

---

# Undistorted structural analysis of soluble proteins by attenuated total reflectance infrared spectroscopy

---

MICHEL E. GOLDBERG AND ALAIN F. CHAFFOTTE

Unité de Repliement et Modélisation des Protéines, Department of Structural Biology and Chemistry, Institut Pasteur, Paris, France

(RECEIVED June 30, 2005; FINAL REVISION August 22, 2005; ACCEPTED August 22, 2005)

## Abstract

Water from the solvent very strongly absorbs light in the frequency range of interest for studying protein structure by infrared (IR) spectroscopy. This renders handling of the observation cells painstaking and time consuming, and limits the reproducibility of the measurements when IR spectroscopy is applied to proteins in aqueous solutions. These difficulties are circumvented by the use of an Attenuated Total Reflectance (ATR) accessory. However, when protein solutions are studied, ATR spectroscopy suffers from several drawbacks, the most severe being nonproportionality of the observed absorbance with the protein concentration and spectral distortions that vary from protein to protein and from sample to sample. In this study, we show (1) that the nonproportionality is due to adsorption of the protein on the ATR crystal surface; (2) that the contribution of the crystal-adsorbed protein can easily be taken into account, rendering the corrected absorbance proportional to the protein concentration; (3) that the observed variable base line distortions, likely due to changes in the penetration depth of the light beam in solutions with the refractive index that depends on the protein concentration, can be easily eliminated; and (4) that ATR IR spectra thus corrected for protein adsorption and light penetration can be used to properly analyze the secondary structure of proteins in solution.

**Keywords:** ATR; FTIR; proteins; secondary structure; adsorption

**Supplemental material:** see [www.proteinscience.org](http://www.proteinscience.org)

Though X-ray crystallography and NMR provide powerful tools to reveal the details of protein structures, these methods are quite demanding and not applicable to proteins that fail to crystallize, or are too large to be amenable to NMR. Other methods, although providing less information on the three-dimensional structure of

proteins, are of easier and more general use, and provide important insights on the protein conformation. In particular, circular dichroism in the near-UV as well as infrared spectroscopy yield precious information on the secondary structure contents of proteins. These two methods are complementary in that circular dichroism provides good estimates of the number of residues that belong to  $\alpha$ -helices but fails to precisely predict the  $\beta$ -strand content, while infrared spectroscopy predicts  $\beta$ -strands much better than  $\alpha$ -helices. Yet, infrared spectroscopy is still scarcely used for protein structural studies, in spite of the considerable improvements to the speed of acquisition, precision, and quality of the spectra brought about by the advent of Fourier transform infrared spectroscopy (FTIR) (Jackson and Mantsch 1995; Backmann et al. 1996; Baello et al. 2000).

---

Reprint requests to: Michel E. Goldberg, Institut Pasteur, 28 rue du Dr. Roux, 75724 Paris Cedex 15, France; e-mail: [goldberg@pasteur.fr](mailto:goldberg@pasteur.fr); fax: +33-1-40-61-30-43.

*Abbreviations:* ATR, attenuated total reflectance; FKPA, a peptidyl-prolyl *cis-trans* isomerase from *Escherichia coli*; FTIR, Fourier transform infrared; MBP, the maltose binding protein from *Escherichia coli*; PDB, the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank; TRX, the oxidized thioredoxin from *Escherichia coli*.

Article and publication are at <http://www.proteinscience.org/cgi/doi/10.1110/ps.051678205>.

The main reason for this trailing of FTIR seems to be the practical difficulties encountered in applying it to proteins in aqueous solutions. Indeed, the information needed to predict the secondary structure contents from infrared spectra is contained in the amide I and amide II absorption bands of the peptide bonds (centered at  $\sim 1650$  and  $1550$   $\text{cm}^{-1}$ , respectively), which overlap the very strong absorption band of water in the  $1600$   $\text{cm}^{-1}$  region. This has two consequences. One is that the optical path used for FTIR measurements must be extremely short (a few microns only) to let enough light emerge from the sample cell. The second is that, in view of the very small concentration of peptide bonds compared to the water concentration ( $\sim 10^{-2}$  M peptide bonds vs. 111 M water O—H bonds in a 1 mg/mL protein solution), the protein signal is extremely small compared to the water signal, which requires the thickness and positioning of the sample cell to be rigorously the same for successive measurements. The conjunction of these two requirements results in severe technical difficulties. Indeed, only demountable cells can be practically used so that the cell windows can be efficiently cleaned after each protein sample. Since the exact path length of such cells depends on the strength put in tightening them upon reassembly, it is extremely difficult to keep the optical path strictly constant. This prompted the design of demountable (for cleaning purposes) cells that can be filled and emptied without disassembling them (for constant optical path purposes) for recording the sample and buffer spectra. But filling, emptying, cleaning, rinsing, and drying very thin cells are delicate and time-consuming operations since residual traces of liquid as well as air bubbles are not easy to avoid. To circumvent these difficulties, an "attenuated total reflectance" accessory (ATR) can be used instead of the usual transmission sample cell. With an ATR, rather than measuring the absorption of the incident light, one measures the absorption of the evanescent wave that penetrates the solution when the incident light is reflected at the crystal/liquid interface. Loading the ATR with a solution, emptying it, cleaning the sample compartment, rinsing, and drying it are extremely rapid and simple. Furthermore, because the ATR geometry is factory defined and remains perfectly constant, the "optical path" depends only on the crystal and solution refractive indices, thus rendering it perfectly constant for a given solution. Measurements are therefore highly reproducible. Moreover, for several commercially available ATRs with a circular well (as opposed to long crystals) only minute volumes of samples are needed since  $10$   $\mu\text{L}$  of solution amply suffice to cover the active surface of the ATR. These major advantages prompted attempts to use ATR/FTIR spectroscopy for structural studies of proteins (Raussens et al. 1998; Smith et al. 2002).

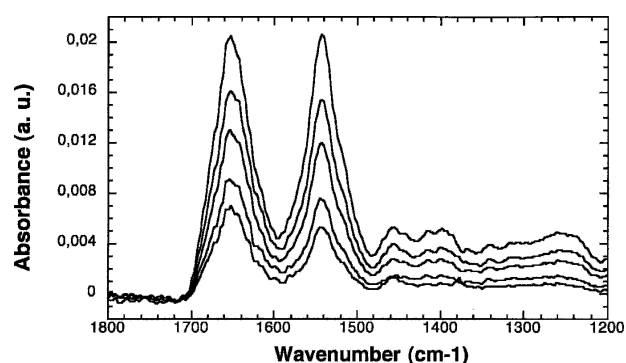
While starting FTIR studies on proteins, we indeed found it very easy and rapid to acquire spectra with an

ATR. However, we also found out that their interpretation is not straightforward and may be badly misleading. In this report, we shall show that the ATR/FTIR spectra as recorded are both biased and distorted, we shall offer a physical interpretation for the bias and distortions we observed, we shall propose a simple method to correct the observed spectra so as to render them reliable, and we shall show that the corrected spectra thus obtained allow for reliable secondary structure predictions.

