Solution NMR Structure and Backbone Dynamics of the PsaE Subunit of Photosystem I from *Synechocystis* sp. PCC 6803[†,‡]

Patrick Barth,[§,||] Philippe Savarin,[⊥] Bernard Gilquin,[⊥] Bernard Lagoutte,[§] and Françoise Ochsenbein*,[@]

Département de Biologie Joliot Curie, Service de Bioénergétique, Service de Biophysique des Fonctions Membranaires, and Département d'ingénierie et d'étude des protéines, CEA de Saclay and URA CNRS 2096, F-91191 Gif sur Yvette Cedex, France

ABSTRACT: PsaE is a small peripheral subunit of photosystem I (PSI) that is very accessible to the surrounding medium. It plays an essential role in optimizing the interactions with the soluble electron acceptors of PSI, ferredoxin and flavodoxin. The solution structure of PsaE from the cyanobacterium *Synechocystis* sp. PCC 6803 has been investigated by NMR with a special emphasis on its protein dynamic properties. PsaE is characterized by a well-defined central core that consists of a five-stranded β-sheet ($+1, +1, +1, -4x$). Four loops (designated the A–B, B–C, C–D, and D–E loops) connect these β-strands, the overall resulting structure being that of an SH3-like domain. As compared to previously determined PsaE structures, conformational differences are observed in the first three loops. The flexibility of the loops was investigated using ^{15}N relaxation experiments. This flexibility is small in amplitude for the A–B and B–C loops, but is large for the C–D loop, particularly in the region corresponding to the missing sequence of *Nostoc* sp. PCC 8009. The plasticity of the connecting loops in the free subunit is compared to that when bound to the PSI and discussed in relation to the insertion process and the function-(s) of PsaE.

Photosystem I (PSI)[1] is one of the two reaction centers of the oxygenic photosynthetic apparatus in cyanobacteria, green algae, and higher plants. PSI is a membrane-embedded protein complex composed of 11 subunits in cyanobacteria (*1*), its main function being that of a light-driven oxidoreductase. Following photon absorption, a series of fast electron transfer reactions occurs in the core of PSI (*2*). This complex process catalyzes the global reduction of ferredoxin (Fd) (or flavodoxin) and oxidation of plastocyanin (or cytochrome c_6) (*3*), using energy from light to drive a reaction that is energetically unfavorable. Reduced Fd then shuttles electrons between PSI and many key enzymes in cellular metabolism

(*4*). Many structural and functional studies suggest that Fd interacts with PSI at a multiple-binding site domain that protrudes from the membrane surface and involves the three stromal subunits PsaC–PsaE (*5–9*). PsaD and PsaE contain no cofactors. PsaC bears two 4Fe–4S centers (F_A and F_B), the terminal electron acceptors of PSI which transfer the electron to Fd. Two of these subunits, PsaC and PsaE, are mainly responsible for the high-affinity binding of Fd (*10, 11*), while PsaD seems to provide a complex electrostatic network to facilitate the approach of the Fd (*12–14*). A model for the structural arrangement of these subunits, which was first deduced from electron microscopy (*9*), was recently improved by X-ray crystallography studies of PSI from the cyanobacterium *Synechococcus elongatus* (*15, 16*). The most recent refinements of this model provide a precise description of the position, orientation, and numerous contacts of these subunits. However, little is known about the assembly of these peripheral polypeptides, the factors controlling the topological constraints required for efficient electron transfer and the fine-tuning of Fd binding. Interestingly, mutated PSI lacking either PsaD or PsaE can be reconstituted in vitro with the isolate subunit from several species (*17–19*). Using a PsaE-less mutant of *Synechocystis* sp. 6803, it was shown the reconstitution resulted in a fully functional system in terms of Fd interaction and reduction (*14*). These results suggest that enough intrinsic structural information is present in the isolated polypeptides for their correct assembly in the PSI complex.

In cyanobacteria, PsaE is a small (8 kDa), slightly basic, and highly conserved protein. Three-dimensional (3D) solution structures have been determined by NMR for two PsaE polypeptides from *Synechococcus* sp. 7002 (S 7002) (*20*) and *Nostoc* sp. 8009 (N 8009) (*21*). A five-stranded β-sheet (named βA–βE), similar to an SH3 domain (*20*), was found

[‡] The coordinates of the structure of S 6803 PsaE are available as PDB entry 1gxi.

[§] Service de Bioénergétique.

[||] Present address: Department of Molecular and Cell Biology, University of California, 229 Stanley Hall, Berkeley, CA 94720.

[⊥] Département d'ingénierie et d'étude des protéines.

[@] Service de Biophysique des Fonctions Membranaires.

[1] Abbreviations: COSY, correlated spectroscopy; CPMG, Carr–Purcell–Meiboom–Gill; CSA, chemical shift anisotropy; DSS, 2,2-dimethyl-2-silapentane-5-sulfonic acid; DTT, dithiothreitol; EDTA, ethylenediaminetetraacetic acid; Fd, ferredoxin; FID, free induction decay; GSH, reduced glutathione; GST, glutathione *S*-transferase; HMQC, heteronuclear multiple-quantum coherence; HSQC, heteronuclear single-quantum coherence; IPTG, isopropyl thiogalactopyranoside; LB, Luria-Bertani; MD, molecular dynamics; N 8009, *Nostoc* sp. PCC 8009; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; OD, optical density; PCR, polymerase chain reaction; PsaX, polypeptide X of photosystem I; PSI, photosystem I; rmsd, root-mean-square deviation; $R_N(N_z)$, longitudinal ^{15}N relaxation rate constant; $R_N(N_{x,y})$, transverse ^{15}N relaxation rate constant; $R_N(H_z^N \leftrightarrow N_z)$, cross-relaxation rate between amide proton and nitrogen; S 6803, *Synechocystis* sp. PCC 6803; S 7002, *Synechococcus* sp. PCC 7002; TFA, trifluoroacetic acid; TOCSY, total correlated spectroscopy; WT, wild type.
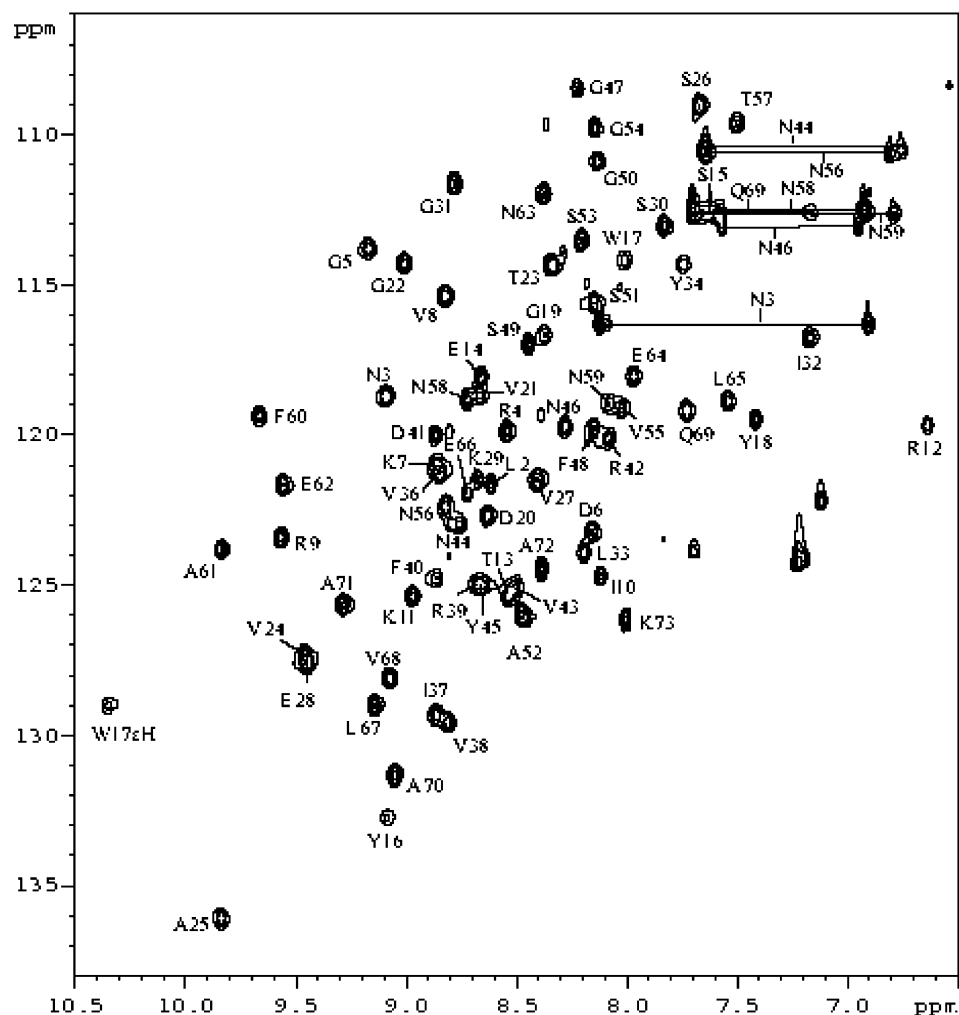
FIGURE 1: $^1H-^{15}N$ HSQC spectrum of ^{15}N-labeled S 6803 PsaE. The assignments are indicated.

as a common, well-defined structural motif. The general features of the intervening loops are also conserved, with the third one, between β-strands C and D, being either short (11 amino acids) or long (18 amino acids) depending on the species, thus defining two main families of PsaE (Figure 2b). From the X-ray data (*15, 16*), this loop seems to be of special importance for the interaction with the core of PSI. Both N- and C-terminal ends of the protein are located close to one another in the structure, a feature also predicted for spinach PsaE based on immunochemical experiments (*22*). In addition, the common acidic C-terminus contains a 3_{10}-helical turn, most often followed by a short sequence that is rich in Ala, Pro, and Lys. The PsaE subunit from S 7002 is in the "long third loop" family and has no extra C-terminal sequence. In contrast, the PsaE from N 8009 is the "short third loop" type but has an extra sequence of eight amino acids, including four lysines, at the C-terminus. In this respect, PsaE from S 6803 appears to be like a chimeric protein, with a long C−D loop and a short C-terminal extra sequence that includes a single lysine. Several mutants of S 6803 PsaE that exhibit considerable modifications in ferredoxin binding have been described (*11, 23*). The structural consequences of these mutations have yet to be studied. We report here detailed NMR structural analysis of S 6803 PsaE in solution, with a special emphasis on its backbone dynamic properties. A strong structural homology with the structures

of S 7002 and N 8009 PsaE is observed and confirms the SH3-like fold. However, conformational variations were found in the connecting A−B, B−C, and C−D loops. As already observed for S 7002 and N 8009 PsaE, the atomic resolution of these three loops is significantly lower than in the protein core. We thus investigated the dynamic behavior of the protein backbone in S 6803 PsaE using ^{15}N relaxation experiments to discriminate between a real flexibility of these loops and an insufficient number of experimental constraints. The flexibility of the various backbone regions of PsaE is compared to similar data obtained with few SH3 domains. A potential general role for flexibility in protein recognition mediated by these domains is highlighted. Finally, the functional implications of PsaE backbone plasticity are discussed specifically in the context of the Fd binding site in PSI.

