Volume 260, number 2, 241-244 FEB 08065 January 1990

Amino acid sequence of photosystem I subunit IV from the cyanobacterium *Synechocystis* PCC 6803

Françoise Rousseau and Bernard Lagoutte

Service de Biophysique, Département de Biologie, C.E.N. SACLAY, Gif sur Yvette, Cedex 91 191, France

Received 10 November 1989

We describe here the complete amino acid sequence of photosystem I subunit IV from *Synechocystis* 6803. The molecular mass of 8.0 kDa is lower than in higher plants and *Chlamydomonas*, due to the lack of a characteristic, proline-rich, N-terminal sequence. The remaining sequence exhibits a good conservation, with a hydrophilic and strongly basic N-terminal head followed by two hydrophobic domains. There is no possibility of classical membrane-spanning alpha helices. This component is likely to be one of the most stroma accessible subunits of photosystem I.

Photosystem I, subunit IV; Photosystem I, subunit VII; Amino acid sequence; (*Synechocystis* 6803)

1. INTRODUCTION

Most of the genes encoding the PSI peripheral subunits have now been sequenced in different organisms, but mainly in higher plants and green algae [1-12]. Full amino acid sequences were obtained for subunit VII (*psaC* gene product) in the case of spinach [13] and barley [14], confirming the DNA sequences and the absence of important post-translational modifications. The *psaD* gene sequence has been described more recently in a cyanobacterium, *Synechocystis* 6803 [10]. Apart from the expected absence of the transit peptide, this sequence also lacks 25 amino acids at the N-terminal end when compared with higher plant genes products. The remaining part of the sequence is fairly well conserved. It was thus interesting to get further comparative information between a cyanobacterium and higher plants concerning another nuclear PSI polypeptide. Subunit IV, the product of the *psaE* gene, is also thought to be located on the stroma side of the thylakoids. It has been described in different higher plants [6,9] and *Chlamydomonas* [11], but just partial amino acid sequences were reported for *Synechococcus* 6301 [15] and *Synochococcus vulcanus* [16]. We describe here the full amino acid sequence of this protein isolated from *Synechocystis* 6803 PSI particles

Correspondence address: F. Rousseau, Service de Biophysique, Département de Biologie, C.E.N. SACLAY, Gif sur Yvette, Cedex 91 191 France

Abbreviations: BME, β-mercapto-ethanol; EDC, ethyldimethylcarbodiimide; PMSF, phenylmethylsulfonidefluoride; PTH, phenylthiohydantoin; PSI, photosystem I; SDS, sodium dodecyl sulfate

2. MATERIALS AND METHODS

Synechocystis 6803 was grown on BG 11 medium at 30°C; cells were harvested by centrifugation at the late log phase.

PSI particles were isolated from frozen cell pellets essentially as described [17], except for the following points; protease inhibitors (PMSF, benzamidine and aminocaproic acid) were added in the first buffer at a final concentration of 1 mM, and the last lyophilization was replaced by a second step of solubilization. The PSI pellet recovered after a 60 min centrifugation at 200 000 × g was resuspended in Tris buffer (50 mM, pH 8.0) at a chlorophyl concentration of 1.5 mg/ml. Small subunits were immediately extracted by adding NaSCN at a final concentration of 2 M, in the presence of BME 0.1%. This extraction step was conducted under gentle shaking, either 1 h at room temperature or overnight at 4°C.

Gel permeation chromatography of the extract was run on a Sephacryl S100 column, 90 cm long and 1.6 cm in diameter. Tris buffer 50 mM, pH 8.0, 1 M NaSCN and 1 mM DTT were used for elution at a flow rate of 0.15 ml/min. Absorbance was recorded at 280 nm. Iodoacetic acid was used to block the cysteine residues [19].

Protein fractions were analyzed by SDS slab gel electrophoresis; the resolving gel contained 15% acrylamide and 0.4% bisacrylamide. Gel buffer was 0.5 M Tris, pH 8.8 and running buffers were Tris, Tricine, SDS (0.1 M, 0.1 M, 0.1%), pH 8.25 for the cathode and Tris 0.2 M, pH 8.9 for the anode [18].