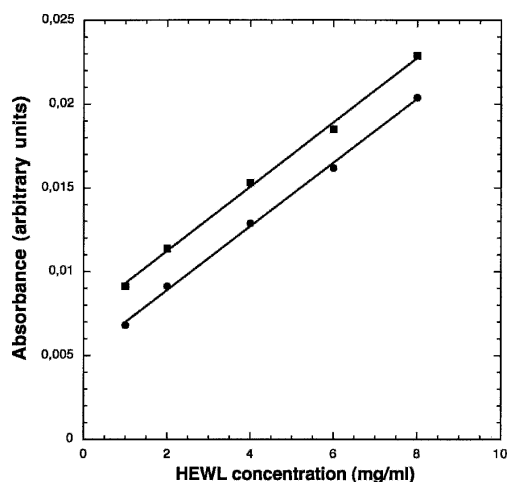
## Results

### *Concentration dependence of the ATR/FTIR spectra*

In order to determine the usable protein concentration range for spectral measurements and to assess the quality of the spectra obtained with our instrument, ATR/FTIR transmission spectra were recorded at different concentrations of hen egg white lysozyme, a protein chosen for its availability, stability, and well-characterized spectral and structural properties. The protein concentrations were 1, 2, 4, 6, and 8 mg/mL. The absorption spectra at these concentrations were constructed, using as a reference the transmission spectrum of the pure buffer (Fig. 1). Visual examination of the spectra revealed that the height of the amide I and amide II peaks were not proportional to the protein concentration. This was confirmed by plotting the observed absorbance at the amide I and amide II peak frequencies as a function of the protein concentration. The straight line thus obtained (Fig. 2) indicates the presence of two components to the absorbance. One is concentration



**Figure 1.** Crude ATR/FTIR absorption spectra of hen lysozyme. The transmission spectra of 0.01 M sodium phosphate buffer pH = 7.0 and of lysozyme solutions at different concentrations in the same buffer were recorded in single beam mode. The corresponding absorption spectra were constructed using the buffer spectrum as the 100% reference. For each protein concentration, the absorbance (in arbitrary units) is shown as a function of the light frequency. The lysozyme concentrations were 1, 2, 4, 6, and 8 mg/mL, respectively, for the spectra in the order of increasing peak height.



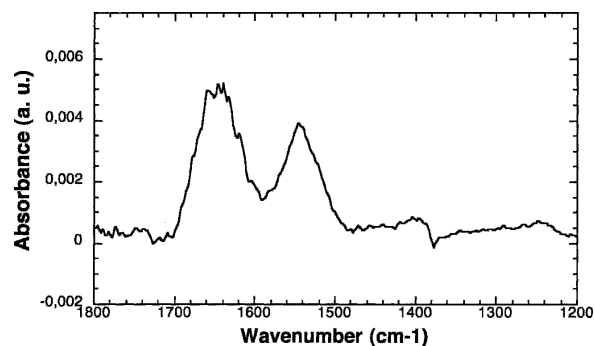
**Figure 2.** Peak values of the amide I and II bands as a function of lysozyme concentration. The values of the absorbances at the peaks of the amide I and II bands in the spectra of Figure 1 are shown as a function of the lysozyme concentration. Squares and circles correspond to the amide I and amide II bands, respectively. The straight lines were obtained by linear regression.

independent and corresponds to the absorbance obtained by extrapolation to zero protein concentration (0.0072 and 0.005 arbitrary units, respectively, for the amide I and amide II bands). The second is proportional to the protein concentration. For the amide I band, the absorbance of the concentration independent component corresponds to that of 3.5 mg/mL of the concentration-dependent component, as estimated from the slope of the amide I straight line. Similar experiments were conducted with bovine Ribonuclease A and horse muscle myoglobin and provided similar results, with the absorbance of the concentration-independent components corresponding to that of 3 and 5 mg/mL, respectively, of the concentration-dependent components. That the slopes of the straight lines obtained for the amide I and amide II bands are very similar, while the extrapolation to zero concentrations of the two lines do not coincide, suggests that the ratios of the amplitudes of the amide I and amide II bands are not the same for the concentration-dependent and concentration independent components.

#### *Nature of the concentration independent absorbing component*

It is part of the biochemists' common knowledge that proteins tend to stick to glass surfaces, and that at protein concentrations of less than 1 mg/mL the surface gets saturated with protein. Because all our ATR/FTIR spectra were recorded at protein concentrations of at least 1 mg/mL, we hypothesized that at all concentrations

investigated the diamond crystal of the ATR might behave as a glass surface and might be saturated with adsorbed protein that the evanescent wave would have to cross before penetrating the solution. Accordingly, the adsorbed protein would be responsible for the concentration independent part of the IR light absorption. To test this hypothesis, 20- $\mu$ L samples of lysozyme solutions at 1, 2, 4, and 8 mg/mL were incubated for 10 min in the ATR, which is amply sufficient to allow for adsorption of the protein onto the crystal surface (Santos et al. 2003). The solution was carefully removed from the ATR well by vacuum aspiration, 20  $\mu$ L of buffer were added, and the IR spectrum immediately recorded with only 100 accumulations to shorten the recording time and hopefully minimize possible protein desorption (Fig. 3). Although the signal-to-noise ratio was rather poor (because of a small signal and limited number of accumulations), the spectra thus obtained clearly showed the characteristics of a protein spectrum, with the two major bands at the amide I and amide II positions. This confirmed that some protein got strongly adsorbed to the crystal during the 10-min incubation of the lysozyme in the ATR. Moreover, as expected from our hypothesis, the amplitudes of the amide I and amide II bands were the same (within the noise) for the four concentrations investigated, and these amplitudes represented over 70% of the concentration-independent component of the absorption discussed in the previous section. This suggested that at least a major part of the adsorbed protein was strongly attached to the ATR crystal. To test the stability of the adsorption, the buffer that had been added after removal of the 8 mg/mL lysozyme solution was removed by aspiration and replaced by 20  $\mu$ L of fresh buffer, rinsing (i.e., aspiration followed by addition of 20  $\mu$ L of buffer) was repeated



**Figure 3.** ATR/FTIR absorption spectrum of crystal-adsorbed lysozyme. Lysozyme (20  $\mu$ L) at 8 mg/mL was incubated for 10 min in the ATR well, carefully removed by aspiration under vacuum, and replaced with 20  $\mu$ L of buffer. The transmission spectrum was recorded in single beam mode with 100 accumulations. The absorption spectrum shown was then constructed using the transmission spectrum of pure buffer as the 100% reference.

twice, and the spectrum recorded. Finally, rinsing was applied three more times and the spectrum recorded again. While the peak amplitude after the first three rinses was slightly reduced compared to the initial spectrum of the adsorbed material, the spectrum after the sixth rinse was unchanged compared to that after three rinses. Thus, about half of the adsorbed protein appeared very tightly bound to the crystal, while the other half could be removed by repeated rinsing. Both the loosely and tightly adsorbed protein molecules could be efficiently removed from the crystal surface by washing the ATR sample compartment twice with a soft detergent solution (2% Hellmanex) followed by thorough rinsing with distilled water, a procedure that lasts  $\sim 2$  min (see Materials and Methods). This was demonstrated by incubating a 10-mg/mL lysozyme solution in the ATR, applying the wash-and-rinse procedure, adding 20  $\mu$ L of buffer to the ATR and recording the spectrum, which showed no detectable absorption in the amide I–amide II region.

That protein adsorption onto a surface results, at saturating concentrations, in the formation of a protein monolayer is a widely accepted notion (for review, see Hlady et al. 1999). The IR absorbances we observed for the concentration-independent component are compatible with this view. Indeed, let us estimate roughly the concentration of protein in solution that would absorb the same amount of light as the monolayer of adsorbed protein. Assuming the thickness of the monolayer to be of the same order of magnitude as the diameter of the protein, it is expected to be  $\sim 20$ – $30$  Å for a small protein like lysozyme, myoglobin or ribonuclease. In order for the evanescent wave to undergo, in the solution, an absorbance equivalent to that caused by the adsorbed material, the protein concentration must be such that the volume occupied by the protein molecules that are on the pass of the evanescent wave in solution (beam diameter  $\times$  penetration depth  $\times$  fraction of the volume occupied by the protein) must be the same as the volume of adsorbed protein it crosses (beam diameter  $\times$  monolayer thickness). The penetration depth of the evanescent wave in the ATR accessory can be calculated (Griffiths and de Haseth 1986) to be  $\sim 0.88$  microns for a frequency corresponding to the amide I region. Therefore, for the evanescent wave to travel across the equivalent of 20 Å of protein molecules in the solution, the fraction of the volume occupied by the protein in the solution must be  $\sim 0.23\%$ – $0.34\%$  (i.e., 20/8800 and 30/8800, respectively). Assuming a standard value of 0.73 mL/g for the partial specific volume of the protein, this corresponds to a protein concentration of the order of 0.31%–0.46% (w/v), which compares very well with the protein concentrations of 0.3%–0.5% (i.e., 3–5 mg/mL) where the absorbance increment is equivalent to the absorbance of the adsorbed protein (see previous section).

