Structural Analysis of the PsbQ Protein of Photosystem II by Fourier Transform Infrared and Circular Dichroic Spectroscopy and by Bioinformatic Methods†

Mónica Balsera,‡ Juan B. Arellano,‡ José R. Gutiérrez,§ Pedro Heredia,‡ José L. Revuelta,§ and Javier De Las Rivas*,‡

Instituto de Recursos Naturales y Agrobiología, Consejo Superior de Investigaciones Científicas, Cordel de Merinas 52, Salamanca 37008, and Instituto de Microbiología y Bioquímica, Universidad de Salamanca, Campus Miguel de Unamuno, Salamanca 37007, Spain

Received August 6, 2002; Revised Manuscript Received October 28, 2002

ABSTRACT: The structure of PsbQ, one of the three main extrinsic proteins associated with the oxygen-evolving complex (OEC) of higher plants and green algae, is examined by Fourier transform infrared (FTIR) and circular dichroic (CD) spectroscopy and by computational structural prediction methods. This protein, together with two other lumenally bound extrinsic proteins, PsbO and PsbP, is essential for the stability and full activity of the OEC in plants. The FTIR spectra obtained in both H_2O and D_2O suggest a mainly α-helix structure on the basis of the relative areas of the constituents of the amide I and I′ bands. The FTIR quantitative analyses indicate that PsbQ contains about 53% α-helix, 7% turns, 14% nonordered structure, and 24% *β*-strand plus other *β*-type extended structures. CD analyses indicate that PsbQ is a mainly α-helix protein (about 64%), presenting a small percentage assigned to *β*-strand (≈7%) and a larger amount assigned to turns and nonregular structures (≈29%). Independent of the spectroscopic analyses, computational methods for protein structure prediction of PsbQ were utilized. First, a multiple alignment of 12 sequences of PsbQ was obtained after an extensive search in the public databases for protein and EST sequences. Based on this alignment, computational prediction of the secondary structure and the solvent accessibility suggest the presence of two different structural domains in PsbQ: a major C-terminal domain containing four α-helices and a minor N-terminal domain with a poorly defined secondary structure enriched in proline and glycine residues. The search for PsbQ analogues by fold recognition methods, not based on the secondary structure, also indicates that PsbQ is a four α-helix protein, most probably folding as an up−down bundle. The results obtained by both the spectroscopic and computational methods are in agreement, all indicating that PsbQ is mainly an α protein, and show the value of using both methodologies for protein structure investigation.

Photosystem II (PSII)[1] is a pigment−protein complex, which consists of at least 25 different protein subunits, at present denoted PsbA−Z according to the genes that encode them (for review see refs *1*−*3*). PSII is located in the thylakoids of higher plants, algae, and cyanobacteria, where it reduces plastoquinone using water as electron donor in a light-dependent multiple reaction that produces molecular oxygen as a subproduct (*4*). The release of molecular oxygen takes place on the lumenal side of PSII after the sequential accumulation of four oxidizing equivalents in an inorganic cluster, formed by four redox-active Mn ions, and the non-redox-active Ca^{2+} and Cl^- ions. The so-called manganese cluster is sandwiched between the lumenal-facing protein matrix of the transmembranal subunits of the PSII core (PsbA, PsbB, PsbC, and PsbD) and the lumenal extrinsic subunits of PSII. The three main extrinsic proteins of PSII are PsbO, PsbP, and PsbQ in higher plants and green algae. These proteins are essential for the stability and full activity of the oxygen-evolving complex (OEC) (for review see ref *5*), although none seem to bind directly to Mn (*6*). However, in cyanobacteria (*7*) and nongreen algae (*8*), PsbO is associated with two different proteins: PsbV (cytochrome *c*-550) and PsbU.

To understand the molecular mechanism of the oxygen-evolving reaction, an essential prerequisite is knowledge of PSII structure. Structural information concerning the PSII complex of cyanobacteria is on its way to being fulfilled, since a PSII structure from *Synechococcus elongatus* has already been resolved to 3.8 Å by X-ray crystallography (*9*). In the case of higher plants, PSII is still far from being resolved, where only 3D maps derived from electron cryomicroscopic images of the spinach PSII dimer, at 10 Å (*10*), and a PSII−LHCII supercomplex, at 17 Å (*11*), have so far been reported. At the same time, there has been little progress in obtaining the 3D structure of the three higher plant extrinsic proteins of PSII. A computational 3D model has been suggested for PsbO from higher plants (*12*) and has been fitted into the higher plant PSII data at 17 Å (*13*).

† This work was supported by funds from the Spanish Ministry of Science and Technology (Project PB1998-0480). M.B. holds a fellowship from the Spanish Ministry of Science and Technology.

* Corresponding author: Phone 34-923-219606 (ext 224); fax 34-923-219609; e-mail jrivas@usal.es.

‡ Instituto de Recursos Naturales y Agrobiología (CSIC).

§ Universidad de Salamanca.

[1] Abbreviations: FTIR, Fourier transform infrared; PSII, photosystem II; 3D, three-dimensional; CD, circular dichroism.

10.1021/bi026575l CCC: $25.00
Published on Web 01/01/2003

It should be noted that the cyanobacterial PSII X-ray structure at 3.8 Å provides incomplete structural data for about half of the PsbO backbone, with no residue identification (*9*). So far, there has been no progress on the 3D structure of PsbP and PsbQ from higher plants and green algae, and only some partial structural studies have been reported (*14*, *15*).

In higher plants and green algae, the three proteins PsbO, PsbP, and PsbQ are proposed to be closely associated, showing a ratio of 1:1:1 among them (*13*, *16*). The stoichiometry of the three extrinsics with respect to PSII is still being debated, given reports that indicate that the number of PsbO bound per PSII can be one (*13*, *16*) or two (*17*). PsbP and PsbQ can be separated from PSII under high ionic strength conditions (*18*). On comparing PsbP- and PsbQ-depleted PSII preparations with fully functional PSII preparations, the oxygen-evolving rate decreases in the former, but the activity can be regained to a certain extent when PsbP and PsbQ are reconstituted to PSII or when millimolar concentrations of Ca^{2+} and Cl^- are added (*18*, *19*). This suggests a relationship between PsbP and PsbQ and the Ca^{2+} and Cl^- requirement for oxygen evolution. Moreover, these two proteins may play a structural role in the OEC by forming a barrier that is open for substrates (*20*) and products (*21*) but is closed to nonphysiological reductant agents (*22*). PsbP and PsbQ also create a low-dielectric medium that is optimal for PSII binding Ca^{2+} (*23*) and Cl^- (*24*), and it has been shown that both proteins affect the magnetic properties of the Mn cluster (*25*). Some authors have also reported that the absence of these two proteins induces conformational changes in peripheral antenna proteins (*26*). Therefore, it is clear that PsbP and PsbQ play important roles in the lumenal oxygen-evolving activity of PSII from higher plants and green algae. In this study, we have focused on PsbQ. By using FTIR and CD spectroscopy and computational structural prediction methods, we present a structural analysis of this protein, calculating it to be mainly an α protein with two structural domains, one of which we predict to form a four-α-helix bundle.