EXPERIMENTAL PROCEDURES

Overexpression and Purification of Wild Type (WT) PsaE. The DNA sequence encoding WT PsaE was cloned into a pET15b plasmid (gift from U. Mühlenoff). For overexpression, this plasmid was transformed into *Escherichia coli* BL21 (DE3)plysS. Labeling experiments were conducted in M9 minimal medium (*24*) with [^{15}N]ammonium sulfate (ISOTEC Inc., Matheson Co., Miamisburg, OH) and 10 mM glucose. A small liquid culture in LB medium, made with
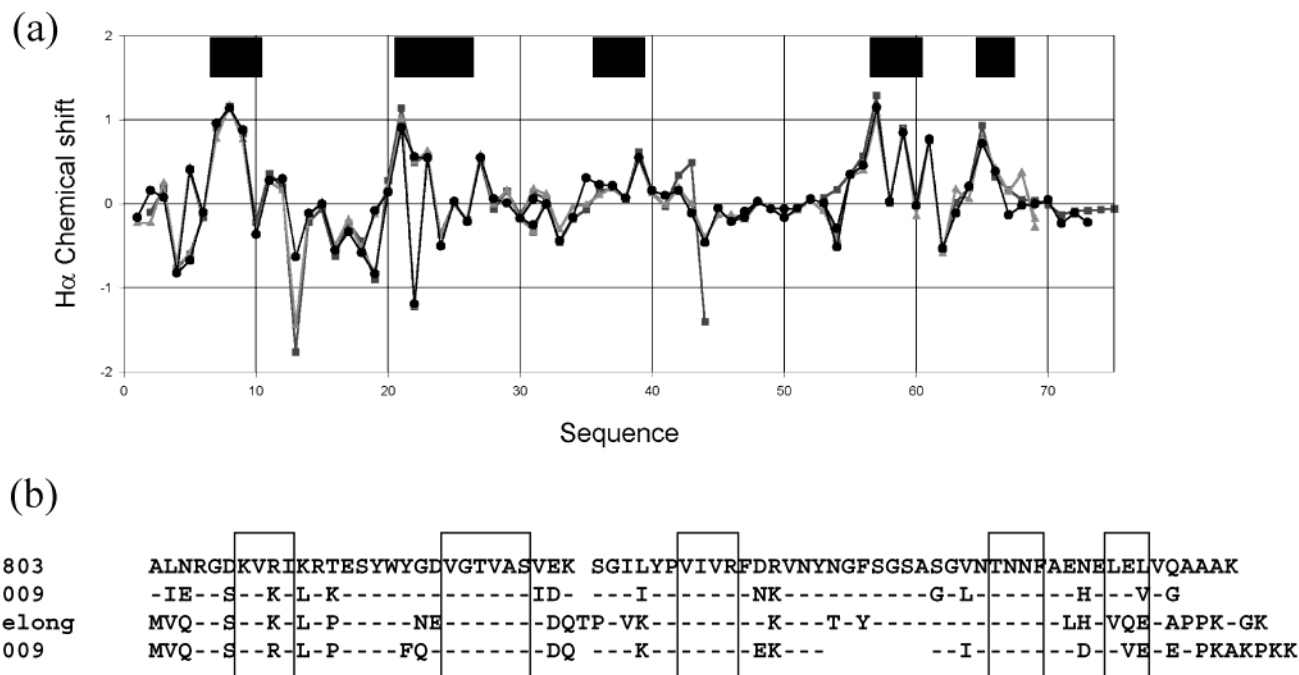
FIGURE 2: (a) Hα chemical shift index of S 6803 PsaE (●) at pH 5.2 at 20 °C. For comparison, values of Hα chemical shifts of the previously determined PsaE structures are also reported [S 7002 (▲) and N 8009 (■)]. The positions of the five β-strands are indicated with boxes at the top. (b) Sequence comparison of the PsaE proteins from S 7002, S 6803, N 8009, and *S. elongatus*. Residues identical to that of S 6803 PsaE are indicated with a dash. The five β-strands are in boxes.

freshly plated transformed cells, was grown up to an OD of 2 and inoculated at a dilution of 1/500 in a large volume of ^{15}N-labeled M9 medium (1−2 L). Incubation was carried out overnight at 28 °C; when the OD at 580 nm reached 1.5−2, induction was initiated by adding 0.5 mM isopropyl β-D-thiogalactoside and the culture kept at 28 °C for 6 h. Harvested cells were resuspended in Tris-HCl buffer [50 mM Tris-HCl (pH 8.5), 0.1 M NaCl, 1 mM EDTA, 5 mM $MgCl_2$, and 0.1% mercaptoethanol] and broken in the presence of protease inhibitors (1 mM phenylmethanesulfonyl fluoride, 1 mM benzamidine, and 1 mM ε-amino-n-caproic acid), using a cell disrupter (Z model, Constant Systems). Cellular debris was sedimented at 9000g for 10 min at 4 °C, and the whole cell extract was precipitated in a stepwise fashion with ammonium sulfate (20, 35, 42, and 75% saturation). The last precipitate containing PsaE is solubilized in denaturing buffer [50 mM Tris (pH 8.5), 1 mM EDTA, 0.1 mM DTT, and 6 M urea]. The solution is then dialyzed rapidly (1−2 h) against renaturing buffer [20 mM Tris (pH 8)]. This procedure allowed a fraction of the high-molecular weight contaminants to precipitate. The dialysis supernatant is injected onto an anion exchange column (DE52, Whatmann) which does not retain PsaE and allows a fast separation of most remaining contaminants, including nucleic acids. The excluded fraction is precipitated at 75% ammonium sulfate saturation, and the precipitate is solubilized in denaturing buffer [100 mM Tris (pH 8.5), 1 mM EDTA, 0.5 mM DTT, and 1 M urea] to minimize the oligomerization of PsaE. This solution is loaded on a gel filtration column (Superdex 75, Pharmacia). PsaE is precipitated at 80% acetone or 75% ammonium sulfate saturation and resolubilized with a small volume of 6 M guanidine chloride. This sample is further loaded onto a C18 reverse phase column equilibrated with H_2O and 0.1% TFA and eluted with a gradient of acetonitrile (22). The fraction containing pure PsaE is then extensively dialyzed overnight against water before lyophilization. The lyophilized PsaE was dissolved in 60 mM phosphate buffer (pH 5.2) containing 10% D_2O and 100% D_2O for H_2O and D_2O spectra, respectively. The PsaE subunit from *Synechocystis* S 6803 has a tendency to aggregate at a concentration higher than 1 mM. The NMR sample was thus prepared at a concentration of 0.5 mM. At this concentration, the protein is momomeric as shown by exclusion chromatography (data not shown).

Overexpression and Purification of PsaE R39Q. The PsaE gene was PCR mutated (R39Q) and fused to the gene encoding glutathione *S*-transferase (25) using the pGex-kT plasmid from Pharmacia. The thrombin site generated for the cleavage of the chimeric protein added an extra dipeptide (GS) at the N-terminus of PsaE. An overnight culture of *E. coli* BL21 cells transformed with the resulting plasmid was inoculated in 1 L of M9 medium. After incubation at 37 °C for 4 h, IPTG (0.5 mM) was added to induce the expression of the GST−PsaE chimera and the culture incubated for 3 h at 22 °C. Cell lysis was performed using the PsaE WT protocol. The soluble lysis extract is purified by one-step affinity chromatography on a glutathione (GSH) column. The chimera is eluted with reduced GSH. The chimeric protein was proteolyzed by thrombin for 1 h at 10 °C, and the GST was retained on the same glutathione affinity column. In a final step, thrombin and R39Q PsaE are separated by reverse phase chromatography as previously described for WT PsaE. The mutated protein was solubilized under the same conditions as the WT protein.

NMR Spectroscopy for Structural Determination. NMR experiments were carried out on Bruker DRX spectrometers operating at 500, 600, and 800 MHz and equipped with 5 mm triple-resonance gradient probes with actively shielded three-axis gradients. All experiments were performed at 293 K and pH 5.3.

Standard experiments were recorded to assign all ^{1}H and ^{15}N resonances (HSQC, HMQC-COSY, TOCSY, NOESY, and NOESY-HSQC). Quadrature detection was achieved by either the TPPI or the TPPI−States method (*26*). ^{15}N decoupling during acquisition was performed using a GARP sequence (*27*). Mixing times for NOESY and TOCSY experiments were 100 and 80 ms, respectively. For these two-dimensional (2D) experiments, the number of scans was between 120 and 160 and the number of points was 2048 in the direct dimension; 450 increments were used in the indirect dimension for homonuclear experiments. The recycling delay was 1 s. Both HSQC and HMQC-COSY experiments were carried out at 600 MHz with 200 increments in the indirect dimension and spectral widths of 3 and 2.4 kHz in the ^{1}H and ^{15}N dimensions, respectively. The 3D ^{1}H−^{15}N NOESY-HSQC spectrum was recorded at 600 MHz with 142 experiments and a spectral width of 7.2 kHz in the indirect ^{1}H dimension, 64 increments and a spectral width of 2.4 kHz in the ^{15}N indirect dimension, and 512 points for the direct ^{1}H dimension and a spectral width of 3 kHz. Thirty-two scans were acquired per free induction decay. The mixing time was 100 ms. The 3D ^{1}H−^{15}N HNHA experiment (*28*) was performed with 100 increments and a spectral width of 6 kHz in the indirect ^{1}H dimension, 60 increments and a spectral width of 2 kHz in the ^{15}N indirect dimension, and 2048 points in the direct ^{1}H dimension with a spectral width of 2 kHz.

All spectra were processed with XWINNMR (Bruker). The superposition of 3D slices necessary for the assignment of resonances was performed using TRIAD (TRIPOS).

Experimental Restraints. The proton−proton distance restraints were deduced from the analysis of the NOESY-HSQC spectrum obtained at 600 MHz and the homonuclear NOESY spectra obtained at 800 MHz in H_2O and D_2O. The volumes of the NOESY cross-peaks were integrated and then classified in several categories: very strong, strong, low strong, medium, low medium, weak, and very weak, for which the corresponding distances were set to 2, 2.5, 2.7, 3, 3.5, 4, and 5 Å, respectively. The error for these distances value was set to $\pm 0.125d^2$ (*29*). When comparison of the distance deduced from peaks found on both sides of the diagonal and in both experiments (or both solvents) showed an error larger than 25%, an error equal to twice the root-mean-square deviation was used (*30*).