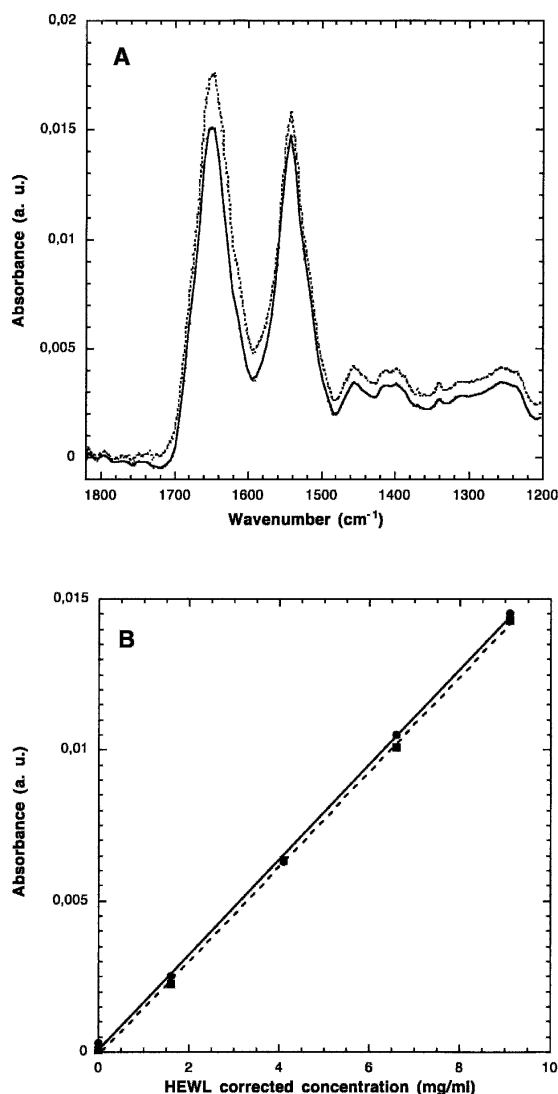
#### *Absorption spectrum of the protein in solution*

As shown above, the contribution of the crystal-adsorbed protein to the absorption spectrum is quite important. The experimental spectrum therefore does not correspond to the real spectrum of the protein in solution. This can be easily corrected by recording two spectra of the same protein, one at high (5–10 mg/mL), the second at low ( $\sim 0.5$ – $1$  mg/mL) protein concentration. Subtracting the absorption spectrum at low concentration from the absorption spectrum at high concentration eliminates the contribution of the adsorbed protein (since it is the same in the two samples) and yields the spectrum of the soluble protein at a concentration equal to the concentration difference between the two solutions. Figure 4A shows the absorption spectrum of lysozyme at 9.1 mg/mL obtained by subtracting the spectrum recorded at 0.9 mg/mL from that recorded at 10 mg/mL. When this was done for lysozyme solutions at various concentrations, the resulting spectra were superimposable when normalized with respect to the concentration. Furthermore, the peak absorbances obtained for the amide I and amide II bands were now strictly proportional to the “apparent” concentration (i.e., the difference between the high and the low concentrations used for recording the spectra) as shown in Figure 4B.

It is noteworthy that the spectrum of the adsorbed protein, determined by replacing the protein solution with buffer but omitting the washings (Fig. 3), differed from the corrected spectrum of the protein in solution in terms of the amide I and amide II band shapes, as well as in relative intensities of these two bands. This indicates that the adsorbed protein differs from the native protein in solution in terms of secondary structure contents, suggesting partial denaturation of the protein upon adsorption. Such post-adsorption changes in conformation were already reported for several proteins (Kondo and Mijara 1996; Zoungrana et al. 1997; Van Tassel et al. 1998; Santos et al. 2003). This emphasizes the need to clear the experimental spectra from the contribution of the nonnative adsorbed protein. The double concentration differential method we propose meets this requirement and renders the absorption measurements quantitative.

#### *Baseline adjustment*

Most of the absorption spectra obtained by this differential method (see, e.g., the continuous line in Fig. 4A) showed a baseline distortion: While the absorbance should be zero throughout the 1700–1850  $\text{cm}^{-1}$  region, it systematically showed a decrease with decreasing frequencies until the region of the amide I band was



**Figure 4.** Corrected ATR/FTIR absorption spectra of a lysozyme solution. The transmission spectra of buffer and of lysozyme solutions at 0.9, 2.5, 5, 7.5, and 10 mg/mL were successively recorded. A new recording was then made with the 0.9-mg/mL solution. The absorption spectra were constructed from the spectrum at each concentration using that at 0.9 mg/mL as the 100% transmission reference. (A) The spectrum constructed from the 10 mg/mL (corresponding to a 9.1-mg/mL solution) is shown in solid line. The spectrum distortion due to the protein refractive index increment was corrected as indicated in Materials and Methods. The corrected spectrum is shown in dotted lines. (B) The amide I (filled circles and solid line) and amide II (filled squares and dotted line) peak heights derived from the spectra corrected only for the adsorption are shown as a function of the lysozyme corrected concentration (i.e., the concentration at which the spectrum was recorded minus 0.9 mg/mL, the concentration of the solution used as the reference). The straight lines were obtained by linear regression.

reached. This unexpected behavior appeared the more pronounced for spectra acquired at the higher protein concentrations. From this, we inferred that this spurious absorption might be due to a very slight reduction in the

depth of penetration of the evanescent wave with increasing protein concentration, presumably due to an increase in the refractive index of the solution. This would result in a slight reduction of the water absorption in the experiment at high protein concentration compared to that at low protein concentration. Using the Spectral Subtract function of the PROTA software (Subtract.AB in GRAMS/32), a variable weight factor was assigned to the buffer versus air absorption spectrum (recorded in the ATR) and the weight factor was optimized so as to minimize the difference between the protein spectrum and the weighted buffer spectrum in the 1720–1850  $\text{cm}^{-1}$  region. Applying this procedure resulted in a corrected lysozyme spectrum (dotted line in Fig. 4A), which shows a flat and constant baseline. The usefulness of such spectra, corrected as indicated above for both the adsorbed protein and the baseline distortion, will now be considered.

#### Secondary structure analysis

Although corrected as described above, the ATR/FTIR spectra obtained still do not coincide with FTIR spectra obtained with a conventional transmission cell, since the ATR introduces some systematic spectral distortions. Some of them, like that caused by the variation of the penetration depth with the light frequency, can be corrected (Griffiths and de Haseth 1986). Others are not readily amenable to mathematical correction. However, these distortions are constant for a given ATR accessory and reproducible from spectrum to spectrum. Thus, differences in the ATR/FTIR spectra (once corrected as discussed above for protein adsorption and baseline distortion) of different proteins should quantitatively reflect differences in their conformations and serve for secondary structure prediction. Based on this reasoning, we constructed a library of ATR/FTIR corrected spectra for proteins of known three-dimensional structures and used it to predict the secondary structure contents of proteins not represented in the library. The transmission spectra of 29 reference proteins, each at low and high concentration, were acquired using the ATR with 1000 accumulations and at 4  $\text{cm}^{-1}$  resolution. For each protein, the absorption spectrum of the protein in solution was constructed, using the transmission spectrum of the diluted sample as a reference for the concentrated sample. The baseline distortion correction was applied as indicated above. Finally, the corrected spectrum of each protein was normalized and introduced in the data library as described in the Materials and Methods section. In order to test the validity of the database thus constructed, three proteins investigated in our laboratory (the maltose binding protein MBP, the peptidyl-prolyl *cis-trans*-isomerase FKPA and thioredoxin TRX

from *Escherichia coli*) were subjected to secondary structure predictions based on their FTIR spectra. Their spectra were acquired and treated as those of the database. The structural parameters, i.e., contents in  $\alpha$ -helices H, extended  $\beta$ -strands E, bends S, and hydrogen bonded turns T as defined by Kabsch and Sander (1983), were obtained from the PDB files indicated in Table 1. Secondary structure predictions were then achieved using an adapted (see Materials and Methods) version of the variable selection method (Manavalan and Johnson 1987). The reconstructed spectrum corresponding to the best fit is shown, for each protein, as a solid line in Figure 5. While the fits seem reasonably good for the maltose binding protein and the FKPA isomerase, it significantly deviates from the experimental data in the region of the amide I band peak for thioredoxin. This rather strong deviation of the thioredoxin fit is confirmed by the high value of its root mean square deviation (0.0222) as indicated in Table 1. Table 1 also shows the secondary structure predictions corresponding to the best fit for each test protein, together with their secondary structure features assessed from their known 3D structures. The comparison between the predicted versus the assessed secondary structure contents indicates that the secondary structure contents of the three proteins were predicted with fairly good accuracy from the ATR/FTIR corrected spectra, in particular for the extended  $\beta$ -

strands of the two first proteins, while their  $\alpha$ -helix, bend and turn contents predictions were less accurate as often reported with transmission cell FTIR.

