The original publication is available at http://www.blackwell-synergy.com/

European Journal of Biochemistry.

270 (19), 3916–3927

doi:10.1046/j.1432-1033.2003.03774.x
The single-tryptophan of the PsbQ protein of photosystem II is at the end of a 4-α-helical bundle domain

Mónica Balsera¹, Juan B. Arellano¹, Florencio Pazos²,³, Damien Devos²,⁴, Alfonso Valencia²*, & Javier De Las Rivas¹

¹Instituto de Recursos Naturales y Agrobiología (CSIC), Cordel de Merinas 52, Salamanca 37008 (Spain); ²Centro Nacional de Biotecnología (CSIC), Cantoblanco, Madrid 28049 (Spain)
³current address: Imperial College, London U.K.
⁴current address: University of California, San Francisco USA

*Corresponding author: Alfonso Valencia. Centro Nacional de Biotecnología (CSIC), Cantoblanco, Madrid 28049 (Spain). Phone +34 91 585 45 70. Fax +34 9 585 45 06. e-mail: valencia@cnb.uam.es.

Running title: 3D structural analysis of PsbQ

†This study was supported by funds from the Spanish Ministry of Science and Technology (project PB1998-0480) and the V European Union Framework Programme. M. Balsera holds a fellowship from the Spanish Ministry of Science and Technology.
Abbreviations:

Chl, chlorophyll; FWHM, full width at half maximum; Gdn-HCl, guanidine hydrochloride; PMSF, Phenylmethylsulphonyl fluoride; PSII, photosystem II;

SUMMARY

We examined the microenvironment of the single tryptophan and the tyrosine residues of PsbQ, one of the three main extrinsic proteins of green algal and higher plant photosystem II (PSII). On the basis of this information and the previous data on secondary structure (Balsera, M., Arellano, J.B., Gutiérrez, J.R., Heredia, P., Revuelta, J.L. & De Las Rivas, J. (2003). *Biochemistry* **42**, 1000-1007), we screened structural models derived by combining various threading approaches. Experimental results show that the tryptophan residue is partially buried in the core of the protein but still in a polar environment, according to the intrinsic fluorescence emission of PsbQ and the fact that fluorescence quenching by iodide was weaker than that by acrylamide. Furthermore, quenching by cesium suggests that a positively-charged barrier shields the tryptophan microenvironment. Comparison of the absorption spectra in native and denaturing conditions indicates that 1-2 out of 6 tyrosines of PsbQ are buried in the core of the structure. By threading methods a 3D structural model was built for the C-terminal domain of the PsbQ protein family (residues 46-149), while the N-terminal domain is predicted to have a flexible structure. The model for the C-terminal domain is based on the 3D structure of cytochrome $b_{562}$, a mainly $\alpha$ protein with a helical up-down bundle folding. Despite the large sequence differences between the template and PsbQ, the structural and energetic parameters for the explicit model are acceptable, as
judged by the corresponding tools. This 3D model is compatible with the experimentally
determined environment of the tryptophan residue and with published structural
information, however, it has a limited value until an experimental 3D structure for PsbQ is
available.

*Keywords: extrinsic proteins, photosystem II, PsbQ, threading, three-dimensional model*
INTRODUCTION

Photosystem II (PSII) is a type-II reaction center found in thylakoids of all oxygenic photosynthetic organisms (cyanobacteria, algae and higher plants), which harnesses light energy to oxidize water, producing molecular oxygen as a side-product [1, 2, 3, 4]. The structure of the core of this pigment-protein complex, which consists of about twenty-five (intrinsic and extrinsic) proteins denoted as PsbA-Z, has been X-ray resolved at 3.8 Å and 3.7 Å for two species of *Synechococcus* [5, 6]. The 3D structures of these two PSII core complexes show the arrangement of some Psb proteins, chlorophylls and other cofactors, and also suggest some possible ligands for the Mn cluster, where water is oxidised. For a functional Mn cluster, other ionic cofactors such as Ca$^{2+}$ and Cl$^-$ are required [7, 8, 9], however, there is no clue as to where these two latter cofactors are localized in the X-ray structure of PSII. The three luminal extrinsic proteins PsbO, PsbV and PsbU, observed in the 3D structure of the PSII core of *Thermosynechococcus vulcanus*, have a role in the stabilization of the Mn cluster and its ionic cofactors Ca$^{2+}$ and Cl$^-$, and also in the overall (thermo)stability of PSII [10, 11, 12]. PsbO is the only orthologous PSII extrinsic protein found in all oxygenic photosynthetic organisms, but PsbV and PsbU are only present in cyanobacterial and red algal PSII. Exceptionally, there is a fourth extrinsic protein of 20 kDa in red algal PSII that is not found in any of the other PSII complexes [13]. PsbP and PsbQ are the counterparts of PsbV and PsbU in green algae and higher plants [10]. All these PSII extrinsic proteins facilitate oxygen evolution, but they differ in their specific binding to PSII. PsbO is the only extrinsic protein totally exchangeable without loss of function, in binding to PSII of any of the former photosynthetic organisms. In contrast, the
red algal PsbU and PsbV are only partially functional, and PsbP and PsbQ are not functional when binding to PSII of either of the other two photosynthetic organisms [14]. Differences in the binding properties of green algal and higher plant PsbP and PsbQ have also been observed [15], suggesting that the former do not need the presence of PsbO when (re)binding to PSII. Moreover, it has been suggested that the structure of some of these extrinsic proteins depends on the organism [15, 16]. The specific binding sites for PsbO, PsbP and PsbQ in the lumenal side of green algal and higher plant PSII is less known than in cyanobacterial PSII. In higher plants, PsbO is believed to have an extended structure that lies on the surface of CP47/D2 (PsbB/PsbD) [17, 18], but also on the surface of CP43/D1 (PsbC/PsbA) [19]. Intriguingly, the arrangement for the higher plant PsbO is slightly different from that observed in the X-ray resolved cyanobacterial PSII. On the other hand, PsbP and PsbQ are positioned at the N-terminus of D1 [17, 20]. In addition, PsbQ requires the presence of PsbP when binding to higher plant PSII, but there is no direct evidence for their mutual interaction [10]. Likewise, the partial degradation of the N-terminal regions of PsbP and PsbQ results respectively in a decrease in, and in a complete loss of, binding affinity for the lumenal side of PSII [21, 22].