The values of *J* coupling were obtained from the HNHA 3D experiment (*28*). ϕ angle restraints were derived from the $J_{HN−Ha}$ coupling value on the basis of the empirical Karplus relation (*31*); they were set to $-70 \pm 40°$ for $J_{HN−Ha}$ values lower than 6 Hz and to $-120 \pm 40°$ for $J_{HN−Ha}$ values higher than 8 Hz.

Structure Calculation. A semiautomated iterative assignment procedure was used to assign the NOE cross-peaks and built the 3D structure simultaneously. This procedure is described in detail in ref *32*. Since the ^{1}H and ^{15}N chemical shift assignments of S 6803 PsaE and S 7002 PsaE are similar (see the Results), a predicted 3D structure of S 6803 PsaE was first built by homology modeling on the basis of the solution structure of S 7002 PsaE (*20*). A first set of NOESY cross-peaks was assigned manually (80%). The assignment of the remaining NOESY cross-peaks was carried out using the semiautomated procedure on the basis of the proton chemical shift list and the predicted model. To avoid

misassignment, the remaining cross-peaks were assigned in the first step of the procedure using a low chemical shift tolerance (0.01 ppm) and a short distance cutoff. As the iteration procedure progressed, the values used for these parameters were progressively set to the standard values (*30*).

Finally, 200 structures were calculated. A restrained molecular dynamic at 600 K, a slow cooling, and a minimization were carried out with a standard energy function [file topallh22x.pro and parallh22x.pro in X-PLOR 3.1 (*33*), comprising an electrostatic term]. The 14 best structures were selected to be analyzed. As shown in ref *32*, this final set of structures should not be affected by the use of a model structure instead of a linear one as a starting point of the procedure.

Backbone ^{15}N Relaxation Measurements. The pulse sequences used to determine ^{15}N $R_N(N_z)$, ^{15}N $R_N(N_{x,y})$, and ^{1}H−^{15}N NOEs were based on those previously described (*34*) with minor modifications. All experiments were recorded with the ^{15}N sweep width set to 2430 Hz for $R_N(N_z)$ and $R_N(N_{x,y})$ and to 2030 Hz for the heteronuclear ^{1}H−^{15}N NOEs and the ^{1}H sweep width to 3000 Hz with the carrier set at the middle of the amide proton resonances. Each 2D data set consisted of 1024 complex data points in t_2 and 60 complex points in t_1 for $R_N(N_z)$ and $R_N(N_{x,y})$ and and 200 complex points in t_1 for the heteronuclear ^{1}H−^{15}N NOEs. Twenty-four scans were acquired per free induction decay, except for the NOE measurements, which used 352 scans due to the lack of an initial polarization transfer. A watergate on resonance sequence was used just prior to acquisition to suppress the solvent resonance (*35*).

For the heteronuclear ^{1}H−^{15}N NOEs, a carefully optimized water flip-back pulse was added before the last proton 90° pulse in the experiment without saturation (*36*). The two experiments with and without proton saturation were carried out in an interleaved manner, FID by FID. A relaxation delay of 10 s was used before the FID of the experiment without saturation. The power used for proton saturation was optimized in a one-dimensional (1D) experiment to minimize sample heating. A GARP sequence with proton pulses at a power of ∼1 W applied during 4 s was sufficient to reach the ^{1}H−^{15}N heteronuclear NOE steady state.

For the other experiments, a recycle delay of 4 s was employed. $R_N(N_z)$ values were obtained at 600 MHz with 12 delays of 12, 58, 96, 156, 180, 300, 396, 480, 540, 696, 1020, and 2004 ms. $R_N(N_{x,y})$ values were obtained at 600 MHz with 10 delays of 8, 24, 64, 80, 144, 200, 280, 400, 496, and 640 ms. In all these experiments, the points corresponding to different relaxation delays were acquired in an interleaved manner to avoid any bias that could arise from long-term instabilities.

Data Processing and Determination of Relaxation Rates. NMR spectra were processed using GIFA (*37*). A pure cosine bell and a pure squared cosine bell were applied along t_1 and t_2, respectively. The data were zero filled to 1024 points along t_1 and to 2048 points along t_2 prior to Fourier transformation. Finally, a baseline correction was applied in both dimensions using the corresponding GIFA baseline routine (*38*). Cross-peak intensities were determined from peak heights using the GIFA peak-picking routine. Relaxation rate determination and error estimation were performed as previously described (*39, 40*). The fitting to a monoexponential decay was achieved using the nonlinear Leven-

burg−Marquardt minimization algorithm (*41*) implemented in MATLAB. All the fitting procedures were written using this software. The quality of the fits and error estimations were obtained using established Monte Carlo procedures, with an experimental Gaussian error set to 5% of the experimental intensities and using a set of 500 synthetic data for each N−H vector. NOE enhancements were obtained as the ratio of the peak heights in the spectra recorded with and without saturation of protons during the relaxation delay. The uncertainties of these NOEs were the sum of the uncertainties of the peak intensities in each experiment. The rate of cross relaxation between the amide proton and nitrogen $R_N(H_z^N \leftrightarrow N_z)$ was then calculated according to the following expression:

$$\text{NOE} = 1 + \frac{\gamma_H}{\gamma_N} \frac{R_N(H_z^N \leftrightarrow N_z)}{R_N(N_z)} \quad (1)$$

Spectral Density Calculation. Spectral densities $J(\omega)$ were calculated using the reduced matrix approach proposed by Peng and Wagner (*42−46*). This approach makes possible the calculation of the spectral densities at frequencies 0 and ω_N and the evaluation of a spectral density $\langle J(\omega_H) \rangle$ at high frequency. $\langle J(\omega_H) \rangle$ is close to $J(\omega_H + \omega_N)$ or $J(0.85\omega_H)$ (*47−50*).

Lipari and Szabo Analysis. The model-free approach proposed by Lipari and Szabo (*51, 52*) and its subsequently extended version proposed by Clore et al. (*53, 54*) were applied. The expression of the spectral density corresponding to the initial model of Lipari and Szabo containing two correlation times is given by

$$J(\omega) = \frac{2}{5}\left[\frac{S^2\tau_R}{1 + (\omega\tau_R)^2} + \frac{(1 - S^2)\tau_i'}{1 + (\omega\tau_i')^2}\right], \frac{1}{\tau_i'} = \frac{1}{\tau_R} + \frac{1}{\tau_i} \quad (2)$$

where τ_i is the effective correlation time for internal motions.

The expression for the spectral density function in the extended version of the model is the sum of three Lorentzians and is given by

$$J(\omega) = \frac{2}{5}\left[\frac{(S_f^2 S_s^2)\tau_R}{1 + (\omega\tau_R)^2} + \frac{S_f^2(1 - S_s^2)\tau_s'}{1 + (\omega\tau_s')^2} + \frac{(1 - S_f^2)\tau_f'}{1 + (\omega\tau_f')^2}\right] \quad (3a)$$

with

$$\frac{1}{\tau_s'} = \frac{1}{\tau_R} + \frac{1}{\tau_s} \text{ and } \frac{1}{\tau_f'} = \frac{1}{\tau_R} + \frac{1}{\tau_f} \quad (3b)$$

The first term of the equation is related to the slowest motion and is weighted by the square of a generalized order parameter for internal motions S^2 $(=S_f^2 S_s^2)$, where S_f^2 and S_s^2 provide a model-independent measurement of the spatial restriction of all fast and slow time scale internal motions, respectively. S_f^2 and S_s^2 are associated with two effective correlation times, τ_f and τ_s, respectively.

In eqs 2 and 3, the highest correlation time τ_R was considered as the overall tumbling correlation time τ_R and was estimated from the ratio $R_N(N_{x,y})/R_N(N_z)$ averaged over the residues located in the most structured segments of the protein (*55*). Some models include the contribution of slow

Table 1: Experimental Restraints and Structural Statistics[a]

no. of experimental distance restraints	
total	1119
manually assigned	892
automatically assigned	157
ambiguous constraints	70
rmsd from distance restraints, distances (Å)	0.08 ± 0.002
rmsd from distance restraints, dihedrals (deg)	2.4 ± 0.7
no. of violation distance restraints higher than 0.75 Å	0
no. of violation dihedral restraints higher than 11°	0
rmsd from ideal values	
bonds (Å)	0.017 ± 0.0004
angles (deg)	3.5 ± 0.07
impropers (deg)	2.3 ± 0.4
energy (kcal/mol)	
bonds	59 ± 40
angles	221 ± 130
VdW[b]	10 ± 2
NOE[c]	137 ± 14
electrostatic	−1025 ± 24
Ramachandran analysis (%)[d]	
residues in the most favored region	58
residues in additional allowed regions	37.8
residues in generously allowed regions	3.7
residues in disallowed regions	0.5
quality index, %ϕ, ψ	95.8
coordinate precision (Å)	
rmsd on backbone atoms	1.3 ± 0.9

[a] All values are averaged on the 14 X-PLOR structures. [b] The van der Waals energy is calculated with a repel function and the parallhpg parameters. [c] The values of the square-well NOE are calculated with force constants of 50 kcal mol^{-1} Å^{-2}. [d] Calculated with Procheck-nmr (*76*).

Table 2: Parameters of the Various Lipari and Szabo Models Used To Fit the Experimental NMR Relaxation Parameters

model	optimized parameters	fixed parameters
1	S^2	τ_R, τ_i
2	S^2, τ_i	τ_R
3	S^2, R_{ex}	τ_R, τ_i
4	S^2, τ_i, R_{ex}	τ_R
5	S_f^2, S_s^2, τ_s	τ_R, τ_f

conformational exchange to $R_N(N_{x,y})$, by the fit of an additional term R_{ex}. For each residue, the different models were tested according to the increasing number of adjustable parameters (from model 1 to model 5, Table 2). A grid search determination of the initial values for the adjustable parameters was followed by the fitting procedure itself. The validity of the fit was checked according to the statistical criteria $\chi_{exp}^2 < \chi_{95\%}^2$ (see below). The simplest model leading to a satisfactory fit was retained, and models including a higher number of adjustable parameters were not considered.