#### *Compared accuracies of ATR and transmission cell based secondary structure predictions*

In order to compare the value of transmission cell versus ATR/FTIR spectroscopy for secondary structure predictions, each of the 20 proteins common to the ATR spectra library we constructed and to the transmission cell spectra library provided with the PROTA software were subjected to prediction based on both libraries. To this effect, the spectra of one of these proteins was removed from the two libraries, and submitted to the secondary structure prediction procedure (to predict the secondary structure of the corresponding protein) with the spectra remaining in one or the other two libraries. This procedure was repeated for each protein common to both libraries. Table 2 shows the predicted secondary structure contents of these 20 proteins using the ATR library as well as their contents assessed from their 3D structure in the PDB. Figure 6 shows, for each database, the plot of the  $\beta$ -strand contents of the 20 proteins assessed versus their predicted contents. For the set of data points obtained with each database, a linear regression was applied. The ordinates at zero abscissa were  $-0.015$  and  $0.026$  for the ATR and transmission cell database, respectively, the slopes were  $1.075$  and  $0.937$ , respectively, and the correlation coefficients were  $0.896$  and  $0.642$ , respectively. Thus, in terms of ordinate at zero abscissa and of slopes, the two fits compare equally well with the expected values (0 and 1, respectively), but the ATR database provides a much better correlation coefficient which reflects a significantly more robust prediction of the  $\beta$ -strand contents with the ATR measurements compared to the conventional transmission cell FTIR measurements. The same method was used to analyze the spectra of the two databases, and the two databases contain the spectra of the same set of proteins. Thus, the better correlation obtained with the ATR—compared to the transmission—database likely reflects larger experimental errors introduced in recording the spectra with transmission cells (used in PROTA) rather than with the ATR.

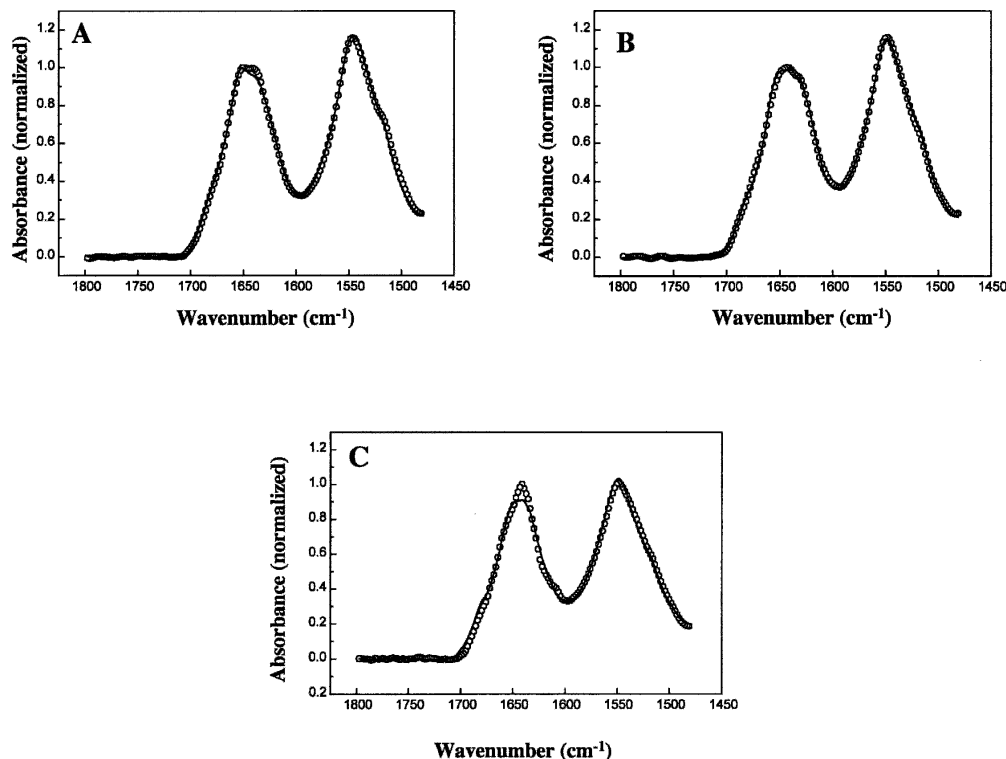
#### Discussion

The results reported in this paper demonstrate that FTIR spectra recorded with an ATR accessory suffer from two types of biases: one is related to the adsorption of proteins onto the ATR crystal, the other to the properties of the evanescent wave at the crystal/solution interface. It is shown that, at protein concentrations

**Table 1.** Secondary structure prediction for three test proteins

	H	E	S	T	O	Sum	RMS
MBP	0.36	0.16	0.08	0.17	0.24	1.01	0.0149
<i>PDB 1LLS</i>	<i>0.43</i>	<i>0.16</i>	<i>0.09</i>	<i>0.13</i>	<i>0.19</i>	<i>1.00</i>	
FKPA	0.41	0.20	0.12	0.8	0.25	1.06	0.0119
<i>PDB 1Q6U</i>	<i>0.35</i>	<i>0.155</i>	<i>0.04</i>	<i>0.15</i>	<i>0.305</i>	<i>1.00</i>	
TRX	0.30	0.18	0.11	0.8	0.31	0.98	0.0222
<i>PDB 2TRX</i>	<i>0.26</i>	<i>0.255</i>	<i>0.05</i>	<i>0.24</i>	<i>0.195</i>	<i>1.00</i>	

The transmission spectra of the *E. coli* maltose binding protein, FKPA *cis-trans* isomerase, and oxidized thioredoxin were recorded both at high (16.5, 10.0, and 10.4 mg/mL, respectively) and low (1.65, 1.0, and 1.04 mg/mL) protein concentrations, respectively. The absorption spectrum of each protein was calculated taking the transmission at low protein concentration as the 100% reference. The spectra were corrected for water distortion, normalized in amplitude, and analyzed with the adapted VARSELEC method selecting the best set of 26 spectra among the 29 present in the FTIR/ATR database (3654 combinations). The results (upper line for each protein) are shown as the fraction of residues in each secondary structure type and compared to the values assessed from the PDB (lower line for each protein, in italics). The names of the PDB files used are indicated in the table under the protein name. The assessed values indicated for thioredoxin are the averages of those reported for the two chains present in the unit cell, for which the helical contents (particularly  $\alpha$ -helices and 3-10 helices) are strikingly different. The secondary structure type symbols are H for  $\alpha$ -helices, E for extended  $\beta$ -strands, S for bends, T for hydrogen bonded turns, and O for others. "Sum" indicates the sum of the relative contents of all structural types. RMS is the root-mean-square deviation calculated for the best fit.