From a functional point of view, there is a consensus that PsbO stabilizes the Mn cluster [10], but several roles have been assigned to the other two (or three) extrinsic proteins. In cyanobacteria, PsbV and PsbU maintain the overall stability of PSII, but PsbU may also optimise the Ca$^{2+}$ and Cl$^{-}$ environment in the Mn cluster [23]. In red algae, oxygen evolution is strongly dependent on Ca$^{2+}$ and Cl$^{-}$ in the absence of PsbV and PsbU, pointing out that they both play a similar role to PsbP and PsbQ in green algae and higher
plants [13, 24, 25, 26, 27, 28]. Other functions proposed for PsbP and PsbQ are: (i) to form a gate that is open for substrates [26] and products [29], but closed to non-physiological reducing agents [30]; (ii) to create a low dielectric medium that is optimal for PSII binding to Ca$^{2+}$ [31] and Cl$^{-}$ [32] and (iii) to tune up the magnetic properties of the Mn cluster [33].

It will be very useful address the analysis of the complex with information about the structure of the individual complexes. Unfortunately little is known about the three-dimensional structure of PsbP and PsbQ, compared with the wealth of information about PsbO, PsbV and PsbU [6, 34, 35, 36]. In our previous report [37] we suggested that the PsbQ protein had two different structural domains: the N-terminal (residues 1-45) with a non-canonical secondary structure and the C-terminal (residues 46-149) with a mostly $\alpha$-helix structure. Now, we propose a 3D model for the C-terminal domain of PsbQ based on its structural analogy with the 3D-known-structure of a protein, using threading and modelling. The resulting model is compatible with the previously obtained information on the secondary structure of the protein and with the experimental results obtained from changes in the absorption of the protein in denaturing conditions, protein tryptophan fluorescence emission and fluorescence quenching.

MATERIALS AND METHODS

Material and chemicals
Spinach leaves were purchased at the local market. Guanidine hydrochloride (Gdn-HCl), CsCl, Na$_2$S$_2$O$_3$ and acrylamide were from Sigma-Aldrich Corp. (St. Louis, MO). KI was
from Merck & Co. Inc. (Whitehouse Station, NJ). All these chemicals were of reagent grade and used without further purification.

Isolation and purification of the PsbQ protein from spinach

PSII-enriched membranes were isolated from spinach leaves as described elsewhere [38] with some modifications [39]. Total chlorophyll (Chl) and Chla/Chlb ratio were determined spectrophotometrically by the method of Arnon [40]. PSII-enriched membranes with a concentration of 4-6 mg Chl·mL⁻¹ were stored frozen at -80 ºC until use. When purifying PsbQ, PSII-enriched membranes were washed in a 10-fold excess of 20 mM Mes pH 6.0 and then centrifuged at 40,000x g for 30 min at 4 ºC. The pellet was suspended in 20 mM Mes pH 6.0 with 10 mM CuCl₂ to a concentration of 2 mg Chl·mL⁻¹. PSII-enriched membranes were incubated for 1 h at room temperature, followed by a centrifugation at 40,000x g for 30 min at 4 ºC. The pH of the supernatant was adjusted to 8.0, by adding alternatively pH-unadjusted stock solutions of Tris and EDTA. The final concentration of EDTA was 5 mM. In these conditions the pale blue of the supernatant at pH 6.0 turned to deep blue at pH 8.0. The supernatant was passed through a 0.45 μm pore-size syringe filter and kept at 4 ºC without any further treatment until chromatography. The chromatographic steps were carried out in an Äktapurifier-100 apparatus, (Amersham Pharmacia Biotech. UK Limited, Buckinghamshire, England). The native PsbQ protein was first passed through a cation-exchange High-Trap SP column (1 mL), Amersham Biosciences AB (Uppsala, Sweden), and then through a gel filtration Superdex-200 column HR 10/30, Amersham Biosciences AB, both pre-equilibrated with 20 mM Tris-HCl, pH 8.0, 35 mM
NaCl, 1 mM EDTA and 1 mM phenylmethylsulphonyl fluoride (PMSF). Further details of these two chromatographic steps are given in [37].

** SDS-PAGE analysis**

The SDS-PAGE analysis was carried out using a Protean II xi Cell, Bio-Rad Laboratories (Hercules, CA), according to Laemmli [41] with a total acrylamide content of 17% in the resolving SDS-polyacrylamide gel. The SDS-polyacrylamide gels were stained with Coomassie R-250.

**Protein concentration and absorbance measurements**

Absorption spectra were recorded in a Cary 100 UV-visible spectrophotometer, Varian Inc. (Palo Alto, CA), using a scan rate of 30 nm·min⁻¹ at 20 °C. The PsbQ protein concentration was determined from the sum of the extinction coefficients of its aromatic amino acids at 276 nm in 6 M Gdn-HCl as described in [42], such determination yielded a molar extinction coefficient of 14100 M⁻¹·cm⁻¹. The degree of tyrosine exposure (α) was calculated from the second-derivative spectrum [43], as follows: \( \alpha = \frac{r_n - r_u}{r_n - r_a} \), where \( r_n \) and \( r_u \) are the experimentally-determined numerical values of the ratio \( a/b \), and \( r_a \) is the theoretical numerical value of ratio \( a/b \) for a mixture of aromatic amino acids (Tyr and Trp), containing the same molar ratio as the protein under study, dissolved in a model solvent (i.e. ethylene glycol), which possesses the same characteristics of the interior of the protein matrix. The script \( a \) is the peak-peak distance between the maximum around 287 nm and the minimum around 283 nm, and the script \( b \) is the peak-peak distance between
the maximum around 295 nm and the minimum around 290 nm in the second derivative absorption spectrum of the protein.

