Proceedings of the Indiana Academy of Science
(1993) Volume 102 p. 93-109

STRUCTURAL ANALYSES OF THE 33 KDA PHOTOSYSTEM II MANGANESE-STABILIZING PROTEIN OF *SYNECHOCOCCUS* SP. PCC 7942

Betsy A. Read
Department of Microbiology
The Ohio State University
Columbus, OH 43210
and
Jo Ann Meunier and Carolyn N. Vann
Department of Biology
Ball State University
Muncie, IN 47306

ABSTRACT: The secondary structure of the 33 kDa manganese-stabilizing protein (MSP) associated with the oxygen-evolving complex of the cyanobacterium *Synechococcus* sp. PCC 7942 was predicted using information from a family of homologous amino acid sequences and applying various algorithms to: 1) determine the similarity between substituted residues in a particular position; 2) predict the propensity of various neighboring residues to contribute to a specific secondary structure configuration; 3) analyze hydropathy, chain flexibility, potential antigenic determinants and; 4) identify a region that may serve to bind this polypeptide to the photosystem II core complex. Nine functionally important regions (FIRs) were identified, one of which exhibited a high degree of flexibility as well as partial sequence similarity to the manganese-binding site of the bacterial superoxide dismutase. The nucleotides encoding two aspartic acid residues in this region have been identified as possible ligands for manganese.

INTRODUCTION

Although experimental evidence indicates that the native conformation of a protein is encoded in its amino acid sequence, it is not possible at this time to predict the tertiary structure of a protein or its function based on knowledge of primary structure alone. Most of what is known concerning the details of folding to produce protein structure has been obtained from X-ray diffraction patterns of protein crystals. Unfortunately, crystallographic analysis is a lengthy process that currently lags behind the accumulation of protein sequence data. The amino acid sequence of a protein, however, may serve as a framework for predicting secondary structure and constructing profiles of properties such as hydropathy, segmental flexibility, bulkiness, and accessibility. A protein profile developed from comparisons of sequences of homologous polypeptides from divergent organisms may permit the identification of potentially important domains or residues to aid in the rational design of site-directed mutagenesis experiments which probe the function of a polypeptide or of a specific region of a polypeptide.

In the present investigation, a systematic study of the secondary structure of a photosynthesis polypeptide, the 33 kDa manganese-stabilizing protein (MSP) from the cyanobacterium *Synechococcus* sp. PCC 7942 (also known as *Anacystis nidulans* R2), was undertaken to characterize important features of the polypeptide and to identify probable functionally important regions as targets for mutagenesis. The study used various computer algorithms to analyze features of secondary structure of the polypeptide and to compare three known MSP amino acid sequences for the extent of sequence conservation within specific regions. Finally, a homology search was performed to compare func-

tionally important regions identified within the MSP with amino acid sequences of other manganese-binding polypeptides.

The MSP is one of several polypeptides of the oxygen evolving complex which are intimately associated with the photolysis of water to molecular oxygen. Other components required for water-oxidizing catalytic activity include a polynuclear manganese complex (2/4 Mn/complex), 2-3 Ca^{2+} (Cammarata and Cheniae, 1987; Ikeuchi and Inoue, 1986; Yasuhiro, *et al.,* 1986), Cl^- (Homann, 1987), and a number of intrinsic photosystem II (PSII) polypeptides (43 kDa, 47 kDa, D_2, D_1, cytochrome b_{559}) (Bertold, *et al.,* 1981; Camm, *et al.,* 1987; Cammarata and Cheniae, 1987; Ohno, *et al.,* 1986). It has been suggested that while the presence or absence of the MSP may not be the primary determinant of residual water-splitting activity, it may nonetheless be responsible for maintaining the necessary conformation of the Mn-cluster required for efficient oxygen evolution (Hunziker, *et al.,* 1987; Kuwabara and Murata, 1982; Kuwabara, *et al.,* 1985; Miyao and Murata, 1984).

The use of the secondary structure analyses to design site-directed mutagenesis experiments targeting functionally important regions within the MSP may clarify some of the following: the role of specific amino acids in the polypeptide function; the identification of residues binding the MSP to the photosystem II (PSII) complex; whether the MSP serves as a dismutase to scavenge free radicals produced during the photolysis process; and whether or not Mn is bound to specific amino acids of the MSP for stabilizing the Mn-cluster in the proper conformation necessary for the efficient oxidation of water (Kuwabara, *et al.,* 1985; Oh-oka, *et al.,* 1986).

The cyanobacterium, *Synechococcus* sp. PCC 7942, was used for the study of the MSP in this investigation for several reasons: 1) the amino acid sequence of the MSP has been determined, and the gene (*psbO*) has been cloned in this well-studied cyanobacterium (Kuwabara, *et al.,* 1987); 2) although *Synechococcus* sp. PCC 7942 is a prokaryotic organism, it is capable of performing oxygenic photosynthesis in a manner similar to higher plants and may serve as a general model system of oxygen evolution in plants; 3) the cyanobacterium is transformable to facilitate future mutagenesis experiments; 4) thylakoid preparations from cyanobacteria have a looser structure than preparations from plant chloroplasts (Pistorius, 1983) which may suggest a different means of sequestering Mn required for oxygen evolution; and 5) in future experiments, the oxygen-evolving capacity of specifically constructed mutants could be compared to mutants being produced in other well-studied organisms including the photoautotrophic cyanobacterium, *Synechocystis* sp. PCC 6803, which appears to be able to evolve oxygen at low levels in the absence of the MSP (Burnap and Sherman, 1991).

The analysis of the secondary structure of the MSP was performed by comparing the three amino acid sequences from *Synechococcus* sp. PCC 7942 (Kuwabara, *et al.,* 1985), pea (Watanabe, *et al.,* 1987), and spinach (Herman, *et al.,* 1987; Oh-oka, *et al.,* 1986) and applying two algorithms: one to evaluate the similarity of physiochemical properties of substituted amino acids between the species (Zvelebil, *et al.,* 1987) and a second to calculate the propensity of a particular region to assume a specific secondary structure (Williams, *et al.,* 1987). This information was used to identify nine conserved, functionally important domains (FIRs) for subsequent mutagenesis experiments. Because of their conservation and specific secondary structure features, these FIRs are believed to include residues that are either responsible for maintaining important structural features or are directly involved in catalysis and/or in ligand binding.