**Figure 5.** Spectral analysis of test proteins. The corrected ATR/FTIR absorption spectra of *E. coli* maltose binding protein (A), FKPA prolyl *cis-trans* isomerase (B), and thioredoxin (C) were constructed and analyzed using the adapted VARSELEC method selecting the best set of 26 spectra among the 29 (spectra) of the complete database. For each protein, the open circles show the experimental data and the solid line corresponds to the theoretical spectrum reconstructed from the best fit.

routinely used for FTIR measurements, the light absorption by the protein adsorbed on the crystal represents a significant fraction of the light absorption by the protein in solution. This is due to the fact that, at such concentrations, the protein covers the crystal surface with a molecular monolayer that the evanescent wave has to cross before penetrating the solution. While the adsorption of proteins at interfaces has been investigated in some details by means of attenuated total reflectance using various detection methods (Hlady et al. 1999; Santos et al. 2003), its influence on FTIR spectra as recorded with an ATR has been generally overlooked (see, e.g., Smith et al. 2002). In this report, it is shown that (1) the crystal-adsorbed protein biases the quantitative estimate of protein concentrations from the FTIR spectra; (2) it also distorts the shape of the spectra since even a protein as stable as hen lysozyme can undergo post-adsorption changes in conformation. Therefore, reliable quantitative measurements as well as secondary structure predictions based on FTIR spectra must eliminate the contribution of the crystal-adsorbed protein to the IR absorption. As shown in this report, this can be easily done by recording the transmission spectra of a concentrated and a diluted protein solution, and cal-

culating the absorption spectrum using the spectrum of the diluted solution as a reference. It should be emphasized that, while the contribution of the adsorbed protein may sometimes be considered as negligible if the FTIR spectra are recorded at very high protein concentrations, it can by no means be overlooked in spectra recorded at protein concentrations usually used for FTIR measurements. Indeed, at protein concentrations of  $\sim 3\text{--}5$  mg/mL, the contribution of the adsorbed protein is as large as that of the soluble protein.

FTIR spectra obtained by use of an ATR are also distorted by a variety of factors influencing the behavior of the evanescent wave. One of them is the variation of the penetration depth with the protein concentration, which results in a difference in the light absorption by water for the concentrated and diluted solutions. This difference is particularly important in the amide I–amide II region, which coincides with a very strong absorption band of water. As shown above, this can be easily corrected by adding to the protein spectrum a water spectrum weighted so as to minimize the absorbance of the protein solution in the  $1720\text{--}1850$   $\text{cm}^{-1}$  region, where the protein absorbance should be zero. It should be noted that, though the physical cause of the spectrum

**Table 2.** Secondary structure prediction for each protein of the database

	H	E	S	T	O	Sum	RMS
Alc. DH	0.27	0.26	0.08	0.14	0.26	1.01	0.0110
<i>PDB 1YE3</i>	<i>0.23</i>	<i>0.22</i>	<i>0.10</i>	<i>0.14</i>	<i>0.31</i>	<i>1.00</i>	
Carban	0.17	0.31	0.11	0.14	0.29	1.01	0.0139
<i>PDB 1CA2</i>	<i>0.08</i>	<i>0.29</i>	<i>0.12</i>	<i>0.13</i>	<i>0.38</i>	<i>1.00</i>	
Catalase	0.23	0.25	0.15	0.13	0.31	1.06	0.0189
<i>PDB 7CAT</i>	<i>0.26</i>	<i>0.14</i>	<i>0.12</i>	<i>0.155</i>	<i>0.325</i>	<i>1.00</i>	
Cit.Synth	0.57	0.04	0.05	0.18	0.16	0.99	0.0098
<i>PDB 1CTS</i>	<i>0.54</i>	<i>0.015</i>	<i>0.065</i>	<i>0.14</i>	<i>0.24</i>	<i>1.00</i>	
Conca. A	-0.14	0.45	0.08	0.12	0.29	0.94	0.0259
<i>PDB 1NLS</i>	<i>0</i>	<i>0.465</i>	<i>0.10</i>	<i>0.125</i>	<i>0.31</i>	<i>1.00</i>	
Chtgen	0.20	0.29	0.08	0.13	0.27	0.98	0.0216
<i>PDB 2CGA</i>	<i>0.07</i>	<i>0.29</i>	<i>0.10</i>	<i>0.175</i>	<i>0.365</i>	<i>1.00</i>	
Cyto. c	0.41	0.03	0.14	0.13	0.24	0.96	0.0222
<i>PDB 1AKK</i>	<i>0.34</i>	<i>0</i>	<i>0.17</i>	<i>0.14</i>	<i>0.35</i>	<i>1.00</i>	
Enolase	0.40	0.15	0.12	0.12	0.26	1.04	0.0074
<i>PDB 1EBH</i>	<i>0.38</i>	<i>0.16</i>	<i>0.07</i>	<i>0.13</i>	<i>0.26</i>	<i>1.00</i>	
Hemoglobin	0.61	-0.01	0.07	0.14	0.17	0.99	0.0092
<i>PDB 1BBB</i>	<i>0.67</i>	<i>0</i>	<i>0.06</i>	<i>0.11</i>	<i>0.16</i>	<i>1.00</i>	
Lysozyme	0.49	0.07	0.08	0.15	0.20	0.99	0.0137
<i>PDB 6LYZ</i>	<i>0.30</i>	<i>0.06</i>	<i>0.07</i>	<i>0.28</i>	<i>0.29</i>	<i>1.00</i>	
Hexokinase	0.42	0.10	0.10	0.12	0.26	1.00	0.0122
<i>PDB 2YHX</i>	<i>0.39</i>	<i>0.15</i>	<i>0.135</i>	<i>0.09</i>	<i>0.235</i>	<i>1.00</i>	
Myoglobin	0.54	0.07	0.07	0.11	0.22	1.01	0.0101
<i>PDB 1WLA</i>	<i>0.74</i>	<i>0</i>	<i>0.025</i>	<i>0.13</i>	<i>0.105</i>	<i>1.00</i>	
Ovalbumin	0.24	0.25	0.12	0.16	0.31	1.08	0.0102
<i>PDB 1OVA</i>	<i>0.295</i>	<i>0.29</i>	<i>0.05</i>	<i>0.15</i>	<i>0.215</i>	<i>1.00</i>	
Pepsin	0.16	0.34	0.11	0.13	0.27	1.01	0.0246
<i>PDB 4PEP</i>	<i>0.11</i>	<i>0.41</i>	<i>0.08</i>	<i>0.14</i>	<i>0.26</i>	<i>1.00</i>	
Pepsinogen	0.20	0.33	0.09	0.12	0.27	1.01	0.0216
<i>PDB 3PSG</i>	<i>0.12</i>	<i>0.36</i>	<i>0.06</i>	<i>0.13</i>	<i>0.33</i>	<i>1.00</i>	
RNase A	0.10	0.25	0.13	0.17	0.35	0.99	0.0161
<i>PDB 3RN3</i>	<i>0.18</i>	<i>0.33</i>	<i>0.095</i>	<i>0.19</i>	<i>0.205</i>	<i>1.00</i>	
RNase S	0.07	0.30	0.10	0.23	0.32	1.01	0.0215
<i>PDB 2RNS</i>	<i>0.18</i>	<i>0.31</i>	<i>0.09</i>	<i>0.105</i>	<i>0.315</i>	<i>1.00</i>	
Sup. dismut	0.15	0.29	0.13	0.13	0.34	1.05	0.0207
<i>PDB 2SOD</i>	<i>0</i>	<i>0.355</i>	<i>0.30</i>	<i>0.05</i>	<i>0.295</i>	<i>1.00</i>	
Thaumatococin	-0.04	0.36	0.16	0.11	0.38	1.00	0.0075
<i>PDB 1RQW</i>	<i>0.105</i>	<i>0.31</i>	<i>0.17</i>	<i>0.145</i>	<i>0.27</i>	<i>1.00</i>	
Triose-P. I.	0.32	0.16	0.11	0.11	0.30	0.99	0.0092
<i>PDB 1YPI</i>	<i>0.38</i>	<i>0.16</i>	<i>0.085</i>	<i>0.11</i>	<i>0.265</i>	<i>1.00</i>	

For each protein that is represented in both the ATR/FTIR database and the PROTA database, the spectrum was removed from the database and the secondary structure of the corresponding protein was predicted by means of the adapted VARSELEC method, selecting the best set of 25 spectra among the 28 remaining in the database (3276 combinations). The results (upper line) are shown as percentage of residues in each secondary structure type and compared to the values assessed from the PDB (lower line, in italics). For each protein, the name of the PDB file used is indicated in the table under the protein abbreviation. The abbreviations are Alc. DH, alcohol dehydrogenase; Carban, carbonic anhydrase; Cit. Synth, citrate synthase; Conca. A, concanavalin A; Chtgen, chymotrypsinogen; Cyto. C, cytochrome c; Sup. dismut, superoxide dismutase; Triose-P. I, triose phosphate isomerase. The symbols for secondary structure types are as in Table 1.

distortion is different when transmission cells are used (irreproducibility of the cell thickness rather than influence of the protein refractive index on the evanescent

wave penetration depth), the same correction procedure is routinely applied in conventional IR spectroscopy.