**Fluorescence emission spectra and fluorescence quenching**

Fluorescence emission spectra were recorded in a steady-state spectrofluorometer Model QM-2000-4, Photon Technology International Inc. (Lawrenceville, NJ), equipped with a refrigerated circulator. Fluorescence emission spectra were recorded in 0.5 cm path quartz cells at 20 °C. Both excitation and emission monochromators were set at 3-nm slit widths. Protein samples were excited at 280 or 295 nm. Fluorescence emission spectra were recorded from 300 to 500 nm with steps of 0.5 nm and an integration time of 2 s, averaged 3 times, and corrected by subtracting the Raman band and the buffer signal. During measurement, stock solutions of PsbQ were diluted in 50 mM Tris-HCl pH 8.0 and their concentration was kept between 3.5-10 μM. A final concentration of 6 M Gdn-HCl was used when denaturing PsbQ. Polar uncharged acrylamide and KI and CsCl salts were used for performing collisional quenching of protein tryptophan fluorescence at 20 °C. NaCl was added to maintain a constant ionic strength. The 4 M stock solution of KI contained 1 mM Na$_2$S$_2$O$_3$ to avoid the formation of I$^-$ [44]. Fluorescence intensities were corrected when adding acrylamide [45]. The fluorescence quenching was analyzed following the classical and modified Stern-Volmer equations, $F_0/F = 1 + K_{sv}[Q]$ or $F_0/F = (1 + K_{sv}[Q]) \times e^{-[Q]}$ [46, 47], where $F_0$ and $F$ are the fluorescence intensities in the absence and presence of the quencher Q, $K_{sv}$ is the collisional quenching constant, and $V$ is the static constant, which is related to the probability of finding a quencher molecule close enough to a newly formed excited state to quench it immediately.
Bioinformatic methods

Multiple sequence alignment and secondary structure prediction of the PsbQ protein family were obtained previously [37]. The threading programs used to predict a fold for PsbQ were: FFAS [48], THREADER2 [49], 3DPSSM [50], FUGUE [51], 123D+ [52] and BIOINBGU [53]. These programs propose a list of protein hits whose 3D-known structure could be similar to the query protein based on features of the PsbQ sequence family such as secondary structure, solvent accessibility, contact potentials, etc. These methods cover a vast range of available threading strategies, based on clearly different principles and libraries. The templates are scored by a reliability index, usually a z-score, which measures the difference in score between the raw score of a query-template alignment and the distribution of the scores for all the templates in the fold library. The protein fold recognition protocol proceeded as follows. First, a set of candidate folds was chosen based not only on the scores of the three best hits proposed by each threading program, but also on the fold similarity among the three best hits using the FSSP database [54]. Then, 1D and 3D alignments of the query protein with each of the hit templates were inspected using the THREADLIZE package [55]. In addition, CLUSTALX [56] was used to align the PsbQ sequence and the profiles of the proposed structures derived from the alignments deposited in the HSSP database [57]. The quality of each alignment was evaluated by the number and distribution of gaps, percentage of identity and distribution of hydrophobic residues. Once a template was chosen, a full atom 3D model based on the threading alignment was obtained with the Swiss-Model automated modelling server [58] and evaluated using the WHATCHECK [59], PROMODII [60], VERIFY3D [61] programs, and the distribution of the conserved residues based on the Xd parameter [62]. This latter parameter measures the
distribution of the distances between the conserved residues and all the residues, since the most conserved residues are those implicated in the structure and/or function and appear clustered in the structure [63].

RESULTS

Isolation and purification of the PsbQ protein

The use of 10 mM CuCl₂ to release the extrinsic PsbQ protein from PSII-enriched membranes was based on the finding of Jegerschöld et al [64]. When adding 6-7 mM CuSO₄ to a PSII preparation to examine the effect of Cu²⁺ on PSII activity by EPR, the former authors reported a concomitant 90% loss of PsbQ, whereas the other two extrinsic proteins (PsbO and PsbP) remained largely bound. This observation gains interest if we also bear in mind that Cu²⁺ at (sub)milimolar concentrations inhibits a specific prolyl-endopeptidase for PsbQ, a protease that cleaves the N-terminus at the carboxyl side of the fourth and twelfth proline residues of PsbQ from spinach [65]. Standards protocols to release the extrinsic peptides of PSII include high salt concentration washes [10]. However, the 1 M NaCl wash, frequently selected to release PsbP and PsbQ, also detaches the prolyl-endopeptidase. When removing NaCl by prolonged dialysis, this protease is activated and cleaves PsbQ at low salt conditions. We circumvented the drawbacks of the 1 M NaCl wash by taking advantage of the Cu²⁺ effect. In this latter case, first, the prolyl-endopeptidase (if present in the supernatant) is expected to be largely inhibited by 10 mM CuCl₂ and second, prolonged dialysis is not required before chromatography due to the very low ionic strength of the 10 mM CuCl₂ washing buffer. Incubation of the PSII-enriched membranes with this buffer yielded a supernatant containing PsbQ, but also some PsbO and a little PsbP (Fig. 1,
The first chromatographic step in the cationic-exchange High-Trap SP column was very similar to the one described in [37], except that larger volumes of the supernatant were loaded due to its lower protein concentration, and also that the High-Trap SP column was thoroughly washed with the pre-equilibrating buffer (10-15 mL) to remove unbound materials and also traces of Cu$^{2+}$. After the linear salt gradient, the PsbQ-enriched fractions were pooled, concentrated and loaded onto the Superdex 200 HR 10/30 column. After filtration through this latter column, the fractions containing PsbQ (Fig. 1, lane d) were kept at 4 °C until use.

**Absorbance spectrum**

The aromatic amino acids (and also cystine if present) are responsible for the absorption band of proteins in the near UV region. The sequence of the PsbQ protein from spinach contains one tryptophan, six tyrosines, and four phenylalanines. Figure 2A shows the overall contribution of these 11 aromatic amino acids to the absorption spectrum of PsbQ in the region 260-310 nm. In native conditions, a maximum at 277.5 nm and two shoulders at ~282 and ~292 nm are inferred from the absorption spectrum of PsbQ. A hypsochromic shift of 1-2 nm is observed in the absorption spectrum of this protein in denaturing conditions (6 M Gdn-HCl). This shift may be due to changes in the microenvironment of tyrosine residues that become more polar following protein denaturation [43]. According to the equation for $\alpha$ (see Material and Methods), the degree of tyrosine exposure can be estimated from the second-derivative of the absorbance spectra of PsbQ when determining the ratio $a/b$ in native and denaturing conditions (Fig. 2A). The values for $r_n$ and $r_c$ were about 2.6 and 3.6, respectively, and the value for $r_a$ was -0.58 (see reference 43). The
resulting value for $\alpha$ was 0.76, indicating that one or two tyrosine residues are not solvent exposed in PsbQ.

