31-protein data set, the accuracy of the secondary-structure prediction should increase. To confirm this, we eliminated three proteins (STI, avidin, and insulin) from the 31 proteins and calculated the performance indices for each of the secondary structures of the remaining 28 proteins. Insulin was eliminated because its spectrum was measured under acidic conditions (pH 2.2) and the content of turns estimated from the CD spectra deviates greatly from the X-ray estimate (Fig. 4c). As indicated in Table 6, eliminating these proteins decreased the δ values of all the component structures, resulting in an improved overall performance (δ = 0.048) compared with the results for the 31 proteins (δ = 0.054).

As another case test, the data set of 31 proteins was divided into two groups to compare the prediction for α-helix–rich and β-strand–rich proteins: 18 proteins whose α-helix contents were greater than their β-strand contents were classified as the α-group, and the remaining 13 proteins were classified as the β-group (see Table 1). The performance indices for each component structure calculated using these two groups of data set are listed in Table 6. For the α-group, the δ value is smaller for only two component structures—α_D (0.037) and β_R (0.054)—than the corresponding values obtained from the 31-protein data set (0.043 and 0.070, respectively), and the overall performance remains unchanged (δ= 0.054). On the other hand, when the 13 proteins in the β-group are used as the reference CD data set, the δ values for all the component structures except α_D decrease significantly and the overall performance is greatly improved (δ = 0.044). The root-mean-square differences in the segment numbers between the CD and X-ray estimates were calculated to be 2.5 for both α-helices and β-strands. Thus the prediction of secondary-structure segments for β-proteins and β-strand–rich proteins can be improved by using the reference data set of the β-group. These results indicate that the type of reference proteins as well as their number is important for secondary-structure analysis based on VUVCD spectra. Such analyses might be further improved by selecting the most appropriate VUVCD database.

CONCLUDING REMARKS

The present paper describes the results of secondary-structure analyses of proteins with VUVCD spectra of 31 proteins, which were successfully measured down to 160 nm using a synchrotron-radiation spectrophotometer. The component spectra of four secondary structures revealed new characteristic higher energy peaks that are useful for the future theoretical assignment of CD spectra. It is evident that the content and segment numbers of secondary structures can be more accurately estimated by extending the short-wavelength limit and increasing the reference VUVCD spectra. The usefulness of synchrotron-radiation VUVCD spectroscopy in structural biology should increase with the further accumulation of VUVCD databases and improvements in analysis methods.

This work was financially supported by a JSPS Research Fellowship for Young Scientists (No. 15008237), and a Grant-in-Aid for Scientific Research from the Ministry of Education, Science, Sports, and Culture of Japan (No. 16350088).

REFERENCES

1. Fasman, G.R. (1996) *Circular Dichroism and the Conformational Analysis of Biomolecules*, Plenum, Press, New York
2. Berova, N., Nakanishi, K., and Woody, R.W. (2000) *Circular dichroism: Principles and Applications* 2nd ed., Wiley-VCH Press, New York
3. Kabsch, W. and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometric features. *Biopolymers* **22**, 2577–2637
4. Frishman, D. and Argos, P. (1995) Knowledge-based protein secondary structure assignment. *Proteins* **23**, 566–579
5. 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.) 2nd ed., pp. 601–620, Wiley-VCH Press, New York
6. Venyaminov, S.Y. and Yang, J.T. (1996) Determination of protein secondary structure in *Circular Dichroism and the Confor-*