MATERIALS AND METHODS

Isolation of Spinach PsbQ Cloned in Escherichia coli.

Coding regions for the mature PsbQ protein were amplified by high-fidelity PCR with the clone pSoc16.11 as a template (*27*). This clone was a kind gift donated by Professor R. G. Herrmann. The PCR primers used were 5′-CCATATGGAG-GCCAGGCCCATCGTTGTt-3′ (forward) and 5′-GGGATC-CTTAACCGAGCTTGGCAAGAAc-3′ (reverse), having restriction sites for *Nde*I and *Bam*HI, respectively (underlined nucleotides). The resulting PCR product was purified and then cloned into the *Eco*RV-digested Bluescript II SK^+ vector (BSK^+). To confirm the nucleotide sequence of the PCR-amplified *psbQ* insert, the multiple cloning site of BSK^+ was sequenced in both directions in an automated sequencer by use of the reverse and universal primers. In the following step, the plasmid DNA was consecutively cut with *Nde*I and *Bam*HI and cloned into the *Nde*I- and *Bam*HI-predigested pET12a (Unigene). The resulting construction was named JR2592. The chosen forward primer including a *Nde*I restriction site introduced a methionine residue in the first N-terminal position of the recombinant PsbQ protein. JR2592 was transferred into *E. coli* BL21(DE3)pLysS. The region of the upstream leader *ompT* of pET12a, encoding the outer membrane protein specified by *ompT*, was removed when the restriction enzymes *Nde*I and *Bam*HI were used. Therefore, the recombinant PsbQ accumulated in the cytosolic space of *E. coli* BL21(DE3)pLysS. The transformed cells were named B96 and stored at −80 °C in 20% glycerol.

Purification of the Recombinant PsbQ Protein.

B96 cells harboring the resulting JR2592 vector were grown at 37 °C in LB medium supplemented with 100 mg/mL ampicillin and 20 mg/mL chloramphenicol. When the optical density of the culture reached the value of 1.2 at 550 nm, the expression of the recombinant PsbQ was initiated by adding 1 mM IPTG. After incubation for 15−18 h at 28 °C, the cells were harvested by centrifugation, suspended in 1 M Tris-HCl, pH 8.4, containing 1 mM phenylmethanesulfonyl fluoride (PMSF), and then sonicated. PMSF was also present in the subsequent buffers. The sonicated cells were centrifuged at 150000*g* for 45 min to remove unbroken cells. The supernatant was collected and partly precipitated with 45% $(NH_4)_2SO_4$. The precipitant was discarded and the supernatant was fully precipitated with 100% $(NH_4)_2SO_4$. This latter precipitant was suspended in 20 mM Tris-HCl, pH 8.0, and 1 mM EDTA (buffer A) and dialyzed against buffer A at 4 °C overnight. Unsolubilized material was removed by centrifugation at 150000*g* for 45 min. The protein solution was first passed through a High-Trap SP column (Amersham Pharmacia Biotech) preequilibrated with buffer A. The PsbQ protein was eluted by means of a salt linear gradient up to 0.4 M NaCl for 30 min at a flow rate of 1 mL/min. The PsbQ-enriched fractions were pooled and passed through a Superdex 200 column (Amersham Pharmacia Biotech) preequilibrated with 20 mM sodium phosphate, pH 8.0, and 1 mM EDTA (buffer B) at a flow rate of 0.2 mL/min. To test the purity of the samples, SDS−polyacrylamide gel electrophoresis (SDS−PAGE) was carried out according to Laemmli (*28*) with a total acrylamide content of 12% in the separating gel.

FTIR Spectroscopic Analyses.

After PsbQ purification, the protein was lyophilized in 20 mM sodium phosphate buffer, pH 8.0, and suspended in MilliQ H_2O or in pure D_2O (Aldrich Chemical Co.) to a protein concentration of about 30 mg/mL for H_2O and about 20 mg/mL for D_2O. Infrared spectroscopy analyses were performed with a Nicolet Nexus 670 spectrometer equipped with a MCT/A detector and with a Ge−KBr beam splitter. Samples were located in a liquid cell between two 32 × 3 mm CaF_2 windows (Wilmad Glass Co. Inc.) with a path-length spacer of 6 μm (Harrick Scientific Co.). For each measurement, a 30 μL sample was used to fill the cell. Measurement conditions were as follows: scan speed 2.53 cm s^{-1}; resolution 2 cm^{-1}; wavenumber range 4000−650 cm^{-1}; number of scans per sample 1000; temperature 21 °C. Background correction was obtained by measuring background before each sample and automatically subtracting. The buffer signal in each sample was removed by subtraction of the IR spectrum of the corresponding buffer from the spectrum of each sample. An adequate removal of the water band was performed by subtraction till a flat baseline in the 1900−1750 cm^{-1} region was achieved. This was equivalent to eliminating the specific band of water at 2125 cm^{-1} (*29*). Fourier deconvolution, derivative, and band decomposition of the original amide I and I′ bands were performed

as described elsewhere (*29*−*32*). The software used for the analysis was SpectraCalc (Galactic Inc., Salem, NH). The mathematical solution to the decomposition of IR amide I band in a sum of Gaussian bands may not be unique, but when restrictions are imposed such as the maintenance of the initial band positions in an interval of ± 1 cm^{-1} and the preservation of the bandwidth within the expected limits, the band decomposition result is reproducible and reliable (*29*). The quantification procedure has an estimated error $\leq 3\%$ (*30*).

CD Spectroscopic Analyses.

An aliquot of purified PsbQ protein was diluted in 20 mM sodium phosphate, pH 8.0, to a protein concentration of 1.0 mg/mL. CD spectra were measured in a J-720 spectropolarimeter in a cell with a 0.1 mm optical path length over a wavelength range from 180 to 260 nm at a temperature of 25 °C. Each CD spectrum was the average of two accumulations at a scanning speed of 20 nm/min and a 1 nm spectral bandwidth. To obtain structural information, the CD data were analyzed by five different programs: CDSSTR, which implements the variable selection method by performing all possible calculations with a fixed number of proteins from a reference set (*33*); SELCON3, which incorporates a self-consistent method together with the singular value decomposition algorithm (SVDA) to assign protein secondary structure (*34*); CONTIN/LL, which implements the ridge regression algorithm of Provencher and Glockner (*35*) incorporating the locally linearized model of Van Stokkum et al. (*36*); LINCOMB, a linear combination least-squares method from Perczel et al. (*37*) with the reference data sets from Yang et al. (*38*); and CDNN, a method for circular dichroism deconvolution by use of back-propagation neural networks (*39*). Most of these methods can be used or downloaded from the Internet (*40*). The quantification of α-helical secondary structure was also determined according to the classical equation defined by Greenfield and Fasman (*41*).