**Fluorescence measurements**

The single tryptophan amino acid present in PsbQ from spinach is fully conserved throughout the PsbQ sequence family [37]. This aromatic amino acid can specifically be excited at an excitation wavelength beyond 295 nm [47]. Therefore, the intrinsic fluorescence emission spectrum of PsbQ depends only on the microenvironment that surrounds the tryptophan residue, so it can indicate the extent to which this residue is exposed to the solvent [45]. The intrinsic fluorescence emission spectrum of PsbQ has a maximum at 327 nm and a full width at half maximum (FWHM) of 53 nm at 20 ºC in native conditions (Fig. 2B). However, quenching of the fluorescence intensity and a bathochromic fluorescence shift of the emission peak from 327 nm to 353 nm are observed in denaturing conditions (6 M Gdn-HCl), suggesting that the microenvironment of the tryptophan residue is exposed to the solvent in the denatured state. At 280 nm, tyrosine (and also tryptophan) residues emit. Thus, the intrinsic fluorescence emission spectrum of PsbQ has a maximum at about 323 nm at 20 ºC. The normalization at 400 nm [66] of the two spectra of PsbQ, at 295 and 280 nm, shows that the fluorescence emission due to tyrosine is weak. This suggests that there is an efficient singlet-singlet energy transfer from Tyr (to Tyr) to Trp. The difference between the two fluorescence emission spectra clearly shows a weak band centered at 304 nm. It corresponds to the fluorescence emission of Tyr residues in PsbQ [66] that did not transfer their excitation energy due to either a long Tyr-Trp distance or an inefficient Tyr-Trp transition dipole orientation.
Quenching of tryptophan fluorescence by iodide, cesium ion and acrylamide

Aqueous fluorescence collision quenchers have extensively been used to measure the exposure of tryptophan residues to the aqueous environment [44, 67]. The efficiencies of the indole fluorescence quenching for acrylamide and I\textsuperscript{–} have been shown to be unity, which is five times higher than the efficiency for Cs\textsuperscript{+} [46]. Cs\textsuperscript{+} and I\textsuperscript{–} are two quenchers that may collide with exposed indole groups, and also with groups located in a negative or positive environment, respectively. Acrylamide can quench both exposed and unexposed residues [67]. Figure 3 shows the dependence of the relative intrinsic fluorescence intensity of PsbQ with the quencher concentration monitored at 320 nm when exciting at 295 nm. No bathochromic fluorescence shift was observed for any quencher when concentration increased, indicating the absence of protein denaturation (data not shown). Whereas a linear dependence inferred between the intrinsic fluorescence intensity of PsbQ and the concentration of Cs\textsuperscript{+} or I\textsuperscript{–}, an upward curve was obtained with increasing concentrations of acrylamide, suggesting some static quenching [47, 67]. The Cs\textsuperscript{+} and I\textsuperscript{–} results were represented with the classical Stern-Volmer plot but a modified plot was used for acrylamide to obtain both the collisional ($K_{sv}$) and static ($V$) quenching constants. All fluorescence measurements for the three quenchers were carried out at the same ionic strength (0.2 M NaCl), though a second ionic strength (1 M NaCl) was used for I\textsuperscript{–}. The collisional quenching constant is greater for the polar uncharged acrylamide ($K_{sv} = 3.2 \pm 0.1 \text{ M}^{-1}$) than the respective ones for the ionic quenchers, and likewise greater for the anionic quencher I\textsuperscript{–} ($K_{sv} = 1.2 \pm 0.1 \text{ M}^{-1}$) than for the cationic one Cs\textsuperscript{+} ($K_{sv} = 0.0 \text{ M}^{-1}$). The modified Stern-Volmer equation gives a static quenching constant ($V$) for acrylamide of
0.11 ± 0.07 M$^{-1}$. $K_{sv}$ for acrylamide did not change when increasing the ionic strength of the solvent, but the collisional quenching constant showed a decrease for I$^-$ at 1M NaCl ($K_{sv} = 0.67 ± 0.04$ M$^{-1}$). All these results suggest that the tryptophan residue is to some extent buried in the PsbQ protein matrix, where the polar uncharged acrylamide can diffuse to, but where the ionic compounds have little access. In addition, the effect of the ionic strength on the quenching of the tryptophan fluorescence by I$^-$ [44] and the lack of quenching by Cs$^+$ indicate that a positive-charged barrier is shielding the tryptophan microenvironment.