Protease from *Staphylococcus aureus* V8 was used at pH 4.5 (ammonium acetate buffer, 75 mM) and 37°C for 20 h [20]. In these conditions, cleavages occur mainly after glutamic residues. Trypsine hydrolysis was performed in ammonium carbonate buffer 100 mM, pH 8.3, at 37°C for 20 h. All peptides were purified by HPLC on a Delta-Pack C18, 300 Å column from Waters. Carboxypeptidase Y degradation was as described [21]. All enzymes were from Sigma.

A pulsed liquid sequencer 477A from Applied Biosystems was used for all sequencing assays; analysis of the PTH amino acids was automatically performed by the 120A on-line HPLC analyzer from the same company. The PTH derivative of carboxymethylcysteine eluted just after PTH glutamine. Protein sequences were mainly analyzed using CITI 2 programs.

3. RESULTS AND DISCUSSION

3.1. *PSI subunits purification*

Chaotropic agents and mild detergents have been us-

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/90/$3.50

ed long term to achieve specific extractions of peripheral subunits from PSI particles (for reviews, see [22,23]. Various chaotropes were recently reported to be particularly efficient in the case of a cyanobacterium [24]. Organic extractions described for spinach [25,26] resulted in poor recovery of subunit VII and we finally selected thiocyanate in a reducing medium (see section 2). This results in the best yield for 3 major subunits, namely subunits II, IV and VII. The recovery of subunit VII, poorly stained by Coomassie blue, was monitored by complexing the cystein residues with monobromobimane [27], a strongly fluorescent reagent. The different polypeptides were separated by gel permeation chromatography in a chaotropic medium. A typical elution profile with the corresponding electrophoretic pattern is shown in fig.1. The first peak to elute from the column is mainly due to pigment-protein complexes of large sizes. Subunit IV elutes later than subunit VII (*psaC* gene products), in contrast with the electrophoretic migration. Collecting peaks at half height provided sufficiently pure fractions for sequencing work; this was particularly clear for the last peak (subunit IV). Subunit II assingment was based on the good agreement between the electrophoretic migration of this component and that found in a previous work [10]. Preliminary N-terminal sequencing determination

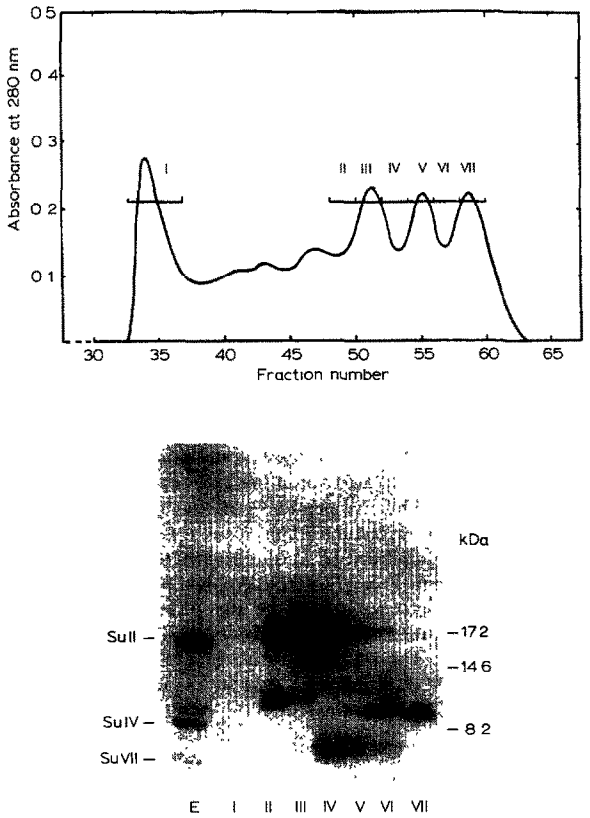

Fig.1. Upper part: elution profile of a NaSCN extract of PSI on a Sephacryl S100 column. Lower part: analysis of the fractions on a 15% polyacrylamide gel electrophoresis. E is for the total extract.