Computational Methods for Structural Prediction and Fold Recognition.

To analyze the structure of this protein by computational methods, the PsbQ sequence of *Spinacea oleracea* was retrieved from SwissProt (P12301), followed by searching sequence homology by BLASTp in public databases [$P(N) < 10^{-7}$] (http://www.ncbi.nlm.nih.gov/blast/) (*42*). A tBLASTn search in a higher plant EST database (http://www.ncbi.nlm.nih.gov/dbEST) was also carried out (*43*). Sequence alignments were performed with both CLUSTAL-X (*44*) and T-COFFEE (*45*) multiple sequence alignment programs.

Secondary structure prediction was made from the collated results of the PsbQ multiple alignment. The spinach sequence was placed first in the multiple alignment and used as a reference. The methods used for secondary structure prediction were PHD (*46*), PROF (*47*), and SSpro (*48*). The accuracy for each of these methods is the so-called Q_3, a statistical value representing the mean of a Gaussian distribution reflecting the number of residues predicted correctly compared to known structures. This accuracy value was 70.6% for PHD, 76.6% for PROF, and 76.0% for SSPRO, being automatically and weekly evaluated by the EVA server (*49*). Predicted solvent accessibility was also calculated with PHD and PROF.

Fold recognition methods were used to look for structures analogous to PsbQ (*50*). These methods give a priority list of structures from their known structure protein libraries (formed by PDB files) where the unknown structure protein is best fitted on the basis of a set of calculated parameters. For this prediction we only applied the threading methods that do not use secondary structure information: THREADER2, where sequence information is combined with pseudoenergies obtained from solvation and contact potentials previously derived from known protein structures (*51*); FFAS, where sequence profiles of structurally known proteins (PDBs) are compared with the query protein profile (such profiles generated for each protein family do not involve the PSI−Blast algorithm) (*52*); and FUGUE, where environment-specific substitution tables derived from the HOMSTRAD database, along with structure-dependent gap penalties, are used to construct profiles (*53*). The resulting accuracy of these programs for any given problem is the so-called *z*-score, which measures the difference in score between the query−template alignment and the mean of the query with all the possible templates in the fold library of each program, taking into account the standard deviation of the score distribution. The *z*-score threshold for an accuracy of 95% for these methods is $z > 3.5$ for THREADER2, $z > 5$ for FUGUE, and $z > 8$ for FFAS.

RESULTS

PsbQ Expression and Purification.

After overexpression of PsbQ, the cells were harvested, suspended, and then sonicated in 0.8 M Tris-HCl, pH 8.4, a buffer that causes the release of the PSII extrinsic proteins from the lumenal side of the thylakoid membrane (*54*). Any possible binding (or affinity) of the recombinant PsbQ protein to the surface of the cellular membrane of *E. coli* was avoided by means of this buffer. The basic p*I* of the PsbQ protein (∼9.2) was found to be advantageous for the purification procedure. After the $(NH_4)_2SO_4$ precipitation and dialysis steps (see Materials and Methods), the use of the cationic-exchange High-Trap SP column gave a solution of PsbQ with a purity of at least 90% (Figure 1, lane c). After filtration through a Superdex 200 column, the purity of the PsbQ solution increased to over 95% (Figure 1, lane d**)**. This latter chromatography step was also used to swap the buffer A (containing Tris) for buffer B (containing phosphate), avoiding the presence of Tris in the IR analysis. The retention times of the recombinant PsbQ in the High-Trap SP and Superdex 200 columns were identical to those of the partly purified native PsbQ protein from spinach, and so was the mobility of the PsbQ polypeptide in SDS−PAGE analysis (data not shown). From SDS−PAGE it was observed that the recombinant PsbQ did not suffer any perceptible proteolytic degradation during either the cell growth period or the isolation procedure. The stability of the recombinant PsbQ was also tested in solution. After more than 48 h at 4−8 °C, SDS−PAGE analysis again showed no degradation products of the recombinant PsbQ protein. The presence of a specific prolyl endopeptidase for PsbQ in thylakoid membranes makes the isolation and purification difficult for the native PsbQ (*55*). In the method reported here, the presence of 1 mM PMSF and 1 mM EDTA was enough to avoid any degradation of PsbQ.

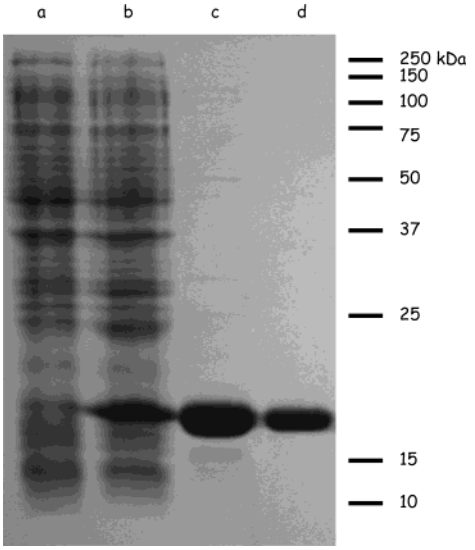

FIGURE 1: Purification steps of the recombinant PsbQ. The cDNA for the mature protein was cloned into pET12a. The recombinant PsbQ has a Met residue in the N-terminus. The SDS−polyacrylamide gel shows **(**a**)** the noninduced cells of *E. coli*, **(**b) the IPTG-induced cells, **(**c**)** post-High-Trap SP column, and **(**d) post-Superdex 200 column.

Quantitative Analyses of the PsbQ IR Amide Band.

Quantitative structural data for the isolated PsbQ preparation were obtained from the conformation-sensitive infrared spectral amide I and I′ bands, arising from the peptide bond of the protein in H_2O and D_2O media, respectively, and located from 1700 to 1600 cm^{-1} (*29*). Figure 2 shows the analyses of infrared amide I and I′ bands of the isolated PsbQ protein in H_2O (left panel) and D_2O (right panel). In each panel, the top graph shows the original amide band and the band mathematically reconstructed by addition of the main Gaussian components. The graph in the middle of each panel shows the deconvolution of the original amide band (with bandwidth 18 cm^{-1} and *k* factor 2), indicating also the peaks of the band components detected. The graph at the bottom of each panel shows the derivative of the original amide band.