Additional information was needed to identify possible ligand binding sites which are expected to be on surface exposed flexible regions of the polypeptide. Crawford, *et al.* (1987) have found good correlation between predicted loop segments, minimal hydropathy, and maximal flexibility. Thus, algorithms were used to prepare a profile of average hydropathy (Kyte and Doolittle, 1982) and chain flexibility values (Karplus and Schultz, 1985). A surface profile of the MSP from *Synechococcus* sp. PCC 7942 was generated following the procedure of Karplus and Schultz (1985) and Parker, *et al.* (1986), which are based on a composite of algorithms assessing hydrophilicity, accessibility, and flexibility parameters. The composite profile was plotted and probable antigenic sites were identified.

In order to identify a potential Mn binding region, the amino acid sequences of our FIRs were compared with sequences known to bind Mn within other proteins. Amino acids in region 249-256 within a predicted FIR were identified as a probable ligand binding region. Partial sequence homology of this region with *Escherichia coli* superoxide dismutase led to the identification of two aspartic acid residues (251 and 253) which might be Mn-binding ligands.

Finally, a region of the polypeptide that might serve to anchor the MSP to the photosystem II core complex was identified using the membrane preference parameters of Argos, *et al.* (1982) with a modified window. The window was shortened since the MSP is not a transmembrane protein but potential membrane anchoring sites might be localized within shorter regions of hydrophobicity.

MATERIALS AND METHODS

Computer analyses. The IBI/Pustell DNA and Protein Sequence Analysis System, obtained from International Biotechnologies Incorporated, was used in the analysis of the nucleic acid and amino acid sequences of the MSP. Quattro, the professional spreadsheet manufactured by Borland International, was employed to perform the necessary mathematical operations dictated by the various algorithms to generate the secondary structure and physiochemical profiles of the MSP.

Secondary Structure Prediction. Homologous sequences of the 33 kDa polypeptide from pea and spinach were aligned with that of the polypeptide from *Synechococcus* sp. PCC 7942 according to Kuwabara, *et al.* (1987). (Since this research was initiated MSP sequences from additional organisms have become available). The degree of sequence conservation at each position along the chain was quantified and assigned a "conservation number," C_i, following the procedures of Zvelebil, *et al.* (1987) in which values of C_i ranging from 0 to 1 were computed based on ten chemical properties of the amino acids at each position. Zvelebil, *et al.* (1987) characterized each amino acid by assigning yes or no values for the following properties: hydrophobic, positive, negative, charged, small, tiny, alipatic, aromatic, proline. The conservation number was calculated from the equation $C_i = 0.9 - 0.1 \times P$, where P represents the total number of variant chemical properties of the amino acids at position i. A high conservation number is yielded when chemically similar types of amino acids occur at a position, and a low number results when there is high variability between substituted residues, or if there is an insertion.

After C_i was calculated for each position, a "smoothed conservation number," CS_i, was figured by averaging C_i over three residues (i - 1, i, i + 1). The CS_i was then plotted along the sequence, and the average conservation number over the entire sequence (C_{av}) was determined.

The single residue method of Williams, *et al.* (1987) was used to predict the secondary structure of the MSP from *Synechococcus* sp. PCC 7942. According to the method of Williams, *et al.* (1987), each amino acid in the MSP from the cyanobacterium was assigned a propensity (P) value for forming an alpha helix, beta sheet, or a turn. The conformational state profiles were then smoothed by averaging over run constants of 6, 5, and 2 for helix, strand, and turn, respectively. Residues were predicted to be in a particular secondary structure if their average preference exceeded the decision constant (DC) for that structure. According to the procedures of Williams, *et al.* (1987), decision constants of 0.99, 0.95, and 1.19 were used for helix, strand, and turn, respectively. In cases of overlap, where the average preference of more than one conformational state exceeded their respective DCs, priority was given to that propensity farthest above its DC. Regions not predicted as either helix, strand, or turn were classified in the random coil conformation having average P_H, P_S, and P_T values less than their respective DCs.

According Zvelebil, *et al.* (1987), an additional 9% improvement in the prediction of secondary structure is afforded if the extent of sequence conservation at each residue is taken into consideration. This was accomplished by calculating $(CS_i - C_{av}) \times A$. The constant A is included to heighten differences in conservation. Optimal values, according to Zvelebil, *et al.* (1987), were found to be $A = 150$ for $C_{av} \leq 0.55$ and $A = 250$ for $C_{av} \geq 0.55$. In the present investigation, the optimal value for A was found to be 0.25. This value was then added to the Williams, *et al.* (1987) conformational state profiles for helix, strand, and turn, prior to assigning residues to a particular secondary structure.

Profiles of average hydropathy and chain flexibility values were also determined. The hydropathy scale of Kyte and Doolittle (1982), which uses a sliding window of seven residues and assigns the value to the fourth residue, was used to generate the hydropathy profile. The profile of chain flexibility was produced using the algorithm of Karplus and Schultz (1985) based on crystallographic temperature factors. The weighted average of a window of seven residues was computed and again was assigned to the fourth residue.

Predicting Functionally Important Residues (FIRs). The conservation plot and the predicted secondary structure were used to identify FIRs of the MSP. Again, the algorithm developed by Zvelebil, *et al.* (1987) was used, whereby an active region was predicted if: 1) within a 5 residue segment, n or more residues were invariant, where $n = 4$, if $C_{av} \geq 0.5$, and $n = 3$, if $C_{av} \leq 0.5$; and 2) in a predicted loop region there was a conserved residue at position i ($C_i = 1.0$) and the $CS_i > 1.5 \times C_{av}$ and $CS_i > 0.7$. The first part of the algorithm attempts to quantify the extent of sequence invariance (n) in relation to the background of overall similarity between aligned sequences, while the second part is founded on the principle that invariant residues in loop regions tend to be more indicative of an FIR than invariant residues within the secondary structure core.

Evaluation of Membrane Preference. In order to obtain information regarding the anchoring of the MSP to the PSII core complex, the membrane preference parameters of Argos, *et al.* (1982) were applied using the function M_5 (window of 5) to localize potential membrane anchoring sites comprised of shorter regions of hydrophobicity. Possible membrane anchoring regions were predicted when $M_5 > 1.0$. Negative M_5 values were regarded as markers for connecting loops or solvent exposed portions of the protein molecule.