Quality of the Fits and Evaluation of Uncertainties. Motion parameters were optimized for each residue for which the data could be collected from nonlinear least-squares minimization of the error function χ_{exp}^2:

$$\chi_{exp}^2 = \left[\frac{R_N(N_{x,y})_{calc} - R_N(N_{x,y})_{exp}}{\delta R_N(N_{x,y})}\right]^2 + \left[\frac{R_N(N_z)_{calc} - R_N(N_z)_{exp}}{\delta R_N(N_z)}\right]^2 + \left(\frac{\text{NOE}_{calc} - \text{NOE}_{exp}}{\delta \text{NOE}}\right)^2 \quad (4)$$

where $\delta(X)$ [X being $R_N(N_{x,y})$, $R_N(N_z)$, or NOE] is the estimated experimental error of the relaxation parameters. The subscripts exp and calc refer to experimental and back-

calculated relaxation parameters, respectively. Initial values for parameters were obtained from a grid search minimization. The quality of the fits and the uncertainties were obtained from statistical analysis (*39, 40*). Five hundred simulated data sets were selected from a Gaussian noise distribution centered on the relaxation rates back-calculated from the optimized parameters and corresponding to the previously estimated errors in experimental $R_N(N_{x,y})$, $R_N(N_z)$, and NOEs. A model was considered satisfactory if the optimized χ_{exp}^2 lay within the 95% confidence limit obtained from the 500 Monte Carlo simulations, i.e., if $\chi_{exp}^2 < \chi_{95\%}^2$. The uncertainties in the values of the model-free parameters were the standard deviation of the simulated parameters.

RESULTS

Assignment of Resonances. The assignment of the chemical shifts of all protons of the S 6803 PsaE was performed using the standard sequential protocol using the 2D and 3D homo- and heteronuclear spectra (*56*). All ^1H and ^{15}N chemical shift assignments are available as Supporting Information. Figure 1 shows an assigned ^{15}N HSQC spectrum. This spectrum is comparable to that of the two previously resolved PsaE proteins, S 7002 PsaE and N 8009 PsaE. Figure 2a shows the Hα chemical shift indexes for S 6803, S 7002, and N 8009 PsaE. The values are very similar for the three proteins ($\Delta\delta < 0.1$ ppm), and this agrees with their high level of sequence identity (Figure 2b). This observation suggests that the 3D structures of the three proteins are likely to be very similar.

From the HSQC spectrum (Figure 1), several residues of S 6803 PsaE display two sets of NMR resonances with unequal (major and minor form) intensities. These residues are mostly located in the region spanning residues 40−53 and also include R_4, G_5, Y_{16}, W_{17}, and T_{23}. Doubling peaks were also reported previously for the N- and C-terminal residues of N 8009 PsaE and were attributed to an incomplete release of the N-terminal methionine and to the cis−trans isomerization of the A70−P71 linkage, respectively (*21*). These explanations are not likely for S 6803 PsaE since the complete release of the N-terminal methionine was confirmed by mass spectroscopy (data not shown) and since no proline residue is found close to any exchanging residues. The molecular basis of the conformation exchange observed in S 6803 PsaE seems more likely to be due to the presence of two conformations in slow exchange (slower than the millisecond time scale) involving the rearrangement of aromatic ring residues W17 and F40. The structure determination and the dynamic study were performed on the major form of the protein.

Conformational Constraints and Structural Determination. The assignment of the NOE constraints from S 6803 PsaE was performed using a combination of manual and semiautomated iterative assignment procedures (see Experimental Procedures). The summary of the constraints used for the structure calculation together with statistical details of the 14 best structures is reported in Table 1. Figure 3b shows a ribbon representation of the mean structure of S 6803 PsaE. This solution structure comprises a five-stranded β-sheet (+1, +1, +1, −4x) (*57*) with four connecting loops. The five strands are composed of residues 7−10 (β_A), 21−26 (β_B), 36−39 (β_C), 57−60 (β_D), and 65−67 (β_E). Superposition of
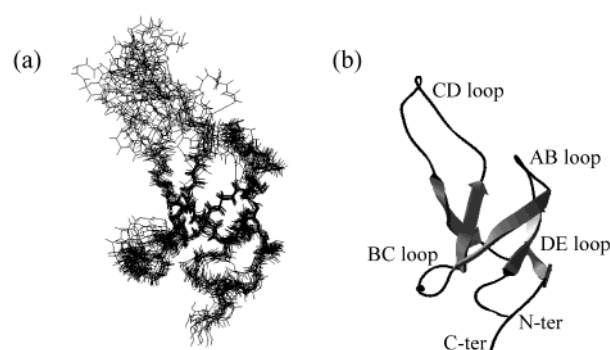

FIGURE 3: Structure of S 6803 PsaE. (a) Backbone superposition of the final 14 structures calculated as described in Experimental Procedures. The backbone heavy atoms of the β-strand were used for the alignment. (b) Ribbon representation of the representative structure of S 6803 PsaE, the closest structure to the mean structure. The positions of the loops are indicated.

the backbone atoms (CO, Cα, and N) of the 14 best structures is shown in Figure 3a. The secondary structural elements are characterized by a well-defined backbone with an rmsd value around the mean coordinate position equal to 0.48 ± 0.10 Å. The definition of the connecting loops is variable. The A−B loop which connects strands β_A and β_B is of medium resolution (rmsd = 1.18 ± 0.9 Å). The B−C loop presents a similar resolution (rmsd = 0.96 ± 0.5 Å). The C−D loop, which is the largest (17 residues), is poorly defined. Residues 48−51 exhibit an rmsd value larger than 3 Å. This is correlated with the low number of NOE restraints in this segment of the protein. The D−E loop is short and well-defined (rmsd = 0.62 ± 0.3 Å). On the basis of specific NOE contacts and ϕ and ψ values, three β-turns were identified: N_3-D_6, $W_{17}-D_{20}$ (type II β-turn), and $S_{26}-K_{29}$ (type I β-turn). The D−E loop forms a type III 3/10-helical turn from residue E_{62} to E_{65}, including one hydrogen bond between the amide proton of D_{64} and the carbonyl of A_{61}. Thirteen residues are buried (relative accessible surface area of <15%) Among these residues, two are polar (D_6 and S_{26}) and seven (V_8, I_{10}, V_{36}, V_{38}, F_{40}, F_{60}, and L_{65}) constitute the main hydrophobic core, surrounded by a limited hydrophobic cluster (L_2, V_{24}, and V_{27},).

Structural Comparison of S 6803 PsaE with Other PsaE Structures. Figure 4 shows the superposition of the mean structures of S 6803 and of the two previously determined PsaE structures, S 7002 and N 8009 PsaE (entries 1PSE and 1QP3, respectively). The secondary structural elements of S 6803 PsaE can be superimposed with those of S 7002 and N 8009 PsaE, with rmsds of 0.93 and 0.85 Å, respectively. However, some significant variations are observed in the loops connecting these secondary structural elements.

The position of the A−B loop is relatively remote from the protein core in the structure of S 6803 as compared to the two other structures (rmsds of 2.28 and 1.95 Å with S 7002 and N 8009, respectively). Its position is mostly influenced by contacts between the W_{17} side chain and residues located in the C−D loop and with F_{40} located in the hydrophobic core. The proximity between some CH groups observed in the previously determined PsaE structures is not observed in the case of S 6803 PsaE. For example, contacts between W_{17} and V_{43} observed in S 7002 PsaE and between W_{17} and V_{43} or Y_{16} and Y_{45} observed in the case of N 8009 PsaE are not present in our homonuclear NOESY
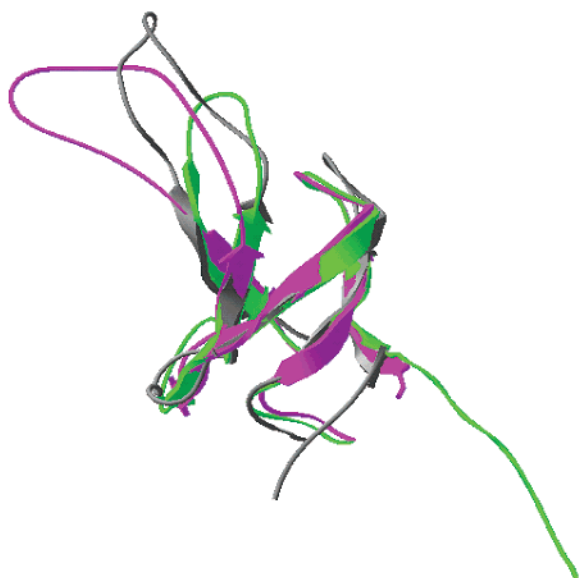
FIGURE 4: Superposition of the representative structures of S 6803 PsaE in gray, S 7002 PsaE (PDB entry 1PSF) in magenta, and N 8009 PsaE (PDB entry 1QP2) in green. The backbone heavy atoms of the β-strand were used for the alignment.

spectra obtained at 800 MHz. These missing NOE constraints may be due either to a significant distance increase between the corresponding residues or to the decrease of the signal-to-noise ratio resulting from the presence of two sets of peaks for W_{17} and F_{40}.

Some minor variations in the conformation of the B–C loop are also observed between the structure of S 6803 PsaE and those of S 7002 and N 8009 PsaE, particularly for residues S_{26}–K_{29}. The conformation of this region is largely influenced by two sets of long-range NOE constraints. A first set of constraints between hydrophobic residues V_{36} and I_{37} (N-terminus of β_C) and polar residues A_{25}, S_{26}, E_{28}, K_{29}, and G_{31} (C-terminus of β_B and the B–C loop) tends to maintain the B–C loop in a closed conformation, as also observed for S 7002 and N 8009 PsaE. A second set includes several NOEs between V_{27} and residues located in the N-terminal tail (L_2, N_3, and R_4). This second set of NOEs is compatible with the formation of a hydrophobic cluster (L_2, V_{24}, and V_{27}) that tends to open the conformation of the B–C loop. Among these residues, L_2 and V_{27} are not conserved (Figure 2b). These sequence variations could explain the differences observed in the conformation of the B–C loop between the structure of S 6803 PsaE and those of S 7002 and N 8009. The conformation of the CD loop varies significantly from one structure to another. The number of constraints per residue in this region is very low, and the high rmsd between the 14 structures of the model suggests that this loop is highly flexible. Finally, the D–E loops from the three structures are superimposable.

