Indian Journal of Biochemistry & Biophysics
Vol. 41, October 2004, pp. 258-262

Comparative modelling of tetramanganese cluster of *Chaetosphaeridium globosum*

L S Rath[1] and M K Raval[2*]

[1]Bioinformatics Research Center, International Institute of Information Technology, Gachibowli, Hyderabad-500 019, India
[2]Post Graduate Department of Chemistry, Rajendra College, Bolangir-767 002, Orissa, India

Received 17 March 2004; revised 19 July 2004

The D1 protein of photosystem II (PS II) complex of a micro-alga *Chaetosphaeridium globosum* has been theoretically modelled from its sequence using comparative modeling with known backbone structure of D1 protein from bacterium *Thermosynechococcus vulcanus* as template. The model is built with missing loops and all side chains, which are not resolved in the structure of the template. The structure of the tetramanganese cluster (TMC) and the ligand forming side chains have been subjected to modeling studies in order to gather more information useful to understanding of the water splitting reactions. Earlier models of TMC have been scrutinized and an insight into the manganese coordination sphere has been provided.

Keywords: Photosystem II, D1 protein, water oxidizing complex, tetra manganese complex, comparative modeling, *Chaetosphaeridium globosum, Thermosynechococcus vulcanus*

IPC Code: C 07 K 14/405

The study of photosystem II (PS II) as a model system for a potent solar fuel cell is of great significance[1-3]. The heart of the water splitting apparatus of plants is the photochemically active reaction center (RC II) of PS II, which contains two major subunits D1 and D2 proteins[4-6]. The photosynthetic oxidation of water to molecular oxygen takes place at the tetramanganese cluster (TMC), which is also called water-oxidizing complex (WOC)[3]. Its chemical formula has been suggested to be $Mn_4O_xCa_1Cl_{1-2}(HCO_3)$ and all these native coordinates, except bicarbonate are required for photo-oxidation of water[7]. It is coordinated to the D1 protein only[5,8]. The mature D1 subunit consists of N-acetyl-Thr2-Ala344 and it is highly conserved among bacteria, algae and plant species. Its molecular mass is 38 kDa and it contains 5-transmembrane α-helices with its N-terminus exposed to the stromal surface of the thylakoid membrane[8]. All the four Mn atoms in TMC are located on the lumenal surface[8,9].

Several structural models have been derived from spectroscopic and X-ray diffraction signatures of TMC[1,10-12]. Earlier, it was reported that D1-Y161 can act as a strong proton-donor and its re-reduction occurs at the same time as the release of molecular oxygen from the TMC, during the last stage of transition ($S_3 \rightarrow S_4 \rightarrow S_0$)[13], which was consistent with the H-abstraction model in which Y161 was placed close to, but not binding to TMC[14]. A number of other D1 and D2 residues have also been proposed as potential ligands for TMC[2,15-23].

The information on distances between Mn atoms of TMC is obtained from extended X-ray absorption fine structure (EXAFS). The EXAFS data of S_1 and S_2 states of TMC show three peaks in the Fourier transformation of the k-space data at 1.8, 2.7 and 3.3 Å[24]. The 1.8 Å peak has been assigned to backscattering from the first coordination sphere of Mn, which is proposed to consist mainly of O-donor ligands. The 2.7 Å peak arises exclusively from backscattering from Mn, implying that TMC contains Mn-Mn separation at this distance, while the 3.3 Å peak has been assigned to backscattering of Mn and Ca. The EXAFS determined structure of S_3-state is distinctly different from S_1 and S_2 in that all Mn-Mn vectors lengthen[24,25]. In this state, the 2.7 Å vector assigned to Mn-Mn scattering lengthens and splits into two vectors at 2.8 Å and 3.0 Å. The 3.3 Å vector also increases and could be resolved as two vectors at 3.4 Å, assigned to Mn-Mn scattering and 3.6 Å assigned to Mn-Ca scattering[1].

Several structural models of TMC have been proposed based on EPR and EXAFS data[26-29]. The recent X-ray crystal structure of PS II from *Thermoynechococcus vulcanus* at 3.7 Å resolution provides an updated structure of TMC[30]. However, it fails to resolve the structure of ligands to the TMC. Positions of Ca^{2+}, Cl^-, bridged oxygen, and substrate water molecules are also not resolved. Knowledge of these structural features is essential for understanding the function of WOC. The present work attempts to generate these missing information by theoretical

*Author for correspondence
Present address: PG Department of Chemistry, Government College, Sundargarh-770 002, Orissa, India
E-mail: mkroval@sancharnet.in

model building. Side chains of ligating amino acid residues are built. Position of Ca^{2+} is estimated in the C-terminal cleft and Cl^- is assigned as a ligand to it.