Composite Surface Profile for Predicting Antigenicity. A surface profile of the MSP from the cyanobacterium was generated following the procedure of Parker, *et al.* (1986), which is based on a composite of algorithms assessing hydrophilicity, accessibility, and flexibility parameters. Hydrophilicity (Parker, *et al.,* 1986) and accessibility (Janin, 1979) profiles were determined by summing the parameters for each residue in a

seven residue segment and assigning the sum to the fourth residue. The previously described algorithm of Karplus and Schultz (1985) was used to determine the flexibility profile. In each profile, surface sites were defined as any residues with a profile value greater than 25% of the average value for that particular parameter. The three surface profile plots were used to produce a composite profile. The surface sites for each profile were scaled from 0 to 100 such that the maximum surface site value in each plot was set equal to 100 and the 25% surface site value was equal to 0. The plots were then superimposed, and the maximum value of each residue was used to give the composite profile value. Composite profiles were plotted, and antigenic sites were identified as regions with surface profiles greater than a value of 50%.

RESULTS

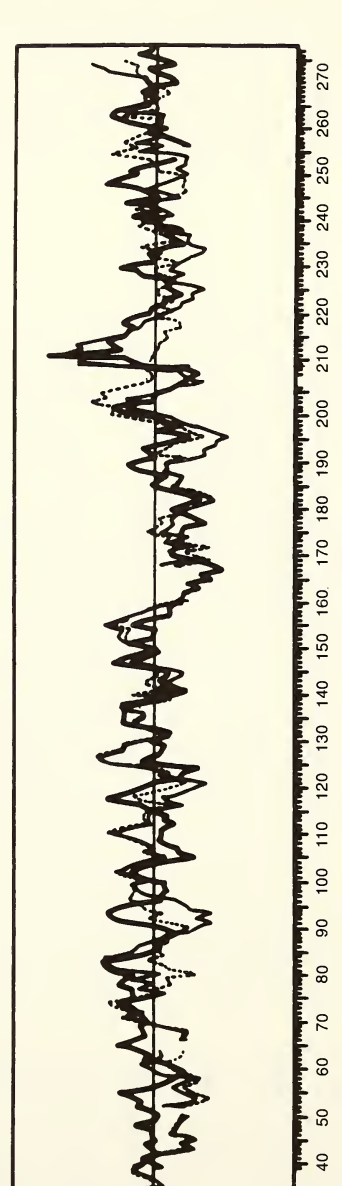
Computer comparisons of the nucleotide sequences of the *psbO* genes and amino acid sequences of three MSPs indicated that there was a high degree of similarity between their secondary structures. The analyses facilitated the construction of a conservation profile of the polypeptide to identify regions of conserved sequences which are believed to be functionally important. Other analyses revealed important regions of probable structure or function including: membrane insertion or contact interfaces, surface exposed flexible regions, antigenic determinants, and residues which might serve as ligands for Mn.

In order to examine the degree of relatedness between MSPs from the cyanobacterium and two higher plant species, the amino acid sequences of the MSPs from pea, spinach, and *Synechococcus* sp. PCC 7942 were aligned as performed previously by Kuwabara, *et al.* (1987). As was shown by these investigators, the MSPs from pea and spinach exhibited significant amino acid sequence homology, with 84% of the 248 residues in the mature protein being conserved between the two species of higher plants. When the MSP sequence from the cyanobacterium was compared to that of either spinach or pea, the homology was significantly lower, with only 48% of the residues being conserved.

When similar chemical and physical properties of the amino acids occurring at each position were taken into account and the extent of sequence conservation quantified according to the method of Zvelebil, *et al.* (1987), the secondary structures of the MSPs appeared to be quite similar. The average conservation number over the entire sequence was found to be 74%, significantly greater than the absolute amino acid conservation. The increase in conservation observed using this method suggested that the majority of amino acid substitutions occurred between residues that share the same physical and chemical properties.

The secondary structures of the three MSPs were predicted using the methods of Williams, *et al.* (1987) and Zvelebil, *et al.* (1987) and are represented in Figure 1, which depicts the propensities of residues to be found in helixes, beta sheets, or turns, respectively. It is evident from the three plots that the MSPs follow the same general trends with respect to their secondary structure propensities, particularly for beta strand and helical propensities. The MSP may be classified according to the method of Levitt and Chothia (1976, 1984) as an alpha/beta protein having mixed and alternating segments of a-helix and B-sheet structures.

The conservation profile generated by comparisons of the three proteins is illustrated in Figure 2 in which the degree of conservation is plotted against the amino acid alignment number. The secondary structure predicted for the MSP from *Synechococcus* sp. PCC 7942 is indicated above the plot. Other comparisons of homologous proteins of known


HELICAL PROPENSITY
PROPENSITY
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
30
40
50
60
70
80
90
100
110
120
130
140
150
160
170
180
190
200
210
220
230
240
250
260
270

TURN PROPENSITY
PROPENSITY
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
30
40
50
60
70
80
90
100
110
120
130
140
150
160
170
180
190
200
210
220
230
240
250
260
270