mational Analysis of Biomolecules (Fasman, G.R., ed.) pp. 69–107, Plenum Press, New York
7. 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
8. 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
9. 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
10. Venyaminov, S.Y. and Vassilenko, K.S. (1994) Determination of protein tertiary structure class from circular dichroism spectra. *Anal. Biochem.* **222**, 176–184
11. Manavalan, P. and Johnson, W.C., Jr. (1983) Sensitivity of circular dichroism to protein tertiary structure class. *Nature* **305**, 831–832
12. 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
13. 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
14. France, L.L., Kieleczawa, J., Dunn, J.J., Hind, G., and Sutherland, J.C. (1992) Structural analysis of an outer surface protein from the Lyme disease spirochete, Borrelia burgdorferi, using circular dichroism and fluorescence spectroscopy. *Biochim. Biophys. Acta* **1120**, 59–68
15. Wallace, B.A. (2000) Conformational changes by synchrotron radiation circular dichroism spectroscopy. *Nat. Struct. Biol.* **7**, 708–709
16. Clarke, D.T. and Jones, G.R. (2004) CD12: a new high-flux beamline for ultraviolet and vacuum-ultraviolet circular dichroism on the SRS, Daresbury. *J. Synchrotron Radiat.* **11**, 142–149
17. Jones, G.R. and Clarke, D.T. (2004) Applications of extended ultra-violet circular dichroism spectroscopy in biology and medicine. *Faraday Discuss.* **126**, 223–236
18. Wallace, B.A., Wien, F., Miles, A.J., Lees, J.G., Hoffmann, S.V., Evans, P., Wistow, G.J., and Slingsby, C. (2004) Biomedical applications of synchrotron radiation circular dichroism spectroscopy: identification of mutant proteins associated with disease and development of a reference database for fold motifs. *Faraday Discuss.* **126**, 237–243
19. 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
20. 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.* **30**, 522–523
21. 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.* **31**, 826–827
22. 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
23. Matsuo, K., Yonehara, R., and Gekko, K. (2004) Secondary-structure analysis of proteins by vacuum-ultraviolet circular dichroism spectroscopy. *J. Biochem.* **135**, 405–411
24. Woody, R.W. and Koslowski, A. (2002) Recent developments in the electronic spectroscopy of amides and α-helical polypeptides. *Biophys. Chem.* 101–102, 535–551
25. Wu, J., Yang, J.T., and Wu, C.S. (1992) Beta-II conformation of all-beta proteins can be distinguished from unordered form by circular dichroism. *Anal. Biochem.* **200**, 359–364
26. Sreerama, N. and Woody, R.W. (2003) Structural composition of β_I- and β_{II}-proteins. *Protein Sci.* **12**, 384–388
27. Greenfield, N. and Fasman, G.D. (1969) Computed circular dichroism spectra for the evaluation of protein conformation. *Biochemistry* **8**, 4108–4116
28. Johnson, W.C., Jr. and Tinoco, I., Jr. (1972) Circular dichroism of polypeptide solutions in the vacuum ultraviolet. *J. Amer. Chem. Soc.* **94**, 4389–4392
29. Brahms, S., Brahms, J., Spach, G., and Brack, A. (1977) Identification of β, β-turns and unordered conformations in polypeptide chains by vacuum ultraviolet circular dichroism. *Proc. Natl Acad. Sci. USA* **74**, 3208–3212
30. Sreerama, N. and Woody, R.W. (1994) Poly(pro)II helices in globular proteins: identification and circular dichroic analysis. *Biochemistry* **33**, 10022–10025
31. Compton, L.A. and Johnson, W.C., Jr. (1986) Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication. *Anal. Biochem.* **155**, 155–167
32. Chen, Y.-H., Yang, J.T., and Chau, K.H. (1974) Determination of the helix and beta form of proteins in aqueous solution by circular dichroism. *Biochemistry* **13**, 3350–3359
33. Chin, D.H., Woody, R.W., Rohl, C.A., and Baldwin, R.L. (2002) Circular dichroism spectra of short, fixed-nucleus alanine helices. *Proc. Natl Acad. Sci. USA* **99**, 15416–15421
34. King, S.M. and Johnson, W.C. (1999) Assigning secondary structure from protein coordinate data. *Proteins* **35**, 313–320
35. Brahms, S. and Brahms, J. (1980) Determination of protein secondary structure in solution by vacuum ultraviolet circular dichroism. *J. Mol. Biol.* **138**, 149–178
36. Johnson, W.C., Jr. (1986) Extending circular dichroism spectra into the vacuum UV and its application to proteins. *Photochem. Photobiol.* **44**, 307–313
37. Bandekar, J., Evans, D.J., Krimm, S., Leach, S.J., Lee, S., McQuie, J.R., Minasian, E., Nemethy, G., Pottle, M.S., Scheraga, H.A., Stimson, E.R., and Woody, R.W. (1982) Conformations of cyclo (L-alanyl-L-alanyl-epsilon-aminocaproyl) and of cyclo (L-alanyl-D-alanyl-epsilon-aminocaproyl); cyclized dipeptide models for specific types of beta-bend. *Int. J. Pept. Protein Res.* **19**, 187–205
38. 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
39. 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