**PsbQ fold recognition**

An exhaustive search for 3D-known structure homologous protein to PsbQ through all known public biological databases did not reveal any possible protein of known 3D-structure on which to build models of the PsbQ. Therefore, a fold recognition approach by threading methods was carried out in the search for remotely related structures, using both the spinach PsbQ sequence and the PsbQ family alignment as references [37]. The three best hits of the threading methods are shown in Table 1A. Most of them pointed mainly to $\alpha$-helix proteins (14 out of 18) as candidate models for PsbQ: the up-down and orthogonal bundles were the most frequent architectures. The threading programs did not identify candidate folds for the region of the sequence corresponding to the N-terminal domain (residues 1-45). As new threading runs excluded this domain, the selection of mainly $\alpha$-helix templates became even clearer (13 out of 15) (Table 1B). Among all the possibilities for PsbQ, the four $\alpha$-helix up-down bundle appeared to be the dominant topology, judging by the proportion (one third of all the cases) and the confidence level of the hits. The hits
1vltB0 and 1aep00 had a confidence level of more than 80%. They correspond to different proteins of the same CATH [68] family (1.20.120.x, Table 1). Although most of the scores of the other predictions are below these confidence levels two other structures, 1ego00 and 1jafA0, were selected by two or more programs (THREADER2, 3D-PSSM, 123D+ and FUGUE, 3D-PSSM respectively). Furthermore, when the prediction was restricted to the second domain (C-terminal), 1qsdA0 and 256bA0 were identified as templates by two methods (BIOINBGU, 3D-PSSM and 3D-PSSM, 123D+ respectively). All these six potential targets have similar topology and structure (see their FSSP database [54] classification of α−helical up-down bundle structures, Fig. 4A). The family includes proteins that are homogeneous in structure but heterogeneous in sequence and function, i.e. 1vlt (aspartate receptor) and 256b (cytochrome b_{256}) are 20% identical. Other structural architecture proposed by several threading methods was a mainly α orthogonal bundle. This architecture (CATH 1.10.x.x) appears in 4 out of 18 candidates when taking the whole PsbQ sequence (Table 1A) and 5 out of 15 when taking only the C-terminal domain (Table 1B). However, the topology of these structures did not correspond to a unique topological family. Based on the predicted secondary structure [37] a clear distribution of amphipathic residues is shown, with the non-polar residues forming one of the faces of the helices (Fig. 4B). This distribution favours a parallel packing of the 4 α-helices, supporting the up-down bundle architecture rather than the orthogonal. The length of the connecting loops between helices also supports the up-down bundle topology.

Selection of the best PDB template for the C-terminal domain of PsbQ
In order to select the best template to construct a remote 3D model for the C-terminal domain (residues 46-149) of PsbQ, the structural alignments between the problem PsbQ protein and each of the threading hits (Table 1) were manually inspected using the THREADLIZE package [55], bearing in mind the compatibility of the predicted [37] and known secondary structures. Also, the quality of the sequence alignment between the families (each threading hit family obtained from HSSP database), the number and distribution of gaps, the sequence homology and the hydropathy profile were analysed. After this manual process, the best fitting between PsbQ and the chain A of 256b (256bA0) was selected. This protein is a periplasmic cytochrome $b_{562}$ of *E. coli* with a mass of 11.78 kDa and unknown function [69]. The 256bA0 structure consists of four main $\alpha$-helices, and a $3_{10}$ helix at the end of the second helix, that fold as a helical up-down bundle. The 1D sequence alignment between C-terminal domain of PsbQ and 256bA0 is shown in Fig. 5A. This alignment is compatible with the complete PsbQ family alignment (data not shown). In spite of the low sequence identity (about 8%), a good match of the corresponding secondary structures and hydropathy profiles was obtained (data not shown).

**3D threading model for the C-terminal domain of PsbQ protein**

A full-atom model for the C-terminal domain of PsbQ was obtained using the SWISSMODEL [58] program based on the threading alignment between PsbQ and cytochrome $b_{562}$ (Fig. 5C). The WHATCHECK [59] and PROMOD [60] programs were used to evaluate the models. The corresponding parameters obtained were: Ramachandran plot=-0.290, backbone conformation=-0.609, chi-1/chi-2 rotamer normality=-0.945, bond lengths=0.791, and bond angles=1.353 and the energetic parameter of the model was $E=$-
3160 kJ·mol\(^{-1}\). Bond lengths and angles in close to the optimal value of 1. Ramachandran plot, backbone conformation and chi-1/chi-2 rotamer normality correspond to z-scores and therefore a positive value indicates better than average and their maximum values are around 4. The values for all these parameters obtained for the PsbQ model are quite good and the programs did not mark any of them as poor or inappropriate. Another structural analysis, obtained by the VERIFY3D program [61], gave an average value of 0.21, which is greater than zero, the quality value indicated by the program. In addition, the distribution of distances between conserved residues and between all the residues was calculated, as was the Xd parameter (see Materials and Methods, [62]). The value of Xd was greater than zero (Xd = 9.5), which indicates that the conserved residues are close to each other in the structure, as is typical for normal proteins. Moreover, visual inspection of the models revealed a correct distribution of conserved residues in the hydrophobic core of the structure and also in the loops connecting helix2 and helix3.

**DISCUSSION**

In a previous paper [37] a secondary structure analysis of the PsbQ spinach protein was carried out by CD and FTIR spectroscopy and bioinformatic tools. It was concluded that PsbQ was mainly α protein, with two different structural domains: a minor N-terminal domain, with a poorly defined secondary structure enriched in proline and glycine amino acids (residues 1-45), and a major C-terminal domain containing four α-helices (residues 46-149). We have now extended the study on PsbQ by building a 3D model based on a fold recognition computational approach. The computational searches did not reveal any
structural template for the N-terminal region of PsbQ, probably due to its apparent lack of stable structure. A search for disorder segments in the PsbQ sequence was also performed using the PONDR program [70]. The result suggested the N-terminal segment (residues 4-27) is the longest and most disorder region of PsbQ (data not shown), confirming our previous results [37]. For the rest of the structure, the C-terminal domain, a four $\alpha$-helical up-down bundle topology is proposed, and in particular the structure of cytochrome $b_{256}$ was selected as template. However, significant differences are expected between this cytochrome structure and the structure of PsbQ. For example, PsbQ has no heme group, so a more compact structure should be expected. Moreover, the sequence alignment between 256b and PsbQ (Fig. 5A) requires the inclusion of a three-residue gap (109-111). Then, the region corresponding to the helix-3 in PsbQ (Fig. 5C) is expected to be continuous and one turn shorter. It is impossible to determine whether PsbQ possesses a $3_{10}$ helix, as the template 256b has between helix-2 and helix-3 (residues PKL). In contrast to 256b, a short $\beta$-strand or a longer loop is suggested for this region of PsbQ [37]. The loops in the model are difficult to predict due to their flexibility, but they are foreseen to be highly charged and solvent exposed and so they could be implicated in the electrostatic binding of PsbQ to PSII.