Backbone Dynamics. NMR Relaxation Parameters and Spectral Densities. We determined the $R_N(N_z)$, $R_N(N_{x,y})$, and ^1H–^{15}N heteronuclear NOE relaxation parameters of 60 of the 73 amide nitrogen nuclei of the protein at 600 MHz (data not shown). Following the reduced density matrix formalism (42–46), three values of the spectral density function have been derived from the relaxation parameters: $J(0)$, $J(\omega_N)$, and $\langle J(\omega_H)\rangle$, where $\langle J(\omega_H)\rangle$ is close to $J(\omega_H + \omega_N)$ or $J(0.85\omega_H)$ (47–50). These values are reported in Figure 5

as a function of the protein sequence. As expected, the spectral density obtained for each residue decreases monotonically as a function of the frequency. Four residues (9, 10, 40, and 61) show large values of $J(0)$ compared to the other residues, most probably reflecting the presence of conformational exchange on the microsecond to millisecond time range for these residues. The general shape of $J(\omega_N)$ is flat along the protein sequence, with a mean value of 0.298 $\pm$ 0.04 ns. This value of the spectral density function reflects the fraction of energy used by motions occurring in the nanosecond range mainly due to the overall tumbling of the molecule. In contrast, significant variations of $\langle J(\omega_H)\rangle$ along the sequence are observed, allowing the description of three distinct behaviors. (i) In regions corresponding to the well-defined β-sheets and the adjacent regions, i.e., the N_3–N_{44} segment, including the three β-strands β_A–β_C together with the V_{55}–Q_{69} segment, including β_D and β_E, the mean value of $\langle J(\omega_H)\rangle$ is 0.0052 $\pm$ 0.0024 ns. (ii) In the regions corresponding to the loop between β_C and β_D and the N_{46}–G_{54} segment, values of $\langle J(\omega_H)\rangle$ are significantly higher with a mean value of 0.014 $\pm$ 0.003 ns. (iii) For residues located at the N- and C-termini of the protein (residues L_2 and residues A_{72} and K_{73}, respectively), higher values of $\langle J(\omega_H)\rangle$ are also observed, with a mean value of 0.016 $\pm$ 0.003 ns.

Model-Free Analysis. The overall tumbling correlation time τ_R was estimated on the basis of the average value of the $R_N(N_{x,y})/R_N(N_z)$ ratio. Residues showing ^1H–^{15}N heteronuclear NOE values lower than 0.75 were not considered since internal motions significantly contribute to their relaxation data. Conversely, residues showing high values of $R_N(N_{x,y})$, thus including a significant contribution of conformational exchange, were also discarded for the calculation of the overall tumbling correlation time τ_R. A value of 6.0 ns was obtained and was used to extract the Lipari and Szabo parameters. This value is consistent with an 8 kDa monomeric protein at 293 K.

The experimental data were fitted using various models differing in the number of adjustable parameters they contain. The characteristics of these models are reported in Table 2. Briefly, models 1–4 are based on the simplest version of the Lipari and Szabo approach (equation 2 in Experimental Procedures), and models 3 and 4 include a term of conformational exchange R_{ex}. Model 5 corresponds to the extended version of the model (eq 3 in Experimental Procedures) and takes into account a fast and a slow internal motion. This latter model is appropriate for residues showing high internal flexibility.

Values of the order parameter S^2, of the internal correlation time τ_i, and of the exchange term R_{ex} obtained for S 6803 PsaE are plotted on a per residue basis in Figure 6c–e, respectively. Residues located in the five β-sheets (β_A–β_E) exhibit high values of S^2 (0.79 $\pm$ 0.05) and low values of τ_i (0.042 $\pm$ 0.005 ns) as expected for residues located in a well-defined tertiary structure. Values of conformational exchange are also low for these residues except in some cases (9, 10, 40, and 61). This is fully consistent with the low atomic rmsd ($\sim$1 Å) obtained in this region (Figure 6b).

The loops and turns that join the five β-sheets show various behaviors. The A–B loop appears to be quite rigid with motion parameter values similar to those observed in regular β-sheets except for one residue, S_{15}, that exhibits an S^2 value of 0.6. Unfortunately, it was not possible to
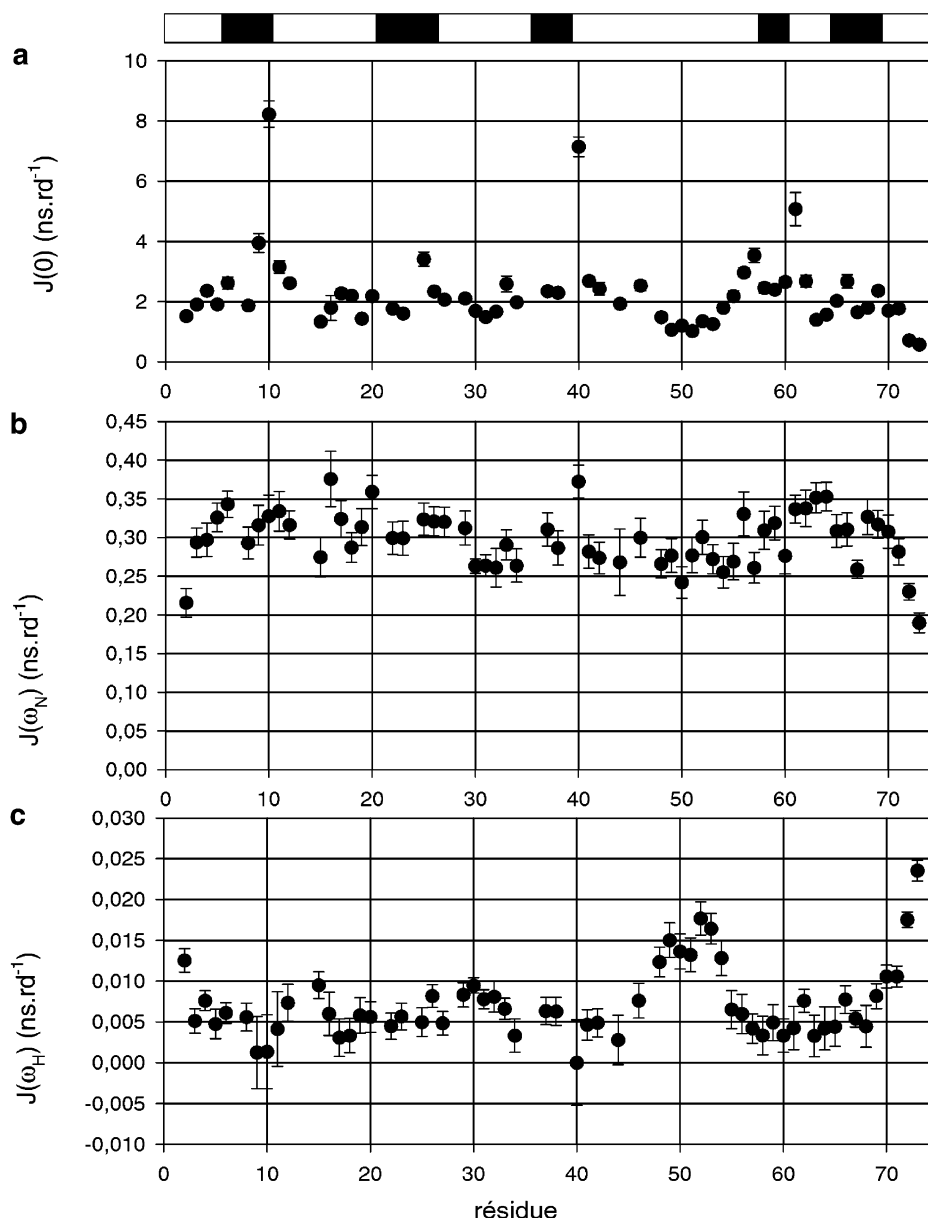
FIGURE 5: Values of the spectral density function (in nanoseconds per radian) determined for S 6803 PsaE at pH 5.2 and 20 °C from ^{15}N relaxation data acquired at 600 MHz as a function of the protein sequence: (a) at zero frequency, including the contribution of conformational exchange, (b) at a frequency of ω_N (60 MHz), and (c) at a frequency of $0.85\omega_H$ (522 MHz). Bars represent the experimental error. Boxes at the top delineate the five β-strands.

determine the relaxation data for T_{13} and E_{14} due to resonance overlaps. Nevertheless, the relatively large atomic rmsd observed in this region (1.2 Å for residues 12−19; see Figure 6b) may not be entirely explained by the presence of local conformational flexibility around residue S_{15}. The structure in this region is mainly influenced by long-range NOEs with the aromatic protons of F_{40} and W_{17}, and both residues show two sets of resonances. Some critical NOEs defining the major conformation of the protein may not have been detected due to the low signal-to-noise ratio for these residues, accounting also for the lower local resolution (Figure 6a). The B−C loop exhibits intermediate values of S^2 (0.71 ± 0.07), showing the presence of a significant but limited flexibility. These results are consistent with the intermediate atomic rmsd found in this region (∼1 Å). In contrast, the C−D loop exhibits significantly low values of S^2 (0.65 ± 0.1) together with high values of the internal correlation time (0.63 ± 0.4 ns). In particular, the F_{48}−S_{53}

segment located in the sequence missing from all eukaryotes and some prokaryotes (1, 21) appears to be remarkably flexible, with a mean S^2 value of 0.54 ± 0.06 and an internal correlation time of 0.75 ± 0.3 ns. This high internal flexibility is consistent with the near-zero chemical shift index values of these residues (Figure 2a) and explains the low number of NOEs per residue and the high atomic rmsd (∼4 Å) in this loop. On the contrary, the motion parameters for the D−E loop correspond to highly immobilized residues (S^2 = 0.89 ± 0.02), consistent with the good structural resolution in this region. Finally, residues located at the N- and C-termini (residues 2 and residues 72 and 73, respectively) are highly flexible (S^2 = 0.37 ± 0.18).