The quantification of the band components and the peak positions is presented in Table 1. A good correspondence between the bands was observed in H_2O and D_2O data. The band at 1687 cm^{-1} in H_2O corresponds with the one at 1681 cm^{-1} in D_2O, and so do the bands at 1677 with 1671, 1656 with 1654, 1633 with 1632, and 1616 with 1615. This correspondence shows the isotopic displacement expected when IR measurements are carried out in D_2O (*56*). This displacement is larger at higher wavenumbers: 1687 and 1677 cm^{-1}. Two new bands appear in the D_2O spectrum. The one at 1608 cm^{-1} is not significant because it represents less than 1% of the total area. The other one at 1643.5 cm^{-1} is located between the major bands at 1654 cm^{-1} (38%) and 1632 cm^{-1} (35%) in the D_2O spectrum and seems to be predominant in the IR original amide I′ spectrum and in the deconvolution (Figure 2, right panel). The derivative (Figure 2, right panel, bottom graph) reveals that this peak at 1643 cm^{-1} is overlapped with the 1654 cm^{-1} band; the decomposition and quantification of the bands indicate that the former represents 14.4% of total structure. In H_2O, the major band is detected at 1656 cm^{-1}, accounting for 67% of the total amide I area. The second main band is detected at 1633 cm^{-1} and represents 24% of the total area. Assignment of these IR bands to specific secondary structure elements is presented in the Discussion.

Secondary Structure of PsbQ from CD Analyses.

Figure 3 shows the CD spectrum of the recombinant PsbQ protein. The CD spectrum is characterized by a strong positive band at 193 ± 1 nm and two strong negative bands at 209 ± 1 and 219 ± 1 nm. These bands are characteristics of α-helical structures (*57*) and suggest that α-helix is the main secondary structure in PsbQ. A quantification of the content of the α-helical secondary conformation by the Greenfield and Fasman method (*41*) resulted in about 63% of α-helix. Further, the overall secondary structure of PsbQ was analyzed by the five CD methods cited under Materials and Methods. The percentages of α-helix, *β*-strand, and turns and nonordered structures were obtained for each method separately. Then, the mean and the standard deviation for each type of structure were calculated from the corresponding percentages given by all five methods. In this way, the achieved data were 63.9% ± 8.8% for α-helix, 7.0% ± 3.7% for *β*-strand, and 29.4% ± 6.8% assigned to turns and nonregular structures. These indicate that PsbQ is a mainly α-helix protein with a significant amount of nonregular structure.

Computational Structural Predictions on PsbQ.

Figure 4 shows the multiple alignment of 12 PsbQ sequences, with 10 sequences derived from higher plants and two from green algae. The nomenclature in the figure relates to the identification given by the SwissProt database for PsbQ from *Spinacea oleracea* (spinach sequence, PSBQ·SPIOL), *Zea mays* (two sequences, PSQ1·MAIZE and PSQ2·MAIZE), *Arabidopsis thaliana* (two sequences, PSQ1·ARATH and PSQ2·ARATH), *Onobrychis viciaefolia* (PSBQ·ONOVI), *Chlamydomonas reinhardtii* (PSBQ·CHLRE), and *Volvox carteri* (PSBQ·VOLCA). The PsbQ sequence from *O. viciaefolia* showed clear disagreement with the rest of the plant sequences in the motif AWPYV (position 73 of the alignment). Indeed, we had detected a frame error in this region of its amino acid sequence when comparing it with those of other proteins of the PsbQ family. A correct translation of the *O. viciaefolia psbQ* DNA sequence allows the detection of this error and, therefore, PSBQ·ONOVI was modified by changing the original sequence LGHMF with the motif AWPYV (PSBQmod·ONOVI). This motif is highly conserved in all PsbQ from higher plants. Four other PsbQ sequences were reconstructed from the plant EST database, corresponding to *Lycopersicon esculentum* (PsbQ·Lyce·est), *Mesembryanthemum crystallinum* (PsbQ·Mese·est), *Glycine max* (PsbQ·Glyc·est), and *Medicago truncatula* (PsbQ·Medi·est). The multiple alignment denotes the fully conserved residues with an asterisk. A high sequence similarity was observed among the higher plant PsbQ proteins, where the percentage of identity with respect to spinach for the most similar sequence was 79.3% (*A. thaliana*) and 68% for the most divergent one (*O. viciaefolia*). On the contrary, the two algae sequences showed a lower percentage of identity with respect to spinach, 26.5% for *C. reinhardtii* and 28.6% for *V. carteri*. This difference is more obvious for the N-terminal region up to the RAKVS motif in position 53 of the alignment.

Secondary structure prediction for the multiple alignment is presented in Figure 4 and the percentages of α-helix and *β*-strand structure predicted are shown in Table 2. The