The 3D model for the C-terminal domain presented here corresponds to PsbQ from spinach, but it would be equally valid for the rest of the PsbQ family. Indeed a similar fold recognition approach performed with Chlamydomonas reinhardtii sequence leads to a similar conclusions (data not shown). The PsbQ family consists of higher plant PsbQ proteins (percentage of identity with respect to spinach >65%) and of green algal PsbQ.
proteins, which are slightly divergent from the former (percentage of identity with respect to spinach <30%) [37]. This higher plant green alga divergence is also observed when other parameters like the theoretical isoelectric point (pI) of the PsbQ proteins are calculated: i.e. the pI is 9.25 for spinach but 5.71 for Chlamydomonas. However, this difference in the pI is less evident when the two domains of PsbQ are considered apart. In this case, the theoretical pI’s are very similar when calculated for each domain: pI (N-t, residues 1-45) = 4.47 and pI (C-t, residues 46-149) = 9.49 for spinach and pI (N-t, residues 1-43) = 4.47 and pI (C-t, residues 44-149) = 8.87 for Chlamydomonas. Moreover, the highest divergence between the higher plant and green algal PsbQ sequences is found in the N-terminal domain (residues 1-45) (Fig. 5B). A bipartite region is inferred in the higher plant PsbQ that consists of a hydrophobic part, enriched in proline and glycine (residues 4-20), followed by a negatively-charged part (residues 21-45). In contrast, the green algal PsbQ sequences have an accumulation of positively and negatively charged amino acids instead of the hydrophobic part. Based on the knowledge that the N-terminal region of PsbQ is essential for its binding to PSII [22], and that the binding properties between the higher plant and algal PsbQ are different, i.e. the former requires PsbP but not the latter [15], the 4-20 amino acids part may be responsible for the different behaviour of the higher plant and green algal PsbQ when binding to the lumenal side of PSII.

The presence of a single tryptophan in the sequence makes possible to study its environment with fluorescence spectroscopy techniques. Based on the classification by Burstein et al [70], the fluorescence emission maximum at 327 nm and FWHM of 53 nm of PsbQ suggest that the tryptophan residue, as a type-I, is located in the protein core, but still in a polar microenvironment. According to the classification by Vivian et al [72], the
tryptophan residue is within class-2, where the edge of the benzene ring is exposed to the solvent. These two classifications, in which the tryptophan residue of PsbQ is proposed to be partially buried are consistent with the results of fluorescence quenching. The fact that I has a smaller $K_{SV}$ constant ($1.2 \text{ M}^{-1}$) than acrylamide ($3.2 \text{ M}^{-1}$) indicates that there is little accessibility for $\Gamma$ to quench the singlet excited tryptophan. In addition, the lack of quenching by $\text{Cs}^+$ and the decrease in the $K_{SV}$ for $\Gamma$ at higher ionic strength suggests not only that the tryptophan residue is hidden from ionic quenchers, but also that its microenvironment is shielded by a positively charged barrier, which cannot be penetrated by cationic quenchers. The tryptophan residue in the model of PsbQ protein is at the beginning of helix-2, pointing toward the core of the protein (Fig. 5C). This position is in full agreement with position $a$ for Trp in amphipathic helices, where they form a lid over the hydrophobic core of the protein [73]. In addition, Trp is surrounded by a positively-charged cluster of residues, mainly located in loop between helix-2 and 3 and that between helix-3 and 4. This microenvironment for the tryptophan residue derived from the 3D model is compatible with the fluorescence data. Regarding the exposure of the tyrosine residues to the solvent, the described changes in the absorption spectrum of PsbQ suggest that 1-2 out of 6 tyrosine residues are buried in the protein. The arrangement of four tyrosines is shown in Fig. 5C, while the other two in the Nt-domain are most likely solvent exposed in the flexible structure of the domain. It is proposed that at least one of them is buried in the core of the protein (Y134) while the others seem to be solvent exposed. Tyr134 and Trp are very close, near one end of bundle, so they could seal the hydrophobic core of PsbQ.
Although PsbQ is proposed to have a role in maintaining an optimal concentration of Cl\(^-\) in PSII, the 3D model for PsbQ adds little information about binding. Based on the difference in the pI of ~4.5-5 units between the N-terminal and C-terminal domains of PsbQ we suggest that, when using extrinsic polypeptide-reconstituted PSII particles, the low requirement of NaCl is due to the ability of the C-terminal region of PsbQ to electrostatically attract Cl\(^-\) and keep it in the neighbourhood of the oxygen-evolving complex. In addition, PsbQ can also play a role in maintaining the overall stability of PSII. PsbQ has been reported to be thermostable, with a melting point of about 65 °C [74]. This is compatible with some of the features found in the PsbQ sequence. PsbQ favors Arg (5.4%), but avoids the thermolabile Cys and His in all its sequence and Pro (except Pro72) in its four \(\alpha\)-helices. The respective frequency of these residues is related to the thermostability of the proteins [75], suggesting that PsbQ could fulfil these requisites. Moreover, salt bridges formed between residues that are relatively close to each other in the sequence are also known to stabilize proteins [76]. Particularly, PsbQ has several Arg and Lys residues sequentially close to Glu and Asp residues that could form salt bridges (i.e. Arg27 and Glu28, Glu36 and Arg37, Glu47 and Arg51, Glu67 and Arg68, Lys102 and Glu106, Asp100 and Lys101, Asp130 and Lys131). This again favours that PsbQ is a thermostable protein. PsbP has also been suggested to be thermostable [77] and, intriguingly, PsbU and PsbV are proposed to play a role in the thermoprotection of PSII proteins [11, 12]. All in all, PsbQ in conjunction with PsbP could play an functional role in keeping Ca\(^{2+}\) and Cl\(^-\) bound to the oxygen-evolving complex, but also a structural role in maintaining the overall (thermo) stability of PSII.
In conclusion, the 3D model for PsbQ suggests that the C-terminal domain has a four-helical bundle folding. The N-terminal domain it is predicted to be flexible without a defined three-dimensional structure. The absorption and fluorescence analyses of PsbQ have revealed the microenvironment of the tryptophan residue and the exposure of tyrosine residues to the solvent. The experimental results support the 3D model proposed for PsbQ as an up-and-down four helical bundle with the single Trp semi-buried at the end of the bundle structure.

We thank the Protein Design Group (Centro Nacional de Biotecnología, Madrid CSIC) for technical assistant and helpful comments.