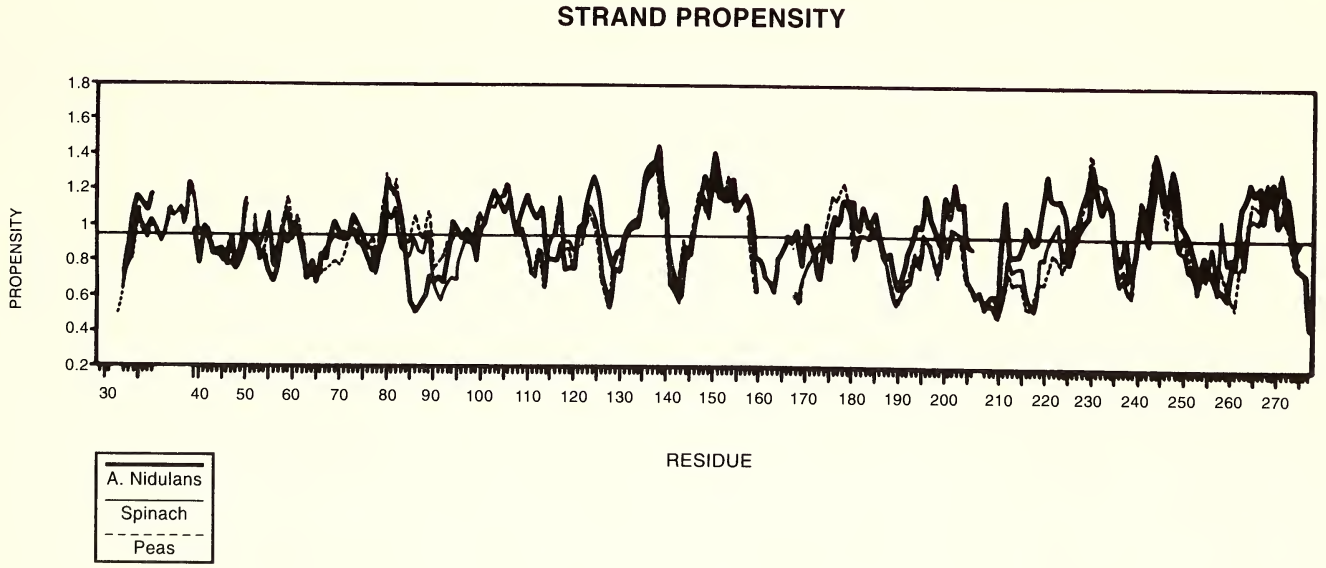


Figure 1. Helical, beta sheet, and turn profiles of the three MSP polypeptides. The propensity profiles of the polypeptide from *Synechococcus* sp. PCC 7942 are indicated by the heavy line, while profiles of the polypeptide from spinach and pea are indicated by the lighter line and the dashed line, respectively. Propensity values are represented along the y-axis and are plotted against the amino acid residues on the x-axis. Numbering on the x-axis corresponds to the amino acid residues in the mature MSP from the cyanobacterium. Insertions that occur in the higher plant species are included and are marked by the blank spaces in the numbering along the x axis. Deletions are indicated by gaps in the profile plots. Decision constants of 0.99, 0.95, and 1.19 were used in the prediction of helical, beta sheets, and turn segments, respectively.

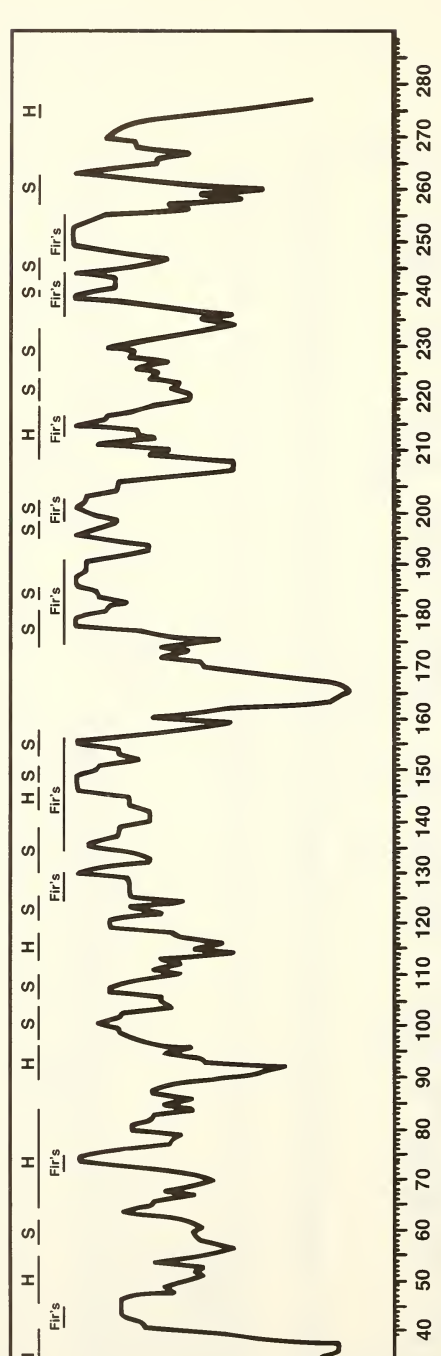


Figure 2. Conservation profile of the MSP polypeptide. The smoothed conservation number (CS_i) is plotted along the MSP sequence from *Synechococcus* sp. PCC 7942. Predicted secondary structure segments are indicated by the horizontal bars labeled S (for strand) and H (for helix). Predicted functionally important residues (FIR's) are also indicated by horizontal bars above the plot.

Table 1. Secondary structure prediction of the MSP sequences from *Synechococcus* sp. PCC 7942, spinach, and pea. The algorithims of Williams, *et al.* (1987) and Zvelebiel, *et al.* (1987) were applied to predict the secondary structures of the three 33 kDa MSP proteins (H = alpha helix, S = beta strand, C = random coil, T = turn). A comparison of the secondary structure composition of the three peptides is also given.

```
                30                  40          50          60          70
                .                   .           .           .           .
Synechoc.       CC  CSSSSSHH        HTTTTTHHHHHHHHHHTTSSSSHTTHHHHHH
Spinach         CC CCCHHHHHHHHSSSSSSTTTTCCCCCCS STTCCTSSHHTTT CCCHH
Pea             CCCCHHHHHHHHHHSSSSSSTTTTCCCCCCC CTTCCTSSSSTTT CCCHH
Conserved       CC          HH          TTT             TSS   TT    HH

                        80          90          100         110         120
                        .           .           .           .           .
                HHHHHHHHHHHHHHTTTCCHHHHHHTTSSSSSSSTSSSSSTTHHHHHHTT
                HHHHHHHHHHHHHHTHHHTTTTTCCCCHHHHHHHSSSSTHHTTHHHHTTTT
                HHHHHHHSSSSTSSSSSCCHHHHHHHHHHHHHSSSSSHHHTTHHHHHHTT
                HHHHHHH                         S SS    TTHHHH  TT

                        130         140         150         160         170
                        .           .           .           .           .
                SSSSSSTTTSSSSSSSSSSTTTHHHHHSSSSTSSSSSSTTTCCCCCCCCC
                SSSSHHHTHHSSSSSSSSSTTTCCCSSSSSSSSSSS TTT       TCC
                SSSSHHHTHHSSSSSSSSSTTHHHSSSSSSSSSS TTT       TTC
                SSSS    T  SSSSSSSSS TT     SSSS SSSS  TT         C

                        180         190         200         210
220
                        .           .           .           .           .
                CTTTSSSSSSSSTSSSTHHHTHHHTSSSTSSSSTTTT  THHHHHHHHHHHS
                CTTSSSSSSSTTTSSSTTTTTTTCTTCCCCHHHHTTTTTTHHHHHHHHHHHH
                CTTSSSSSSSTTTSSSTTTTTTCCTTCTCHHHHHTTTTTTHHHHHHHTHHHH
                CTT SSSSSS  TSSST   T   T          TTT  THHHHHHH HHH

                230         240         250         260         270
                .           .           .           .           .
        SSSSTSSSSSSSSSTTTTTSSTSSSSSHHTTCCCCTTHHHHHHSSSTSSTTSSHHHHH
        SSSTTSSSSSSSSSTTTTTSSSSSSSSSSTTCCCCTCCCCCCCCCTSSSSSSHHHH
        HTTTTSSSSSSSSSTTTTTSSSSSSSSSSTTCCCH HHHHHHHSSSSSSSSSHHHH
            TSSSSSSSSSTTTTTSS SSSSS  TTCCC             SS  S HHH
```