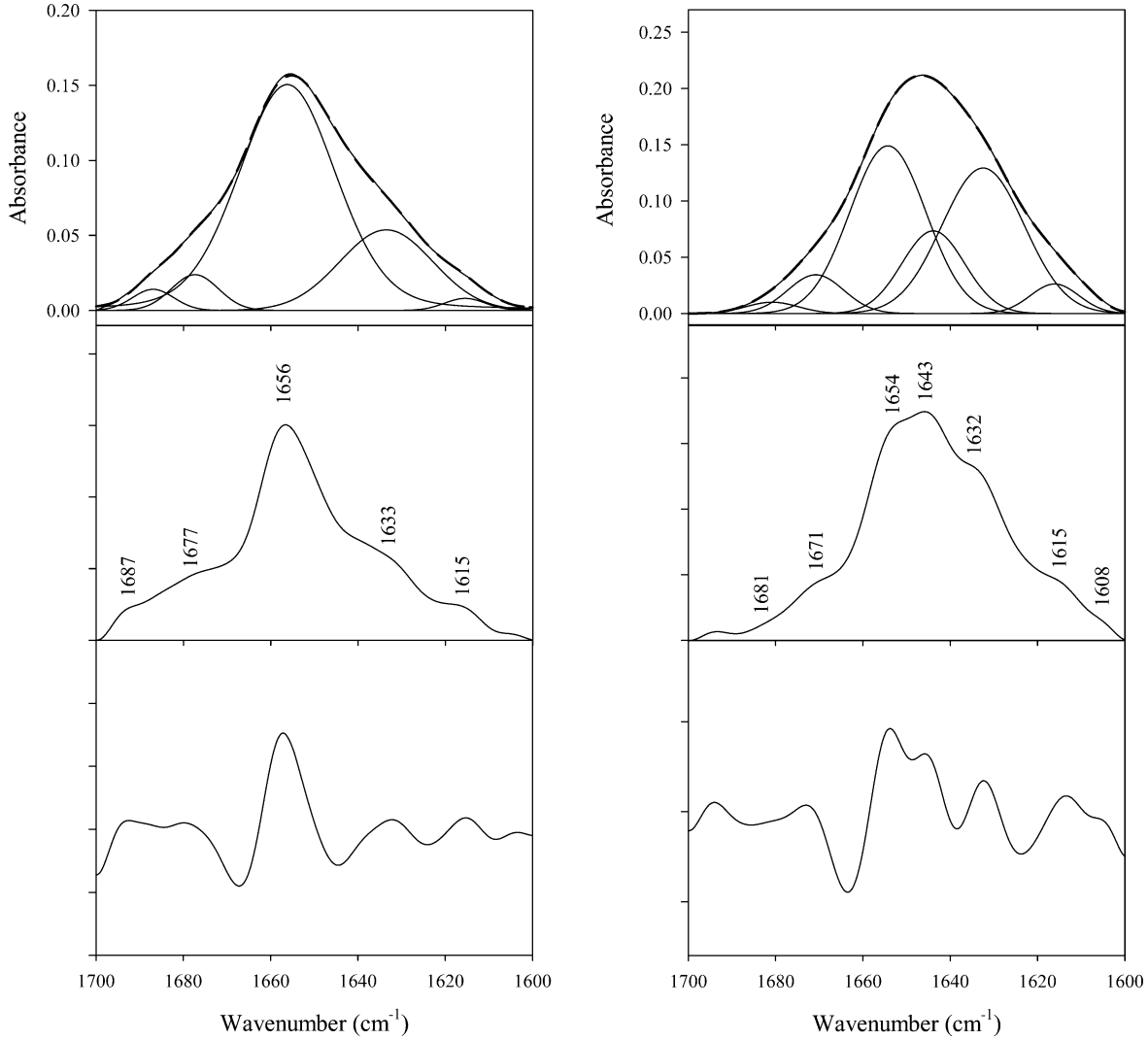

FIGURE 2: Band decomposition of the original amide I band at 21 °C of PsbQ protein (left panel) in H_2O buffer and (right panel) in D_2O buffer. The top graphs include the original spectra, the reconstructed spectra, and the five and seven main Gaussian bands that compose the spectrum in H_2O and D_2O, respectively. The middle graphs include the deconvolved spectra (bandwidth of 18 cm^{-1} and *k* factor of 2) with the peak position of each band indicated. The bottom graphs present the derivatives of the spectra. The spectra were baseline-corrected prior to the decomposition procedure. The numerical values obtained are reflected in Table 1. Contours: experimental, solid line; reconstructed by addition of band components, dashed line.

Table 1: Values Corresponding to Band Position and Percentage Area Obtained after Decomposition of the Amide I and I′ Bands of PsbQ Protein in H_2O and D_2O at Room Temperature[a]

H_2O		D_2O	
band position (cm^{-1})	band area (%)	band position (cm^{-1})	band area (%)
1687.2 ± 1.1	2.74 ± 0.36	1681.1 ± 0.3	1.08 ± 0.47
1677.3 ± 0.5	4.75 ± 0.73	1670.9 ± 0.2	6.81 ± 0.37
1656.2 ± 0.4	67.57 ± 2.67	1654.1 ± 0.3	38.32 ± 0.89
		1643.5 ± 0.4	14.40 ± 1.34
1633.1 ± 1.2	23.72 ± 3.79	1632.1 ± 0.6	34.66 ± 1.39
1615.7 ± 0.8	1.38 ± 0.54	1615.3 ± 0.6	3.93 ± 1.98
		1607.8 ± 0.5	0.57 ± 0.22

a The data correspond to the average of six different samples ± the standard deviations in H_2O and to the average of four different samples ± the standard deviations in D_2O.

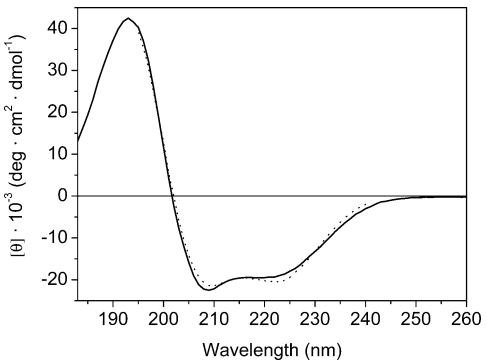

FIGURE 3: Circular dichroic spectra of PsbQ protein in 20 mM sodium phosphate, pH 8.0, at 25 °C. The dotted line represents the fitting of the experimental values by the LINCOMB method (*37*, *38*). The spectra is characterized for a maximum at 193 nm and two minima at 209 and 219 nm, with ellipticity values of 42 470, −22 508, and −19 504 deg·cm^2·$dmol^{-1}$, respectively.

percentages indicate that PsbQ is mainly α, accounting at least for 50.3% of the total secondary structure of PsbQ. The alignment suggests that PsbQ contains four main α-helices in the C-terminal domain. The N-terminal domain is predicted to contain a minor percentage of β-strand (3−7%) and a more significant one (20−26%) which corresponds not to α and β structures but to other structures that include mainly turns and loops (designated L). A low-