Methodology

The D1 subunit of *Chaetosphaeridium globosum* (TREMBL ID Q8M9W3, gene name PSBA, molecular mass 38.849 kDa and sequence length 352 amino acids) was modelled by using *Thermoynechococcus vulcanus* as a reference template (PDB ID 1izl, gene bank ID: NP681245), whose crystal structure is known at 3.7 Å resolution[30]. Alignment of sequences is done by BLAST[31]. The 3D structure of D1 protein of *C. globosum* is generated by the MODELLER program[32]. This approach uses the method of satisfaction of spatial restraints for model building with the structure from the structural database as the template and the amino acid sequence of the studied protein as the target. The missing loops 88-91 and 241-250 in the template are inserted into the model using LOOP module of the MODELLER[32]. The coordinates of four Mn atoms were extracted from 1izl and superimposed on the active site of the model.

Generated model is evaluated using the program PROCHECK[33]. The coordination sphere of Mn cluster is determined by locating ligand forming atoms within a radius of 3.5 Å from the central metal atom by an in-house program PDBENV. Visualization of model and measurement of distances are done with Insight II (Accelrys), RasMol (Roger Sayle, Glaxo Wellcome Research and Development, Stevenage, Hertfordshire, U.K.), and also MODELYN (provided by C N Mandal, Indian Institute of Chemical Biology IICB, Kolkata).

Results and Discussion

The sequence alignment of our quarry sequence with D1 of 1izl is shown in Fig. 1. The BLAST result shows 83% identities. The backbone conformation in form of Ramachandran plot (figure not shown) generated by PROCHEK showed 73.8% of the amino acid residues located in the most favoured regions, 17.1%