Overall Consensus = 55.6%

COMPARISON OF SECONDARY STRUCTURE COMPOSITION

Protein Character	*Synechoc*	Spinach	Pea
Residue limits	30-278	30-276	30-276
Total Amino Acids	249	248	248
% Alpha Helix	32.5	26.6	31.4
No. Helical Regions	12	12	13
Ave. Helix Length	6.8	5.5	5.9
% Beta Sheet	37.3	30.5	35.1
No. Sheet Regions	19	13	13
Ave. Sheet Length	4.8	5.9	6.7
% Turn	22.4	24.9	22.2
No. Turns	26	22	20
Ave. Turn Length	2.2	2.8	2.8

structure have shown that most insertions and deletions are accommodated in loop or turn regions rather than in well defined secondary structures (Thorton, *et al.,* 1988). This is exemplified by the conservation plot of Figure 2 in which most of the minima can be seen to coincide with loop regions between helical and strand regions (indicated above the plot). However, the first prominent minimum in the plot occurring after residue 38 in the cyanobacterial sequence represents an insertion of eight residues that occurs in both polypeptides from higher plant species of pea and spinach. The second prominent minimum that occurs between residues 160-168 represents a deletion in the two higher plant sequences.

Using the algorithm of Zvelebil, *et al.* (1987), nine functionally important regions (FIRs) were identified. These are represented above the conservation plot in Figure 2. Seven of these nine FIRs appear to be located partially, if not entirely, within loop or turn segments. This is not surprising as in most globular proteins where loops form more than 30% of the structure, loop regions tend to play a key role in recognition between molecules and may often incorporate essential catalytic residues as well (Thorton, *et al.,* 1988). These nine FIRs which are located between residues 41-45, 71-76, 125-130, 135-156, 176-191, 199-203, 213-217, 236-244, and 247-255 are prime targets for site-directed mutagenesis experiments which may provide insight into the structure and function of this important protein.

The information contained in these figures is summarized in Table 1. As expected for evolutionarily related proteins, the secondary structure homology is greater than that of the primary structure. While the absolute overall consensus in secondary structure was found to be only 55% (12% greater than the absolute amino acid homology), it is important to note that the true borders between alpha helixes, beta strands, and loops may vary somewhat from one sequence to another, since the predictions are based on average properties. The predicted number of helical and beta sheet segments indicate that in this respect considerable homology exists. The number of helical regions are 12, 12, and 13, and the number of beta sheet regions are 19, 13, and 13 for the cyanobacterium, spinach, and pea, respectively. Particularly striking in all 3 proteins is the large number of predicted beta turns. The number and position of the turns are conserved for the most part among the 3 different species. Like many of the other Mn- and Fe-binding superoxide dismutases, the MSP is expected to be highly folded (Marres, *et al.,* 1985).

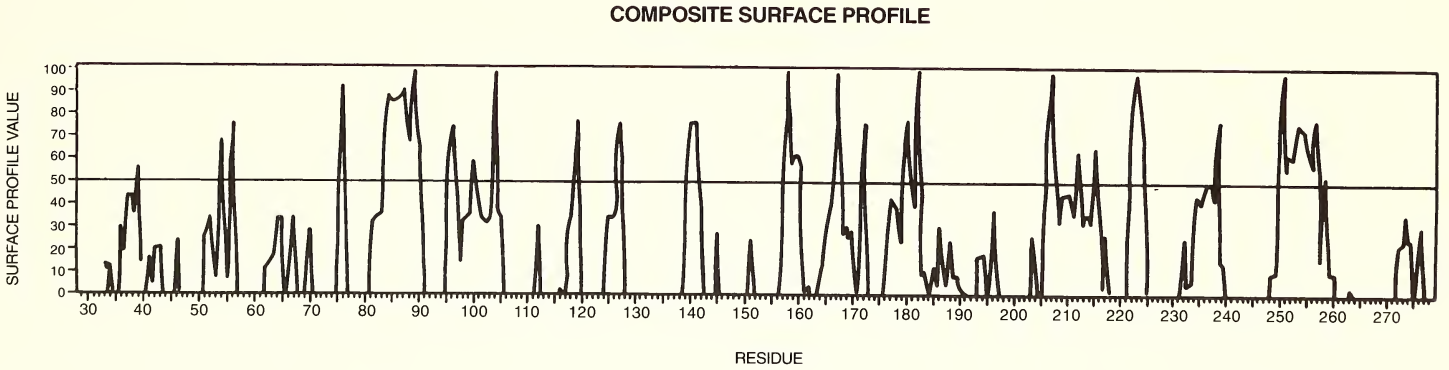


Figure 3. Composite surface profile of the MSP from *Synechococcus* sp. PCC 7942. A combination of algorithms assessing hydrophillicity, accessibility, and mobility parameters were used to produce the composite surface profile. The solid line represents a 50% decision constant for probable antigenic sites.

An analysis of the secondary structure of the cyanobacterial MSP was performed to determine the conformational preferences for sheet, helix, and turn. The polypeptide was found to be 32% alpha helical and 37% beta sheet. Twelve alpha helixes, 19 beta sheets, and 26 turns were predicted in the cyanobacterial protein. Considerable helical potential was found to exist in the amino terminal region, while several potential alpha/beta/alpha units exist in the central portion of the protein.

Analyses of the average values of both hydropathy and chain flexibility for the cyanobacterial MSP were also performed. Using the profiles of hydropathy and chain flexibility as complementary, reciprocal indicators of polypeptide segment position with respect to the hydrophobic interior of the protein works well. Caution should be exercised when predicting secondary structural elements in this manner, however. Minimal hydropathy and maximal flexibility values correlated with most (17 out of 26) of the predicted loop segments, while maximum hydropathy and minimum chain flexibility correlated with only some (7 out of 19) of the predicted beta strands. According to Crawford, *et al.* (1987), the amplitudes of both hydropathy and chain flexibility are comparatively small and variable for predicted alpha helixes. Unfortunately, alpha helixes in the MSP could not be predicted with any confidence following these criteria.