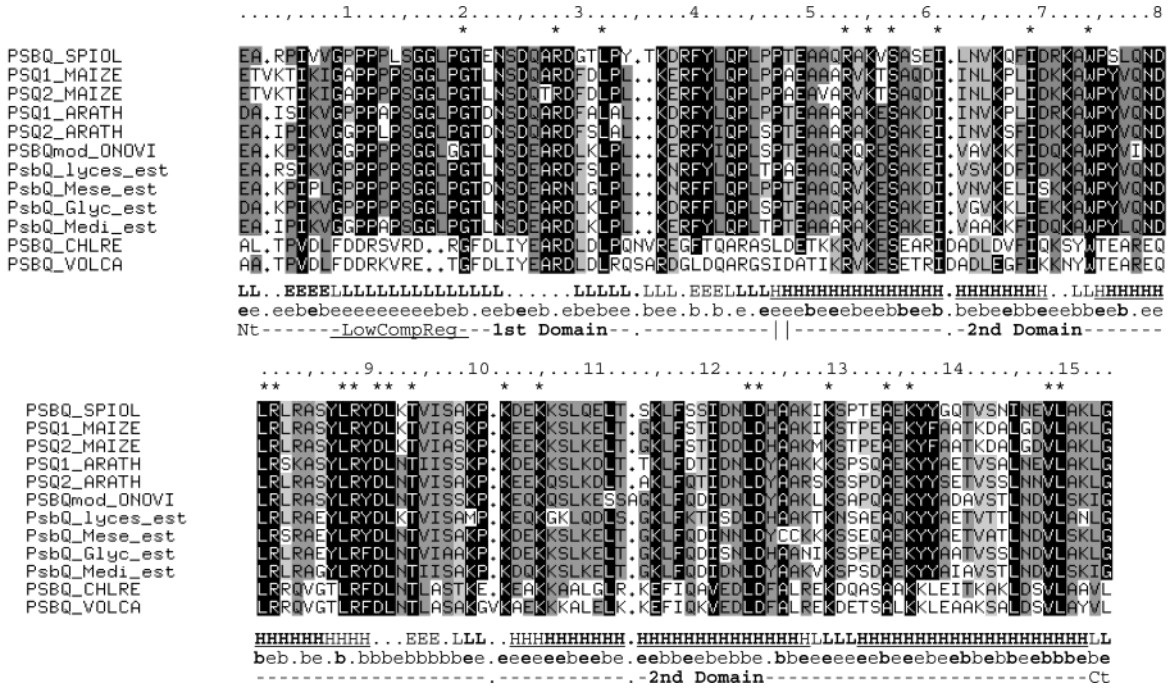

FIGURE 4: Multiple sequence alignment for the PsbQ family. An asterisk indicates a totally conserved amino acid and the color corresponds to the average similarity by the blosum 62 matrix. ONOVI PsbQ sequence was modified comparing to the other proteins in the family and translating its DNA sequence. Four sequences were reconstructed from plants EST database. The first line below the alignment shows the secondary structure prediction by PHD, PROF, and SSpro. The letters correspond to different secondary structure elements: H = helix, E = strand, and L = others. The second line indicates the predicted solvent accessibility of each amino acid in the protein (e = exposed; b = buried). Two different structural domains are proposed: 1−47, with no regular secondary structure where a low complexity region is found; 48−153, in which four α helices are predicted. The most accurate prediction is shown in black.

Table 2: Secondary Structure Prediction by Computational Methods

components	min−max percentage
α-helix	50.3−58.4
β-sheet	2.7−6.7
other structures	20.1−26.8
nonpredicted	26.8−8.1

[a] Computational methods PHD, PROF, and SSpro were compared. The most accurate prediction was used to calculate the minimum percentage for each secondary structure component according to Figure 4. The maximum percentage was obtained by basal prediction.

complexity motif enriched in Pro and Gly, G-P(4)-L-S-G(2)-L-P-G, is also detected in the N-terminus of higher plants PsbQ. This kind of polyproline motif is a good candidate to form polyproline type II structures (called PPII), which are characterized by having a left-handed helix conformation (*58*). The amount of this type of structure in PsbQ from the alignment can be estimated to be 8−10%.

The searching for PsbQ analogues by those threading methods (THREADER2, FFAS, and FUGUE) that are not based on the secondary structure of proteins was used as an additional computational methodology to predict PsbQ whole structure. Two methods (FFAS and FUGUE) were also used specifically for the prediction of the C-terminal domain (from residue 48 to the end). Four out of these five predictions suggest that PsbQ is a four-α-helices protein forming an up−down bundle fold (according to CATH nomenclature; *59*). The four templates proposed are 1vlt (PDB code) from THREADER2 (*z*-score 3.79); 1sct (for the C-terminal domain) from FFAS (*z*-score 5.62); 1jaf from FUGUE (*z*-score 3.91); and 1aep (for the C-terminal domain) from FUGUE (z-score 5.12). A fifth prediction suggests that PsbQ is an α/*β* complex with a *z*-score of 6.09 for 1dkg from FFAS. On the basis of the *z*-score of each method, the folding templates 1vlt and 1aep correspond to the most accurate predictions.

DISCUSSION

Three different approaches have been carried out to determine the secondary structure of the PsbQ protein of PSII: two spectroscopic techniques, FTIR and CD, and one computational method. These techniques agree in their predictions that PsbQ is a mainly α-helix protein. The CD analyses gave 64% ± 9% for α-helix, 7% ± 4% for *β*-strand, and 29.4% ± 6.8% assigned to turns plus nonregular structures. The secondary structure quantification obtained by bioinformatic methods indicates that PsbQ contains about 58% α-helix, 7% *β*-sheet, and 27% other structures (see Table 2).

The assignment of the IR bands to specific secondary structures is a key point of the discussion of FTIR data (*29*, *60*). The secondary structure elements of proteins absorb at different positions in the amide I region of the IR spectrum, but sometimes several structural elements can vibrate at very similar positions, making difficult the assignment. It is also known that quantitative IR analysis is difficult to determine when samples are only measured in D_2O (*56*). Looking at our FTIR results for the PsbQ protein, the bands at 1656 and 1654 cm^{-1} are the major ones both in water, 67%, and in heavy water, 38%, respectively (Figure 2 and Table 1). The band around 1656 cm^{-1} is often attributed as α-helix plus nonordered peptide structures in H_2O. In D_2O, the nonordered component moves to 1643.5 cm^{-1} (*60*) and corresponds mainly to loops, which do not present a regular repetitive conformation. Following this reasoning, it is expected that in the FTIR analyses of PsbQ the area at 1656

cm^{-1} in H_2O (67%) was sum of the two areas at 1654 cm^{-1} (38%) and 1643 cm^{-1} (14%) in D_2O. However, this was not the case here, since a difference of about 15% in area was observed (Table 1). On top of this, there was a difference of about 11% between the bands at 1633 cm^{-1} in H_2O (24%) and at 1632 cm^{-1} in D_2O (35%). If we correlate the differences detected when samples are changed from water to heavy water as solvent, it is likely that 11−15% of the area at 1656 cm^{-1} in H_2O moves to 1632 cm^{-1} in D_2O. In general, vibrations at 1635−1630 cm^{-1} are assigned to β-sheet structures, but some elements that belong to α-helix structures can also give signal in this region of the IR spectrum. For example, cytochrome *c* is an α-helix protein (composed of three major α-helices and two minor ones) that presents an IR band at 1633 cm^{-1} that has been attributed to α-helix distortion or bending provoked by helix−helix interactions (*61*). Myoglobin and hemoglobin are also α-helix proteins, which show IR spectra with a strong band at about 1638−1632 cm^{-1} that has been assigned to vibrational motion of helical segments of their respective structures (*62*, *63*). In these examples the overall content of the α-helix secondary structure is the sum of the two band areas at 1655 and 1638−1632 cm^{-1}. If we consider that about 14% of the strong band at 1656 cm^{-1} in H_2O is due to nonordered segments (band at 1643 cm^{-1} in D_2O), one might propose that a part of the α-helix structure component (about 15%) has downshifted in D_2O. This assignment would give an overall estimation of α-helix structure of about 53%. The validity of this consideration is supported by the other two techniques (CD and bioinformatic), lending further confirmation that PsbQ is a protein composed predominantly of α-helix.