DISCUSSION

The structure and dynamic properties of S 6803 PsaE in solution have been investigated in this work. The structure is very similar to that already determined by NMR spec-
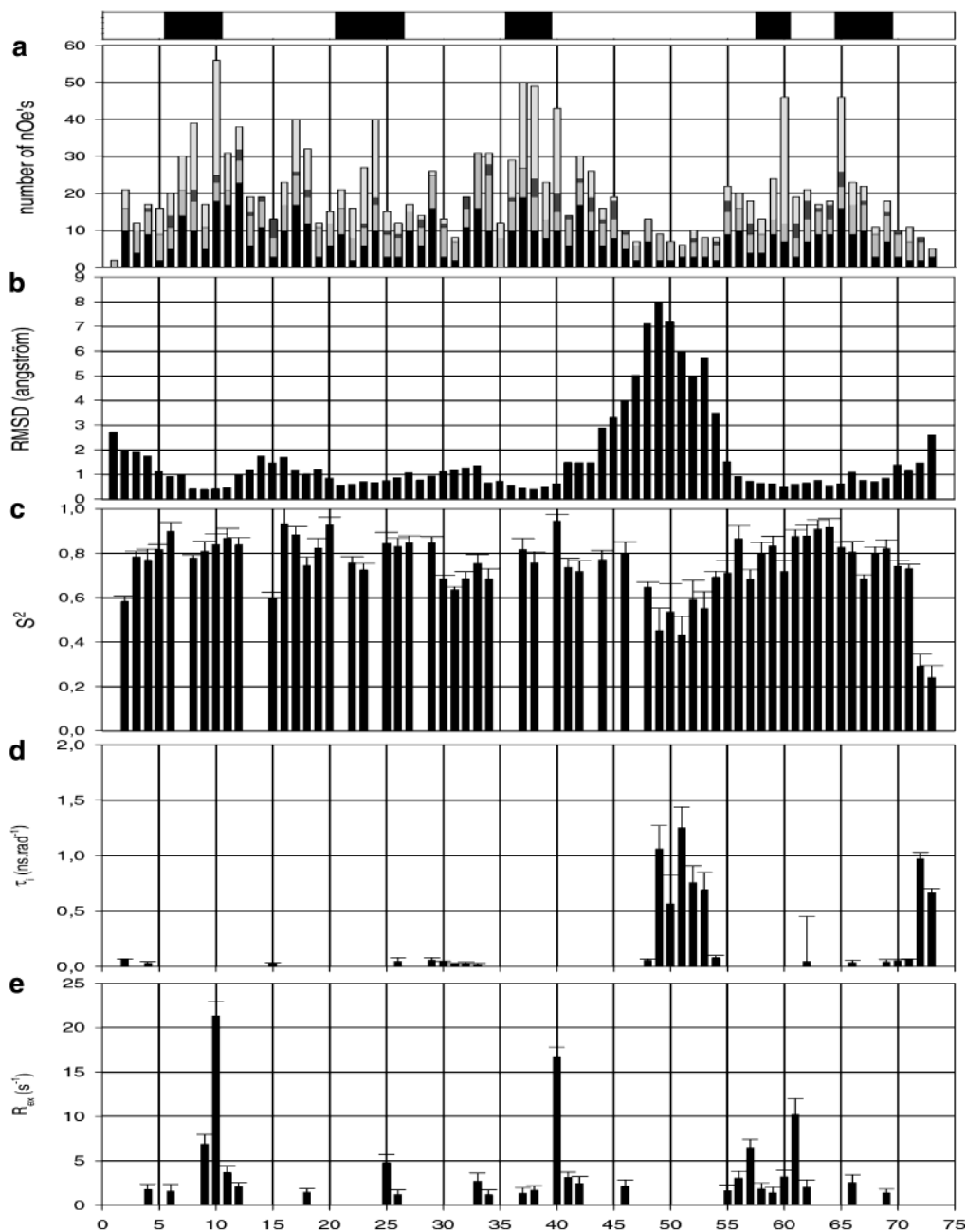
FIGURE 6: (a) Distribution of $^1H-^1H$ NOEs used as restraints in the calculation of S 6803 PsaE structure: black, intraresidue NOE; light gray, sequential NOE; dark gray, medium-range NOE ($2 \leq |i - j| \leq 4$); and white, long-range NOE ($|i - j| \leq 5$). (b) Average rmsd for the backbone heavy atoms of the 14 structures. (c−e) Model-free parameters derived from ^{15}N relaxation measurements obtained for S 6803 PsaE at pH 5.2 and 20 °C as a function of the protein sequence. The overall tumbling correlation time τ_R was fixed at 6 ns/rad for all residues. Bars represent the standard deviation of simulated parameters: (c) generalized order parameter S^2, (d) correlation time τ_i corresponding to internal motions, and (e) contribution of the conformational exchange parameter (R_{ex}) to the ^{15}N transverse relaxation. Boxes at the top delineate the β-strands.

troscopy of PsaE from the two other species (*20, 21*). The overall topology of the protein is conserved, particularly the well-defined five-β-strand motif. Conformational variations were found in the connecting loops (A−B, B−C, and C−D). As already observed for PsaE from S 7002 and N 8009, the atomic resolution of these three loops is significantly lower than in the protein core. We investigated the dynamic behavior of the S 6803 PsaE backbone using ^{15}N relaxation experiments to determine whether this lower resolution was due to the flexibility of these loops or to an insufficient number of experimental constraints. Flexibility in the A−B loop was shown to be limited to the close environment of residue S_{15}. It is likely that the relatively high atomic rmsd

in the entire loop also results from an insufficiently low number of constraints due to the presence of two conformations, especially for W_{17} (see the Results). The B−C loop also exhibits a moderate internal flexibility, in full agreement with the resolution of the structure in this region. On the other hand, the C−D loop clearly appears to be highly flexible, particularly in the region of seven amino acids that is missing in the PsaE polypeptides from some species (i.e., the short third loop PsaE family).

PsaE belongs to the SH3-like superfold family which includes many different members with no sequence homology (*58, 59*). SH3-like domains are often part of modular proteins showing a high functional versatility like kinases
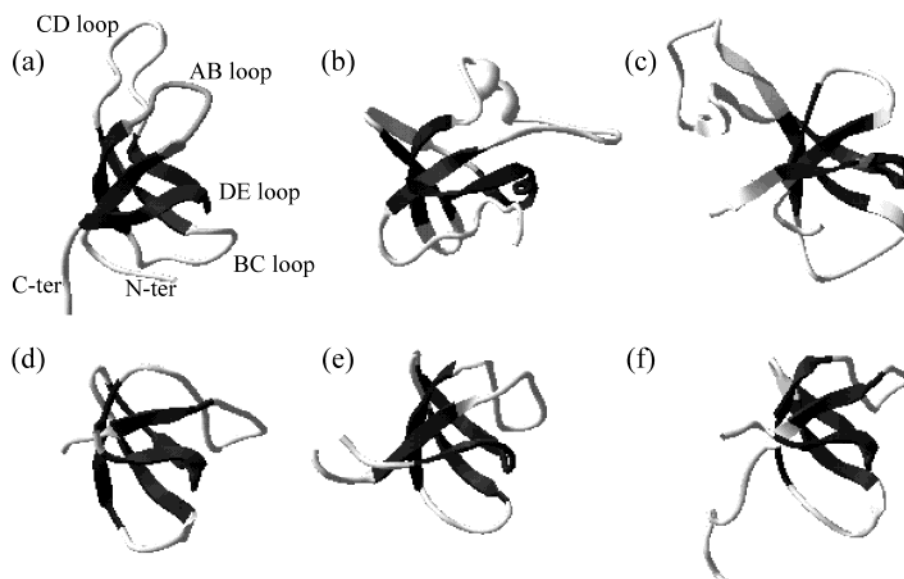
FIGURE 7: Ribbon representation of different proteins or domains sharing the SH3-like fold. The five β-strands are in black: (a) S 6803 PsaE, (b) the C-terminal domain of the diphtheria toxin repressor protein (PDB entry 1BYM), (c) CcdB, a topoisomerase poison from *E. coli* (PDB entry 3VUB), (d) the Src SH3 domain (PDB entry 1SRL), (e) the Hck SH3 domain (PDB entry 5HCK), and (f) the Btk SH3 domain (PDB entry 1AWX).

involved in signal transduction, nucleic acid binding proteins, and redox enzymes involved in cell metabolism. Their topology is characterized by a five-β-strand motif and four connecting loops. Figure 7 shows a structural alignment of some SH3-like domains. The five β-strands highlighted in black are well aligned among the entire superfamily. In contrast, the lengths of the three A−B, B−C, and C−D loops are highly variable. Sequence insertions can be present: in the A−B loop as for the SH3-like domain of the diphtheria toxin repressor, DtxR (Figure 7b) (*60*), and for the SH3 domains of protein kinases (for example, from Src, Hck, and Btk, panels d−f of Figure 7, respectively) (*61−63*); in the B−C loop as for the *E. coli* topoisoremase poison CcdB (Figure 7c) (*64*); and in the C−D loop as for CcdB and PsaE (panels a and c of Figure 7, respectively).

In contrast, the D−E loop is structurally very conserved. It has been suggested that the $\beta_D−3_{10}$-helix−β_E motif acts as a lock for the barrel topology, thus playing a key role in preserving the SH3 fold. Due to this folding constraint, a long variable loop in the D−E region is probably not allowed (*65, 66*).

SH3-like domains are generally found in multidomain proteins and contribute to the specific recognition of partners. In the case of SH3 domains of protein kinases, a single domain can participate successively in various activation or inhibition processes involving different partners (*67*). Large variations in sequence, length, and flexibility of the loops are responsible for the specificity of each SH3 domain. Some structural features of the loops have been determined from ^{15}N relaxation in two cases: the Hck and Btk SH3 domains (*63, 68*). The highest flexibility was observed for the B−C loop, the so-called Src loop, with an S^2 of 0.5 for the Hck SH3 domain. The flexibility of the other loops decreases in the following order: A−B loop (RT loop), C−D loop (distal loop), and D−E loop. Interestingly, the flexible SH3 B−C loop is responsible for most of the interactions with the usual polyproline motif recognized by the SH3 domain. This structural plasticity is usually thought to be a major advantage for promoting interactions with multiple partners (*67, 69*).

PsaE appears to share some general features with SH3 domains, its three main loops, and especially the most flexible one (CD), interacting with protein partners (i.e., PSI subunits). An obvious difference with respect to the SH3 properties is that the interacting surfaces in PsaE have likely evolved to strengthen its association rather than to interact with multiple partners. PsaE behaves as a stable part of PSI, its loops becoming poorly accessible for interactions with multiple partners once it is inserted in the complex. This association, indeed, survives a long incubation time at low PSI concentrations (0.2 μM), as determined from absorption spectroscopy experiments (*70, 71*). The fast light-induced electron donation to ferredoxin, which requires a fine spatial tuning of many PSI residues (*11*), remains unaffected after several hours at room temperature. This excludes any significant PsaE release, which otherwise induces a clearly detectable 200-fold increase in the dissociation constant (*14*). Yet PsaE is one of the most accessible subunits of PSI, being the first to be released upon the action of chaotropic salts and the last to be inserted during biogenesis (*72*). In this context, the high flexibility of the C−D loop of PsaE may be a way to facilitate its association with the preformed photosystem by allowing its adaptation to a quite rigid macromolecular complex.