Since the MSP is not an intrinsic membrane protein, a membrane-buried helix propensity profile was constructed using the parameters of Argos, *et al.* (1982) with a sliding window of only five residues (H_5). This function was introduced for two reasons: 1) to identify possible membrane anchoring segments, i.e., short hydrophobic alpha helixes that may serve to bind the MSP to the PS II core complex; and/or 2) to localize the position of hydrophilic solvent-exposed connecting loops that may serve as possible protein-protein contact interfaces. As expected, the vast majority of the cyanobacterial MSP exhibited a low preference for membrane insertion, and apart from the signal sequence, which is represented by the first 28 amino acids, the MSP does not appear to have any membrane-buried helical segments.

Residues 135-160, which exhibited some probability for membrane insertion, may represent a contact interface with another molecule. The stearic hindrance of the two prolines in this region prohibit the segment from adopting an alpha helical structure, and hence, it was not predicted to be membrane-immersed. Secondary structure analysis indicated that this sequence is likely to form a beta/alpha/beta structure. The presence of nonpolar residues and the absence of many charged residues suggested that this region, which is one of the longest of nine identified FIRs, might be membrane immersed. The few charged residues within this region, two glutamic acid residues and one lysine, may be important in terms of charge pairing, perhaps with another protein associated with the PSII core complex.

Contact interfaces between molecules are often as closely packed as the interior of a protein molecule. Contacts constitute a precise fit between hydrophobic side chains, charge pairings, side chains, and backbone hydrogen bonds (Argos and MohanaRao, 1982; von Heijne, 1987). Since this region is relatively long in terms of its conservation, exhibits minimal to moderate flexibility, and is the most likely region to be membrane-immersed, it may indeed play an important role in anchoring the MSP within the thylakoid membrane or to another protein. On the other hand, the association of the MSP with the PSII reaction center may be primarily ionic, since treatment with high salt will remove the MSP.

The most prominent negative depression in the membrane-buried helix profile was found between residues 82-92. This region is likely to be a connecting loop on the surface of the protein and is one of the most highly charged regions of the polypeptide. The

concentration of charge and the hydrogen bonding potential of the residues in this region could also be important in forming a contact interface with another protein molecule. This region, however, is not a FIR.

Figure 3 is a composite surface profile of the *Synechococcus* PCC 7942 MSP. Profile plots of accessibility, HPLC-hydrophilicity, and flexibility parameters were combined to produce the composite surface profile. The surface site values (a surface site is defined as any residues with a profile greater that 25% above the average) for each profile were scaled from 0 to 100 with the maximum surface site value in each plot set to 100 and the 25% surface site value set to 0. The profiles were then superimposed, and the maximum value of each residue was taken as the composite profile value. Although only a small number of antigenic sites are composed of residues that are continuous in the primary sequence, according to Parker, *et al.* (1986), these sites will have composite surface values greater than 50%. The two most prominent antigenic sites produced by contiguous amino acids within the MSP were formed by residues 82-90 and 249-256. Antibodies might be raised to synthetic peptides mimicking these portions of the polypeptide.

While several other regions exhibited surface profile values that exceeded 50%, these regions are less likely to be antigenic because they encompass few residues. It is worth noting that the highly antigenic region, which encompasses residues 249-256, was a predicted FIR. Since this region also exhibited partial similarity to the Mn-binding site of superoxide dismutase from *E. coli,* it was identified for future saturation mutagenesis.

If the region from 249-256 is indeed a ligand binding domain, it must be accessible, and a certain degree of segmental flexibility would be expected to allow the protein to undergo conformational changes upon binding ligands or substrates. Flexibility can be advantageous in aiding the access of ligands to the active site and may also affect the rate and equilibrium constant for the ligand binding. If the MSP binds Mn which cycles between various oxidation states, flexibility of the protein structure would be necessary for electron transfer as the protein is reversibly oxidized and reduced. It would be essential that the structure be able to relax in order to accommodate changes in geometry around the Mn atom as it changes its oxidation state.

It is important to note that there are four Mn associated with the PSII complex, and twenty to twenty-four ligands are expected in the coordination sphere of the Mn-cluster. How many ligands, if any, are provided by the MSP to bind Mn or to hold the oxygen evolving complex in the proper conformation is not known (Philbrick, *et al.,* 1991). At this time, only one residue, the Glu-60 of the reaction center D2 polypeptide, is known to provide a ligand to the Mn-cluster (Vermaas, 1990), but work by Diner, *et al.* (1988) suggests that processing of the carboxy terminus of the D1 polypeptide is also required for proper Mn binding and assembly of the oxygen-evolving complex.

Other potential Mn-binding domains on the MSP include the site that corresponds to that previously identified by Oh-oka, *et al.* (1986) in pea, which also exhibited partial sequence homology with the bacterial superoxide dismutase, and the FIR between amino acids 176 and 192. The MSP sequence from *Chlamydomonas reinhardtii* (Philbrick and Zilinskas, 1988) was compared to that of spinach, and a stretch of 33 identical residues in this region (139 to 171) was observed which was 70% conserved in the two known cyanobacterial sequences (*Synechocystis* sp. PCC 6803 and *Synechococcus* sp. PCC 7942). The region identified by Oh-oka, *et al.* (1986) in *Synechococcus* sp. PCC 7942 encompasses the first FIR and a few neighboring residues in our analyses. It was predicted to be in a loop or turn in all three of the MSPs but did not appear to be particularly accessible or flexible. The FIR at 175-191 was predicted to be in a beta sheet.

Until recently, the MSP was believed to be required for the lysis of water and oxygen evolution. Mayfield, *et al.* (1987) found that expression of the MSP was required for stability of PSII particles by examining *Chlamydomonas reinhardtii* mutants. In this organism, only about 25% of the core PSII proteins accumulated in mutants deficient for expression of the polypeptide. However, *psbO* deletion and insertion mutants were constructed in *Synechocystis* sp. PCC 6803, which resulted in the lack of expression of the MSP (Mayes, *et al.,* 1991; Philbrick, *et al.,* 1991; Burnap and Sherman, 1991). The mutants exhibited normal expression and assembly of the polypeptides of PSII, but oxygen evolution was 70% of that seen in wild-type organisms, and mutants were more susceptible to photoinhibition. Thus, the results of these recent experiments suggest that the MSP may play only a minor role, if any, in providing Mn ligands.

Our analyses also suggested that the MSP is a highly folded protein which has a beta/alpha/beta super-secondary structure. These recurrent motifs are energetically favorable configurations, and the beta/alpha/beta structure is common to many globular proteins.