REFERENCES


32. Wincencjusz, H., Yocum, C.F. & van Gorkom, H.J. (1999). Activating anions that replace Cl\textsuperscript{-} in the O\textsubscript{2}-evolving complex of photosystem II slow the kinetics of the terminal step in water oxidation and destabilize the S\textsubscript{2} and S\textsubscript{3} states. *Biochemistry* **38**, 3719-3725.


**Supplementary material**
The 3D coordinates of the PsbQ model presented in this paper is available at the EJB web site.
Table 1. Templates proposed for the PsbQ protein by different threading methods.

A. Prediction for the complete PsbQ sequence (residues 1-149)

<table>
<thead>
<tr>
<th>Method</th>
<th>PDB</th>
<th>Score</th>
<th>CATH or SCOP</th>
<th>Structural classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>THREADER2</td>
<td>1vltB0</td>
<td>3.79</td>
<td>C 1.20.120.30</td>
<td>Mainly α; up-down bundle; 4 helices</td>
</tr>
<tr>
<td></td>
<td>1ego00</td>
<td>3.02</td>
<td>C 1.20.120.10</td>
<td>Mainly α; up-down bundle; 4 helices</td>
</tr>
<tr>
<td></td>
<td>256bA0</td>
<td>2.71</td>
<td>C 1.20.120.10</td>
<td>Mainly α; up-down bundle; 4 helices</td>
</tr>
<tr>
<td>FFAS</td>
<td>1dkg b</td>
<td>6.09</td>
<td>S</td>
<td>Coiled-coil; parallel</td>
</tr>
<tr>
<td></td>
<td>1setG0</td>
<td>5.17</td>
<td>C 1.10.490.10</td>
<td>Mainly α; orthogonal bundle; globin like</td>
</tr>
<tr>
<td></td>
<td>1dg4A0</td>
<td>4.99</td>
<td>C 2.60.34.10</td>
<td>Mainly β; sandwich; complex</td>
</tr>
<tr>
<td>FUGUE</td>
<td>1jafA0</td>
<td>3.91</td>
<td>C 1.20.120.10</td>
<td>Mainly α; up-down bundle; 4 helices</td>
</tr>
<tr>
<td></td>
<td>1gsa02</td>
<td>3.06</td>
<td>C 3.30.470.20</td>
<td>α β; 2-layer sandwich</td>
</tr>
<tr>
<td></td>
<td>1g59</td>
<td>3.05</td>
<td>S</td>
<td>All α; multihelical 2 all-α domains</td>
</tr>
<tr>
<td>BIOINBGU</td>
<td>1fzp b</td>
<td>9.5</td>
<td>S</td>
<td>All α; up-down bundle; 3 helices</td>
</tr>
<tr>
<td></td>
<td>1b0nA0</td>
<td>8.6</td>
<td>C 1.10.260.10</td>
<td>Mainly α; orthogonal bundle; repressor</td>
</tr>
<tr>
<td></td>
<td>1qsdA0</td>
<td>6.5</td>
<td>C 1.20.1040.50</td>
<td>Mainly α; up-down bundle; spectrin</td>
</tr>
<tr>
<td>3D-PSSM</td>
<td>1d7ma</td>
<td>1.59</td>
<td>S</td>
<td>Coiled-coil; parallel</td>
</tr>
<tr>
<td></td>
<td>1ego00</td>
<td>1.79</td>
<td>C 1.20.120.10</td>
<td>Mainly α; up-down bundle; 4 helices</td>
</tr>
<tr>
<td></td>
<td>1jafA0</td>
<td>2.07</td>
<td>C 1.20.120.10</td>
<td>Mainly α; up-down bundle; 4 helices</td>
</tr>
<tr>
<td>123D+</td>
<td>1wdeB1</td>
<td>4.20</td>
<td>C 1.10.238.10</td>
<td>Mainly α; orthogonal bundle; recoverin</td>
</tr>
<tr>
<td></td>
<td>1ego00</td>
<td>3.87</td>
<td>C 1.20.120.10</td>
<td>Mainly α; up-down bundle; 4 helices</td>
</tr>
<tr>
<td></td>
<td>1zymA2</td>
<td>3.72</td>
<td>C 1.10.274.10</td>
<td>Mainly α; orthogonal bundle; enzyme i</td>
</tr>
</tbody>
</table>
### B. Prediction for the C-terminal domain of PsbQ (residues 46-149)

<table>
<thead>
<tr>
<th>Method</th>
<th>PDB</th>
<th>Score</th>
<th>CATH or SCOP</th>
<th>Structural classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>FFAS</td>
<td>1sctG0</td>
<td>5.62</td>
<td>C 1.10.490.10</td>
<td>Mainly α; orthogonal bundle; globin like</td>
</tr>
<tr>
<td></td>
<td>1gevA0</td>
<td>5.27</td>
<td>C 1.10.490.10</td>
<td>Mainly α; orthogonal bundle; globin like</td>
</tr>
<tr>
<td></td>
<td>1dkg b</td>
<td>5.15</td>
<td>S</td>
<td>Coiled-coil; parallel</td>
</tr>
<tr>
<td>FUGUE</td>
<td>1aep00</td>
<td>5.12</td>
<td>C 1.20.120.10</td>
<td>Mainly α; up-down bundle; 4 helices</td>
</tr>
<tr>
<td></td>
<td>1jafA0</td>
<td>4.30</td>
<td>C 1.20.120.10</td>
<td>Mainly α; up-down bundle; 4 helices</td>
</tr>
<tr>
<td></td>
<td>1gltn04</td>
<td>3.67</td>
<td>C 1.10.8.70</td>
<td>Mainly α; orthogonal bundle; helicase</td>
</tr>
<tr>
<td>BIOINBGU</td>
<td>1qsdA0</td>
<td>9.5</td>
<td>C 1.20.1040.50</td>
<td>Mainly α; orthogonal bundle; spectrin</td>
</tr>
<tr>
<td></td>
<td>1fzp b</td>
<td>9.4</td>
<td>S</td>
<td>All α; up-down bundle; 3 helices</td>
</tr>
<tr>
<td></td>
<td>2crxA1</td>
<td>7.5</td>
<td>C 1.10.443.10</td>
<td>Mainly α; orthogonal bundle; integrase</td>
</tr>
<tr>
<td>3D-PSSM</td>
<td>1d7ma</td>
<td>1.75</td>
<td>S</td>
<td>Coiled-coil; parallel</td>
</tr>
<tr>
<td></td>
<td>256ba0</td>
<td>2.28</td>
<td>C 1.20.120.10</td>
<td>Mainly α; up-down bundle; 4 helices</td>
</tr>
<tr>
<td></td>
<td>1qsdA0</td>
<td>3.1</td>
<td>C 1.20.1040.50</td>
<td>Mainly α; up-down bundle; spectrin</td>
</tr>
<tr>
<td>123D+</td>
<td>256ba0</td>
<td>4.08</td>
<td>C 1.20.120.10</td>
<td>Mainly α; up-down bundle; 4 helices</td>
</tr>
<tr>
<td></td>
<td>1abvo0</td>
<td>3.62</td>
<td>C 1.20.520.20</td>
<td>Mainly α; orthogonal bundle; peroxidase</td>
</tr>
<tr>
<td></td>
<td>1fxc c</td>
<td>3.40</td>
<td>S</td>
<td>All α; up-down long long α-hairpin; 2</td>
</tr>
</tbody>
</table>