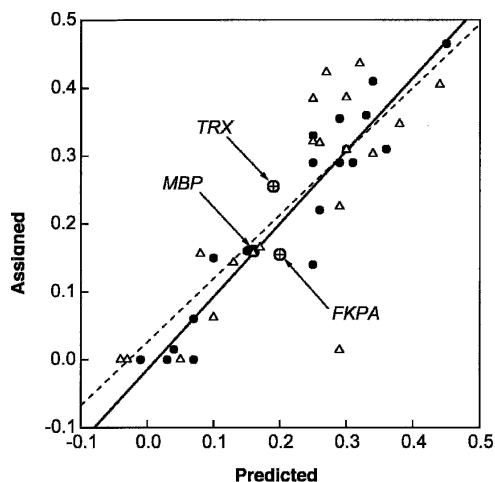
A second factor influencing the absorption of the evanescent wave is related to the change in the penetration depth of the evanescent wave with the light frequency. This affects all spectra recorded by means of an ATR, regardless of the solute and its concentration. Because the penetration depth is proportional to the wavelength (assuming the refractive indices of the ATR crystal and of the solution to be frequency independent, which is not rigorously correct), the penetration depth can be calculated (see formula in Griffiths and de Haseth 1986) to be ~8%–9% longer at the wavelength of the amide II band than at that of the amide I band, thus increasing artificially the apparent intensity of the amide II, compared to amide I, band.

Another important source of spectral distortion is the so-called “anomalous dispersion.” This effect was shown (Grdadolnik 2002) to account for 10%–15% of the amide II signal observed for bovine serum albumin in an ATR. Taken together, the penetration depth dependence on the wavelength and the anomalous dispersion thus account for a 20%–25% overestimate of the amide II band intensity compared to the amide I. These two effects account for the major part, if not the totality, of the lower ratio of the amide I to amide II amplitudes observed in the ATR (~0.9–1.1) compared to a classical transmission cell (~1.3–1.4).

Though a mathematical formula (Griffiths and de Haseth 1986) can easily correct the experimental ATR spectra for the wavelength dependence of the penetration depth, and a procedure was developed to correct for the anomalous dispersion (Grdadolnik 2002), these corrections were not applied in the present study because the same solvent was used throughout our experiments, resulting in the same spectral distortions for all the samples analyzed, and because systematic spectral distortions, i.e., that are the same for the proteins under investigation and for the proteins in the database, do not affect the spectrum analysis procedure we used. Thus, FTIR spectra recorded using an ATR accessory, once corrected for the crystal-adsorbed protein and for the distortion caused by the protein refractive index increment, can be readily used for secondary structure predictions, as documented by the results in Table 2. Indeed, the robustness of secondary structure predictions from the spectra recorded with the ATR appears better than that of predictions made from conventional spectra recorded with transmission cells, as indicated by a much better correlation coefficient for the ATR data compared to the transmission cell data shown in Figure 6.

The fit between predicted and assessed secondary structure contents (see Tables 1,2) is far from being perfect. For some proteins, the discrepancy between the predicted and the assessed values is rather high. This





**Figure 6.** Compared robustness of the ATR and transmission cell FTIR-based predictions of  $\beta$ -strand contents. The  $\beta$ -strand content of each protein represented in both the ATR and PROTA FTIR spectra database were predicted by removing its spectrum from each database and analyzing either the ATR or the transmission cell FTIR spectrum with the adapted VARSELEC program using the cognate reduced database. The  $\beta$ -strand content assessed from the PDB is plotted vs. the predicted content. The open triangles and dotted line correspond to the PROTA-based prediction, and the closed circles and solid line correspond to the ATR-based prediction. The two straight lines were obtained by linear regression. The crossed circles correspond to the three test proteins indicated by the arrows.

may be due in part to the fact that the IR absorption bands of several amino acid side chains overlap the amide I and II bands, and thus contribute to the part of the IR absorption spectra used in the spectral analysis. Careful spectral analysis should therefore subtract the side chain contributions from the spectra of the unknown proteins as well as from those of the database. This could, in principle, be done by recording the spectra of the side chains alone, computing a “side-chain spectrum” from the known amino acid composition of each protein, and subtracting it from the experimental spectrum. It is likely that such side-chain corrections, included in the PROTA software for the correction of conventional transmission cell spectra, would improve the robustness of the predictions. Another way to improve the secondary structure prediction from the FTIR spectra might be to include more secondary structure types in the analysis. Indeed, the analysis reported here was based on only the following structural features (as defined by Kabsch and Sander 1983):  $\alpha$ -helices, extended  $\beta$ -strands, hydrogen bonded turns and bends, all residues not belonging to one or the other of these four types being qualified as “others.” Including 3–10 helices,  $\pi$ -helices and/or isolated  $\beta$ -bridges in the database might improve the overall secondary structure predictions. The limited number of secondary structure

types we used might account for the rather poor prediction of the  $\beta$ -strand content of *E. coli* thioredoxin, for which the 3D structure reported in the PDB indicates the presence of six residues (i.e., 5.5%) belonging to 3–10 helices in one of the two chains per unit cell. It is plausible that this relatively high 3–10 helix content is responsible for the impossibility to obtain a better fit of the theoretical spectrum of thioredoxin to the experimental data (Fig. 5C), and hence, better secondary structure predictions. Considering the 3–10 helices as a distinct type of secondary structure (while in our analysis these helices were included in the “other” type), or grouping all the helices ( $\alpha$ ,  $\pi$ , and 3–10) in the same secondary structure type, could be easily done through minor modifications of the VARSELEC program, and might improve the robustness of the secondary structure prediction based on FTIR spectra.

There exists a variety of procedures, other than VARSELEC, that have been developed for determining the secondary structure of proteins from FTIR spectra (for reviews, see Surewicz and Mantsch 1988; Arrondo et al. 1993). Some are based on band decomposition and curve fitting including frequency and/or bandshape analysis (Cameron et al. 1982; Byler and Susi 1986; Goormaghtigh et al. 1990), others are based on the use of calibration sets (Dousseau and Pézolet 1990; Lee et al. 1990; Sarver and Krueger 1991a). A systematic comparison of their accuracy in predicting the secondary structure from FTIR spectra on the basis of data reported in the literature is not an easy task. Indeed, the studies reported rely on different procedures, but also on different databases, which makes it difficult to assess which of the procedures or the databases is responsible for the quality of the results. However, based on our experience with CD, the most critical point for obtaining good predictions from CD spectra comparison is the quality of the experimental data (both that of the sample to analyze and of the proteins in the database) as well as the proper choice of the proteins in the database to be taken into account, the latter being optimized by the Variable Selection procedure (Manavalan and Johnson 1987). We therefore chose, in our study, the VARSELEC program, which is based on the “singular value decomposition” method initially proposed for CD by Hennessy and Johnson (1981), and includes the Variable Selection procedure (Manavalan and Johnson 1987). In spite of the fact that running it is much less tedious and time consuming than using VARSELEC, the program contained in the PROTA package, which is based on the “principal component factor analysis” method (Pancoska et al. 1991), was not retained because it does not include the variable selection procedure. Indeed, when tested on individual spectra extracted from the PROTA database using either the PROTA or the VARSELEC algorithm, VARSELEC provided a

better secondary structure prediction, justifying further that the VARSELEC procedure was preferred.