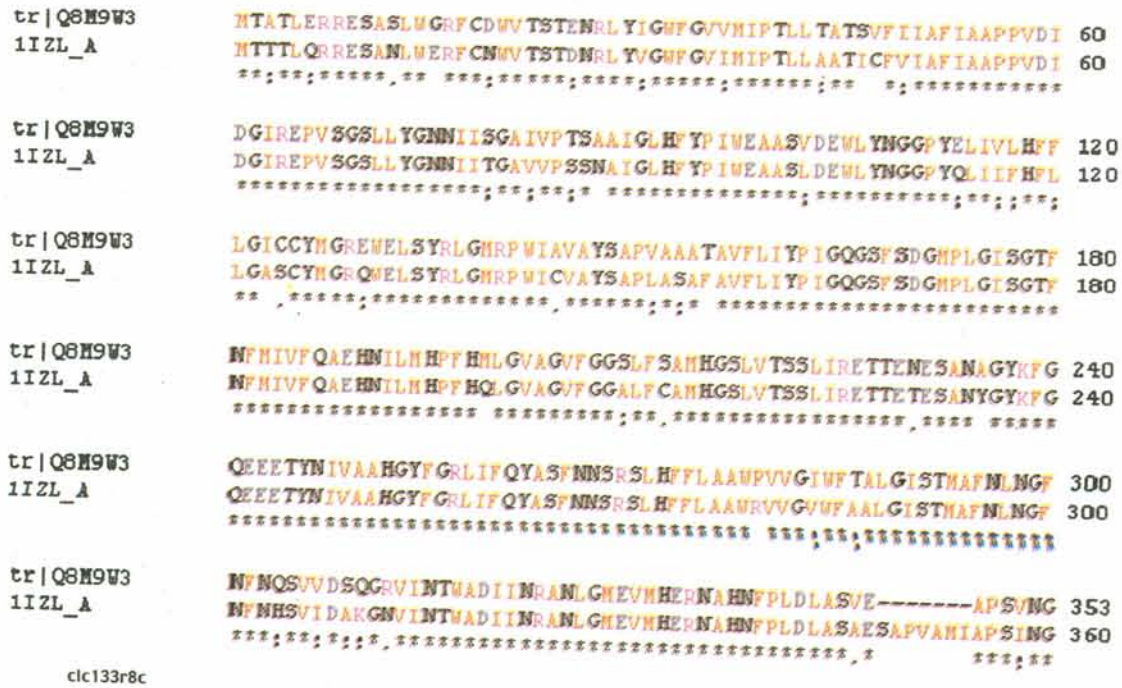

Fig.1—Sequence alignment of D1 protein of *Thermosynechococcus elongatus* (PDB ID 1izl) with the D1 protein of *Chaetosphaeridium globosum* [The x-ray crystallographically determined structures do not contain segments 1-31, 88-91, 241-250, and 345-353. Post-translational modification hydrolyzes peptide linkage between 344 and 345 residues. In the bottom line of alignment, the similar residues are marked by '*', very similar residues by ':', semi conserved residues by '.' and no marks indicate different amino acid residues with respect to the template. Different colours of amino acid residues are as per the following scheme: AVFPMILW, red, small (small+hydrophobic (include aromatic-Y)); DE, blue, acidic; RK, magenta, basic; STYHCNGQ, dark green, hyderoxyl+amine+basic-Q; others, grey]

of the residues in additionally allowed regions, 4.7% in generously allowed regions and only 4.4% remain in disallowed regions.

The 3D structure of our model is compared with the X-ray structure of D1 protein of 1izl. In the crystal structure of 1izl of D1, the following segments of sequence are missing 1-31, 88-91, and 241-250. The comparison by structural superposition gave RMS value of 0.6568 for Cα atoms, suggesting a very good structural homology (Fig. 2). Based on the results of PROCHECK the built model structure could be characterized as a good structure. The secondary structure of the modeled protein is shown in Fig. 3.

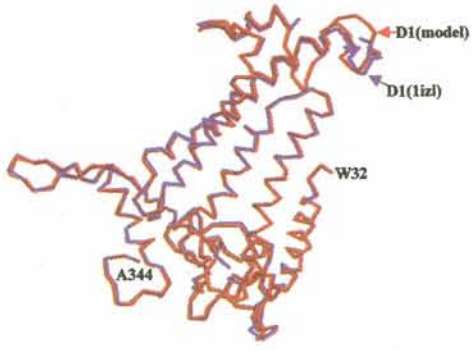

Fig. 2—The superposition of Cα chain of D1 protein of 1izl with the model D1 protein of *C. globosum* using MODELLER[32] [(RMS=0.6568). The backbone structures are drawn using RasMol]

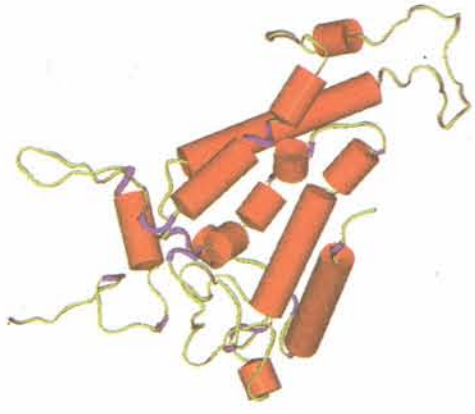

Fig. 3—The secondary structure of *C. globosum* after MODELLER run [Cylinders represent α-helices and arrows β-strands. Graphics are generated using Insight II]

In crystal structure of 1izl[30], the distances between Mn1-Mn2 and Mn2-Mn3 are 2.692 and 2.710 Å. These distances suggest existence of di-μ-oxo bridges[34-36]. Mn2-Mn4 distance is 2.749 Å. This may be attributed to the presence of mono-μ-oxo or carboxylato bridges between these atoms. The distance between Mn1-Mn4 is 3.348 Å, which may be attributed to the presence of di-μ-aqua bridges. The aqua bridges may be the substrate water molecules undergoing photo-oxidation to molecular oxygen.

In our model of TMC, we measured the proximity of either oxygen or nitrogen atoms of the amino acid residues (Figs. 4, 5). We found that OD1(D170) is at 3.004 Å from Mn3, OE1(E333) at 3.107 Å from Mn4, OE2(E333) at 3.326 Å from Mn2, OE2(E189) at 2.519 Å from Mn4, O(A344) at 2.452 Å from Mn1, NE2(H190) at 3.448 Å from Mn4, NE2(H337) at 3.332 from Mn1. The visual inspection of the 3D model of WOC suggests a cleft formed by C-terminal end, where Ca may reside. ND1 (H337), O(L343) and OXT(A344) may form ligand to Ca^{2+} (Fig. 5). The other ligands to metal clusters may be molecules and oxo bridges. The controversy over Cl^- as a ligand to metal cluster exists[37]. However, in the present model, Cl^- is tentatively assigned to Ca^{2+} as a ligand (Fig. 5). The ligand distances are expected to further decrease to less than 3 Å on energy minimization with metal ions. The study in this direction will be taken up in future.