* The PDB codes are presented according to the CATH nomenclature, which includes two more cases to specify the subunit and the domain (i.e. 1xxxA2 = PDB file 1xxx, subunit A, domain 2).

* The score thresholds for each method with a certainty >80% are: >3.5 for THREADER2, >8 for FFAS, >5.0 for FUGUE, >10 for BIOINBGU, <1.0 for 3D-PSSM and >5.0 for 123D+. 
FIGURE LEGENDS

Fig. 1. Purification steps of the native PsbQ protein from spinach. The SDS-PAGE shows (a) control PSII-enriched membranes, (b) 10 mM CuCl₂-washed PSII-enriched membranes, (c) supernatant of the 10 mM CuCl₂-washed PSII-enriched membranes, (d) post-Superdex 200 HR 10/30 column.

Fig. 2. (A) Absorption spectra (thick traces) and the second derivative of the absorption spectra (thin traces) of the PsbQ protein under native (solid lines) and denaturing (dashed lines), 6 M Gdn-HCl, conditions. The arrows indicate the peak-peak distances between maxima and minima that required to determine the values for $a$ and $b$ according to [43]. (B) Intrinsic fluorescence emission spectra of PsbQ. When exciting at 295 nm under both native (thick solid line) and denaturing (thick dashed line), 6 M Gdn-HCl, conditions, and when exciting at 280 nm under native conditions (thin solid line). The difference fluorescence emission spectrum between excitations at 280 and 295 nm when normalizing at 400 nm (thin dashed line).

Fig. 3. Stern-Volmer analyses of the quenching of the single tryptophan containing PsbQ protein by Acrylamide (●), Iodide (○, 0.2 M NaCl; ●, 1M NaCl) and Cesium (■). The experimentally determined collisional and static quenching constants, $K_{SV}$ and $V$, are included in the text.
**Fig. 4.** (A) 3D superposition of the templates 1aep, 1ego, 1jaf, 1qsd, 1vls, 256b as indicated in FSSP database (B) Helical wheel diagram for the 4 $\alpha$-helices of PsbQ. The first residue of each wheel is numbered according to the spinach PsbQ sequence (T46, W71, S193 and T131); the hydrophobic residues of the internal faces are filled in grey. The single tryptophan is filled in black and the four tyrosines are surrounded by circular dotted lines.

**Fig. 5.** (A) Sequence alignment of the template (cytochrome $b_{562}$) and the C-terminal domain of PsbQ (residues 46-149). The ruler starts in 46 according to the first residue of the mature PsbQ protein. (B) Sequence alignment of the N-terminal domain of PsbQ from spinach and *Chlamydomonas reinhardtii* where the charged residues are indicated. (C) View of the 3D model for PsbQ, where the non-modelled N-terminal domain of PsbQ (residues 1-45) is shown as a string and the wireframe of the aromatic amino acids W71, Y84, Y87, Y133, Y134 are outlined. The gap between the residues 109 and 111 in helix, numbered according to the PsbQ-256b alignment (A), is labeled, as well as, the helix $3_{10}$ present in the template but not predicted in PsbQ.
Fig. 1

a  b  c  d

97 kDa
66
45
30
20.1
14.4
Fig. 3
Fig. 4
**Fig. 5**

**A**

c562_ecoli ADLEDNMETLNDNLKVIEKADNAAQVKDALTKMRAAALDAQKATPPKLED
PSBQ_SPIOL TEAAQRAKVSASEILNVKQFIDRKAWPSILNDLRASYLKLTVISA

...50,...60,...70,...80,...90,...,

\[\text{c562_ecoli} \quad KS.PDSPEMKDFRHGFDILVGQIDDALKLANEGKVKEAQAAEQLKTTRNAYHCKYR\]
PSBQ_SPIOL \[\quad KPKDEKSLQELT...SKLFSSIDNLHAAIKSPTEA{\text{E}}KYGT{\text{V}}SNINEVLAKLG\]

..100...,..110...,..120...,..130...,..140...,..150..

**B**

\[
\begin{align*}
\text{PSBQ_SPIOL} & \quad 1\quad E{\text{A}}\text{RPTIKPFPPL\ldots CLESE}^\text{N}\text{S}^{\text{C}}\text{K}\text{G}T\text{K}^\text{E}^\text{N}L\text{SP},\text{PP}^\text{45} \\
\text{PSBQ_CMLRE} & \quad 1\quad \ldots\text{LTP}FLFD{\text{E}}^\text{R}^\text{P}^\text{R}^\text{D}^\text{E}\text{I}^\text{N}V{\text{E}}\text{R}^\text{E}^\text{R}^\text{T}^\text{O}\text{R}^\text{A}^\text{S}^\text{L}^\text{43}
\end{align*}
\]

**C**

[Diagram of protein structure with labeled residues and helices]