It has been reported that improved predictions can be obtained, using a principal component regression-based computer program that includes an "inside model space" bootstrap to take into account the variability of the amide I and amide II bands introduced by the various levels of hydration of the amide bonds (Smith et al. 2002). This is of particular interest in view of the significant differences in the absorption of exposed and buried amide groups that are in  $\alpha$ -helices (Walsh et al. 2003). There also exists a variety of experimental approaches that were reported to improve the reliability and precision of secondary structure predictions from FTIR spectroscopy. Thus, coupling FTIR with circular dichroism (Sarver and Krueger 1991b), or including hydrogen exchange in the FTIR measurements (Baello et al. 2000) were shown to improve the secondary structure predictions. None of these approaches was attempted in the present study since our goal was limited to finding out whether or not ATR/FTIR can be substituted for transmittance FTIR for the sake of simplifying and eventually improving data collection. The results we obtained clearly indicate that, provided appropriate caution is exerted in collecting and analyzing the data, the use of an ATR accessory leads to secondary structure predictions that are at least as good as those obtained by conventional transmission cell spectroscopy, with the considerable advantage that sample handling is dramatically easier, more reproducible, and much faster. It is likely that, based on this conclusion, the use of ATR/FTIR spectroscopy will become of more widespread use in investigations on protein structure.

## Materials and methods

### Proteins

The proteins chosen to construct the database were alcohol dehydrogenase from horse liver, carbonic anhydrase from bovine erythrocyte, catalase from bovine liver, citrate synthase from porcine liver, bovine  $\alpha$ -chymotrypsinogen, concanavalin A, cytochrome *c* from horse liver, enolase from baker yeast, glyceraldehyde-3-phosphate dehydrogenase from rabbit muscle, human hemoglobin, hexokinase from baker yeast, soybean trypsin inhibitor, bovine pancreatic trypsin inhibitor, human  $\alpha$ -lactalbumin, lactate dehydrogenase from rabbit muscle, bovine  $\beta$ -lactoglobulin, hen egg white lysozyme, myoglobin from horse muscle, hen ovalbumin, pepsin from porcine stomach, porcine pepsinogen, horse radish peroxidase, phosphoglycerate kinase from yeast, bovine ribonucleases A and S, human serum albumin, superoxide dismutase from bovine erythrocyte, thaumatin from *Thaumatococcus daniellii*, and triose phosphate isomerase from baker yeast. They were all obtained from Sigma-Aldrich in the highest purity grade available and used without further purification. The test proteins were the *E. coli* maltose binding protein MBP and prolyl *cis-trans*-

isomerase FKPA, both purified in our unit by Dr. J.-M. Betton, and thioredoxin purified and kindly provided by the laboratory of J. Beckwith (Harvard); 0.5 to 1 mL of a solution of each protein, prepared at a concentration expected to be  $\sim 10$  mg/mL, was dialyzed for at least 15 h at 4°C against 500 mL of 0.01 M sodium phosphate pH 7.0 (except for pepsin, where the pH was 5.5). The protein concentration in the dialysate was determined by spectrophotometry, using either published specific extinction coefficients, or extinction coefficients calculated from the amino acid composition according to Pace et al. (1995). A 10-fold dilution of the dialysate in the dialysis buffer was prepared and used for the spectrophotometric concentration determination as well as to serve as the 100% transmission reference solution for the FTIR measurements (see below).

### Acquisition of ATR/FTIR spectra

The FTIR spectra were recorded on a PROTA FTIR Protein Analyzer from ABB-Bomem, which corresponds to an ABB-Bomem MB104 infrared spectrometer supplemented with software specifically dedicated to the analysis of infrared spectra of proteins. The PROTA software is based on the GRAMS program from Galactic Industries Corporation and includes, among other features, a database containing the FTIR absorption spectra (obtained by using conventional transmission cells) of 33 proteins of known 3D structures, a routine allowing one to subtract the contribution of the protein amino acid side chains from its absorption spectrum, and a program to predict the secondary structure contents of a protein from its IR spectrum by comparison to the database. The spectrophotometer was equipped with a DuraSamplIR II, 9 reflection diamond ATR accessory from SensIR Technologies, and with a conventional, low-sensitivity, DTGS Deuterated Triglycine Sulfate (DTGS) detector. Higher sensitivity, improved signal-to-noise ratios and shorter accumulation times would have been obtained using a narrow range Mercury Cadmium Telluride (MCT) detector. The spectrophotometer was set up in a room where the temperature was maintained at  $20^\circ \pm 2^\circ\text{C}$ . The volume of the sample deposited in the ATR well was 20  $\mu\text{L}$ . The ATR well was covered with a slightly convex glass cover (supplied with the ATR) to prevent solvent evaporation. All spectra were acquired in the single beam mode with a  $4\text{-cm}^{-1}$  resolution. For all preliminary experiments (unless otherwise stated) the data from 300 scans were accumulated and a buffer transmission spectrum was systematically recorded (300 scans) before and after each series of measurements. For the construction of the database as well as for the analysis of the three test proteins, accumulation was over 1000 scans and took  $\sim 50$  min for each spectrum. The spectra of the dialysate and its 10-fold dilution were recorded one after the other without delay so as to minimize a possible drift in the output of the light source. The absorption spectrum of each solution was constructed from its single beam spectrum using the appropriate reference spectrum and the Absorbance function of the PROTA software. Uncorrected absorption spectra were constructed using the buffer single beam spectrum as a reference. Absorption spectra corrected for protein adsorption on the ATR crystal were constructed from the spectrum of the dialysate using the spectrum of the 10-fold diluted dialysate as the reference. The baseline distortion resulting from the refractive index increment due to the protein was corrected by means of the Spectral Subtract function of the PROTA software, using as subtrahend the absorption spectrum of the buffer (obtained

from the single beam spectra of buffer and air, i.e., with the empty ATR) and the 1850–1720  $\text{cm}^{-1}$  frequency range for the minimization. Unless otherwise stated, after each spectrum recording, the ATR well was emptied by means of an Eppendorf pipette tip connected to a vacuum pump. Next, the ATR window was washed as follows: 50  $\mu\text{L}$  of a 2% Hellmanex (Hellma GmbH) detergent solution were introduced in the ATR well. After 1 min, the ATR window was thoroughly rubbed with the extremity of a cotton tip previously soaked in the detergent solution. The excess detergent was removed from the compartment by aspiration and the whole washing operation was repeated a second time. The ATR well was then thoroughly rinsed by flushing it with distilled water with simultaneous aspiration of the excess water. The window and well were then carefully dried by aspiration of all traces of residual water.

### Estimation of secondary structure

The secondary structure contents of the proteins were estimated using the single value decomposition with the variable selection (VARSELEC) procedure proposed by Manavalan and Johnson (1987). Two databases were used, one consisting in the 33 reference spectra included in the PROTA software, the second consisting in the 29 protein spectra acquired with the ATR accessory. In all cases, the spectra were normalized to 1 at the peak of the amide I band and only the 165 values corresponding to the spectral range between 1797 and 1481  $\text{cm}^{-1}$  were taken into account for the analysis. Four structural types were considered:  $\alpha$ -helix, extended  $\beta$ -strands, bends, and H-bonded turns. The fraction of residues belonging to each structural type was assessed from the DSSP analysis (Kabsch and Sander 1983) of the PDB entry corresponding to each protein. For each secondary structure prediction, the best fit was selected from the combination(s) of spectra corresponding to the smallest RMS value and the sum of the structural fractions (between 0.90 and 1.10) closest to 1.0. In order to test the reliability of the secondary structure determination with either the PROTA database or the database we constructed, each of the 20 spectra of the reference proteins common to both databases were submitted to the VARSELEC procedure after its removal from the corresponding database.