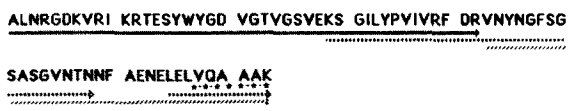

Fig.2. Amino acid sequence of PSI subunit IV. The solid arrow is for direct sequencing from the N terminal; V8 peptides are underlined with dotted arrows and the overlapping tryptic peptide with a slashed one. Asterisks are for the C-terminal degradation by carboxypeptidase Y.

confirmed subunit VII assignment; the following sequence has been found, with the classical cystein arrangement:

SHSVKIYDTC IGCTQCVRAC P[T/E]? DVL[E/A]? MV

Cysteine residues were identified as carboxymethyl derivatives. When considering the upper letters for the two uncertainties, the homology is almost identical to other sequences of plant subunit VII [1-3,7,14,16], except for the last valine instead of isoleucine.

3.2. *Sequencing strategy*

Direct determination of the first 52 amino acid residues was possible using an increased cleavage time at proline 35. Peptides from V8 protease and trypsine digestions were separated by HPLC and selected from their amino acid composition. Two peptides from the V8 experiment reached the C-terminal of the protein. The correct alignment was confirmed by sequencing an overlapping tryptic peptide (fig.2). Sequential degradation from the C-terminal by carboxypeptidase Y confirmed the correct attribution of the last V8 peptide, but precise interpretation in terms of kinetics was very difficult due to the alanine tripeptide. Particularly, the exact ratio of terminal lysine to alanine remains uncertain. On the other hand, the yield of the last amino acid during Edman degradation is often poor, so that it is not possible to exclude a microheterogeneity at the C-terminal, some chains lacking the terminal lysine after a partial degradation or a maturation process.

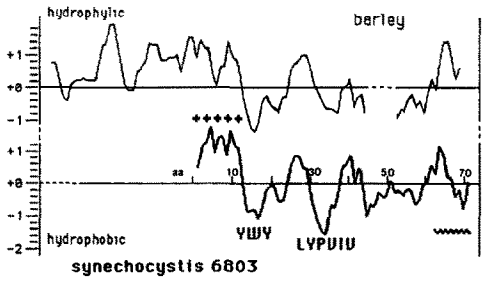

Fig.3. Hydrophilicity plot of subunit IV from barley (normal line) and *Synechocystis* 6803 (bold line), according to Hopp and Woods [28]; the averaging window was on 6 amino acids. The predicted helix is delineated by a wavy line. A gap has been introduced in the barley sequence corresponding to the extra sequence of *Synechocystis*.

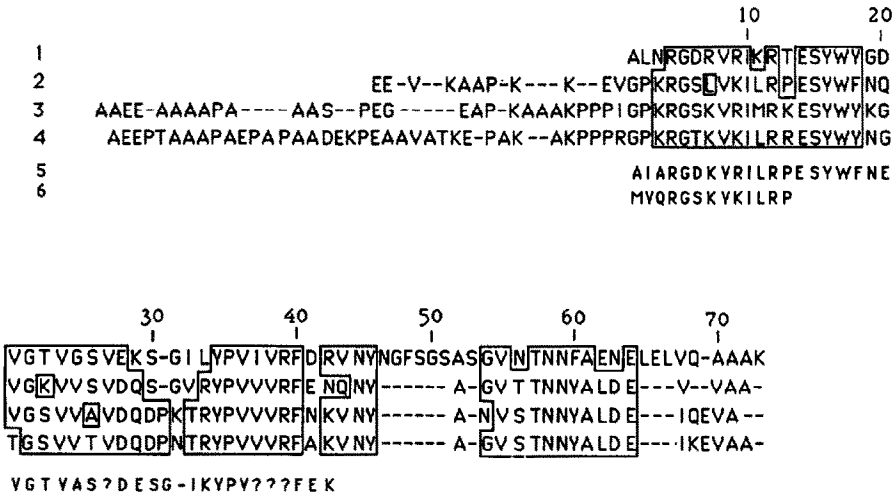

Fig.4. Comparison of full sequences of the PSI subunit IV in different organisms: 1' *Synechocystis*, 6803; 2' *Chlamydomonas reinhardtii* [11]; 3' spinach [8]; 4' barley [6]. Partial sequences are for: 5' *Synechococcus* 6301 [15]; and 6' *Synechococcus vulcanus* [16]. The most conserved sequences are inside the large boxes, just non-conservative substitutions are excluded.