The recent improvement to 2.5 Å resolution of the structural model of *S. elongatus* PSI (*16*) allows a detailed structural comparison between the free and the PSI-bound PsaE subunit. *S. elongatus* PsaE belongs to the long C−D loop family, and its sequence is highly similar to that of S 6803 PsaE (Figure 2b). From the high-resolution PSI structure, it appears that PsaE establishes interatomic contacts with the other subunits exclusively through its loops. These contacts can be grouped into three distinct regions as highlighted in Figure 8c. The major region of contacts involves both the A−B and C−D loops, and these participate in a dense network of buried hydrogen bonds involving residues such as the two tyrosines (Y_{15} and Y_{34}) that are conserved in all cyanobacteria. Remarkably, these contacts connect PsaE with the six PSI subunits A−F (region I of
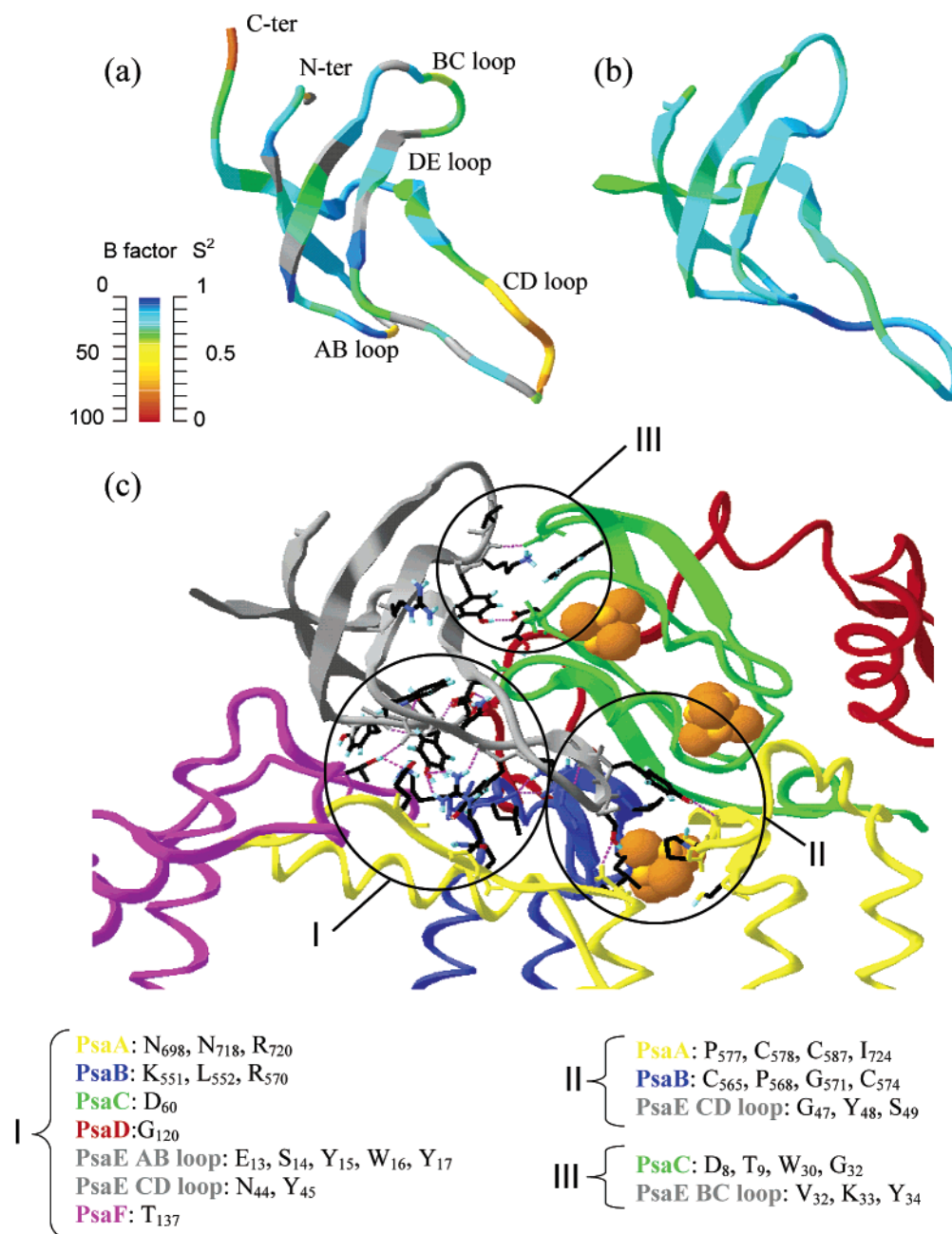
FIGURE 8: Comparison of free and bound PsaE. (a) Ribbon representation of the closest structure to the mean structure of free S 6803 PsaE. The ribbon is colored according to the order parameter as indicated in the corresponding scale. (b) Ribbon representation of *S. elongatus* PsaE bound to PSI (PDB entry 1C51) in the same orientation as free S 6803 PsaE. The ribbon is colored according to the *B*-factor as indicated in the corresponding scale. (c) Close-up view of the structure of *S. elongatus* PSI (PDB entry 1C51) centered on bound PsaE with the same orientation as in panels a and b. The different PSI subunits are represented with a colored ribbon: PsaA in yellow, PsaB in blue, PsaC in green, PsaD in red, PsaE in gray, and PsaF in magenta. Residues located on the contact surface between PsaE and the other subunits are shown with CPK color. Hydrogen bounds are represented with magenta dotted lines. Three regions of interactions noted I−III can be defined and are circled. The identities of the residues involved in the three regions of contact are indicated. Finally, the position of residue R_{39} is also indicated.

Figure 8c). A second region of interaction (region II of Figure 8) involves the C−D loop and both subunits PsaA and PsaB. Upon integration of PsaE, the C−D loop undergoes a structural rearrangement, generating a short β-sheet structure in contact with the region of PsaB that binds the iron−sulfur center Fx. This region which is structurally constrained in PSI has the lowest B factor values and is the best defined part of PsaE in the crystal structure (Figure 8b), whereas it is the most flexible region of the free PsaE (Figure 8a). Finally, the most external region of contacts between PsaE and PSI is located in the B−C loop that interacts only with

PsaC mainly through residues K_{33} and Y_{34} located close to the iron−sulfur cluster FB (region III of Figure 8c).

The regions of PsaE interacting with the other PSI subunits (in particular in region I of Figure 8c) involve complex networks of buried hydrogen bonds that are known to be highly sensitive to the orientation of the partners. Such interfaces are specific and structurally constrained. Moreover, PsaE is in close contact with the two cofactors of the PSI core which are involved in electron transfer to ferredoxin: Fx (region II in Figure 8c) and F_B (region III in Figure 8c). The integration of PsaE into PSI could thus allow a fine-

tuning of the distances between these cofactors and optimize the efficiency of the electron transfer. In this context, the high flexibility of the C−D loop of the free PsaE may be a structural advantage for this region which has to accommodate the preformed architecture during the assembly of this multimodular binding site.

The surface of PsaE buried (*73, 74*) in the complex is estimated to be around 350 Å². The loss of solvent accessibility mostly affects the loops, especially the A−B and C−D loops. The high stability of integrated PsaE is surprising when one is simply considering the total buried surface of 700 Å², which is significantly lower than the average value of 1600 ± 400 Å² reported for other binding interfaces (*75*). One must keep in mind that many other factors might contribute to the energetics of this insertion. The complex network of interactions among the A−D and F subunits is possibly not fully optimized in the absence of PsaE. This is suggested by the increased turnover rate of PsaD in PSI deprived of the PsaE subunit (*18*).

Since the interactions of PsaE with the other PSI subunits involve exclusively residues located in PsaE's loops, one could postulate that the core of the protein simply acts as a scaffold supporting the interacting loops. However, a recent study (*11*) has shown that one conserved arginine residue of PsaE located in strand β_C at position 39 plays a crucial role in the optimization of the interactions of PSI and ferredoxin. The absence of positive charge at this position was able to induce a 50-fold decrease in PSI affinity for ferredoxin, an effect similar in amplitude to that observed for the deletion of the full-length protein. The side chain of the arginine at position 39 is accessible to the solvent in the protein in solution and points to the stroma when in situ in PSI (Figure 8c). The huge functional effect induced by the mutation at this position raises the question of a possible misfold of the mutant protein. Figure 9 shows the superposition of the heteronuclear HSQC spectra recorded for the WT PsaE and the mutated protein R39Q. Both spectra are quasi-identical, with just residues A$_{25}$ and F$_{40}$ showing a small change in their chemical shift values. A significant modification of the general structure of the modified protein is thus excluded. These results are consistent with the accessibility of the arginine side chain to the solvent and the absence of NOE contacts between the side chain protons (other than Hα and Hβ) and protons belonging to other residues. This structural control confirms that arginine 39 at the C-terminal end of the β_C strand is a key site for tuning the interactions with ferredoxin.

PsaE can be described as having a modular architecture with a rigid central β-sheet motif supporting key residue(s) for the interaction with soluble electron acceptors (in particular R$_{39}$ for the interaction with ferredoxin) and several flexible regions that allow this polypeptide to integrate into the PSI complex. Two essential functions of PsaE now become obvious. (i) It accurately locates a key basic residue for the efficient trapping of ferredoxin, and (ii) it bridges six PSI subunits, probably improving the conformational stability of many residues involved in electron tranfer to ferredoxin. The multiple connections of PsaE on the acceptor site of PSI also suggest a third role as a sensor of the correct structure on this side of PSI. The release of PsaE after a limited number of hydrogen bonds are broken could be a signal for a turnover or reparation process to occur. The
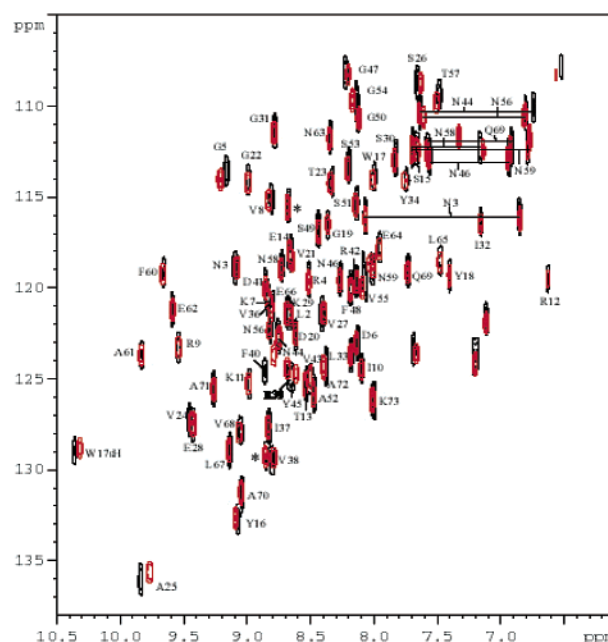

FIGURE 9: Superposition of the HSQC spectrum of wild-type S 6803 PsaE at pH 5.3 and 20 °C (black) and of the mutant R39Q (red). In these experiments, both proteins have been overexpressed using the GST fusion system. An asterisk indicates the two supplementary residues at the N-terminus used to generate the thrombin cleavage site (see Experimental Procedures).

diversity and complexity of the interactions involving PsaE and the number of interacting partners in the final PSI complex could reflect the central role of this small protein module in the formation of the entire functional PSI complex.