DISCUSSION

Clearly, there is sufficient information embedded in an amino acid sequence to determine the final folding of the polypeptide. However, predictions of the tertiary structure of a polypeptide from the primary sequence are currently not accurate because of two important considerations (King, 1986): 1) little is known about the rules regulating the assembly of sequential intermediate forms of the polypeptide that are short-lived and unstable but which are necessary to achieving the final tertiary structure; and 2) almost all algorithms which attempt to relate structure to sequence implicitly assume that each residue carries information for the folding process, but experimental evidence has shown that many residues are not involved.

The identification of specific residues as targets for mutagenesis, however, did not require knowledge of the tertiary structure of the polypeptide. Recognition of important structural and functional residues was performed not only by the application of algorithms to predict a structural profile of secondary structure but also by the comparison of homologous polypeptides to identify conserved residues which play important structural and/or functional roles.

Having built a structural model by analysis of secondary features and by examination of the homology of one sequence to another, it is necessary to consider the validity of this approach. Although secondary structure analyses are still unable to delineate the full three-dimensional structure of a protein molecule from its primary amino acid sequence alone, regions of ordered secondary structure may be delineated with reasonable accuracy. Predictions of tertiary structure will become more feasible as x-ray crystallography data accumulates and methods of analysis of structure and predictions of polypeptide folding continue to rapidly evolve.

Statistical propensities for "alpha helix forming" and "beta sheet forming" abilities based on known X-ray crystallography data have been extended to arrive at empirical prediction schemes such as the well-known Chou-Fasman method of predicting secondary structure (Chou and Fasman, 1974a, 1974b, 1978a, 1978b). While the method of Chou-Fasman is only 49% accurate, a modified version of this analysis developed by Williams, *et al.* (1987) was used in this investigation. This method eliminates many of the Chou-Fasman rules, chooses the best decision constants for structure assignments, and has been shown to be 57% correct in assigning helical, strand, and turn conformational states.

The inaccuracy of prediction methods is often associated with delineating the termini of sheets and helixes (Argos, *et al.,* 1982). It is difficult to define the exact residue span encompassing a certain ordered structure, and often the transition from coil to order is made in a continuous fashion requiring several residues. Furthermore, the rules governing predictive methods can be ambiguous and lead to overlapping helix and sheet determinations. Thus, it is difficult for secondary structure prediction methods which only consider local information to accurately delineate structure. Using these kinds of predictions on known structures, the number and relative positions of helixes and strands are correctly predicted but are often slightly displaced or of the wrong length (Taylor, 1987).

Chothia and Lesk (1986) have analyzed the accuracy of secondary structure predictions by examining pairs of homologous sequences of known structures. They observed that above 50% residue homology the differences between predicted and actual secondary structure are slight and that a model built structure can be almost as good as a moderate resolution crystal structure. The degree of homology between the polypeptides in this investigation is very close to 50%, and the secondary structure predictions of the 33 kDa MSP of *Synechococcus* PCC 7942 were enhanced by comparisons of the polypeptide to the two known MSP sequences in higher plants.

The secondary structure analyses and identification of potential ligands for manganese binding will permit construction of site-directed cyanobacterial mutants which can be examined for altered function to determine if the analyses correctly identified a functionally or structurally important region within the MSP.

ACKNOWLEDGMENTS

This work was supported by grants from the Research Corporation, Indiana Academy of Science, Ball State Internal Granting Program, and Sigma Xi Research Society.

LITERATURE CITED

Argos, P. and J.K. MohanaRao. 1982. Prediction of protein structure. *In:* J.H. Hash (Ed.), *Methods in Enzymology,* Academic Press, New York, 130: 185-207.

_____, J.K. Rao, and P.A. Hargrave. 1982. Structural prediction of membrane-bound proteins. Eur. J. Biochem. 128: 565-575.

Bertold, D.A., G.T. Babcock, and C.F. Yocum. 1981. A highly resolved, oxygen-evolving photosystem II preparation from spinach thylakoid membranes: EPR and electron transport studies. FEBS Lett. 134: 321-324.

Burnap, R.L. and L.A. Sherman. 1991. Deletion mutagenesis in *Synechocystis* sp. PCC 6803 indicates that the Mn-stabilizing protein of photosystem II is not essential for O_2 evolution. Biochem. 30: 440-446.

Camm, E.L., B.R. Green, D.R. Allred, and L.A. Staehelin. 1987. Association of the 33 kDa polypeptide (water-splitting) with PS II particles: Immunochemical quantification of residual polypeptide after membrane extraction. Photosyn. Res. 13(1): 69-80.

Cammarata, K.V. and G.M. Cheniae. 1987. Studies on 17,24 kD depleted photosystem II membranes. I. Evidence for high and low affinity calcium sites in 17,24 kD depleted membranes from wheat versus spinach. Plant Physiol. 84: 587-595.

Chothia, C. 1984. Principles that determine the structure of proteins. Annu. Rev. Biochem. 53: 537-572.

_____ and A.M. Lesk. 1986. The relation between the divergence of sequence and structure in proteins. EMBO J. 5(4): 823-826.

Chou, P.Y. and G.D. Fassman. 1974a. Conformation parameters for amino acids in helical, alpha-sheet, and random coil regions calculated from proteins. Biochem. 13: 211-222.

_____ and _____. 1974b. Prediction of protein conformation. Biochem. 13: 222-245.

_____ and _____. 1978a. Empirical prediction of protein conformations. Annu. Rev. Biochem. 47: 251-276.

_____ and _____. 1978b. Prediction of the secondary structure of proteins from their amino acid sequence. Adv. Enzymol. Relat. Areas. Mol. Biol. 47: 45-148.

Crawford, I.P., T. Niermann, and K. Kirschner. 1987. Prediction of secondary structure by evolutionary comparision: Application to the subunit of tryptophan synthase. Proteins Struc. Funct. Genet. 2(2): 118-129.

Diner, B.A., J.G. Metz, D.F. Ries, and B.N. Cohen. 1988. Carboxy-terminal processing of polypeptide D1 of the photosynthesis II reaction center of *Scenedesmus obliquus* is necessary for assembly of the oxygen-evolving complex. J. Biol. Chem. 253: 8972-8980.

Hermann, R.G., A. Tyagi, J. Hermans, C. Jansson, and F. Vater. 1987. Nucleotide sequence of cDNA clones encoding the complete "33 kDa" precursor protein associated with the photosynthetic oxygen-evolving complex from spinach. Mol. Gen. Genet. 207: 288-293.