3.3. *Sequence description*

PSI subunit IV from *Synechocystis* 6803 is 73 amino acids long with a net positive charge of +1 at pH 7. The calculated molecular mass is 8000, in good agreement with the electrophoretic migration of this protein (about 9 kDa); this was not the case for all other subunit IVs so far described and will be discussed later. The full sequence fits well with partial sequences previously published for other cyanobacteria [15,16]. The hydrophilicity profile [28] is presented in fig.3, in comparison with the corresponding profile of the barley polypeptide. The N-terminal dodecapeptide is the most hydrophilic part of the protein, with a high density of basic amino acids. This sequence is likely to protrude from the thylakoid membrane, and could serve as a strong accessible electrostatic anchor for acidic polypeptides. Such a basic cluster was also found for subunit II [5]. Two hydrophobic domains lie further in the sequence (fig.3). The first one is mainly due to a typical aromatic tripeptide: tyrosine-tryptophan-tyrosine (amino acids 16–18), the second is a longer stretch of non-polar amino acids (amino acids 33–38). These three well-defined domains are also found in *Chlamydomonas* and higher plants (fig.4). Another well-conserved domain, of intermediate hydrophilicity, is also present close to the C-terminal, between residues 54 and 64. An alpha-helix prediction [29–31] has been found for the C-terminal decapeptide (fig.3), arguing for a possible interaction with membrane components, but not for a membrane spanning localization due to the short length of this putative helix.

Two features make the *Synechocystis* sequence different from others: (i) extra sequences exist near the C-terminal, one of 7 amino acids especially rich in serine residues (amino acids 46–53); (ii) the N-terminal part is shortened by 11 to 37 residues when compared with other organisms. When present, and mainly in higher plants, this N-terminal sequence is characteristic and deserves some detailed comments. It is essentially built of alanine, proline and charged amino acids; noticeable is a proline tripeptide located just before the beginning of the common sequence shared by all organisms. A close correlation is observed between the length of this sequence and the discrepancy between apparent and calculated molecular mass. For higher plants, a mean apparent molecular mass of 16–17 kDa has been reported instead of the calculated values of 9.7 for spinach and 10.8 for barley. In the case of *Chlamydomonas*, which has a shorter N-terminal sequence, an intermediate situation is observed with apparent and true molecular masses of 11 and 8.1 kDa, respectively. The difference becomes almost non-significant for *Synechocystis*, where this special sequence no longer exists. Similar observations can be made concerning PSI subunit II, for which extra N-terminal sequences of the same type are found in higher plants but not in *Synechocystis* [10]. The high proline content of all these sequences is likely to keep them in a non-linear conformation, even in the presence of SDS, thus disturbing the expected electrophoretic migration. The structural constraints given by these N-terminal sequences could play a role in the routing process through the chloroplast envelope, in close association with the transit peptides. In cyanobacteria, all these sequences are of no use for stroma-located PSI subunits, and consequently are no longer present.

Transit peptides deduced from *psaE* cDNAs in higher plants and *Chlamydomonas* are in favor of a stroma localization of subunit IV. No binding with added ferredoxine has been observed so far using the cross-linking reagent EDC [32,33]. To be more convincing, this apparent lack of vicinity has still to be corroborated with other cross-linkers of different length. An exclusive structural role for this subunit is also possible,

the positive hydrophylic head interacting with acidic counterparts like the N-terminal sequences of PSI subunits 1a or 1b. The exact function of this subunit thus remains to be found, but it is likely to be involved close to the ferredoxine binding site.

Acknowledgements: We wish to particularly thank Dr A. Menez and F. Bouet for extended and efficient sequencing facilities. This work was partly supported by a grant from CNRS (URA 1290).