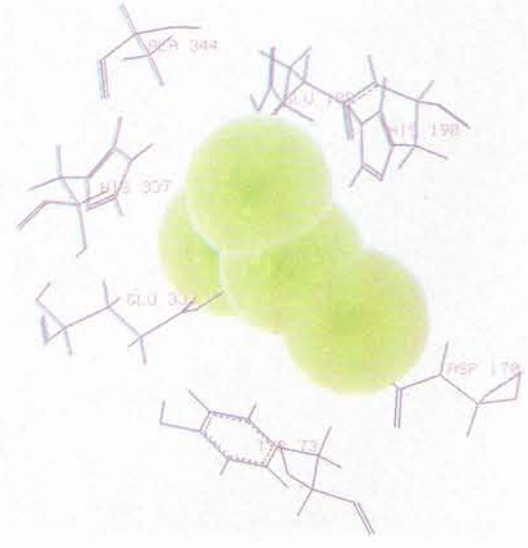

Fig. 4—The Mn-atoms of 1izl superimposed on the active site of D1 protein of *C. globosum* [The neighboring amino acid side chains are depicted in wire frame along with their sequence number, while the Mn cluster is represented by spacefill model. Tyr 73 is too far to form ligand to Mn cluster]

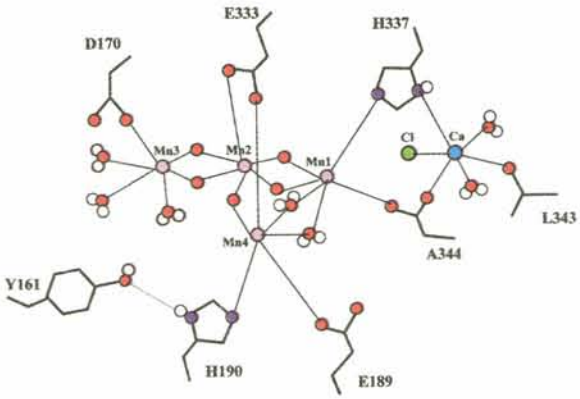

Fig. 5—A schematic diagram of the funnel model of Mn-Ca cluster of D1 protein with probable coordination to ligand atoms [Amino acid residues and their sequence numbers are indicated. The association of cluster with one Cl^- is also shown. Mn, Ca, Cl, O, N, and H are represented as violet, light blue, green, red, deep blue, and white color filled circles, respectively. Broken lines represent H-bonding]

Oxygen atom in phenolic group of Y161 is proposed to be in close proximity (<5 Å) from Mn cluster and acts as H-abstractor in free radical mechanism of water splitting[14]. But, the crystallographic structure as well as the model shows the closest distance of OT (Y161) to be 6.835 Å from Mn2. OT(Y161) is proximal to ND1(H190) (2.840 Å) and may form H-bonding with it. NE2(H190) may form ligand to Mn cluster. Thus, the route of electron transfer from WOC to oxidized special pair chlorophylls of RC II ($P680^+$) may be through H190-Y161 (Fig. 5).

In conclusion, it may be mentioned here that the present resolution and crystallographic goodness of fit (R factor = 0.53), compared to earlier structure of *Synechococcus elongatus*[8] (R factor = 0.59) is certainly better, yet the detailed structure of TMC in terms of ligands, oxo-bridges, and binding H_2O cannot be resolved due to incomplete information and conclusions at atomic dimensions cannot be drawn, without resorting to molecular modelling. However, until a better resolution structure is available, the present theoretical model of TMC may be useful in understanding the mechanism of photosynthetic oxidation of water.

Acknowledgement

LSR is thankful to Profs. C Mitra of Hydrabad Central University, India, and A. Mitra of Bioinformatics Research Center, International Institute of Information Technology (IIIT), Hyderabad for discussion and guidance, to GVK BioScience, Hyderabad, India for providing access to Acclerys. We are also grateful to Prof. A Sali, California Institute for Quantitative Biomedical Research, University of California, San Francisco, for permission to use MODELLER6V2. Permission to use the preview version of MODELYN by Dr. C N Mandal, IICB, Kolkata, India is gratefully acknowledged.