With respect to the band at 1633 cm^{-1}, its center was difficult to fix, as indicated by its standard deviation in water (15% of the signal) and its broad bandwidth (25.6 cm^{-1}). These parameters indicate that the broad band at 1633 cm^{-1} does not correspond to a unique structural vibration but to a mixture of two or more structures absorbing at similar wavenumbers. The CD analyses and the computational predictions showed that PsbQ has a small proportion of well-defined β-sheet: no more than 7%. From the FTIR in H_2O, the signal observed at 1633 cm^{-1} (about 24%) can be assigned to β-strands plus other β-type extended structures (*58*). An interesting β-type extended structure is the polyproline type II structure (PPII) (*64*). In the PsbQ sequence alignment from higher plants (Figure 4), a low-complexity region is detected in the N-terminal domain. This region is enriched in proline and glycine residues and is a good candidate to form a PPII structure. The estimation of this type of structure from the alignment indicates about 8−10%. Given the above, we suggest that the IR band of PsbQ observed at 1632−1633 cm^{-1} in D_2O (which represents 35%; Table 1) corresponds to a mixture of β-strand, β-type extended chains, and vibrations coming from α-helix due to helix distortion or bending.

The three minor bands at 1687−1681, 1677−1671, and 1615 cm^{-1} were respectively assigned to high-frequency vibration of β-strand (3−1%), turns (5−7%), and lateral side chains (less than 1%) (*30*, *65*). These assignments can be added to the other quantitative data from FTIR to give about 24% total β-strand plus other β-type extended structures and about 7% of turns.

A previous FTIR study on the isolated PsbQ protein has been published (*15*). The authors indicated that the band at 1645 cm^{-1} has a relative area of 31%, which is quite different from the 14% that we observe at 1643.5 cm^{-1} in D_2O. Another discrepancy is that they directly assigned β-sheet to all the signal coming from the band at 1632 cm^{-1} and therefore concluded that PsbQ has 28% β-sheet. In their study, the IR spectrum of PsbQ protein was not recorded in H_2O solvent but only in D_2O, and therefore they were unable to analyze differences between the two solvents as done here. However, the main reason for the disagreement about the intensity of the 1643 cm^{-1} band is likely to be due to the deconvolution method used. The other publication performed curve-fitting on deconvoluted spectra with a variable *k* factor from 1.5 to 2.5 (*15*, *66*). Moreover, they adjust the initial frequency of each band manually. The method performed here is different, because only the original IR amide bands are used for the curve-fitting and we have obtained the peak of each band from the agreement between the deconvolution and the derivative (as shown in Figure 2). This method has been proven accurate and successful in the location and quantification of the components of the IR amide band for many proteins (*29*−*32*).

The prediction of four long α-helices by the bioinformatic methods is consistent with the general conclusion that the major part of PsbQ has α-helix structure. This is further supported by the prediction of the residue solvent accessibility shown in Figure 4. In such a prediction, the α-helices observed present alternation of exposed (e) and buried (b) residues. This alternation is in good agreement with four α-helices, packed together, and presenting alternative residues that face the solvent and others that are imbedded in a more hydrophobic protein core. As a whole, the secondary structure and solvent accessibility predictions allow us to propose two distinct domains in PsbQ: a first N-terminal domain rich in extended structures and a second C-terminal domain rich in α-helix structure. These domains are indicated in Figure 4. In addition, the major proportion of PsbQ protein folding templates proposed by threading methods are α-helical proteins with up−down bundle folds, based on CATH classification (*59*). The 3D coordinate superimposition of these templates shows that the core of these proteins belongs to the same structural family as formed by several α-helices interacting in a very compact way. Therefore, these predictions support the experimental results obtained by FTIR and CD and have allowed us to present a new structural view of the PsbQ protein.

ACKNOWLEDGMENT

We thank Dr. Jon Nield for critical reading of the manuscript. We are grateful to Professor Dr. José M. Andreu and Rubén M. Buey for helping with the CD spectroscopy and also to Professor Dr. Alfonso Valencia and the Protein Design Group for helping in the bioinformatic analyses.

REFERENCES

1. Rhee, K. H. (2001) *Annu. Rev. Biophys. Biomol. Struct. 30*, 307−328.
2. Hankamer, B., Morris, E., Nield, J., Carne, A., and Barber, J. (2001) *FEBS Lett. 504*, 142−151.
3. Barber, J., and Kuhlbrandt, W. (1999) *Curr. Opin. Struct. Biol. 9*, 469−475.