REFERENCES

[1] Ohyama, K., Fukuzawa, H., Takayuki, K., Hiromasa, S., Sano, T., Sano, S., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S.-I., Inokuchi, H. and Ozeki, H. (1986) Nature 322, 572–574.
[2] Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., Zaita, N., Chunwongse, J., Obokata, J., Yamaguchi-Shinozaki, K., Ohto, C., Torazawa, K., Meng, B.Y., Sugita, M., Deno, H., Kamogashira, T., Yamada, K., Kusuda, J., Takaiwa, F., Kato, A., Tohdoh, N., Shimada, H. and Sugiura, M. (1986) EMBO J. 5, 2043–2049.
[3] Dunn, P.P.J. and Gray, J.C. (1988) Plant Mol. Biol. 11, 311–319.
[4] Hoffman, N.E., Pichersky, E., Malik, V.P., Ko, K. and Cashmore, A.R. (1988) Plant Mol. Biol. 10, 435–445.
[5] Lagoutte, B. (1988) FEBS Lett. 232, 275–280.
[6] Okkels, J.S., Jepsen, L.B. Hønberg, L.S., Lehmbeck, J., Scheller, H.V., Brandt, P., Høyer-Hansen, G., Stummann, B., Henningsen, K.W., Von Wettstein, D. and Møller, B.L. (1988) FEBS Lett. 237, 108–112.
[7] Schantz, R. and Bogorad, L. (1988) Plant Mol. Biol. 11, 239–247.
[8] Steppuhn, J., Hermans, J., Nechustai, R., Ljungberg, U., Thümmler, F., Lottspeich, F. and Herrmann, R.G. (1988) FEBS Lett. 237, 218–224.
[9] Münch, S., Ljungberg, U., Steppuhn, J., Schneiderbauer, A., Beyreuther, K. and Herrmann, R.G. (1988) Curr. Genet. 14, 511–518.
[10] Reilly, P., Hulmes, J.D., Pan, Y.-C.E. and Nelson, N. (1988) J. Biol. Chem. 263, 17658–17662.
[11] Franzén, L.-G., Frank, G., Zuber, H. and Rochaix, J.-D., Plant Mol. Biol., in press.
[12] Franzén, L-G., Frank, G., Zuber H. and Rochaix, J-D., Mol. Gen. Genet., in press.
[13] Oh-oka, H., Takahashi, Y., Kuriyama, K., Saeki, K. and Matsubara, H. (1988) J. Biochem. 103, 962–968.
[14] Vibe Scheller, H., Svensen, I. and Lindberg Møller, B. (1989) Carlsberg Res. Commun. 54, 11–15.
[15] Alhadeff, M., Lundell, D.J. and Glazer, A.N. (1988) Arch. Microbiol. 150, 482–488.
[16] Koike, H., Ikeuchi, M., Hiyama, T. and Inoue, Y. (1989) FEBS Lett. 253, 257–263.
[17] Biggins, J. and Mathis, P. (1988) Biochemistry 27, 1494–1500.
[18] Schägger, H. and Von Jagow, G. (1987) Anal. Biochem. 166, 368–379.
[19] Crestfield, A.M., Moore, S. and Stein, W.H. (1963) J. Biol. Chem. 238, 622–627.
[20] Drapeau, G.R. (1977) Methods Enzymol. 47 E, 189–191.
[21] Hayashi, R. (1977) Methods Enzymol. 47 E, 84–96.
[22] Golbeck, J.H. (1988) Biochim. Biophys. Acta 895, 167–204.
[23] Lagoutte, B. and Mathis, P. (1989) Photochem. Photobiol. 49, 833–844.
[24] Parrett, G.K., Mehari, T., Warren, P.G. and Golbeck, J.H. (1989) Biochim. Biophys. Acta 973, 324–332.
[25] Oh-oka, H., Takahashi, Y., Wada, K., Matsubara, H., Ohyama, K. and Ozeki, H. (1987) FEBS Lett. 218, 52–54.
[26] Wynn, R.M. and Malkin, R. (1988) FEBS Lett. 229, 293–297.
[27] Kosower, N.S. and Kosower, E.M. (1987) Methods Enzymol. 143, 76–84.
[28] Hopp, T.P. and Woods, H. (1981) Proc. Natl. Acad. Sci. USA, 78, 3824–2838.
[29] Chou, P.Y. and Fasman, G.D. (1977) J. Mol. Biol. 115, 135–175.
[30] Chou, P.Y. and Fasman, G.D. (1980) Annu. Rev. Biochem. 47, 251–276.
[31] Garnier, J., Osguthorpe, D.J. and Robson, B. (1978) J. Mol. Biol. 120, 97–120.
[32] Zanetti, G. and Merati, G. (1987) Eur. J. Biochem. 169, 143–146.
[33] Zilber, A.L. and Malkin, R. (1988) Plant Physiol. 88, 810–814.