Homann, P.H. 1987. The relations between the chloride, calcium and polypeptide requirements of photosynthetic water oxidation. J. Bioenerg. Biomemb. 19: 105-123.

Hunziker, D., D.A. Abramowicz, R. Damoder, and G.C. Dismukes. 1987. Evidence for an association between a 33 kDa extrinsic membrane protein, manganese and photosynthetic oxygen evolution. I. Correlation with the S_2 multiline EPR signal. Biochim. Biophys. Acta 890: 6-14.

Ikeuchi, M. and Y. Inoue. 1986. Characterization of O_2 evolution by a photosystem II reaction center complex isolated by a simplified method: Disjunction of secondary acceptor quinone and enhanced Ca^{2+} demand. Arch. Biochem. Biophys. 247: 97-107.

Janin, J. 1979. Surface and inside volumes in globular proteins. Nature 277: 491-492.

Karplus, P.A. and G.E. Schultz. 1985. Prediction of chain flexibility in proteins: A tool for the selection of peptide antigens. Naturwissenschaften 72: 212-213.

King, J. 1986. Genetic-analysis of protein folding pathways. Biotechnol. 4: 191-222.

Kuwabara, T. and N. Murata. 1982. Separation and characterization of thylakoid and cell envelopes of the blue-green alga (Cyanobacteria) *Anacystis nidulans.* Plant Cell Physiol. 23: 533-539.

_____, M. Miyao, and N. Murata. 1985. The function of the 33-kilodalton protein in photosynthetic oxygen evolution system studied by reconstitution experiments. Biochim. Biophys. Acta 806: 283-289.

_____, K.J. Reddy, and L.A. Sherman. 1987. Nucleotide sequence of the gene from the cyanobacterium *Anacystis nidulans* R2 encoding the Mn-stabilizing protein involved in photosystem II water oxidation. Proc. Nat. Acad. Sci. USA 84: 8230-8234.

Kyte, J. and R.F. Doolittle. 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157: 105-132.

Levitt, M. and C. Chothia. 1976. Structural patterns in globular proteins. Nature 261: 552-558.

Marres, C.A., A.P. Van Loon, P. Oudshoorn, H. Van Steeg, and L.A. Grivell. 1985. Nucleotide sequence analysis of the nuclear gene coding for manganese superoxide dismutase of yeast mitochondria, a gene previously assumed to code for the Rieske iron-sulphur protein. Eur. J. Biochem. 147: 153-161.

Mayes, S.R., K.M. Cook, S.J. Self, A. Zhang, and J. Barber. 1991. Deletion of the gene encoding the photosystem II 33 kDa protein from *Synechocystis* sp. PCC 6803 does not inactivate water-splitting but increases vulnerability to photoinhibition. Biochim. Biophys. Acta 1060: 1-12.

Mayfield, S.P., P. Bennoun, and J.D. Rochaix. 1987. Expression of the nuclear encoded OEE1 protein is required for oxygen evolution and stability of photosystem II particles in *Chlamydomonas reinhardtii.* EMBO J. 6: 313-318.

Miyao, M. and N. Murata. 1984. Role of the 33-kDa polypeptide in preserving Mn in photosynthetic oxygen-evolution systems and its replacement by chloride ions. FEBS Lett. 170: 350-354.

Ohno, T., K. Satoh, and S. Katoh. 1986. Chemical composition of purified oxygen-evolving complexes from the thermophilic cyanobacterium *Synechococcus* sp. Biochim. Biophys. Acta 852: 1-8.

Oh-oka, H., S. Tanaka, K. Wada, T. Kuwabara, and N. Murata. 1986. Complete amino acid sequence of the 33 kDa protein isolated from spinach photosystem II particles. FEBS Lett. 197: 63-66.

Parker, J.M., D. Guo, and R.S. Hodges. 1986. New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: Correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. Biochem. 25: 5425-5432.

Philbrick, J.B., B.A. Diner, and B.A. Zilinskas. 1991. Construction and characterization of cyanobacterial mutants lacking the manganese-stabilizing polypeptide of photosystem II. J. Biol. Chem. 266: 13370-13376.

_____ and B.A. Zilinskas. 1988. Cloning, nucleotide sequence and mutational analysis of the gene encoding the photosystem II manganese-stabilizing polypeptide of *Synechocystis* 6803. Mol. Gen. Genet. 212: 418-425.

Pistorius, E.K. 1983. Effects of Mn^{2+}, Ca^{2+}, and chlorpromazine on photosystem II of *Anacystis nidulans:* An attempt to establish a functional relationship of amino acid oxidase to photosystem II. Eur. J. Biochem. 135: 217-222.

Taylor, W.R. 1987. Protein structure prediction. *In:* M.J. Bishop and C.J. Rawlings (Eds.), *Nucleic Acid and Protein Sequence Analysis: A Practical Approach,* pp. 285-321, IRL Press, Washington D.C., 397 pp.

Thorton, J.M., B.L. Sibanda, M.S. Edwards, and D.J. Barlow. 1988. Analysis, design and modification of loop regions in proteins. BioEssays 8(2): 63-69.

Vermaas, W.F.J., J. Charite, and G.Z. Schen. 1990. Glu-69 of the D2 protein in photosystem II is a potential ligand to Mn involved in photosynthetic oxygen evolution. Biochemistry 29: 5325-5332.

von Heijne, G. 1987. Sequence analysis in molecular biology: Treasure trove of trivial pursuit. Academic Press, Inc, New York, 184 pp.

Watanabe, A., E. Minami, M. Murase, K. Shindohara, T. Kuwabara, and N. Murata. 1987. Biogenesis of photosystem II complex in spinach chloroplasts. Prog. Photosyn. Res. 4: 629-636.

Weiss, W. and G. Renger. 1986. Studies on the nature of the water-oxidizing enzyme. II. On the functional connection between the reaction-center complex and the water-oxidizing enzyme system Y in photosystem II. Biochim. Biophys. Acta 850: 173-183.

Williams, R.W., A. Chang, D. Juretic, and S. Loughran. 1987. Secondary structure predictions and medium range interactions. Biochim. Biophys. Acta 916: 200-204.

Yasuhiro, K., K. Satoh, and S. Katoh. 1986. A simple procedure to determine Ca^{2+} in oxygen-evolving preparations from *Synechococcus sp.* FEBS Lett. 205: 150-154.

Zvelebil, M.J., G.J. Barton, W.R. Taylor, and M.J. Sternberg. 1987. Prediction of protein secondary structure and active sites using the alignment of homologous sequences. J. Mol. Biol. 195: 957-961.