References

1 Carrell T G, Tyryshkin A M & Dismukes G C (2002) *J Biol Inorg Chem* 7, 2-22
2 *Nucleonics Week* (2002), McGraw Hill, New York
3 Ruettinger W & Dismukes G C (1997) *Chem Rev* 97, 1-24
4 Rath L S & Raval M K (2001) *Indian J Biochem Biophys* 38, 75-77
5 Barber J (2002) *Curr Opin Struc Biol* 12, 523-30
6 Dismukes G C, Klimov V V, Baranov S V, Kozlov Y N, Dasgupta J & Tryrshkin A (2001) *Proc Natl Acad Sci (USA)* 98, 2170-75
7 Ananyev G, Zaltsman L, Vasko C & Dismukes G C (2001) *Biochim Biophys Acta* 1503, 52-68
8 Zouni A, Witt H-T, Kern J, Fromme P, Krauss N, Saenger W & Orth P (2001) *Nature (London)* 409, 739-43
9 Sauer K, Guiles R D, McDermott A E, Cole J L, Yachandra V Y, Zimmermann J L, Klein M P, Dexheimer S L & Britt R D (1988) *Chem Scripta* 28, 87-91
10 Hansson O, Aasa R & Vanngard T (1987) *Biophys J* 51, 825-32
11 Guiles R D, Zimmermann J L, McDermott A E, Yachandra V K & Cole J L (1990) *Biochemistry* 29, 471-85
12 Britt RD (1996) in *Oxygenic Photosynthesis: The light reactions* (D R Ort & C F Yocum, eds.), pp 137-59, Kluwer Academic Press, Dordrecht
13 Gilchrist M L, Ball J A, Randall D W, Britt R D (1995) *Proc Natl Acad Sci (USA)* 92, 9545-49
14 Tommos C, Tang X S, Warncke K, Hoganson C W, Styring S, McCracken J, Diner B A & Babcock G T (1995) *J Am Chem Soc* 117, 10325-35
15 Boerner R J, Nguyen A P, Barry B A & Debus R J (1992) *Biochemistry* 31, 6660-6672
16 Nixon P J, Chisholm D A & Diner B A (1992) in *Plant Protein Engineering* (P R Shewry & S Gutteridge, eds.), Vol. 2, pp 93-141, Cambridge University Press, Cambridge
17 Vermaas W F J, Styring S, Schroder W P & Andersson B (1993) *Photosynth Res* 38, 249-63
18 Chu H A, Nguyen A P & Debus R J (1995a) *Biochemistry* 34, 5839-58
19 Chu H A, Nguyen A P & Debus R J (1995b) *Biochemistry* 34, 5859-5882
20 Debus R J, Campbell K A, Gregor W, Li Z-L, Burnap R L & Britt R D (2001) *Biochemistry* 40, 3690-3699
21 Hays A-M A, Vassilier I R, Golbeck J H & Debus R J (1998) *Biochemistry* 37, 11352-11365
22 Debus R J (1992) *Biochem Biophys Acta* 1102, 269-352

23 Baranov S V, Ananyev G M, Klimov V V & Dismukes G C (2000) *Biochemistry* 39, 6060-6065
24 Yachandra V K, Sauer K & Klein M P (1996) *Chem Rev* 96, 2927-2950
25 Dau H, Andrews J, Yachandra V K, Roelofs T R, Latimer M J, Liang W, Sauer K & Klein M P (1995) *Biochemistry* 34, 5274-5287
26 Zhang M, Khangulov S V, Dismukes G C & Barynin V V (1994) *Inorg Chem* 33, 382
27 Zhang M & Dismukes G C (1996) *Inorg Chem* 35, 3307-19
28 Law N A, Kampf J W & Pecoraro V L (2000) *Inorg Chim Acta* 297, 252-64
29 Cinco R M, Rompel A, Visser H, Aromi G, Christou G, Sauer K, Klein M P & Yachandra V K (1999) *Inorg Chem* 38, 5988-98
30 Kamiya N & Shen J-R (2003) *Proc Natl Acad Sci (USA)* 100, 98-103
31 Altschul S F, Gish W, Miller W, Myer E W & Lipman D J (1990) *J Mol Biol* 215, 403-410
32 Sali A & Blundell T L (1993) *J Mol Biol* 234, 779-815
33 Laskowski R A, MacArthur M W, Moss D S & Thrnton J M (1993) *J Appl Crystallogr* 26, 283-291
34 Robblee J H, Cinco RM, Yachandra V K (2001) *Biochim Biophys Acta* 1503, 7-23
35 Yachandra V K, Sauer K & Klein M P (1996) *Chem Rev* 96, 2927-50
36 DeRose V J, Mukerji I, Latimer M J, Yachandra V K, Sauer K & Klein M P (1994) *J Am Chem Soc* 116, 5239-49
37 Hasegawa K, Kimura Y, Ono T A (2004) *Biophys J* 86, 1042-50