The PDB files used for assessing the secondary structure contents in the ATR database are as follows: alcohol dehydrogenase, 1YE3; carbonic anhydrase, 1CA2; catalase, 7CAT; citrate synthase, 1CTS;  $\alpha$ -chymotrypsinogen, 2CGA; concanavalin A, 1NLS; cytochrome *c*, 1AKK; enolase, 1EBH; glyceraldehyde-3-phosphate dehydrogenase, 1J0X; hemoglobin, 1BBB; hexokinase, 2YHX; soybean trypsin inhibitor, 1AVU; bovine pancreatic trypsin inhibitor, 4PTI;  $\alpha$ -lactalbumin, 1HFZ; lactate dehydrogenase, 6LDH;  $\beta$ -lactoglobulin, 1BEB; lysozyme, 6LYZ; myoglobin, 1WLA; ovalbumin, 1OVA; pepsin, 4PEP; pepsinogen, 3PSG; peroxidase, 1HCH; phosphoglycerate kinase, 3PGK; ribonuclease A, 3RN3; ribonuclease S, 2RNS; serum albumin, 1AO6; superoxide dismutase, 2SOD; thaumatin, 1RQW; and triose phosphate isomerase, 1YPI.

### Electronic supplemental material

The information provided in the electronic supplementary material consists in three files, all in Microsoft Word. One, named INSTRUCTIONS FOR USE.doc, is a complete description of the way in which (1) the data corresponding to

the ATR/FTIR spectra corrected for protein adsorption and water-related baseline distortion should be formatted for use in the VARSELEC program, and (2) the various files introduced in the VARSELEC program have to be adapted for the analysis of FTIR spectra. A second file, S.DTA, contains the spectra of the 29 proteins of the ATR/FTIR database formatted for use in the VARSELEC program. The last file, named SS.DTA, contains the secondary structure content of the 29 proteins of the ATR/FTIR database, formatted for use in the VARSELEC program.

### Acknowledgments

This paper is dedicated to Robert L. Baldwin in recognition of his precious contribution, as Ph.D. advisor 40 years ago, in making one of us (M.E.G.) aware of the risk of protein adsorption to glass surfaces. This research was supported by grants from the Institut Pasteur, the CNRS (National Center for Scientific Research, URA 2185) and the French Ministry of Research and Technology (ACI Prions, SDV 03/1347).

### References

- Arrondo, J.L.R., Muga, A., Castresana, J., and Goni, F.M. 1993. Quantitative studies of the structure of proteins in solution by Fourier-transform infrared spectroscopy. *Prog. Biophys. Mol. Biol.* **59**: 23–56.
- Backmann, J., Schultz, C., Fabian, H., Hahn, U., Saenger, W., and Naumann, D. 1996. Thermally induced hydrogen exchange processes in small proteins as seen by FTIR spectroscopy. *Proteins* **24**: 379–387.
- Baello, B.I., Pancoska, P., and Keiderling, T.A. 2000. Enhanced prediction accuracy of protein secondary structure using hydrogen exchange Fourier transform infrared spectroscopy. *Anal. Biochem.* **280**: 46–57.
- Byler, D.M. and Susi, H. 1986. Examination of the secondary structure of proteins by deconvolved FTIR spectra. *Biopolymers* **25**: 469–487.
- Cameron, D.G., Kauppinen, J.K., Moffat, D.J., and Mantsch, H.H. 1982. Precision in condensed phase vibrational spectroscopy. *Appl. Spectrosc.* **36**: 245–250.
- Dousseau, F. and Pézolet, M. 1990. Determination of the secondary structure content of proteins in aqueous solutions from their amide I and amide II infrared bands. Comparison between classical and partial least-squares methods. *Biochemistry* **29**: 8771–8779.
- Goormaghtigh, E., Cabiaux, V., and Ruyschaert, J.M. 1990. Secondary structure and dosage of soluble and membrane proteins by attenuated total reflection Fourier-transform infrared spectroscopy on hydrated films. *Eur. J. Biochem.* **193**: 409–420.
- Grdadolnik, J. 2002. An attenuated total reflection infrared spectroscopy of water solutions. *Int. J. Vib. Spec.* **6**: 2–6.
- Griffiths, P.R. and de Haseth, J. 1986. Surface analysis. In *Fourier transform infrared spectrometry*. John Wiley & Sons, New York.
- Hennessey Jr., J.P. and Johnson Jr., W.C. 1981. Information content in the circular dichroism of proteins. *Biochemistry* **20**: 1085–1094.
- Hlady, V., Buijs, J., and Jennissen, H.P. 1999. Methods for studying protein adsorption. *Methods Enzymol.* **309**: 402–429.
- Jackson, M. and Mantsch, H.H. 1995. The use and misuse of FTIR spectroscopy in the determination of protein structure. *Crit. Rev. Biochem. Mol. Biol.* **30**: 95–120.
- Kabsch, W. and Sander, C. 1983. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**: 2577–2637.
- Kondo, A. and Mihara, J. 1996. Comparison of adsorption and conformation of hemoglobin and myoglobin on various inorganic ultrafine particles. *J. Colloid. Interface Sci.* **177**: 214–221.
- Lee, D.C., Haris, P.I., Chapman, D., and Mitchell, R.C. 1990. Determination of protein secondary structure using factor analysis of infrared spectra. *Biochemistry* **29**: 9185–9193.
- Manavalan, P. and Johnson Jr., W.C. 1987. Variable selection method improves the prediction of protein secondary structure from circular dichroism spectra. *Anal. Biochem.* **167**: 76–85.

- Pace, C.N., Vajdos, F., Fee, L., Grimsley, G., and Gray, T. 1995. How to measure and predict the molar absorption coefficient of a protein. *Protein Sci.* **4**: 2411–2423.
- Pancoska, P., Yasui, S.C., and Keiderling, T.A. 1991. Statistical analyses of the vibrational circular dichroism of selected proteins and relationship to secondary structures. *Biochemistry* **30**: 5089–5103.
- Raussens, V., de Jong, H., Pezolet, M., Ruysschaert, J.M., and Goormaghtigh, E. 1998. Secondary structure of the intact H<sup>+</sup>, K<sup>+</sup>-ATPase and of its membrane-embedded region. An attenuated total reflection infrared spectroscopy, circular dichroism and Raman spectroscopy study. *Eur. J. Biochem.* **252**: 261–267.
- Santos, J.H., Matsuda, N., Qi, Z.M., Yoshida, T., Takatsu, A., and Kato, K. 2003. Adsorption behavior of cytochrome *c*, myoglobin and hemoglobin in a quartz surface probed using slab optical waveguide (SOWG) spectroscopy. *Anal. Sci.* **19**: 199–204.
- Sarver Jr., R.W. and Krueger, W.C. 1991a. Protein secondary structure from Fourier transform infrared spectroscopy: A database analysis. *Anal. Biochem.* **194**: 89–100.
- . 1991b. An infrared and circular dichroism combined approach to the analysis of protein secondary structure. *Anal. Biochem.* **199**: 61–67.
- Smith, B.M., Oswald, L., and Franzen, S. 2002. Single-pass attenuated total reflection Fourier transform infrared spectroscopy for the prediction of protein secondary structure. *Anal. Chem.* **74**: 3386–3391.
- Surewicz, W.K. and Mantsch, H.H. 1988. New insight into protein secondary structure from resolution enhanced infrared spectroscopy. *Biochim. Biophys. Acta* **952**: 115–130.
- Van Tassel, P.R., Guemouri, L., Ramsden, J.J., Tarjus, G., Viot, P., and Talbot, J. 1998. A particle-level model of irreversible protein adsorption with a postadsorption transition. *J. Colloid. Interface Sci.* **207**: 317–323.
- Walsh, S.T., Cheng, R.P., Wright, W.W., Alonso, D.O., Daggett, V., Vanderkooi, J.M., and DeGrado, W.F. 2003. The hydration of amides in helices: a comprehensive picture from molecular dynamics, IR, and NMR. *Protein Sci.* **12**: 520–531.
- Zougrana, T., Findenegg, G.H., and Norde, W. 1997. Structure, stability, and activity of adsorbed enzymes. *J. Colloid. Interface Sci.* **190**: 437–448.