4. Rutherford, A. W. (1989) *Trends Biochem. Sci. 14*, 227−232.
5. Seidler, A. (1996) *Biochim. Biophys. Acta 1277*, 35−60.
6. Gregor, W., and Britt, D. (2000) *Photosynth. Res. 65*, 175−185.
7. Shen, J. R, Burnap, R. L., and Inoue, Y. (1995) *Biochemistry 34*, 12661−12668.
8. Enami, I., Murayama, H., Ohta, H., Kamo, M., Nakazato, K., and Shen, J. R. (1995) *Biochim. Biophys. Acta 1232*, 208−216.
9. Zouni, A., Witt, H. T., Kern, J., Fromme, P., KrauB, N., Saenger, W., and Orth, P. (2001) *Nature 409*, 739−743.
10. Morris, E. P., Hankamer, B., Zheleva, D., Friso, G., and Barber, J. (1997) *Structure 5*, 837−849.
11. Nield, J., Funk, C., and Barber, J.(2000) *Philos. Trans. R. Soc. London B, Biol. Sci. 355*, 1337−1344.
12. Pazos, F., Heredia, P., Valencia, A., and De Las Rivas, J. (2001) *Proteins: Struct., Funct., Genet. 45*, 372−381.
13. Nield, J., Balsera, M., De Las Rivas, J., and Barber, J. (2002) *J. Biol. Chem. 277*, 15006−15012.
14. Zhang, H., Ishikawa, Y., Yamamoto, Y., and Carpentier, R. (1998) *FEBS Lett. 426*, 347−351.
15. Zhang, H., Yamamoto, Y., Ishikawa, Y., and Carpentier, R. (1999) *J. Mol. Struct. 513*, 127−132.
16. Murata, N., Miyao, M., Omata, T., Matsunami, H., and Kuwabara, T. (1984) *Biochim. Biophys. Acta 765*, 363−369.
17. Betts, S. D., Hachigian, T. M., Pichersky, E., and Yocum, C. F. (1994) in *Photosynthesis: from Light to Biosphere* (Mathis, P., Ed.) pp 385−388, Kluwer Academic Publishers, Dordrecht, The Netherlands.
18. Ghanotakis, D. F., Babcock, G. T., and Yocum, C. F. (1984) *FEBS Lett. 167*, 127−130.
19. Ljungberg, U., Jansson, C., Andersson, B., and Akerlund, H. E. (1983) *Biochim. Biophys. Acta 113*, 738−744.
20. Hillier, W., Hendry, G., Burnap, R. L., and Wydrzynski, T. (2001) *J. Biol. Chem. 50*, 46917−46924.
21. Anderson, J. M. (2001) *FEBS Lett. 488*, 1−4.
22. Vander Meulen, K. A., Hobson, A., and Yocum, C. F. (2002) *Biochemistry 41*, 958−966.
23. Vrettos, J. S., Stone, D. A., and Brudvig, G. (2001) *Biochemistry 40*, 7937−7945.
24. Wincencjusz, H., Yocum, C. F., and van Gorkom, H. J. (1999) *Biochemistry 38*, 3719−3725.
25. Campbell, K. A., Gregor, W., Pham, D. P., Peloquin, J. M., Debus, R. J., and Britt, R. D. (1998) *Biochemistry 37*, 5039−5045.
26. Boekema, E. J., van Bremen, J. F. L., van Roon, H., and Dekker, J. P. (2000) *Biochemistry 39*, 12907−12915.
27. Jansen, T., Rother, J., Steppuhn, H., Reinke, K., Beyreuther, K., Jansson, C., Andersson, B., and Herrmann, R. G. (1987) *FEBS Lett. 216*, 234−240.
28. Laemmli, U. K. (1970) *Nature 227*, 680−685.
29. Arrondo, J. L. R., Muga, A., Castresana, J., and Goñi, F. M. (1993) *Prog. Biophys. Mol. Biol. 59*, 23−56.
30. Arrondo, J. L., Castresana. J., Valpuesta, J. M., and Goñi, F. M. (1994) *Biochemistry 33*, 11650−11655.
31. Echabe, I., Haltia, T., Freire, E., Goni, F. M., and Arrondo, J. L. (1995) *Biochemistry 34*, 13565−13569.
32. De Las Rivas, J., and Barber, J. (1997) *Biochemistry 36*, 8897−8903.
33. Johnson, W. C. (1999) *Proteins: Struct., Funct., Genet. 35*, 307−312.
34. Sreerema, N., and Woody, R. W. (1993) *Anal. Biochem. 209*, 32−44.
35. Provencher, S. W., and Glockner, J. (1981) *Biochemistry 20*, 33−37.
36. Van Stokkum, I. H. M., Spoelder, H. J. W., Bloemendal, M., Van Grondelle, R., and Groen, F. C. A. (1990) *Anal. Biochem. 191*, 110−118.
37. Perczel, A., Park, K., and Fasman, G. D. (1992) *Anal. Biochem. 203*, 83−93.
38. Yang, J. T., Wu, C.-S. C., and Martinez, H. M. (1986) *Methods Enzymol. 130*, 208−269.
39. Dalmas, B., Hunter, G. J., and Bannister, W. H. (1994) *Biochem. Mol. Biol. Int. 34*, 17−26.
40. Lobley, A., Whitmore, L., and Wallace, B. A. (2002) *Bioinformatics 18*, 211−212.
41. Greenfield, N., and Fasman, G. D. (1969) *Biochemistry 8*, 4108−4116.
42. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res. 25*, 3389−3402.
43. Banfi, S., Guffanti, A., and Borsani, G. (1998) *Trends Genet. 14*, 80−81.
44. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997) *Nucleic Acids Res. 22*, 4673−4680.
45. Notredame, C., Higgins, D. G., and Heringa, J. (2000) *J. Mol. Biol. 302*, 205−217.
46. Rost, B. (1996) *Methods Enzymol. 266*, 525−539.
47. Rost, B. (2001) *J. Struct. Biol. 134*, 204−218.
48. Baldi, P., Brunak, S., Frasconi, P., Pollastri, G., and Soda, G. (1999) *Bioinformatics 15*, 937−946.
49. Eyrich, V. A., Marti-Renom, M. A., Przybylski, D., Madhusudhan, M. S., Fiser, A., Pazos, F., Valencia, A., Sali, A., and Rost, B. (2001) *Bioinformatics 17*, 1242−1243.
50. Fischer, D., Elofsson, A., Rychlewski, L., Pazos, F., Valencia, A., Rost, B., Ortiz, A. R., and Dunbrack, R. L., Jr. (2001) *Proteins: Struct., Funct., Genet. 45*, 171−183.
51. Jones, D., Taylor, W., and Thornton, J. (1992) *Nature 358*, 86−89.
52. Rychlewski, L., Jaroszewski, L., Li, W., and Godzik, A. (2000) *Protein Sci. 9*, 232−241.
53. Shi, J., Blundell, T. L., and Mizuguchi, K. (2001) *J. Mol. Biol. 310*, 243−257.
54. Ljungberg, U., Akerlund, H. E., and Andersson, B. (1986) *Eur. J. Biochem. 158*, 477−82.
55. Kuwabara, T., Murata, T., Miyao, M., and Murata, N. (1986) *Biochim. Biophys. Acta 146*, 146−155.
56. Jackson, M., and Mantsch, H. H. (1995) *Crit. Rev. Biochem. Mol. Biol. 30*, 95−120.
57. Holzwarth, G. N., and Doty, P. (1965) *J. Am. Chem. Soc. 87*, 218−228.
58. Stapley, J. S., and Creamer, T. P. (1999) *Protein Sci. 8*, 587−595.
59. Orengo, C. A., Michie, A. D., Jones, S., Swindells, M. B., and Thornton, J. M. (1997) *Structure 5*, 1093−1108.
60. Haris, P. I., and Chapman, D. (1995) *Biopolymers 37*, 251−263.
61. Paquet, M. J., Laviolette, M., Pézolet, M., and Auger, M. (2001) *Biophys. J. 81*, 305−312.
62. Byler, D. M., and Susi, H. (1986) *Biopolymers 25*, 469−487.
63. Heimburg, T., Schünemann, J., Weber, K., and Geisler, N. (1999) *Biochemistry 38*, 12727−12734.
64. Dukor, R. K., and Keiderling, T. A. (1991) *Biopolymers 31*, 1747−1761.
65. Barth, A. (2000) *Prog. Biophys. Mol. Biol. 74*, 141−173.
66. Ahmed, A., Tajmir-Riahi, H. A., and Carpentier, R. (1995) *FEBS Lett. 363*, 65−68.

BI026575L