SUPPORTING INFORMATION AVAILABLE

Table of assignments of all proton and nitrogen resonances of S 6803 PsaE in 80 mM phosphate buffer at pH 5.3 at 293 K. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Goldbeck, J. H. (1994) in *Advances in Photosynthesis, The Molecular Biology of Cyanobacteria* (Bryant, D. A., Ed.) pp 319−360, Kluwer Academic Publishers, Dordrecht, The Netherlands.
2. Brettel, K. (1997) *Biochim. Biophys. Acta 1318*, 322−373.
3. Setif, P. (2001) *Biochim. Biophys. Acta 1507*, 161−179.
4. Knaff, D. B., and Hirasawa, M. (1991) *Biochim. Biophys. Acta 1056*, 93−125.
5. Zanetti, G., and Merati, G. (1987) *Eur. J. Biochem. 169*, 143−146.
6. Zilber, A. L., and Malkin, R. (1988) *Plant Physiol. 88*, 810−814.
7. Lelong, C., Setif, P., Lagoutte, B., and Bottin, H. (1994) *J. Biol. Chem. 269*, 10034−10039.
8. Lelong, C., Boekema, E. J., Kruip, J., Bottin, H., Rogner, M., and Setif, P. (1996) *EMBO J. 15*, 2160−2168.
9. Kruip, J., Chitnis, P. R., Lagoutte, B., Rogner, M., and Boekema, E. J. (1997) *J. Biol. Chem. 272*, 17061−17069.

10. Fischer, N., Hippler, M., Setif, P., Jacquot, J. P., and Rochaix, J. D. (1998) *EMBO J. 17*, 849−858.

11. Barth, P., Guillouard, I., Setif, P., and Lagoutte, B. (2000) *J. Biol. Chem. 275*, 7030−7036.

12. Hanley, J., Setif, P., Bottin, H., and Lagoutte, B. (1996) *Biochemistry 35*, 8563−8571.

13. Chitnis, V. P., Jungs, Y. S., Albee, L., Golbeck, J. H., and Chitnis, P. R. (1996) *J. Biol. Chem. 271*, 11772−11780.

14. Barth, P., Lagoutte, B., and Setif, P. (1998) *Biochemistry 37*, 16233−16241.

15. Klukas, O., Schubert, W. D., Jordan, P., Krauss, N., Fromme, P., Witt, H. T., and Saenger, W. (1999) *J. Biol. Chem. 274*, 7351−7360.

16. Jordan, P., Fromme, P., Witt, H. T., Klukas, O., Saenger, W., and Krauss, N. (2001) *Nature 411*, 909−917.

17. Cohen, Y., Steppuhn, J., Herrmann, R. G., Yalovsky, S., and Nechushtai, R. (1992) *EMBO J. 11*, 79−85.

18. Chitnis, P. R., and Nelson, N. (1992) *Plant Physiol. 99*, 239−246.

19. Jin, P., Sun, J., and Chitnis, P. R. (1999) *Biochim. Biophys. Acta 1410*, 7−18.

20. Falzone, C. J., Kao, Y. H., Zhao, J., Bryant, D. A., and Lecomte, J. T. (1994) *Biochemistry 33*, 6052−6062.

21. Mayer, K. L., Shen, G., Bryant, D. A., Lecomte, J. T., and Falzone, C. J. (1999) *Biochemistry 38*, 13736−13746.

22. Lagoutte, B., and Vallon, O. (1992) *Eur. J. Biochem. 205*, 1175−1185.

23. Rousseau, F., Setif, P., and Lagoutte, B. (1993) *EMBO J. 12*, 1755−1765.

24. McIntosh, L. P., and Dahlquist, F. W. (1990) *Q. Rev. Biophys. 23*, 1−38.

25. Hakes, D. J., and Dixon, J. E. (1992) *Anal. Biochem. 202*, 293−298.

26. Marion, D., Kay, L. E., Sparks, S. W., Torchia, D. A., and Bax, A. (1989) *J. Am. Chem. Soc. 111*, 1515−1517.

27. Shaka, A. J., Barker, P. B., and Freeman, R. (1985) *J. Magn. Reson. 64*, 547−552.

28. Vuister, G. W., and Bax, A. (1994) *J. Biomol. NMR 4*, 193−200.

29. Nilges, M., Macias, M. J., O'Donoghue, S. I., and Oschkinat, H. (1997) *J. Mol. Biol. 269*, 408−422.

30. Gilquin, B., Lecoq, A., Desne, F., Guenneugues, M., Zinn-Justin, S., and Menez, A. (1999) *Proteins 34*, 520−532.

31. Pardi, A., Billeter, M., and Wuthrich, K. (1984) *J. Mol. Biol. 180*, 741−751.

32. Savarin, P., Zinn-Justin, S., and Gilquin, B. (2001) *J. Biomol. NMR 19*, 49−62.

33. Brünger, A. T. (1992) *X-PLOR, Version 3.1. A system for X-ray Crystallography and NMR*, Yale University Press, New Haven, CT.

34. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994) *Biochemistry 33*, 5984−6003.

35. Piotto, M., Saudek, V., and Sklenar, V. (1992) *J. Biomol. NMR 2*, 661−665.

36. Grzesiek, S., and Bax, A. (1993) *J. Am. Chem. Soc. 115*, 12593−12594.

37. Pons, J. L., Malliavin, T. E., and Delsuc, M. A. (1996) *J. Biomol. NMR 8*, 445−552.

38. Rouh, A., Delsuc, M. A., Bertrand, G., and Lallemand, J. Y. (1993) *J. Magn. Reson. 102*, 357−359.

39. Palmer, A. G., III, Rance, M., and Wright, P. E. (1991) *J. Am. Chem. Soc. 113*, 4371−4380.

40. Mandel, A. M., Akke, M., and Palmer, A. G., III (1995) *J. Mol. Biol. 246*, 144−163.

41. Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (1988) *Numerical recipes in C: The art of scientific computing*, Cambridge Unviversity Press, Cambridge, England.

42. Peng, J. W., and Wagner, G. (1992) *J. Magn. Reson. 98*, 308−332.

43. Peng, J. W., and Wagner, G. (1992) *Biochemistry 31*, 8571−8586.

44. Farrow, N. A., Zhang, O., Forman-Kay, J. D., and Kay, L. E. (1995) *Biochemistry 34*, 868−878.

45. Lefevre, J. F., Dayie, K. T., Peng, J. W., and Wagner, G. (1996) *Biochemistry 35*, 2674−2686.

46. van Heijenoort, C., Penin, F., and Guittet, E. (1998) *Biochemistry 37*, 5060−5073.

47. Farrow, N. A., Zhang, O., Szabo, A., Torchia, D. A., and Kay, L. E. (1995) *J. Biomol. NMR 6*, 153−162.

48. Ishima, R., and Nagayama, K. (1995) *Biochemistry 34*, 3162−3171.

49. Ishima, R., Yamasaki, K., Saito, M., and Nagayama, K. (1995) *J. Biomol. NMR 6*, 217−220.

50. Peng, J. W., and Wagner, G. (1995) *Biochemistry 34*, 16733−16752.

51. Lipari, G., and Szabo, A. (1982) *J. Am. Chem. Soc. 104*, 4546−4559.

52. Lipari, G., and Szabo, A. (1982) *J. Am. Chem. Soc. 104*, 4559−4570.

53. Clore, G. M., Driscoll, P. C., Wingfield, P. T., and Gronenborn, A. M. (1990) *Biochemistry 29*, 7387−7401.

54. Clore, G., Szabo, A., Bax, A., Kay, L. E., Driscoll, P. C., and Gronenborn, A. M. (1990) *J. Am. Chem. Soc. 112*, 4989−4991.

55. Kay, L. E., Torchia, D. A., and Bax, A. (1989) *Biochemistry 28*, 8972−8979.

56. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley and Sons, New York.

57. Richardson, J. S. (1976) *Proc. Natl. Acad. Sci. U.S.A. 73*, 2619−2623.

58. Murzin, A. G., Brenner, S. E., Hubbard, T., and Chothia, C. (1995) *J. Mol. Biol. 247*, 536−540.

59. Lo Conte, L., Brenner, S. E., Hubbard, T. J., Chothia, C., and Murzin, A. G. (2002) *Nucleic Acids Res. 30*, 264−267.

60. Wang, G., Wylie, G. P., Twigg, P. D., Caspar, D. L., Murphy, J. R., and Logan, T. M. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 6119−6124.

61. Yu, H., Rosen, M. K., Shin, T. B., Seidel-Dugan, C., Brugge, J. S., and Schreiber, S. L. (1992) *Science 258*, 1665−1668.

62. Horita, D. A., Baldisseri, D. M., Zhang, W., Altieri, A. S., Smithgall, T. E., Gmeiner, W. H., and Byrd, R. A. (1998) *J. Mol. Biol. 278*, 253−265.

63. Hansson, H., Mattsson, P. T., Allard, P., Haapaniemi, P., Vihinen, M., Smith, C. I., and Hard, T. (1998) *Biochemistry 37*, 2912−2924.

64. Loris, R., Dao-Thi, M. H., Bahassi, E. M., Van Melderen, L., Poortmans, F., Liddington, R., Couturier, M., and Wyns, L. (1999) *J. Mol. Biol. 285*, 1667−1677.

65. Murzin, A. G., Lesk, A. M., and Chothia, C. (1994) *J. Mol. Biol. 236*, 1369−1381.

66. Murzin, A. G., Lesk, A. M., and Chothia, C. (1994) *J. Mol. Biol. 236*, 1382−1340.

67. Dalgarno, D. C., Botfield, M. C., and Rickles, R. J. (1997) *Biopolymers 43*, 383−400.

68. Horita, D. A., Zhang, W., Smithgall, T. E., Gmeiner, W. H., and Byrd, R. A. (2000) *Protein Sci. 9*, 95−103.

69. Sundberg, E. J., and Mariuzza, R. A. (2000) *Struct. Folding Des. 8*, 137−142.

70. Setif, P. Q., and Bottin, H. (1994) *Biochemistry 33*, 8495−8504.

71. Setif, P. Q., and Bottin, H. (1995) *Biochemistry 34*, 9059−9070.

72. Nechushtai, R., and Nelson, N. (1985) *Plant Mol. Biol. 4*, 377−384.

73. Vriend, G., and Sander, C. (1991) *Proteins 11*, 52−58.

74. Voorintholt, R., Kosters, M. T., Vegter, G., Vriend, G., and Hol, W. G. (1989) *J. Mol. Graphics 7*, 243−245.

75. Lo Conte, L., Chothia, C., and Janin, J. (1999) *J. Mol. Biol. 285*, 2177−2198.

76. Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) *J. Biomol. NMR 8*, 477−486.